UNIVERSITY OF COPENHAGEN FACULTY OF SCIENCE

PhD Dissertation

A LIS. SIGIPTINA A SISUAL AND A SIGNAL

Analysis of synchrotron X-ray tomographic data Reconstruction and application





By Leise Borg

Academic advisor Jon Sporring

UNIVERSITY OF COPENHAGEN August 2018

Abstract

Synchrotron X-ray tomography allows to non-destructively image the interior of three-dimensional objects at a very high resolution. Users of synchrotron facilities depend on reliable reconstructions (images) of the specimen they brought, and subsequent analysis of the images. My dissertation addresses both of these analysis steps through mathematical modeling and specific approaches to practical solutions.

Many in situ X-ray CT experiments entail rigs to control the in situ environment. In some cases, the X-ray beam is fully or partly occluded by the rig leading to "polluted" or missing data for some projection angles. This may lead to artifacts in the tomographic reconstructions, which complicate the subsequent image analysis. We have analysed a specific tomographic experiment including a percolation cell for controlling fluid flow in chalk samples. The images provided by the synchrotron facilities suffered from severe reconstruction artifacts due to occluded projections. Our goal was to find methods to suppress the artifacts and to generalize our analyses to make them applicable to a wider audience. We restricted ourselves to consider artifact reduction methods only in conjunction with the filtered backprojection algorithm because, with these large amounts of data, the use of simple and fast algorithms is crucial. The methods we proposed removed or suppressed the artifacts successfully. The work was followed by a theoretical characterization of all types of incomplete data artifacts and proofs of the advantages of including a smooth cutoff across singularities in incomplete data.

Establishing image analysis in X-ray synchrotron tomographic images was the other focus of this work. Synchrotron tomography is widely used in medicine and was established many years ago. However, specialized methods are required to fully exploit the potentials of the three-dimensional (3D) nature of the images. We analysed 3D images of muscle biopsies by measuring the muscle fiber morphology. The biopsies were from healthy participants and participants with one of three different types of neuromuscular diseases. Generally, these types of measures are carried out by light microscopy in two-dimensional (2D) images, and we showed that 3D image analysis may be preferred in specific situations. Further, we provided a measure of the orientation consistencies of the muscle fibers, which is a measure of how aligned or organized the fibers are related to each other. We showed that participants with cerebral palsy exhibited a lower degree of muscle fiber orientation consistency than the healthy participants did.

One of the main challenges in synchrotron data is the amount of data. Conversely, it is also the main advantage. The degree of detail and the 3D nature provided by the modality is what makes it superior to many other imaging modalities. We focused on methods that were able to handle these large amounts of data and at the same time were able to extract important details. The main contributions of this thesis can be seen as an approach towards better understanding the implications of synchrotron X-ray tomographic data.

Preface

This Ph.D. thesis is the result of my work at the Image Section, Department of Computer Science, University of Copenhagen from December 2014 to August 2018. My supervisors were François Lauze and Jon Sporring.

The first part of my work was carried out for the P^3 project, which is an inter-institutional collaboration between the University of Copenhagen, Maersk Oil and Gas, and Innovation Fund Denmark. During this project, I worked 6 months in Per Christian Hansens research group in Department of Applied Mathematics and Computer Science, DTU Compute, Technical University of Denmark, as a part of my change of scientific environment.

The second part of my work was carried out for the MAX4Imagers project, which is a collaboration between researchers from the newly opened synchrotron, MAX IV in Lund, Sweden, the Capital Region hospitals, University of Copenhagen, and the Technical University of Denmark. It is funded by the Capital Region Research Foundation for Healthcare.

My main research outputs are the three following articles on which this thesis is based.

- L. Borg, J. S. Jørgensen, J. Frikel, and J. Sporring. Reduction of variable-truncation artifacts from beam occlusion during in situ x-ray tomography. *Measurement Science* and Technology, 28(12), 2017. [1]
- L. Borg, J. S. Jørgensen, J. Frikel and E. T. Quinto. Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data [2]. Under review at the Society for Industrial and Applied Mathematics (SIAM) Journal on Imaging Sciences.
- 3. L. Borg, J. Sporring, V. A. Dahl, A. B. Dahl, E. B. Dam, and J. Pingel. Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography [3]. *Submitted to Computers in Biology and Medicine*.

The inclusion of all three articles in this dissertation is permitted. In addition to the articles, I published a technical report [4], which was the preliminary work of our article [1]. Moreover, I attended conferences and contributed to a number of poster presentations.

The thesis is written in such a way that a Ph.D. student in computer science should be able to follow. Therefore, a soft introduction to computed tomography is presented in Section 2. Further, our article [2] in Section 4.1 comes along with an introductory description and some figures for visualizing the geometries and some of the main points. This should support the readability of the article for non-mathematician readers.

The three articles are incorporated in the thesis as they are and as their own entity, with separate sectioning systems and paginations. Therefore, there are several cases of the same section numbers and page numbers. When I refer to sections, the referred section number is related to the thesis, as given in the table of contents. Otherwise, when referring to sections in the articles, it is always stated which article I refer to.

Acknowledgements

There are so many people I would like to thank, and I almost do not know where to start! From the very beginning of my work in academia, I should soon find out that everyone I met was helpful, curious, open-minded, always open for collaboration, and (often) very, very sweet.

My journey started when professor Mads Nielsen gave me the opportunity of pursuing my dream to work in academia. Computer science it was - and what a world! I have tried to put words on my gratitude, but it never really felt sufficient, so now, in writing, I hope I will settle: Thank you so much, Mads - you are the greatest! I also want to thank my two supervisors, François Lauze and Jon Sporring, who supported me through mountain-like academic obstacles. Already soon in my project, Jon established a study group, which gave me a superb foundation for the subsequent steps I was to take within computed tomography. Your distinct ability to hunt details has been invaluable through the years of our collaboration. The work during my stay at Technical University of Denmark (DTU), was supervised by Jakob Sauer Jørgensen. I am truly thankful for all that you have done for me. You are one of the most thorough, patient, and clever persons I have ever met. I learned so much from you, and on top of that, you were really a friend during difficult times. Erik Bjørnager Dam, you brought in great advice, a reliable help, and super-positive energy in the last period of my Ph.D., for which I am so grateful.

I want to thank E. Todd Quinto and Jürgen Frikel for being really cool collaboration partners, and for our friendship that hopefully will stand many years from now. I also want to thank Per Christian Hansen for welcoming me in his group at DTU, for his super courses and lively communication skills, Kim Steenstrup Pedersen for being a he-man, Vedrana Andersen Dahl for sharing cool stuff with me, and for being the sweetest of all, Anders Bjorholm Dahl for his support, Tim Dyrby for his ever-contagious enthusiasm, Hans Martin Kjer for being so helpful, and all my colleagues in the Image Section, who were always up for good discussions.

Last but not least, I want to thank my family. Peter, you have supported me like no one else. During stormy weather and sunny days, I knew I could *always* count on you. Thank you for making me feel that I am the luckiest woman in the world for having you in my life. Our cute, brave and sweet kids, Villads and Birk, also deserve a big thank you. Despite your young ages you managed to give me space during the last period of work-marathon - I know it has been hard on you. You even cheered me up: "you are doing good, mom", "your thesis looks awesome". You, little guys, helped me keep my feet on the ground and brought in small sparkles of complete happiness in the everyday life of a Ph.D. student. Dear family: thank you!

Contents

Al	bstract	ii					
Pr	reface	iii					
A	cknowledgements	\mathbf{iv}					
1	Introduction 1.1 Aims of the thesis 1.2 Structure of the thesis	1 . 2 . 2					
2	Computed tomography in short2.1Data acquisition2.2Reconstruction methods2.3Practical aspects	4 . 5 . 7 . 9					
3	 3 Variable-truncation artifacts 3.1 Article # 1: Reduction of variable-truncation artifacts from beam occlusion during in situ X-ray tomography 3.2 Supplementary material 3.3 Summary 						
4	 Characterizing artifacts from incomplete data 4.1 Article # 2: Characterizing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data	34 . 37 . 66					
5	Image analysis of structures in 3D, part 1 5.1 Data	67 . 67 . 69 . 69 . 75 . 76					
6	 Image analysis of structures in 3D, part 2 6.1 Article # 3: Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography 6.2 Supplementary material 6.3 Summary 	77 . 77 . 88 . 96					
7	Conclusions and perspectives 97						

Introduction

X-ray computed tomography (CT) is an imaging technique providing the possibility of seeing the *inside* of an object. For data acquisition, the object is irradiated by X-rays at different angles, and the transmitted beam intensity is detected. Tomographic reconstruction entails methods to convert these raw CT measurements into images of the objects. To take full advantage of the potential of such three-dimensional (3D) images, subsequent image analysis must be carried out. Users of synchrotron facilities bring samples to be imaged and expertise in their own area of science, and often depend on both tomographic reconstruction and subsequent analysis of the images. My dissertation addresses both of these analysis steps through mathematical modeling and specific approaches to practical solutions.

Acquisition of synchrotron X-ray data is where extremities meet; tiny particles (electrons) are accelerated in gigantic storage rings by strong magnetic fields to reach near-relativistic speeds. X-rays are produced by collisions between the particles and matter when quantum states are changed in the atoms. An image of the European Synchrotron Radiation Facility located in Grenoble in France is shown in Figure 1.1.



Figure The European Synchrotron Radiation Facility Grenoble, 1.1:inimage France. The is copied from https://www.ngicorp.com/portfolio/ esrf-european-synchrotron-radiation-facility/

"Big data" is a concept associated with extremely large and complex datsets. The definition has changed over the time and according to [5] one definition of big data is based on volume, variety, and velocity:

- 1. Volume refers to the magnitude of data and sizes are reported in multiple terabytes and petabytes.
- 2. Variety refers to the structural heterogeneity in a dataset. Text, images, audio, and video are examples of unstructured data, which sometimes lack the structural organization required by machines for analysis.
- 3. Velocity refers to the rate at which data are recorded.

Synchrotrons are big, and they definitely produce big data according to the definition just given and according to authors of [6]. Big data requires processing methods suitable for handling the inherent level of detail and complexity of the specific problem, which leads us to the aims of this thesis.

1.1 Aims of the thesis

The aim of this thesis is to add to the applications in the analysis of synchrotron X-ray tomographic data. Due to a large amount of data and the complexity often inherent in tomographic data, important factors for the methods are well-understood behavior, robustness, and computational efficiency. The aims fall into two categories: Tomographic reconstruction and image analysis. In the category of tomographic reconstruction, I worked on two separate projects: a practical and an analytic approach for artifact reduction.

Typically, synchrotron facilities provide reconstructions based on the data measured at the site. According to the variety of experimental setups and in specimens to be imaged, the reconstruction methods must be rather generic. This means that a high reconstruction quality is not guaranteed for more specialized imaging experiments.

In the P^3 project, we had raw synchrotron CT measurements and corresponding reconstructions available from the SPring-8 facilities in Japan. However, the experimental setup was highly specialized, and the reconstructions provided by SPring-8 contained streak artifacts, which initiated the first project aim; Our aim was to develop a method to obtain reconstructions without streak artifacts from the raw data. At the same time, the method was to be able to handle the large amounts of data from the synchrotron and needed to be relatively fast as the real-time analysis was crucial.

This work led to new ideas for a more theoretical approach for characterizing artifacts from incomplete data. The aim was to use formulations from microlocal analysis to characterize not only streaks similar to those in the P^3 project, but also for artifacts from arbitrary incomplete datasets.

In the last study, we analysed 3D structures in images from synchrotron X-ray CT of muscle biopsies. This was a unique dataset in many ways; Muscle biopsies are usually not imaged in three dimensions, and transmission synchrotron X-ray CT has to our knowledge never before been used for analysing muscle fibers. The aim was to apply suitable algorithms to explore 3D features of the muscle fibers. Our approach was segmentation followed by 3D morphologic analyses and measuring the muscle fiber orientations with respect to each other.

1.2 Structure of the thesis

Section 2 covers a short introduction to CT; historical perspectives, mathematical models of the data acquisition and reconstruction methods, and lastly, practical aspects. The following four sections, Section 3, 4, 5, and 6, represent the four main works of the thesis. The main structure of these sections is such that each section contains an article, supplementary material, and a summary (except Section 5 which contains yet unsubmitted work and no supplementary material). The summaries are extractions of the results, discussions, conclusions, and outlook of the different works. The supplementary material in Section 4 is part of the introduction of the section and thus does not have its own section.

The work conducted relevant for the P^3 project is presented in Section 3, and is an analysis of variable-truncation artifacts. Section 3.1 contains the article *Reduction of variabletruncation artifacts from beam occlusion during in situ x-ray tomography.* The article is followed by supplementary material: a short description of how the sinogram is linked to the raw data is presented in Section 3.2.1, in Section 3.2.2 experiments on synthetic data serve to give an intuition about the nature of the truncated-projection artifacts, Section 3.2.3 discusses an algebraic solution for the truncated-projection problem, and Section 3.3 is a summary of Section 3.

The theoretical approach for characterizing artifacts from arbitrary incomplete data is laid out in Section 4. It has an introductory part with figures for visualizing the geometries of the mathematical descriptions and the most important points. Section 4.1 contains the article Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data and Section 4.2 is a summary for the analyses in Section 4.

The initial set of analyses on the images of muscle biopsies were based on a Chan-Vese segmentation [7] followed by a measure of the anisotropy in the 3D volumetric segmentations. A similar study was carried out on rat muscle biopsies [8], with which we wanted to compare our results for the humane dataset. These analyses are presented in Section 5.1 to Section 5.4 and will be submitted for publication as part of a future article. Section 5.5 is a summary of Section 5. Section 6 is a direct follower of the work presented in Section 5. A snake method [9] enabled analyses on each single muscle fibers, which led to the work presented in the article *Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography*, and is contained in Section 6.1. Supplementary material is presented in Section 6.2.2 contains supplementary discussions not given in the article. Section 6.3 is a summary of Section 6. Section 7 draws conclusions and sets the four works described in Chapters 3, 4, 5, and 6 into perspective.

Computed tomography in short

The very first steps towards transmission tomography date back to the 19'th and 20'th centuries. The German physicist Wilhelm Conrad Röntgen discovered the X-rays in 1895 and won a Nobel Prize for the realization of X-ray imaging in 1901 [10]. Figure 2.1 shows one of the first X-ray images ever taken. It shows the hand of Anna Bertha Röntgen, who was the wife of Wilhelm Conrad Röntgen.



Figure 2.1: One of the first X-ray images in history: The hand of Anna Bertha Röntgen with a ring on the fourth finger.

Later, in the year of 1917, the Austrian mathematician Johan Radon introduced the Radon transform [11], which defines a projection of a two-dimensional object, f(x, y), by line integrals of the object. The natural next step was the definition of the inverse Radon transform, implying that f(x, y) can be exactly reconstructed from the complete set of all possible line integrals of the object. It took more than 40 years before the physicist, Allan MacLeod Cormack, in 1963 applied the solutions to the inverse problem, to X-ray images. In 1971, Hounsfield built a prototype head scanner, which was the introduction of CT in medical practice. Hounsfield and Cormack were awarded the Nobel Prize in Physiology in 1979 for their work on X-ray CT. Over the years, CT application has expanded to cover a range of applications, from medical imaging to spectroscopic analysis, materials science, and geophysics, among others. The first types of CT scanners required long scanning times and high radiation dose but yet provided images of relatively low resolution. While the first generation of scanners used a parallel-beam configuration, subsequent scanner generations now use 2D detectors for parallel-beam configuration, fan-beam configuration, circular

cone-beam configuration, and helical cone-beam configuration providing substantially higher temporal and spatial resolution, while requiring a lower radiation dose. For details on the beam configurations refer to [12] or [13]. While modern lab tomography provides images of high quality, synchrotron tomography takes the quality and possibilities to completely new levels.

The tunable wavelengths for monochromaticity, high intensity, and penetration power allows for investigation of a wide range of specimens in various depths and resolutions. The high emitted flux and brilliance increase the signal-to-noise ratio and the resolution. On top of this, a high coherence of the beam, which is the property that enables the emitted radiation to produce wave-like phenomena, allows for diffraction used for phase contrast imaging [14, 15]. The most recent synchrotron sources are "fully coherent" down to wavelengths in the ultraviolet and soft-X-ray range [16]. In 2017 [15] reached a resolution of ~ 20 nm resolution, by magnifying the projected image with a Fresnel zone plate and enhancing the contrast with a Zernike type phase ring. This is absolutely amazing! Not only can we see the inside of a specimen, but we can also go down to a 20-nanometer resolution!

As history has shown, CT is subject to continuous research and development which has led the way to spatial and temporal resolution never seen before, capturing changes down to tens of nanometers and sub-nanoseconds [17]. The MAX IV Laboratory in Lund is currently the synchrotron X-ray source with the beam of highest brilliance (2016) [18].

There exist many variations of X-ray synchrotron imaging, such as X-ray holography, magnetic Compton scattering, resonant X-ray scattering, high-energy fluorescent X-ray, ultra-small angle scattering, just to mention some. In this thesis, we consider only transmission X-ray synchrotron imaging with a monochromatic beam in a parallel-beam configuration.

In the following sections, mathematical models of the data acquisition and how to obtain an image from this data are presented, followed by practical aspects in CT.

2.1 Data acquisition

Different types of materials attenuate different amounts of X-rays according to the photonmatter interaction. This property is exploited in the transmission X-ray CT-imaging technique, where an object is irradiated by X-rays and the amount of transmitted X-rays is measured by a detector. The amount of X-rays measured by the detector carries information about the attenuation coefficients inside the material, but this is only the integrated information from that specific X-ray penetrating the entire object from one specific angle. To refine the information of the attenuation coefficients, the object must be irradiated from different angles and by a number of X-rays covering the entire object. Gathering all this information with proper amounts of radiation angles and beam-detector pairs covering the entire object makes it possible to obtain a 3D image of the object. The pixel values in the image are directly correlated to the physical X-ray attenuation coefficient for that specific material.

The Radon transform [11], R, is a mathematical model widely used in CT, which integrates a function, f, over the straight line, L. In CT, it is defined by the line integrals of the attenuation coefficients through an object, which is essentially the negative logarithm of the ratio of the incoming and outgoing number of photons, as we will see in the following. The Radon transform is closely related to the Lambert-Beer model [19, 20, 21], which describes the behavior of the photon attenuation when a beam is penetrating an object. Following derivations can be found in any introductory book on CT, such as in [12], on which these derivations are based. In inhomogeneous objects, the attenuation coefficients, f, are spatially variant, as sketched in Figure 2.2. In the figure, I_0 is the incoming beam and I_n is the beam measured in the detector after having traveled the distance t in step lengths of Δt . The thought behind the derivation is that we allow the attenuation coefficients, f, to vary as indicated in Figure 2.2 and then let the width Δt go towards zero.



Figure 2.2: Lambert-Beer's law of attenuation.

The intensity of the beam after passing through the first portion of the object, Δt , is given by

$$I_1 = I_0 - f_1 I_0 \Delta t = I_0 (1 - f_1 \Delta t).$$

Likewise, the intensity I_2 is

$$I_2 = I_1(1 - f_2\Delta t) = I_0(1 - f_1\Delta t)(1 - f_2\Delta t)$$

The intensity after having passed $n \cdot \Delta t$ is

$$I_n = I_0(1 - f_1\Delta t)(1 - f_2\Delta t)\cdots(1 - f_n\Delta t)$$

The Taylor expansion to first order of the exponential e^{-x} around 0 is

$$e^{-x} \approx 1 - x, \tag{2.1}$$

so the terms in the parantheses can be expressed as a Taylor expansion

$$I_n \approx I_0 \mathrm{e}^{-\sum\limits_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{f}_{\mathrm{i}} \Delta \mathrm{t}}.$$

If the discretization of the object is very fine, i.e. Δt is very small, $f_i \Delta t$ is close to zero, and the equality holds:

$$I_n = \lim_{f_i \Delta t \to 0} I_0 e^{-\sum_{i=1}^n f_i \Delta t}$$

When Δt is small, we can convert the summation into an integral,

$$\lim_{\Delta t \to 0} I_0 e^{-\sum\limits_{i=1}^n f_i \Delta t} = I_0 e^{-\int \limits_0^t d f(t) dt}$$

where t is the position at the detector, and finally we have

$$I_n = I_0 e^{-\int_0^{t_d} f(t)dt},$$
 (2.2)

known as Lambert-Beer's law of attenuation. By rearranging equation 2.2, we define the projection integral

$$-\ln\frac{I_n}{I_0} = \int_0^{t_d} f(t) dt,$$
 (2.3)

which is the Radon transform. We will use following notation for the Radon transform

$$Rf(\varphi, p) = -\ln \frac{I_n}{I_0} = \int_{L(\varphi, p)} f(p) dp, \qquad (2.4)$$

where φ is the rotation angle, p is the displacement of the detector or the signed distance from origo, parameterized as

$$p = x_1 \cos \varphi + x_2 \sin \varphi$$

and L is the sets of points $\mathbf{x} = (x_1, x_2)$:

$$L(\varphi, p) = \{ \mathbf{x} \in \mathbb{R}^2 : x_1 \cos \varphi + x_2 \sin \varphi = p \}.$$
 (2.5)

Figure 2.3 visualizes the geometry of the set-up and the related the variables. The dotted lines are the projection lines, $L(\varphi_0, p)$.



Figure 2.3: The Radon transform showing one projection for the angle φ_0 .

 $s_{\varphi_0}(p)$ is denoted as the projection obtained from the angle, φ_0 , and variable p:

$$s_{\varphi_0}(p) = Rf(\varphi_0, p).$$

While acquiring the data, the source-detector system rotates by a small angle to obtain data from different angles until 180 degrees is covered.

2.2 Reconstruction methods

Typical reconstruction methods fall into two major categories: Fourier-based and algebraicbased. The approaches are quite different as Fourier-based methods are based on continuous formulations and algebraic methods are based on a discretized formulation. The following is a very short presentation of the two types of methods. One can read about the methods in any introductory book on the mathematics behind CT, such as [12], and these descriptions are based on [12, 22, 23, 24]. **Fourier-based reconstruction** There exist a number of Radon inversion formulas, where the most common is the filtered backprojection (FBP). The Radon transform and FBP are closely related to the Fourier transform. The Fourier slice theorem states that the 1D Fourier transform of a projection of an object is equal to the slice parallel to the projection line taken from the 2D Fourier transform of the object. This basically means that the object can be reconstructed from its projections (the Radon transform). I.e. the reconstruction, \tilde{f} , is calculated

$$\tilde{f}(x_1, x_2) = \int_0^\pi \int_{-\infty}^{-\infty} \mathcal{F}_1 s_{\varphi}(\omega) |\omega| e^{2\pi i p \omega} d\omega d\varphi,$$

where

$$\mathcal{F}_1 s_{\varphi}(\omega) = \int_{-\infty}^{-\infty} s_{\varphi}(p) e^{-2\pi i p \omega} dp$$

is the 1D Fourier transform of a projection $s_{\varphi}(\omega)$ and ω is the frequency. $|\omega|$ is a ramp filter that emphasizes the high-frequency components of the signal and it derives from the change of Cartesian coordinates to polar coordinates. In practice, a low-pass filter is added for deemphasizing the high frequencies in the reconstruction. The intuition behind this method is to "smear" the projections back along the line they came from, which can be captured in one word: backprojection. All backprojections are added and the result is normalized by the number of projections to obtain the reconstruction. Prior to backprojection, one needs to filter the projections for covering change of coordinate systems and to de-emphasize high-frequency components. So, the name "filtered backprojection" covers the act quite well.

Algebraic reconstruction The algebraic reconstruction techniques represent the reconstruction problem as a linear system of equations in a matrix-vector form,

$$A\tilde{f} = b$$

A is the system matrix, representing the weights from the projection lines through the pixels of the discretized object, f, consisting of the elements

$$a_{i,j} = \frac{\text{illuminated area of pixel } j \text{ by ray } i}{\text{total area of pixel } j}.$$

b is the measurements, and \tilde{f} is the solution to the inverse problem. In practice, iterative solutions are typically preferred according to the relatively large sizes of CT problems. Two classes of iterative methods commonly used are the algebraic reconstruction techniques (ART) and the simultaneous iterative reconstruction techniques (SIRT) [22]. The reconstruction problem may be formulated as the minimization of a functional, where the goal is to find a solution \tilde{f} that minimizes the distance between $A\tilde{f}$ and b. Prior knowledge about noise, missing data, etc. may be incorporated in the formulations (but will not be covered here). The basic ART method use the updating rule:

$$\tilde{f}^{(k,i)} = \tilde{f}^{(k-1,i)} + \lambda \frac{b_i - a_i^T \tilde{f}^{(k-1,i)}}{||a_i||_2^2} a_i, k = 1, 2, ..., i = 1, 2, ..., M,$$

where k is the iteration step number, λ is the relaxation parameter, a_i^T is the *i*'th row of A, and M is the total number of rows in A. The SIRT methods replace the one-by-one application of rows in ART by a single simultaneous iteration step, by accessing the rows simultaneously and computing the next iteration vector:

$$\tilde{f}^{(k)} = \tilde{f}^{(k-1)} + \lambda T_1 A^T T_2 (b - A u^{(k)})$$

One example of SIRT is the Cimmino method for which

$$T_1 = I$$
 and $T_2 = \frac{1}{m} \text{diag}(\frac{1}{||a_i||_2^2}),$

and m is the total number of rows in A.

2.3 Practical aspects

The imaging technique and reconstruction methods described in the previous sections were from a completely theoretical perspective. In practice, when working with real acquisition setups and real data, however, one must consider additional aspects. These aspects are abundant and include source and detector conditions, photon starvation, detector element cross-talk, scattering, and conditions in acquisition setups leading to different types of limited data.

These practical issues all lead to reconstruction artifacts which differ in nature according to their cause. Artifacts include ring artifacts, beam hardening artifacts, misalignment artifacts, metal artifacts, motion artifacts and artifacts related to limited data, among others. Improving the acquisition hardware or post-processing of the acquired data are two common ways to address the issues.

Examples of limited data include limited-angle (LA), region-of-interest (ROI), exterior data, sparse data, and variations of these types of limited data. Acquisition of only a range of angular measurements is referred to as LA. ROI, also called interior tomography is when only a part of the object is covered by the detector panel and exterior data is when the center part of the object is not covered by the detector panel. Sparse data is obtained when only a sparse number of projections is available or the spatial resolution is low. Our article in Section 3.1, address the issues in an example of an in-situ setup, leading to a combination ROI- and LA-data. We refer to the resulting artifacts as variable-truncation artifacts. There exist two different schemes often used when approaching limited-data problems:

- Repair data + Fourier-based reconstruction: Repairing the data may, for instance, be approached by inpainting methods, which fill in the missing pixels values, or by handling the discontinuities in the data. After repairing the data, FBP may be used to obtain the image.
- Use data as it is + algebraic reconstruction with a suitable regularization method: The regularization method can be constructed in such a way as to handle the missing data and to obtain a higher numerical stability. Some examples are total variation [25], downweighing affected sinogram pixels [26], or shearlet-based regularization [27], among others.

It may be possible to incorporate regularization in the integrals in the Fourier-based approach. In practice, though, this is probably less frequent. One of the two schemes must be chosen according to the specific data at hand; Numerical methods may handle issues in a more clever way as they can target the issues more specifically, FBP, however, can handle very large problems in relatively short time.

The P^3 project investigates the possibilities of being able to extract larger amounts of oil than currently possible in oil reservoirs in the Danish sector of the North Sea. The key is to search underground chalk reservoirs. Chalk is up to 40-50 percent full of tiny cavities or pores, in which oil, water, CO_2 or any other type of fluid can be contained [28]. Modeling grain and pore structure in rock samples can help to determine if it is worth going after oil in some of those marginal oil fields. The reconstructions of the chalk provided by the SPring-8 facilities in Japan contained severe artifacts due to a specialized imaging setup. The work outlined in this section serves to reconstruct the images and at the same time suppress artifacts. We restricted ourselves to consider methods in conjunction with the FBP method. Iterative reconstruction methods typically require computing times several orders of magnitude longer than FBP, and with these large amounts of data, the use of simple and fast algorithms is crucial. Section 3.1 contains our article *Reduction of variable-truncation artifacts from beam occlusion during in situ X-ray tomography*, Section 3.2 is supplementary material and Section 3.3 is a discussion of the work.

3.1 Article # 1: Reduction of variable-truncation artifacts from beam occlusion during in situ X-ray tomography

Meas. Sci. Technol. 28 (2017) 124004 (19pp)

Reduction of variable-truncation artifacts from beam occlusion during *in situ* x-ray tomography

Leise Borg¹, Jakob S Jørgensen^{2,4}, Jürgen Frikel³ and Jon Sporring¹

¹ Department of Computer Science, University of Copenhagen, Sigurdsgade 41, 2200 Kbh. N, Denmark

² Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard

Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark

³ Department of Computer Science and Mathematics, OTH Regensburg, Universitätsstraße 31, Regensburg, Germany

E-mail: lebo@di.ku.dk

Received 31 May 2017, revised 14 August 2017 Accepted for publication 13 September 2017 Published 22 November 2017



Abstract

Many in situ x-ray tomography studies require experimental rigs which may partially occlude the beam and cause parts of the projection data to be missing. In a study of fluid flow in porous chalk using a percolation cell with four metal bars drastic streak artifacts arise in the filtered backprojection (FBP) reconstruction at certain orientations. Projections with non-trivial variable truncation caused by the metal bars are the source of these variable-truncation artifacts. To understand the artifacts a mathematical model of variable-truncation data as a function of metal bar radius and distance to sample is derived and verified numerically and with experimental data. The model accurately describes the arising variable-truncation artifacts across simulated variations of the experimental setup. Three variable-truncation artifact-reduction methods are proposed, all aimed at addressing sinogram discontinuities that are shown to be the source of the streaks. The 'reduction to limited angle' (RLA) method simply keeps only non-truncated projections; the 'detector-directed smoothing' (DDS) method smooths the discontinuities; while the 'reflexive boundary condition' (RBC) method enforces a zero derivative at the discontinuities. Experimental results using both simulated and real data show that the proposed methods effectively reduce variable-truncation artifacts. The RBC method is found to provide the best artifact reduction and preservation of image features using both visual and quantitative assessment. The analysis and artifact-reduction methods are designed in context of FBP reconstruction motivated by computational efficiency practical for large, real synchrotron data. While a specific variable-truncation case is considered, the proposed methods can be applied to general data cut-offs arising in different in situ x-ray tomography experiments.

Keywords: computed tomography, *in situ* measurements, filtered backprojection, sinogram discontinuities, variable-truncation artifacts

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. In situ x-ray tomography

X-ray computed tomography (CT) has been an indispensable non-invasive imaging technique in medical imaging, ⁴ Present address: School of Mathematics, University of Manchester, M13 9PL, United Kingdom.

1361-6501/17/124004+19\$33.00

1

non-destructive testing, materials science and numerous other areas for many years. Traditionally, an object has been subjected to x-ray CT to provide images of its (static) internal three-dimensional structure, or to study the effect of treatments such as heating by comparing an x-ray CT scan before and after the treatment. Since only an image of the sample after the treatment is obtained, such studies only provide an indirect window into the processes taking places during the



Figure 1. Top row: Shepp–Logan phantom sinograms with different cut-offs. Bottom row: Corresponding reconstructions. (a) Complete: No artifacts. (b) ROI: Region-of-interest caused by smaller detector width than sample size. (c) LA: Limited angle, where projections from a range of angles are missing. (d) VT: Variable truncation, where symmetrically placed metal objects outside the sample occlude the beam causing projections with variable amount of truncation. The gray-scale window of all reconstructions is [0, 0.6].

treatment. In recent years, *in situ* x-ray CT has become an increasingly wide-spread technique to study processes as they happen within material samples subject to various environments such as high temperature, high pressure, controlled atmosphere, fluid flow, etc. Some noteworthy examples in materials science include studies of rock fractures during freeze-thaw cycles [1], deformation and liquid flow in aluminum alloys under temperature control [2], evolution of fatigue cracks in magnesium [3] and fracture evolution in cement under compression [4].

In many *in situ* x-ray CT experiments the imaging equipment is complemented by additional equipment, often called rigs, to control the *in situ* environment, such as furnaces, tension rigs, fluid flow tubes, etc. Often, this equipment can be arranged to remain outside the x-ray beam from source to detector to prevent any effect on the acquired data. However, in some cases this is not possible, and the x-ray beam will either pass through or be fully or partially occluded by the equipment at some of the projection angles. This results in parts of the tomographic data being either 'polluted' by the additional attenuation through the equipment or, in the case of occlusion, completely missing. Depending on the severity of this effect, the resulting tomographic reconstruction may suffer from artifacts ranging from minor to completely destructive for the desired imaging task.

The present work is motivated by *in situ* x-ray CT studies of fluid flow through porous chalk. The experiment entails a percolation cell with four metal bars to sustain high temperature and pressure, as detailed in section 2 and illustrated in figure 2. At some projection angles the metal bars cause full or partial occlusion of the x-ray beam, which results in missing data and drastic streak artifacts in the reconstruction, see figures 3 and 4. The goal of the present work is to understand the cause of these artifacts and determine methods to reduce them.

The most common reconstruction method in routine use at synchrotrons remains the filtered backprojection (FBP) algorithm due to its versatility, robustness, computational efficiency and well-understood behavior. In recent years a variety of, mainly iterative, reconstruction methods based on an algebraic imaging model and using statistical and a priori information have been proposed and potential improvements over FBP demonstrated, see e.g. [5-8]. However, major challenges remain before these methods are suitable for larger datasets in practical synchrotron routine use, which often uses high resolution and thousands of projections. With these large amounts of data, the use of simple/fast algorithms is crucial. Iterative reconstrutction methods typically require computing times several orders of magnitude longer than FBP. They are more complex and typically involve tweaking of several parameters such as the number of iterations and any regularization parameters before improvements are obtained. Further, a multitude of different methods exist each with strengths and weaknesses for different types of images, making it difficult for non-specialists to choose a suitable method for a particular data set. Therefore, to be of direct relevance and easy to adopt in practice, in the present work we restrict ourselves to consider artifacts and reduction methods in conjunction with FBP reconstruction.

1.2. Typical reconstruction artifacts

A variety of reconstruction artifacts can be encountered in case of incomplete data. Figure 1 illustrates a number of typical cases with different forms of incomplete data and corresponding reconstruction artifacts. Shown first is the complete-data case without artifacts for the classical Shepp– Logan phantom. In the sinogram the columns contain projections obtained at angles from 0° to 180° in a parallel-beam configuration. Second, region-of-interest (ROI) data and artifacts arise from a smaller detector width than sample size causing projections to be truncated and detector-directed discontinuities to arise in the sinogram. The effect is a bright ring around and a pronounced cupping artifact across the ROI. The third case is limited angle (LA) data and artifacts, where projections are missing for a range of angles, causing angular discontinuities in the sinogram and major streak artifacts emerge from edges in the reconstruction. The fourth case, which is the focus of the present work, can be seen as a combination of LA data and ROI data, in which the ROI truncation varies across projections from fully missing through partial to complete projections. In the sinogram, both angular and detector-directed discontinuities are present. The corresponding reconstruction contains LA artifacts as well as new regularly spaced streaks seemingly unrelated to image features. We refer to this case as variable-truncation data and artifacts. In this article we focus on the variable-truncation (VT) artifacts, which we will show derive from the staircasing shape of the missing-data region in the sinogram with angular and detector-directed discontinuities, as seen in the zoom-inset in figure 1(d).

A closely related class of artifacts, mainly described in the medical imaging literature, is known as metal artifacts. Metal artifacts are often observed in medical imaging when metal is present, e.g. as implants in body parts and dental fillings. One can distinguish between two types of metal effects. First, the presence of highly-attenuating metal within tissue or bone can lead to physical effects, such as high attenuation coefficients, high noise-to-primary-signal ratio, high scatterto-primary-signal ratio, beam hardening, and non-linear partial volume effects [9, 10]. Second, the metal edges lead to sinogram discontinuities either directly present in the raw data or user-induced, e.g. through discarding the affected data [9]. The second type of metal artifacts are very similar to the VT artifacts in that they are thin regular streaks extending tangentially from the metal edges [10, 11].

Metal-artifact reduction is often treated as missing-data problems. Typical methods are projection completion methods, which fill in missing or deteriorated data points by synthetic data [12], and algebraic reconstruction methods, which may completely omit these data points [13, 14]. Omitting single data points is not possible when using FBP, as this method needs complete projections. The methods proposed in this article provide a possibility of disregarding missing or deteriorated data in conjunction with FBP. The simplest possible idea, and the starting point for the present work, is to replace any missing data points by zero values, as illustrated in the zoom-inset of figure 1(d), simply to have complete projections without blanks in order for FBP to process the data. Doing this has two consequences:

(i) It essentially removes the first type of metal artifacts, i.e. any physical effects caused by the presence of metal which corrupts the data, as such effects primarily relate to the projection lines traversing the metal (see for instance [9] and [10] for a description of metal artifacts). This means that we do not need to account further for potential metal artifacts of this type.

(ii) It introduces discontinuities in the detector direction of the sinogram. Such discontinuities will be drastically emphasized by the filtering step of FBP to cause an artificial over- and under-shooting effect, which in turn will be back-projected and create pronounced streak artifacts.

Many metal-artifact reduction methods are aimed at reducing these streaks by handling the discontinuities. Similarly, in the present work, we focus on addressing the discontinuities in the VT data to reduce the arising streak artifacts.

1.3. Contributions and organization of the present work

Much work has been devoted to characterize and conceive reduction methods for both LA artifacts, see e.g. [15–19], and ROI artifacts [20–23]. On the other hand the VT artifacts arising for the particular *in situ* x-ray tomography setup have to the best of our knowledge not been addressed in the literature.

To this end, the present work presents two contributions. First, after describing the experimental setup and data in section 2, we develop in section 3 a mathematical model of the considered setup to describe precisely how and where in the reconstruction the VT artifacts arise. The model is used to simulate VT data for variations of the experimental setup using small and large radii and distances of the metal bars, in order to describe the artifacts across a more general class of problems. We also explain that the artifacts arise from sinogram discontinuities during the filtering step of FBP.

Second, in section 4 we propose three methods to reduce the VT artifacts – all of them simple to implement as a preprocessing step for standard FBP to allow efficient reconstruction, essential for large real synchrotron data sets. Insertion of zeros values (IZV) for the blank data points provides the starting point, on which we want to improve by following proposed methods:

- (i) Reduction to limited angle (RLA), which simply discards all truncated projections to obtain an LA data set. This is a very easy way to remove the detector-directed discontinuities, but the drawback is that it discards useful data.
- (ii) Detector-directed smoothing (DDS), which applies local smoothing to remove the detector-directed discontinuities.
- (iii) Reflexive boundary condition (RBC), which handles the detector-directed discontinuities by imposing a reflexive boundary condition.

In section 5 we report qualitative and quantitative results to assess the proposed methods on simulated and real VT data sets. The results show that all methods can remove the VT artifacts effectively, while DDS and especially RBC allow further reconstruction improvement by using data from the partial projections, compared to RLA, which discards this data. Finally, section 6 discusses the methods and results in a wider context before section 7 concludes the study.



Figure 2. (a) Side-view of the percolation cell. (b) Top-view of the setup, where the specimen is placed in the center between four metal bars (not to scale). Four angular positions for the detector (the black bar) are shown: For two of them, the projections are complete. For the detector position along the SW–NE diagonal, the beam is occluded completely by the metal bars and no signal is measured at the detector. In between one position is shown at which the outermost parts of the beam are occluded by two of the metal bars, leading to a projection with truncation from both sides.

2. Data

2.1. Acquisition set-up

The motivating case for the present study is in situ x-ray microtomography imaging of fluid flow through porous chalk, in which the goal is to recover oil from the North Sea underground. In situ x-ray tomography data was obtained for a cylindrical porous chalk sample of diameter 0.6 mm using beamline BL20XU of the SPring-8 Synchrotron Radiation Facility, Japan using a monochromatic (28 keV) parallel-beam scan configuration. Fluid is forced through the sample by a percolation cell, seen in figure 2(a), by applying a pressure of 50 bars imitating the underground conditions. The goal is to model the structural changes of the sample during the fluid flow and a series of scans are acquired continuously over the experiment. Structural changes are slow compared to the acquisition time of each complete scan and any sample deformations within each scan can be neglected. The percolation cell is equipped with four metal bars which can sustain pressures of 200 bar and temperatures of 100 °C. The metal bars have a radius of 1 mm and are positioned in a square around and at approximately distance of 15.6 mm from the sample. The number of detector pixels is 2048×2048 , providing in each horizontal slice a field of view (FOV) of approximately 0.5 mm in diameter [24]. The detector is positioned outside the percolation cell and 1800 projections are collected covering 0 to 180 degrees. As seen in figure 2(b) most projections are complete, some are fully occluded by the metal bars, while some projections are partially occluded. These partial projections are the focus of the present work. In addition, since the sample is larger than the FOV, all projections are slightly truncated.

2.2. Reconstruction provided by the synchrotron site

At SPring-8, FBP is used slice-by-slice to produce a 2048³ reconstructed volume based on the acquired CT data [25]. Due to slices being independently reconstructed thanks to the parallel-beam geometry we consider without loss of generality the FBP reconstruction of a single 2048×2048 -pixel slice. Artifacts are clear, especially towards the outlying region of the reconstruction, see figure 3. They appear as streaks with angles of approximately 45 and 135 degrees measured in the counter-clockwise direction from the horizontal axis. The artifacts cause problems for subsequent analysis of the data: From the reconstruction, an automated segmentation of the chalk pores is needed as a step towards the fluid flow model. This requires a high-quality artifact-free reconstructed image. The present work is motivated by identifying the cause of the streaks and finding a way to reduce them.

2.3. Processing the raw data

From the complete raw data set the data corresponding to slice number 1000 out of 2048 is extracted and consists of 1800 projections of length 2048 pixels. These are the transmission values measured at the detector which we denote by $I(\phi, p)$, where ϕ denotes the rotation angle and p the detector position. In addition, two flat fields $I_{0,\text{before}}$ and $I_{0,\text{after}}$ are recorded before and after $I(\phi, p)$ was recorded, i.e. with the x-ray beam on, but the sample out of the field of view, as well as a dark field I_{bg} , i.e. with the source off to account for any background radiation. Flat and dark fields are used for flat- and dark-field correction of all projections, i.e.

$$Z(\phi, p) = \frac{I(\phi, p) - I_{\rm bg}(\phi, p)}{I_0(\phi, p) - I_{\rm bg}(\phi, p)}.$$
 (1)

As the source conditions such as voltage and current may change slightly during a scan, $I_0(\phi, p)$ is an interpolated flat field calculated for each projection from the two flat fields $I_{0,\text{before}}$ and $I_{0,\text{after}}$ by weighting by projection angle:

$$I_0(\phi, p) = \left(1 - \frac{\phi}{180^\circ}\right) \cdot I_{0,\text{before}}(p) + \frac{\phi}{180^\circ} \cdot I_{0,\text{after}}(p).$$
⁽²⁾

Figure 4 shows the flat- and dark-field corrected sinogram $Z(\phi, p)$. It represents the number of counted photons relative to the emitted photons and is the sinogram from which the SPring-8 reconstruction in figure 3(a) is calculated. The horizontal axis is projection angles, ϕ , over 0 to 180 degrees and the vertical axis is detector element position, *p*. Because the metal bars are positioned relatively far from the chalk sample and the detector width is small compared to the metal bar radius, the full sinogram resembles the LA sinogram in figure 1(c). However, on close inspection, as seen in the zooms of figure 4, clear differences are apparent. Instead of the immediate transition from a full to a complete blank projection in the LA case, projections

4



Figure 3. (a) FBP reconstruction provided by the SPring-8 synchrotron facility with zoom areas indicated by black squares. (b) Zooms of the reconstruction.



Figure 4. Flat- and dark-field corrected sinogram $Z(\phi, p)$ and two zooms. The white rectangles in (a) and (b) denote the zoom areas in (b) and (c), respectively. Gray scale window is [0, 0.53], corresponding to the fraction of photons detected. (a) Transmission sinogram. (b) Zoom 1. (c) Zoom 2.

are increasingly truncated over an angular range, eventually becoming completely blank, when the beam is fully occluded, before gradually becoming decreasingly truncated again. This corresponds to the case in figure 1(d). Due to the symmetry of the metal bar positions, the patterns are symmetric around the detector center, seen in zoom 1. Taking a more narrow look at the sinogram in zoom 2, stair-casing is revealed with vertical and horizontal boundaries. In the idealized case of continuous ϕ and p, the width of the truncated projections would be decreasing or increasing continuously. Since a finite number of projections are recorded in practice, a stepwise change is observed with angulardirected discontinuities, i.e. across vertical line segments.

In the detector direction a smooth transition is seen from the full signal to zero photon counts across the detector elements in the partially occluded projections. The smooth transition indicates that the occlusion does not occur abruptly but slightly gradually over some detector pixels. The lower counts recorded in the transition region are not representative of the sample and need to be handled along with blank measurements due to beam occlusion.

2.4. Handling small or zero transmission values in FBP reconstruction

For a monochromatic x-ray beam, Lambert–Beer's law, here rewritten as

$$\int_{L} \mu(s)ds = -\log\frac{I}{I_0},\tag{3}$$

where $\mu(s)$ is the linear attenuation coefficient at spatial position *s* to be recovered, provides a decent model of x-ray attenuation as a basis for reconstruction. Dark field and variables ϕ and *p* have been omitted here for simplicity. For parallel-beam data reconstruction can be done using the filtered backprojection (FBP) algorithm applied to a sinogram consisting of the right-hand side of (3) for all measured angles ϕ and detector positions *p*.

To arrive at such a sinogram, we need to apply the negative logarithm to $Z(\phi, p)$, and we denote the resulting sinogram as $S(\phi, p) = -\log Z(\phi, p)$. Zero- and close-to-zero values in $Z(\phi, p)$ are not handled well by this model, as they are mapped



Figure 5. Sinogram $S(\phi, p)$ after negative logarithm transform and insertion of zero values in blank pixels, along with two zooms. The white rectangles in (a) and (b) denote the zoom areas in (b) and (c), respectively. Gray scale window is [0, 1.43]. (a) Attenuation sinogram. (b) Zoom 1. (c) Zoom 2.



Figure 6. (a) FBP reconstruction based on the sinogram $S(\phi, p)$ with zeros inserted in blank positions, with zoom areas indicated by black squares. (b) Zooms of the reconstruction.

to infinity or very large values by the negative logarithm and will dominate the reconstruction [10]. Values of $Z(\phi, p) \in [0, 0.53]$ smaller than an empirically chosen threshold, 0.24, are discarded, leaving blank areas in the sinogram. FBP cannot handle pixels without an assigned value, so some values need to be assigned to complete the projections. The simplest choice, that does not make any assumptions, is to assign the value zero in the blanks of $S(\phi, p)$, resulting in the attenuation sinogram, which we still denote $S(\phi, p)$, shown in figure 5. In practice this operation was carried out by replacing all entries of $S(\phi, p)$ for which $Z(\phi, p)$ was smaller than 0.24 by a zero. A few, less than 50, isolated pixels within the existing projections were hereby unintentionally set to zero. These pixels were assigned values by linear interpolation in the corresponding projection.

Following this, an FBP reconstruction was computed using the iradon function of MATLAB R2014b employing a Hamming filter for noise reduction, see figure 6. The FBP reconstruction is seen to contain thin streak artifacts distributed regularly over the image with angles of approximately 45 degrees and 135 degrees. As such, the streaks have some resemblance with the SPring-8 reconstruction in figure 3, only less pronounced. As we do not have access to details of the synchrotron reconstruction method, including possible pre- and postprocessing steps, it is difficult to determine the cause of the slightly different appearance. However, since the streaks in both reconstructions occur at the same angles, we believe their origin in both cases to lie with the VT data.

As seen in figure 5, the assignment of zero-values in $S(\phi, p)$ introduces detector-directed discontinuities in the sinogram. As we will describe further in the following section, it is these discontinuities that cause the regular streak artifacts in the FBP reconstruction.

3. Analysis

3.1. Modeling the data cut-off

In order to understand the arising VT artifacts we devise a mathematical model of the missing sinogram data caused by the metal bars. We use an indicator function $m(\phi, p)$, which



Figure 7. Metal bars 1–4 centered at (x, y) = (d, d), (-d, d), (-d, -d), and <math>(d, -d).

we refer to as the mask, to represent whether a data point exists, $m(\phi, p) = 1$, or not, $m(\phi, p) = 0$, for $\phi \in [0^{\circ}, 180^{\circ})$ and $p \in \mathbb{R}$. The case $m(\phi, p) = 0$ corresponds to those projection lines which are blocked by the metal bars. Figure 7 shows the considered setup. Four metal bars with radius *r* and indexed by i = 1, 2, 3, 4 are positioned at $(x_1, y_1) = (d, d)$, $(x_2, y_2) = (-d, d), (x_3, y_3) = (-d, -d)$ and $(x_4, y_4) = (d, -d)$, i.e. at a distance of $\sqrt{2d}$ from the origin, which is the position of the sample and assumed to be the center of rotation. The *p*-axis, which represents the detector, is oriented at an angle ϕ measured counterclockwise from the positive *x*-axis.

Each line of integration (along an x-ray) can be parametrized as a function of ϕ and p as

$$L(\phi, p) = \{(x, y) \in \mathbb{R}^2 : x \cos \phi + y \sin \phi = p\}.$$
 (4)

The line of integration at angle ϕ which intersects the center of bar *i* is denoted $L_i(\phi)$. From knowledge of metal bar *i*'s center coordinates (x_i, y_i) the corresponding *p*-value $p_i(\phi)$ of $L_i(\phi)$ can be found as

$$p_i(\phi) = x_i \cos \phi + y_i \sin \phi. \tag{5}$$

Using the trigonometric identity $x \cos \phi + y \cos \phi = \sqrt{x^2 + y^2} \sin(\phi + \operatorname{atan2}(y, x))$ the expression for $p_i(\phi)$ can be simplified to a single sine function,

$$p_i(\phi) = \sqrt{2d}\sin(\phi + \phi_i),\tag{6}$$

where $\phi_i = \operatorname{atan2}(y_i, x_i) = 45^\circ, 135^\circ, -135^\circ, -45^\circ$ for i = 1, 2, 3, 4.

Since the metal bar radius is *r*, all parallel rays within a distance of *r* from $L_i(\phi)$ will be blocked by metal bar *i*, in other words, for all *p* in the range

$$p_{\text{range}_i}(\phi) =]p_i(\phi) - r, p_i(\phi) + r[$$
 for $i = 1, 2, 3, 4,$
(7)

no signal will be measured. As an example, the range $p_{range_4}(\phi)$ is sketched in figure 7. At a fixed angle ϕ a ray will be blocked if its *p*-value is in (at least) one of these ranges, or equivalently, in the union of the ranges. Hence we can write the mask combining all four metal bar ranges explicitly as

$$m(\phi, p) = \begin{cases} 0 & \text{when } p \in \bigcup_{i=1}^{4} p_{\text{range}_i}(\phi), \\ 1 & \text{otherwise.} \end{cases}$$
(8)

Figure 8 shows a synthesized image consisting of four metal bars seen from above, the corresponding sinogram mask extending to fully including the metal bars, and a zoom of the mask corresponding to the actual detector width. The metal bar radius is 1 mm and the distance to the rotation center is $\sqrt{2} \cdot 11 \text{ mm} = 15.6 \text{ mm}$ to closely approximate the actual experimental set-up. The full mask contains four bands of missing data, one for each metal bar. Each band has a distinct sinusoidal appearance, which is in agreement with (8), which prescribes four superpositioned sinusoidal bands of vertical thickness 2r. Due to the symmetry of our setup, the two sets of opposite metal bars result in two pairwise crossings of the bands over the 180° range.

When the detector only covers 0.5 mm instead of the entire image, this is equal to zooming in on the sinogram, seen in figure 8(c). The zoom is indicated by the centered rectangle in figure 8(b). The missing-data region in the sinogram zoom closely resembles the motivating real-data case in figure 4 and can be used to simulate missing data from a complete sinogram.

3.2. Varying metal bar radius and distance to object

Using the derived mathematical model we can simulate the effect of variations of the experimental setup (metal bar radius and distance) on the arising data truncation. We do this to show how the derived mathematical model describes a larger class of potentially interesting imaging scenarios with variable beam occlusion. By considering this extended set of test problems we hope to illustrate that our proposed VT artifact reduction methods can enable useful reconstructions under more challenging data truncation than that of the originally motivating case of figure 3.

Figure 9 shows four cases of possible acquisition set-ups (combinations of radius 1 mm and 2 mm, and distance 3 mm and 11 mm) with images and corresponding masks applied to the sinogram of the Shepp–Logan phantom. To verify that the analytically computed mask expression (8) is correct, the boundary curves of the intervals in (7) as function of ϕ for all bars are plotted on top of sinogram masks simulated numerically by applying MATLAB's radon function to the images in the top row in figure 9. The boundary curves follow the numerical mask boundaries exactly, and further illustrate

L Borg et al

(c)

(a)

Figure 8. (a) Top-view of four metal bars with radius 1 mm. (b) The corresponding sinogram mask (white is 1, black is 0). The horizontal axis is projection angles, ϕ , over 0 to 180 degrees and the vertical axis is detector element position, *p*. The two horizontal lines denote the zoom of the sinogram in (c), imitating a detector width of 0.5 mm. (c) A zoom of the mask, which reproduces the variable truncation of projections in the real data in figure 4.

(b)



Figure 9. The top row shows images of four metal bars seen from above. Metal bar radii r and distances d are (r, d) = (1, 11), (1, 3), (2, 11), (2, 3), respectively. This gives rise to four sinogram masks (similar to the mask shown to the right in figure 8): Mask 1, 2, 3, and 4, which are applied to the sinogram of the Shepp–Logan phantom, shown in the bottom row. The detector width is 0.5 mm. (a) Mask 1. (b) Mask 2. (c) Mask 3. (d) Mask 4.

the contribution of each of the four metal bar shadows in the sinogram.

The original case corresponds to Mask 1 and is seen to entail the smallest amount of missing data.

Increasing radius at fixed distance is seen to produce a wider band of missing data, while the cut-off slope appears to be unchanged. This is in agreement with behavior predicted by (7) that the $p_i(\phi)$ follows a sine curve independent of r, while the vertical band thickness grows linearly with r, which causes a wider horizontal gap of missing data. Increasing distance at fixed radius is seen to reduce the missing data gap width and increase cut-off steepness. This can also be understood from (7), since the sinusoidal amplitude grows linearly with d causing the derivative at the crossings to be of larger magnitude, which is equivalent to a steeper cut-off and a narrower missing data gap.

3.3. How variable-truncation artifacts arise in filtered backprojection

To understand how streaks arise, we first briefly remind the reader that reconstruction by FBP consists of two steps: Filtering and backprojection. The filtered sinogram will be referred to as $H(\phi, p)$.

Any introductory textbook on the mathematics of CT, such as [10], explains that the filtering step of FBP emphasizes discontinuities to yield very large values with opposite signs on each side of the discontinuity in the filtered projection. This is illustrated in figure 11. Backprojecting such a filtered projection produces a pair of a dark and bright streak along $L(\phi, p)$.

A truncated projection viewed over the full projection's domain contains a discontinuity at the truncation point. If the truncation is constant over all projections, as is the case with



Figure 10. Reconstructions based on sinograms in figure 9 with Mask 1, 2, 3, and 4 from left to right, respectively. The white squares indicate the positions of the zooms in the bottom row. The gray-scale windows are for the full images [-0.23, 1.03], [-0.49, 1.03], [-0.37, 1.02], [-0.76, 0.98] and the zooms [0.10, 0.30], [0.05, 0.30], [-0.01, 0.20], [0.10, 0.30], respectively. Different windows are used, since different amounts of missing data cause different reconstruction intensities. (a) Reconstruction 1. (b) Reconstruction 2. (c) Reconstruction 3. (d) Reconstruction 4.

ROI data as shown in figure 1(b), these streaks still 'line up' and cancel each other out (in the continuous or highly sampled case) yielding no streak artifacts but only a circle of discontinuities at the ROI border and a cupping artifact as seen in figure 1(b). However, if truncation is not constant over all projections, as in the case in figure 1(d), the resulting streaks will no longer line up and cancel each other out. In particular, in case of a finite number of projections, the angular increment is finite and hence even a gradual truncation width change in the continuous domain is approximated by steps in the discrete practical case. This means that adjacent truncated projections have different width, which results in streaks occurring at different *p*-positions, preventing them from canceling with each other. The result, as shown in the following section, is regular streaks at certain angles across the entire reconstruction.

3.4. Streak artifacts caused by data with variable truncation

Figure 10 shows the reconstructions when applying mask 1–4 to the sinogram of the Shepp–Logan phantom. In the top row, the full reconstructions are seen, and in the bottom row the zooms specified by the white rectangles in the top row are seen. To emphasize VT artifacts the zooms are shown in narrower gray-scale window, as specified in the figure caption. The full reconstruction images mainly show LA artifacts, i.e. few large streaks emerging from edges in the image, especially from the skull. The zooms additionally show regularly spaced VT artifacts. In the reconstructions in figure 10, there is exactly one streak for each detector-directed discontinuity in the sinograms in figure 9, and its angle and position depends

on the position of the detector-directed discontinuity. For further discussions and illustrations of this, see [26].

Four different angles of streaks are seen in Reconstruction 1: either just under or just over 45 degrees or just under or just over 135 degrees. The reason of this is found in the corresponding sinogram in figure 9 where the detector-directed discontinuities are seen to occur exactly around these angles. Because the sinogram discontinuities occur with approximately regular spacing in the *p*-direction, the streaks also appear regularly spaced and with angles corresponding to the angles of the truncated projections.

The streaks in case 3 have the same spacing as in case 1, because the cut-off slope is the same. However, since the detector-directed discontinuities occur farther from 45° and 135° the two sets of VT artifacts are further from being pairwise parallel.

For mask 2 and 4 the metal bars are much closer to the rotation center than for mask 1 and 3. This leads to a less steep cut-off of the data, and more plentiful detector-directed discontinuities, which again means that more densely spaced streaks are present in the reconstructions, as observed in the second and fourth columns of figure 10. As before, the streaks in case 2 and 4 occur with the same spacing but at different angles, since the detector-directed discontinuities occur at different angles.

As a final remark, we note that in each case four seemingly parallel sets of streaks can be observed, however since the detector-directed discontinuities occur over an angular range, only the streaks coming from the same projection are in fact perfectly parallel.



Figure 11. The proposed methods illustrated on a 1D truncated projection before and after filtering. (a) Projections before filtering, $S(\phi, p)$. Note, that RBC is shifted downwards for illustration purposes as indicated by the arrow and coincides with COM left of the discontinuity. (b) Projections after filtering, $H(\phi, p)$. (c) Zoom of projections are filtering. Note, that in RBC the padded values will be set to zero before backprojection, see figure 13 for an illustration on the original data.

4. Methods

4.1. Overview of proposed methods

Having completed our analysis and description of the VT artifacts caused by detector-directed discontinuities we now proceed to propose three methods to reduce them. The proposed methods are motivated by being straightforward to implement in conjunction with FBP to allow fast reconstruction using a well-understood algorithm already in routine use at synchrotron beamlines. In the following subsections we describe each of the proposed methods but we first give a brief overview of the methods considered:

- Complete dataset (COM): in case of simulated data use the complete sinogram as 'ground truth' without applying any missing-data mask.
- Insertion of zero values (IZV): fill all missing data points by zeros.
- Reduction to limited angle (RLA): set all truncated projections to zero, thus reduce to limited-angle problem.
- Detector-directed smoothing (DDS): multiply all truncated projections with a function for local smoothing of detector-directed discontinuities.
- Reflexive boundary condition (RBC): introduce a reflexive boundary condition at the detector-directed discontinuities.

Figure 11 illustrates the methods on a 1D truncated projection before (left) and after (middle and right) the filtering step of FBP. The methods RLA, DDS and RBC are motivated by handling the detector-directed discontinuities in order to reduce VT artifacts that arise when simply filling missing values by zeros in the IZV method.

4.2. IZV: insertion of zero values

The simplest choice that does not make any assumptions is to assign zero values in the log-transformed sinogram domain. However, this introduces discontinuities in the truncated projection causing oscillations after filtering, as illustrated in figure 11. Figure 11(b) illustrates the extreme overshoot followed by an undershoot near the discontinuity, resulting in pairs of dark and bright streaks across the reconstructed image. The overshoot even offsets long before the discontinuity, as seen in figure 11(c). Typical ROI-artifacts appear as a bright ring in the outskirts of the reconstruction (where the truncation discontinuities are), which is also explained by this effect. IZV is the starting point that we want to improve on by the following methods.

4.3. RLA: reduction to limited angle

Replacing all truncated projections completely by zero values eliminates sinogram stair-casing, thereby removing the detector-directed discontinuities and therefore also the VT artifacts. Setting truncated projections to zero reduces the problem to a pure LA problem, which as seen in figures 1(c) and (d) effectively removes streaks. In figure 11, this method is seen as the graph which is zero for all *p*-values. It is the impression of the authors that RLA is in practical use among x-ray CT experimentalists as a simple method to avoid the streak artifacts caused by variable data truncation. While simple and effective, this method completely discards the potentially usable data in the truncated projections. For an example of this, see [26, figure 19].

4.4. DDS: detector-directed smoothing

This method multiplies each of the truncated projections with a function, which dampens intensity values near the detectordirected discontinuities, thereby obtaining smooth transitions, as seen in figure 11. This method is motivated by [15, 27], where smoothing is applied in the angular direction to reduce LA artifacts. In our case we do not apply angular smoothing, but smoothing in the detector direction. At angle ϕ_0 we have a projection truncated at $p = a_0$ and $p = b_0$ in $S(\phi_0, p)$. We modify the smoothing function used in [27] for detectordirected smoothing by defining

$$g_{\epsilon}(p) = \left(\frac{1}{\epsilon^2}p(2\epsilon - p)\right)^2,\tag{9}$$

where ϵ specifies the width of the smoothing region, i.e. the number of pixels over which projection values will be smoothed to decay to zero, and

$$\kappa_{\epsilon}^{a,b}(p) = \begin{cases} g_{\epsilon}(p-a) & \text{if } p \in [a, a+\epsilon], \\ 1 & \text{if } p \in]a+\epsilon, b-\epsilon[, \\ g_{\epsilon}(p-(b-2\epsilon)) & \text{if } p \in [b-\epsilon, b], \\ 0 & \text{else.} \end{cases}$$
(10)

Then the smoothed projection is obtained as

$$Q(\phi_0, p) = \kappa_{\epsilon}^{a, b}(p) S(\phi_0, p). \tag{11}$$

The function $g_{\epsilon}(p)$ is two times differentiable and the artifacts are two orders smoother after applying the smoothing, meaning that they are not completely smoothed, so they are mathematically still present [28]. This is in line with what we observe: after filtering of the DDS-projection, there still remain some oscillations, however substantially dampened compared to the IZV-projection, as seen in figure 11(c). We have empirically found a smoothing region of $\epsilon = 30$ pixels to provide a suitable level of smoothing and use this value throughout the presented results.

4.5. RBC: reflexive boundary condition

This method introduces a reflexive boundary condition at the detector-directed discontinuities. The high-pass ramp filter in FBP emphasizes discontinuities, and the low-pass filter, e.g. a Hamming filter, often used together with the ramp filter may not dampen the oscillations sufficiently to avoid streaks. To counteract this, a reflexive boundary condition [29] is imposed to the truncated projections by padding by existing data values reflected in the discontinuity point. This induces a Neumann boundary condition with zero slope at the boundary. This method is similar to typical handling of ROI-problems where the outermost detector element values are used as padding values. After filtering of the projections, the padding is set to zero again before backprojecting, because we want to avoid introducing any artificially created data. This method works on the discontinuity itself and does not alter any existing pixel values explicitly as the RLA method and DDS methods do. As observed in figure 11(c), this is the only method that continues to coincide with the COM signal to the left of the cut-off after filtering, except for very few pixels \simeq 4 just before the cut-off. The discrepancy at these last few pixels is caused by the filter width in FBP: at and near the discontinuity, the filter uses pixel values from the right side of the discontinuity to calculate values to the left of the discontinuity. When reflected pixel values are used instead of pixel values from COM, this alters the pixel values left of the discontinuity.

5. Results

5.1. Overview and purpose of experiments

The proposed artifact reduction methods are assessed in the following three studies:

(i) In the original dataset in figure 3 we want to reduce the VT artifacts to enable subsequent segmentation as part

of an automated data processing pipeline designed for improving analysis of fluid flow through the pores of the chalk. The first study shows how the proposed methods perform on the original data set. Since a ground truth is not available for quantitative comparisons, we evaluate the methods qualitatively, i.e. using visual inspection.

- (ii) Second, we compare the methods on the well-known Shepp-Logan phantom with simulated missing data in order to evaluate the methods in a controlled noise-free setting sufficiently simple that the effect of the different methods can be clearly seen. The four masks from figure 9 are applied to a complete sinogram to assess the proposed methods across a range of possible experimental setups. The full-angle reconstruction will be used as the reference when assessing the methods quantitatively.
- (iii) Finally, an equivalent real-data study is carried out using a separate chalk data set with a complete set of non-truncated projections acquired under comparable imaging conditions as the original truncated chalk data set. This study is included to evaluate the methods on a real, noisy, high-complexity problem, comparable to the original dataset, but with the advantage that a full-angle reconstruction can be determined and used as reference image for quantitative assessment.

5.2. Quantitative image quality assessment

The following image quality measures are chosen to evaluate the methods quantitatively:

(i) The root-mean-square error (RMSE):

RMSE =
$$\left(\frac{1}{NM}\sum_{i=1}^{M}\sum_{j=1}^{N}|R(i,j) - R_0(i,j)|^2\right)^{\frac{1}{2}}$$
, (12)

where *N* and *M* are number of rows and columns in the images *R* and R_0 , representing the reconstruction and the reference image, respectively. This measure is a standard measure, used for pixel-wise comparison in images.

(ii) Spectral magnitude distortion (SMD) [30]:

$$SMD = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |P_R(i,j) - P_{R_0}(i,j)|^2, \quad (13)$$

where P_R and P_{R_0} are the power spectra of the reconstruction and reference image, respectively. Since VT streaks occur with regular spacing, the power spectrum of a reconstruction with such artifacts will contain distinct pronounced peaks at particular frequencies. In addition, the streaks are discontinuous features relative to the underlying image and as such show up in the power spectrum as high-frequency components. In contrast, the power spectrum of the reference image contains only the frequencies of the underlying image, and the SMD precisely captures these differences. This measure, which is not as commonly used as the RMSE in the tomographic imaging literature, is chosen because it is particularly

21



(a) RLA

(b) DDS

(c) RBC

Figure 12. Top row: Full reconstructed images by RLA, DDS and RBC with white squares indicating region of interest. Middle row: Zoom to region of interest. Size of region of interest is 200 by 200 pixels. Bottom row: Difference images (zoom) between the RLA, DDS, and RBC reconstructions and IZV reconstruction. Gray-scale windows: Full reconstructions and zooms: $[-1 \cdot 10^{-4}, 10 \cdot 10^{-4}]$, difference images: $[-2 \cdot 10^{-4}, 2 \cdot 10^{-4}]$.

suited to detect changes in regular streak artifacts and hence well aligned with the purpose of the paper.

5.3. Original chalk dataset with variable-truncation data

Figure 12 shows the results of applying RLA, DDS and RBC to the original chalk data set with variable data truncation. The top row shows the full reconstructions for RLA, DDS, and RBC from left to right. The white squares indicate the position of a zoom to a region of interest shown in the middle row. The bottom row shows difference images between RLA, DDS and RBC and the IZV reconstruction in figure 6(b) to emphasize improvements.

The top and middle rows of figure 12 illustrate that the VT streaks are reduced to such an extent that they are no longer visible to the naked eye. Streaks, consisting of a dark and a bright line side by side, are present in all three difference

images in the bottom row of figure 12, corresponding to only being present in the IZV reconstruction. This means that they have effectively been removed by RLA, DDS and RBC. The RLA difference image has additional vague underlying streak structures with same directions of the streaks. Since these are only present here, they must correspond to the data lost when the truncated projections are set to zero. The discarded projections all correspond to angles near 45° and 135°, explaining the angles of the underlying structures. The background of the RBC difference image appears constant, whereas smooth variations appear in the gray squares for DDS. We can explain this, since RBC works only on the discontinuity itself and does not damp or enhance the projection values near the discontinuities as DDS does. In conclusion, all methods remove streaks satisfactorily, while we consider RBC to be superior to DDS, which is in turn superior to RLA, in terms of not modifying reconstruction background structure.



Figure 13. Illustration of application of IZV (left) and RBC (center and right) on a zoom of the sinogram in the *H*-domain (after filtering). The left and right sinograms are the ones used when backprojecting. (a) IZV after filtering. (b) RBC after padding and filtering. (c) RBC after padding, filtering, and reinsertion of zeros.

Given that some projections are completely missing, especially so for RLA, one could have expected LA artifacts similar to the ones in figure 10 but none are seen. The intensity of LA streak artifacts scale with the contrast between the features from which they emerge [28]. Unlike the Shepp–Logan phantom, which contains large contrasts and thus clear LA artifacts, the chalk-data contrast is relatively low, which we believe can explain the absence of pronounced LA streak artifacts.

Figure 13 shows zooms of the $H(\phi, p)$ sinogram, i.e. after filtering, for IZV and RBC applied to the original dataset. Figure 13(a) illustrates the projections of IZV after filtering, just before backprojection. The extreme overshoot/undershoot shown in figure 11 is also seen in this image. In the missingdata region next to the discontinuities of the truncated projections, the pixels are very dark (undershoot), while just within the existing-data region the pixels are very bright (overshoot). Figure 13(b) shows the same part of the sinogram for RBC with reflexive padding and after filtering, while figure 13(c)further shows when zeros have been reinserted in place of the padding prior to backprojection. We note that detectordirected discontinuities remain in the RBC sinogram and will thus be backprojected. However, they do not generate VT artifacts in the RBC reconstruction in figure 12. This emphasizes that it is the filtering step of FBP that leads to streaks from detector-directed discontinuities, whereas the backprojection step does not.

5.4. Simulated variable-truncation data using the Shepp– Logan phantom

The four different masks are applied to the Shepp–Logan sinogram. As was seen in figure 10, the IZV reconstructions suffer from both LA and VT artifacts. RLA, DDS and RBC are applied to the four cases and zooms of reconstructed images are shown in figure 14 along with the same zoom for IZV repeated for ease of comparison.

RLA, DDS and RBC reconstructions all show substantial reductions of VT artifacts. Because of the phantom simplicity, the noise-free simulation study and the use of a narrow grayscale window, differences between the methods are highlighted, and the LA artifacts stand out clearly. The DDS method does not remove the streaks as successfully as the other methods do, which is most clear when comparing the reconstructions in the first and third columns. Figure 11(c) shows that the DDS signal still has an overshoot followed by and undershoot near the discontinuity. In other words, DDS smooths the discontinuities meaning that it also smoothies the streaks and dampens their effect, but does not remove them completely. There is a trade-off such that if ϵ is too small, streaks are not sufficiently reduced, and if too large, streaks are reduced but too much of the actual image is lost as well.

Comparison of the reconstructions in the fourth column illustrates that DDS and RBC have fewer LA artifacts than RLA has: the black blob in the reconstructions would have been an ellipse if data were not missing; the blob in the RLA reconstruction is larger, less closed and more smeared than for DDS and RBC, because more data has been erased in the RLA method. Similarly, the small bright ellipse in the bottom-right corner is more well-defined for DDS and RBC than for RLA. Comparable observations can be made for the three other cases but to a lesser extent since data truncation is most severe in the fourth case.

RLA removes all detector-directed discontinuities completely, meaning that VT artifacts are not present in RLA reconstructions. The RBC reconstructions are similar to the RLA reconstructions in that no streaks are visible, indicating that RBC removes – or at least substantially reduces – the VT artifacts. RBC thus has fewer LA artifacts than RLA and better streak reduction than DDS, and from a qualitative perspective the results render it the superior method of the three.

For quantitative assessment, table 1 reports RMSE and SMD values for IZV, RLA, DDS and RBC applied to the four cases. The computed values are global, i.e. computed over the full reconstruction images using the full-angle COM reconstruction as reference. In all cases, the global RMSE values are very similar for all methods and hence do not reflect the visual assessment that all methods improve over the IZV method: in terms of global RMSE, RLA appears to worsen the reconstruction, while DDS and RBC only offer marginal gains. The global SMD measures, on the other hand, are in better alignment with the visual assessment, confirming RBC as the best performing method. We believe the discrepancy for RMSE is



Figure 14. Shepp–Logan reconstruction zooms based on methods (vertical) and different masks imposed on the sinogram (horizontal). The position of the zooms are shown in figure 10. Gray-scale windows are [0.10, 0.30], [0.05, 0.30], [-0.01, 0.20], and [0.10, 0.30] for the reconstructions in columns 1 to 4, respectively.

Table 1. Quantitative image quality results for IZV, RLA, DDS and RBC reconstruction from simulated variable-truncation data of the Shepp–Logan phantom. RMSE and SMD values are computed over the entire reconstructed image using the full-angle COM reconstruction as reference image.

	RMSE				SMD			
	Mask 1	Mask 2	Mask 3	Mask 4	Mask 1	Mask 2	Mask 3	Mask 4
IZV	0.0422	0.0930	0.0634	0.1361	5.3092	17.0283	8.5020	27.3855
RLA	0.0461	0.1043	0.0669	0.1478	4.3620	15.1321	7.5844	26.0166
DDS RBC	0.0416 0.0414	0.0922 0.0925	0.0631 0.0630	0.1356 0.1364	3.7973 3.7827	12.8935 12.8449	6.9873 6.9641	23.4302 23.3903

due to quite large differences between the full-angle reconstruction and any of the other reconstructions caused by the missing data and LA artifacts. The improvements that RLA, DDS, and RBC offer in terms of RMSE are so small compared to the global quality drop from the full reconstruction to either of the missing-data reconstructions, that they may be completely dominated by other artifacts. Measuring the RMSE only over a smaller region, where LA artifacts are not dominating, may better capture the reduction of VT streaks. We therefore replace the global measures by computing RMSE and SMD locally over the zooms in figure 14 and report values in table 2. These local RMSE results are in all cases much better aligned with the visual assessment, indicating an improvement by RLA over IZV, and further improvements by DDS and RBC. Similar conclusions are obtained for local SMD. In general, the measures do not indicate a noteworthy difference between DDS and RBC. However, as a consistent trend, RBC performs marginally better when Mask 1 and 3

|--|

	RMSE				SMD			
	Mask 1	Mask 2	Mask 3	Mask 4	Mask 1	Mask 2	Mask 3	Mask 4
IZV	0.0385	0.1038	0.0567	0.1607	0.0955	0.2381	0.1319	0.2905
RLA	0.0132	0.0374	0.0132	0.0368	0.0450	0.1059	0.0439	0.0947
DDS RBC	0.0101 0.0098	0.0174 0.0195	0.0102 0.0099	0.0185 0.0211	0.0355 0.0345	0.0666 0.0679	0.0348 0.0346	0.0696 0.0719



Figure 15. Reconstructions from sinograms in figure 9, with varying radii and distances. The white squares indicate the zooms in figure 16. Gray-scale windows are $[1.0 \cdot 10^{-4}, 6.0 \cdot 10^{-4}]$, $[0, 6.0 \cdot 10^{-4}]$, $[-1.0 \cdot 10^{-4}, 6.0 \cdot 10^{-4}]$, and $[1.0 \cdot 10^{-4}, 6.0 \cdot 10^{-4}]$ for each of the columns 1 to 4, respectively—the same as in the zooms of figure 16.

are applied, whereas DDS performs a little better when Mask 2 and 4 are applied. The reconstructed images indicate that RBC is superior to DDS for Mask 1 and 3, while DDS and RBC reconstructions with Mask 2 and 4 appear of comparable quality. Overall, the quantitative results support the previous conclusion that RBC is the best performing method.

5.5. Simulated variable-truncation data using complete chalk data set

In addition to the original chalk data set with VT data we have access to a comparable but complete chalk data set without data truncation. We can use this data set to simulate data truncation as we did for the synthetic Shepp–Logan phantom in order to assess the proposed methods on a real data set with the same complexity as the original truncated chalk data. Crucially, by using the complete chalk data, we can determine a full-angle FBP reconstruction for use as reference image in quantitative assessment.

Applying the four different sinogram masks, e.g. applying IZV to the complete sinogram and reconstructing from this, gives images with VT artifacts, seen in figure 15 and zoom-ins to a region of interest in the top row of figure 16. In particular, we note the similarity of the streaks occurring for mask 1 to the ones in the original truncated data set seen in figure 6, for which the experimental setup (metal bar radius and distance to sample) are comparable.

After applying methods RLA, DDS, and RBC, the streaks are reduced as seen in figure 16. Visually, all three methods improve on the IZV method. Due to noise it is difficult to see differences in the performance of the methods. However, for the second and especially the fourth column, DDS and RBC reproduce the dark round objects substantially better than RLA, for which objects are distinctively more square-like. The data cut-off in case 2 and 4 is flatter, meaning that more projections are truncated and thus discarded by RLA. It therefore makes sense that the largest advantage of DDS and RBC over RLA is seen in these cases.

Quantitative results in the form of local RMSE and SMD values are reported in table 3 calculated only over the zooms shown in figure 16 using the full-angle chalk reconstruction as reference. Both RMSE and SMD values are in alignment with the Shepp–Logan study in that RLA improves compared to IZV, while DDS and RBC provide further (and comparable) improvements.

We emphasize that the fourth case was deliberately chosen to be very challenging, which is reflected in the severe LA artifacts. Nevertheless, DDS and RBC manage to remove the VT artifacts and preserve the main features. Hence, DDS and RBC may in specific cases, where a coarse reconstruction is sufficient for the relevant imaging task, enable a successful imaging experiment in the face of severe beam occlusion.

6. Discussion

The results show that in all considered cases DDS and RBC outperform RLA. As the VT data cut-off becomes increasingly steep, i.e. for metal bars being further away from the sample, the number of VT artifacts is reduced as the data approaches the LA case without data truncation. In these cases, RLA discards only a small amount of data and the relative advantage of DDS and RBC over RLA diminishes, and at a certain point one may obtain equally good results by RLA.

In the other extreme, as the distance of the metal bars to the sample is decreased, the data cut-off becomes less steep and many, possibly all, projections will suffer from variable truncation. In this case RLA will have few, or even no,



Figure 16. Mask 1, 2, 3, and 4 imposed to full dataset of chalk, reconstructed by use of methods IZV, RLA, DDS, and RBC. Zoom to region of interest of size 409×409 pixels. The gray-scale windows are the same as those for the full reconstruction in figure 15.

Table 3. Quantitative image quality results for IZV, RLA, DDS and RBC reconstruction from simulated variable-truncation data usingthe complete chalk data set. RMSE and SMD values are computed locally over only the region of interest shown in the zooms in figure 16using the full-angle COM reconstruction as reference image.

	RMSE				SMD				
	Mask 1	Mask 2	Mask 3	Mask 4	Mask 1	Mask 2	Mask 3	Mask 4	
IZV	0.0628	0.1274	0.0789	$0.1845 \cdot 10^{-3}$	0.2790	0.7042	0.3798	$1.0406 \cdot 10^{-3}$	
RLA	0.0550	0.1188	0.0557	$0.1357 \cdot 10^{-3}$	0.2215	0.4917	0.2241	$0.5190 \cdot 10^{-3}$	
DDS	0.0514	0.0932	0.0518	$0.0977 \cdot 10^{-3}$	0.1956	0.3899	0.1994	$0.4009 \cdot 10^{-3}$	
RBC	0.0489	0.0966	0.0492	$0.1069 \cdot 10^{-3}$	0.1901	0.3869	0.1929	$0.3982 \cdot 10^{-3}$	

projections available, and therefore only allow a very poor reconstruction, if one at all. DDS and RBC, on the other hand, will be able to use the truncated projections and continue to produce a reconstruction. An indicative example of this was seen using Mask 4 in figures 14(d) and 15(d), but for thinner metal bars closer to the sample, the cut-off will be even flatter and result in a larger advantage of DDS and RBC over RLA. A typical answer to streaks in CT reconstructions is to increase the number of angular steps [10]. To illustrate the effect hereof on VT artifacts, we used 2, 4, 8, 16, and 32 times the number of projections originally used, namely 1800, and reconstructed from the Shepp–Logan sinogram with simulated VT by Mask 1 from figure 9(a). The top row in figure 17 shows zooms of sinograms with different numbers of projections. The bottom row shows zooms of



Figure 17. Top row: Zooms of Shepp–Logan phantom sinograms with varied number of angular measurements as indicated below each experiment. The zooms are as in figure 4(c), all in the angular range of 41.1 to 41.5 degrees. The number of projections varies in each sinogram in that each step corresponds to one projection. Bottom row: Zooms of corresponding reconstructions. The zooms are as in figure 10. Gray-scale window of the reconstructions is [0.4, 0.12]. A zoom inset in the figure (e) is seen with increased constrast (gray-scale window is here [0.71, 0.8]), indicating that vague shadows of the VT artifacts are still present.

the corresponding reconstructions, to be compared with figure 10(a). The number of streaks doubles, when the number of projection doubles, however with weakened strength for each projection number increase. Approximately 16 times more projections are needed to smooth out the VT artifacts. The streaks may eventually cancel, but even with 32 times more projections $(32 \cdot 1800 = 57600 \text{ projections})$ at the considered resolution of 2048, vague shadows from the VT artifacts are still present as seen in the zoom inset of figure 17(e). In practice, one cannot always increase the number of projections due to experimental constraints, e.g. one may wish to keep acquisition time down to enable doing more scans within an allocated beam time. For a fixed acquisition time, one could increase the number of projections while reducing exposure time of each projection accordingly. However, this would reduce the signal to noise ratio in each projection and thus reduce reconstruction quality. In practice a trade-off is chosen when setting up the scan (which was out of our control) and any VT artifacts have to be dealt with subsequently.

In the present work we deliberately restricted our focus to reconstruction using the FBP algorithm to avoid increasing reconstruction time and complexity. It is quite likely that algebraic or regularization-based reconstruction methods can be developed to achieve further improvements. However, this may entail longer development time, more difficult integration in beamline reconstruction software, as well as longer computing time. We leave this for future investigations.

We also did not investigate sinogram inpainting methods, but focused on enabling use of all the existing data through handling of the sinogram discontinuities, without introducing artificial data which might produce new and more subtle artifacts. Nevertheless, a comparison of the methods proposed here with sinogram inpainting methods is an interesting future direction. The proposed methods are completely automatic requiring no user input, with the sole exception of setting the threshold value below which to discard data values. This value is highly dependent on the data set and should be chosen to preserve as much of the data as possible, while discarding unreliable data. Through analysis of a histogram of the sinogram values it should be possible to determine a suitable threshold automatically, but developing a robust method for this was beyond the scope of the present work.

In the present work we focused on the detector-directed discontinuities to reduce VT artifacts, while leaving LA artifacts unreduced. We believe a straightforward extension would be to combine the methods proposed here with LA reduction techniques such as the angular smoothing proposed in [15] to achieve simultaneous LA and VT artifact reduction.

While the analysis and mathematical model for VT data developed are specific for a parallel-beam configuration, the same principles can be used to model other cases such as a cone-beam geometry which will also experience streaks from VT data. The proposed artifact reduction methods are not tied to the parallel-beam geometry and can be directly applied to other configurations. The methods are also not specific to the particular data cut-off but applicable to general data cut-offs, including asymmetric truncation, internal gaps and combinations hereof. As such, the methods may be capable of reducing VT artifacts across a range of *in situ* x-ray tomography setups.

7. Conclusion

In an *in situ* x-ray tomography experiment of fluid flow in porous chalk the reconstructed cross sections were found to suffer from drastic streak artifacts with certain preferred directions. The streaks were found to arise from detectordirected discontinuities in the sinogram introduced by variable truncation of projections caused by partial occlusion of the x-ray beam by four metal bars of the percolation cell used in the experiment. A mathematical model describing variable truncation as a function of metal bar radius and distance to sample was derived and verified numerically and in comparison with the experimental data. Using the mathematical model, the origin and location of variable-truncation artifacts were established.

Three methods for variable-truncation artifact reduction were proposed addressing in different ways the detectordirected discontinuities: by discarding all truncated projections (reduction to limited angle), by smoothing the discontinuities (detector-directed smoothing), and by enforcing a zero derivative at the discontinuities (reflexive boundary condition).

All methods successfully reduced the streak artifacts in the original variable-truncation chalk data set. In more extensive tests across a range of simulated variable-truncation data cases using both synthetic and real data, the reflexive boundary condition method was found to outperform the other methods in terms of both qualitative and quantitative assessment of reconstruction image quality. The advantage of this method and the detector-directed smoothing method was found to be larger in more challenging cases involving larger proportion of truncated projections.

The proposed artifact-reduction methods are conceptually simple, computationally efficient and intended to be easy to incorporate with existing FBP implementations and apply to large real synchrotron data sets. The methods can handle quite general data cut-offs and may enable successful reconstruction across a variety of conceivable *in situ* x-ray tomography experiments with partial beam occlusion leading to variabletruncation data.

Acknowledgments

JSJ was supported by the project 'High-Definition Tomography' funded by Advanced Grant No. 291405 from the European Research Council. We thank the Japan Synchrotron Radiation Research Institute for the allotment of beam time on beamline BL20XU of SPring-8 (Proposal 2015A1147). Valuable and constructive discussions with Robinson professor Eric Todd Quinto have been a great help in the development of this research work. LB and JS acknowledge funding from the Innovation Fund Denmark and Mærsk Oil and Gas A/S, for the P³ Project. The work of JF was partially supported by the HC Ørsted Postdoc programme, co-funded by Marie Curie Actions at the Technical University of Denmark. We thank Henning Osholm Sørensen for giving us access to the experimental tomography data.

ORCID iDs

Jakob S Jørgensen https://orcid.org/0000-0001-9114-754X

References

 De Kock T, Boone M A, De Schryver T, Van Stappen J, Derluyn H, Masschaele B, De Schutter G and Cnudde V 2015 A pore-scale study of fracture dynamics in rock using x-ray micro-CT under ambient freeze–thaw cycling *Environ. Sci. Technol.* **49** 2867–74

- [2] Cai B, Lee P D, Karagadde S, Marrow T J and Connolley T 2016 Time-resolved synchrotron tomographic quantification of deformation during indentation of an equiaxed semisolid granular alloy *Acta Mater.* **105** 338–46
- [3] King A, Ludwig W, Herbig M, Buffi J-Y, Khan A A, Stevens N and Marrow T J 2011 Three-dimensional *in situ* observations of short fatigue crack growth in magnesium *Acta Mater.* 59 6761–71
- [4] Yang Z, Ren W, Sharma R, McDonald S, Mostafavi M, Vertyagina Y and Marrow T J 2017 *In situ* x-ray computed tomography characterisation of 3D fracture evolution and image-based numerical homogenisation of concrete *Cement Concrete Compos.* **75** 74–83
- [5] Sidky E Y and Pan sep X 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* 53 4777–807
- [6] Batenburg K J and Sijbers J 2011 DART: a practical reconstruction algorithm for discrete tomography *IEEE Trans. Image Process.* 20 2542–53
- [7] Gao H, Yu H, Osher S and Wang G 2011 Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM) *Inverse Problems* 27 115012
- [8] Andersen A H and Kak A C 1984 Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm *Ultrason. Imaging* 6 81–94
- [9] Boas F E and Fleischmann D 2012 CT artifacts: causes and reduction techniques *Imaging Med.* 4 229–40
- [10] Buzug T H 2008 Computed Tomography: From Photon Statistics to Modern Cone-Beam CT (Berlin: Springer)
- [11] De Man B, Nuyts J, Dupont P, Marchand G and Suetensl P 1999 Metal streak artifacts in x-ray computed tomography: simulation study *IEEE Trans. Nucl. Sci.* 46 691–6
- [12] Veldkamp W J H, Joemai R M S, van der Molen A J and Geleijns J 2010 Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT Med. Phys. 37 620–8
- [13] Boas F E and Fleischmann D 2011 Evaluation of two iterative techniques for reducing metal artifacts in computed tomography *Radiology* 259 894–902
- [14] Gjesteby L, Man B D, Jin Y, Paganetti H, Verburg J, Giantsoudi D and Wang G 2016 Metal artifact reduction in CT: where are we after four decades? *IEEE Access* 4 5826–49
- [15] Frikel J and Quinto E T 2013 Characterization and reduction of artifacts in limited angle tomography *Inverse Problems* 29 125007
- [16] Frikel J 2013 Sparse regularization in limited angle tomography Appl. Comput. Harmon. Anal. 34 117–41
- [17] Delaney A H and Bresler Y 1998 Globally convergent edgepreserving regularized reconstruction: an application to limited-angle tomography *IEEE Trans. Image Process.* 7 204–21
- [18] Persson M, Bone D and Elmqvist H 2001 Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography *Phys. Med. Biol.* 46 853
- [19] Rantala M, Vanska S, Jarvenpaa S, Kalke M, Lassas M, Moberg J and Siltanen S 2006 Wavelet-based reconstruction for limited-angle x-ray tomography *IEEE Trans. Med. Imaging* 25 210–7
- [20] Noo F, Clackdoyle R and Pack J D 2004 A two-step Hilbert transform method for 2D image reconstruction *Phys. Med. Biol.* 49 3903–23
- [21] Zou Y, Pan X and Sidky E Y 2005 Image reconstruction in regions-of-interest from truncated projections in a reduced fan-beam scan *Phys. Med. Biol.* 50 13–27

- [22] Müller M and Arce G R 1996 Truncation artifacts in tomographic reconstructions from projections Appl. Opt. 35 3902–14
- [23] Van G 2009 Towards accurate image reconstruction from truncated X-ray CT projections *PhD Thesis* University of Antwerp, Antwerp, The Netherlands
- [24] Yang Y, Hakim S S, Bruns S, Dalby K N, Uesugi K, Stipp S L S and Sørensen H O 2016 Wormholes grow along paths with minimal cumulative surface (arXiv:1704.01062)
- [25] Uesugi K, Hoshino M, Takeuchi A, Suzuki Y, Yagi N and Nakano T 2010 Development of fast (sub-minute) microtomography AIP Conf. Proc. 1266 47–50
- [26] Borg L, Jørgensen J S and Sporring J 2017 Towards characterizing and reducing artifacts caused by varying projection truncation *Technical Report* 2017/1, Department

of Computer Science, University of Copenhagen (available online: http://static-curis.ku.dk/portal/files/181386847/ teknisk_rapport_leise_borg1.pdf)

- [27] Barannyk L L, Frikel J and Nguyen L V 2016 On artifacts in limited data spherical Radon transform: curved observation surface *Inverse Problems* 32 015012
- [28] Nguyen L V 2015 How strong are streak artifacts in limited angle computed tomography? *Inverse Problems* 31 055003
- [29] Hansen P C, Nagy J G and O'Leary D P 2006 Deblurring Images: Matrices, Spectra, and Filtering (Philadelphia: SIAM)
- [30] Kanters F M W 2007 Towards object-based image editing PhD Thesis Technische Universiteit Eindhoven Eindhoven, The Netherlands

3.2 Supplementary material

This section provides the link between the raw data and the sinogram (Section 3.2.1), a description of three types of artifacts (Section 3.2.2), and an algebraic approach for handling truncated-projection artifacts (Section 3).

3.2.1 Linking the sinogram to the raw data

The descriptions in this section are based on the work presented in the technical report [4]. As described in our article *Reduction of variable-truncation artifacts from beam occlusion during in situ X-ray tomography* in Section 3.1 the transmissions are zero due to shadowing metal bars in the setup, see Figure 2b in the article. Please note that transmissions are what is measured by the detector, whereas projections are the negative logarithm of the transmissions. The recorded 2D transmissions are measured by a detector panel consisting of 2048×2048 detector elements¹. Figure 3.1 shows transmissions for six angular measurement with numbers, $a = \{438, 442, 446, 459, 454, 462\}$ out of $a_{\text{max}} = 1800$ angular measurements, covering 180 degrees. The transmissions are already corrected for dark and white fields (see Section 2.3 in the article *Reduction of variable-truncation artifacts from beam occlusion during in situ X-ray tomography* for details). The figure shows that the transmissions are increasingly truncated by the metal bars over the angular measurements until the beam is fully occluded in measurement number 462.



Figure 3.1: 2D transmissions, corrected for dark and white fields. The dimensions are 2048×2048 according to the number of detector elements. Data shown in this figure are obtained at SPring-8.

¹We thank the Japan Synchrotron Radiation Research Institute for the allotment of beam time on beamline BL20XU of SPring-8 (Proposal 2015A1147).

Picking out row number 1000 (approximately the ones in the middle) of all 1800 transmissions, and collecting them side-by-side, yields the transmission sinogram for slice number 1000. Each transmission slice (one for each angle) appear as a column in the 1800×2048 transmission sinogram, $T(\phi, p)$, shown in Figure 3.2. The angular variable is on the abscissa and the detector displacement is on the ordinate (which is how sinograms are presented through the entire dissertation).



Figure 3.2: Transmission sinogram for slice number 1000, which is obtained at SPring-8.

3.2.2 Artifacts

Examples of FBP reconstructions (MATLAB, version R2017a) are shown in Figure 3.3 from a range of sinograms. The sinograms are shown in the top row, and from left to right, a full sinogram, a limited-angle sinogram, and a sinogram with truncated projections are shown. In the bottom row, the respective reconstructions are shown. The sinograms were obtained from 51 angles over 180 degrees from a 200×200 image of the Shepp-Logan phantom. The reconstructions have the same dimensions as the original image.



Figure 3.3: Sinograms are shown in the top row, from left to right: full sinogram, limitedangle sinogram, sinogram with truncated projections. The bottom row shows the corresponding FBP reconstructions.
There are many types of artifacts in the image. The pronounced artifacts outside the phantom head in the left reconstruction is caused by the sparse number of angular measurements. The LA artifacts in the center reconstruction of the figure are caused by the LA cut-offs and missing data in the sinogram. They appear along the projection lines L with angles perpendicular to the cut-off angles in the sinogram but emerge only from edges in the reconstruction (see [29] for details). Further, we see that edges in the reconstructed image with angles corresponding to the angles of the missing projections are not reconstructed. The rightmost reconstruction of the figure has truncated projection leads to a streak artifact along the corresponding projection line. This means that:

- 1. The artifacts appear as streaks independently of the object
- 2. The slopes of the streak artifacts are perpendicular to the angle by which the truncated projections appear in the sinogram.
- 3. Their translative positions in the reconstructions correspond to the detector-wise translation of the truncations in the sinogram.

A generalization of these observations and other types of missing data artifacts (not artifacts caused by low resolution) are presented in our article *Characterizing Reconstruction* Artifacts from Arbitrary Incomplete X-ray CT Data, in Section 4.1.

3.2.3 An algebraic approach

The left image in Figure 3.4 shows a reconstruction based on the sinogram with truncated projections in the right panel of Figure 3.3. The reconstruction is based on the Cimmino implementation in AIRtools [23] and was stopped after 50 iterations. It is very similar to the rightmost FBP reconstruction in Figure 3.3; The reconstruction qualities and appearances do not differ much between the reconstruction methods. However, the Cimmino reconstruction is a little more smoothed than that of the FBP reconstruction. This is probably due to the relatively low iteration number, and the regularization in the Cimmino method. The right image of the figure is a Cimmino reconstruction - also based on the sinogram with truncated projections to the right in Figure 3.3. In this version, the rows in the A-matrix corresponding to the missing data points were removed prior to reconstruction.



Figure 3.4: Left: Cimmino reconstruction (similar to the rightmost plot in the middle row in Figure 3.4). Right: Cimmino reconstruction where rows in A-matrix corresponding to the missing data were removed prior to reconstruction.

The results in the figures imply that zeros in a sinogram do not represent missing data. This is not an issue in standard limited-angle problems, but definitely is for truncated-projection cases; backprojecting a projection of zeros in FBP adds nothing to the reconstruction. The Cimmino method does not need full projections, however, FBP does, which is one of the reasons why algebraic methods handle truncated projections better than FBP. Another advantage of the Cimmino method is that it does not cause overexposure as FBP does; Backprojecting the truncated projections adds to the pixel values only across parts of the reconstruction and not on the zero-values parts.

3.3 Summary

In Section 3.1 reconstructions from an in situ X-ray tomography experiment of fluid flow in porous chalk was analysed. The reconstructions suffered from drastic streak artifacts, which were found to arise from variable truncations of projections. The truncations were caused by partial occlusion of the X-ray beam by four metal bars in the experimental setup. A mathematical model describing variable truncation as a function of metal bar radius and distance to sample was derived and verified. Based on this model, various masks were used to simulate variable-truncation data and imposed on synthetic and real data. We proposed four methods for variable-truncation artifact reduction and the most successful method was the Reflexive Boundary Condition method. The four methods are all based on the FBP algorithm, which is the most common reconstruction method in routine use at synchrotrons. This is due to its robustness, computational efficiency, and well-understood behavior. The proposed methods are conceptually simple, computationally efficient and intended to be easy to incorporate with existing FBP implementations and apply to large real synchrotron data sets. The methods can handle a variety of in situ X-ray tomography experiments with partial beam occlusion.

An example of an algebraic solution (Section 3.2.3) suggests removal of rows in the A matrix and b vector corresponding to those projection lines affected by the beam occlusion. It is quite likely that algebraic or regularization-based reconstruction methods can be developed to handle variable-truncation artifacts. However, this typically involves tweaking of several parameters, more difficult integration in beamline reconstruction software, as well as longer computing time. We leave this for future investigations.

Characterizing artifacts from incomplete data 4

This section is a theoretical approach to describing reconstruction artifacts from arbitrary incomplete data. Section 4.1 contains the article *Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data* and Section 4.2 is a summary of the work. Following is a presentation of some of the variables and outlines in the article, supported by example figures.

In the article, we consider only continuous problems where the object is a function on \mathbb{R}^2 and the data (sinogram) is a function on $[0, 2\pi] \times \mathbb{R}$, respectively, rather than finite data on a grid. Therefore, we do not analyze artifacts caused by numerical discretization (as exemplified in the left panel of Figure 3.3 in this thesis). Each projection line, $L(\varphi, p)$, in the object space correlates to one single point in the data space, as illustrated in Figure 4.1. The two projection lines in the figure are parallel.



Figure 4.1: Linking object space and data space.

The projection lines defined in this thesis in Section 2.1,

$$L(\varphi, p) = \{x \in \mathbb{R}^2 : x_1 \cos \varphi + x_2 \sin \varphi = p\}$$

are expressed a little differently in the article, namely:

$$L(\varphi, p) = \{ x \in \mathbb{R}^2 : x \cdot \bar{\theta}(\varphi) = p \},\$$

where $\bar{\theta}(\varphi) = (\cos(\varphi), \sin(\varphi))$ is the unit vector in the direction of φ . A line $L(\varphi, p)$ contains the point $p(\varphi)\bar{\theta}(\varphi)$ and is parallel to $\theta^{\perp}(\varphi) = (-\sin(\varphi), \cos(\varphi))$, which is the unit vector $\pi/2$ radians counterclockwise from $\bar{\theta}(\varphi)$, as sketched in Figure 4.2. In the figure, the dots represent the points $p(\varphi)\bar{\theta}(\varphi)$ and x_b (x_b will be introduced later).



Figure 4.2: Correlating the variables in the geometrical setting of the object space.

The complete dataset, $Rf(\varphi, p)$, is given over all $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$, where the incomplete dataset is the subset, A, of $[0, 2\pi] \times \mathbb{R}$. The characteristic function of the set A is the function that is equal to one on A and zero outside of the set, and is denoted $\mathbb{1}_A$. The incomplete CT data can then be modeled as

$$R_A f(\varphi, p) = \mathbb{1}_A(\varphi, p) R f(\varphi, p).$$

 $\mathbb{1}_A(\varphi, p)Rf(\varphi, p)$ sets the data to zero outside of A since we don't know the data there. An example of the subset A is sketched in Figure 4.3. $R_A f$ equals Rf inside the subset A and is zero off of A. In the figure, the boundary of A is indicated as $\mathrm{bd}(A)$. I will in a bit explain the other elements in the figure.



Figure 4.3: The subset A, the boundary of A is bd(A), the wavefront set of A is $WF(\mathbb{1}_A)$, and examples of the 3 cases where streaks may occur in the reconstruction, depending on $Rf(\varphi, p)$ on bd(A).

Microlocal analysis provides a mathematical framework for describing the artifacts. In particular, the definitions of singularities, wavefront sets, and how singularities are transformed by Fourier integral operators [30, 31] (in FBP) are useful in this context:

• Singularities are edges or density jumps described by the tuple (t,ξ) , where t is the singular position and ξ is the singular direction. Consider bd(A) in Figure 4.3. Let $\mathbb{1}_A$

be the function that is equal to one on A and zero outside of A, then the points on bdA are all singular positions. Take a singular position, $t_0 = (\varphi_0, p_0) \in bdA$. $\xi_0 \neq 0 \in bd(A)$, is a vector pointing in the normal direction to the boundary of A at the singular position. The black arrows in Figure 4.3 indicate singular directions and the red arrows point to the describing text corresponding to the geometric examples in the figure.

- Wavefront sets are sets of the singularities, as exemplified in Figure 4.3 and indicated by WF($\mathbb{1}_A$). Wavefront sets are in this context used to state that artifacts only occur on lines $L(\varphi, p)$ with $(\varphi, p) \in \mathrm{bd}A$.
- The reconstruction from a complete dataset is $\mathcal{L}f = R^*(\Lambda Rf)$, where R^* is the backprojection operator and Λ is the standard FBP filter. The reconstruction from an incomplete data set is $\mathcal{L}_{\mathcal{A}}f = R^*(\Lambda R_A f) = R^*(\Lambda \mathbb{1}_A Rf)$. Singularities in a reconstruction, may it be $\mathcal{L}f$ or $\mathcal{L}_{\mathcal{A}}f$, can ONLY come from singularities in the original dataset Rf or $R_A f$.

The artifacts are either object-dependent or object-independent, determined by the geometry of bd(A). Object-dependent artifacts are caused by the object, f, being scanned AND the geometry of bd(A), whereas object-independent artifacts solely depend on the geometry of bd(A) (if Rf is non-zero). The article describes 3 main cases of artifacts. If we consider a point $(\varphi_0, p_0) \in bd(A)$, then artifacts may occur in following cases:

Case I: The boundary of A is smooth with a finite slope. An artifact can occur along the curve:

$$x_b = x_b(\varphi) = p(\varphi)\overline{\theta}(\varphi) + p'(\varphi)\theta^{\perp}(\varphi).$$

See Figure 4.2 for an example of the position of point x_b , and please note that $x_b \in L(\varphi, p)$. This type of artifact is object independent because it can occur whether f is smooth normal to $L(\varphi_0, p_0)$ or not. Being smooth normal to $L(\varphi_0, p_0)$ means that f has no wavefront set $(x, \pm \bar{\theta}(\varphi))$ for any x on $L(\varphi_0, p_0)$ (note that $\bar{\theta}(\varphi)$ is normal to $L(\varphi_0, p_0)$). An example of this is shown in Figure 4.1: f is smooth normal to the projection line $L(\varphi_l, p_m)$.

Case II: The boundary of A is smooth with a finite slope AND f has a singularity normal to $L(\varphi_0, p_0)$. Then, an artifact can occur along the line $L(\varphi_0, p_0)$. This type of artifact is object dependent because it will not occur unless f has a singularity normal to $L(\varphi_0, p_0)$. In Figure 4.1 f has a singularity normal to $L(\varphi_l, p_n)$, indicated by the red arrow.

Case III: The boundary of A is non-smooth at $(\varphi_0, p_0) \in bd(A)$. Then, an artifact can occur all along $L(\varphi_0, p_0)$. If bd(A) comes to a corner at (φ_0, p_0) , f is smooth normal to $L(\varphi_0, p_0)$, AND $Rf(\varphi, p) \neq 0$, then an artifact occurs along the line $L(\varphi_0, p_0)$. This type of artifact is object independent because it can occur whether f is smooth normal to $L(\varphi_0, p_0)$ or not. A corner in this context is when a singularity has *several* singular directions rather than only one. The "star" of arrows in Figure 4.3 indicate a singularity with several directions.

Objects that have singularities normal to a line in the boundary of the image might not contribute artifacts. This is the reason why we say that artifacts *can* occur in the three cases just described. A remark to case I: If the slope of bd(A) at (φ_0, p_0) is small enough, i.e.

$$|p'(\varphi_0)| < \sqrt{1 - p_0^2}$$

then the artifact curve is visible in the reconstruction, otherwise it may appear outside the typical reconstruction area (outside the unit disk). This is exemplified in Figure 3A in our

article Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data. In the limited-angle case $p'(\varphi_0)$ is infinitely large meaning that $x_b(\varphi_0)$ is a point infinitely far away. Limited-angle artifacts are object dependent and such artifacts occur along the line L.

Considering artifacts in case II and III, the general rule is that an artifact can occur if a *corner* in the sinogram, $R_A f$, is present. In a complete sinogram, Rf, corners do not occur due to the nature of the acquisition scheme (in a continuous setting). Artifacts emerging because of corners are thus only associated with points on bdA, $(\varphi, p) \in bdA$.

4.1 Article # 2: Characterizing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data

Analyzing Reconstruction Artifacts from Arbitrary Incomplete X-ray CT Data*

Leise Borg[†], Jürgen Frikel[‡], Jakob Sauer Jørgensen[§], and Eric Todd Quinto[¶]

- Abstract. This article provides a mathematical analysis of artifacts from arbitrary incomplete X-ray computed tomography (CT) data when using the classical filtered backprojection algorithm. We prove that artifacts arise from points at the boundary of the data set. Our results show that, depending on the geometry of this boundary, two types of artifacts can arise: object-dependent and object-independent artifacts. The object-dependent artifacts are generated by singularities of the object being scanned and these artifacts can extend all along lines. They generalize the streak artifacts observed in limited-angle tomography. The article also characterizes two new types of artifacts that are essentially independent of the object; they occur along lines if the boundary of the data set is not smooth at a point and along curves if the boundary is smooth locally. In addition to the geometric description of artifacts, the article apply to the well-known incomplete data problems, including limited-angle and region-of-interest tomography, as well as to unconventional X-ray CT imaging setups. Reconstructions from simulated and real data are analyzed to illustrate our theorems, including the reconstruction that motivated this work–a synchrotron data set in which artifacts appear along lines that have no relation to the object.
- Key words. X-ray tomography, incomplete data tomography, limited angle tomography, region of interest tomography, reconstruction artifact, wavefront set, microlocal analysis, Fourier integral operators

AMS subject classifications. 44A12, 92C55, 35S30, 58J40

1. Introduction. Over the past decades computed tomography (CT) has established itself as a standard imaging technique in many areas, including materials science and medical imaging. Its principle is based on collecting numerous X-ray measurements of the object along all possible directions (lines) and then reconstructing the interior of the object by using an appropriate mathematical algorithm. In classical tomographic imaging setups, this procedure works very well because the data can be collected all around the object (i.e., the data are complete) and standard reconstruction algorithms, such as filtered backprojection (FBP), provide accurate reconstructions [29, 36]. However, in many recent applications of CT, some data are not available, and this leads to incomplete (or limited) data sets. The reasons for data incompleteness might be health-related (e.g., to decrease dose) or practical (e.g., when the scanner cannot image all of the object as in digital breast tomosynthesis).

Classical incomplete data problems have been studied from the beginning of tomography, including *limited-angle tomography*, where the data can be collected only from certain view-

^{*}Submitted to the editors DATE.

Funding: This work was partially funded by: Innovation Fund Denmark and Maersk Oil and Gas (Borg); HC Ørsted Postdoc programme, co-funded by Marie Curie Actions at the Technical University of Denmark (Frikel); ERC Advanced Grant No. 291405 and EPSRC Grant EP/P02226X/1 (Jørgensen); Otto Mønsteds Fond, Tufts FRAC and NSF grants DMS 1311558 and DMS 1712207 (Quinto)

[†]University of Copenhagen (lebo@di.ku.dk).

[‡] OTH Regensburg, (juergen.frikel@oth-regensburg.de)

[§] University of Manchester, (jakob.jorgensen@manchester.ac.uk).

[¶]Tufts University, (todd.quinto@tufts.edu, http://math.tufts.edu/faculty/equinto/).



Figure 1: Left: A small part of the sinogram of the chalk sample analyzed in Section 7. Notice the rough boundary. Right: Small central section of a reconstruction of the chalk. Monochromatic parallel beam data were taken of the entire cross section of the chalk over 1800 views covering 180 degrees, and there were 2048×2048 detector elements with a 0.5 mm field of view, providing micrometer resolution of the sample. Data [51] obtained, with thanks from the Japan Synchrotron Radiation Research Institute from beam time on beamline BL20XU of SPring-8 (Proposal 2015A1147). For more details, see Section 7 and [5, ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved].

angles [27, 22]; interior or region-of-interest (ROI) tomography, where the X-ray measurements are available only over lines intersecting a certain subregion of the object [9]; or exterior tomography, where only measurements over all lines outside a certain subregion are available [28, 41].

In addition, new scanning methods generate novel data sets, such as the synchrotron experiment [6, 5] in Section 7 that motivated this research. That reconstruction, in Figure 1, includes dramatic streaks that are independent of the object and were not described in the mathematical theory at that time but are explained by our main theorems. A thorough practical investigation of this particular problem was recently presented in [5].

Regardless of the type of data incompleteness, in most practical CT problems a variant of FBP is used on the incomplete data to produce reconstructions [36]. It is well-known that incomplete data reconstruction problems that do not incorporate a priori information (as is the case in all FBP type reconstructions) are severely ill-posed. Consequently, certain image features cannot be reconstructed reliably [40] and, in general, artifacts¹, such as the limitedangle streaks in Figure 2 can occur. Therefore, reconstruction quality suffers considerably, and this complicates the proper interpretation of images.

1.1. Related research in the mathematical literature. Our work is based on microlocal analysis, a deep theory that describes how singularities are transformed by Fourier integral operators, such as the X-ray transform. In the 1990s microlocal analysis was used to characterize visible singularities from X-ray CT data and their strength was given in [40, 24].

 $^{^{1}}Artifacts$ are image features (singularities), e.g. streaks or points, that are added by the algorithm and that are not part of the original object.

Subsequently, artifacts were extensively studied in the context of limited-angle tomography, e.g., [22, 12]. The strength of added artifacts in limited-angle tomography was analyzed in [31]. Similar characterizations of artifacts in limited-angle type reconstructions have also been derived for the generalized Radon line and hyperplane transforms as well as for other Radon transforms (such as circular and spherical Radon transform), see [13, 14, 32, 1, 33].

Metal in objects can corrupt CT data and create dramatic streak artifacts [3]. This can be dealt with as an incomplete data problem by excluding data over lines through the metal. Recently, this problem has been mathematically modeled in a sophisticated way using microlocal analysis in [37, 44, 35], all of which are in the spirit of our article. A related problem is studied in [34], where the authors develop a streak reduction method for quantitative susceptibility mapping (see also [7]). Moreover, microlocal analysis has been used to characterize properties of related integral transforms in pure and applied settings [4, 15, 47, 10, 43].

To the best of our knowledge, the mathematical literature up until now, e.g., [22, 31, 12, 13, 33] used microlocal and functional analysis to explain streak artifacts along lines that are generated by singularities of the object, *and* they were for specific problems, primarily limited-angle tomography. Even classical setups, such as region-of-interest tomography, had not yet been thoroughly explored microlocally, not to mention nonstandard imaging setups such as the one presented in Figure 1.

1.2. Basic mathematical setup and our results. In this article, we present a unified approach to characterize reconstruction artifacts for arbitrary incomplete X-ray CT data that are caused by the choice of data set. We not only consider all of the above mentioned classical incomplete data problems but also emerging imaging situations with incomplete data. To this end, we employ tools from microlocal analysis and the calculus of Fourier integral operators.

If f is the density of the object to be reconstructed, then each CT measurement is modeled by a line integral of f over a line in the data set. As we will fully describe in Section 2.1, we parametrize lines by $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$, and the CT measurement of f over the line $L(\varphi, p)$ is denoted $Rf(\varphi, p)$. With complete data, where $Rf(\varphi, p)$ is given over all $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$, accurate reconstructions can be produced by the FBP algorithm. In incomplete data CT problems, the data are taken over lines $L(\varphi, p)$ for (φ, p) in a proper subset, A, of $[0, 2\pi] \times \mathbb{R}$ and, even though FBP is designed for complete data, it is still one of the preferred reconstruction methods in practice. As a result, incomplete data CT reconstructions usually suffer from artifacts.

We prove that incomplete data artifacts arise from points at the boundary or "edge" of the data set, bd(A), and we show that there are two types of artifacts: *object-dependent* and *object-independent* artifacts. The object-dependent artifacts are caused by singularities of the object being scanned. In this case, artifacts can spread all along a line $L(\varphi_0, p_0)$ (i.e., a streak) if $(\varphi_0, p_0) \in bd(A)$ and if there is a singularity of the object along the line (such as a jump or object boundary tangent to the line)—this singularity of the object "generates" the artifact. The streak artifacts observed in limited-angle tomography are special cases of this type of artifact.

In addition, we characterize two new phenomena that are (essentially) independent of the object being scanned. If the boundary of A is smooth near a point $(\varphi_0, p_0) \in bd(A)$, then we prove that artifacts can appear in the reconstruction along curves generated by bd(A) near

 (φ_0, p_0) , and they occur whether the object being scanned has singularities or not. We also prove if bd(A) is not smooth (see Definition 3.2) at a point (φ_0, p_0) then, independent of the object, an artifact line can be generated all along $L(\varphi_0, p_0)$.

We will illustrate our results using classical problems including limited-angle tomography and region-of-interest tomography, as well as problems with novel data sets, including the synchrotron data set in Figure 1. In addition to the geometric characterization of artifacts, we also provide a characterization of their strength in Sobolev scale, and we review a general method to suppress the artifacts.

1.3. Organization of the article. In Section 2, we provide notation and some of the basic ideas about distributions and wavefront sets. In Section 3 we give our main theoretical results, and in Section 4, we apply them to classical examples to explain added artifacts. In Section 5, we describe the strength of added artifacts in Sobolev scale. Then, in Section 6, we describe a straightforward way to decrease the added artifacts. We provide more details of the synchrotron experiment in Section 7 and observations and generalizations in Section 8. Finally, in the appendix, we give some technical theorems and then prove the main theorems.

2. Mathematical basis. Much of our theory can be made rigorous for distributions of compact support (see [11, 45] for an overview of distributions), but we will consider only functions in $L^2(D)$, the set of absolutely square integrable functions supported in the open unit disk $D = \{x \in \mathbb{R}^2 : ||x|| < 1\}$. This setup is realistic in practice, and our theorems are simpler in this case than for general distributions. Remark A.4 provides perspective on this.

2.1. Notation. For $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$, we define

(2.1)
$$\overline{\theta}(\varphi) = (\cos(\varphi), \sin(\varphi)) \text{ and} \\ \theta^{\perp}(\varphi) = (-\sin(\varphi), \cos(\varphi)),$$

so $\overline{\theta}(\varphi)$ is the unit vector in the direction of φ and $\theta^{\perp}(\varphi)$ is the unit vector $\pi/2$ radians counterclockwise from $\overline{\theta}(\varphi)$. The line perpendicular to $\overline{\theta}(\varphi)$ and containing $p\overline{\theta}(\varphi)$ is denoted

(2.2)
$$L(\varphi, p) = \left\{ x \in \mathbb{R}^2 : x \cdot \overline{\theta}(\varphi) = p \right\}$$

This line is parameterized by $t \mapsto p\overline{\theta}(\varphi) + t\theta^{\perp}(\varphi)$. Note that $L(\varphi, p) = L(\varphi + \pi, -p)$.

We define the X-ray transform or Radon line transform of $f \in L^2(D)$ to be

(2.3)
$$Rf(\varphi, p) = \int_{-\infty}^{\infty} f(p\overline{\theta}(\varphi) + t\theta^{\perp}(\varphi)) \,\mathrm{d}t$$

For functions g on $[0, 2\pi] \times \mathbb{R}$, the dual Radon transform or backprojection operator is defined

(2.4)
$$R^*g(x) = \int_0^{2\pi} g(\varphi, x \cdot \overline{\theta}(\varphi)) \,\mathrm{d}\varphi.$$

2.2. Wavefront sets. In this section, we define some important concepts needed to describe singularities in general. Sources, such as [11], provide introductions to microlocal analysis. Generally cotangent spaces are used to describe microlocal ideas, but they would

complicate this exposition, so we will identify a covector $(x, \xi dx)$ with the associated ordered pair of vectors (x, ξ) . The book chapter [23] provides some basic microlocal ideas and a more elementary exposition adapted for tomography.

The wavefront set is a deep concept that makes singularities of functions precise, and we take the definition from [11].

Definition 2.1 (Wavefront set). Let $x_0 \in \mathbb{R}^2$ and $\xi_0 \in \mathbb{R}^2 \setminus \mathbf{0}$. A cutoff function at x_0 is any C^{∞} -function of compact support that is nonzero at x_0 . Let f be a locally integrable function. We say f is smooth at x_0 in direction ξ_0 if there is a cutoff function ψ at x_0 and an open cone V containing ξ_0 such that the Fourier transform $\mathcal{F}(\psi f)(\xi)$ is rapidly decaying at infinity.²

If f is not smooth at x_0 in direction ξ_0 , then we say f has a singularity at x_0 in direction ξ_0 . The wavefront set of f is the set WF(f) of all such (x_0, ξ_0) .

For example, if A is a subset of the plane with a smooth boundary, and f is equal to 1 on A and 0 off of A, then WF(f) is the set of all points (x, ξ) where x is in the boundary of A and ξ is normal to the boundary at x.

We parametrize lines on $[0, 2\pi] \times \mathbb{R}$, rather than $S^1 \times \mathbb{R}$, because, in practice, the data space consists of points $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$ and the sinogram is represented as a picture in the (φ, p) -plane, not on a cylinder. We will consider only distributions on $[0, 2\pi] \times \mathbb{R}$ that are restrictions of distributions for $(\varphi, p) \in \mathbb{R}^2$ that are 2π -periodic in φ . The Radon transform and our other operators on $[0, 2\pi] \times \mathbb{R}$ are all 2π -periodic in φ , and so this is no real restriction. For this reason, we can use the same definition of wavefront set as for \mathbb{R}^2 , and we just apply it to the 2π -periodic extensions.

Our next definition allows us to express efficiently the correspondence between singularities of the object and of its Radon transform.

Definition 2.2. Let $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$. The 6 normal space of the line $L(\varphi, p)$ is

(2.5)
$$N(L(\varphi, p)) = \left\{ (x, \omega \overline{\theta}(\varphi)) : x \in L(\varphi, p), \, \omega \in \mathbb{R} \right\}$$

and the set of singularities of f normal to $L(\varphi, p)$ is

(2.6)
$$WF_{L(\varphi,p)}(f) = WF(f) \cap (N(L(\varphi,p))).$$

Let f be a locally integrable function on \mathbb{R}^2 . We say f is smooth normal to the line $L(\varphi, p)$ if $WF_{L(\varphi,p)}(f) = \emptyset$.

For $x_0 \in \mathbb{R}^2$, we let

$$WF_{x_0}(f) = WF(f) \cap (\{x_0\} \times \mathbb{R}^2).$$

Now, let g be a locally integrable function on $[0, 2\pi] \times \mathbb{R}$ and $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$. We define

(2.7)
$$WF_{(\varphi,p)}(g) = WF(g) \cap \left(\{(\varphi,p)\} \times \mathbb{R}^2\right).$$

It is important to understand each set introduced in Definition 2.2: $N(L(\varphi, p))$ is the set of all (x,ξ) such that $x \in L(\varphi, p)$ and the vector ξ is normal to $L(\varphi, p)$ at x. Therefore,

²That is, for every $k \in \mathbb{N}$, there is a constant $C_k > 0$ such that $|\mathcal{F}(\psi f)(\xi)| \leq C_k/(1 + ||\xi||)^k$ for all $\xi \in V$.

 $WF_{L(\varphi,p)}(f)$ is the set of wavefront directions of f above points $x \in L(\varphi,p)$ that are normal to this line.

If g is a locally integrable function on $[0, 2\pi] \times \mathbb{R}$, then $WF_{(\varphi,p)}(g)$ is the set of wavefront directions with base point (φ, p) . The sets defined in Definition 2.2 have an important relationship that we will exploit starting in the next section.

3. Main results. In contrast to limited-angle characterizations in [12], our main results characterize arbitrary incomplete data reconstructions, including classical problems such as region-of-interest tomography, exterior tomography, and limited-angle tomography. Our results are formulated in terms of the wavefront set (Definition 2.1).

In many applications, reconstructions from incomplete CT data are calculated by the filtered backprojection algorithm (FBP), which is designed for complete data (see [36] for a practical discussion of FBP in practice). In this case, the incomplete data is often implicitly extended by the algorithm to a complete data set on $[0, 2\pi] \times \mathbb{R}$ by setting it to zero off of the set A (*cutoff region*) over which data are taken. Therefore, the incomplete CT data can be modeled as

(3.1)
$$R_A f(\varphi, p) = \mathbb{1}_A(\varphi, p) R f(\varphi, p),$$

where $\mathbb{1}_A$ is the characteristic function of A.³ Thus, using the FBP algorithm to calculate a reconstruction from such data gives rise to the reconstruction operator:

(3.2)
$$\mathcal{L}_A f = R^* \left(\Lambda R_A f \right) = R^* \left(\Lambda \mathbb{1}_A R f \right),$$

where Λ is the standard FBP filter (see e.g., [29, Theorem 2.5] and [30, §5.1.1] for numerical implementations) and R^* is defined by (2.4).

Our next assumption collects the conditions we will impose on the cutoff region A. There, we will use the notation int(A), bd(A), and ext(A) to denote the interior of A, the boundary of A, and the exterior of A, respectively.

Assumption 3.1. Let A be a proper subset of $[0, 2\pi] \times \mathbb{R}$ (i.e., $A \subseteq [0, 2\pi] \times \mathbb{R}$) with a nontrivial interior and assume A is symmetric in the following sense:

In addition, assume that A is the smallest closed set containing int(A), i.e. A = cl(int(A)).

We now justify using this assumption. Since A is proper, data over A are incomplete. Being symmetric means that the part of A in $[0, \pi] \times \mathbb{R}$ determines the data set. We assume that A is the smallest closed set containing int(A) in order to exclude degenerate cases, such as when A includes an isolated curve.

Our next definition gives us the language to describe the geometry of bd(A).

Definition 3.2 (Smoothness of bd(A)). Let $A \subset [0, 2\pi] \times \mathbb{R}$ and let $(\varphi_0, p_0) \in bd(A)$.

• We say that bd(A) is smooth near (φ_0, p_0) if this boundary is a C^{∞} curve near (φ_0, p_0) . In this case, there is a unique tangent line in (φ, p) -space to bd(A) at (φ_0, p_0) .

³The characteristic function of a set A is the function that is equal to one on the set and zero outside of A.

- If this tangent line is vertical (i.e., of the form $\varphi = \varphi_0$), then we say the boundary has infinite slope at (φ_0, p_0) .
- If this tangent line is not vertical, then bd(A) is defined near (φ_0, p_0) by a smooth function $p = p(\varphi)$. In this case, the slope of the boundary at (φ_0, p_0) will be the slope of this tangent line, which is given by $p'(\varphi_0)$.
- We say that bd(A) is not smooth at (φ_0, p_0) if it is not a smooth curve at (φ_0, p_0) .
 - We say that bd(A) has a corner at (φ_0, p_0) if the curve bd(A) is continuous at (φ_0, p_0) , is smooth at all other points in some neighborhood of (φ_0, p_0) , and has one-sided slopes at (φ_0, p_0) but those slopes are different.

3.1. Basic theorems on artifacts and visible singularities. In this section we begin to characterize singularities and artifacts, and we show that artifacts occur only on lines $L(\varphi, p)$ for $(\varphi, p) \in bd(A)$. This material is either known (e.g., [40]) or it follows from [40].

Definition 3.3 (Artifacts and visible singularities). Every singularity of $\mathcal{L}_A f$ that is not a singularity of f is called an artifact (i.e., any singularity in WF($\mathcal{L}_A f$) \ WF(f)). A streak artifact is a line of artifacts.

Every singularity of $\mathcal{L}_A f$ that is a singularity of f is called a visible singularity (from data on A) (i.e., any singularity in WF($\mathcal{L}_A f$) \cap WF(f)). Other singularities of f are called invisible (from data on A).⁴

Theorem 3.4 (Localization of artifacts to lines from bd(A)). Let A satisfy Assumption 3.1. If $(x,\xi) \in WF(\mathcal{L}_A f) \setminus WF(f)$, then there is a $(\varphi_0, p_0) \in bd(A)$, such that $x \in L(\varphi_0, p_0)$ and ξ is normal to $L(\varphi_0, p_0)$. That is, if a singularity of $\mathcal{L}_A f$ at (x,ξ) is an artifact, then it, necessarily, comes from bd(A).

This theorem follows from the microlocal analysis in [40, 42]. More generally, every singularity of $\mathcal{L}_A f$ must come from a singularity of $\mathbb{1}_A R f$, and if the singularity does not come from R f, it must come from a singularity of $\mathbb{1}_A$, and these singularities are on $\mathrm{bd}(A)$.

Our next theorem gives an analysis of singularities in $\mathcal{L}_A f$ corresponding to lines not in $\mathrm{bd}(A)$. It shows that visible singularities are along lines $L(\varphi, p)$ when $(\varphi, p) \in \mathrm{int}(A)$ and invisible singularities are on lines for $(\varphi, p) \in \mathrm{ext}(A)$.⁵

Theorem 3.5 (Visible and invisible singularities in the reconstruction). Let $f \in L^2(D)$ and let $A \subset [0, 2\pi] \times \mathbb{R}$ satisfy Assumption 3.1.

- (i) If $(\varphi, p) \in int(A)$ then $WF_{L(\varphi,p)}(f) = WF_{L(\varphi,p)}(\mathcal{L}_A f)$. That is, all singularities of f normal to $L(\varphi, p)$ are visible from data on A.
- (ii) If $(\varphi, p) \notin (A \cap \operatorname{supp}(Rf))$, then $\operatorname{WF}_{L(\varphi,p)}(\mathcal{L}_A f) = \emptyset$. In this case, $\mathcal{L}_A f$ is smooth normal to $L(\varphi, p)$. That is, all singularities of f normal to $L(\varphi, p)$ are invisible from data on A.
- (iii) Now, let $x \in D$ and assume that all lines through x are parameterized by points in int(A) (i.e., $\forall \varphi \in [0, 2\pi]$, $(\varphi, x \cdot \overline{\theta}(\varphi)) \in int(A)$). Then all singularities of f at x are

⁴Invisible singularities are smoothed by \mathcal{L}_A and reconstruction of those singularities is in general is exponentially ill-posed in Sobolev scale.

⁵Visible singularities of f can appear on lines $L(\varphi_0, p_0)$ for $(\varphi_0, p_0) \in bd(A)$, but these can cause artifacts all along the line $L(\varphi_0, p_0)$ as noted in Theorem 3.6 and seen in the reconstructions in Section 4. In addition, artifacts along such lines can mask visible singularities on those lines.

visible in $\mathcal{L}_A f$:

(3.4)
$$WF_x(f) = WF_x(\mathcal{L}_A f).$$

Therefore, artifacts occur only on lines in bd(A).

This theorem follows directly from [40, Theorem 3.1] or [42], and the proof will be sketched in Section A.2 since it is not exactly given in the literature.

3.2. Characterization of artifacts. We now characterize artifacts in FBP reconstructions from incomplete data given by (3.2). In particular, we show that the nature of artifacts depends on the smoothness and geometry of bd(A) and, in some cases, singularities of the object f. Our next two theorems show that artifacts arise in two ways:

- **Object-independent artifacts:** those caused essentially by the geometry of bd(A).
- Object-dependent artifacts: those caused by singularities of the object f that are normal to L(φ, p) for some (φ, p) ∈ bd(A).

As a first generalization of the limited-angle characterizations in [22, 12] we consider the case where bd(A) is locally smooth. This theorem describes the range of artifacts that can occur in this case, and it guarantees they will occur in certain cases.

Theorem 3.6 (Artifacts for locally smooth boundary). Let $f \in L^2(D)$ and $A \subset [0, 2\pi] \times \mathbb{R}$ satisfy Assumption 3.1. Let $(\varphi_0, p_0) \in bd(A)$ and assume that bd(A) is smooth near (φ_0, p_0) .

I. Assume bd(A) has finite slope at (φ_0, p_0) and let I be a neighborhood of φ_0 such that bd(A) is given by a smooth curve $p = p(\varphi)$ near (φ_0, p_0) . Let

 $\operatorname{Du}(A)$ is given by a smooth curve $p = p(\varphi)$ heat (φ_0, p_0) . Let

(3.5)
$$x_b = x_b(\varphi) = p(\varphi)\theta(\varphi) + p'(\varphi)\theta^{\perp}(\varphi) \quad \text{for } \varphi \in I.$$

A. An object-independent artifact can occur along the curve $I \ni \varphi \mapsto x_b(\varphi)$. B. If f has no singularity normal to $L(\varphi_0, p_0)$ then

(3.6)
$$\operatorname{WF}_{L(\varphi_0,p_0)}(\mathcal{L}_A f) \subset \left\{ (x_b, \omega \overline{\theta}(\varphi_0)) : \omega \neq 0 \right\}.$$

That is, $\mathcal{L}_A f$ is smooth normal to $L(\varphi_0, p_0)$, except perhaps at the single point $x_b(\varphi_0) \in L(\varphi_0, p_0)$.

- (i) If, in addition, $Rf(\varphi_0, p_0) \neq 0$, then equality holds in (3.6) and the $x_b(\varphi)$ artifact curve will be visible near $x_b(\varphi_0)$.
- (ii) If Rf is zero in a neighborhood of (φ_0, p_0) , then the $x_b(\varphi)$ artifact curve will not be visible near $x_b(\varphi_0)$.
- II. In all cases with smooth boundary at (φ_0, p_0) , if f has a singularity normal to $L(\varphi_0, p_0)$, then an object-dependent streak artifact can occur along $L(\varphi_0, p_0)$.
 - A. If f is smooth conormal to $L(\varphi_0, p_0)$ and bd(A) is vertical at (φ_0, p_0) , then $\mathcal{L}_A f$ is smooth normal to $L(\varphi_0, p_0)$.
 - B. If bd(A) is not vertical then $\mathcal{L}_A f$ is smooth normal to $L(\varphi_0, p_0)$ except possibly at $x_b(\varphi_0)$.

Theorem 3.6 part II. gives a necessary condition under which there can be an objectdependent artifact. This is observed in practice (see Figure 2). However, Remark A.6 shows such object-dependent streak artifacts might not occur, even if there is a singularity of f normal to this line. This is why we state that artifacts *can occur* in parts I.A. and II. of Theorem 3.6.

The proof is given in Section A.3 and we will present reconstructions illustrating this theorem in Section 4.2.

Remark 3.7. Assume bd(A) is smooth at with finite slope at (φ_0, p_0) . Let I be a neighborhood of φ_0 and $p: I \to \mathbb{R}$ describes bd(A) near (φ_0, p_0) . Note that

$$x_b(\varphi) \in L(\varphi, p)$$
 for $\varphi \in I$.

If the slope of bd(A) at (φ_0, p_0) is small enough, i.e.,

(3.7)
$$|p'(\varphi_0)| < \sqrt{1 - p_0^2}$$

holds, then the artifact curve $\varphi \to x_b(\varphi)$ will be inside the unit disk, D, near (φ_0, p_0) . If not, this curve will be outside, D, at least at φ_0 .

If bd(A) is smooth and vertical at (φ_0, p_0) (infinite slope), then there will be no objectindependent artifact on the line $L(\varphi_0, p_0)$ because $x_b(\varphi_0)$ is a "point at infinity" since " $p'(\varphi_0)$ " is infinite. This is proven using the arguments in the proof of Theorem 3.6 B.

When bd(A) is not smooth at (φ_0, p_0) , the next theorem shows there can be a streak artifact in $\mathcal{L}_A f$ along $L(\varphi_0, p_0)$ that is independent of the object f.

Theorem 3.8 (Artifacts for nonsmooth boundary). Let $f \in L^2(D)$ and $A \subset [0, 2\pi] \times \mathbb{R}$ satisfy Assumption 3.1. Let $(\varphi_0, p_0) \in bd(A)$ and assume that bd(A) is not smooth at (φ_0, p_0) . Then, $\mathcal{L}_A f$ can have a streak artifact on $L(\varphi_0, p_0)$ independent of f.

If f is smooth normal to $L(\varphi_0, p_0)$, $Rf(\varphi_0, p_0) \neq 0$ and bd(A) has a corner at (φ_0, p_0) (see Definition 3.2), then $\mathcal{L}_A f$ does have a streak artifact all along $L(\varphi_0, p_0)$:

$$WF_{L(\varphi_0,p_0)}(\mathcal{L}_A f) = N(L(\varphi_0,p_0)).$$

This theorem is proven in Section A.3.

The streak artifacts (artifacts along lines) in this theorem are object-independent, and they are illustrated in a simple case in Section 4.4 and on real data in Section 7.

In Theorem 5.2, we will describe the strength of the artifacts in Sobolev scale in specific cases of Theorems 3.6 and 3.8.

Object-dependent streak artifacts along lines were analyzed for limited-angle tomography in articles such as [22, 12, 31], but we are unaware of a general reference to Theorem 3.6, part II. for general incomplete data problems, although this is not surprising given the limitedangle result. We are not aware of a previous reference in the mathematical literature to the x_b curve artifact in Theorem 3.6 part I. or to the corner artifacts from Theorem 3.8.

4. Numerical illustrations of our theoretical results. We now consider a range of wellknown incomplete data problems as well as unconventional ones to show how the theoretical results in Section 3 are reflected in practice. All sinograms have $\varphi \in [0, \pi]$ on the horizontal axis and $p \in [-\sqrt{2}, \sqrt{2}]$ on the vertical. Except for the center picture in Figure 3(A), the reconstruction is displayed on $[-1, 1]^2$. **4.1. Limited-angle tomography.** First, we analyze limited-angle tomography, a classical problem in which part II. of Theorem 3.6 applies. In this case bd(A) consists of vertical lines given by $\varphi = \varphi_0$ (infinite slope). Taking a closer look at the statement of Theorem 3.6 and the results of [12, 14] one can observe that, locally, they describe the same phenomena, namely: whenever there is a line $L(\varphi_0, p_0)$ in the data set with $(\varphi_0, p_0) \in bd(A)$ and which is normal to a singularity of f, then a streak artifact can be generated along $L(\varphi_0, p_0)$ in the reconstruction $\mathcal{L}_A f$. Therefore, Theorem 3.6 generalizes the results of [22, 12] as it also apply to cutoffs with non-vertical slope.

It is important to note that, with limited-angle data, there are no object-independent artifacts since bd(A) is smooth and the artifact curve $\varphi \mapsto x_b(\varphi)$ is not defined.



Figure 2: *Left:* Limited-angle data (cutoff with infinite slope). *Center:* FBP reconstruction. *Right:* Reconstruction highlighting object-dependent artifact lines tangent to skull corresponding to the four circled points in the sinogram.

Figure 2 illustrates limited-angle tomography. The boundary consists of the vertical lines $\varphi = 4\pi/9$ and $\varphi = 5\pi/9$. The artifact lines are exactly the lines with $\varphi = 4\pi/9$ or $5\pi/9$ that are tangent to boundaries in the object (i.e., wavefront directions are normal to the line). The four circled points on the sinogram correspond to the object-dependent artifact lines at the boundary of the skull. The corresponding lines are tangent to the skull and have angle $\varphi = 4\pi/9$ and $\varphi = 5\pi/9$. One can also observe artifact lines tangent to the inside of the skull with these same angles.

4.2. Smooth boundary with finite slope. We now consider the general case in Theorem 3.6 by analyzing the artifacts for a specific set A which is defined as follows. It will be cut in the middle so that the left-most cutoff boundary occurs at $\varphi = a := \frac{4}{9}\pi$; the right-most boundary is constructed as $\varphi = b := \frac{5}{9}\pi$ for $p \leq 0$ and

(4.1)
$$p(\varphi) = c\sqrt{\varphi - b}, \quad \varphi > b$$

for p > 0 such that the two parts join differentiably at $(\varphi, p) = (0, 0)$. The steepness of the curved part of the right-most boundary is governed by the constant c (as seen in the two sinograms in Figure 3).

ARTIFACTS FROM ARBITRARY INCOMPLETE CT DATA

According to the condition (3.7), the curved part of bd(A) is the only part that can potentially cause object-independent artifacts in D, since the other parts are vertical. In Figure 3, we consider two data sets A with smooth boundary, the top where the x_b artifact curve is outside the unit disk and the bottom where it meets the object.



(A) Left: Boundary of A with large slope (c = 1.3). Center: FBP reconstruction over the larger region $[-2, 2]^2$ to show that the x_b artifact is outside of the region displayed in the right frame. Right: Reconstruction highlighting object-dependent artifact lines tangent to the skull corresponding to the four circled points in the sinogram.



(B) Smooth boundary with small slope (c = 0.65) causing a prominent curve of artifacts in *D*. Left: Sinogram with part of boundary that could causes artifacts in the reconstruction region highlighted in magenta. The solid part of the curve indicates the artifacts that are realized in the reconstruction. The dotted curve at the right end of the sinogram indicates potential artifacts that are not realized because the corresponding part of bd(A) is outside supp(Rf) (see Theorem 3.5(ii)). Center: FBP reconstruction. Right: The magenta curve in the reconstruction is the curve of artifacts $\varphi \mapsto x_b(\varphi)$ and the yellow artifact lines are object-dependent artifacts similar to those in Figure 3(A).

Figure 3: Illustration of artifacts with smooth boundary given by (4.1). The artifact curve $\varphi \mapsto x_b(\varphi)$ is outside the reconstruction region in the top figure and the curve meets the object in the bottom picture.

Figure 3(A) provides a reconstruction with data set defined by c = 1.3 in (4.1). Many artifacts in the reconstruction region are the same as in Figure 2 because the boundaries of

the cutoff regions are substantially the same: the artifacts corresponding to the circles with $\varphi = 4\pi/9$ and the lower circle with $\varphi = 5\pi/9$, are the same limited-angle artifacts as in Figure 2 because those parts of the boundaries are the same. However, the upper right circled point in the sinogram has $\varphi > 5\pi/9$ so the corresponding artifact line has this larger angle, as seen in the reconstruction. The center reconstruction in Figure 3(A) shows the x_b artifact curve, but it is far enough from D that it is not visible in the reconstruction on the far right.

Figure 3(B) provides a reconstruction with data set defined by c = 0.65 in (4.1). In this case, the object-dependent artifacts are similar to those in Figure 3(A), but the lines for (φ, p) defined by (4.1) are different because bd(A) is different. The highlighted part of the boundary of A defined by (4.1) indicates the boundary points that create the part of the $x_b(\varphi)$ artifact curve that is in the reconstruction region. The highlighted curve in the right-hand reconstruction of Figure 3(B) is this part of the $x_b(\varphi)$ curve. Note that this curve is calculated using the formula (3.5) for $x_b(\varphi)$ rather than by tracing the physical curve on the reconstruction. That they are substantially the same shows the efficacy of our theory. A simple exercise shows that, for any c > 0, the x_b curve changes direction at $x_b(1/2 + 5\pi/9)$.

Let (φ_0, p_0) be the coordinates of the circled point in the upper right of the sinogram in Figure 3(B). This circled point is on the boundary of $\operatorname{supp}(Rf)$ so $L(\varphi_0, p_0)$ is tangent to the skull and an object-dependent artifact is visible along $L(\varphi_0, p_0)$ in the reconstruction. The x_b curve ends at $x_b(\varphi_0)$ (as justified by Theorem 3.5 part (ii)) and so the x_b curve seems to blend into this line. If $\operatorname{supp}(f)$ were larger and the dotted part of the magenta curve on the sinogram were in $\operatorname{supp}(Rf)$, the x_b curve would be longer.

4.3. Region-of-interest (ROI) tomography. The ROI problem, also known as interior tomography, is a classical incomplete data tomography problem in which a part of the object (the ROI) is imaged using only data over lines that meet the ROI. ROI data are used when the detector width is not large enough to contain the complete object or when researchers would like a higher resolution scan of a small part of the object. We now demonstrate how our theoretical results predict precisely which reconstruction artifacts occur for two choices of detector width.

Note that Theorem 3.5 part (iii) implies that if the ROI is convex, then all singularities of f in the interior of the ROI are recovered. The reason is that, in this case, all wavefront directions at all points in the interior are normal to lines in the data set. This is observed in both reconstructions in Figure 4.

The boundary of the sinogram are given by horizontal lines $p = \pm p_0$ where $p_0 = 0.4$ in Figure 4(A) and $p_0 = 0.8$ in Figure 4(B). The points given by (3.5) are $x_b(\varphi) = p_0\overline{\theta}(\varphi) + 0\theta^{\perp}(\varphi)$ since p' = 0, i.e., a circle of radius p_0 . The ring around the ROI in Figure 4(A) is exactly the curve $\varphi \mapsto x_b(\varphi)$ for this ROI. There are no object-dependent streak artifacts in this picture because there are no artifacts of the object tangent to lines for $(\varphi, p) \in \mathrm{bd}(A)$.

More generally, if the ROI is smooth and strictly convex then the artifact curve x_b traces the boundary of the ROI. The proof is an exercise using the parametrization in (φ, p) of tangent lines to this boundary.

Analyzing the reconstruction in Figure 4(B), one sees the top and bottom of an x_b artifact circle which is analogous to the one in Figure 4(A). However, the artifact circle does not extend outside the object (as represented by the dotted magenta curve) because Rf is zero



(A) Sinogram and ROI reconstruction within a disk of radius 0.4 centered at the origin, $p \in [-0.4, 0.4]$.



(B) Sinogram and ROI reconstruction within a disk of radius 0.8 centered at the origin, $p \in [-0.8, 0.8]$.

Figure 4: ROI reconstructions with small and large ROI. In each subfigure top left is the sinogram, top right the reconstruction and bottom the same with sinogram boundary and reconstruction artifacts plotted. Note that the part of the sinogram for $\varphi \in [0, \pi]$ is shown.

near the lines corresponding to $p = \pm 0.8$, $\varphi < 0.7121$ and $\varphi > 2.429$ (this is indicated by the dotted part of bd(A) in the sinogram). Therefore, the reconstruction is smooth there as explained by Theorem 3.6 part (ii).

One also sees object-dependent artifacts described by Theorem 3.6 part II. in Figure 4(B). These occur along lines $L(\varphi_0, p_0)$ where $p_0 = \pm 0.8$ that are tangent to the two boundaries of the skull. The angles for these lines are $\varphi_0 = 0.7121$ and $\varphi_0 = 2.429$ (these four points (φ_0, p_0) are circled on the sinogram), and they are where the support of Rf intersects the line $p = \pm 0.8$; these are lines at the boundary of the data set at which Rf has wavefront set (so f has wavefront set normal to the line $L(\varphi, p)$). The same phenomenon happens on the line with $p = \pm 0.8$ that is tangent to the inside of the skull, which is what causes the second set of four visible line artifacts indicated in yellow.

4.4. The general case. The reconstruction in Figure 5 illustrates all our cases in one. In that figure, we consider a general incomplete data set with a rectangular region cut out of the

sinogram leading to all considered types of artifacts. Now, we describe the resulting artifacts. In Figure 5 the horizontal sinogram boundaries at $p = p_0 = \pm 0.35$ for $\phi \in \left[\frac{7}{18}\pi, \frac{11}{18}\pi\right]$ are displayed in solid magenta line. As in the ROI case, along these boundaries, we have p' = 0 and thus circular arcs of radius p_0 for the given interval for φ are added in the reconstruction (as indicated by solid magenta). As predicted by Theorem 3.8, each of the four corners produce a line artifact as marked by the yellow solid lines in the right-hand reconstruction, and they align tangentially with the ends of the curved artifacts.



Figure 5: Left: The sinogram for a general incomplete data problem in which the cutoff region, A, has a locally smooth boundary with zero and infinite slope as well as corners. The cutout from the sinogram is at $\frac{7\pi}{18}$ and $\frac{11\pi}{18}$, $p = \pm 0.35$. Center: FBP reconstruction. Right: Reconstruction with the circular x_b curve artifacts highlighted in magenta and object-independent "corner" streak artifacts highlighted in yellow.

The circular arc between those lines corresponds to the top and bottom parts of bd(A) as the data are, locally, constrained as in ROI CT (see Section 4.3).

In Figure 5, there are other object-dependent streaks corresponding to the vertical lines in the sinogram at $\varphi = \frac{7\pi}{18}$ and at $\varphi = \frac{11\pi}{18}$ as predicted by Theorem 3.6II., but they are less pronounced and more difficult to see.

4.5. Summary. We have presented reconstructions that illustrate all of types of incomplete data and each of our theorems from Section 3. All artifacts arise because of points $(\varphi_0, p_0) \in \mathrm{bd}(A)$, and they fall into two categories.

- Streak artifacts along the line $L(\varphi_0, p_0)$:
 - Object-dependent streaks when bd(A) is smooth at (φ_0, p_0) and a singularity of f is normal to $L(\varphi_0, p_0)$.
 - Object-independent streaks when bd(A) is nonsmooth at (φ_0, p_0) .
- Artifacts on curves:
 - They are object-independent, and they are generated by the map $\varphi \mapsto x_b(\varphi)$ from parts of bd(A) that are smooth and of small slope.

5. Strength of added artifacts. Using the Sobolev continuity of Rf, one can measure the strength in Sobolev scale of added artifacts in several useful cases. First, we define the Sobolev norm [45, 38]. We state it for distributions, so it applies to functions $f \in L^2(D)$.

Definition 5.1 (Sobolev wavefront set [38]). For $s \in \mathbb{R}$, the Sobolev space $H_s(\mathbb{R}^n)$ is the set of all distributions with locally square integrable Fourier transform and with finite Sobolev norm:

(5.1)
$$\|f\|_{s} := \left(\int_{y \in \mathbb{R}^{n}} |\mathcal{F}f(y)|^{2} (1 + \|y\|^{2})^{s} \, dy\right)^{1/2} < \infty.$$

Let f have locally square integrable Fourier transform and let $x_0 \in \mathbb{R}^n$ and $\xi_0 \in \mathbb{R}^n \setminus \mathbf{0}$. We say f is in H^s at x_0 in direction ξ_0 if there is a cutoff function ψ at x_0 and an open cone V containing ξ_0 such that the localized and microlocalized Sobolev seminorm

(5.2)
$$\|f\|_{s,\psi,V} := \left(\int_{y\in V} |\mathcal{F}(\psi f)(y)|^2 (1+\|y\|^2)^s \, dy\right)^{1/2} < \infty.$$

On the other hand, if (5.2) does not hold for any cutoff at x_0 , ψ or conic neighborhood V of ξ_0 , then $(x_0, \xi_0) \in WF_s(f)$, the Sobolev wavefront set of f of order s.

An exercise using the definitions shows that $WF(f) = \bigcup_{s \in \mathbb{R}} WF_s(f)$ [11].

The Sobolev wavefront set can be defined for periodic distributions on $[0, 2\pi] \times \mathbb{R}$ by considering the periodic extension to $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$ as discussed for C^{∞} wavefront set in Section 2.2.

Note that this norm on distributions on $[0, 2\pi] \times \mathbb{R}$ is not the typical $H_{0,s}$ norm used in elementary continuity proofs for the Radon transform (see e.g., [19, equation (2.11)]), but this is the appropriate norm for the continuity theorems for general Fourier integral operators [20, Theorem 4.3.1], [8, Corollary 4.4.5].

Our next theorem gives the strength in Sobolev scale of added singularities of $\mathcal{L}_A f$ under certain assumptions on f. It uses the relation between microlocal Sobolev strength of f and of Rf, [40, Theorem 3.1] and of g and R^*g , which is given in Proposition A.7.

Theorem 5.2. Let $f \in L^2(D)$ and let $A \subset [0, 2\pi] \times \mathbb{R}$ satisfy Assumption 3.1. Let $(\varphi_0, p_0) \in bd(A)$ and assume $Rf(\varphi_0, p_0) \neq 0$ and f is smooth normal to $L(\varphi_0, p_0)$, i.e., $WF_{L(\varphi_0, p_0)}(f) = \emptyset$.

- (i) Assume bd(A) is smooth and not vertical at (φ_0, p_0) . Let $x_b = x_b(\varphi_0)$ be given by (3.5) and let $\omega \neq 0$. Then, $\mathcal{L}_A f$ is in H_s for s < 0 at $\xi_0 = (x_b, \omega \overline{\theta}(\varphi_0))$ and $\xi_0 \in WF_0(\mathcal{L}_A f)$. Thus, there are singularities above x_b in the 0-order wavefront set of $\mathcal{L}_A f$.
- (ii) Now, assume bd(A) has a corner at (φ_0, p_0) . Then for each $(x, \xi) \in N(L(\varphi_0, p_0))$, $(x, \xi) \in WF_1(\mathcal{L}_A f)$ and, except for two points on $L(\varphi_0, p_0)$, $\mathcal{L}_A f$ is in H_s for s < 1 at (x, ξ) .
- (iii) Now, assume $(\varphi_0, p_0) \in int(A)$. Then, for every $(x, \xi) \in N(L(\varphi_0, p_0))$, $(x, \xi) \in WF_s(f)$ if and only if $(x, \xi) \in WF_s(\mathcal{L}_A f)$.

This theorem provides estimates on smoothness for more general data sets than the limited-angle case, which was thoroughly considered in [22, 31]. In contrast to part (i) of this theorem, if bd(A) has a vertical tangent at (φ_0, p_0) , then, under the smoothness assumption on f, there are no added artifacts in $\mathcal{L}_A f$ normal to $L(\varphi_0, p_0)$ (see Theorem 3.6 part II.). Part (i) of this theorem is a more precise version of the last assertion of Theorem 3.8. In cases (i) and (ii), bd(A) will cause specific singularities in specific locations on $L(\varphi_0, p_0)$. The two

more singular points in part (ii) will be specified in equation (A.16). If one part of bd(A) is vertical at (φ_0, p_0) , then there is only one more singular point.

This theorem will be proven in Section A.4 of the appendix.

6. Artifact reduction. In this section, we briefly describe a simple method to suppress the added streak artifacts described in Theorems 3.6 and 3.8. As outlined in Section 3, the application of FBP to incomplete data extends the data from $A \subset [0, 2\pi] \times \mathbb{R}$ to all of $[0, 2\pi] \times \mathbb{R}$ by padding it with zeros off of A. This hard truncation can create discontinuities along bd(A)and that explains the artifacts. These jumps are stronger singularities than those of Rf for $Rf \in H_{1/2}([0, 2\pi] \times \mathbb{R})$ since $f \in L^2(D) = H_0(D)$.

One obvious way to get rid of the jump discontinuities of $\mathbb{1}_A$ is to replace $\mathbb{1}_A$ by a smooth function on $[0, 2\pi] \times \mathbb{R}$, ψ , that is equal to zero off of A and equal one on most of int(A) and smoothly transitions to zero near bd(A). We also assume ψ is symmetric in the sense $\psi(\varphi, p) = \psi(\varphi + \pi, -p)$ for all (φ, p) .

This gives the forward operator

(6.1)
$$R_{\psi}f(\varphi,p) = \psi(\varphi,p)Rf(\varphi,p)$$

and the reconstruction operator

(6.2)
$$\mathcal{L}_{\psi}f = R^* \left(\Lambda R_{\psi}f\right) = R^* \left(\Lambda \psi Rf\right).$$

Because ψ is a smooth function, R_{ψ} is a standard Fourier integral operator and so \mathcal{L}_{ψ} is a standard pseudodifferential operator. This allows us to show that \mathcal{L}_{ψ} does not add artifacts.

Theorem 6.1 (Artifact Reduction Theorem). Let $f \in L^2(D)$ and $A \subset [0, 2\pi] \times \mathbb{R}$ satisfy Assumption 3.1. Then

(6.3)
$$\operatorname{WF}(\mathcal{L}_{\psi}f) \subset \operatorname{WF}(f).$$

Therefore, \mathcal{L}_{ψ} does not add artifacts to the reconstruction.

Let $x \in D$, $\varphi \in [0, 2\pi]$, and $\omega \neq 0$ and assume $\psi(\varphi, x \cdot \overline{\theta}(\varphi)) \neq 0$. Then,

(6.4)
$$(x, \omega \overline{\theta}(\varphi) \mathrm{dx}) \in \mathrm{WF}(\mathcal{L}_{\psi} f) \text{ if and only if } (x, \omega \overline{\theta}(\varphi) \mathrm{dx}) \in \mathrm{WF}(f).$$

Proof. We provide an outline using arguments from [14]. First, \mathcal{L}_{ψ} is a standard pseudodifferential operator (Ψ DO) because it is the composition of R^* , the Ψ DO $\Lambda\psi$, and R, and because R satisfies the global Bolker assumption (see (A.7)). When the top-order symbol of \mathcal{L}_{ψ} is nonzero, the operator \mathcal{L}_{ψ} is elliptic. The symbol is calculated more generally in Theorems 6.1 and 6.2 in [14] or [39]. For reference, the symbol of \mathcal{L}_{ψ} is

$$\sigma\left(\mathcal{L}_{\psi}\right)\left(x,\omega\overline{\theta}(\varphi)\right) = 4\pi\psi\left(\varphi,x\cdot\overline{\theta}(\varphi)\right),\,$$

and this follows from the symbol calculation in [14, (6.3) and (A.11)] with $\xi = \omega \overline{\theta}(\varphi)$, with weights $\mu = 1$ and $\nu = 1$, $P = ||\xi||$, and one uses that ψ is symmetric.

Many practitioners include a smooth cutoff function in incomplete data algorithms, but others do not. Theorem 6.1 shows the advantages of including such a cutoff, and it has been



Figure 6: Left: Smoothed sinogram. Center: Smoothed reconstruction with suppressed artifacts. Right: Reconstruction using \mathcal{L}_A , with sharp cutoff.

suggested in several settings, including limited-angle X-ray CT [22, 12] and more general tomography problems [14, 13, 46]. More sophisticated methods are discussed in [6, 5] for the synchrotron problem that is described in Section 7.

Figure 6 illustrates the effects of our smoothing algorithm. At the end of Section 7, we will illustrate the efficiency of this simple artifact reduction method on real synchrotron data.

7. Application: a synchrotron experiment. Figure 7 shows tomographic data of a chalk sample (Figure 7(A) and 7(B)) that was acquired at a synchrotron experiment [6, 5] (see [25] for related work) and a zoom of the corresponding reconstruction (Figure 7(C)). As can be clearly observed, the reconstruction includes dramatic streaks that are independent of the object. These streaks motivated the research in this article since they were not explained by the mathematical theory at that time (such as in [22, 31, 12, 13, 14]).

Taking a closer look at the attenuation sinogram, Figures 7(A)-7(B), a staircasing is revealed with vertical and horizontal boundaries. This is a result of X-rays being blocked by four metal bars that help stabilize the percolation chamber (sample holder) as the sample is subjected to high pressure during data acquisition, see Figure 8. More details are given in [5].

Because the original reconstructions of this synchrotron data used a sharp cutoff, the original reconstructions suffer from severe streak artifacts, see Figure 9(A). These artifacts are exactly described by Theorem 3.8 in that each corner of the sinogram gives rise to a line artifact in the reconstruction (cf. Figures 7(A)-7(B)). The authors of [5] then use a smooth cutoff at bd(A) that essentially eliminates the streaks. The resulting reconstruction is shown in Figure 9(B).

8. Discussion. We first make observations about our results for \mathcal{L}_A and then discuss generalizations.

8.1. Observations. Theorem 3.8 is valid as long as $WF_{(\varphi_0,p_0)}(\mathbb{1}_A) = \{(\varphi_0,p_0)\} \times (\mathbb{R}^2 \setminus \mathbf{0}),$ but the analogous theorem for Sobolev singularities, Theorem 5.2(ii), assumes that A has a corner at (φ_0, p_0) . If A has a weaker singularity at (φ_0, p_0) , then an analogous theorem would hold but one would need to factor in the Sobolev strength of the wavefront of $\mathbb{1}_A$ above (φ_0, p_0) .



Figure 7: Left and Center: The truncated attenuation sinogram (after processing to get Radon transform data) and an enlargement of a central section of bd(A). Right: Reconstruction given in Figure 1 [5, ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved].



Figure 8: Data acquisition setup for the synchrotron experiment [5, ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved].

Our artifact reduction method, which is motivated by Theorem 6.1, works well for the synchrotron data as was shown in Figure 9. The article [5] provides more elaborate artifact reduction methods that are even more successful. We would also like to point out that in other incomplete data tomography problems, this simple technique might not work as efficiently as in the presented problem. In general, the artifact reduction methods need to be designed for each particular incomplete data tomography problem, but Theorem 6.1 implies that smooth cutoffs are better than abrupt cutoffs.

There are other methods to deal with incomplete data. In other approaches, authors have completed incomplete data so that the completed data is in the range of the Radon transform



(A) Standard FBP reconstruction



(B) FBP reconstruction with artifact reduction (cf. Theorem 6.1).

Figure 9: Reconstructions from synchrotron data without smoothing (top) and with smoothing (bottom)[5, ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved].

with full data (e.g., [26, 2, 49]). In [34] and [7], the authors develop artifact reduction methods using quantitative susceptibility mapping. For metal artifacts, there is vast literature (see, e.g., [3]) for artifact reduction methods, and we believe that those methods might also be useful for certain other incomplete data tomography problems. Authors [37, 44, 35] have effectively used microlocal analysis to understand these related problems.

Our theory is developed based on the continuous case – we view the data as functions on $[0, 2\pi] \times \mathbb{R}$, not just defined at discrete points. As shown in this article, our theory predicts and

explains the artifacts and visible and invisible singularities. In practice, real data are discrete, and discretization may also introduce artifacts, such as undersampling streaks. Discretization in our synchrotron experiment could be a factor in the streaks in Figure 7. Furthermore, numerical experiments have finite resolution, and this can cause (and sometimes de-emphasize) artifacts. For all these reasons, further analysis is needed to shed light on the interplay between the discrete and the continuous theory for CT reconstructions from incomplete data.

8.2. Generalizations. Our theorems were proven for $\mathcal{L}_A = R^*(\Lambda(\mathbb{1}_A R))$, but the results hold for more general filtering operators besides Λ . One key to our proofs is that Λ satisfies the ellipticity condition in Remark A.5, but many other operators satisfy this condition. For example, the operator, $L = -\frac{\partial^2}{\partial p^2}$, in Lambda CT [9] satisfies this condition, and the only difference comes in our Sobolev Continuity Theorem 5.2. Since L is order two, the operator R^*LR is of order 1 and the smoothness in Sobolev scale of the reconstructions would be one degree lower than for \mathcal{L}_A .

Our theorems hold for fan-beam data when the source curve γ is smooth and convex and the object is compactly supported inside γ . This is true because, in this case, the fan-beam parameterization of lines is diffeomorphic to the parallel beam parametrization we use and the microlocal theorems we use are invariant under diffeomorphisms.

Theorems 3.6 and 3.8 hold verbatim for generalized Radon transforms with smooth measures on lines in \mathbb{R}^2 because they all have the same canonical relation, given by (A.4), and the proofs would be done as for \mathcal{L}_A but using the basic microlocal analysis in [39].

Analogous theorems hold for other Radon transforms including the hyperplane transform and the spherical transform of photoacoustic CT. The proofs would use our arguments here plus the proofs in [14, 13]. These generalizations are the subject of ongoing work. In incomplete data problems for R, the artifacts are either along curves $\varphi \mapsto x_b(\varphi)$ or they are streaks along the lines corresponding to points on bd(A). However, in these higher dimensional cases, the results will be more subtle because artifacts can spread on *proper subsets* of the surface over which data are taken, not necessarily the entire set (see [13, Remark 4.7]).

Analogous theorems should hold for cone beam CT, but this type of CT is more subtle because the reconstruction operator itself can add artifacts, even with complete data [15, 10].

Appendix A. Proofs. Here, we provide some basic microlocal analysis and then use this to prove our theorems. In this section, we adapt the standard terminology of microlocal analysis and consider wavefront sets as subsets of cotangent spaces [50]. Elementary presentations of microlocal analysis for tomography are in [23] and [18, Section 2.2]. Standard references include [11, 48].

A.1. Building blocks. Our first lemma gives some basic facts about wavefront sets.

Lemma A.1. Let $x_0 \in \mathbb{R}^2$. Let u and v be locally integrable functions or distributions.

- (i) Let U be an open neighborhood of x_0 . Assume that u and v are equal on U, then $WF_{x_0}(u) = WF_{x_0}(v)$.
- (ii) Let ψ be a smooth, compactly supported function, then $WF_{x_0}(\psi u) \subset WF_{x_0}(u)$. If ψ is nonzero at x_0 then $WF_{x_0}(u) = WF_{x_0}(\psi u)$.
- (iii) $WF_{x_0}(u) = \emptyset$ if and only if there is an open neighborhood U of x_0 on which u is a smooth function.

ARTIFACTS FROM ARBITRARY INCOMPLETE CT DATA

The analogous statements hold for functions on $[0, 2\pi] \times \mathbb{R}^{6}$.

These basic properties are proven using the arguments in Section 8.1 of [21], in particular, Lemma 8.1.1, Definition 8.1.2, and Proposition 8.1.3.

Our next definition will be useful to describe how wavefront set propagates under R and R^* .

Definition A.2. If $C \subset T^*([0, 2\pi] \times \mathbb{R}) \times T^*(\mathbb{R}^2)$ then,

$$C^t = \left\{ (x, \xi, \varphi, p, \eta) : (\varphi, p, \eta, x, \xi) \in C \right\}.$$

If $B \subset T^*([0, 2\pi] \times \mathbb{R})$, then

$$C^{t} \circ B = \left\{ (x,\xi) \in T^{*}(\mathbb{R}^{2}) : (x,\xi,\varphi,p,\eta) \in C^{t} \text{ for some } (\varphi,p,\eta) \in B \right\}.$$

The function g on $[0, 2\pi] \times \mathbb{R}$ will be called *symmetric* if

(A.1)
$$\forall (\varphi, p) \in [0, 2\pi] \times \mathbb{R}, \ g(\varphi, p) = g(\varphi + \pi, -p).$$

If $f \in L^2(D)$, then Rf and $\Lambda \mathbb{1}_A Rf$ are both locally integrable functions that satisfy this symmetry condition. For such functions,

(A.2)
$$(\varphi_0, p_0, \omega_0(-\alpha d\varphi + dp)) \in WF(g) \Leftrightarrow (\varphi_0 + \pi, -p_0, -\omega_0(\alpha d\varphi + dp)) \in WF(g).$$

For these reasons, we will identify cotangent vectors

(A.3)
$$(\varphi_0, p_0, \omega_0(-\alpha d\varphi + dp)) \Leftrightarrow (\varphi_0 + \pi, -p_0, -\omega_0(\alpha d\varphi + dp)).$$

Our next proposition is the main technical theorem of the article. It provides the wavefront correspondences for R and R^* which we will use in our proofs.

Proposition A.3 (Microlocal correspondence of singularities [42]). The X-ray transform, R, is an elliptic Fourier integral operator with canonical relation

(A.4)
$$C = \left\{ \left(\varphi, x \cdot \overline{\theta}(\varphi), \omega(-x \cdot \overline{\theta}(\varphi) \mathrm{d}\varphi + \mathrm{d}p), x, \omega \overline{\theta}(\varphi) \right) \\ : \varphi \in [0, 2\pi], x \in \mathbb{R}^2, \omega \neq 0 \right\}.$$

Let $f \in L^2(D)$ and let g be a locally integrable function on $[0, 2\pi] \times \mathbb{R}$ that is symmetric by (A.1). Let $x_0 \in \mathbb{R}^2$, $\varphi_0 \in [0, 2\pi]$, and let p, α , and ω be real numbers with $\omega \neq 0$.

The X-ray transform R is an elliptic FIO with canonical relation C. Therefore,

(A.5)
$$WF(Rf) = C \circ WF(f) \quad and \\ C \circ \left\{ (x_0, \omega \overline{\theta}(\varphi) d\mathbf{x}) \right\} = \left\{ \left(\varphi_0, x_0 \cdot \overline{\theta}(\varphi_0), \omega(-x_0 \cdot \theta^{\perp}(\varphi_0) d\varphi + d\mathbf{p}) \right) \right\}$$

under the identification (A.3).

The dual transform R^* is an elliptic Fourier integral operator associated with C^t . Then,

(A.6)

$$WF(R^*g) = C^t \circ WF(g) \quad and$$

$$C^t \circ \{(\varphi, p, \omega(-\alpha d\varphi + dp))\} = \{(x_0(\varphi, p, \alpha), \omega\overline{\theta}(\varphi))\}$$

$$where \quad x_0(\varphi, p, \alpha) = \alpha \theta^{\perp}(\varphi) + p\overline{\theta}(\varphi).$$

⁶This lemma is valid on $[0, 2\pi] \times \mathbb{R}$ since we consider only distributions on $[0, 2\pi] \times \mathbb{R}$ that are the restriction to $[0, 2\pi] \times \mathbb{R}$ of distributions for $(\varphi, p) \in \mathbb{R}^2$ that are 2π -periodic in φ .

Proof. The facts about R are directly from [40, Theorem 3.1] or [42, Theorem A.2], and they use the calculus of the FIO R [16, 17] (see also [39]). Note that the crucial point is that R is an elliptic Fourier integral operator that satisfies the global Bolker Assumption: the natural projection

(A.7)
$$\Pi_L: C \to T^*(Y) \text{ is an injective immersion,}$$

so (A.5) holds for R. Note that we are using the identification (A.3) in asserting that (A.5) is an equality. The proofs for R^* are parallel to those for R except they involve the canonical relation for R^* , C^t , rather than C.

Remark A.4. In order to multiply distributions u and v on $[0, 2\pi] \times \mathbb{R}$ as distributions one needs the non-cancellation condition

(A.8)
$$\forall (\varphi, p, \xi) \in WF(u), \ (\varphi, p, -\xi) \notin WF(v)$$

to hold. Then uv is a distribution and an upper bound for WF(uv) can be calculated in terms of WF(u) and WF(v) [21, Theorem 8.2.10]. However, this non-cancellation condition does not hold, in general for $\mathbb{1}_A$ and Rf when $\mathbb{1}_A$ either is smooth with finite slope or is not smooth at (φ_0, p_0) . That is why we consider functions $f \in L^2(D)$ since $\mathbb{1}_A Rf$ will be a function in $L^2([0, 2\pi] \times \mathbb{R})$ even if [21, Theorem 8.2.10] does not apply.

Our next remark will be used in ellipticity proofs that follow.

Remark A.5. The operator Λ is elliptic in all cotangent directions except $d\varphi$ because the symbol of Λ is $|\tau|$ where τ is the Fourier variable dual to p. However, the $d\varphi$ direction will not affect our proofs. This is true because, for any function $f \in L^2(D)$, the covector $(\varphi, p, \omega d\varphi)$ is not in WF(Rf) because WF(Rf) = $C \circ WF(f)$ (use the definition of composition and (A.4)). So, for each $f \in L^2(D)$, WF(ΛRf) = WF(Rf). Because $C^t \circ \{(\varphi, p, \alpha d\varphi)\} = \emptyset$ by (A.4), even if $(\varphi, \omega d\varphi) \in WF(\mathbb{1}_A Rf)$, that covector will not affect the calculation of $C^t \circ WF(\Lambda \mathbb{1}_A Rf)$. Therefore, Λ is elliptic on all cotangent directions that are preserved when composed with C^t , and these are all the directions we need in our proofs.

Our theorems will be valid for any 2π -periodic Ψ DO on $[0, 2\pi] \times \mathbb{R}$ that is invariant under the symmetry condition (A.1) and satisfies this ellipticity condition (although the Sobolev results will depend on the order of the operator).

A.2. Proof of Theorem 3.5. Assume $(\varphi_0, p_0) \notin \operatorname{bd}(A)$. Then, there are two cases: either $(\varphi_0, p_0) \in \operatorname{int}(A)$, the interior of A, or $(\varphi_0, p_0) \in \operatorname{ext}(A)$, the exterior of A.

If $(\varphi_0, p_0) \in \text{int}(A)$ then there is a neighborhood U of (φ_0, p_0) that is contained in A. By Lemma A.1 part (i), since Rf and $\mathbb{1}_A Rf$ agree in U, $WF_{(\varphi_0, p_0)}(Rf) = WF_{(\varphi_0, p_0)}(\mathbb{1}_A Rf)$. The proof is finished using (A.6) and (A.9). The proof of part (ii) follows the same argument except that $\mathbb{1}_A Rf$ is zero (so smooth) in a neighborhood of (φ_0, p_0) .

The last part of the theorem is proven using part (i) and the fact that every line through x is parameterized by points in int(A).

A.3. Proof of Theorems 3.6 and 3.8. In the proofs of Theorems 3.6 and 3.8, we use Proposition A.3 to analyze how multiplication by $\mathbb{1}_A$ adds singularities to the data Rf and

then to the reconstruction, $\mathcal{L}_A f$. We first make observations that will be useful in the proofs of both theorems.

Let A satisfy Assumption 3.1 and let $f \in L^2(D)$. Let

 $G = \mathbb{1}_A R f$ then $R^* \Lambda G = \mathcal{L}_A f.$

By Remark A.5 and the statements in Proposition A.3,

(A.9)
$$WF(\mathcal{L}_A f) = C^t \circ WF(G).$$

Using the expression (A.4) for C, one can show for $(\varphi_0, p_0) \in [0, 2\pi] \times \mathbb{R}$ that

(A.10)
$$C \circ (N^*(L(\varphi_0, p_0)) \setminus \mathbf{0}) = T^*_{(\varphi_0, p_0)}([0, 2\pi] \times \mathbb{R}) \setminus P$$

and
$$P = \{(x, \omega \overline{\theta}(\varphi_0) dx) : x \in L(\varphi_0, p_0), \ \omega \in \mathbb{R}\}$$
$$P = \{(\varphi, p, \omega d\varphi) : (\varphi, p) \in [0, 2\pi] \times \mathbb{R}, \ \omega \in \mathbb{R}\}.$$

Because WF(Rf) = $C \circ WF(f)$, equation (A.10) implies that if f is smooth conormal to $L(\varphi_0, p_0)$, then Rf is smooth at (φ_0, p_0) .

Using analogous arguments for C^t , one shows for $(\varphi, p) \in [0, 2\pi] \times \mathbb{R}$ that

(A.11)
$$C^{t} \circ \left(T^{*}_{(\varphi_{0},p_{0})}([0,2\pi] \times \mathbb{R})\right) = N^{*}(L(\varphi_{0},p_{0})) \setminus \mathbf{0}.$$

By (A.9), if G is smooth near (φ_0, p_0) then $\mathcal{L}_A f$ is smooth conormal to $L(\varphi_0, p_0)$.

Proof of Theorem 3.6. Assume bd(A) is smooth with finite slope at (φ_0, p_0) . Therefore, there is an open neighborhood I of φ_0 and a smooth function $p = p(\varphi)$ for $\varphi \in I$ such that $(\varphi, p(\varphi)) \in bd(A)$. A straightforward calculation shows for each $\varphi \in I$ and each $\omega \neq 0$ that

$$\eta(\varphi) = \left(\varphi, p(\varphi), \omega\left(-p'(\varphi)\mathrm{d}\varphi + \mathrm{d}p\right)\right)\right)$$

is conormal to bd(A) at $(\varphi, p(\varphi))$. A calculation using (A.6) and (A.9) shows that

(A.12)
$$\eta(\varphi) \in WF(G)$$
 if and only if $(x_b(\varphi), \omega\overline{\theta}(\varphi)d\mathbf{x}) \in WF(\mathcal{L}_A f)$,

where $x_b(\varphi)$ is given by (3.5). Then, $(x_b(\varphi_0), \omega \overline{\theta}(\varphi_0) dx)$ is the possible object-independent artifact that could occur on $L(\varphi_0, p_0)$.

If f has no singularities conormal to $L(\varphi_0, p_0)$, then Rf is smooth so $WF(G) \subset WF(\mathbb{1}_A)$ by Lemma A.1 part (ii). This proves (3.6).

If $Rf(\varphi_0, p_0) \neq 0$, then $WF_{(\varphi_0, p_0)}(G) = \{\eta(\varphi_0)\}$ by Lemma A.1 part (ii). Now, by (A.12), $(x_b(\varphi_0), \omega\overline{\theta}(\varphi_0)d\mathbf{x}) \in WF(\mathcal{L}_A f)$. This proves part (i).

If Rf is zero in a neighborhood of (φ_0, p_0) , then G is smooth near (φ_0, p_0) so, by the note below (A.11), $\mathcal{L}_A f$ is smooth. This proves (ii) and finishes the proof of part I.

To prove part II. we make a simple observation. Singularities of f conormal to $L(\varphi_0, p_0)$ can cause singularities in G only above (φ_0, p_0) and those can cause singularities of $\mathcal{L}_A f$ only conormal to $L(\varphi_0, p_0)$.

Proof of Theorem 3.8. The first observation is straightforward: if bd(A) is not smooth at (φ_0, p_0) , then that singularity can cause singularities in G at (φ_0, p_0) which cause singularities of $\mathcal{L}_A f$ conormal to $L(\varphi_0, p_0)$.

Assume f is smooth conormal to $L(\varphi_0, p_0)$, $Rf(\varphi_0, p_0) \neq 0$, and A has a corner at (φ_0, p_0) (see Definition 3.2). Then, by Lemma A.1, $WF_{(\varphi_0, p_0)}(G) = WF_{(\varphi_0, p_0)}(\mathbb{1}_A)$ which is equal to $T^*_{(\varphi_0, p_0)}([0, 2\pi] \times \mathbb{R}) \setminus \mathbf{0}$. Therefore, by (A.11), $WF_{L(\varphi_0, p_0)}(\mathcal{L}_A f) = N^*(L(\varphi_0, p_0)) \setminus \mathbf{0}$. This finishes the proof of Theorem 3.8

Remark A.6. Here we provide an example for which Theorems 3.6 and 3.8 apply but $\mathcal{L}_A f$ has no artifacts. This explains why one can, in general, only say artifacts "can occur."

Let

$$A = \{(\varphi, p) : \varphi \in [0, \pi/2], \ p \le 0 \text{ or } \varphi \in [\pi, 3\pi/2], \ p \ge 0\}.$$

Then the set of lines in A do not meet the open first quadrant of the coordinate plane. Let f be the characteristic function of $[0, 1]^2$, then $\mathcal{L}_A f = 0$.

Although A has a non-vertical boundary, there is no x_b curve. Although f has singularities conormal to L(0,0) and $(0,0) \in bd(A)$, there is no object-dependent artifact in $\mathcal{L}_A f$. Furthermore, bd(A) has a corner, but there are no corner streak artifacts in $\mathcal{L}_A f$.

A.4. Proof of Theorem 5.2. We first prove a proposition giving the correspondence between Sobolev wavefront set and R^* .

Proposition A.7 (Sobolev wavefront correspondence for R and R^*). Let $(\varphi_0, p_0) \in [0, 2\pi] \times \mathbb{R}$, $\omega_0 \neq 0$, and let s and α be real numbers. Let

$$\eta_0 = \omega_0(-\alpha \mathrm{d}\theta + \mathrm{d}p), \quad x_0 = p_0\overline{\theta}(\varphi_0) + \alpha\theta^{\perp}(\varphi_0), \text{ and } \xi_0 = \omega_0\overline{\theta}(\varphi_0)\mathrm{d}x.$$

Let f be a distribution on \mathbb{R}^2 and g a distribution on $[0, 2\pi] \times \mathbb{R}$. Then,

(A.13)
$$(x_0,\xi_0) \in WF_s(f) \iff (\varphi_0,p_0,\eta_0) \in WF_{s+1/2}(Rf),$$

(A.14)
$$(\varphi_0, p_0, \eta_0) \in \mathrm{WF}_s(g) \Longleftrightarrow (x_0, \xi_0) \in WF_{s+1/2}(R^*g).$$

Proof. Equivalence (A.13) is given [40, Theorem 3.1], however the proof of the \Leftarrow implication for R was left to the reader.

The proof of the \Rightarrow implication of (A.14) is completely analogous to the proof given in [40] for R. For completeness, we will prove the \Leftarrow implication. Assume g is in H_s at (φ_0, p_0, η_0) . By [38, Theorem 6.1, p. 259], we can write $g = g_1 + g_2$ where $g_1 \in H_s$ and $(\varphi_0, p_0, \eta_0) \notin WF(g_2)$. The operator R^* is continuous in Sobolev spaces from H_s to $H_{s+1/2}^{\text{loc}}$ by [48, Theorem VIII 6.1] since C^t is a local canonical graph. Therefore $R^*g_1 \in H_{s+1/2}^{\text{loc}}$. Since $(\varphi_0, p_0, \eta_0) \notin WF(g_2)$, $(x_0, \xi_0) \notin WF(R^*g_2)$ by the wavefront correspondence (A.6). An exercise using Definition 5.1 and the Fourier transform shows that $R^*g = R^*g_1 + R^*g_2$ is in H_s at (x_0, ξ_0) .

Proof of Theorem 5.2. Let $f \in L^2(D)$ and let A satisfy Assumption 3.1. Let $(\varphi_0, p_0) \in$ bd(A) and assume $Rf(\varphi_0, p_0) \neq 0$ and f is smooth conormal to $L(\varphi_0, p_0)$. Because f is smooth conormal to $L(\varphi_0, p_0)$, WF $_{(\varphi_0, p_0)}(Rf) = \emptyset$ so Rf is smooth in a neighborhood of (φ_0, p_0) by Lemma A.1 part (iii). Since $Rf(\varphi_0, p_0) \neq 0$, for each s,

(A.15)
$$(WF_{s-1})_{(\varphi_0,p_0)}(\Lambda \mathbb{1}_A Rf) = (WF_s)_{(\varphi_0,p_0)}(\mathbb{1}_A Rf) = (WF_s)_{(\varphi_0,p_0)}(\mathbb{1}_A);$$

the left-hand equality is true because Λ is an elliptic pseudodifferential operator of order one in these directions and, the right-hand equality is true by arguments similar to the proof of part (ii) of Lemma A.1.

To prove part (i) of the theorem, assume $\operatorname{bd}(A)$ is smooth and has finite slope at (φ_0, p_0) . Because the Sobolev wavefront set is contravariant under diffeomorphism [48], we may assume $\operatorname{bd}(A)$ is a horizontal line, at least locally near (φ_0, p_0) . Let $\eta_0 = \operatorname{dp}$. We claim that $(\varphi_0, p_0, \pm \eta_0) \in \operatorname{WF}_{1/2}(\mathbb{1}_A)$ and, for s < 1/2, $\mathbb{1}_A$ is in H_s at $(\varphi_0, p_0, \pm \eta_0)$. Furthermore $\mathbb{1}_A$ is smooth in every other direction above (φ_0, p_0) . The proofs of these two statements are now outlined. Using a product cutoff function $\psi = \psi_1(\varphi)\psi_2(p)$ to calculate $\mathcal{F}(\psi\mathbb{1}_A)$ and integrations by parts, one can show that this localized Fourier transform is of the form $S(\nu)T(\tau)$ where S is a smooth, rapidly decreasing function and T is $\mathcal{O}(1/|\tau|)$. Therefore $S(\nu)T(\tau)$ is rapidly decaying in all directions but the vertical. This implies that $\mathbb{1}_A$ is in H_s for s < 1/2 at $(\varphi_0, p_0, \pm \eta_0)$ and $(\varphi_0, p_0, \pm \eta_0) \in \operatorname{WF}_{1/2}(\mathbb{1}_A)$. This also shows that this localized Fourier transform is rapidly decaying in all directions except $\pm \eta_0$. Now, using (A.15) one sees that $(\varphi_0, p_0, \pm \eta_0) \in \operatorname{WF}_{-1/2}(\Lambda \mathbb{1}_A Rf)$; $\Lambda \mathbb{1}_A Rf$ is in H_s for s < -1/2 at $(\varphi_0, p_0, \pm \eta_0)$; and (φ_0, p_0, η_0) .

Now, by Proposition A.7, $\mathcal{L}_A f = R^* \Lambda \mathbb{1}_A R f$ is in H_s at $(x_b(\varphi_0), \pm \overline{\theta}(\varphi_0))$ for s < 0 and

$$(x_b, \pm \overline{\theta}(\varphi_0)) \in WF_0(\mathcal{L}_A f)$$

where $x_b(\varphi_0)$ is given by (3.5). Using this theorem again, one sees that for any $x \in L(\varphi_0, p_0)$, if $x \neq x_b(\varphi_0)$,

$$(x, \pm \overline{\theta}(\varphi_0) \mathrm{dx}) \notin \mathrm{WF}(\mathcal{L}_A f).$$

Therefore, the only covectors in $N^*(L(\varphi_0, p_0)) \cap WF(\mathcal{L}_A f)$ are $(x_b(\varphi_0), \alpha \overline{\theta}(\varphi_0) dx)$ for $\alpha \neq 0$.

To prove part (ii), assume bd(A) has a corner at (φ_0, p_0) . Let α_1 and α_2 be the slopes at (φ_0, p_0) of the two parts of bd(A). Let

(A.16)
$$\eta_j = -\alpha_j d\varphi + dp, \quad x_{bj} = p_0 \overline{\theta}(\varphi_0) + \alpha_j \theta^{\perp}(\varphi_0), \qquad j = 1, 2.$$

An argument similar to the diffeomorphism/integration by parts argument in the last part of the proof is used. First a diffeomorphism is used to transform the corner so, *locally* Abecomes $A = \{(\varphi, p) : \varphi \ge 0, p \ge 0\}$. Then one uses a product cutoff $\psi = \psi_1(\varphi)\psi_2(p)$ at (0, 0). Then, the Fourier transform can be written $\mathcal{F}(\psi \mathbb{1}_A) = S(\nu)T(\tau)$ where $S(\nu) = \mathcal{O}(1/|\nu|)$ and $T(\tau) = \mathcal{O}(1/|\tau|)$. So, the localized Fourier transform is decreasing of order -1 in the dp (vertical) and d φ (horizontal) directions and -2 in all other directions.

Note that η_1 and η_2 are the images of dp and d φ under the diffeomorphism back to the original coordinates. By contravariance of Sobolev wavefront set under diffeomorphism, $(\varphi_0, p_0, \pm \eta_j) \in WF_{-1/2}(\Lambda \mathbb{1}_A R f)$ and, for s < -1/2, $\Lambda \mathbb{1}_A R f$ is in H_s at (φ_0, p_0, η_j) . Other covectors are in $WF_{1/2}(\Lambda \mathbb{1}_A R f)$. One finishes the proof using (A.14).

This proof shows that $C^t \circ \{(\varphi_0, p_0, \eta_j)\} \in WF_0(\mathcal{L}_A f)$, and these are the "more singular points" referred to after the statement of Theorem 5.2. If one part of bd(A) is vertical at (φ_0, p_0) , then for one value of j, η_j is parallel $d\varphi$ and $C^t \circ \{(\varphi_0, p_0, \eta_j)\} = \emptyset$ so there is only one point, not two on $L(\varphi_0, p_0)$ on which f is more singular.

Part (iii) is valid for the following reasons. Since $(\varphi_0, p_0) \in \operatorname{int}(A)$, $\mathbb{1}_A Rf = Rf$ in a neighborhood of (φ_0, p_0) so $(WF_{s-1/2})_{(\varphi_0, p_0)}(\Lambda \mathbb{1}_A Rf) = (WF_{s+1/2})_{(\varphi_0, p_0)}(Rf)$ (ignoring $d\varphi$ directions). One finishes the proof using Proposition A.7.

Acknowledgements. We thank the Japan Synchrotron Radiation Research Institute for the allotment of beam time on beamline BL20XU of SPring-8 (Proposal 2015A1147) that provided the raw data described in Section 7. Todd Quinto thanks John Schotland and Guillaume Bal for a stimulating discussion about multiplying distributions that relates to Remark A.4 and the reasons we consider only functions in this article. He thanks the Technical University of Denmark and DTU Compute for a wonderful semester during which this research was being done and he is indebted to his colleagues there for stimulating, enjoyable discussions that influenced this work.

In addition, the authors thank the funding agencies listed at the start of the article. Quinto also thanks the Tufts University Deans of Arts and Sciences for their support.

The authors thank the referees for important comments that greatly improved the exposition and theorems, and in particular, the proofs of Theorems 3.6 and 3.8.

REFERENCES

- L. L. BARANNYK, J. FRIKEL, AND L. V. NGUYEN, On artifacts in limited data spherical radon transform: curved observation surface, Inverse Problems, 32 (2016), p. 015012.
- [2] R. H. T. BATES AND R. M. LEWITT, Image reconstruction from projections: I: General theoretical considerations, III: Projection completion methods (theory), IV: Projection completion methods (computational examples), Optik, 50 (1978), pp. I: 19–33, II: 189–204, III: 269–278.
- [3] F. E. BOAS AND D. FLEISCHMANN, CT artifacts: Causes and reduction techniques, Imaging Med., 4 (2012), pp. 229–240.
- [4] J. BOMAN AND E. T. QUINTO, Support theorems for real analytic Radon transforms, Duke Math. J., 55 (1987), pp. 943–948.
- [5] L. BORG, J. S. JØRGENSEN, J. FRIKEL, AND J. SPORRING, Reduction of variable-truncation artifacts from beam occlusion during in situ x-ray tomography, Measurement Science and Technology, 28 (2017), p. 124004, http://stacks.iop.org/0957-0233/28/i=12/a=124004.
- [6] L. BORG, J. S. JØRGENSEN, AND J. SPORRING, Towards characterizing and reducing artifacts caused by varying projection truncation, tech. report, Department of Computer Science, University of Copenhagen, 2017/1, http://static-curis.ku.dk/portal/files/181386847/teknisk_rapport_leise_borg1. pdf. 42pp.
- [7] J. K. CHOI, H. S. PARK, S. WANG, Y. WANG, AND J. K. SEO, Inverse problem in quantitative susceptibility mapping, SIAM J. Imaging Sci., 7 (2014), pp. 1669–1689.
- [8] J. J. DUISTERMAAT, Fourier integral operators, vol. 130 of Progress in Mathematics, Birkhäuser, Inc., Boston, MA, 1996.
- [9] A. FARIDANI, E. L. RITMAN, AND K. T. SMITH, Local tomography, SIAM J. Appl. Math., 52 (1992), pp. 459–484.
- [10] D. V. FINCH, I.-R. LAN, AND G. UHLMANN, Microlocal Analysis of the Restricted X-ray Transform with Sources on a Curve, in Inside Out, Inverse Problems and Applications, G. Uhlmann, ed., vol. 47 of MSRI Publications, Cambridge University Press, 2003, pp. 193–218.
- [11] F. G. FRIEDLANDER, Introduction to the theory of distributions, Cambridge University Press, Cambridge, second ed., 1998.
- [12] J. FRIKEL AND E. T. QUINTO, Characterization and reduction of artifacts in limited angle tomography, Inverse Problems, 29 (2013), p. 125007.
- [13] J. FRIKEL AND E. T. QUINTO, Artifacts in incomplete data tomography with applications to photoacoustic tomography and sonar, SIAM J. Appl. Math., 75 (2015), pp. 703–725.
- [14] J. FRIKEL AND E. T. QUINTO, Limited data problems for the generalized Radon transform in \mathbb{R}^n , SIAM J. Math. Anal., 48 (2016), pp. 2301–2318.
- [15] A. GREENLEAF AND G. UHLMANN, Non-local inversion formulas for the X-ray transform, Duke Math. J., 58 (1989), pp. 205–240.

ARTIFACTS FROM ARBITRARY INCOMPLETE CT DATA

- [16] V. GUILLEMIN, On some results of Gelfand in integral geometry, Proc. Symp. Pure Math., 43 (1985), pp. 149–155.
- [17] V. GUILLEMIN AND S. STERNBERG, Geometric Asymptotics, American Mathematical Society, Providence, RI, 1977.
- [18] B. N. HAHN AND E. T. QUINTO, Detectable singularities from dynamic Radon data, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1195–1225, https://doi.org/10.1137/16M1057917.
- [19] M. G. HAHN AND E. T. QUINTO, Distances between measures from 1-dimensional projections as implied by continuity of the inverse Radon transform, J. Wahrscheinlichkeitstheorie verw Gebiete, 70 (1985), pp. 361–380, https://doi.org/10.1007/BF00534869.
- [20] L. HÖRMANDER, Fourier Integral Operators, I, Acta Math., 127 (1971), pp. 79–183.
- [21] L. HÖRMANDER, The analysis of linear partial differential operators. I, Classics in Mathematics, Springer-Verlag, Berlin, 2003. Distribution theory and Fourier analysis, Reprint of the second (1990) edition [Springer, Berlin].
- [22] A. KATSEVICH, Local tomography for the limited-angle problem, Journal of Mathematical Analysis and Applications, 213 (1997), pp. 160 – 182, https://doi.org/10.1006/jmaa.1997.5412.
- [23] V. P. KRISHNAN AND E. T. QUINTO, Microlocal Analysis in Tomography, in Handbook of Mathematical Methods in Imaging, O. Scherzer, ed., Springer Verlag, New York, Berlin, 2015, ch. 18, pp. 847–902.
- [24] P. KUCHMENT, K. LANCASTER, AND L. MOGILEVSKAYA, On local tomography, Inverse Problems, 11 (1995), pp. 571–589.
- [25] F. LAUZE, Y. QUÉAU, AND E. PLENGE, Simultaneous reconstruction and segmentation of ct scans with shadowed data, in Scale Space and Variational Methods in Computer Vision, F. Lauze, Y. Dong, and A. B. Dahl, eds., Cham, 2017, Springer International Publishing, pp. 308–319.
- [26] A. K. LOUIS, Ghosts in tomography, the null space of the Radon transform, Math. Methods Appl. Sci., 3 (1981), pp. 1–10.
- [27] A. K. LOUIS, Incomplete data problems in X-ray computerized tomography I. Singular value decomposition of the limited angle transform, Numer. Math., 48 (1986), pp. 251–262.
- [28] F. NATTERER, Efficient implementation of 'optimal' algorithms in computerized tomography, Math. Methods Appl. Sci., 2 (1980), pp. 545–555.
- [29] F. NATTERER, The mathematics of computerized tomography, Classics in Mathematics, Society for Industrial and Applied Mathematics (SIAM), New York, 2001.
- [30] F. NATTERER AND F. WÜBBELING, Mathematical methods in image reconstruction, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.
- [31] L. V. NGUYEN, How strong are streak artifacts in limited angle computed tomography?, Inverse Problems, 31 (2015), p. 055003.
- [32] L. V. NGUYEN, On artifacts in limited data spherical radon transform: flat observation surfaces, SIAM J. Math. Anal., 47 (2015), pp. 2984–3004, https://doi.org/10.1137/140980740.
- [33] L. V. NGUYEN, On The Strength Of Streak Artifacts In Filtered Back-projection Reconstructions For Limited Angle Weighted X-Ray Transform, J. Fourier Anal. Appl., 23 (2017), pp. 712–728.
- [34] B. PALACIOS, G. UHLMANN, AND Y. WANG, Reducing Streaking Artifacts in Quantitative Susceptibility Mapping, SIAM J. Imaging Sci., 10 (2017), pp. 1921–1934.
- [35] B. PALACIOS, G. UHLMANN, AND Y. WANG, Quantitative analysis of metal artifacts in X-ray tomography, tech. report, University of Washington, 2018. arXiv:1712.01974 [math.AP].
- [36] X. PAN, E. Y. SIDKY, AND M. VANNIER, Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?, Inverse Problems, 25 (2009), p. 123009.
- [37] H. S. PARK, J. K. CHOI, AND J. K. SEO, Characterization of metal artifacts in x-ray computed tomography, Comm. Pure Appl. Math., 70 (2017), pp. 2191–2217, https://doi.org/10.1002/cpa.21680.
- [38] B. PETERSEN, Introduction to the Fourier Transform and Pseudo-Differential Operators, Pittman, Boston, 1983.
- [39] E. T. QUINTO, The dependence of the generalized Radon transform on defining measures, Trans. Amer. Math. Soc., 257 (1980), pp. 331–346.
- [40] E. T. QUINTO, Singularities of the X-ray transform and limited data tomography in R² and R³, SIAM J. Math. Anal., 24 (1993), pp. 1215–1225.
- [41] E. T. QUINTO, Exterior and Limited Angle Tomography in Non-destructive Evaluation, Inverse Problems,

14 (1998), pp. 339–353.

- [42] E. T. QUINTO, An Introduction to X-ray tomography and Radon Transforms, in The Radon Transform and Applications to Inverse Problems, G. Ólafsson and E. T. Quinto, eds., vol. 63 of AMS Proceedings of Symposia in Applied Mathematics, Providence, RI, USA, 2006, American Mathematical Society, pp. 1–23.
- [43] E. T. QUINTO, Support Theorems for the Spherical Radon Transform on Manifolds, International Mathematics Research Notices, 2006 (2006), pp. 1–17. Article ID = 67205.
- [44] G. RIGAUD, On Analytical Solutions to Beam-Hardening, Sens. Imaging, 18: 5 (2017), https://doi.org/ 10.1007/s11220-016-0153-4.
- [45] W. RUDIN, Functional analysis, McGraw-Hill Book Co., New York, 1973. McGraw-Hill Series in Higher Mathematics.
- [46] L. SHEN, E. T. QUINTO, S. WANG, AND M. JIANG, Simultaneous reconstruction and segmentation with the Mumford-Shah functional for electron tomography, in 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 5909–5912, https://doi. org/10.1109/EMBC.2016.7592073.
- [47] P. STEFANOV AND G. UHLMANN, Is a curved flight path in SAR better than a straight one?, SIAM J. Appl. Math., 73 (2013), pp. 1596–1612.
- [48] F. TRÈVES, Introduction to Pseudodifferential and Fourier Integral Operators, Volume 2: Fourier Integral Operators, Plenum Press, New York and London, 1980.
- [49] J. VOGELGESANG AND C. SCHORR, Iterative Region-of-Interest Reconstruction from Limited Data using Prior Information, Sens. Imaging, 18: 16 (2017), https://doi.org/10.1007/s11220-017-0165-8.
- [50] F. W. WARNER, Foundations of differentiable manifolds and Lie groups, vol. 94 of Graduate Texts in Mathematics, Springer-Verlag, New York-Berlin, 1983. Corrected reprint of the 1971 edition.
- [51] Y. YANG, S. S. HAKIM, S. BRUNS, K. N. DALBY, K. UESUGI, S. L. S. STIPP, AND H. O. SØRENSEN, Effect of cumulative surface on pore development in chalk, tech. report, University of Copenhagen and SPring-8, 2018, https://arxiv.org/abs/1704.01062. submitted, 22pp.

4.2 Summary

In Section 4.1 we presented characterizations of all types of incomplete data artifacts. All artifacts arise because of points $(\varphi_0, p_0) \in bdA$, and they fall into two categories: artifacts on lines and artifacts on curves:

- 1. Streak artifacts along the line $L(\varphi_0, p_0)$:
 - Object-dependent streaks when bd(A) is smooth at (φ_0, p_0) and a singularity of f is normal to $L(\varphi_0, p_0)$
 - Object-independent streaks when bd(A) is nonsmooth at (φ_0, p_0) .
- 2. Artifacts on curves:
 - They are object-independent, and they are generated by the map $\varphi \to x_b \bar{\theta}(\varphi)$ from parts of bd(A) that are smooth and of small slope.

The article described the strength of the artifacts in Sobolev scale in specific cases. Further, we provided illustrations of the theoretical results and proved the advantages of including a smooth cutoff across singularities in the incomplete data.

Image analysis of structures in 3D, part 1

The aim of establishing comprehensible and easily accessible image analysis in synchrotron imaging for the MAX4Imagers was directly in line with the aims of this thesis. The following work relates to analyses of images of muscle biopsies taken from humans, and is covered in Section 5 and in Section 6. Parts of the work in the present section will be submitted for publication together with other analyses not ready yet.

5

Neuromuscular diseases are caused by an unbalanced signaling between the brain and the muscles, which often leads to involuntary passive or active muscle contractions. One symptom often seen in neuromuscular diseases is muscle atrophy which is reflected as smaller muscle volume, and hence smaller cross-section areas of the individual muscle fibers [32].

The motivation of the study in this section is described in the following. In [8], the authors used an isotropy measure for distinguishing between states of muscle tissues. The study was carried out on rat muscle images from healthy rats and rats that had had botox injected. The isotropy index was measured by a stand-alone software tool, Quant3D [33]. Our aim was to conduct similar analyses on the humane biopsies and compare their results with ours. We followed exactly the same procedures with only one exception; In [8] they used a type of Markov random field for segmentation (see [8] for details), where we used a Chan-Vese segmentation [7]. Apart from measuring the isotropy index, we included two measures for assessing the atrophy of the muscles. The two measures were based on two different segmentation methods, the Chan-Vese segmentation, and a 2D snake segmentation method [9]. This section contains subsections describing the data, methods, results, discussion, and a summary.

5.1 Data

A total of 41 biopsies were taken from the m. Gastrocnemius in four classes of participants: 3 biopsies from participants with apoplexy (A), 21 biopsies from participants with cerebral palsy (CP), 7 biopsies from participants with spinal cord injury (S), and 10 biopsies from healthy control participants (H). The biopsies were then embedded in Bouin's fluid and stored at 4 degrees Celsius for 48 hours. After that, the biopsies were washed with 98% Ethanol for 5 min. six times. Subsequently, the biopsies were stored in 98% EtOH at 4 degrees Celsius until imaging. A subpart of size $844 \times 844 \times 713 \ \mu\text{m}3$ of each muscle biopsy was imaged by the TOMCAT beamline at the Swiss Light Source [34]. The images consisted of 2160 slices, each with a resolution of 2560×2560 pixels. Stacking these images provided a 3D image of size $2560 \times 2560 \times 2160$ pixels. One example slice (number 1000) from each class is shown in Figure 5.1. The shape of the muscle fibers is generally elongated. The data was captured in such a way that many fibers are extending perpendicularly to volume slices, while few are extending in the plane of the slices. From the examples, it is difficult to see the difference between the A, CP, and H-classes. However, the S-class looks different since the number of muscle fibers seems to be smaller, they are not lying side-by-side in clusters, and "granular" tissue is in the space between the muscle fibers.


Figure 5.1: The 1000'nd slice from the four 3D images - one from each class. The participant ID is indicated in the respective captions. The images are reconstructions provided by the Swiss Light Source Facilities.

5.2 Methods

The resolution was reduced by a factor of four in all three directions, resulting in a $640 \times 640 \times 540$ image. After that, the outermost pixels were cut out to reduce the image size to $303 \times 303 \times 530$ voxels, corresponding to a volume of interest (VOI) of size $0.4 \times 0.4 \times 0.7$ mm³. This reduction was done to be able to compare the analysis to that of [8], where they reduced similarly. We used a 2D Chan-Vese method [7] to segment the fibrillar tissue in a slice-by-slice setting. For measuring the isotropy we used the same procedure as described in [8]: The non-fibrillar from the segmentations were used as inputs, where 10,000 points with 513 random orientations were used to calculate the mean object lengths of the objects. They were used to find the so-called star length distribution (SLD), which provides the anisotropy of the objects in the segmented image. A value close to one reflects a more isotropic structure, indicating a higher degree of disorganization. The volume fraction is the fraction of fibrillar tissue over the entire volume. To find the volume fraction, the number of pixels belonging to the fibrillar volume (according to the Chan-Vese segmentation) was counted and divided by the number of pixels in the entire slice. This was done for all segmented slices in each sample.

As a supplement to the Chan-Vese segmentation, a snake method [9] was used for providing cross-section areas of the fibers. The method was used on slice number 1000 in each sample. The area of four fibers in slice number 1000 was measured for all samples where muscle fibers were present and clear to see. Four random points on the slice were used to pick the fibers to measure. If the random points were not on top of a fiber or hit a fiber that was already chosen, the nearest fiber was chosen. Conditions for the fibers to enter in the area measurements were: 1) the entire fiber had to be visible on the slice, 2) it had to (seemingly) extend perpendicularly to the slice plane of the image, 3) the fiber had to be inside the circular area defined by the larges circle that fits the images. The means, the standard deviations (SD), and standard error of the means (SEM) were calculated for the isotropies and volume fractions for each of the four classes, where SEM = $\frac{SD}{\sqrt{n}}$.

5.3 Results

A zoom of the images in Figure 5.1 are shown in Figure 5.2. The size of the zoom is $0.4 \times 0.4 \text{ mm}^2$, corresponding to the size of the slices after the described reduction. Four examples of the 2D segmentations are shown in Figure 5.3 and can be compared with the images in Figure 5.2. The 3D segmentations of the fibrillar tissue are shown in Figure 5.4, Figure 5.5 shows the non-fibrillar 3D segmentations, and Figure 5.6 shows the SLD for the non-fibrillar tissue. An example of segmentations by use of the snake is shown in Figure 5.7.



(c) H6

(d) S8

Figure 5.2: The 1000'nd slice of the original images, zoom of slices in Figure 5.1. The images are reconstructions provided by the Swiss Ligth Source Facilities.



Figure 5.3: The 245'th slices of the segmented images: reduced resolution, reduced area, segmented in fibrillar and non-fibrillar tissue. These slices have the same position in the samples as the high-resolution companion slices shown in Figure 5.2 and are therefore directly comparable.



(a) A3

(b) CP10



Figure 5.4: The 3D volumes of the segmented fibrillar tissue.



Figure 5.5: The 3D volume of the segmented non-fibrillar tissue.



Figure 5.6: SLD of the non-fibrillar tissue.



Figure 5.7: Example of snake segmentations. The underlying image is a reconstruction provided by the Swiss Ligth Source Facilities.

The isotropy indices for the samples are shown in Figure 5.8. The classes are plotted on the x-axis, and the isotropy index is on the y-axis. In Figure 5.8a, the isotropy indices for each of the individual samples are plotted, in Figure 5.8b the mean and STD are shown and in Figure 5.8c the mean and SEM are shown. A similar structure is used in Figures 5.9 and Figure 5.10, which show the volume fractions and the 2D cross-section areas. Please note the different ranges of y-axes for the measures.



Figure 5.8: Isotropy of the non-fibriliar tissue. A value closer to one indicates larger degree of disorganization.



Figure 5.9: The volume fraction.



Figure 5.10: Cross-section areas.

5.4 Discussion

The Chan-Vese method used for segmentation was based on 2D analysis of the images, and a 3D implementation would most probably have improved the results as it would analyse surfaces in 3D instead of contours in 2D. Another disadvantage of the method was that it did not separate the individual fibers from each other. Therefore, we could not directly use these segmentations to calculate the cross-section areas. With the implementation of the 2D snake segmentation method, however, the curves are closed contours around each of the fiber edges. The method performed excellently and captured the edges of the fibers almost perfectly. Therefore, for measuring the cross-section areas, we used the snake segmentation on single slices of the samples. The fiber selection method for measuring the cross-section areas had a number of disadvantages; First of all, cross-section areas should be measured perpendicularly to the extension direction of the fibers. However, from a 2D view, we could not be certain that this demand was satisfied. Another disadvantage of this method was that it was more likely to choose larger fibers than smaller ones; random points in the images would be more likely to hit larger fibers than smaller ones. This would skew the cross-section areas measures towards higher values, however, all classes were affected equally by this effect. As an inter-class comparison was more important than absolute values of the cross-section areas for this analysis, we did not consider this as a major issue.

The mean isotropy indices differed only a little between the classes as seen in Figure 5.8. This observation also holds for the two other measures in Figure 5.9 and Figure 5.10. Though the differences were not significant, there are tendencies which are in accordance with our expectations: On average, H had a little lower isotropy index with a mean of around 0.6, where S had the highest mean isotropy index at around 0.7, as seen in plot 5.8b and 5.8c. This implies that the healthy control participants have a higher degree of linear structure of the muscle tissue. In the study of rat muscle biopsies, the samples from the healthy controls also had higher mean isotropy index than that of the rats that were injected by botox.

According to plot 5.8a, the variance of the isotropy index for CP was smaller compared to H - especially if disregarding the bottom point (which may be an outlier) - implying that the structures were more similar among the participants with cerebral palsy.

Figure 5.9 shows that the mean volume fraction of the H-class was higher than in any of the other classes. This was in accordance with our assumption, as atrophy is typically observed in neuromuscular diseases [32]. The mean volume fraction of the S-class was 0.45 and lower than for the other classes which were all above 0.6. This implied that the participants with spinal cord injury on average suffered from the largest degree of atrophy compared with the other classes in this study.

The fiber cross-section areas ranged from 300 to 15,000 μ m², as indicated in Figure 5.10, implying diameters of 20 to 140 μ m if the fiber cross-sections were shaped like perfect circles. The mean fiber cross-section area for H at 6600 μ m² was higher than the average of the other classes, which were 4700, 4900, and 5000 μ m² for A, CP, and S, respectively. The corresponding mean diameters, if the fibers were perfectly circular, were 92, 78, 78, and 80 μ m for H, A, CP, and S, respectively. This supports the assumption that healthy participants have larger muscle fibers compared to the diseased participants.

5.5 Summary

The structure of the muscle tissue for the neuromuscular diseased participants compared to healthy conditions was quantified by measuring the isotropy index, the volume fraction, and the fiber cross-section areas. The differences between healthy and diseased participants were not reflected very well, as only subtle differences in the measures were present. However, those subtle differences were in alignment with our assumptions of 1) loss of linear structure of muscle tissue in the diseased participants (in line with the findings in [8], where the rats that had botox injected also expressed loss of linear structure) and 2) atrophy in muscle tissue for the diseased participants compared to healthy participants [32], measured by volume fractions and muscle fiber cross-section areas. In this section, we quantify the orientation consistencies and carry out 3D morphological analyses on a subset of the images used in Section 5. Section 6.1 contains our article *Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography*, Section 6.2 provides supplementary details and discussions not given in the article, and Section 6.3 is a summary of the work.

6

6.1 Article # 3: Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography

Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron X-ray computed tomography

Leise Borg^{*1}, Jon Sporring¹, Erik Bjørnager Dam¹, Vedrana Andersen Dahl², Tim Bjørn Dyrby^{2,4}, Anders Bjorholm Dahl², and Jessica Pingel³

¹Department of Computer Science, University of Copenhagen, Denmark ²Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark ³Department of Neuroscience, University of Copenhagen, Denmark ⁴Danish Research Centre for Magnetic Resonance, Center for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre,

Hvidovre, Denmark

August 24, 2018

Abstract

Cerebral palsy lead to morphological changes in muscle fibers, and traditionally this has been investigated by quantifying muscle fiber morphology using 2D imaging methods. This will however only allow analyzing cross-sections through the tissue and not enable a precise characterization of the muscle fibers. Three-dimensional (3D) microscopic imaging from synchrotron X-ray computed tomography (SXCT) allows for investigations at a sub-micrometer resolution where precise morphological measures of individual muscle fibers can be done. In this study, we analyzed 3D SXCT images of muscle biopsies taken from participants with cerebral palsy and from healthy controls. By slicing the volumetric images we carried out 2-dimensional (2D) analysis and compared the obtained characteristics to the results obtained using full 3D analysis. We demonstrate an increased variation in muscle fiber orientations in samples from participants with cerebral palsy. Further, our analysis shows that the morphological characterization obtained using 2D imaging is not able to distinguish variation in the orientation of fibers from variation in size. Therefore, the traditional understanding that the size of muscle fibers from people with cerebral palsy varies more than healthy controls was not confirmed by our study, but the muscle fiber orientation varies more. In addition to these findings, our study shows the great potential in using 3D SXCT compared to 2D microscopy for ex-vivo microscopy.

Keywords: Synchrotron computed tomography, muscle biopsies, 3-dimensional images, cerebral palsy, orientation consistency, morphology

1 Introduction

Analysis of muscle fiber biopsies is most often done using 2D slice-based microscopy methods to measure morphological features such as average fiber cross-section area and its standard deviation.

^{*}Corresponding author's e-mail address: lebo@di.ku.dk

This approach is sensitive to the slicing direction because if the biopsy is not sliced orthogonal to the fiber orientation, it will overestimate the cross-section area. Instead, the cross-section area can be computed with much higher precision if 3D microscopy imaging is employed, which is exactly what is demonstrated in this paper.

Imaging the microstructure of tissue in 3D at a resolution similar to optical microscopy used in histopathology has become possible with the development of large-scale synchrotron X-ray imaging facilities. In this study, images were acquired at the Swiss Light Source (SLS) at the TOMCAT beamline [1], that allow for volumetric X-ray CT imaging at sub-micrometer resolution. This sophisticated imaging technique combined with advanced quantitative image analysis has allowed us to quantify the uncertainty when using 2D morphological characterization. This is demonstrated on tissue biopsies from people with cerebral palsy (CP) and healthy controls (H) in a study with a total of 30 participants.

The contributions of this work are threefold, in that we show:

- 1. Analysis of muscle fiber morphology from 2D microscopy requires very careful alignment of the biopsy with the muscle fiber direction to avoid overestimating the fiber area. Due to variation in the alignment of fibers this analysis may be associated with uncertainties. Instead, analyzing the morphology in 3D microscopy images allows an accurate morphological quantification of muscle fibers.
- 2. People with CP have smaller cross-section areas of their muscle fibers compared to healthy people, which is in line with previous findings [2]. The variance in their cross-section areas is however not significantly different between CP and healthy people, which is in contrast to previous findings. This can be explained by variation in the orientation of the fibers, resulting in an overestimation of fibers that have not been cut orthogonal to the fiber orientation.
- 3. We introduce a novel biomarker of muscle integrity, which we refer to as orientation consistency. The muscle fiber orientation consistency is a quantification of how aligned or ordered the fibers are organized in a sample. We show that the orientation consistency is significantly larger for the healthy participants than it is for the participants with CP.

This demonstrates the potential in using 3D synchrotron imaging combined with advanced image analysis for understanding the 3D morphology of muscle fibers as shown here, but it also demonstrates the potential in ex-vivo bioimaging.

Synchrotron X-ray computed tomography (SXCT) allows for three-dimensional (3D) imaging of objects at a very high resolution. The high photon flux, monochromatic, tunable beams in synchrotrons together with methods like phase contrast allows for contrast in tissue despite little difference attenuation, and therefore permit non-destructive ex-vivo examination of various specimens in various depths and at different scales [3, 4]. X-ray computed tomography is most widely used in clinical practice for in-vivo examination, but the use of SXCT for bioimaging is at a resolution that is comparable to optical microscopy, and at best only slightly lower than electron microscopy. The sample size and radiation dose will most often only allow for ex-vivo investigations, but the technique is very well suited for studying muscle fiber biopsies as done here. 3D structures of single muscle fibers on a sub-micrometer scale were clearly visible, but it did not reveal sub-structures such as myofibrils.

Muscle biopsies play an integral role in the evaluation of patients with neuromuscular diseases. Morphologic analyses are used in research settings for cerebral palsy, and common pathologic findings include a decrease in fiber size and an increase in fiber size variation [2]. The evaluations are often carried out on two-dimensional (2D) histochemical-based images or other image modalities in 2D. For reliable morphological results, the biopsy must be orientated such that muscle fibers extend perpendicularly to the slice plane [5]. However, it may occur that the biopsy has been tilted during tissue preparation or the samples have low orientation consistency, i.e. the muscle fibers are not organized parallelly and point in different directions. Then, a subsequent morphologic analysis will be prone to errors because those fibers not extending perpendicularly to the slice plane appear with altered cross-section areas and shapes. Much work has been devoted to analysing muscle biopsies and 2D microscope images of muscle biopsies (see for instance [6]. However, SXCT images of humane muscle biopsies have to our knowledge been addressed only a few times in the literature [7, 8]. This means that the appearances and relative pixel values of blood cells, muscle fibers and other tissues in SXCT scans are not yet fully established. 3D shapes may be used as an indication for determining the tissue type, though. The authors of [8] compare light microscopy histology and a corresponding phase contrast CT image and identify nerve tissue and muscle spindles in a CT scan of a mouse soleus muscle sample. Also, they identify blood vessels, blood cells, and empty capillaries.

For 2D analysis of muscle fibers, there exist several automatic and semi-automatic segmentation methods [9, 10, 11], some of which are based on active contours [12, 13, 14] whereas others combine threshold selection and morphology techniques [15, 16]. To our knowledge, there are only two reported papers of segmented muscle tissue in SXCT images [8, 17]. Some of the studies just mentioned provide segmentations only, whereas other studies provide subsequent analyses on the segmented fibers relating to muscle diseases, functional properties, degeneration and regeneration of muscles. There exists a vast amount of segmentation methods, and the optimal method must be chosen according to the data at hand and to the intended subsequent analysis.

Our aim is to make a 3D analysis on muscle fibers and to test whether a 3D analysis should be considered in place of a 2D analysis for fiber morphology. To evaluate this, we calculate the muscle fiber cross-section areas in 2D and in 3D from the CP-class and the H-class and we validate by the following hypothesis, often reported in the literature [2]: The muscle fiber cross-section areas from participants with cerebral palsy are smaller than those from healthy participants. Also, we calculate the standard deviation of the cross-section areas (STD) both in 2D and in 3D. We introduce a novel method for measuring the orientation consistency, which is based on 3D muscle fiber orientations of the samples.

2 Materials and methods

This study focuses on a detailed morphological study of muscle fibers and the fiber orientation consistency in the 3D volumes. Since the data set is very large, we found it sufficient to investigate parts of the acquired volumetric images.

Data Our dataset consisted of images of muscle biopsies. A total of 30 biopsies was taken from m. Gastrocnemius: 20 biopsies from participants with cerebral palsy and 10 from healthy participants. The biopsies were then embedded in Bouin's fluid and stored at 4 degrees Celsius for 48 hours. After that, the biopsies were washed with 98% Ethanol for 5 min six times. Subsequently, the biopsies were stored in 98% EtOH at 4 degrees Celsius until analysis. Imaging was done at the TOMCAT beamline at the Swiss Light Source [1] using synchrotron X-ray CT. Only the central part of the biopsy was imaged, resulting in an image of $2560 \times 2560 \times 2160$ voxels. The voxel size is $(330 \text{ nm})^3$, so the depicted volume is $844 \times 844 \times 773 \ \mu\text{m}$. For details of image acquisition, please refer to [17]. The slice planes of the imaged biopsy were determined by how the biopsy was oriented in the SXCT scanner. One example slice from each of the two classes is shown in Figure 1. The shape of the muscle fibers is generally elongated. The data were captured in such a way that many fibers are extending perpendicularly to the slices in sample H6 and it seems that sample CP10 exhibits a lower orientation consistency than that of sample H6.

The image quality is much lower in the periphery because these parts of the sample were not traversed by X-rays from all directions when obtaining the projections in the SXCT-scan. Therefore, we only consider the interior of the largest circle that fits in the example images of Figure 1.

Snake segmentation Automated segmentation of individual fibers is challenging because the intensity of the fibers vary and the boundary is often not clear all the way along the periphery of the



(a) CP10

(b) H6

Figure 1: Slice number 1000 from each class. CP10 is from sample number 10 in the cerebral palsy class, and H6 is sample number 6 in the healthy class.

fiber. Therefore, we chose to use a semiautomated method based on a 2D snake [18]. The 2D snake method used the unfolded representation of the curve [19] to segment the fibers slice-by-slice. Each fiber segmentation was seeded by clicking the approximate center in a graphical user interface, and if needed, the segmentations were adjusted semi-automatically. The segmentations from one slice were passed on to the following slice and used as an offset to the subsequent segmentation. For every 20 slices, the segmentations were semi-automatically adjusted and new segmentations were initialized if new fibers had entered the image. The 2D segmentations of slices number 100 to 300 were stacked to form 3D segmentations of 66 μ m in height and of 844 μ m in diameter.

2D morphology The 2D analysis was based on the segmentations in the 201 slices. In each slice, the cross-section areas of individual fibers and the STD of the cross-section areas of the fibers were calculated, and the means of both measures were calculated from the 201 slices. The cross-section areas were calculated by counting the number of pixels inside the fiber segmentations in each slice and STD was calculated by taking the standard deviation of the cross-section areas in each slice.

3D morphology The 3D segmentations were presented as surface renderings of the segmentations in the 201 slices. 3D surface normal vectors were computed from the segmented muscle fibers in each node of the segmented surface. To calculate the cross-section areas, we identified each fiber direction by analyzing its normal vectors N_k ; The outer product of the set of normals was summarized

$$M = \sum_{k=1}^{K} N_k N_k^T$$

to obtain the matrix, M, known as the Structure Tensor [20]. In this case, it is a 3×3 matrix. For a perfectly straight fiber, most of its normals will be perpendicular to the fiber's direction, and the two eigenvectors of M corresponding to the two largest eigenvalues in absolute sense span the cross section of the fiber, and the eigenvector, d, corresponding to the smallest eigenvalue is parallel to the overall direction of the fiber. With the fiber direction at hand, we estimated the cross-section area as follows:



5

Figure 2: Two examples of 3D volumetric segmentations from 201 slices.

1. Calculate the residuals points, R(l), of the segmented points on a fiber's surface by projecting the points, P(l) of the segmentation onto the plane perpendicular to d,

$$R(l) = P(l) - (P(l) \cdot d)d.$$

- 2. Make a principal component analysis on the cloud of residual points, R(l), to obtain the principal component variances a and b in the plane perpendicular to d.
- 3. Assume the cross-section is approximately elliptical and estimate the area of the cross-section as the area of an ellipse with \sqrt{a} and \sqrt{b} as the lengths of its main axes,

$$A = \frac{\pi\sqrt{ab}}{4}.$$

Orientation consistency We developed an algorithm to quantify the fiber orientation consistency. It was estimated by the entropy of the fiber orientation distribution for each sample: All fiber orientations were represented by two vectors, v and -v, and plotted on a sphere. A 2D histogram for the azimuth-elevation coordinates with 12 bins in each angular direction was evaluated by estimating the entropy. The method was tested by comparing the true entropy on the unit sphere from a uniform distribution with 1000 random samples drawn from a uniform distribution, and the error was less than 2.1%. For further details, please refer to [21].

3 Results

Figure 2 shows two examples of the segmented volumes (note that it is not the same samples as those shown in Figure 1). The segmentations are from 201 slices, corresponding to 66 μ m in height, and the samples are tilted so they are seen from below. Figure 3 shows the results of measuring the cross-section areas in 2D and in 3D for both classes. The red stars indicate the mean measures of the samples and the error bars indicate the means \pm STD or \pm SEM. SEM is calculated as $\frac{\text{STD}}{\sqrt{S}}$, where S is the number of samples. Mean 2D cross-section areas were 6300 and 7400 μ m² for CP and H, respectively. Mean 3D cross-section areas were 2600 and 4500 μ m² for CP and H, respectively. Evaluating the differences between the classes was estimated by a two-sample t-test. The resulting p-values are shown in the respective plots. The means and standard deviations of the STD of the cross-section areas in 2D and



Figure 3: Cross-section areas in 2D and 3D.

	Cross-section area		STD of cross-section area		Entropy
	μ_{2D}	μ_{3D}	μ_{2D}	μ_{3D}	h
CP	6300	2600	5000	2400	1.04
Η	7400	4500	4700	2900	0.58
p-value	0.5	0.0004	0.9	0.4	0.04

Table 1: This table shows the mean cross-section areas in 2D and 3D in μm^2 , mean STD of the cross-section areas in 2D and 3D in μm^2 , and entropy of the fiber orientations in nits (nits = natural units of information).



Figure 4: Examples of the fiber organization in samples CP3 and H5.

in 3D are shown in Table 1. Figure 4 shows two examples of the fiber orientation consistency in the samples. The plots are in 3D but viewed from above (according to the imaging direction) with the same angle as in Figure 1. The fibers are transparent grey, and on top of the fibers, their respective directions are indicated by a randomly colored line. The mean orientation consistencies for the two classes are indicated in Table 1.

4 Discussion

In this section, we discuss the methods and results of the segmentations, the morphology, and the orientation consistency, and present the limitations and outlook of the work.

Segmentation 201 slices were chosen to ensure that enough data were analyzed to obtain reliable results from the data analysis. This number of slices covered approximately all fibers cross-section areas across all samples. The snake method allowed for analysis on single fibers since they were separated from one another in the representation. Segmenting slice by slice in 2D allowed for inspecting and correcting the segmentation quality in the slices on a regular basis. This gave very accurate segmentations allowing for precise 3D characterization of the muscle fibers. The images were processed relatively fast despite the size of the images, and this even without the use of GPUs or large storage space. For speeding up the performance further, the data fitting term and regularization term in the graph cut optimization algorithm were split into two consecutive steps instead of being combined in one single step. The drawback was that fibers with high curvatures in the slices contracted a little in each slice, which sometimes led to losing track of the fiber edges. Another drawback of using this segmentation method was the amount of time spent on each segmentation, which ranged from 10 minutes to 1 hour of focused work per sample, due to the semi-automatic initializations and adjustments of segmentations. In some samples, the fibers were difficult to distinguish from one another due to too weak edge information. Therefore, we did not consider lowering the resolution.

Morphology The mean 2D cross-section areas were in general larger than the mean 3D cross-section areas, as indicated in Figure 3 and in the two left columns of Table 1. The 2D cross-section areas of CP and H were distributed over a large range of areas, whereas the 3D measures were more densely distributed around the mean, as seen in Figure 3. This is because the cross-section areas appear larger and with greater variance when measuring in 2D slices than when measuring the true cross-section area and taking 3D effects into account.

The 2D cross-section areas for CP and H was almost indistinguishable (p-value of 0.5). However, when correcting for the fiber directions, the mean cross-section area was significantly larger for H than for CP, as observed in the literature[2]. The p-value was 0.0004. In 2D, the mean of the STD of the cross-section area shown in Table 1 was slightly increased for CP, agreeing with what is observed in the literature [2]. However, in the 3D analysis, the mean was a little increased for H, meaning that we could not reestablish the increased cross-section variation reported in the literature.

This indicates that in order to obtain reliable morphological results, the biopsy must be orientated such that muscle fibers extend perpendicularly to the slice plane. If not, a subsequent morphologic analysis will be prone to errors because those fibers not extending perpendicularly to the slice plane appear with altered cross-section areas and shapes, and 3D aspects must be considered.

Orientation consistency Measuring the orientation consistency is not trivial, but the method introduced in this work seemed to capture our observations of the differences between the classes very well. The quantification of the orientation consistency as indicated in Table 1 demonstrates that it was significantly higher in H than it was in CP, which is to the best of our knowledge a novel observation. The 2D STD of the cross-section areas may be viewed as a measure inversely related to the orientation consistency; If a sample contains fibers extending in many different directions, and therefore has a low orientation consistency, the 2D STD of the cross-section has a high value. In this view, the larger

8

cross-section variation for CP than for H reported in the literature [2], could be an indication of lower orientation consistency of the fibers, rather than a measure of the cross-section area variation. This viewpoint is in line with the present results showing a larger 2D cross-section variation for CP. This increased variation was not observed in 3D which is considered being more robust. On top of that, we measured a lower orientation consistency in CP, also supporting this argument.

We showed that morphological analysis is only correct if the fibers extend perpendicularly to the 2D slices. Also, we showed that the orientation consistency is smaller for CP. In practice, this means the 2D slice analysis is not reliable for CP participants, but - to a greater extend - reliable for healthy participants. Employing a method which is only reliable when measured on healthy participants entails a fundamental problem of the analysis.

Limitations and outlook We segmented only 201 slices from each of our 30 samples due to time constrictions. If more slices were segmented, the results could have been more precise and could have altered the quantification results. However, it could also have worsened the results as they relied on main-direction measures of the fibers. If more slices were segmented, the fibers would extend over larger volumes and one vector could not have captured its direction as well as it did with fewer slices, which could potentially lead to erroneous morphological measures. Although 201 slices may seem like a lot, they only exhibit 66 μ m, corresponding to approximately two times an average healthy muscle fiber diameter. If another subvolume of the image had been segmented, the results could have looked different. On the other hand, the number of fibers and samples should mount to a reasonable dataset size, and should, therefore, average out statistical differences. We carried out analysis on all available segmentations - also fibers touching the edges. Some samples contained almost only fibers extending in the slice planes, almost all of them touching the edges of the image. These segmentations were kept in order to keep as much information as possible. The consequence was probably smaller cross-section areas in general across all samples and in both 2D and 3D measures. As only a small fraction of the total amount of fibers touched the edges, we anticipate that the measures are still robust.

The structure tensor analysis allowed for finding the main fiber directions, which were used to calculate the 3D cross-section areas. For the vast majority of fibers, it seemed to capture the directions well. However, in rare cases, they were incorrect. As the number of incorrect fiber orientation estimates were relatively low, we considered the morphology measures to be rather reliable. The structure tensor analysis could have allowed the fiber direction to vary along its extension by splitting the analysis into separate parts of the fiber. However, we considered this method to be a decent first approach, which showed the expected differences in the cross-section areas between the classes. Obtaining the cross-section from the 3D volumes, an elliptic shape of the cloud of residual vectors was anticipated. If a fiber changes direction or has a flat contour, this is not strictly correct, but yet a sensible approximation. Because there were no outliers in the 3D cross-section areas (see Figure 3b), we felt convinced that measurements carried out on fibers were usable even if they extended in other directions. During the fiber segmentation, semi-automatic adjustments were required, which was time-consuming.

The differences revealed on the μ m scale are surprisingly small considering the condition of participants with cerebral palsy. Other imaging modalities such as magnetic resonance imaging (MRI) could reveal other and/or more statistically significant differences between the classes, such as whole-muscle cross-section areas. A combination between the two imaging modalities - and scales - would probably find differences between the classes more efficiently.

In the present work, we focused 3D analysis of muscle fibers on a μ m scale, which, to our knowledge, has not been done before. While the methods and analysis were applied to muscle fibers, the same principles are applicable to other types of fibrous materials and on other scales, such as nerve fibers or carbon fibers. Possible extensions of the 3D analysis include more morphological measures such as roundness, a skeletonization of the segmentation and staining the samples to highlight specific types of tissue, such as fiber type or cell nuclei. We leave such investigations for future work.

5 Conclusions

We carried out analyses on SXCT images of muscle biopsies in 2D slices and in 3D volumes. Common pathologic findings in CP include a decrease in fiber cross-section area and an increase in fiber size variation [2]. A comparison of the morphology between participants with cerebral palsy and healthy participants measured in the 2D slices did not show the expected cross-section area difference. A similar analysis carried out on 3D volumes did, however, show significant smaller cross-section areas for CP than for healthy participants. We used this to demonstrate that the 2D estimates of the morphology of muscle fibers are difficult to obtain because the fibers must be relatively unidirectional and they must be sliced orthogonal to their direction. With the variation in the muscle fiber direction of cerebral palsy, this is not possible. Instead, imaging in 3D gives an accurate morphological description directly from the biopsies, independent on their alignment of with the muscle fiber direction. The present study did not show significantly larger cross-section area variations in cerebral palsy when measuring in 2D, and when measuring in 3D the cross-section variation was even (insignificantly) larger in healthy participants. Interestingly, the orientation consistency – a measure of how ordered the fibers are organized in the sample – was significantly larger for the healthy participants than for the participants with cerebral palsy. Therefore, the traditional understanding that the size of muscle fibers from people with cerebral palsy varies more than healthy controls was not confirmed by our study, but the muscle fiber orientation varies more. Our study demonstrated the gain in information that is possible using SXCT because the image resolution is similar to 2D microscopy but in full 3D. With further developments in SXCT techniques, this has great promise of morphological analysis of tissue.

6 Acknowledgements

TD was supported by the Capital Region Research Foundation (grant number: A5657) on the MAX4Imagers project. JP was supported by the Danish research Council (DFF-1333-00197), and the Elsass Foundation.

References

- M. Stampanoni, A. Groso, A. Isenegger, G. Mikuljan, Q. Chen, A. Bertrand, S. Henein, R. Betemps, U. Frommherz, P. Böhler, D. Meister, M. Lange, and R. Abela. Trends in synchrotron-based tomographic imaging: the SLS experience, 2006.
- [2] M. A. Mathewson and R. L. Lieber. Pathophysiology of muscle contractures in cerebral palsy. *Phys. Med. Rehabil. Clin. N. Am.*, 26:57–504, 2015.
- [3] Y. Hwu, G. Margaritondo, and A. Chiang. Q&A: Why use synchrotron x-ray tomography for multi-scale connectome mapping? *BMC Biology*, 15:122, 2017.
- [4] Y. Hwu, W.L. Tsai, A. Groso, G. Margaritondo, and J. H. Je. Coherence-enhanced synchrotron radiology: Simple theory and practical applications. *Journal of Physics D Applied Physics*, 35, 2002.
- [5] M. S. Milenkovic, K Snezana, and S. Dragos. The modern approach to the histopathological diagnosis of muscle diseases. *Skeletal Muscle: From Pharmacology to Clinical Practice*, pages 155–170, 2015.
- [6] J. Ito, A. Araki, H. Tanaka, T. Tasaki, K. Cho, and R. Yamazaki. Muscle histopathology in spastic cerebral palsy. *Brain Dev.*, 18(4):299–303, 1996.

- [7] B. Zeller-Plumhoff, J. Mead, D. Tan, T. Roose, G. Clough, R. Boardman, and P. Schneider. Soft tissue 3d imaging in the lab through optimised propagation-based phase contrast computed tomography. *Opt. Express*, 25:33451–33468, 2017.
- [8] B. Zeller-Plumhoff, T. Roose, O. L. Katsamenis, M. N. Mavrogordato, C. Torrens, P. Schneider, and G. F. Clough. Phase contrast synchrotron radiation computed tomography of muscle spindles in the mouse soleus muscle. J. Anat., pages 859–865, 2017.
- [9] J. Mula, J. D. Lee, F. Liu, L. Yang, and C.A. Peterson. Automated image analysis of skeletal muscle fiber cross-sectional area. J. Appl. Physiol., 1(114):148–155, 2013.
- [10] P. Poggi, C. Marchetti, and R. Scelsi. Automatic morphometric analysis of skeletal muscle fibers in the aging man. Anat. Rec., 217:30–34, 1987.
- [11] L. R. Smith and E. R. Barton. Smash semi-automatic muscle analysis using segmentation of histology: a matlab application. Skeletal Muscle, 4(1), 2014.
- [12] F. Liu, A. Mackey, R. Srikuea, K. Esser, and L. Yang. Automated image segmentation of haematoxylin and eosin stained skeletal muscle cross-sections. *Journal of Microscopy*, 252:275–285, 2013.
- [13] Y. Kim, T. Brox, W. Feiden, and J. Weickert. Fully automated segmentation and morphometrical analysis of muscle fiber images. *Cytometry*, 71A:8–15, 2007.
- [14] O. Röhrle, H. Köstler, and M. Loch. Segmentation of Skeletal Muscle Fibres for Applications in Computational Skeletal Muscle Mechanics. Springer, New York, NY, in: Wittek A., Nielsen P., Miller K. (eds) edition, 2011.
- [15] Z. Wang. A semi-automatic method for robust and efficient identification of neighboring muscle cells. *Pattern Recognition*, 53:300–312, 2016.
- [16] Z. Wang and H. Li. Generalizing cell segmentation and quantification. BMC Bioinformatics, 18(1), 2017.
- [17] J. Pingel, M.S.Nielsen, T. Lauridsen, K. Rix, M. Bech, T. Alkjaer, and R. Feidenhans'l. Injection of high dose botulinum-toxin a leads to impaired skeletal muscle function and damage of the fibrilar and non-fibrilar structures. *Scientific Reports*, 7(14746), 2017.
- [18] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. Int J Comput Vision, 1(321), 1988.
- [19] Kang Li, Xiaodong Wu, Danny Z Chen, and Milan Sonka. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):119–134, 01 2006.
- [20] J. Bigun and G. Granlund. Optimal orientation detection of linear symmetry. Thesis Report, Linkoping studies in science and technology, 85, 1986.
- [21] J. Sporring. Estimating the continuous entropy of a discrete set of orientations in R³. Technical Report 2018/1, Department of Computer Science, University of Copenhagen, 2018. Available online: To come.

6.2 Supplementary material

Section 6.2.1 provides supplementary analysis and full segmentation of three samples. Section 6.2.2 gives supplementary discussions related to the methods used in our article in Section 6.1, and Section 6.2.3 compares the isotropy analyses from Sections 5 and 6.

6.2.1 Supplementary analysis and full segmentations

In our article in Section 6.1 we calculated the orientation consistency based on a measure of the isotropy on a sphere. It captured the differences quite well, but prior to this, we examined another method. In this section, this method is presented together with visualizations of three full segmentations. It is organized in the following parts: Data, Methods, Results, and Discussion.

Data In the supplementary analysis, the dataset was identical to the dataset in our article in section 6.1. The full segmentations visualized in Figure 6.2, however, where samples A3, H8, and S8.

Method In the search for capturing the differences of the orientation consistencies between the classes, we examined the following method, which we refer to as orientation consistency with weight. All fiber directions found by the structure tensor analysis as described in the article in Section 6.1, $v_1, v_2, v_3, ..., v_N$, were dotted pairwisely and weighted by the inverse of the inter-fiber distance. The result was normalized by the number of dotted fiber pairs:

$$Q = \frac{1}{N^2 - 1} \sum_{i=1}^{N} \sum_{j=1}^{N} |v_i \cdot v_j| \frac{1}{dist(c_i, c_j)}, i \neq j,$$

where the inter-fiber distance, $dist(c_i, c_j)$, was the Euclidian distance between the fiber centers of gravity. The main fiber directions, v, given by the structure tensor analysis were parallel to the fibers. However, both v and -v were solutions for the orientation of the fiber. Therefore, we used the norm of the inner product of the vector-pairs to account for both directions of v_n , namely v_n and $-v_n$; The inner product can be represented by the cosine,

$$v_i \cdot v_j = |v_i| |v_j| \cos(\theta)$$

where θ is the angle between the two vectors v_i and v_j . Since cosine is an even function, both v_n and $-v_n$ are accounted for by taking the norm of $\cos \theta$, as illustrated in Figure 6.1, where the blue line would map to the red line. Taking the norm of $\cos \theta$ is essentially the same as taking the norm of the dot product. Samples with high orientation consistencies exhibit high values of Q.



Figure 6.1: To account for both v_i and $-v_i$, the norm if $\cos \theta$ is taken.

Results In Figure 6.2, segmentations are shown for samples A3, H8, and S8, running through 1900 slices (from number 100 to number 2000). In Figure 6.2a the fibers in the cluster towards the background of the image exhibit mainly one single direction. Another cluster of fibers twirls around the cluster in the back altering direction along its extension. The fibers of H8 appear larger and with one main direction that, however, seems to alter continuously a little from left (almost vertical) to right (more vertical) in the image. The fibers in Figure 6.2c look fragmented and sizes seem to vary more. Figure 6.3 shows the measures of orientation consistency with weight.



(a) A3



(b) H8



(c) S8

Figure 6.2: Full segmentations of three samples. $\begin{array}{c} 90 \end{array}$



Figure 6.3: Orientation consistency with weight.

Discussion The orientation consistency with weight incorporated a down weight on gradual changes, as opposed to the method used in our article in section 6.1. This means that if inter-fiber directions changed gradually over the image, the Q-value would still be high even though many different fiber directions were present in the sample. With a p-value of 0.44, there was no significant difference in the orientation consistencies with weight among the classes. This was in contrast with the orientation consistency measured by anisotropy in the article. One reason for the discrepancy could be that samples with gradual changes (as exemplified in sample A3) exhibit a lower degree of orientation consistency for the measure using the anisotropy than for the measure using the weight. For capturing the differences in the orientation consistencies, an altering of the distance measure might be helpful. This is left for future work.

6.2.2 Supplementary discussions

This part contains a deeper discussion of the snake segmentation method and the structure tensor analysis. The choice of reading the images in a slice-by-slice manner as provided by the synchrotron facility and perspectives regarding machine learning used for segmentation of the muscle fibers are also discussed.

The snake segmentation method The snake method was already discussed in our article in section 6.1. Further details are given here. Each segmentation was seeded by a click in the approximate fiber center in a graphical user interface. If necessary, the segmentation was adjusted semi-automatically by 1) dragging in the graph combined with 2) automatic edge detection. Segmentations were (automatically) adjusted continuously from slice to slice and took onsets from the segmentations in previous slices. We used an unfolded version of each fiber segmentation where the coordinate system was rotated according to the segmentation curve [35]. The size of the unfolded image was defined by the number of normals of the detected edges (x-axis) and a specified range across the edges (y-axis). Figure 6.4a shows the normals across a fiber segmentation from the previous slice with lengths of 200 pixels (100 pixels away from the edge in each direction). Figure 6.4b shows the unfolded version where a change of coordinate systems was carried out. The bottom part corresponds to the inside of the muscle fiber and the top part corresponds to the outside of the muscle fiber. The negative values of the gradient taken only in the y-direction (from bottom to top) of the unfolded image are shown in Figure 6.4c. Figure 6.4d is generated by taking the negative gradient magnitude of the image in 6.4a and then unfold it.



(a) Image with normals. The reconstruction was provided by the Swiss Ligth Source Facilities.



(b) Unfolded image.

(c) Negative gradient values of (d) Unfolded image of the negathe unfolded. tive absolute gradient values.

Figure 6.4: Figure 6.4a shows an image with a fiber in the center. On top of the image, normals are plotted, ranging 100 pixels in each direction from the edge of the fiber. Figure 6.4b shows the unfolded image. Figure 6.4c shows the negative gradient values of the unfolded image, taken *only* in the vertical direction (from bottom to top). Figure 6.4d is generated by taking the negative gradient magnitude of the image and then unfold it. Using the negative gradient magnitudes in Figure 6.4d for minimizing the curve energy is the classical approach for segmenting (not in the unfolded version, though). However, in this approach, the segmentation would occasionally snap to edges of adjacent fibers because of the presence of the two dark curves (of low pixel values) in 6.4d. The bottom curve corresponds to the fiber edge, whereas the top curve corresponds to edges of adjacent fibers or other structures. When using the negative gradients only in the bottom-to-top direction illustrated in Figure 6.4c, the segmentation was repelled from adjacent fiber edges due to their high-level pixel values, and, therefore, stuck to the intended fiber edge. If, however, the segmentation would wander off and the normals would unintentionally rotate by 90 degrees tangent to the fiber edge, the use of the negative of the gradient values in all directions would be a better choice. However, a behavior like this was never observed during segmentation of the samples.

In the article, it was mentioned that the data fitting term and regularization term in the graph optimization were split into two consecutive steps instead of being combined in one single step. The regularization imposed elasticity and rigidity on the segmentations by multiplication with a matrix, confining the segmentation curvatures. Two examples of segmentations are shown in Figure 6.5: One example of 5 fibers with low curvature (6.5a), and a fiber segmentation with high curvature (6.5b).



segment

Figure 6.5: Two examples of segmentations. The fiber to the right has a saw-like shape due to too strict regularization for this specific fiber.

The reason for the saw-like shape in the reconstruction in 6.5b is twofold: 1) The fiber has a "flat-looking" profile when viewed in the slice plane (if one could see the image in 6.5b from above), resulting in high curvatures in the parts farthest away from the center. This caused contractions for every new slice (from the bottom towards the top) according to the regularization, which was too prominent for a fiber with such high curvatures. 2) For every nslices a semi-automatic correction to the true fiber periphery was carried out, broadening the fiber segmentations again. The choice of regularization strength was adjusted to handling the majority of the fibers. The consequence was saw-like segmentations for a minor part of the fibers. Adaptive regularization depending on the curvature sized could be an answer to better handling fibers like the one in Figure 6.5b.

The method did not handle fibers extending in the slice planes well, as the contours would change more drastically from slice to slice. As information was already carried from slice to slice there was some sort of 3D aspects. However, a pure 3D snake would probably be an answer to the problem and might have saved us time spent on segmentation adjustments. An estimation of the time spent on implementing a new method vs. spending time on semiautomatic adjustments made us arrive at the decision to stay using the 2D version. Also, it seemed to quite accurately detect the fiber edges based on the studies in Section 5. The choice also allowed for a rather high degree of control.

The structure tensor analysis The structure tensor used our article in Section 6.1 was the sum of the outer products of all of the segmentation surface normals (see the article for details). In [36], the authors applied the structure tensor in a 2D image. We summed up for larger parts of the fibers and applied the analysis on the 3D object rather than in the image coordinates. Figure 6.6 shows two cases of segmented fibers (black) with surface normals (red), and calculated main directions (green). The structure tensor analysis found the fiber directions quiet robustly, both for fibers extending in the slice planes and perpendicularly to the slice planes. Consider the fiber in the bottom part of Figure 6.6; Even though normals pointing upwards and downwards were missing, the overall fiber direction was calculated satisfactorily. The main fiber directions, v, given by the structure tensor analysis point in the directions parallel to the fiber extensions. However, both v and -v are solutions for the orientation of the fiber.

Prior to the structure tensor analysis, we investigated a principal component analysis (PCA) for finding the directions. It worked excellent for fibers like the bottom one in Figure 6.6, but not for fibers extending perpendicularly to the slice planes as the top one in Figure 6.6. PCA finds the principal components, where data in an *n*-dimensional space has the largest variance. For a fiber like the bottom one in Figure 6.6, the principal direction is similar or close to the extension of the fiber. However, for fibers like the top one in Figure 6.6 the principal component would find the largest distance between the bottom and top parts of the fiber, and would therefore probably extend diagonally in the fiber.

Choosing sample orientation Reading the 3D images in a side-to-side fashion instead of a bottom-to-top fashion could have been an answer to better handle fibers extending in the slice planes. For fibers changing directions, such as in Figure 6.2a, one would have to change the reading direction adaptively. Further, for samples containing fibers with various directions (also in Figure 6.2a), one would need to read the image several times and keep track of fibers already segmented. Reading the images in other directions would imply regridding and interpolation on top of the complications mentioned above. The samples were positioned in the CT-scanner such as the majority of the fibers extended perpendicularly to the slice planes, so according to the complications discussed above, the decision fell on reading the slices as provided by the synchrotron facilities.

Future work In future work could include a machine learning approach trained in edge detection, curvatures, texture, fiber size, and shapes for segmentation. To avoid compromising the high-resolution nature of the images, the method could work on subareas of the full image. A supervised machine learning approach could follow the same ideas as the 2D snake by segmenting slice-by-slice. The algorithm could be trained to look for typical changes from one slice to another and integrating information from several slices behind. Considerable amounts of training data are crucial for the performance of machine learning methods, and



Figure 6.6: Fiber surface segmentations (black) with surface normals (red) and main directions (green).

with the snake segmentations at hand, we already have immense amounts of annotated data that could be used for the training purpose.

6.2.3 Comparison between Image analysis of structures in 3D, part 1 and part 2

Both studies in Section 5 and in Section 6 analyse the isotropic behaviors of the fibers in the samples, but with very different approaches. Both approaches agree about increased fiber isotropy (which is the same as the larger degree of variations) in diseased participants compared to healthy participants. However, the results are non-significant in the first set of analyses (Section 5), but significant in the second set of analyses (Section 6) due to refined techniques and more detailed information. The snake segmentations were the main reason for the high degree of details, in that it provided clearly distinguishable segmentations of the fibers. This allowed for very precise 3D characterization of the muscle fibers.

6.3 Summary

In cerebral palsy, muscle atrophy is often observed, which is reflected as decreased crosssection areas of the individual muscle fibers [32]. We used this background knowledge for assessing the findings in our article in Section 6.1, where we compared analysis on 2D slices and 3D images of muscle biopsies. Further, we compared the orientation consistency between the healthy participants and participants with cerebral palsy. In the 3D analysis, we accounted for the various directions of the fibers and showed that the fiber cross-section areas of the healthy participants were significantly larger than they were for the participants with cerebral palsy (p = 0.0004). This result could not be re-established by analysis conducted in 2D (p-value = 0.5), as the fibers did not extend perpendicularly to the slice planes. Therefore, we argued that 3D aspects should be accounted for if fibers do not extend perpendicularly to the slice planes when carrying out morphologic analysis on muscle fibers. We also showed that orientation consistency was significantly higher in healthy participants than it was in participants with cerebral palsy (p-value = 0.04).

The structure tensor analysis used for finding the directions of the muscle fibers was quite robust.

The 2D implementation of the snake method had some problems, and in future work, we would consider a 3D snake implementation or a machine learning approach for segmentation. From this work, we already have large amounts of annotated data which could be used for training a neural network. However, in a clinical setting, it may be difficult to reproduce results across different datasets when using a machine learning approach.

The main contributions of this thesis can be seen as an approach towards better understanding the implications of synchrotron X-ray tomographic data. The work was quite interdisciplinary, where computer science and mathematical modeling met for targeting underlying practical applications in CT and medicine. A workflow diagram of CT image analysis is shown in Figure 7.1.



Figure 7.1: Workflow diagram

The topics relating to this thesis are marked as boldface text in the figure; **tomographic reconstruction** relates to the works described in Sections 3 and 4, and **segmentation**, **feature extraction**, and **analysis based on features** relate to the works described in Sections 5 and 6. The specific results described in the sections were obtained in the general framework of analysis of synchrotron X-ray tomographic data. Studying problems in specific datasets and developing practical solutions to them may take a very long time. However, if generalizations can be made, the time and effort spent can be applicable to a range of related datasets and be of interest to a broader audience. In the reconstruction problem, we analysed variable-truncation artifacts and solved the specific problem. Alongside, we generalized the underlying problem, which makes our work applicable in related settings. In the image analysis problem, we analysed muscle fiber morphology and orientation consistency in 3D. We demonstrated the gain in information from using SXCT in medical imaging and revealed interesting, new knowledge about the orientation consistency, which was higher in healthy participants than in participants with cerebral palsy. Hence, we produced valuable new knowledge, which can be adopted in CT communities and in clinical settings.

One of the main challenges when working with synchrotron data is the amount of data. Controversially, it is also the main advantage. The degree of detail and the 3D nature provided by the modality is what makes it superior to many other imaging modalities. In our work, we focused on methods that were able to handle these large amounts of data but at the same time could extract the important details.

For further development of the image analysis work, a higher degree of automatization and a 3D snake model would be of very high priority. Further, it would be interesting to implement a machine learning approach for segmentation and to compare with the 2D/3Dsnake versions. Automatic image analysis almost always demand human interaction, and tweaking parameters is easier when we can use our intuition (prior knowledge); what is the noise level, how large are the objects we are looking for, how sharp are the edges, how do the shapes look, how many phases are there, etc. The use of machine learning is often an excellent choice in many segmentation and data analysis cases due to its versatility and ability to explore high-dimensional spaces. However, the number of tunable parameters may be plentiful, and the hunt for an optimum set of parameters may be challenging. It may be difficult to reproduce results across datasets, and the discriminating features for classification and prediction are impossible to extract, which makes it difficult to use in clinical a setting. Therefore, it may be more robust to use well-known methods on new datasets in a clinical context. Tomographic reconstruction may also be a parameter hunt; parameter tuning in iterative methods with, say, just two regularization terms may not be very intuitive and difficult to balance. There exist methods that can be used for searching the parameter space, however, if a set of parameters are found as optimum for one dataset it is most likely not optimum for other datasets.

About interdisciplinarity and sharing knowledge. Methods for synchrotron tomographic reconstruction are developed at many places; developers of TomoPy and researchers at the iMinds Vision Lab at the University of Antwerp develop ASTRA, workers at the Karlsruhe Institute of Technology have developed UFO, and the Helmholtz Centers have been working on HDRI [6] to mention some. According to the authors of [6] there is a significant amount of duplication of effort between facilities, and many of the efforts are still not taking a "big data" approach to data analysis. They hope that file formats and approaches to software design will be based on some standards, such that it can be easily shared.

Users of synchrotron facilities often need not only reconstruction but also image analysis for carrying out their research. They bring samples to be imaged and their expertise in their own area of science. Providing the appropriate software and expertise for synchrotron imaging analysis, which is the goal of MAX4Imagers, is a highly interdisciplinary field. The image analysists need to learn about the data to be able to interpret them and to find suitable methods for analysing it.

Therefore, it is absolutely necessary that we share knowledge within disciplines and cooperate across disciplines to keep transcending the limits in synchrotron imaging. Where would we be if the physicists Röntgen and Cormack, the mathematician Radon, and the engineer Hounsfield had not stood on each other's shoulders?

Bibliography

- L. Borg, J.S. Jørgensen, J. Frikel, and J. Sporring. Reduction of variable-truncation artifacts from beam occlusion during in situ x-ray tomography. *Measurement Science* and *Technology*, 28(12), 2017.
- [2] L. Borg, J. Frikel, J. S. Jørgensen, and E. T. Quinto. Analyzing reconstruction artifacts from arbitrary incomplete x-ray ct data. Under review at the Society for Industrial and Applied Mathematics (SIAM) Journal on Imaging Sciences.
- [3] J. Sporring L. Borg, V. A. Dahl, A. B. Dahl, E. B. Dam, and J. Pingel. Muscle fiber morphology and orientation consistency in cerebral palsy from 2- and 3-dimensional images obtained by synchrotron x-ray computed tomography. *submitted*.
- [4] L. Borg, J. S. Jørgensen, and J. Sporring. Towards characterizing and reducing artifacts caused by varying projection truncation. Technical Report 2017/1, Department of Computer Science, University of Copenhagen, 2017. Available online: http://static-curis.ku.dk/portal/files/174175124/Borg_2017_Towards_ characterizing_and_reducing.pdf.
- [5] A. Gandomi and M. Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137 144, 2015.
- [6] D. Y. Parkinson, K. Beattie, X. Chen, J. Correa, E. Dart, B. J. Daurer, J. R. Deslippe, A. Hexemer, H. Krishnan, A. A. MacDowell, F. R. N. C. Maia, S. Marchesini, H. A. Padmore, S. J. Patton, T. Perciano, J. A. Sethian, D. Shapiro, R. Stromsness, N. Tamura, B. L. Tierney, C. E. Tull, and D. Ushizima. Real-time data-intensive computing. *AIP Conference Proceedings*, 1741(1):050001, 2016.
- [7] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), 2001.
- [8] J. Pingel, M. S. Nielsen, T. Lauridsen, K. Rix, M. Bech, T. Alkjaer, and R. Feidenhans'l. Injection of high dose botulinum-toxin a leads to impaired skeletal muscle function and damage of the fibrilar and non-fibrilar structures. *Scientific Reports*, 7(14746), 2017.
- [9] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. Int J Comput Vision, 1(321), 1988.
- [10] W. C. Röntgen. Ueber eine neue art von strahlen. Sonderdruck aus den Sitzungsberichten der Wuzburger Physik.-medic. Gesellschaft, 1896.
- [11] J. Radon. Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser manningsfaltigkeiten. Berichte Sachsische Akadamie der Wissenschaften, Leipzig, 69(5):262-277, 1917.
- [12] T. H. Buzug. Computed Tomography: From Photon Statistics to Modern Cone-Beam CT. Springer, Berlin, Germany, 2008.

- [13] A. C. Kak and M. Slaney. Principles of Computerized Tomographic Imaging. IEEE Press, 1988.
- [14] Y. Hwu, W. L. Tsai, A. Groso, G. Margaritondo, and J. H. Je. Coherence-enhanced synchrotron radiology: Simple theory and practical applications. *Journal of Physics D Applied Physics*, 35, 2002.
- [15] Y. Hwu, G. Margaritondo, and A. Chiang. Q&a: Why use synchrotron x-ray tomography for multi-scale connectome mapping? *BMC Biology*, 15:122, 2017.
- [16] G. Margaritondo, Y. Hwu, and G. Tromba. Synchrotron light: From basics to coherence and coherence-related applications. *Synchrotron Radiation: Fundamentals, Methodolo*gies and Applications., pages 25–53, 2003. eds. Mobilio and Vlaic (SIF Bologna, 2003).
- [17] A.S. Tremsin, G.V. Lebedev, O.H.W. Siegmund, J.V. Vallerga, J.S. Hull, J.B. McPhate, C. Jozwiak, Y. Chen, J.H. Guo, Z.X. Shen, and Z. Hussain. High spatial and temporal resolution photon/electron counting detector for synchrotron radiation research. *Nuclear Instruments and Methods in Physics Research.*, 580:853–857, 2007.
- [18] Z. Matej, R. Mokso, K. Larsson, V. Hardion, and D. Spruce. The MAX IV imaging concept. Adv Struct Chem Imag, 2(16), 2016.
- [19] P. Bouguer. Essai d'optique sur la gradation de la lumière. Paris, France: Claude Jombert, 1729.
- [20] J.H. Lambert. Photometria sive de mensura et gradibus luminis, colorum et umbrae. 1760.
- [21] Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. Annalen der Physik, 162(5):78–88, 1852.
- [22] J. S. Jørgensen. Sparse Image Reconstruction in Computed Tomography. PhD thesis, Technical University of Denmark, 2013. http://orbit.dtu.dk/files/57895386/ phd293_Jorgensen_JS.pdf.
- [23] P. C. Hansen and M. Saxild-Hansen. Air tools a matlab package of algebraic iterative reconstruction methods. *Journal of Computational and Applied Mathematics*, 236(8):2167 – 2178, 2012.
- [24] P. C. Hansen. Lecture notes: Algebraic methods for computed tomography. DTU Compute Department of Applied Mathematics and Computer Science Technical University of Denmark.
- [25] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß. Improved total variationbased CT image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6), 2011.
- [26] W. A. Kalender, R. Hebel, and J. Ebersberger. Recuction of CT artifacts caused by metallic implants. *Radiology*, 165:576–577, 1987.
- [27] T.A. Bubba, D. Labate, G. Zanghirati, and S. Bonettini. Shearlet-based regularized reconstruction in region-of-interest computed tomograph. *Math. Model. Nat. Phenom.*, 13(4), 2018.
- [28] Susan Stipp: "We solve challenges for industry as well". KU news: 07 September 2015. https://www.science.ku.dk/english/press/news/2015/susan_stipp_ solves_challenges/.

- [29] J. Frikel and E. T. Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Problems*, 29(12):125007, 2013.
- [30] F. G. Friedlander. *Introduction to the theory of distributions*. Cambridge University Press, Cambridge, second edition, 1998.
- [31] L. Hörmander. The analysis of linear partial differential operators. I. Classics in Mathematics. Springer-Verlag, Berlin, Germany, 2003.
- [32] M. A. Mathewson and R. L. Lieber. Pathophysiology of muscle contractures in cerebral palsy. *Phys. Med. Rehabil. Clin. N. Am.*, 26:57–504, 2015.
- [33] R. Ketcham and T. Ryan. Quantification and visualization of anisotropy in trabecular bone. J. Microsc., 213:158—-71, 2004.
- [34] M. Stampanoni, A. Groso, A. Isenegger, G. Mikuljan, Q. Chen, A. Bertrand, S. Henein, R. Betemps, U. Frommherz, P. Böhler, D. Meister, M. Lange, and R. Abela. Trends in synchrotron-based tomographic imaging: the SLS experience. *Proc.SPIE*, 6318:6318 – 6318 – 14, 2006.
- [35] Kang Li, Xiaodong Wu, Danny Z Chen, and Milan Sonka. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE transactions on pattern* analysis and machine intelligence, 28(1):119–134, 01 2006.
- [36] Y. Kim, T. Brox, W. Feiden, and J. Weickert. Fully automated segmentation and morphometrical analysis of muscle fiber images. *Cytometry*, 71A:8–15, 2007.