

TECHNISCHE UNIVERSITÄT MÜNCHEN
Physik Department
Lehrstuhl für Biomedizinische Physik

Statistical signal processing and reconstruction algorithms for grating-based X-ray imaging and computed tomography

Maximilian Teuffenbach

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Martin Zacharias
Prüfer der Dissertation: 1. Prof. Dr. Franz Pfeiffer
2. Priv.-Doz. Dr. Tobias Lasser

Die Dissertation wurde am 17.02.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 02.11.2020 angenommen

Abstract

Grating-based X-ray interferometry is a novel imaging technique that offers great potential for the visualization of materials and tissues that are not easily depicted using conventional X-ray imaging methods. Tomographic reconstruction based on interferometric data provides not only access to the distribution of an object's attenuation but also to its refraction and ultra-small-angle scattering power. These images provide valuable additional information that could well expand the diagnostic capabilities of a clinical computer tomography (CT) scanner.

One of the main reasons why this technique has not yet been implemented in a modern CT scanner is that the improved functionality comes at the cost of longer measurement times. Existing projection-based processing algorithms require not a single measurement per projection angle but several measurements with precise grating movements in between. A further reason is that these signal estimation algorithms are also very sensitive to changes in the system alignment due to mechanical vibrations or thermal drifts, which are abundant in a clinical high-power CT using a continuously rotating gantry.

Several solutions for these problems have been proposed but all suffer from major drawbacks. In this thesis first two simple improvements to existing signal estimation methods are presented and then a novel direct reconstruction method is introduced. A fast algorithm for reconstructions using this method is developed, the technique is tested at scans using just a single measurement per angular position and further enhancements that make the reconstructions robust to vibrations and drifts are implemented and tested.

The results in this thesis demonstrate that it is possible to successfully reconstruct the attenuation, refraction, and ultra-small-angle scattering of an object using only a single measurement per projection angle on a system influenced by significant vibrations and drifts.

This is a milestone for the future implementation of a grating interferometer onto a continuously rotating clinical CT scanner.

Zusammenfassung

Röntgen-Interferometrie mit Gittern ist eine neue Bildgebungstechnik die großes Potential in der Darstellung von Materialien und Geweben hat, die sich mit der konventionellem Röntgenbildung nicht zufriedenstellend abbilden lassen. Die tomographische Rekonstruktion basierend auf den Daten von Interferometermessungen erlaubt es nicht nur die Verteilung des Absorptionskoeffizienten für Röntgenstrahlen darzustellen, sondern man erhält auch eine genaue Karte des Brechungsindex und der Kleinwinkelstreuung des Materials. Bei diesen zusätzlichen Bildern handelt es sich um wertvolle Informationen, die die diagnostischen Möglichkeiten eines Computertomographen stark verbessern könnten.

Einer der Hauptgründe warum diese Technik noch nicht in moderne CT Scanner eingebaut wurde ist, dass die erhöhte Funktionalität bis jetzt mit einer längeren Messzeit erkaufte werden musste. Konventionelle Algorithmen zur Signalverarbeitung von Gitterinterferometrie-Datensätzen benötigen nämlich nicht nur eine einzelne Messung pro Winkel, sondern mehrere Messungen pro Winkel und zwischen diesen Messungen müssen die Gitter noch zusätzlich präzise positioniert werden. Ein weiterer Grund ist, dass diese Algorithmen sehr anfällig für kleinste Änderungen der Gitterpositionen durch Vibrationen und Erwärmung des Systems sind. Bei einem leistungsstarken CT Scanner mit einer kontinuierlich rotierenden Gantry kommt es zwangsläufig zu starken Vibrationen und thermischen Drifts.

In der Literatur wurden bereits mehrere Lösungen für diese Probleme vorgeschlagen aber alle diese Lösungen haben spezifische Nachteile. In dieser Arbeit werden zunächst zwei einfache Verbesserungen für bereits existierende Signalverarbeitungsalgorithmen entwickelt und dann wird eine neue direkte Rekonstruktionstechnik vorgestellt. Im Folgenden wird ein schneller Algorithmus zur Verwendung dieser neuen Rekonstruktionstechnik entwickelt, die Technik wird an Messungen mit nur einer Messung pro Winkelschritt getestet und es werden Erweiterungen der Technik implementiert und getestet, die die Rekonstruktion robust gegen Vibrationen und Drifts machen.

Die Ergebnisse dieser Arbeit zeigen, dass es möglich ist die Absorption, Brechung und die Kleinwinkelstreuung eines Objekts mit nur einer einzelnen Messung pro Winkelschritt, auch auf einem System, das von Vibrationen beeinträchtigt wird, zu rekonstruieren.

Diese Erkenntnisse stellen einen bedeutenden Schritt für die zukünftige Verwendung eines Gitterinterferometers auf einem klinischen CT mit kontinuierlich rotierender Gantry dar.

Contents

Abstract	i
Zusammenfassung	iii
1 Introduction	1
2 Background	4
2.1 Principles of X-ray imaging	4
2.1.1 Complex index of refraction	4
2.2 Grating-based X-ray imaging	7
2.2.1 Talbot effect	7
2.2.2 Grating interferometer	8
2.2.3 Moiré pattern	11
2.2.4 Signal formation	12
2.3 CT reconstruction	16
2.3.1 Analytic reconstruction	18
2.3.2 Filtered backprojection	19

2.3.3	Sampling	20
2.3.4	Statistical iterative reconstruction	23
2.3.5	Reconstruction for dark-field	31
2.3.6	Reconstruction for differential phase contrast	32
2.4	Optimization transfer principle	33
2.4.1	General principle	34
2.4.2	Quadratic surrogates	35
2.4.3	Separable quadratic surrogates	36
2.4.4	SQS for CT reconstruction	39
3	Advanced signal estimation for grating-based imaging	43
3.1	Maximum likelihood estimator for GBI signal estimation	44
3.1.1	Bias of the maximum-likelihood estimator	46
3.2	Bias-corrected maximum likelihood estimator	47
3.2.1	Regularized BCMLE	48
3.3	Results	49
3.3.1	Dark-field estimator	49
3.3.2	DPC estimator	53
3.4	Discussion	54
4	Improved sliding window interpolation	56
4.1	Introduction	56

4.2	Robust signal estimation	57
4.3	Results	64
4.4	Discussion	66
5	Statistical iterative reconstruction for grating-based CT	68
5.1	Introduction	68
5.2	Formulation	69
5.2.1	Data term	70
5.2.2	Regularizer	71
5.3	Algorithms	72
5.3.1	Separable paraboloid surrogate for IBSIR	73
5.3.2	Choice of curvature	81
5.3.3	Ordered subsets	86
5.3.4	Fourier space preconditioning	87
5.3.5	Momentum acceleration	88
5.3.6	Simulations	89
5.4	Single shot acquisition	93
5.4.1	Sampling	95
5.4.2	Single-shot simulations	97
5.4.3	Single-shot measurements	101
5.5	Internal parameter estimation during reconstruction	102
5.5.1	Extension of the forward model	103

5.5.2	Grating position estimation	104
5.5.3	Higher order phase drift estimation	111
5.5.4	Flux and visibility drifts	116
5.6	Discussion	120
6	Conclusion and outlook	122
A	Signal estimation	125
A.1	Linearized signal estimation	125
A.2	Bias estimation of the linearized signal estimation	127
A.2.1	Derivatives	127
A.3	Bias corrected maximum likelihood estimator	128
B	Intensity-based statistical iterative reconstruction	134
B.1	Curvatures	134
B.1.1	Hessian matrix	134
B.1.2	Precomputed curvature	135
	Bibliography	136
	List of publications and scientific presentations	147
	Acknowledgements	149

List of Tables

4.1	Error metrics for sinograms estimated using l_2 - and l_1 -optimization	66
5.1	Error metrics for reconstructions of different single-shot scans.	100

List of Figures

2.1	Angle of refraction between two X-ray beams after transversing a medium. . . .	5
2.2	Sketch of a Talbot interferometer.	8
2.3	The stepping curve of an empty scan.	9
2.4	Sketch of a Talbot-Lau interferometer	10
2.5	The overlay of periodic structures generates Moiré patterns.	11
2.6	Moiré fringe pattern as observed at a grating interferometer.	12
2.7	Sketch of an interferometer with objects in the beam path.	13
2.8	Effect of an object on the stepping curve.	14
2.9	CT sampling in Radon space	21
2.10	Estimation of required sampling density.	23
2.11	Comparison of regularization methods in the estimation of a step function. . . .	29
3.1	Bias and variance in the estimation of the dark-field signal for different numbers of photons.	50
3.2	Bias and variance in the estimation of the dark-field signal for different signal strengths.	51

3.3	Bias and variance in the estimation of the differential-phase signal for different numbers of photons.	53
4.1	Signal estimation at the edge of an attenuating object using a sliding window scan and l_2 -optimization.	58
4.2	Signal estimation at the edge of a scattering object using a sliding window scan and l_2 -optimization.	59
4.3	Signal estimation at the edge of a refracting object using a sliding window scan and l_2 -optimization.	60
4.4	Signal estimation at the edge of an object using a sliding window scan and l_1 -optimization.	63
4.5	Sinograms of the estimated signals from a sliding window scan using l_2 - and l_1 -optimization.	66
5.1	Visualization of attenuation and dark-field dimension of the per pixel cost function.	81
5.2	Visualization of the phase dimension of the per pixel cost function.	82
5.3	Visualization of a lookup table for an optimal curvature.	85
5.4	Intermediate results of the reconstruction of δ using IBSIR.	88
5.5	Error metrics for different choices of curvatures for reconstructions using IBSIR.	90
5.6	Error metrics for momentum acceleration and Fourier preconditioning for reconstructions using IBSIR.	91
5.7	Error metrics for subset acceleration for reconstructions using IBSIR.	93
5.8	Sampling of expanded Radon space for different acquisition protocols.	95
5.9	Reconstructing from a scan without phase stepping.	97

5.10	Comparison of different single-shot acquisition protocols and reconstructions. . .	98
5.11	Reconstructions from a single-shot scan of a mouse.	101
5.12	Reconstructions of a simulation with and without grating position estimation. .	108
5.13	Reconstructions of a real scan with and without grating position estimation. . .	110
5.14	Error metrics for reconstructions with and without phase drift estimation. . . .	114
5.15	Reconstructions with and without phase shift estimation and post correction. . .	116
5.16	Error metrics for reconstructions with and without parameter drift estimation. .	119

Chapter 1

Introduction

Wilhelm Conrad Röntgen was one of the first persons in the world that experienced the imaging power of X-rays when he discovered that radiation emitting from a discharge tube was able to penetrate through a seemingly opaque matter and bring a fluorescence screen to glow. Shortly after this observation he used a fine-grained photographic plate to take the first radiography of a human body part, the hand of his wife Anna Bertha [1]. From this moment on there was a strong interest in research on all applications of X-ray radiation and imaging methods utilizing the penetration power of X-rays. These methods quickly became, and still are today, the most important tools in medical diagnostic imaging. X-rays not only proved their merit in medical applications, but they also became one of the most important tools in industrial non-destructive testing.

A shortcoming of X-ray radiography is its two-dimensional nature. An X-ray radiograph collapses the three-dimensional structure of an object onto a two-dimensional image, which makes it impossible to guess the distance between multiple parts behind each other and can cause different objects to show identical radiographs. A more suited depiction of the structure of a three-dimensional object would be a three-dimensional image. This task has been solved by the invention of computed tomography (CT) by Godfrey N. Hounsfield and Allan M. Cormack in 1973 [2]. Using a series of two-dimensional projections images from different directions they were able to reconstruct the three-dimensional distribution of the X-ray attenuation of

the scanned object. CT images have proven to be far more powerful for the diagnostics of various medical conditions than X-ray radiographs and have found extremely wide usage in an uncountable number of medical diagnosis tasks.

For decades the reconstruction of CT images was performed by using algorithms based on a direct analytic solution to a simplified form of the CT reconstruction problem. A major change in the development of CT reconstruction algorithms has been a move away from direct, one-step reconstruction algorithms towards iterative reconstruction algorithms. These algorithms are more computationally expensive but allow to accurately model the real physics behind the X-ray imaging system and to correctly consider the statistical nature of the measurements by employing models of the noise specific to the imaging process. These iterative reconstruction methods allowed for a significant reduction of the necessary X-ray dose applied to patients for a specific exam [3].

A much more fundamental issue inherent to all widely used X-ray imaging techniques is that the contrast of X-ray imaging is solely based on the difference in attenuation of X-ray photons by different materials: bones or metals strongly attenuate X-rays, whereas soft tissue like muscle or organs are more transparent to X-rays. This makes a differentiation between different soft tissues using X-rays a challenging task. Especially X-ray microscopy of biological samples, like autopsies, is severely hindered by this weak contrast. A solution to this problem has been proposed by using not only the attenuation of X-ray photons for imaging, but also utilizing other optical phenomena, like refraction and scattering. Several techniques providing additional contrasts based on these phenomena have been demonstrated successfully under laboratory conditions of varying rigidity [4], i.e. some requiring synchrotron X-ray sources. A promising technique that can exploit these effects for imaging and that has been demonstrated in various laboratory experiments is called grating-based X-ray imaging (GBI). Apart from the conventional attenuation contrast, it gives access to differential phase-contrast, which depends on differences in the refractive index of the specimen, and to the so-called dark-field contrast, which is based on small-angle scattering in the specimen.

Various studies have demonstrated phase-contrast and dark-field X-ray radiography and computer tomography (GBCT) using grating-based X-ray imaging setups but few of these works

have used iterative reconstruction during the CT reconstruction [5, 6, 7, 8, 9, 10, 11, 12].

The scope of this work is to develop novel algorithms for GBI signal extraction and GBCT reconstruction and verify their utility compared to the use of existing signal extraction and reconstruction algorithms.

In Chapter 2 an introduction to the necessary theories needed to understand the physical working principles and the image formation process of GBI will be given. Then known analytic and tomographic reconstruction algorithms for attenuation, differential phase and dark-field CT will be explained and a general method for the development of novel reconstruction algorithms will be presented.

Chapter 3 will review existing techniques for signal estimation from GBI measurements and propose new methods with reduced estimation bias.

Problematic use scenarios of the sliding window interpolation technique, and possible ways to ease the problems with them, will be discussed in Chapter 4.

Chapter 5 is the largest chapter of this thesis and introduces a novel iterative model-based reconstruction method for GBCT. The derivation of efficient algorithms to calculate the solution of this method and the application of the algorithm for the reconstruction of previously hard reconstruction scenarios will be shown. This chapter is concluded with a further expansion of the reconstruction model that allows to reduce the effects of vibrations and various system drifts on the reconstruction.

The thesis is concluded by a final chapter that contains a general discussion of the results and gives an outlook into further areas that would benefit from additional future work.

Chapter 2

Background

In this chapter a brief summary of some of the key concepts necessary to understand the work presented in this thesis will be given. First the principles of X-ray imaging are reviewed, then the fundamentals of GBI are introduced and the formation of the measured signal using a grating interferometer is explained. The third subsection introduces the most widely used analytic CT reconstruction algorithm and explains the working principles of model-based statistical iterative reconstruction. The chapter is concluded by the introduction of a general optimization framework based on optimization of surrogate functions that can be used for the derivation of CT reconstruction algorithms.

2.1 Principles of X-ray imaging

2.1.1 Complex index of refraction

The simplest way to describe X-ray wave propagation in vacuum and matter is by using the projection approximation. This means an incident plane X-ray wave is following a ray path and that a scatterer, if present, is so weak that he does not perturb the ray path. This way any amplitude and phase shifts the X-ray wave experiences while traversing from one point to another can be simply treated as a sum of these shifts along the wave path. The material

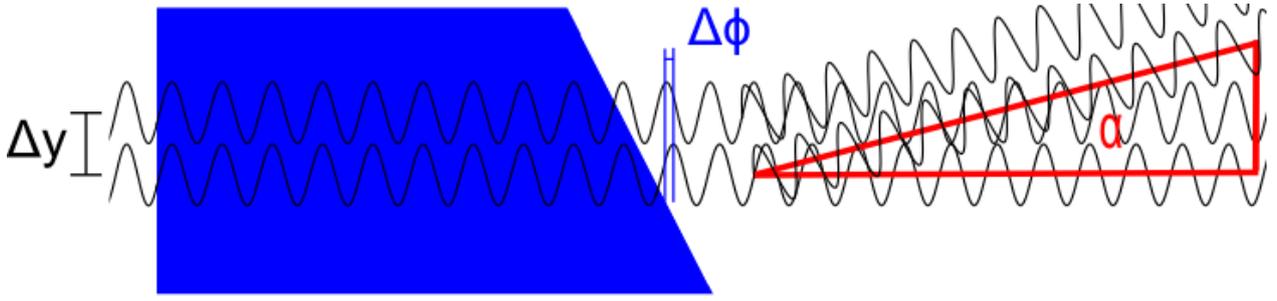


Figure 2.1: Angle of refraction α between two parallel X-ray beams separated by distance Δy after transverseing a wedge-shaped medium with real index of refraction greater than unity.

property that contains information about the change in amplitude and phase a material causes, while a wave transversees it, is called the complex refractive index n . The complex refractive index is often written as

$$n = 1 - \delta + i\beta , \quad (2.1)$$

where δ is called the refractive index decrement and β is the imaginary part of the refractive index. In vacuum the refractive index is unity, resulting in a δ and β equal to zero, but in matter δ and β are different from zero. Usually δ is negative and β is positive but in the X-ray regime δ is positive for energies well away from an objects electron binding energies (e.g. for polymethylmethacrylate (PMMA): $\delta_{PMMA}(30keV) = 2.9 \cdot 10^{-7}$ and $\beta_{PMMA} = 1.18 \cdot 10^{-10}$). Using the refractive index a plane wave ψ_0 with wave number $k = \frac{2\pi}{\lambda}$, where λ is the wave length of the X-rays, traveling in direction $\frac{\mathbf{k}}{k}$ through a homogeneous medium with refractive index n from a starting point $\mathbf{r}_0 = 0$ to a point \mathbf{r} can be described using the following equation:

$$\psi(\mathbf{r}) = \psi_0 e^{i\mathbf{k}\cdot\mathbf{r}} = \psi_0 e^{i(1-\delta)\mathbf{k}\cdot\mathbf{r}} e^{-\beta\mathbf{k}\cdot\mathbf{r}} . \quad (2.2)$$

It is obvious from this equation that passing through this medium has two separate effects on the wave. The first exponential describes a phase shift of the incident wave ψ_0 depending on the real part of the refractive index, whereas the second exponential describes a reduction of the incoming waves amplitude, depending on the imaginary part of the refractive index.

The change of amplitude can be characterized by the ratio of the incident waves intensity and

the exit waves intensity, which is called the transmission T

$$\begin{aligned}
 T &= \frac{I_{out}}{I_{in}} \\
 &= \frac{|\psi_{out}|^2}{|\psi_{in}|^2} \\
 &= e^{-2\beta \mathbf{k} \cdot \mathbf{r}} \\
 &= e^{-2k\beta d} .
 \end{aligned} \tag{2.3}$$

In the last line we rewrote the product of $\frac{\mathbf{k}}{k}$ and \mathbf{r} with the distance d the X-ray traveled through the medium. If we rewrite the last line using the linear attenuation coefficient $\mu = 2k\beta$ and replace the distance d with the integral along the beam path l and allow for a spatially varying index of refraction, we arrive at the famous Lambert-Beer law governing X-ray transmission through an attenuating medium:

$$T = e^{-\int \mu(l) dl} . \tag{2.4}$$

In an analogue fashion, the change of the X-ray waves phase, while traversing along the beam path l , is given by the following equation

$$\Delta\Phi = \int (1 - \delta(l)) k dl . \tag{2.5}$$

Of further interest, especially in the context of grating-based X-ray imaging, is the comparison of the phase change between two X-rays separated only by a small distance Δy . As one can see in Fig. 2.1, two neighboring X-ray waves are not in phase anymore, compared in the original direction, after passing through a medium. By applying Huygens principle one can treat the exit points of both waves as new point sources of X-rays, and in the far-field these new X-rays will again form a plane wave (meaning neighboring rays are in-phase) propagating in a new direction, which has an angle α to the old propagation direction. For small refractive angles α

and looking at infinitesimal small distances Δy , α is given as

$$\begin{aligned}
 \tan \alpha &= \frac{\Delta\Phi \frac{\lambda}{2\pi}}{\Delta y} \\
 \approx \alpha &= \frac{\Delta\Phi \frac{\lambda}{2\pi}}{\Delta y} \\
 \alpha &= \frac{\Delta\Phi}{\Delta y} \frac{\lambda}{2\pi} = \frac{\Delta\Phi}{\Delta y} \frac{1}{k} \\
 \alpha &= \frac{\partial\Phi}{\partial y} \frac{1}{k} = \frac{\partial}{\partial y} \frac{1}{k} \int \delta(x, y) k \, dx \\
 \alpha &= \frac{\partial}{\partial y} \int \delta(x, y) \, dx .
 \end{aligned} \tag{2.6}$$

2.2 Grating-based X-ray imaging

Grating-based X-ray imaging (GBI) is one of the most promising phase-sensitive X-ray imaging techniques [13] because it can be used with conventional X-ray tubes and achieves a high sensitivity of the signal to phase effects. The main working principle is based on the Talbot effect, which is a near-field refraction effect that produces a reference intensity pattern. This reference pattern is locally sensitive to attenuation, phase-shifts, and small-angle scattering from objects in the beam path. In the following the Talbot effect will be briefly introduced and the functional principle of a grating interferometer as well as the signal formation in GBI will be discussed.

2.2.1 Talbot effect

Henry Fox Talbot discovered [14], that a periodic structure that is illuminated by a coherent light wave produces a self-similar intensity pattern at certain propagation distances after the periodic structure. These distances are called the Talbot distances z_T and are given as

$$z_T = \frac{\lambda}{1 - \sqrt{1 - \frac{\lambda^2}{a^2}}} \underset{\text{for } \lambda \ll a}{\approx} \frac{2a^2}{\lambda} \tag{2.7}$$

for an absorbing periodic structure with grating period a illuminated by light of wavelength λ [15]. Interestingly, the Talbot effect is not only found for attenuating gratings but also for

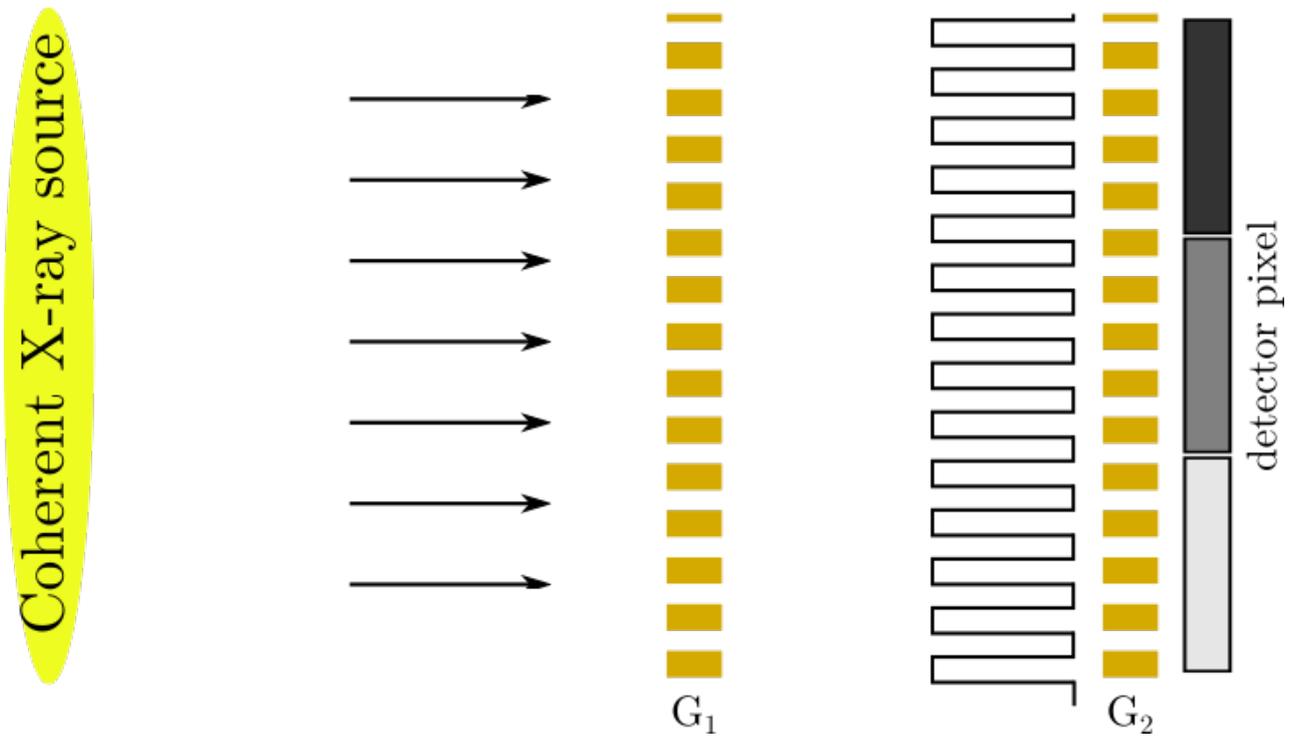


Figure 2.2: The X-ray intensity reaching the detector pixels depends on the position of the analyzer grating G_2 and on the alignment of the interference pattern, which depends on the position of the interference grating G_1 .

pure phase gratings. Here the appearance of the self-images does not simply follow the formula given above, but the self-images appear at so-called fractional Talbot distances, which are, as the name implies, specific fractions of the original Talbot distances. A thorough study of possible fractional Talbot distances can be found in [16].

2.2.2 Grating interferometer

The simplest X-ray grating interferometer would consist of just a single attenuating or phase-shifting periodic grating illuminated by a coherent X-ray beam. As explained by the Talbot effect, there would be self-images of the grating at the Talbot distances down the beam path. But, if one looks at Eq. 2.7 and plugs in realistic numbers for an X-ray imaging setup powered by a conventional X-ray tube, one sees that a single grating interferometer would hardly work: To maintain a high enough flux from a conventional X-ray source and to be able to fit the setup in conventional laboratory spaces, it is unrealistic to work with Talbot distances larger than two meters. Combined with a necessary X-ray energy for medical purposes of at very

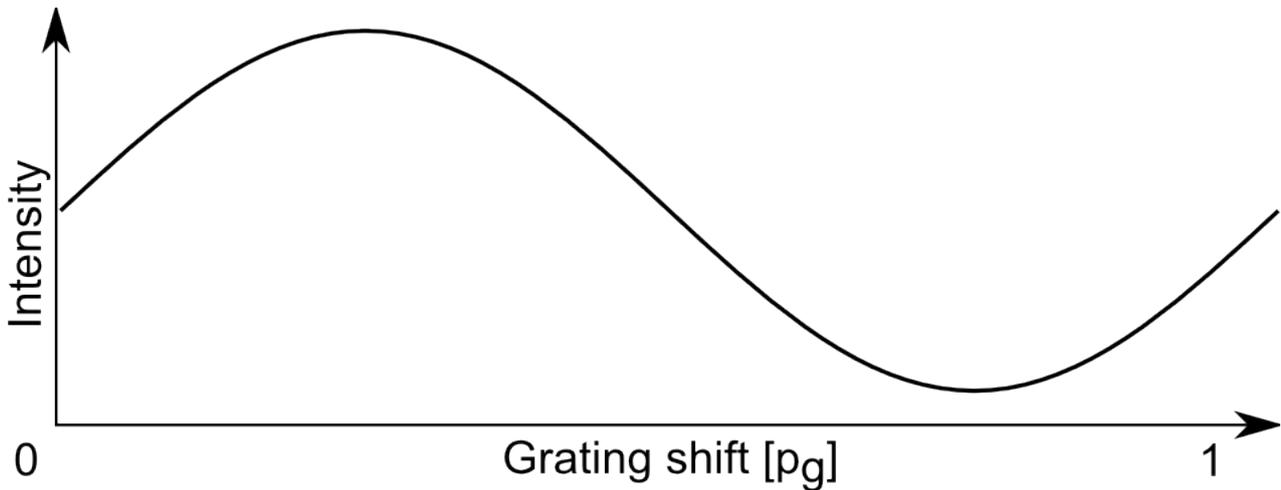


Figure 2.3: The intensity recorded in a single detector pixel changes periodically, with a periodicity of the grating pitch p_g , as function of the analyzer grating position.

least 40 keV, the necessary grating pitch and thus the resulting interference pattern pitch at the first Talbot distance is around $5.5\ \mu\text{m}$. To be able to measure the exact location of the interference pattern it would be necessary to sample the pattern with a resolution at least three times smaller than the interference pattern pitch. As a conventional medical X-ray detector has a pixel size of several hundreds of micrometers directly resolving the interference pattern is not possible with a conventional medical X-ray detector. Similar setups have been demonstrated but require highly specialized laboratory equipment [17, 18, 19].

To solve the problem of a too low detector resolution, a second grating with a grating pitch close to the interference pattern pitch is introduced into the beam path. This grating is called the analyzer grating and, if it is placed parallel to the intensity pattern produced by the interference grating, it will block different parts of the interference pattern, see Fig. 2.2. If the grating is shifted perpendicular to the grating bars, the intensity reaching the detector will change periodically, see Fig. 2.3. Measuring the shape of this so-called stepping curve for each pixel is the aim of most GBI experiments, as it contains averaged information about the interference patterns in front of each detector pixel.

A second restriction has to be considered when designing an X-ray grating interferometer: To be able to witness interference effects sufficient beam coherence is necessary. But, as the grating interferometer is only relying on interference effects in one spatial direction, it is enough to have sufficient transverse coherence in this direction. One way to quantify this necessity is to

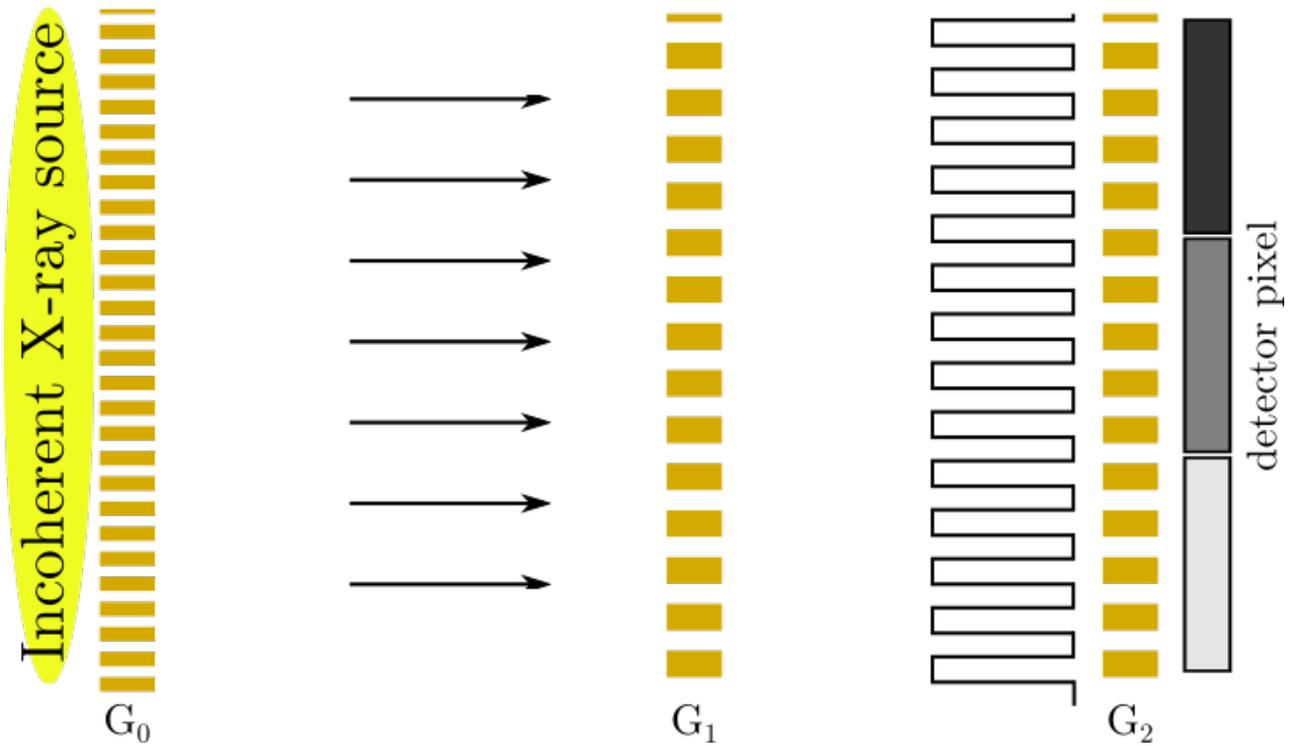


Figure 2.4: The incoherent X-ray source is turned into a series of small slit sources by the source grating G_0 . These slit sources show sufficient coherence to produce an interference pattern after the interference grating G_1 .

consider the X-ray source to be made up of numerous small slit sources [20]. Each of these slit sources will produce an interference pattern at a distance d from the interference grating and L from the slit source. If the next slit is placed at a distance ϵ to the first one it will also cause the same interference pattern at the distance d but shifted by a distance $\epsilon \frac{L}{L-d}$. If this distance is exactly half the pitch of the interference pattern, the signals will cancel each other out. From this rough estimation we can infer that the distance between neighboring slit sources should be much smaller than $\frac{g_2}{2}$ to avoid a loss of intensity of the interference pattern. One way to construct such a multi-slit source is to place an absorption grating G_0 with a period p_0 satisfying

$$p_0 \ll p_2 \frac{1}{2} \frac{L}{L-d}, \quad (2.8)$$

in front of the conventional X-ray source.

A schematic of such a three-grating Talbot-Lau interferometer is given in Fig. 2.4 and was first proposed in [21].

2.2.3 Moiré pattern

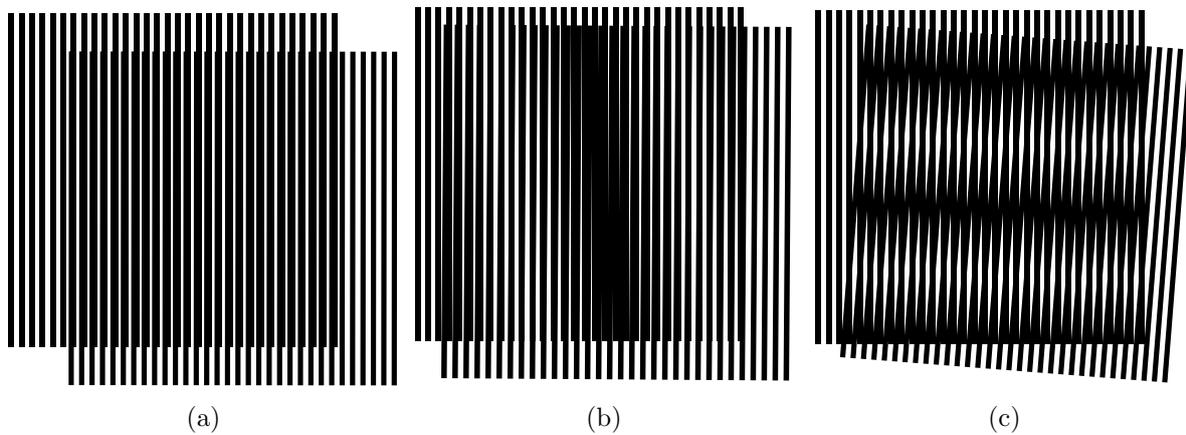


Figure 2.5: An overlay of two periodic structures with matching pitches (a) and two Moiré pattern caused by slight mismatch of the pitches (b) and rotation of the patterns against each other (c).

In an ideal world the pitch and positioning of the interference pattern and the analyzer grating could perfectly match, resulting in identical stepping curves for all pixels of an extended detector. But in real measurements, the periods will always be slightly mismatched, and if two fine pitched periodic structures with similar pitches are placed on top of each other a larger pitched periodic structure can be observed. This is called the Moiré effect and is illustrated in Figure 2.5 for three different alignments of a periodic illumination with a periodic absorption grating, each resulting in a different pattern of transmitted intensity. If the periods of the patterns match perfectly and are aligned in the same direction no pattern is visible, see Fig. 2.5A. But if the pitch of the patterns is slightly mismatched, see Fig. 2.5B, large periodic structures in the direction of the periodic structures appear. The periodicity of these larger structures becomes faster and faster with increasing mismatch between the periods of the fine pitched structures. In grating interferometry this effect is observable if the periodicity of the Talbot carpet (which is given by the periodicity of the interference grating and a possible optical magnification) does not exactly match the periodicity of the analyzer grating and can be easily created by moving the analyzer grating slightly out of the Talbot distance. Large periodic structures in the direction perpendicular to the periodic structures are created if the fine pitched structures are slightly rotated against each other (Fig. 2.5C), which causes a periodic

matching and mismatching of the structures. This effect is also readily observed at a grating interferometer if the gratings are slightly rotated with respect to each other. If these patterns are measured with a resolution much larger than the fine structures, only the larger pitched structures will be visible in the resulting image. In addition to the regular patterns with a constant pitch often patterns with a spatially varying periodicity can be observed, see Fig. 2.6. This can be caused by a multitude of effects: Slight bending of the gratings in the direction of the beam, slight twisting of the grating structures due to fixation on a grating holder or simply due to manufacturing errors resulting in a not perfect periodicity of the gratings.

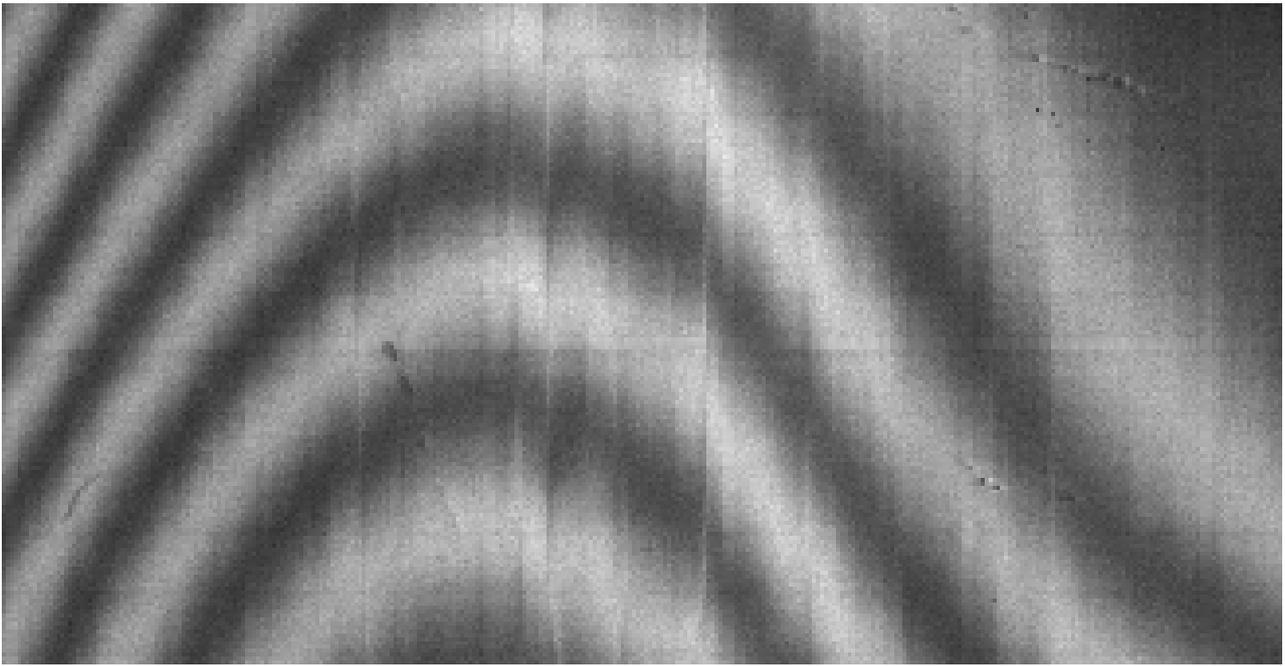


Figure 2.6: A typical Moiré fringe pattern as observed at a grating interferometer.

2.2.4 Signal formation

Introducing a sample into the beam path changes the intensity, shape, and location of the interference pattern depending on the structure of the sample, see Fig. 2.7. As mentioned in the last section it is usually not possible to directly resolve the interference pattern on a conventional detector, but the analyzer grating G_2 is placed in front of the detector to raster over the interference pattern. Near the Talbot distances, shifting the analyzer grating by one period causes the readout of the detector pixels to change in a nearly sinusoidal pattern [20].

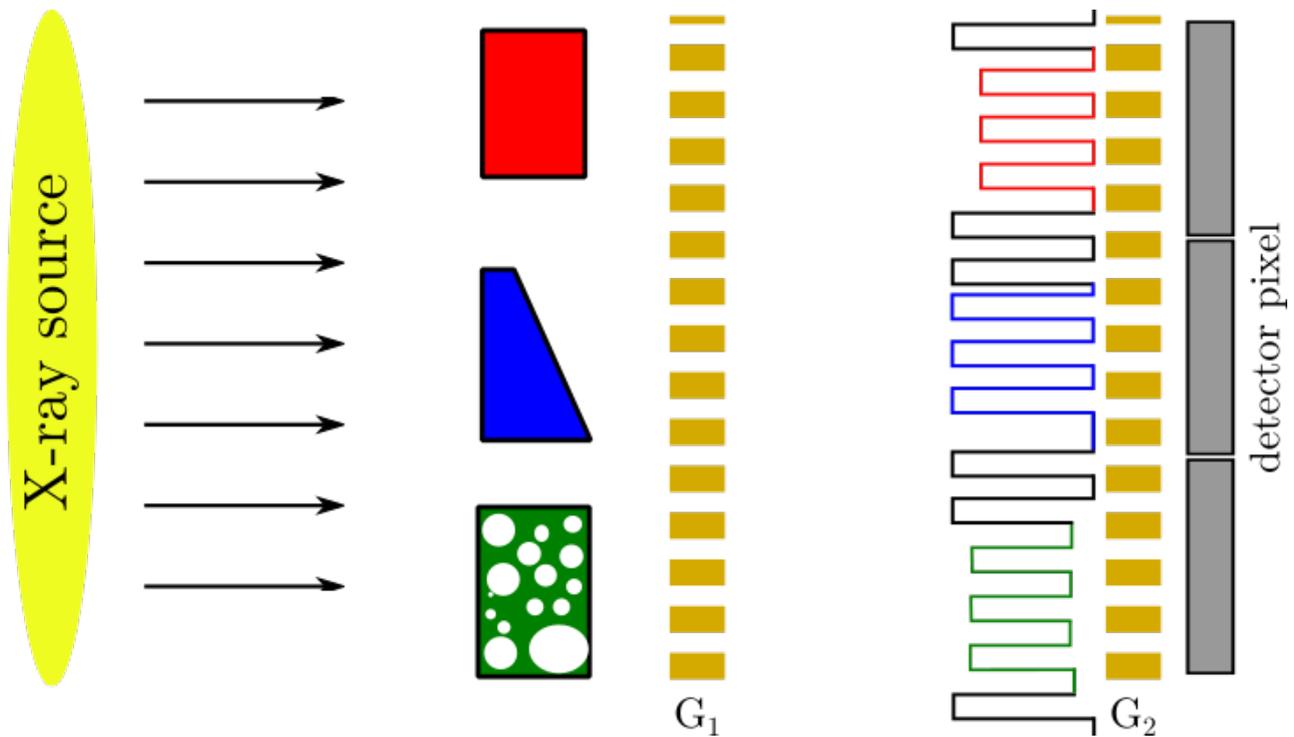


Figure 2.7: The shape of the interference pattern is influenced by the object in the beam path. The red object is a purely attenuating object, which causes a decrease in the mean value of the interference pattern. The blue object is a purely phase-shifting object, which causes a shift of the interference pattern and the green object is a purely scattering object, which causes a decrease in the relative amplitude of the interference pattern.

Alternatively, shifting the source grating or interference grating causes the same effect (in this case the interference pattern is moved over the stationary analyzer grating instead of vice-versa). This process of shifting the analyzer grating multiple times and performing subsequent measurements of points on the sinusoidal curve is called phase-stepping because each step causes a change of the phase of the sinusoidal curve. The thus acquired phase-stepping curve contains information about three mean quantities of the interference pattern in front of each pixel: The mean intensity a_0 , the mean amplitude a_1 and the mean phase ϕ_1 (at a specific grating position). These quantities can be compared between a scan with a sample in the beam path (denoted by superscript s) and a reference scan without a sample in the beam path (denoted by superscript r) to gain information about the optical properties of the sample, see Fig. 2.8.

The relative change in the mean intensity $\frac{a_0^s}{a_0^r}$ directly relates to the X-ray transmission T through the sample, which is governed by Lambert-Beers law, and connects directly to the line integral

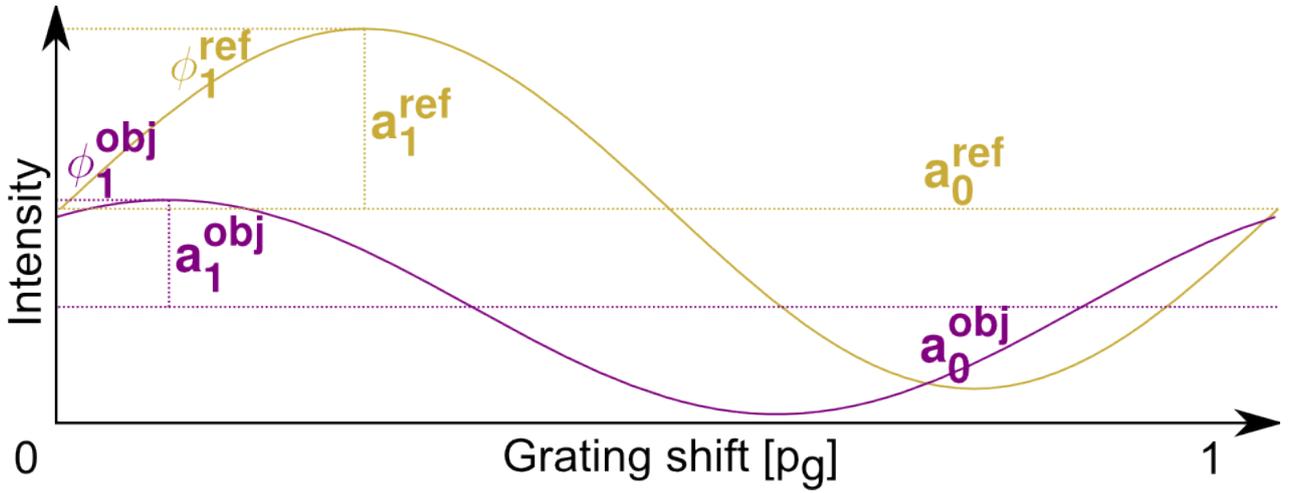


Figure 2.8: The shape of the stepping curve is changed in three different ways by an object present in the beam path: The mean intensity a_0 is reduced, the amplitude a_1 is reduced, and the phase ϕ_1 of the curve is changed relative to a reference scan without an object in the beam path.

over the linear attenuation coefficient μ of the sample:

$$T = \frac{a_0}{a_0^r} = \exp \left[- \int_l \mu(l) dl \right] . \quad (2.9)$$

The change of the stepping curves phase,

$$\Delta\phi_1 = \phi_1^s - \phi_1^r , \quad (2.10)$$

is directly proportional to the mean lateral shift of the interference pattern

$$S = \frac{\Delta\phi_1}{2\pi} g_2 , \quad (2.11)$$

which in turn can be expressed as the angle between the propagation direction of the X-ray after the sample to the X-ray propagation direction before the sample:

$$\alpha \approx \tan \alpha = \frac{S}{d} = \frac{\Delta\phi_1}{2\pi} \frac{g_2}{d} , \quad (2.12)$$

where d is the distance between sample and detector. This refraction angle α was already introduced in Eq. 2.6 and directly connects the phase shift of the phase stepping curves $\Delta\phi_1$

with the real part of the refractive index and the differential phase shift of the wave fronts phase Φ in the direction y , which is perpendicular to the grating bars:

$$\frac{\partial}{\partial y} \int \delta \, dx = \frac{\Delta\phi_1}{2\pi} \frac{g_2}{d} = \frac{\partial\Phi}{\partial y} . \quad (2.13)$$

The relative change of the amplitude $\frac{a_s}{a_1}$ of the phase stepping curve is called the dark-field contrast. The main cause for the change of the amplitude of the stepping curve is a change in the mean amplitude of the interference pattern caused by small-angle X-ray scattering from randomly distributed unresolved structures in the sample [22]. But, also locally unresolved edges or structures that cause a non-constant, but possibly zero-sum, shift of the interference pattern can result in a local extinction of the interference pattern, which results in a lowered mean amplitude of the interference pattern [23].

For the scope of this work, the origin of the dark-field contrast will be modeled by a single simple material coefficient, analogous to the attenuation coefficient, the linear diffusion coefficient ϵ [20]. A single sharp X-ray beam traversing a thin slice of a coherently scattering material will not be sharp anymore after exiting the material but will be transformed into an angular distribution of intensity $A_1(\theta)$. If this distribution is assumed to be Gaussian shaped, one can write

$$A_1(\theta) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{\theta^2}{2\sigma_1^2}} , \quad (2.14)$$

where σ_1 is the width of the scattering distribution.

If another slice of a scattering material is added directly behind the first slice, the scattering distribution $A_1(\theta)$ will be further broadened, which is mathematically expressed as a convolution of $A_1(\theta)$ with the scattering distribution $A_2(\theta)$ caused by the second slice. More general one can write for N slices of materials behind each other

$$A(\theta) = A_1(\theta) * A_2(\theta) * \dots * A_N(\theta) . \quad (2.15)$$

If all these distributions are Gaussian shaped, the resulting angular beam profile $A(\theta)$ again

takes Gaussian shape

$$A(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\theta^2}{2\sigma^2}} , \quad (2.16)$$

with

$$\sigma^2 = \sum_i^N \sigma_i^2 \quad (2.17)$$

for discrete scattering materials or

$$\sigma^2 = \int \frac{\sigma^2(x)}{\Delta x} dx = \int \epsilon dx \quad (2.18)$$

for a series of infinitesimal small scattering structures with extent Δx along the beam path in x -direction. On a detector in distance d behind the scattering object, this will smear out an interference pattern $I^r(x)$ as follows

$$I^s(y) = I^r(y) * A(y) , \quad (2.19)$$

with the scattering distribution $A(\theta)$ propagated onto the detector (assuming small θ) as

$$A(y) = \frac{1}{\sigma d\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2 d^2}} . \quad (2.20)$$

For a sinusoidal interference pattern $I^r(y)$ with period $\frac{2\pi}{p}$ this smearing will result in a new sinusoidal pattern with the same mean value and phase but with its amplitude reduced by

$$D = e^{-\frac{2\pi^2 d^2}{p^2} \sigma^2} = e^{-\frac{2\pi^2 d^2}{p^2} \int \epsilon dx} , \quad (2.21)$$

which can be identified with the dark-field contrast.

2.3 CT reconstruction

X-ray radiography was a breakthrough in medical diagnosis when it was discovered but it was still hampered in its efficiency by its lack of depth information. A shortcoming, which

was solved by an invention of the engineers Godfrey N. Hounsfield and Allan M. Cormack in 1973 [2], and for which they were awarded the Nobel prize for physiology or medicine in 1979. They called their invention computerized tomography (CT). This technique consists of two main steps: Signal acquisition and CT image reconstruction. During the acquisition multiple two-dimensional projections of an object are acquired from a number of different angular positions and in the subsequent image reconstruction an algorithm is used to estimate the three dimensional spatial distribution of the attenuation coefficients of the scanned sample. Since its invention, CT has become one of the most important medical examinations worldwide, e.g. 143 CT exams per 1000 inhabitants were performed in Germany in 2015 [24]. This number shows the importance of CT exams for medical diagnosis today, but it does not come without a drawback: Medical applications account for 98% of the public exposure to ionizing radiation from non-natural sources [25], and of these CT examinations are the largest single contributor [25].

These two sides of the same coin have also been the major drivers for research and development in the field of CT imaging: Improving the image quality allows for more precise diagnosis and allows to use CT imaging to diagnose ever more pathologies but at the same time the necessary level of patient irradiation to achieve images with a predefined quality has to be reduced.

Next to hardware improvements that encompassed nearly all components of a CT scanner, a major contributor to increasing CT imaging power are improvements of the reconstruction algorithms.

As this thesis is mainly concerned with developing novel CT reconstruction algorithms for data acquired using GBI, the following section will present the necessary previous works and theories to understand the most recent techniques for reconstruction of conventional attenuation CT. In addition a short overview of previous approaches for the reconstruction of the differential phase-contrast signal for phase-contrast CT (PCCT) and of the dark-field signal for dark-field CT (DFCT) will be given.

2.3.1 Analytic reconstruction

Decades before Hounsfield and Cormack performed and reconstructed the first CT scan, the Austrian mathematician Johann Radon laid the theoretical foundation for an analytic reconstruction algorithm in his work “Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten” [26]. Johann Radon introduces what is known today as Radon transform, proves that it, and its inverse, are uniquely defined, and gives a solution on how to calculate the inverse Radon transform.

The Radon transform describes the transformation of a given two-dimensional function $f(x, y)$ into the space spanned by the line integrals $p(s)$ over $f(x, y)$ along straight lines s . For a single straight line s_i we can write

$$p(s_i) = \int_{s_i} f(x, y) dl . \quad (2.22)$$

If we identify $f(x, y)$ with the spatially varying linear attenuation coefficient $\mu(x, y)$ and the straight lines s with the straight lines from an X-ray source to an X-ray detector and solve Lambert-Beers law, Eq. 2.45, for the line integral over μ we arrive at

$$-\log \frac{I_i}{I_i^0} = \int_{s_i} \mu(x, y) dl . \quad (2.23)$$

This equation shows that, for an ideal noiseless CT scan, the detector pixel measurements (after normalization and log-transform) are points in the Radon transformed space of the scanned object. Consequently, CT reconstruction can now simply be seen as the process of performing an inverse Radon transformation!

To find such an inversion to the Radon transform we will define a coordinate system in which the function f is a function of the coordinates x and y and the lines s can be describes by the coordinates θ and r , where θ is the angle of the normal with respect to the line which crosses the origin and r is the distance of the line to the origin. Using this notation we can define all line integrals under a certain angle θ as follows

$$p_\theta(r) = \int_{s(\theta,r)} f(x, y), dl \quad (2.24)$$

This set of line integrals for all relevant r at a given θ is called a parallel projection. Using a parallel illumination, or by scanning a finite illumination collimated to a pencil beam shape over the sample, allows to measure such a parallel projection. In conventional medical X-ray systems fan- or cone-beam geometries are more common, but to keep this chapter concise it will be limited to the case of a parallel beam as the more general geometries become increasingly complex and do not add much value in the understanding of analytic CT reconstruction. More information on this topic can be found in [27].

2.3.2 Filtered backprojection

The most widely used analytic reconstruction algorithm is called filtered backprojection (FBP) because it relies on two operations, a filter operation applied to all projections p_θ and the backprojection operation. The backprojection operation means that the values of all projections are simply smeared out along their respective lines and added up. Mathematically this backprojection $g(x, y)$ of p can be written as

$$g(x, y) = \int_0^\pi p_\theta(r) d\theta . \quad (2.25)$$

The rigorous derivation of the algorithm is elaborate, see [27], but is quickly summarized in the following.

We can write the function $f(x, y)$ as a function of its two-dimensional Fourier transform $F(u, v)$ and perform a change of the coordinate system of the Fourier space to polar coordinates (ρ, θ)

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{2\pi i(xu+yv)} du dv \quad (2.26)$$

$$f(x, y) = \int_{-\pi}^{\pi} \int_0^{\infty} F(\rho \cos(\theta), \rho \sin(\theta)) \rho e^{2\pi i(x\rho \cos(\theta)+y\rho \sin(\theta))} d\rho d\theta \quad (2.27)$$

$$\text{with } u = \rho \cos(\theta) \text{ and } v = \rho \sin(\theta) . \quad (2.28)$$

Employing the symmetries of the equation above [27], one arrives at the following form

$$f(x, y) = \int_0^\pi \int_{-\infty}^{\infty} F(\rho \cos(\theta), \rho \sin(\theta)) |\rho| e^{2\pi i \rho(x \cos(\theta) + y \sin(\theta))} d\rho d\theta . \quad (2.29)$$

Now one can utilize the Fourier slice theorem [27], which states that the values along one line through the origin of $F(u, v)$ under angle θ is equal to the one dimensional Fourier transform of the parallel projection of $f(x, y)$ along r under the angle θ ,

$$\begin{aligned} F(\rho \cos(\theta), \rho \sin(\theta)) &= P_\theta(\rho) \\ F(\rho \cos(\theta), \rho \sin(\theta)) &= F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i(xu + yv)} dx dy \\ P_\theta(\rho) &= \int_{-\infty}^{\infty} p_\theta(r) e^{-2\pi i r \rho} dt . \end{aligned} \quad (2.30)$$

Inserting Eq. 2.30 into Eq. 2.29 to substitute $F(\rho \cos(\theta), \rho \sin(\theta))$ yields

$$f(x, y) = \int_0^\pi \int_{-\infty}^{\infty} P_\theta(\rho) |\rho| e^{2\pi i \rho(x \cos(\theta) + y \sin(\theta))} d\rho d\theta , \quad (2.31)$$

which can be understood as a filter operation in Fourier space on the projections,

$$\bar{p}_\theta(r) = \int_{-\infty}^{\infty} P_\theta(\rho) |\rho| e^{2\pi i \rho r} d\rho , \quad (2.32)$$

and a subsequent backprojection of the filtered projections,

$$f(x, y) = \int_0^\pi \bar{p}_\theta(x \cos(\theta) + y \sin(\theta)) d\theta . \quad (2.33)$$

2.3.3 Sampling

The previous derivation of the filtered backprojection algorithm assumed the full continuous projection space $p(r, \theta)$ to be known, which, of course, is unrealistic for a real reconstruction problem: During a scan, the Radon space is only sampled at discrete positions. Now a fundamental question is: How many measurements are necessary to achieve a certain image quality

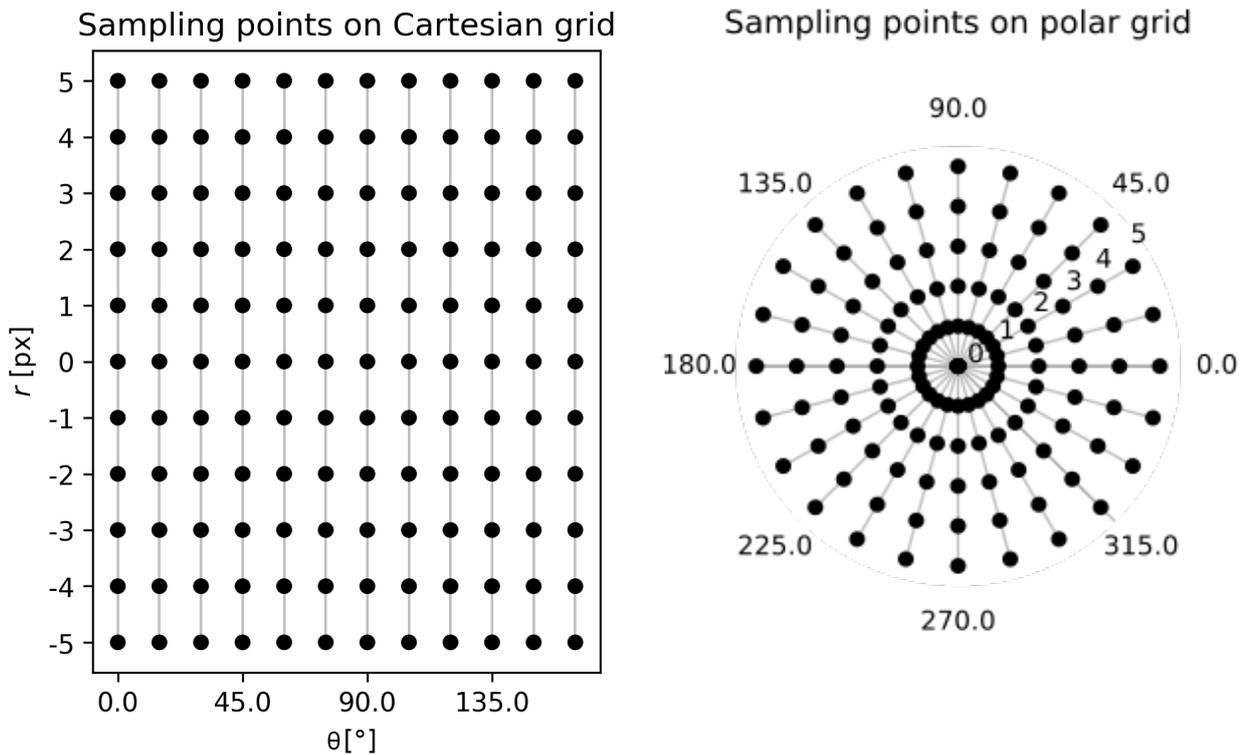


Figure 2.9: Sampling points for a CT scan in parallel beam geometry with equidistantly distributed angular positions over 180 deg in Radon space with coordinates (θ, r) on a Cartesian and a spherical grid.

and how should these be distributed in the sampling space? To answer this question it is helpful to visualize the sampling space. Two different visualization of the Radon space are commonly used: One using Cartesian coordinates (θ, r) , which is called the sinogram and is often used to visualize the projection data acquired during a CT scan, and one using the coordinates (θ, r) in a polar coordinate system, see Fig. 2.9. For parallel beam illumination the measurements fall on regular points inside of a rectangle in the Cartesian grid and on straight lines through the origin in the spherical grid. The Cartesian grid is ideally suited for visualization of measured parallel-beam data as pixels without requiring interpolation. Scans acquired using equiangular and equidistant Fan-beam geometries are situated on curved lines through the origin in the spherical coordinate system, see [28] for further information. For sampling considerations the visualization using the spherical coordinate is quite helpful because, if one recalls the Fourier slice theorem employed in the derivation of the FBP algorithm, see Eq. 2.30. One can realize that the visualization of Radon space sampling is equivalent to visualizing points sampled on the 2D Fourier transform of our object.

The first consideration about sampling can be derived from the Nyquist-Shannon sampling theorem [29], which states that a band-limited function with a maximum frequency of f_{max} can only be exactly recovered if it is sampled with a frequency $f_{sampling}$ of at least two times f_{max} . This is not directly applicable to the ideal sampling situation, as real objects and their projections may have sharp edges and thus their frequency representation is not necessarily band limited, but it is applicable to the real case where we have a pixelated detector with a certain pixel size B and our measurements are actually a convolution of the ideal signal with the rectangular window function of the detector, which is causing a low-pass filtering that limits the resolvable bandwidth of our signal. It can be calculated [28] that the low-pass filter through convolution with a box of size B limits the maximum frequency of the signal to $f_{max} = \frac{1}{B}$. Thus, we should sample the projections at least twice on the length of a detector pixel. An easy way to achieve this is by constructing the CT scanner with a quarter detector offset [28], which means that the central pixel on the detector is offset by one quarter of the detector spacing from the optical axis. This results in a sampling pattern that after rotation of the setup by 180° is again sampling under the same angle but at sampling points shifted by a distance of $\frac{1}{2}B$.

The second consideration [28] about sampling can be derived directly from the Fourier slice theorem: The Fourier transforms of two adjacent projections measured with an angular distance of $\delta = \frac{\pi}{N_{proj}}$ and with N_{samp} sampling points each correspond to the same number of measurements along two lines through the origin of the two-dimensional Fourier transform of the scanned slice of our sample, see Fig. 2.10. As the arrangement of sampling points does not change between projections the highest frequency f_{max} measured in each radial line of the Fourier space is the same for each projection, thus the highest frequencies measured in all projections lie on a circle around the origin with a radius of f_{max} in the Fourier space of the measured object. For small δ the distance between two measurements with frequency f_{max} is $\Delta_{ang} = \delta f_{max}$ and the distance between two measurements on the radial lines is $\Delta_{rad} = \frac{2f_{max}}{N_{samp}}$. If we want to achieve a consistent estimate of both the features with highest angular and radial resolution, we should sample the Fourier space in angular direction at least as well as in radial direction (which is given by the density of sampling points per projection as in the first

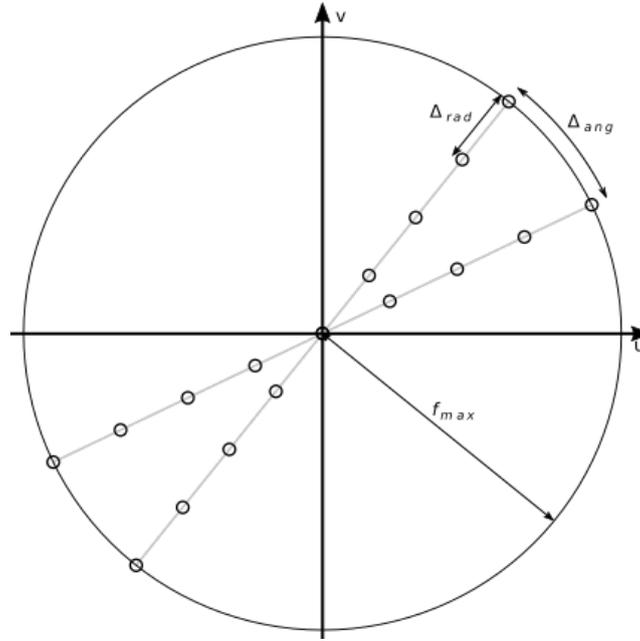


Figure 2.10: Sampling points on the two-dimensional Fourier space of a scanned object slice acquired by one dimensional Fourier transforms of two projections with an angular distance δ and $\frac{2}{f_{max}}$ sampling points per projection (adapted from [28]).

consideration), so the following relation should be fulfilled:

$$\begin{aligned}
 \Delta_{ang} &\lesssim \Delta_{rad} \\
 \delta f_{max} &\lesssim \frac{2f_{max}}{N_{samp}} \\
 \frac{\pi}{N_{proj}} &\lesssim \frac{2}{N_{samp}} \\
 \frac{N_{proj}}{N_{samp}} &\gtrsim \frac{\pi}{2}.
 \end{aligned} \tag{2.34}$$

2.3.4 Statistical iterative reconstruction

A more sophisticated approach to CT reconstruction is to treat the reconstruction as a statistical signal estimation from a series of measurements that are generated from stochastic processes. Following this school of thought allows to take into account the probability distributions of the measurements and results in reconstructions with reduced noise.

The most common statistical estimation approach for CT reconstruction is derived from the

maximum a posteriori probability (MAP) principle based on Bayesian statistics.

2.3.4.1 Maximum a posteriori probability

Assuming a CT scan with measurements \mathbf{y} is the result of a stochastic process which depends on the (non-stochastic) spatial distribution of the attenuation coefficient $\boldsymbol{\mu}$ one can write the probability p of recording a measurement \mathbf{y} given the distribution of $\boldsymbol{\mu}$ as follows: $p(\mathbf{y}|\boldsymbol{\mu})$. This way of writing p should emphasize that p is the probability for measuring the specific realization of a stochastic process \mathbf{y} , given that $\boldsymbol{\mu}$ is the true value of the attenuation coefficients that governs the shape of the probability distributions of our measurements [30].

Using Bayes' theorem [30], one can rewrite this equation:

$$p(\mathbf{y}|\boldsymbol{\mu}) = p(\boldsymbol{\mu}|\mathbf{y}) \frac{p(\mathbf{y})}{p(\boldsymbol{\mu})}. \quad (2.35)$$

The first term on the right side seems much more suitable for the goal of statistical estimation of $\boldsymbol{\mu}$ because it describes the probability that a certain $\boldsymbol{\mu}$ is the true value of $\boldsymbol{\mu}$ given an observed measurement \mathbf{y} . The terms $p(\mathbf{y})$ and $p(\boldsymbol{\mu})$ are called prior probabilities because they are the probability, based on prior knowledge of the world, of witnessing the measurement \mathbf{y} regardless of the object present, and of the measured object having the shape $\boldsymbol{\mu}$, regardless of the performed measurement.

The MAP estimator $\hat{\boldsymbol{\mu}}$ is now simply the value of $\boldsymbol{\mu}$ that maximizes the probability $p(\boldsymbol{\mu}|\mathbf{y})$:

$$\hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} p(\boldsymbol{\mu}|\mathbf{y}). \quad (2.36)$$

Using Eq. 2.35 this becomes

$$\hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} p(\mathbf{y}|\boldsymbol{\mu}) \frac{p(\boldsymbol{\mu})}{p(\mathbf{y})} \quad (2.37)$$

which can be simplified to

$$\hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} p(\mathbf{y}|\boldsymbol{\mu}) p(\boldsymbol{\mu}) \quad (2.38)$$

as $p(\mathbf{y})$ does not depend on $\boldsymbol{\mu}$. Often finding the maximum of the product of two functions is harder than finding the maximum of the sum of two functions. Fortunately, the logarithm of a function still has the same optima as the original function, which one can use to rewrite the problem:

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \arg \max_{\boldsymbol{\mu}} \ln [p(\mathbf{y}|\boldsymbol{\mu})p(\boldsymbol{\mu})] \\ &= \arg \max_{\boldsymbol{\mu}} \ln [p(\mathbf{y}|\boldsymbol{\mu})] + \ln [p(\boldsymbol{\mu})] \\ &= \arg \min_{\boldsymbol{\mu}} L(\mathbf{y}|\boldsymbol{\mu}) + R(\boldsymbol{\mu}) .\end{aligned}\tag{2.39}$$

As most optimization literature is written in the view of minimizing objective functions, a cosmetic change from maximization to minimization was performed in the last equation. To be able to numerically solve this optimization problem one first needs to find suitable expressions for the data fidelity term L , which is the negative logarithm of the likelihood term $p(\mathbf{y}|\boldsymbol{\mu})$, and for the regularizer term R , which is the negative logarithm of the prior probability $p(\boldsymbol{\mu})$. It should be noted that the MAP estimator is identical to the maximum likelihood estimator if the prior knowledge R is constant for all possible $\boldsymbol{\mu}$, which means that no prior knowledge about the object, that allows us to prioritize any specific $\boldsymbol{\mu}$ is known at all.

2.3.4.2 Likelihood terms

The statistical distribution of the measurements is quite complex if analyzed in-depth [27], but can in many cases be approximated very well by independent Poisson or Gaussian distributions [31].

2.3.4.2.1 Poisson distributed noise In the case of Poisson distributed measurements \mathbf{y} , the probability of measuring a certain number of photons y_i at one measurement i given the expected number of Photons of this measurement \bar{y}_i is given by the following probability

$$p_P(y_i|\bar{y}_i) = \frac{\bar{y}_i^{y_i}}{y_i!} e^{-\bar{y}_i} .\tag{2.40}$$

The probability for a specific scan with n independent measurements follows as:

$$p_P(\mathbf{y}|\bar{\mathbf{y}}) = \prod_{i=0}^n \frac{\bar{y}_i^{y_i}}{y_i!} e^{-\bar{y}_i}. \quad (2.41)$$

Taking the negative logarithm of this probability gives us the data fidelity term for Poisson distributed measurements

$$L_P(\mathbf{y}|\bar{\mathbf{y}}) = -\ln p_P(\mathbf{y}|\bar{\mathbf{y}}) = \sum_{i=0}^n \bar{y}_i - y_i \ln \bar{y}_i, \quad (2.42)$$

after dropping constant terms.

2.3.4.2.2 Gaussian distributed noise For Gaussian distributed noise there are two parameters that describe the shape of the probability distribution for a measurement y_i : The expected (or mean) measurement \bar{y}_i and the measurement's variance σ_i^2 . The corresponding probability distribution reads

$$p_G(y_i|\bar{y}_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - \bar{y}_i)^2}{2\sigma_i^2}}. \quad (2.43)$$

If the measurements are independently distributed, the joint negative log-likelihood without constant factors is given as

$$L_G(\mathbf{y}|\bar{\mathbf{y}}, \boldsymbol{\sigma}^2) = -\ln p_G(\mathbf{y}|\bar{\mathbf{y}}, \boldsymbol{\sigma}^2) = \sum_i (y_i - \bar{y}_i)^2 \frac{1}{\sigma_i^2} + \sum_i \ln \sigma_i^2. \quad (2.44)$$

2.3.4.3 Forward model

The data fidelity terms for both Poisson and Gaussian distributed measurements are dependent on the expected measurements $\bar{\mathbf{y}}$, which in their turn are dependent on the expected distribution of the attenuation coefficient $\boldsymbol{\mu}$ of the scanned object. The connection between $\boldsymbol{\mu}$ and $\bar{\mathbf{y}}$ (as well as $\bar{\boldsymbol{\sigma}}$, if applicable) is called the forward model of the reconstruction task. For the attenuation of the intensity of a monochromatic electromagnetic wave passing through matter this forward

model is simply the Lambert-Beer law, see Eq. 2.4:

$$\bar{I}_i = \bar{I}_i^r e^{-\int \mu \, d\ell_i}, \quad (2.45)$$

where \bar{I}_i^r is the expected intensity of a reference scan without any object present and $\int d\ell_i$ represents an integration along a ray pointing from the X-ray source to the detector pixel i . For ideal detectors, the expected intensities $\bar{\mathbf{I}}$ are equal to the expected measurements $\bar{\mathbf{y}}$ and one can directly use this equation, with $\bar{\mathbf{y}} = \bar{\mathbf{I}}$ as the forward model. But for realistic, non-ideal detectors, there are many detector specific effects [27] (offsets, detector blur, non-linearities, afterglow and many more) that must be corrected before the measured data is quantitative. So many CT systems only give access to the measurement data after corrections for these effects are applied, the measurements are normalized with respect to the reference scan intensity, and after taking the negative logarithm. The mean value of this data should then be proportional to the expected line integrals $\bar{\mathbf{I}}$

$$\bar{l}_i = -\ln \frac{\bar{I}_i}{\bar{I}_i^0} = \int \mu \, d\ell_i \quad (2.46)$$

It should be noted that the noise of these line integrals is not Poisson distributed even if the intensity measured by the detector was Poisson distributed before the transformation because of the non-linearity of the logarithm (and of any other non-linear correction steps before).

The forward model for the variance, if using a Gaussian noise model, is extremely dependent on the detector used. In the case of an ideal detector and a monochromatic X-ray source with sufficient flux and the expected variance would be directly proportional to the expected measured intensity,

$$\bar{\sigma}_i^2 = \bar{I}_i. \quad (2.47)$$

For a realistic energy integrating detector this noise model will not be enough and the noise model will have to incorporate a sensitivity s_i and detector readout noise with variance Σ_i^2 for each individual detector element, which would lead to the following noise model,

$$\bar{\sigma}_i^2 = s_i \bar{I}_i + \Sigma_i^2. \quad (2.48)$$

Depending on the detector utilized and on whether non-linear correction steps are performed on the data prior to reconstruction different combinations of noise model and forward model will constitute the most accurate data fidelity term L . For example combining uncorrected intensity measurements, see Eq. 2.45, with an ideal detector with Gaussian noise approximation, see Eq. 2.47, would result in the following likelihood term

$$L(\mathbf{I}|\boldsymbol{\mu}) = \sum_i \left(I_i - I_i^r e^{-\int \boldsymbol{\mu} \, d\mathbf{l}_i} \right)^2 \frac{1}{I_i^r e^{-\int \boldsymbol{\mu} \, d\mathbf{l}_i}} - \int \boldsymbol{\mu} \, d\mathbf{l}_i . \quad (2.49)$$

It should be noted that including a forward model of the measurements variances into the data fidelity term makes the term much more complicated to optimize. The most common solution to this problem is to estimate the expected variances from the actual measurements prior to reconstruction. In the case of Poisson noise an unbiased and efficient estimator of the variance is the mean value of the measurements, or in the case of one measurement simply the value of this measurement

$$\hat{\sigma}_i^2 = I_i. \quad (2.50)$$

Using this, the data fidelity term from Eq. 2.51 becomes much simpler,

$$L(\mathbf{I}|\boldsymbol{\mu}) = \sum_i \left(I_i - I_i^r e^{-\int \boldsymbol{\mu} \, d\mathbf{l}_i} \right)^2 \frac{1}{\hat{\sigma}_i^2}. \quad (2.51)$$

If the estimate for $\boldsymbol{\mu}$ is almost converged, the estimate for $\hat{\sigma}_i^2$ can be updated periodically according to Eq. 2.47 for higher accuracy [31].

2.3.4.4 Regularizer term

Formulating a function that describes the prior probability for the reconstruction estimate to exist is much less straight forward than devising a physical and statistical model for the measurements, as was outlined in the last subsection. Although this can be achieved by using probabilistic approaches [27, 32], this procedure is not particularly intuitive, nor does it seem to be strongly motivated towards a specific solution but mostly seems to be an ex post justification of a successful regularization term.

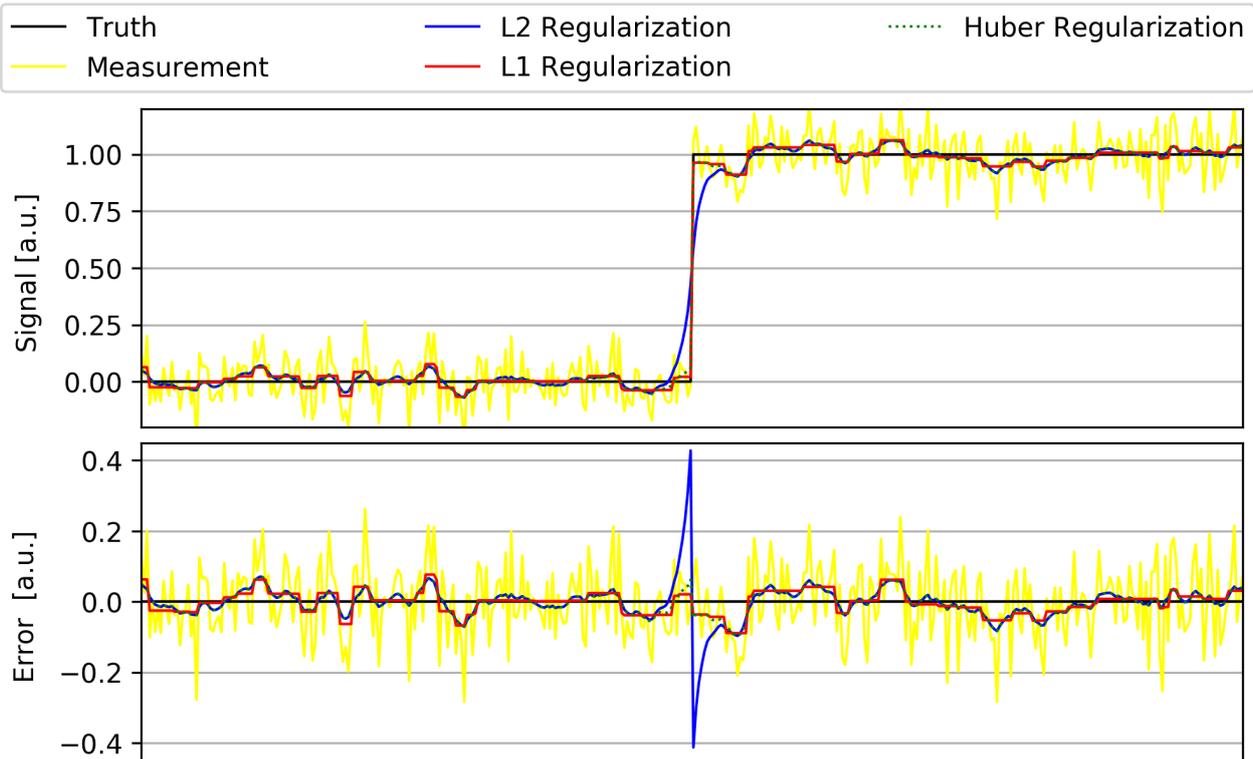


Figure 2.11: Estimation of a step function from measurements corrupted by Gaussian distributed noise. The estimation is regularized by penalizing nearest neighbor differences using different potential functions. The regularization strength is chosen to give the same mean variance in the constant area.

It is more intuitive to see the regularization term as a counterweight to an over-fitting of the estimated quantity to the noise in the measurements: Near the minimum of $L(\mathbf{I}|\boldsymbol{\mu})$ there are a vast number of different possible estimates of $\boldsymbol{\mu}$ that all reduce L to a comparable level. One (or possibly more) of these estimates actually minimizes $L(\mathbf{I}|\boldsymbol{\mu})$ and is thus the maximum likelihood estimate,

$$\hat{\boldsymbol{\mu}}_{ML} = \arg \min_{\boldsymbol{\mu}} L(\mathbf{I}|\boldsymbol{\mu}) , \quad (2.52)$$

for this scan. But for a second identical scan of the same object there would be a different $\hat{\boldsymbol{\mu}}_{ML}$ from the pool of possible estimates simply because the noise of the second scan realized differently. Much more desirable than letting chance decide which estimate the algorithm favors would be designing a criterion which lets the user decide which estimate he favors and thus reducing the effect of chance on the estimate.

A vast number of different criteria based on different assumptions and motivations have been designed by various authors, see [32]. The most widely used criterion is that we assume

our image to be locally smooth. The simplest way to enforce this criterion is to penalize the differences in neighboring voxel values:

$$R(\mathbf{x}) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi(x_j - x_k) , \quad (2.53)$$

where \mathcal{N}_j is a small neighborhood of j and w_{jk} are distance dependent weights to make the small neighborhood more isotropic [33]. The potential function $\psi(r)$ penalizes large values and can take many different forms. The simplest choice for the potential function is the quadratic function

$$\psi_{L2}(r) = r^2 . \quad (2.54)$$

This is a choice which is easy for optimization as the function is convex and quadratic, but it is also problematic as it blurs out sharp edges over several pixels, as dividing the sharp edge into several small steps drastically reduces the sum of the squared differences. A potential function that does not suffer from this problem is the absolute value function,

$$\psi_{L1}(r) = |r| . \quad (2.55)$$

It preserves edges quite well but suffers from two drawbacks: First it produces a piece-wise constant noise pattern that can look quite unnatural, see Fig. 2.11, and it is not differentiable at each point which can cause problems for minimization. Fortunately, there is a group of functions which combines the advantages of regularizers using the quadratic function and of regularizers using the absolute value function. These functions behave like the quadratic function for small arguments and like the absolute value function for large arguments. One of them is the Huber potential function

$$\psi_H(r) = \begin{cases} \frac{1}{2}r^2 & \text{if } |r| < \delta \\ \delta(|r| - \frac{1}{2}\delta) & \text{else} \end{cases} , \quad (2.56)$$

where δ is a shape parameter defining the turning point from quadratic to linear behavior.

Convex potential functions behaving like the absolute value function for large differences and

like the quadratic function are the most often used potential functions in CT reconstruction, e.g. in [34, 35, 36, 11, 12, 37, 32], because they are easy to optimize, work well for diverse objects, and their behavior during the reconstruction is easy to understand.

2.3.5 Reconstruction for dark-field

In first approximation, reconstructions from the dark-field signal \mathbf{D} can be treated completely analogous to the reconstruction of the linear attenuation coefficient from the transmission signal. Under some assumptions [38, 39], the small angle scattering power in a volume element of the sample can be quantified by the linear diffusion coefficient ϵ .

For this quantity the forward model connecting \mathbf{D} and ϵ is given as, see Sec. 2.2.4,

$$D_i = e^{-\frac{2\pi^2 d^2}{p^2} \int \epsilon \, d\ell_{\mathbf{i}}} , \quad (2.57)$$

which can be further transformed to a function of the line integrals over ϵ

$$\int \epsilon \, d\ell_{\mathbf{i}} = -\ln(D_i) \frac{p^2}{2\pi^2 d^2} . \quad (2.58)$$

In this form the linear diffusion coefficient can be reconstructed using any reconstruction algorithm suitable for transmission CT, e.g. FBP or SIR.

It should be noted that the approximations in the introduction of the linear diffusion coefficient are not viable for all materials and setup combinations. Various publications have been published investigating the scattering distribution beyond the Gaussian approximation [40, 41] as well as the directionality of dark-field signal caused by non-isotropic scattering in samples [42, 43], and its connection to the well-established fields of small angle X-ray scattering (SAXS) and ultra-small angle X-ray scattering (USAXS) [44].

2.3.6 Reconstruction for differential phase contrast

The reconstruction of the refractive index from the differential phase contrast (DPC) signal $\frac{\partial\Phi}{\partial y}$ has to be treated differently to the reconstruction from the attenuation or dark-field signal, because of the differential nature of the measured signal.

Again one can directly infer the forward model connecting the refractive index decrement δ with the DPC signal $\frac{\partial\Phi}{\partial y}$, respectively with the shift of the phase stepping curve $\Delta\phi$, from Sec. 2.2.4:

$$\frac{\partial}{\partial y} \int \delta \, d\ell_i = \frac{\Delta\phi_i g_2}{2\pi d} = \frac{\partial\Phi_i}{\partial y} . \quad (2.59)$$

To transform the DPC signal into measurement points of the Radon space of δ an integration along the y -direction, which is perpendicular to the grating bars, is necessary. Any numerical integration algorithm is suitable for this task. However, the constant of integration has to ensure that the phase-contrast signal Φ is zero in regions outside of the sample. For the phase-contrast signal Φ we can write

$$\int \delta \, d\ell_i = \Phi_i , \quad (2.60)$$

which is easily reconstructable using any reconstruction algorithm.

If the reconstruction is performed using the FBP algorithm there is a particularly fast and convenient choice of integration algorithm: Integration in Fourier space.

The Fourier transform of the derivative of a function $f(y)$ can be written as

$$\mathcal{F}_{y\rho} \left[\frac{\partial f(y)}{\partial y} \right] = 2\pi i \rho F(\rho) , \quad (2.61)$$

where $F(\rho)$ is the Fourier transform of $f(y)$, ρ is the spatial frequency in y -direction. Consequently, the integral over a function $f(y)$ can also be expressed in Fourier space:

$$\mathcal{F}_{y\rho} \left[\int f(y) \, dy \right] = \frac{F(\rho)}{2\pi i \rho} + c \delta(\rho) \quad (2.62)$$

where $\delta(u)$ is the delta-distribution and c is an arbitrary constant of integration.

Applying this to the case of DPC data one can write the Fourier transform of the integrated

PC projection under angle θ as follows

$$P_{\theta}(\rho) = \mathcal{F}_{y\rho} [\Phi(y, \theta)] = \mathcal{F}_{y\rho} \left[\int \frac{\partial\Phi}{\partial y}(y, \theta) dy \right] = \mathcal{F}_{y\rho} \left[\frac{\partial\Phi}{\partial y}(y, \theta) \right] \frac{1}{2\pi i\rho} + c \delta(\rho) . \quad (2.63)$$

Now if this equation is inserted into the equation used for the derivation of the FBP, see Eq. 2.31, and c set equal to zero, one arrives at the following equation

$$\begin{aligned} f(x, y) &= \int_0^{\pi} \int_{-\infty}^{\infty} \mathcal{F}_{y\rho} \left[\frac{\partial\Phi}{\partial y}(y, \theta) \right] \frac{1}{2\pi i\rho} |\rho| e^{2\pi i\rho(x \cos(\theta) + y \sin(\theta))} d\rho d\theta \\ &= \int_0^{\pi} \int_{-\infty}^{\infty} \mathcal{F}_{y\rho} \left[\frac{\partial\Phi}{\partial y}(y, \theta) \right] \frac{\text{sgn}(\rho)}{2\pi i} e^{2\pi i\rho(x \cos(\theta) + y \sin(\theta))} d\rho d\theta , \end{aligned} \quad (2.64)$$

where $\text{sgn}(\rho)$ is the sign-function that simply returns the sign of its argument. Now the FBP algorithm for differential data looks very similar to the FBP algorithm for conventional projection data: the Fourier transform of the data is multiplied by $\frac{\text{sgn}(\rho)}{2\pi i}$, the so-called Hilbert filter [45], instead of $|\rho|$, and subsequently backprojected.

2.4 Optimization transfer principle

Almost any general purpose optimization algorithm that requires only the value and gradient of the cost function, can be used to minimize the cost functions of the various statistical iterative CT reconstruction problems, e.g. gradient descent, L-BFGS, (non-linear) conjugate gradient, and many more. Thorough information about these algorithms can be found in all textbooks on numerical optimization [46].

Various algorithms have been specifically designed for the use in CT reconstruction and of these the family of so-called separable quadratic (or paraboloid) surrogate (SQS / SPS) algorithms have found widespread usage. They are all designed using the optimization transfer principle, which will be explained in the following for the case of CT reconstruction.

The optimization transfer principle is a very general framework for the design of optimization algorithms. It tries to solve a seemingly hard optimization problem by repeatedly solving easier optimization problems. Many existing optimization algorithms can be explained using the

optimization transfer principle and it is also a very helpful tool in creating custom algorithms specifically designed for a specific optimization problem.

It was first introduced into medical image reconstruction by De Pierro [47] and Lange [48] and subsequently made extremely popular for CT image reconstruction by various work by Fessler et al. [48, 49, 35].

A general treatment of the framework (under the name "majorization minimization" algorithms) is given by Lange in [50]. A theoretical treatment of convergence is provided by Jacobsen in [51].

2.4.1 General principle

As the name majorization minimization implies, the algorithm consists of two steps. The first step is the majorization step, where a suitable surrogate function $Q^{(n)}(\mathbf{x}|\mathbf{x}^{(n)})$ of the objective function $L(\mathbf{x})$ at the current iteration step n and variable estimate $\mathbf{x}^{(n)}$ is found. $Q^{(n)}$ implies that even the general shape of the surrogate function can change between iterations, not just its constants as implied by $|\mathbf{x}^{(n)}$. Such a function must fulfill two conditions: The tangency condition

$$Q^{(n)}(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) = L(\mathbf{x}^{(n)}) \quad (2.65)$$

and the domination condition

$$Q^{(n)}(\mathbf{x}|\mathbf{x}^{(n)}) > L(\mathbf{x}) \text{ for all } \mathbf{x} . \quad (2.66)$$

These two conditions guarantee that minimizing the surrogate function in the second step (the minimization step) causes the original objective function to decrease:

$$L(\mathbf{x}^{(n+1)}) \leq Q^{(n)}(\mathbf{x}^{(n+1)}|\mathbf{x}^{(n)}) \leq Q^{(n)}(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) = L(\mathbf{x}^{(n)}) . \quad (2.67)$$

It would be sufficient to find an $\mathbf{x}^{(n+1)}$ with a lower function value than $\mathbf{x}^{(n)}$ but for most useful algorithms the surrogate function is something easy to minimize or even directly solvable.

The hard part in designing majorization-minimization algorithms is finding suitable surrogate

functions that fulfil the tangency and domination conditions, are easy to minimize, and stay as close as possible to the original function.

2.4.2 Quadratic surrogates

Often the simplest valid surrogate for a more complex function $h(x)$ is a quadratic function. All continuous functions can be approximated by a second-order Taylor expansion of the objective at the current estimate,

$$h(x) = h(x^{(n)}) + \left. \frac{\partial h(x)}{\partial x} \right|_{x^{(n)}} (x - x^{(n)}) + \frac{1}{2} \left. \frac{\partial^2 h(x)}{\partial x^2} \right|_{x^{(n)}} (x - x^{(n)})^2 + \mathcal{O}(x^3) . \quad (2.68)$$

If $\frac{\partial^2 h(x)}{\partial x^2}$ is bounded for all viable x than there is a quadratic surrogate $q(x|x^{(n)})$ for $h(x)$ at $x^{(n)}$ that looks very similar to the second order Taylor series [50],

$$q(x|x^{(n)}) = h(x^{(n)}) + \left. \frac{\partial h(x)}{\partial x} \right|_{x^{(n)}} (x - x^{(n)}) + \frac{1}{2} c^{(n)} (x - x^{(n)})^2 , \quad (2.69)$$

with a positive curvature $c^{(n)}$. By construction this quadratic function fulfils the tangency condition at $x^{(n)}$

$$q(x^{(n)}|x^{(n)}) = h(x^{(n)}) \quad (2.70)$$

and

$$\left. \frac{\partial h(x)}{\partial x} \right|_{x^{(n)}} = \left. \frac{\partial q(x)}{\partial x} \right|_{x^{(n)}} \quad (2.71)$$

as well as the domination condition for a large enough curvature $c^{(n)}$. Minimizing this surrogate function is directly achieved by a Newton step

$$x^{(n+1)} = x^{(n)} - \frac{1}{c^{(n)}} \left. \frac{\partial h(x)}{\partial x} \right|_{x^{(n)}} . \quad (2.72)$$

From this equation it can be seen that a small curvature $c^{(n)}$ of the surrogate function causes a large update step, thus one would want to choose a $c^{(n)}$ that is as low as possible to achieve a fast convergence of the algorithm, see [52] for an in depth discussion on the optimal choice

of $c^{(n)}$.

Optimization problems in imaging are typically multivariate so we can apply the same procedure to create a multivariate quadratic surrogate function $Q(\mathbf{x})$ to our objective function $H(\mathbf{x})$,

$$Q(\mathbf{x}|\mathbf{x}^{(n)}) = H(\mathbf{x}^{(n)}) + \nabla H(\mathbf{x})^T \Big|_{\mathbf{x}^{(n)}} (\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})^T \mathbf{C}^{(n)} (\mathbf{x} - \mathbf{x}^{(n)}) . \quad (2.73)$$

Here the matrix $\mathbf{C}^{(n)}$ has to be chosen carefully to ensure that the domination condition is fulfilled for all \mathbf{x} . This is the case for a choice of \mathbf{C} that ensures a positive definite matrix $\mathbf{C} - \mathbf{H}_{H(\mathbf{x})}$, where $\mathbf{H}_{H(\mathbf{x})}$ is the Hessian matrix of H for all possible \mathbf{x} . Although this might not be the sharpest quadratic surrogate, finding a general way of constructing sharp multivariate quadratic surrogates for (not necessarily convex) functions is still an open problem [52]. Choosing the optimal \mathbf{C} is not as straight forward for this case as it was for the one-dimensional case, but choosing a viable \mathbf{C} with low diagonal values will generally lead to a faster convergence of the Newton iteration

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \mathbf{C}^{(n)-1} \nabla H(\mathbf{x}) \Big|_{\mathbf{x}^{(n)}} . \quad (2.74)$$

2.4.3 Separable quadratic surrogates

Minimizing a quadratic surrogate can still be quite challenging if the problem at hand is large and the non-diagonal entries of \mathbf{C} are non-zero, because then we need to calculate the inverse of \mathbf{C} to minimize our surrogate function.

Fortunately for many problems in imaging, e.g. CT reconstruction, the cost function can be written in the following way

$$H(\mathbf{x}) = \sum_i h_i(l_i(\mathbf{x})) , \quad (2.75)$$

with $l_i = \sum_j a_{ij} x_j$.

This form allows to construct a surrogate function that separates the variables from each other [53]. First, one chooses quadratic surrogate functions q_i (as in the last subsection) for each original potential function h_i , which results in a quadratic surrogate function Q_{QS} for the

whole objective function,

$$H(\mathbf{x}) \leq Q_{QS}(\mathbf{x}|\mathbf{x}^{(n)}) = \sum_i q_i(l_i(\mathbf{x})|l_i^{(n)}) . \quad (2.76)$$

Then one rewrites $l_i(x)$ using the helper variables α_{ij} with $\alpha_{ij} \geq 0$ ($\alpha_{ij} = 0$ only for $a_{ij} = 0$) and $\sum_j \alpha_{ij} = 1$:

$$\begin{aligned} l_i &= \sum_j a_{ij}x_j \\ &= \sum_j a_{ij}x_j - a_{ij}x_j^{(n)} + a_{ij}x_j^{(n)} \\ &= \sum_j a_{ij}(x_j - x_j^{(n)}) + \sum_j a_{ij}x_j^{(n)} \\ &= \left[\sum_j a_{ij}(x_j - x_j^{(n)}) \right] + l_i^{(n)} \\ &= \sum_j \alpha_{ij} \frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^{(n)}) + \underbrace{\sum_j \alpha_{ij} l_i^{(n)}}_{=1} \\ &= \sum_j \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^{(n)}) + l_i^{(n)} \right) . \end{aligned} \quad (2.77)$$

With this the quadratic surrogate function can be rewritten as

$$\begin{aligned} Q_{QS}(\mathbf{x}|\mathbf{x}^{(n)}) &= \sum_i q_i(l_i(\mathbf{x})|l_i^{(n)}) \\ &= \sum_i q_i \left(\sum_j \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^{(n)}) + l_i^{(n)} \right) | l_i^{(n)} \right) \end{aligned} \quad (2.78)$$

and, after applying Jensen's inequality [50], one ends up with a new surrogate function,

$$Q_{SQS}(\mathbf{x}|\mathbf{x}^{(n)}) = \sum_i \sum_j \alpha_{ij} q_i \left(\frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^{(n)}) + l_i^{(n)} | l_i^{(n)} \right) , \quad (2.79)$$

that dominates the quadratic surrogate function.

Looking at this functions value and gradient at $\mathbf{x}^{(n)}$ reveals that it does fulfil the tangency conditions by construction and looking at its Hessian matrix reveals that all variables are

separated:

$$\begin{aligned}
Q_{SQS}(\mathbf{x}^{(n)}|\mathbf{x}^{(n)}) &= \sum_{i,j} \alpha_{ij} q_i \underbrace{\left(\frac{a_{ij}}{\alpha_{ij}} (x_j^{(n)} - x_j^{(n)}) + l_i^{(n)} | l_i^{(n)} \right)}_{=0} \\
&= \sum_i \sum_j \underbrace{\alpha_{ij}}_{=1} \underbrace{q_i(l_i^{(n)} | l_i^{(n)})}_{=h_i(l_i^{(n)})} = L(\mathbf{x}^{(n)})
\end{aligned} \tag{2.80}$$

$$\begin{aligned}
\left. \frac{\partial Q_{SQS}(\mathbf{x}|\mathbf{x}^{(n)})}{\partial x_k} \right|_{\mathbf{x}^{(n)}} &= \sum_i \alpha_{ik} \frac{\partial q_i \left(\frac{a_{ik}}{\alpha_{ik}} (x_k - x_k^{(n)}) + l_i^{(n)} | l_i^{(n)} \right)}{\partial x_k} \\
&= \sum_i \underbrace{q_i(l_i^{(n)} | l_i^{(n)})}_{=h_i(l_i^{(n)})} a_{ik} = \left. \frac{\partial L(\mathbf{x})}{\partial x_k} \right|_{\mathbf{x}^{(n)}}
\end{aligned} \tag{2.81}$$

$$\left. \frac{\partial^2 Q_{SQS}(\mathbf{x}^{(n)}|\mathbf{x})}{\partial x_k \partial x_l} \right|_{\mathbf{x}^{(n)}} = \begin{cases} \sum_i \frac{a_{ik}^2}{\alpha_{ik}} c_i^{(n)} & \text{if } k = l \\ 0 & \text{else} \end{cases} . \tag{2.82}$$

This means the separable quadratic surrogate function can be written as

$$Q_{SQS}(\mathbf{x}|\mathbf{x}^{(n)}) = H(\mathbf{x}^{(n)}) + \nabla H(\mathbf{x})^T \Big|_{\mathbf{x}^{(n)}} (\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})^T \mathbf{C}^{(n)} (\mathbf{x} - \mathbf{x}^{(n)}) , \tag{2.83}$$

with the diagonal curvature $\mathbf{C}^{(n)}$. For CT reconstruction α_{ij} are commonly [54] chosen as

$$\alpha_{ij} = \frac{a_{ij}}{\sum_{j'} a_{ij'}} , \tag{2.84}$$

with which the entries of the diagonal curvature $\mathbf{C}^{(n)}$ follow as

$$C_{kk}^{(n)} = \sum_i a_{ik} c_i^{(n)} \sum_{j'} a_{ij'} . \tag{2.85}$$

Performing an update using the Newton iteration, see Eq. 2.74, is now trivial.

2.4.4 SQS for CT reconstruction

As written in Sec. 2.3.4, the various possible SIR problems can generally be written as an optimization problem of the following form

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu}} L(\boldsymbol{\mu}|\mathbf{y}) + R(\boldsymbol{\mu}) . \quad (2.86)$$

One can design separable quadratic surrogate functions of the cost function $C = L + R$ by independently finding separable quadratic surrogate functions for the likelihood term L and the regularizer term R . The curvature \mathbf{C} of the cost function is simply the sum of \mathbf{C}_L and \mathbf{C}_R .

2.4.4.1 SQS for the data term

For CT reconstruction problems the data fidelity term L can be written as

$$L(\boldsymbol{\mu}|\mathbf{y}) = \sum_i \psi_i(l_i|y_i) , \quad (2.87)$$

where $l_i = \sum_j a_{ij}\mu_j$ is the forward projection of the current estimate of $\boldsymbol{\mu}$ to a certain detector element at a specific angular position and y_i is the measurement during a scan at that detector element and angular position.

Depending on the noise model for the measurements different potential functions ψ_i are suited for the estimation of $\boldsymbol{\mu}$, see Sec. 2.3.4.2.

One of the simplest forms $\psi_i(l_i|y_i)$ can take is

$$\psi_i(l_i|y_i) = \frac{1}{2}(l_i - y_i)^2 w_i , \quad (2.88)$$

which is the statistically correct term for the maximum-likelihood estimation from measurements y_i with Gaussian distributed noise with variance $\sigma_i^2 = \frac{1}{w_i}$, see Eq. 2.44. Finding a quadratic surrogate q_i for this case is trivial as ψ_i is already quadratic, thus no surrogate is

necessary. But for instructional purposes one can employ Eq. 2.69 to construct q_i :

$$q_i(l_i|l_i^{(n)}) = \frac{1}{2}(l_i - l_i^{(n)})^2 w_i + (l_i^{(n)} - y_i) w_i (l_i - l_i^{(n)}) + \frac{1}{2}(l_i - l_i^{(n)})^2 c_i^{(n)}. \quad (2.89)$$

The curvature $c_i^{(n)}$ has to be chosen such, that q_i fulfills the domination condition for all l_i

$$\begin{aligned} q_i(l_i|l_i^{(n)}) &\geq \psi_i(l_i|y_i) \\ \frac{1}{2} l_i^2 (c_i^{(n)} - w_i) - l_i l_i^{(n)} (c_i^{(n)} - w_i) + \frac{1}{2} l_i^{(n)2} (c_i^{(n)} - w_i) &\geq 0 \\ (l_i - l_i^{(n)})^2 (c_i^{(n)} - w_i) &\geq 0. \end{aligned} \quad (2.90)$$

Obviously, any $c_i^{(n)} \geq w_i$ fulfills this equation and is thus a viable curvature, but the sharpest curvature is achieved for

$$c_i^{(n)} = c_i = w_i. \quad (2.91)$$

Using this, the separable quadratic surrogate function $L_{SQS}(\boldsymbol{\mu}|\mathbf{y})$ for $L(\boldsymbol{\mu}|\mathbf{y})$ is given as

$$L_{SQS}(\boldsymbol{\mu}|\boldsymbol{\mu}^{(n)}) = L(\boldsymbol{\mu}^{(n)}|\mathbf{y}) + \nabla L(\boldsymbol{\mu}|\mathbf{y})^T \Big|_{\boldsymbol{\mu}^{(n)}} (\boldsymbol{\mu} - \boldsymbol{\mu}^{(n)}) + \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}^{(n)})^T \mathbf{C}^L (\boldsymbol{\mu} - \boldsymbol{\mu}^{(n)}) , \quad (2.92)$$

with

$$C_{Lkl} = \begin{cases} \sum_i a_{ik} w_i \sum_j a_{ij} & \text{if } k = l \\ 0 & \text{else} \end{cases}. \quad (2.93)$$

2.4.4.2 SQS for the regularizer term

The same procedure as for the likelihood term has to be performed for the regularizer term,

$$R(\boldsymbol{\mu}) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\mu_j - \mu_k). \quad (2.94)$$

ψ depends on two variables but to construct a separable quadratic update a ψ is needed that depends only on a single variable. This can be achieved using Jensen's inequality,

$$\begin{aligned}
R(\boldsymbol{\mu}) &= \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\mu_j - \mu_k) \\
&= \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi\left(\frac{1}{2} \left[2\mu_j - \mu_j^{(n)} - \mu_k^{(n)}\right] - \frac{1}{2} \left[2\mu_k - \mu_j^{(n)} - \mu_k^{(n)}\right]\right) \\
&\leq \frac{1}{4} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \left[\psi(2\mu_j - \mu_j^{(n)} - \mu_k^{(n)}) + \psi(2\mu_k - \mu_j^{(n)} - \mu_k^{(n)}) \right] \\
&= R_{SS}(\boldsymbol{\mu} | \boldsymbol{\mu}^{(n)}) .
\end{aligned} \tag{2.95}$$

Now one just needs to find a quadratic surrogate $\psi^{QS}(r)$ for $\psi(r)$ if $\psi(r)$ is not the quadratic penalty. The usual shape for a quadratic function that fulfills the tangency condition looks like this,

$$\begin{aligned}
\psi(2\mu_j - \mu_j^{(n)} - \mu_k^{(n)}) &\leq \psi_{jk}^{QS}(2\mu_j | \mu_j^{(n)}, \mu_k^{(n)}) \\
\psi_{jk}^{QS}(2\mu_j | \mu_j^{(n)}, \mu_k^{(n)}) &= \psi(\mu_j^{(n)} - \mu_k^{(n)}) + \dot{\psi}(\mu_j^{(n)} - \mu_k^{(n)})(\mu_j - \mu_j^{(n)}) + c_{jk}^{(n)}(\mu_j - \mu_j^{(n)})^2 .
\end{aligned} \tag{2.96}$$

Now a viable value for $c_{jk}^{(n)}$ has to be found that fulfills the domination condition. The simplest choice that ensures this is to take the maximum value of $\ddot{\psi}(r)$

$$c_{jk}^{(n)} = c = \max_r \ddot{\psi}(r) . \tag{2.97}$$

Combining these two surrogates, one finally arrives at the separable quadratic surrogate function for the regularizer

$$\begin{aligned}
R_{SQS}(\boldsymbol{\mu} | \boldsymbol{\mu}^{(n)}) &= \frac{1}{4} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \left[\psi_{jk}^{QS}(\mu_j | \mu_j^{(n)}, \mu_k^{(n)}) + \psi_{kj}^{QS}(\mu_k | \mu_k^{(n)}, \mu_j^{(n)}) \right] \\
&= \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi_{jk}^{QS}(\mu_j | \mu_j^{(n)}, \mu_k^{(n)}) \\
&= \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi_{jk}^{QS}(\mu_j | \mu_j^{(n)}, \mu_k^{(n)}) ,
\end{aligned} \tag{2.98}$$

with a matrix of second derivatives C_R

$$C_{Rkl} = \begin{cases} \sum_{k \in \mathcal{N}_j} w_{jk} c_{jk}^{(n)} & \text{if } k = l \\ 0 & \text{else} \end{cases} . \quad (2.99)$$

Chapter 3

Advanced signal estimation for grating-based imaging

The estimation of the transmission T , dark-field D , and differential phase contrast ϕ images is the main step of generating useful information from a set of interferograms acquired at different grating positions. The first publications on GBI used a signal estimation based on the Fourier transformation [55, 56] but this method only works for equidistantly distributed grating positions and ignores the statistics of the measurement. An approach to calculate T , D , and ϕ motivated by the maximum-likelihood principle is the weighted least-squares fitting of a sine-wave to the measurements. This approach has been used regularly [57, 12] and also an estimate of the variances of T , D , and ϕ has been proposed [57, 12]. The bias of this maximum-likelihood estimation was discussed at least in one study [58].

In the following, the full analytic solution to the maximum-likelihood estimate through linearization will be presented and analytical expressions for the variance of the maximum-likelihood estimate will be presented.

The expressions for the maximum-likelihood estimates have already been derived in a similar fashion by Hahn et al. [12] but the proposed estimations of the variances are novel findings that generalize findings and correct errors made in [12]. After this, the bias of the MLE will be calculated and novel bias-corrected MLE as well as regularized bias-corrected MLE will be

proposed. In the last section the variance and bias of all estimators will be compared using numerical simulations.

3.1 Maximum likelihood estimator for GBI signal estimation

As argued in Section 2.2.4, the possible intensity measurements of a detector pixel behind an X-ray grating interferometer lie on a phase stepping curve. This means a single intensity measurement I_i depends on the characteristics of the stepping curve without an object in the beam path at the moment the measurement was taken: the mean intensity of the curve I_i^r , its relative amplitude (the visibility) V_i^r , and its phase Φ_i^r at the current grating positioning. But it also depends on characteristics of the object in the beam path: the transmission T , which reduces the mean intensity of the curve, the dark-field signal D which reduces the visibility of the curve, and the phase shift ϕ which shifts the phase position of the stepping curve. This behavior is described in the following equation:

$$I = I_i^r T (1 + V_i^r D \cos(\Phi_i^r + \phi)) . \quad (3.1)$$

During signal estimation, one is interested in estimating T , D , and ϕ as accurate as possible from a series of noisy measurements I_i . It is obvious that even for completely noise-free measurements, at least three measurements with different Φ_i^r (or possibly differing I_i^r and/or V_i^r) are necessary to estimate the three object properties T , D , and ϕ . This is usually performed by changing Φ_i between the measurements by a movement of the gratings. The log-likelihood for a series of N measurements of the same object corrupted by independent Gaussian distributed noise we can be written as

$$L(T, D, \phi) = \sum_i^N (I_i^r T (1 + V_i^r D \cos(\Phi_i^r + \phi)) - y_i)^2 w_i , \quad (3.2)$$

with weights $w_i = \frac{1}{\sigma_i^2}$, where σ_i^2 is the variance of the noise of the measurement.

The maximum-likelihood estimate of the transmission \hat{T}^{ML} , dark-field signal \hat{D}^{ML} and differential phase signal $\hat{\phi}^{ML}$ is given by

$$\hat{T}^{ML}, \hat{D}^{ML}, \hat{\phi}^{ML} = \arg \min_{T, D, \phi} L \quad (3.3)$$

This optimization problem is non-linear and does not have a direct solution. But instead of trying to find an estimate for T , D , and ϕ one can perform variable transformations as follows

$$M = T, \quad C = TD \cos(\phi), \quad S = TD \sin(\phi) \quad (3.4)$$

$$m_i = I_i^r, \quad c_i = I_i^r V_i^r \cos(\phi_i^r), \quad s_i = -I_i^r V_i^r \sin(\phi_i^r) \quad (3.5)$$

and estimate the transformed variables M , C , and S :

$$\hat{M}_{ML}, \hat{C}_{ML}, \hat{S}_{ML} = \arg \min_{M, C, S} \sum_i (M m_i + C c_i + S s_i - y_i)^2 w_i. \quad (3.6)$$

This weighted-least-squares optimization problem can now be solved using weighted linear regression, which provides the minimum variance unbiased estimator. The step-by-step solution for the estimates, their variances, and covariances can be found in App. A.1.

From these maximum-likelihood estimates of M , C , and S , the maximum-likelihood estimates for T , D , and ϕ can be calculated by the inverse variable transform:

$$\hat{T}_{ML} = \hat{M}_{ML}, \quad \hat{D}_{ML} = \frac{\sqrt{\hat{C}_{ML}^2 + \hat{S}_{ML}^2}}{\hat{M}_{ML}}, \quad \hat{\phi}_{ML} = \arctan \left(\frac{\hat{C}_{ML}}{\hat{S}_{ML}} \right) \quad (3.7)$$

The ML-estimate for a linear estimation problem (as the estimation of M , C , and S under Gaussian noise) is automatically unbiased and is also equal to the minimum variance unbiased estimator (MVUE), meaning that no other unbiased estimator can have a lower variance. But, for more general non-linear estimation problems (such as the estimation of T , D , and ϕ under Gaussian noise), the ML-estimator is not necessarily unbiased, it is only asymptotically unbiased, meaning that it only becomes unbiased for a large number of measurements. It follows

that the ML-estimator is not necessarily the MVUE for low numbers of measurements, but it approaches the MVUE asymptotically for large numbers of measurements. This might seem like a minor issue, but as medical X-ray imaging is always about achieving an imaging goal as dose-efficient as possible, an arbitrarily large number of measurements is never available.

3.1.1 Bias of the maximum-likelihood estimator

It has been shown [58] that the ML-estimation of the dark-field signal tends to be biased towards larger values for high noise scenarios and also for the phase signal a break-down of signal estimation has been observed [58] for high noise scenarios and attributed to bias in the ML-estimation of ϕ . A bias in the estimation of a signal is a serious problem, especially if the output of the signal estimation is displayed as an image with quantitatively meaningful values. This is the case because a noisy image can always be filtered with a low-pass filter to reduce the noise in the image (of course with an associated loss of fine detail) but there is no simple filter operation that can reduce the error in an image that is caused by a biased estimation. The human visual perception itself is acting like a low-pass filter: We can see large structures even in a very noisy image, but if the signal is corrupted by bias our visual perception will not be able to make this correction.

The bias $b_\theta(\hat{\theta})$ of an estimator $\hat{\theta}$ of a variable θ is defined as

$$b_\theta(\hat{\theta}) = \mathbb{E}[\hat{\theta}(\mathbf{y}) - \theta] = \int (\hat{\theta}(\mathbf{y}) - \theta) P(\mathbf{y}, \theta) d\mathbf{y} . \quad (3.8)$$

This can often not be explicitly solved, but the expression can be approximated by performing a Taylor expansion of the estimator around the expected values of the measurements $\bar{\mathbf{y}}$ [58]:

$$\hat{\theta}(\mathbf{y}) \approx \hat{\theta}(\bar{\mathbf{y}}) + \sum_i \left. \frac{\partial \hat{\theta}}{\partial y_i} \right|_{\bar{\mathbf{y}}} (y_i - \bar{y}_i) + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 \hat{\theta}}{\partial y_i \partial y_j} \right|_{\bar{\mathbf{y}}} (y_i - \bar{x}_i)(y_j - \bar{y}_j) + \mathcal{O}(y_i^3) . \quad (3.9)$$

For a consistent estimator the following equation approximates the expected value of the estimator:

$$\mathbb{E}[\hat{\theta}(\mathbf{y})] \approx \underbrace{\mathbb{E}[\hat{\theta}(\bar{\mathbf{y}})]}_{\theta} + \sum_i \frac{\partial \hat{\theta}}{\partial y_i} \bigg|_{\bar{\mathbf{y}}} \underbrace{\mathbb{E}[(y_i - \bar{y}_i)]}_{=0} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \hat{\theta}}{\partial y_i \partial y_j} \bigg|_{\bar{\mathbf{y}}} \underbrace{\mathbb{E}[(y_i - \bar{y}_i)(y_j - \bar{y}_j)]}_{K(y_i, y_j)}, \quad (3.10)$$

with the measurement covariance matrix K . From this follows the approximate bias of the estimator as

$$b_{\theta}(\hat{\theta}) = \mathbb{E}[\hat{\theta}(\mathbf{y})] - \theta \approx \frac{1}{2} \sum_{i,j} \frac{\partial^2 \hat{\theta}}{\partial y_i \partial y_j} \bigg|_{\bar{\mathbf{y}}} K(y_i, y_j). \quad (3.11)$$

The calculations for the derivatives necessary for the approximate bias of \hat{T}_{ML} , \hat{D}_{ML} , $\hat{\phi}_{ML}$ can be found in the Appendix.

3.2 Bias-corrected maximum likelihood estimator

Fortunately, it is possible to reduce the bias of the ML-estimator: Cox and Snell [59] published a general formula to calculate the second-order bias-corrected maximum likelihood estimator (BCMLE) and showed that it is the estimator with the lowest variance of all estimators with the same bias as the BCMLE. The BCMLE for the estimated parameter θ_s is given by

$$\hat{\theta}_s^{BC} = \hat{\theta}_s^{ML} - \hat{b}_s^{(2)} \quad (3.12)$$

with

$$\hat{b}_s^{(2)} = \sum_{r,t,u} \frac{1}{2} I_{rs}^{-1} I_{tu}^{-1} (K_{rtu} + 2J_{t,ru}) \quad (3.13)$$

where I_{rs}^{-1} is the inverse of the Fisher information matrix I_{rs} ,

$$I_{rs} = \mathbb{E} \left[\frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right], \quad (3.14)$$

and the tensors $J_{t,ru}$ and K_{rtu} are given by

$$J_{r,st} = \mathbb{E} \left[\frac{\partial L}{\partial \theta_r} \frac{\partial^2 L}{\partial \theta_s \partial \theta_t} \right] \quad (3.15)$$

and

$$K_{rst} = \mathbb{E} \left[\frac{\partial^3 L}{\partial \theta_r \partial \theta_s \partial \theta_t} \right]. \quad (3.16)$$

Now one can use estimators of I_{rs} , $J_{r,st}$, and K_{rst} to calculate an estimated second-order bias correction $b_s^{(2)}$ to the MLE. Formulas for the calculation of I_{rs} , $J_{r,st}$, and K_{rst} are provided in the Appendix.

3.2.1 Regularized BCMLE

Calculating the BCMLE requires the calculation of I_{rs} , $J_{r,st}$, and K_{rst} , which are for the case of grating-based imaging signal estimation, not independent of the measurements. This means that they are not exactly known, but only estimated from the measurements. The estimators provided in the appendix are consistent, meaning they become more accurate with increasing number of measurements, but for a low number of measurements they might be biased themselves. The probably most important term in the calculation, because it affects it quadratically, is the inverse Fisher matrix I_{rs}^{-1} , which is strongly dependent on the estimate of the Fisher matrix I_{rs} . Especially, if diagonal elements of I_{rs} are close to zero, its inverse will be strongly corrupted by inaccuracies. If one looks at the diagonal elements of the Fisher matrix,

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 L(T, D, \phi)}{\partial T^2} \right] &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)^2 w_i \\ \mathbb{E} \left[\frac{\partial^2 L(T, D, \phi)}{\partial D^2} \right] &= \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)^2 w_i \\ \mathbb{E} \left[\frac{\partial^2 L(T, D, \phi)}{\partial \phi^2} \right] &= \sum_i (-TD \sin(\phi) c_i + TD \cos(\phi) s_i)^2 w_i, \end{aligned} \quad (3.17)$$

one can see that they all consist of sums over squared terms. From this, it is evident that the diagonal values can only get small if the summands get small, which is only the case for low values of T and D (and not at all for the second derivative relative to transmission). A simple

way to reduce this problem is to replace low values in the diagonal elements of the Fisher matrix with a given value:

$$I_{rs}^C = \begin{cases} c_{rr} & \text{if } r = s \text{ and } I_{rr} < c_{rr} \\ I_{rs} & \text{else} \end{cases} . \quad (3.18)$$

This results in a smaller inverse Fisher matrix for measurements with low T and D , i.e. high noise scenarios. It should be noted that choosing cutoff values of zero results in the unchanged BCMLE. The result of a smaller inverse Fisher matrix is simply that the bias correction term gets smaller for medium noise regions and approaches a relatively constant value (it still depends on the estimates of J and K , which are of minor influence) for high noise regions. It stays unchanged from the BCMLE for regions with relatively high statistics.

3.3 Results

3.3.1 Dark-field estimator

Simulation studies for two different situations are shown in Fig. 3.1 and Fig. 3.2. In both studies $2 \cdot 10^8$ consecutive phase steppings were performed and the dark-field signal was estimated from each phase stepping using several different estimators: The MLE, the BCMLE with varying regularization strengths and the BCMLE with a clipping operation applied to estimates of D below zero. For all estimators the bias, variance and mean squared error (MSE) was calculated from the samples. In the first study no object was present in the beam path but 45 different numbers of incoming photons per phase stepping (evenly distributed per phase step) were simulated, resulting in varying noise levels of the scans. In the second study a constant flux of 10^3 photons per phase stepping was simulated and 45 different objects in the beam path with varying dark-field signal from 0.01 to 1 and constant transmission ($T = 1$) and differential phase contrast ($\phi = 0$) were simulated. For both studies, five equidistant phase steps per stepping were performed and the reference scan phase was altered between five values between

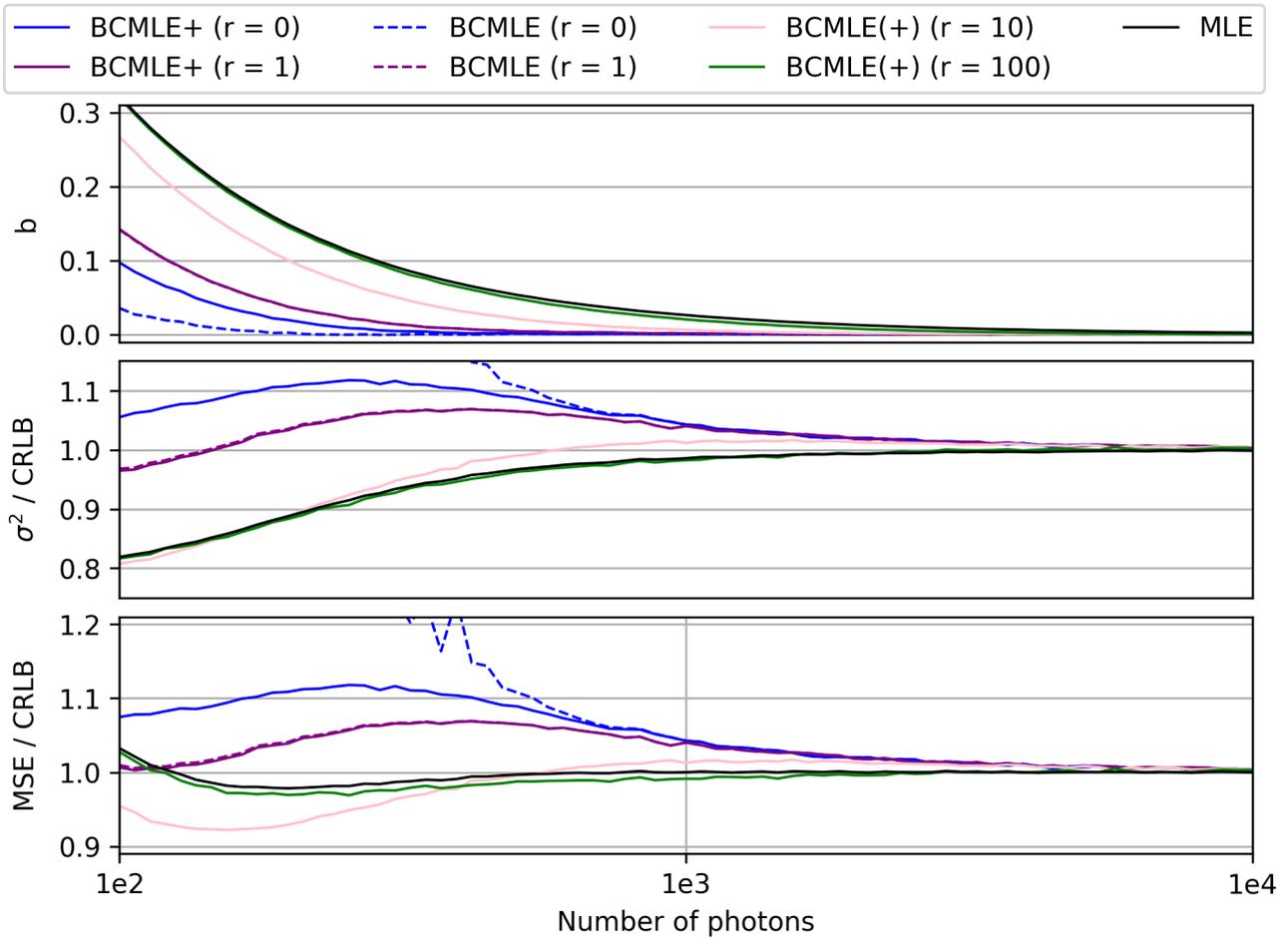


Figure 3.1: The bias b , variance σ^2 , and MSE for the estimation of the dark-field signal for different estimators and different numbers of incoming photons per complete phase stepping was calculated from a total of $2e8$ simulated phase steppings with five stepping positions and a true dark-field signal of unity for each simulation. The variance and MSE are normalized to the true Cramer-Rao lower bound (CRLB) for each scan to be able to better compare the different scenarios. A + behind the name of an estimator means that clipping was performed for this estimator. A (+) means that the curves of the clipped and unclipped estimators are visually identical.

the phase steppings. The interferometer for both simulations had a visibility of 0.2.

The BCMLE generally reduces the bias of the estimation compared to the MLE for all regularization values. The smallest bias is achieved for a regularizer value of zero, i.e. the unregularized BCMLE. Increasing the regularizer value generally reduces the bias reduction. The bias reduction is especially getting smaller with increasing R for high-noise situations, whereas in low-noise regions the regularized estimators behave more like the unregularized estimator. The clipping of negative estimates to zero reduces the bias reduction for the unregularized estimator and for very low regularization values, for higher values clipping does not have any effect.

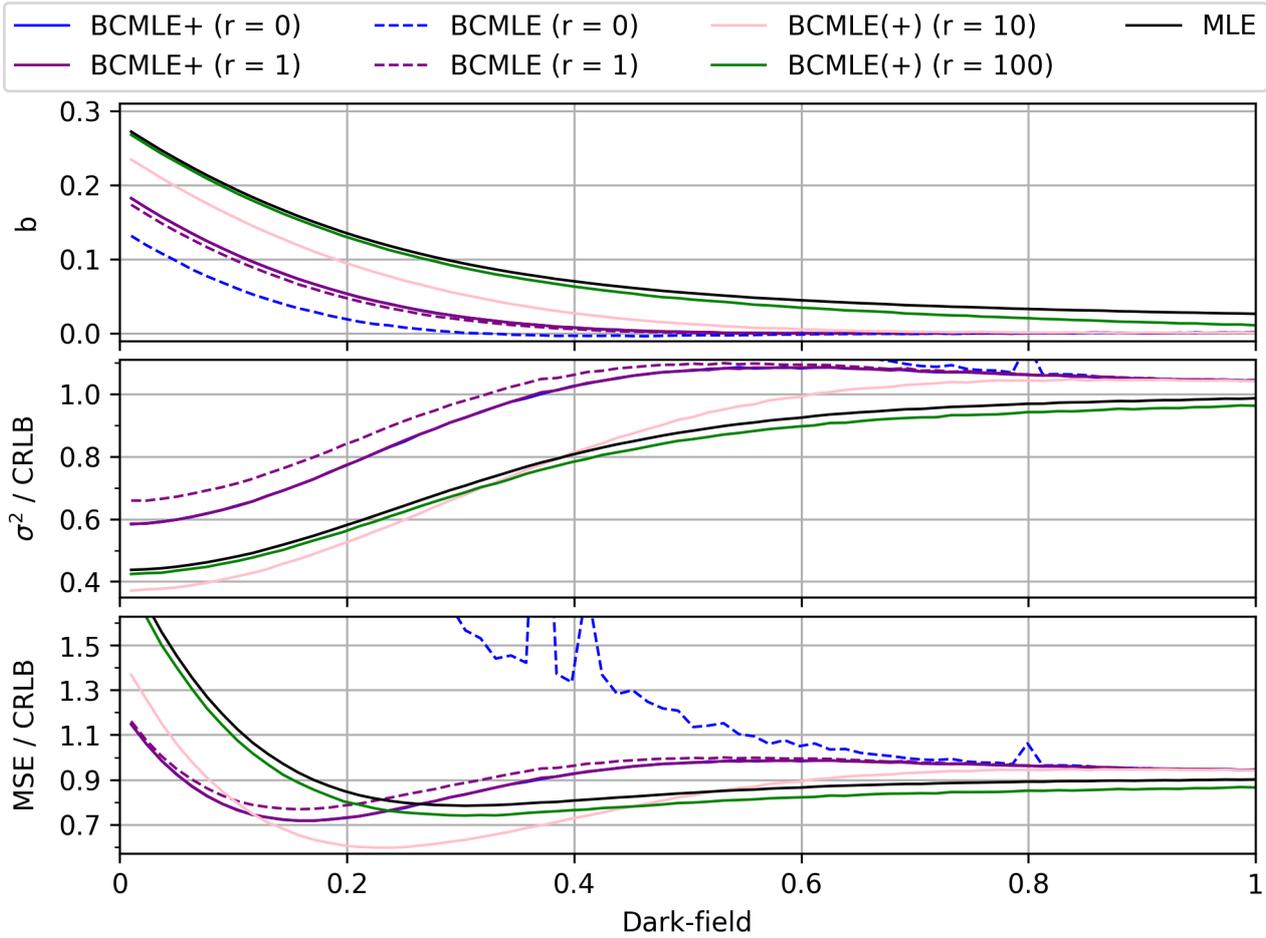


Figure 3.2: The bias b , variance σ^2 , and MSE for the estimation of the dark-field signal for different estimators and different true dark-field values was calculated. The variance and MSE are normalized to the true Cramer-Rao lower bound (CRLB) for each scan to be able to better compare the different scenarios. A + behind the name of an estimator means that clipping was performed for this estimator. A (+) means that the curves of the clipped and unclipped estimators are visually identical.

The variance of the unregularized and only slightly regularized BCMLE is generally above the variance of the MLE. More precisely the variance is well above the MLE for high-noise scenarios and approaching the variance of the MLE for high noise scenarios. For intermediate values of r , the variance of the BCMLE is even lower than the variance of the MLE up to a certain noise level and then approaches the variance of the MLE for low-noise scenarios. When the regularization gets stronger the point where the variance of the BCMLE estimator surpasses the variance of the MLE moves further into lower noise regions but the maximum variance reduction of the BCMLE against the MLE reduces until the BCMLE approaches the MLE for high regularization values. Clipping the negative values generally reduces the variance of the

BCMLe estimator but has little effect on the variance from a certain regularization strength on. This combination of bias reduction and variance increase and decrease (depending on the noise scenario) results in quite complicated behavior of the MSE, which can generally be calculated from the variance and the bias by the following equation

$$\text{MSE}(\theta) = b^2(\theta) + \sigma^2(\theta) . \quad (3.19)$$

Depending on the regularization strength, the BCMLe may have a decreased MSE for high-noise scenarios and an increased MSE for low-noise scenarios. The noise-level where the MSE of the BCMLe is equal to the MLE generally moves further in direction of low-noise scenarios for increasing regularization strength until the MSE approaches the MSE of the MLE for all scenarios for very high regularization strength. As the noise level of a measurement is most often not known before imaging, as the object in the beam path is not known beforehand, and the noise level of the measurement is never uniform across the field of view, as the object is never completely uniform, it is not a simple task to choose a suitable regularization parameter r before the estimation of the signal. But from two observations a good rule of thumb can be proposed. The first observation is that generally an object in the beam path will increase the noise level of the measurement, meaning that a known region without an object or a reference scan without an object in the beam path will always have a lower noise level than the sample scan. The second observation is that the MSE of the BCMLe has a noise level below which the MSE of the BCMLe is lower than of the MLE and that this noise level increases with increasing r . From these two observations one can conclude that there is a value for r where the MSE of the BCMLe crosses the MSE of the MLE at the noise level in background region of the scan. The BCMLe with this regularization strength will decrease the MSE over the whole projection and increase it in no region. If the bias-reduction of this estimator is not sufficient one can define another target function instead of the MSE in the shape of $b + a\sigma^2$ (with $a < 1$) and apply the same routine.

If the choice of BCMLe or BCMLe+ makes a difference one should keep in mind the following: A dark-field signal below zero has no plausible physical meaning and should thus be avoided as a result but on the other hand cutting the range of possible estimates at zero guarantees that

the estimator is biased for low values of D , e.g. if the true value of D is zero any estimate will be either zero or larger than zero inevitably resulting in a mean estimate above zero and thus a bias.

3.3.2 DPC estimator

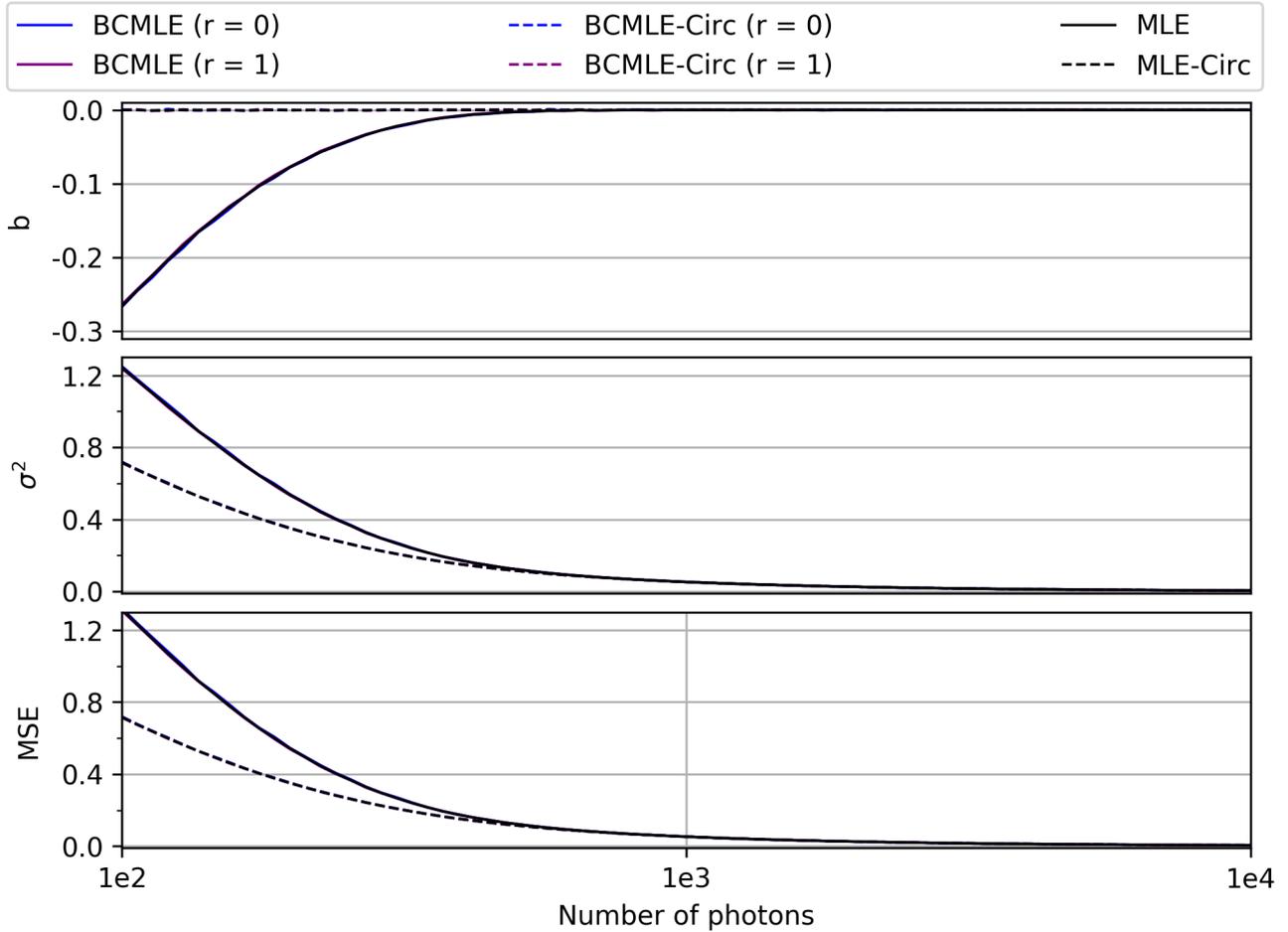


Figure 3.3: The bias b , variance σ^2 , and MSE for the estimation of the differential phase signal for different estimators and different numbers of incoming photons per complete phase stepping was calculated.

The BCMLE correction term is extremely small for the estimation of the DPC signal and thus there is no difference in the variance and bias between the BCMLE and the MLE. But calculating the bias of $\hat{\theta}$ from a sample of measurements \mathbf{y}_i according to

$$b_{\theta}(\hat{\theta}) = \mathbb{E}[\hat{\theta}(\mathbf{y}) - \theta] = \sum_i \hat{\theta}_i(\mathbf{y}_i) - \theta \quad (3.20)$$

clearly shows a bias in the estimation of the DPC signal. To explain the discrepancy between the theoretically expected value and the measured value of the bias one should reconsider if the equations are really suited to calculate the bias in this case: The DPC signal is a circular variable, as one is estimating the phase and thus only the direction of a vector with a length given by the visibility and the dark-field signal, and the estimated expectation value of a circular variable is not simply the mean of the individual estimates but the mean of unit vectors with this phase.

If one calculates the expected value of the DPC signal using the following formula

$$\bar{\phi} = \arg \left(\sum_i \exp(j \phi_i) \right), \quad (3.21)$$

where j is the imaginary unit, the bias suddenly disappears.

This clearly demonstrates that one should keep in mind the nature of the signal one is working with. For DPC imaging scenarios this implies that denoising by use of conventional low-pass filters does not provide good results (as conventional low-pass filter does not treat the DPC signal as a circular variable). A way to denoise images of circular variables would be to replace each pixel with the value of the (weighted) circular mean or circular median of its neighbors.

3.4 Discussion

The chapter showed that the construction of expressions for the estimation of the dark-field signal superior to the MLE is possible. The unregularized BCMLE suffers from an extreme increase in variance and is thus unsuitable for many imaging tasks. The proposed regularized BCMLE allows to smoothly transition from the unregularized BCMLE to the MLE, which results not only in an estimator with reduced bias compared to the MLE but also with a reduced variance for certain noise scenarios. With careful calibration of the regularization strength, the BCMLE can be strictly superior to the MLE in imaging scenarios affected by bias.

Further work is needed to examine the performance of the BCMLE in CT imaging. It should

be noted that here special care must be taken when performing reconstructions using the FBP algorithm, as the FBP algorithm is based on the negative logarithm of the dark-field signal and not on the dark-field signal itself. Calculating the logarithm of the dark-field signal is impossible for estimated dark-field signals below zero. For such a task it would be more suitable to design a BCMLE directly for the estimation of the logarithm of D (which is, differing from the MLE, not the same as the estimator for D). Alternatively, iterative reconstruction has to be performed using a SIR algorithm with a pre-log data model.

The perceived bias of the DPC signal in [58] does not seem to be a real manifestation of bias but a failure to treat a random circular variable with methods of directional statistics.

Chapter 4

Improved sliding window interpolation

4.1 Introduction

The signal estimation step, as presented in the last chapter, requires at least three measurements per projection angle. For CT acquisition it would be highly beneficial to be able to perform a signal estimation using only a single projection per projection angle. A suitable high-quality method for this problem is proposed with the IBSIR-algorithm in Ch. 5 but the IBSIR method is a slow iterative method and often a fast direct method would be desirable for a quick overview and as a starting point for the high-quality method. One possible candidate for a fast reconstruction method using only a single projection per angle is the sliding window acquisition scheme combined with interpolation between neighboring projection angles [60, 61]. This method provides good results in an acquisition scenario where the difference between the line integrals of neighboring projections is only small. This is the case for dense angular sampling and objects with low or slowly changing contrasts. In most realistic use scenarios of single-shot GBCT it is evident that the angular sampling is limited by a multitude of factors. To minimize detector readout noise, one generally wants to limit the number detector readouts to the lowest possible number to achieve a sampling artifact free reconstruction, i.e. around the

Nyquist sampling frequency $N_y = \frac{\pi}{2} N_{px}$, where N_{px} is the number of detector pixels, for FBP reconstruction. Also for high-flux setups with high rotation speeds, i.e. clinical CT systems or time-resolved CT applications, the construction of detectors systems with very high readout speeds and the capability to transfer the measured data to a storage device is challenging. In addition, the requirement for low or slowly changing contrasts is hard to satisfy, i.e. all imaging applications require strong contrasts as for the detection of low contrast features long imaging times are required (which is even more prohibitive in medical X-ray imaging). In the context of medical imaging many structures, i.e. bones, have both sharp delineations and a high contrast and are thus bound to create very strong artifacts if imaged using sliding window interpolation. Interestingly, an interpolated phase stepping that contains information from both sides of a sharp edge of one of the contrasts, i.e. the pixel in question was measuring air for four of five projections and a material that is only attenuating for one of the projections, will cause artifacts not only in the estimation of the attenuation for this pixel but also in the dark-field and differential phase estimation. A visualization of this effect for three different objects can be found in Figures 4.1, 4.2, and 4.3.

The figures show that the signal extraction process for the transmission, dark-field and differential phase contrast by minimizing the least-squares distance to the measurements is not able to handle this situation without producing strong artifacts in all three channels.

In the following methods to mitigate this effect will be explored and evaluated in simulations.

4.2 Robust signal estimation

To extract the transmission T , dark-field D and differential phase signal ϕ from a phase stepping measurement usually a method that minimizes the (possibly weighted) sum of the squared distance of the measurements to the simulated measurements is chosen:

$$\hat{T}, \hat{D}, \hat{\phi} = \arg \min_{T, D, \phi} \sum_i r_i^2 w_i, \quad (4.1)$$

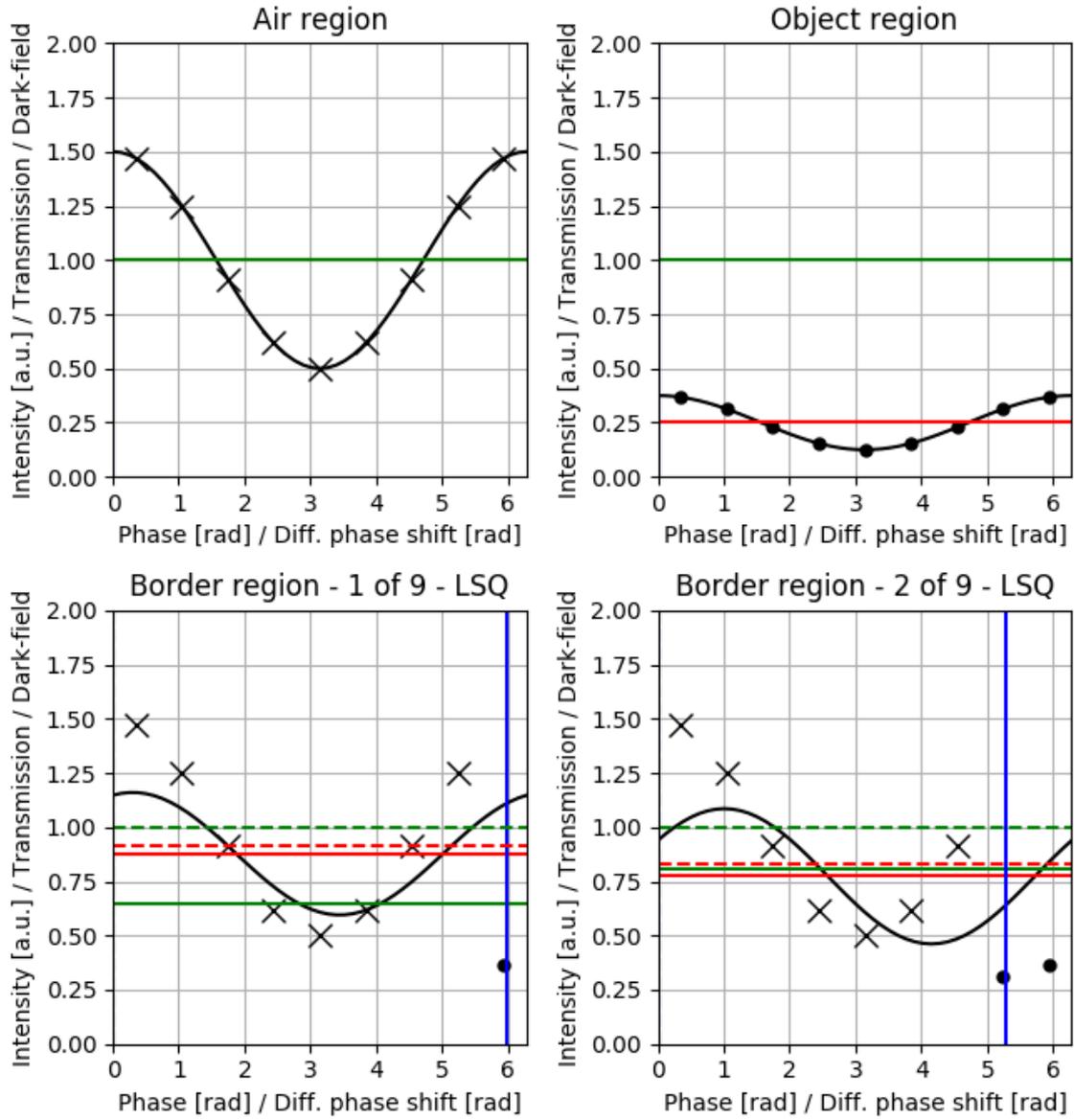


Figure 4.1: Signal estimation for four different situations during the sliding window scan of a sample showing only attenuation. Crosses signify measurements without the sample in the beam path and dots measurements with a sample in the beam path. The solid red lines are the estimated attenuation signal, green lines the estimated dark-field signal and blue lines the differential phase contrast signal. The dashed lines show the mean true signal. The blue dashed line is at zero and the dashed green and red lines are at the same place as the solid lines.

with

$$r_i(T, D, \phi) = I_i^r T (1 + V_i^r D \cos(\Phi_i^r + \phi)) - y_i . \quad (4.2)$$

Minimizing this least-squares problem is the maximum likelihood solution for the case of measurements affected by uncorrelated Gaussian noise with a known variance. In the case of sliding window interpolation, measurement noise is often not the main contributor to errors in the es-

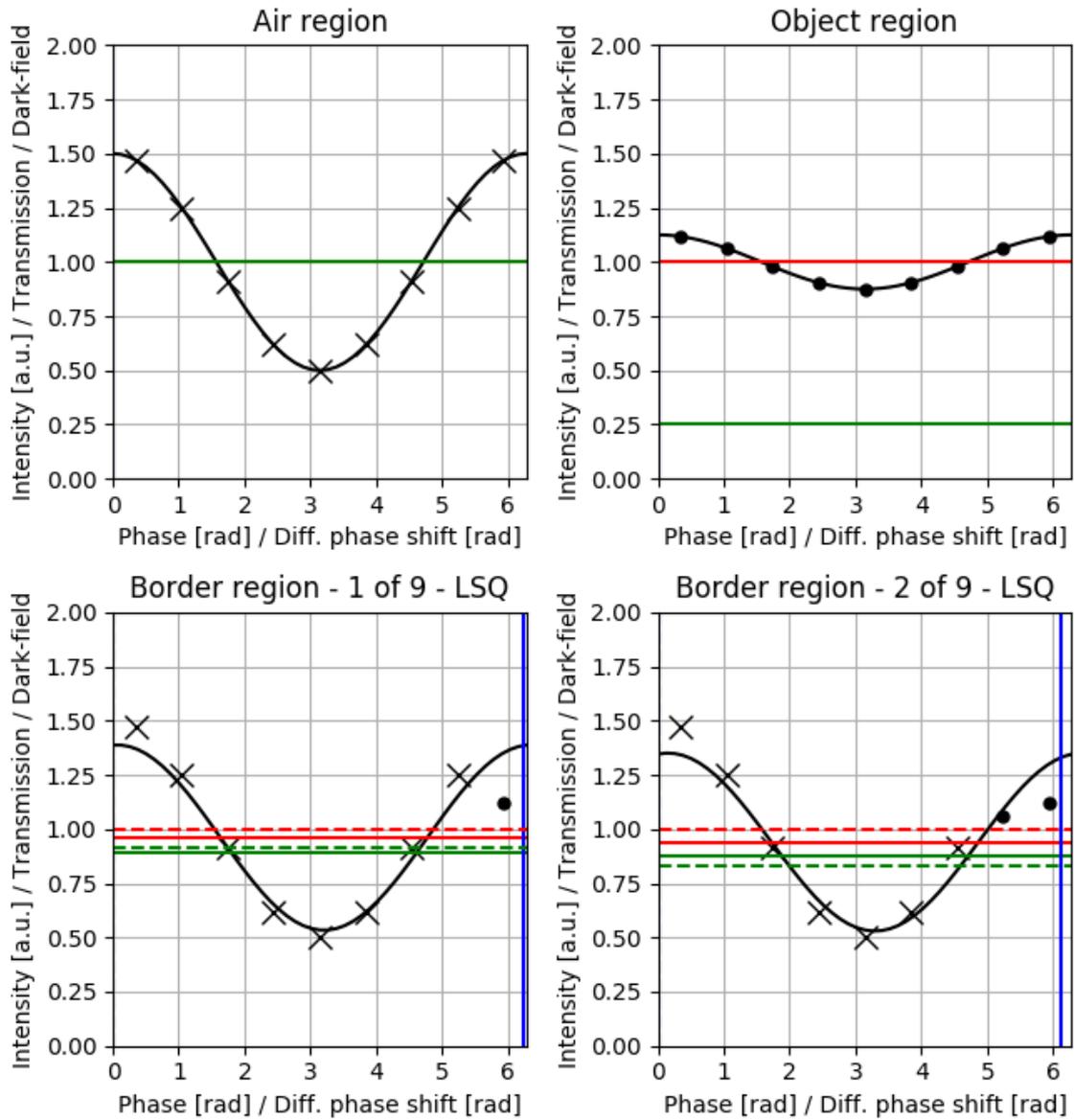


Figure 4.2: Signal estimation for four different situations during the sliding window scan of a sample showing only a dark-field signal. Crosses signify measurements without the sample in the beam path and dots measurements with a sample in the beam path. The solid red lines are the estimated attenuation signal, green lines the estimated dark-field signal and blue lines the differential phase contrast signal. The dashed lines show the mean true signal. The blue dashed line is at zero and the dashed green and red lines are at the same place as the solid lines.

timination of the signal. The problem is more profound because for the different measurements i the underlying true values of T , D , and ϕ are not constant but different for each measurement. As one cannot estimate a set of parameters for each measurement, one must make some assumptions about the measurement. If one assumes that for each estimation most measurements have very similar common true values of T , D , and ϕ and only few of the measurements can

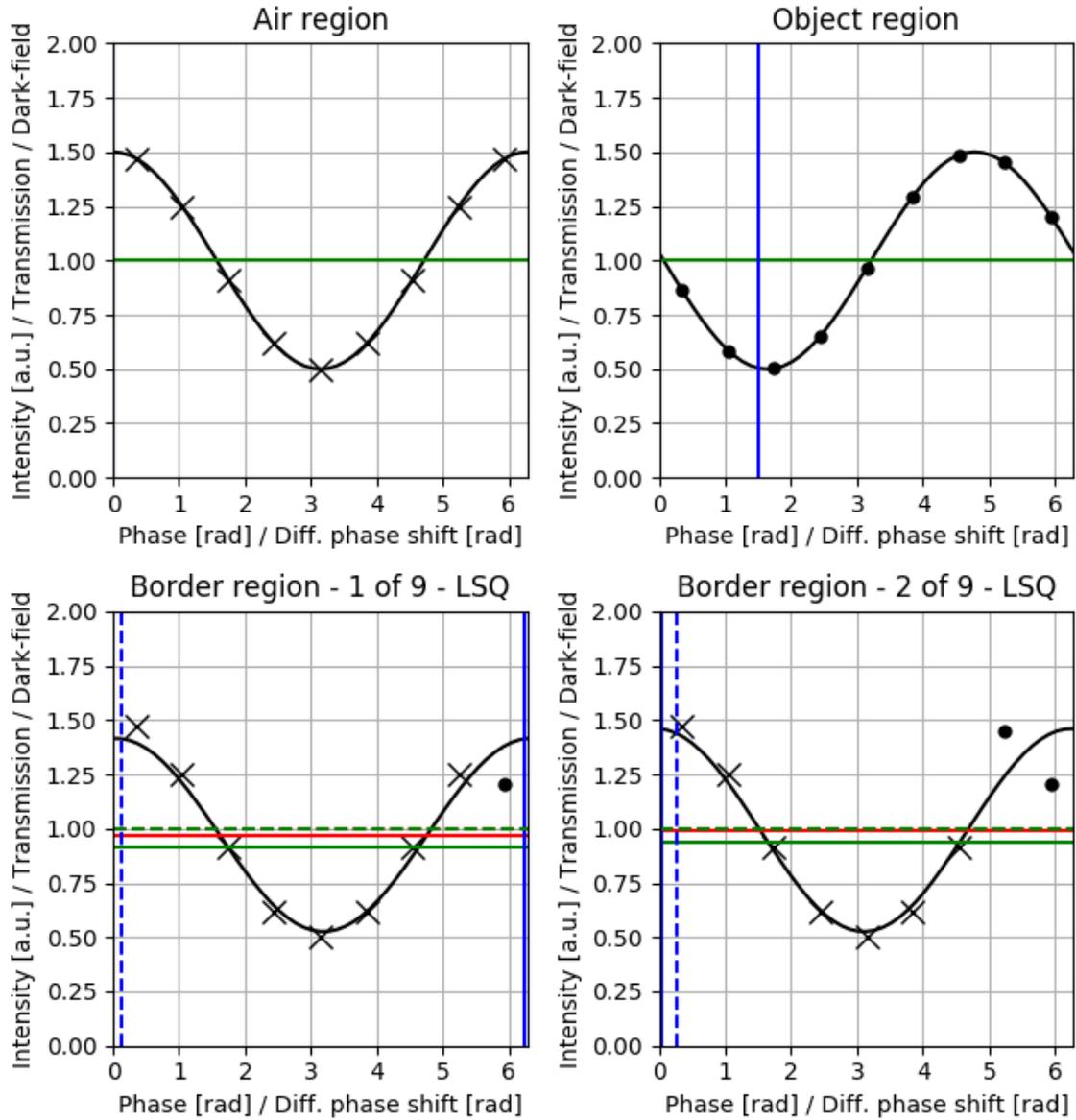


Figure 4.3: Signal estimation for four different situations during the sliding window scan of a sample showing only a differential phase signal. Crosses signify measurements without the sample in the beam path and dots measurements with a sample in the beam path. The solid red lines are the estimated attenuation signal, green lines the estimated dark-field signal and blue lines the differential phase contrast signal. The dashed lines show the mean true signal. The blue dashed line is at zero and the dashed green and red lines are at the same place as the solid lines.

not be explained using these three values, one can treat these measurements as outliers.

A common way to estimate a signal in the presence of outliers is to minimize a different function of the absolute errors of the estimation than the squared errors. A suitable function that is more robust to outliers is, for example, the minimization of the sum of absolute errors. This is often called minimization of the l_1 norm of the error vector (as opposed to the minimization of

the l_2 norm of the error vector in the least-squares problem),

$$\hat{T}, \hat{D}, \hat{\phi} = \arg \min_{T, D, \phi} \sum_i |r| w_i . \quad (4.3)$$

This approach is less susceptible to large errors because they are not squared and thus are not as significant to the cost function as during l_2 -norm minimization. However this l_1 loss function is non-convex at zero and is not as efficient at estimation as the l_2 loss function in the presence of Gaussian noise. A loss function that combines the advantages of both the l_2 - and the l_1 loss function is the Huber loss function, it is equivalent to the l_2 -loss function for small errors and equivalent to the l_1 loss function for large errors:

$$\hat{T}, \hat{D}, \hat{\phi} = \arg \min_{T, D, \phi} \sum_i h(r_i) w_i , \quad (4.4)$$

with

$$h(r) = \begin{cases} \frac{1}{2}r^2 & \text{if } |r| < \delta \\ \delta(|r| - \frac{1}{2}\delta) & \text{else} \end{cases} . \quad (4.5)$$

The parameter δ provides a smooth variation between robustness and efficiency in the presence of Gaussian noise. In addition the Huber function is convex.

For the minimization of the resulting cost function one should note that even though $h(r)$ is convex with respect to r the cost function is not convex in respect to the variables of interest T , D , and ϕ as r is not a convex function of these. Using a gradient based optimization algorithm for this problem could be slow and is not guaranteed to converge to the global minimum. Fortunately many optimization problems of this kind can be solved using the method of iteratively re-weighted least-squares (IRLS). This method iteratively approaches the solution of the optimization problem by solving continuously re-weighted least-squares problems. This is extremely helpful for the problem as there exists a unique, analytic solution for the minimization of the weighted least-squares estimation problem, see Sec. 3.1. Thus one can optimize the non-convex cost function by repeatedly solving convex problems with unique solutions. For

further information on the IRLS algorithm see [62].

Various other non-convex loss functions exist that are even better at rejecting outliers [63]. But these loss functions will generally have multiple minima and there is no guarantee to find the best solution during optimization. Especially if the optimization of these non-convex loss functions is started at the l_2 -minimum it is even highly probable that the new minimum will always be a minimum near the l_2 -minimum and not the globally optimal one.

A visualization of the effectiveness of suppressing outliers in the same situations as in Figures 4.1, 4.2, and 4.3 can be found in Fig. 4.4. From these graphs a shortcoming of this approach becomes evident: Even if the measurements inside the sample are treated as outliers and almost ignored by the signal estimation, the estimated values are not correct. They approach the value of the signal in the region of which most measurements were taken, not a weighted mean value of the values in these two regions. During a tomographic reconstruction this will cause a blurring of the edge. On the other hand, the values of the l_2 -optimization are often not even between the values in both regions but far larger or smaller, resulting in sharp artifacts in the reconstruction.

A further step should be employed to reduce the effect of edge artifacts in the signal extraction: As the non-linear signal estimation is often causing very strong artifacts at the border the use of a median filter along the angular direction of the sinogram will greatly reduce these artifacts without strongly compromising the resolution along this direction. The quality of the signal estimation in each pixel is not equal, as some contain edges, and some do not. Various goodness-of-fit measures are available from the signal estimation process and this information can be used during the filter step. A possible way to do this is by using a weighted median filter [64, S. 111], instead of a plain median filter, along the angular direction of the signal sinograms. As weights w^g for this filter one can use the inverse of the mean of the absolute values of the residuals of the fit in each pixel. As one optimally wants to use a median filter that has a range smaller than the sliding window distance this would often result in a very small number of values for each filter process (usual numbers if stepping positions are between three and ten) which make the use of a median filter very unstable. To counteract this problem, another weighting factor w^d , which decrease quickly with distance to the central pixel, is added.

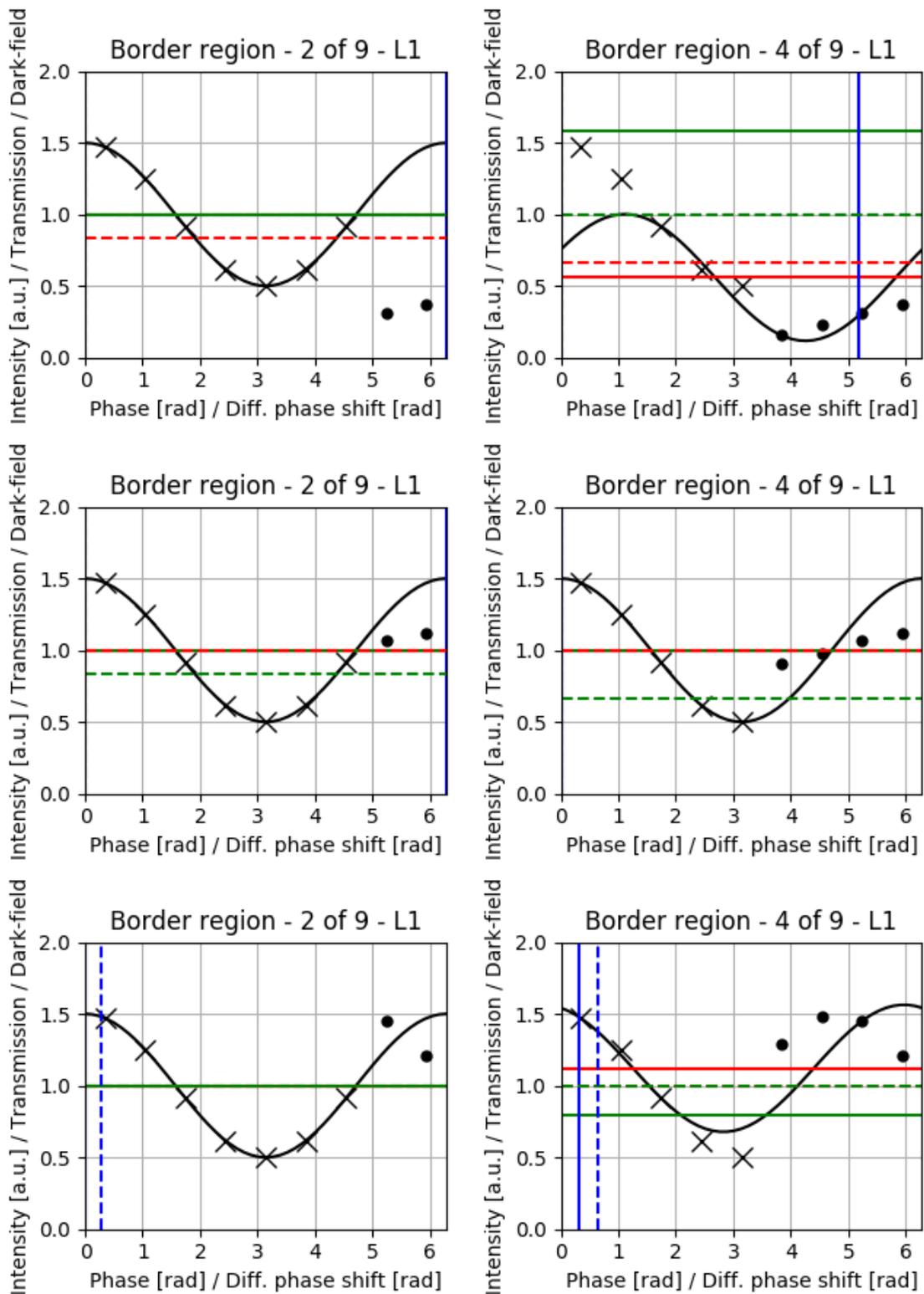


Figure 4.4: Signal estimation for two different situations during the sliding window scan of three different samples showing only attenuation, dark-field or differential phase signal (top to bottom).

Using this factor one can increase the reach of the filter in regions with low w^q while keeping it very local in the case of relatively high w^q . The result of the weighted median filter is the

minimum of the following optimization problem,

$$\hat{\beta}_i = \arg \min_{\beta_i} \sum_{j \in \mathcal{N}_i} |x_j - \beta_i| w_j^a w^d(i, j) \quad (4.6)$$

with

$$w_d(i, j) = s^{-|i-j|}. \quad (4.7)$$

For information on the computation of the weighted median filter see [64, S. 111].

4.3 Results

To evaluate the performance of the proposed improvements numerical simulations were performed. As no general numerical phantoms consisting of the map of attenuation coefficient μ , refractive index decrement δ , and linear diffusion coefficient ϵ are readily available, a suitable numerical phantom was created. As the performance at a very specific situation was not of interest but the performance for a general, diverse and challenging imaging task, an anthropomorphic shape of the phantom was not necessary. Nevertheless, the basic shape and contrast ratios of the well-established FORBILD head phantom [65] were chosen, because this phantom consists of several different sets of features: Many small, grouped, high contrast regions, which are prone to undersampling artifacts and require a high resolution to properly reconstruct, and several larger low contrast regions which require high sensitivity to properly reconstruct. The shape and contrast ratios of this phantom were used for all three material properties. If all the material properties are perfectly co-aligned it is hard to differentiate which material property is responsible for causing a specific artifact as the artifacts are all originating from the same region. Although this is most often the case in real measurements, more insight about the origin of artifacts can be gained from the simulations if the maps of the three material properties are simply rotated with respect to each other. Of course, this destroys any anthropomorphology of the phantom but now one has regions in the phantom which are only attenuating, refracting or scattering and regions with an overlap of these effects. The signal level in each channel was chosen to result in a minimum transmission and dark-field signal of 50% and a maximum differ-

ential phase-contrast of 0.1. The geometry of the imaging setup was simulated to be in parallel beam. The simulations were performed on a 512x512x100 voxel grid and the image formation process was assumed to follow the equations given in Ch. 2. We assumed a monochromatic mean reference scan intensity per pixel of 10^9 photons, resulting in virtually noise-free images, a uniform setup visibility of 0.4 and a distribution of low frequency Moiré fringes on the detector with a period of 18 pixels in both detector dimensions. Measurements from 1205 angles were taken from equidistant positions between 0 and 360 degrees and the grating period of one grating was simulated to move by $\frac{1}{5}$ of a grating position per change of projection angle. The transmission, differential phase contrast, and dark-field contrast were estimated from the resulting interferograms using sliding window interpolation with a window size of five projections. Once the l_2 -optimization objective, see Eq. 4.1, and once the l_1 -optimization objective, see Eq. 4.3, was used. All estimations were then filtered using a median filter with a range of five projections in the angular projection direction and afterwards with a Gaussian filter with a width of $\sigma = 1$ and a cutoff after 2 projections. The resulting sinograms, respectively their distance to the ground truth, can be seen in Fig. 4.5. Upon close inspection it is observable that some of the estimation errors in the dark-field and differential phase-contrast image have a slightly smaller local extent in the l_1 estimation than in the l_2 estimation, the black and white pattern of the artifact has a small gray area between them. This is likely caused by l_1 estimation being able to remove a not-consistent measurement from the estimation in these pixels. A comparison of the filtered images reveals that these regions are filtered by the median filter more effectively. A look at the MSE and the MAE indicates a strong error reduction for D and Φ for the choice of l_1 -estimation over l_2 -estimation if no filtering is applied. After filtering this difference in the characteristics is not as strongly pronounced anymore.

The transmission image on the other hand does not benefit from l_1 estimation over l_2 estimation, both show very comparable estimation errors that are about an order of magnitude lower than the estimation error of the dark-field signal (at equal signal power). This indicates that the sliding window interpolation does not strongly harm the estimation of transmission images. Filtering the attenuation sinograms worsens the estimation as it causes a loss of resolution.

Both findings are supported by the calculated error statistics, which can be found in Tab. 4.1.

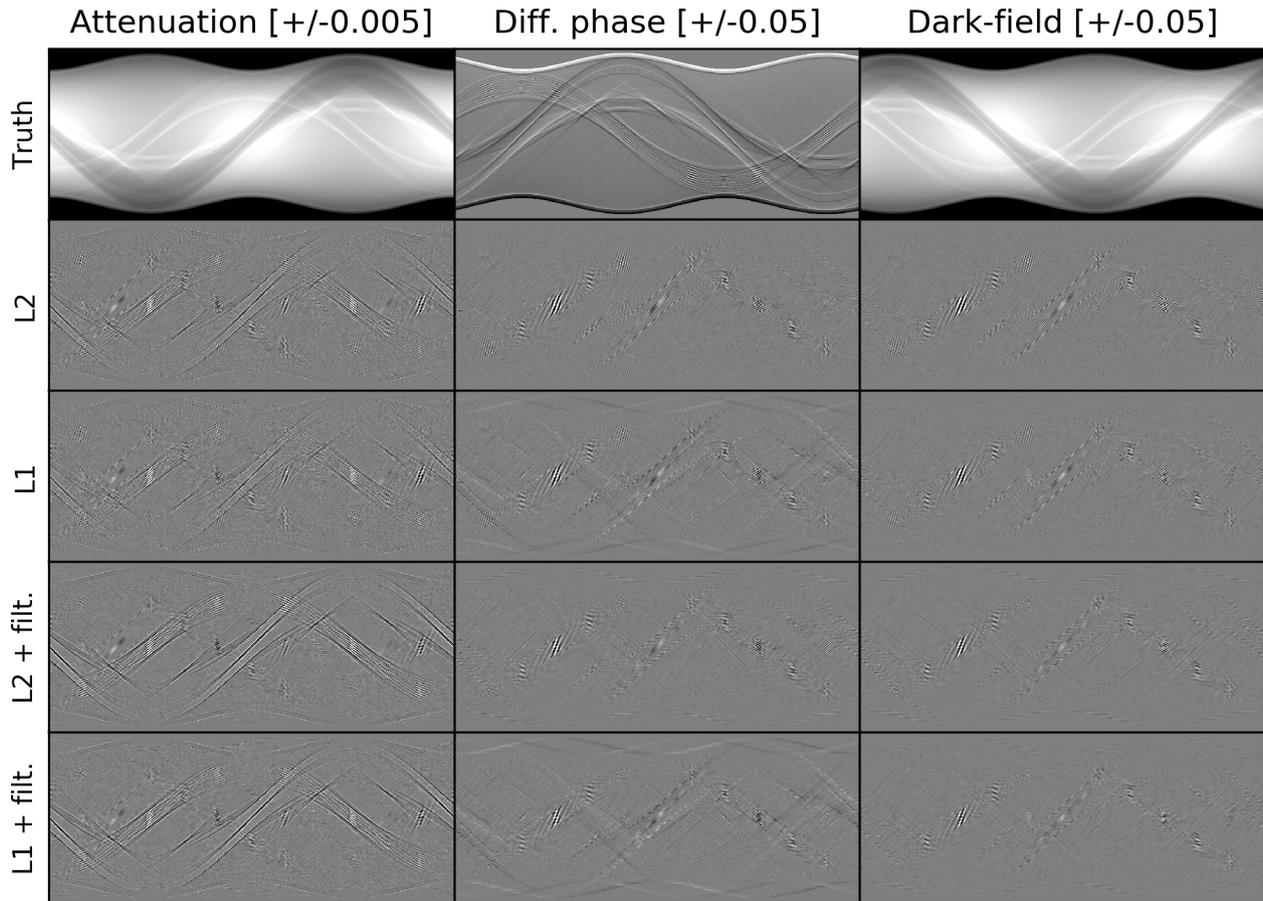


Figure 4.5: Sinograms of the difference to the ground truth for estimated attenuation, differential phase and dark-field signal using l_2 - or l_1 -optimization after and before filtering.

4.4 Discussion

The numerical simulation indicated that l_1 estimation of signals can significantly reduce the estimation error of dark-field and differential phase projections acquired with sliding window interpolation, this is probably caused by an increased robustness to attenuation edges. The attenuation signal estimation is not as strongly affected by SWI but does not benefit from l_1

	NMSE in %			NMAE in %		
	T	ϕ	D	T	ϕ	D
l_2	1.0	1.0	1.0	1.0	1.0	1.0
l_1	1.06	0.601	0.595	1.05	0.829	0.848
l_2 -F	1.24	0.251	0.185	0.969	0.567	0.470
l_1 -F	1.14	0.245	0.159	0.947	0.582	0.462

Table 4.1: Normalized mean squared error (NMSE) and mean average error (NMAE) for sinograms estimated using l_2 - or l_1 -optimization before and after filtering.

estimation in the example. It is recommended to utilize the l_2 estimation or the estimation of the attenuation signal, that is readily available at the first iteration of the l_1 estimation via IRLS.

Filtering the attenuation and differential phase-contrast projections in angular direction further decreases the estimation error. Even a small filter size causes a significant error reduction while not strongly affecting the signal resolution as it already compromised at edges by edge-artifacts from the signal estimation.

While the ability to use only a single projection per projection angle is highly beneficial the resolution acquired using SWI is significantly reduced even with the use of the advanced methods proposed in this chapter. I assume that further possible improvements of the estimation are possible but will probably not allow a completely artifact-free signal estimation.

Taking into account that more efficient reconstruction algorithms for single-shot acquisition schemes (including sliding window acquisition) have been developed in the scope of this thesis, see Ch. 5, a utilization of advanced signal estimation from sliding window interpolation is probably most suited for the generation of starting images for the intensity-based statistical iterative reconstruction introduced in the next chapter.

Chapter 5

Statistical iterative reconstruction for grating-based CT

5.1 Introduction

In the previous sections improvements to the conventional workflow for image reconstruction from GBCT scans were introduced. The conventional workflow consists mainly of the signal estimation step and the actual image reconstructions based on the estimated attenuation, dark-field, and differential phase projections. In the following a reconstruction method that is performing the image reconstruction in a single step by directly estimating the distribution of material properties from the measured interferograms without any intermediate signal estimation in the projection domain will be introduced. The general idea of this method has been proposed in [10] and recently been employed in several publications by the author and others [66, 67, 68, 69]. It will be referred to as intensity-(or interferogram-)based statistical iterative reconstruction (IBSIR) in the following.

Basing the reconstruction directly on the raw measured interferograms provides several strong advantages over the conventional workflow of signal estimation in the projection domain and subsequent reconstructions into image space: The statistical distribution of the measurements is much clearer for the unprocessed detector output than for the estimated projections, lead-

ing to superior image quality in the reconstructions. Furthermore, the process is very easy to adapt to faulty data points or projections without corrupting larger parts of the signal estimation process or rendering parts of the acquired data points unusable. But the most important advantage of IBSIR over any other reconstruction technique proposed yet for GBCT is its ability to reconstruct images acquired with any scan pattern without employing interpolation or restricting assumptions. It will be shown that this allows highest quality GBCT reconstructions even from scans with only a single projection per projection angle.

In the following sections first the model and optimization problem behind IBSIR is introduced. Then an algorithm specifically designed to efficiently solve this optimization problem is developed. After this the reconstruction from scans with a single projection per projection angle is demonstrated and a loose sampling criterion for GBCT is proposed. Finally, expansions of the algorithm allowing the estimation of various system drifts individually for each projection are proposed and tested.

The results in this chapter are a major part of the novel findings presented in this thesis. Huge parts of it have already been published in previous works by the author [68, 69, 70, 71] and are presented in a merged and amended version in the following.

5.2 Formulation

The reconstruction method is formulated as an optimization problem, which requires the minimization of a cost function by iterative adaptation of the values of the spatial distribution of the linear attenuation coefficient $\boldsymbol{\mu}$, the refractive index decrement $\boldsymbol{\delta}$, and the linear diffusion coefficient $\boldsymbol{\epsilon}$. The cost function C is derived from the principle of maximum-a-posteriori probability (MAP), see Sec. 2.3.4.1: It consists of two terms, the likelihood term L , which is a measure of the probability that the measurements originates from the estimated reconstruction, and the regularization term R that incorporates prior knowledge about the image

$$C(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) = L(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) + R(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) . \quad (5.1)$$

Now the solution of our reconstruction method $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\delta}}$, and $\hat{\boldsymbol{\epsilon}}$ are the vectors $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$, that minimize this cost function

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\epsilon}} = \underset{\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}}{\operatorname{arg\,min}} C . \quad (5.2)$$

5.2.1 Data term

The likelihood term L is the negative log-likelihood of the reconstruction given the measurements \mathbf{y} and thus depends on the underlying statistical distribution of the measurements. Without limitation of generality, we will assume that the measured intensity y_i at each data point i is Gaussian distributed around its mean value \bar{I}_i with variance σ_i^2 . After normalizing for pre-factors, this provides the following log-likelihood term L

$$L(\mathbf{y}|\mathbf{I}) = \sum_i \frac{1}{\sigma_i^2} (I_i(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) - y_i)^2 , \quad (5.3)$$

where I_i is the mean intensity of data point i as predicted by a forward model of the imaging system.

The forward model for \mathbf{I} contains the influence of the variables $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ on the measurements and it is a direct combination of the model underlying signal extraction using the weighted least squares approach, see Sec. 3.1, and the iterative reconstruction algorithms for transmission, dark-field and phase-contrast computed tomography, see Sec. 2.3:

$$I_i(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) = I_i^0 T_i(\boldsymbol{\mu}) [1 + V_i^0 D_i(\boldsymbol{\epsilon}) \cos(\phi_i(\boldsymbol{\delta}) + \Phi_i^0)] . \quad (5.4)$$

The additional parameters \mathbf{I}^0 , \mathbf{V}^0 , $\boldsymbol{\Phi}^0$ are not dependent on the variables to be estimated and represent the mean intensities, visibilities, and phase positions of the stepping curves of a hypothetical scan without any sample present in the beam path. They are normally estimated from a reference scan shortly before or after the scan. The transmission \mathbf{T} , dark-field \mathbf{D} , and differential phase contrast $\boldsymbol{\phi}$ on the other hand are directly related to the object estimation.

Following Lambert-Beers-Law, see Eq. 2.4, the transmission in one pixel is given by the negative

exponential of the line integral over the attenuation coefficient μ :

$$T_i = \exp \left[- \int_{l_i} \mu \, d\mathbf{x} \right], \quad (5.5)$$

where $\int_{l_i} d\mathbf{x}$ represents the line integral from the X-ray source to a specific detector pixel at a specific angular position.

The dark-field signal can be modeled in the same way as the attenuation signal, see Eq. 2.21:

$$D_i = \exp \left[- \int_{l_i} c_D \epsilon \, d\mathbf{x} \right] \quad (5.6)$$

with a setup geometry dependent constant c_D for quantitatively comparable results [38].

The differential phase contrast is dependent on the distribution of the refractive index decrement δ as follows, see Eq. 2.13:

$$\Phi_i = c_\Phi \partial_{\perp l_i} \int_{l_i} \delta \, d\mathbf{x}, \quad (5.7)$$

where $\partial_{\perp l_i}$ is the partial derivative perpendicular to the direction of the grating bars and the direction of the line integral l_i and c_Φ is a setup geometry specific constant.

For brevity, the explicit dependence of \mathbf{I} , \mathbf{T} , \mathbf{D} , and ϕ on μ , δ , and ϵ will be dropped in the following.

5.2.2 Regularizer

The regularizer term R of the cost function represents prior knowledge about the characteristics of the sample. Various different choices for this term are possible, see Sec. 2.3.4.4. To cut down the number of choices only regularizer terms that are separate for each material property will be used in this thesis:

$$R = \sum_{\theta \in \{\mu, \delta, \epsilon\}} \beta_\theta R_\theta. \quad (5.8)$$

It is immediately clear that this choice of regularizer is not optimal for the reconstruction problem at hand, as it is not possible to include a very simple fact about all real objects: The spatial distributions of the attenuation and refractive index are not independent of each

other. But the choice is necessary, as the design of suitable multi-channel regularizers for CT reconstruction is still an open research field and was not part of the work for this thesis. Also, the range of single channel regularizers is already quite diverse and a huge amount of different regularizers have been employed for iterative reconstruction of attenuation CT [32]. The simplest regularizers penalize differences in the value of neighboring voxels:

$$R_{\theta} = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\theta_j - \theta_k) , \quad (5.9)$$

where \mathcal{N}_j is the neighborhood of voxels around voxel j , w_{jk} is a weighting function dependent on the distance between voxels j and k , and $\psi(\theta_j - \theta_k)$ is a potential function penalizing value differences between the voxels j and k .

The advantages of the Huber potential function were discussed in Sec. 2.3.4.4 and as it (and very similar potential functions) is an extremely common choice of regularizer for CT reconstruction problems [32]. It was used (with a neighborhood of 26 voxels and weighting dependent on the inverse distance between the voxels) for all following reconstructions.

5.3 Algorithms

General purpose optimization algorithms can be utilized to minimize the IBSIR optimization problem [10, 71, 70]. But a specialized solver designed specifically for this problem is necessary to address the specific challenges inherent to this non-linear, non-convex optimization problem. The problem size of CT reconstruction generally prohibits the application of conventional Newton methods that require storage and computation of complete second order information about the optimization problem. But the IBSIR optimization problem is not only large-scale but also non-linear, meaning that a step in the Newton direction (inverse of the Hessian multiplied with the gradient) could not guarantee to monotonically decrease of the cost function but would require the additional application of a line search algorithm along the direction of the step to ensure monotonic decrease of the cost function. Applying a line search at every iteration is extremely expensive as just for the evaluation of the cost function a forward projection of

each volume is necessary. From this follows that a suitable algorithm should ensure monotonic decrease of the cost function without line searches while still maintaining a fast convergence speed comparable to Newton methods.

Another necessary requirement for the fast convergence of an optimization algorithm is that the algorithm makes sure that the optimization problem is well conditioned to guarantee that that all parts of the optimization problem converge at a similar speed. As all CT reconstruction problems, the IBSIR problem is quite badly conditioned in various aspects: First of all, between voxels, meaning that a small change in one specific voxel generally carries a different influence on the cost function as the same change in another voxel. Another important conditioning problem can be found in the response of the cost function to changes of different spatial frequencies [72, 73]: In the attenuation and dark-field channel this means that the convergence of low frequencies is much faster than the convergence of small details contained in the high frequencies, because large structures have a much higher impact on the cost function than small details. In refractive index channel the dependency is the other way around: The differential operation after the forward projection of the refractive index decrement results in the frequency response of a very strong high-pass filter. This results in the effect that low frequency components of the refractive index reconstruction have almost no effect on the cost function, resulting in extremely slow convergence of low image frequencies.

But the IBSIR problem is not only badly conditioned in the spatial and the frequency domain but also across the different channels: A small change in the attenuation of one voxel will always carry a larger influence on the cost function than a small change in the refractive index decrement or linear diffusion coefficient of this voxel.

In the following an optimization algorithm specifically build for IBSIR that addresses all these challenges will be proposed.

5.3.1 Separable paraboloid surrogate for IBSIR

The majorization-minimization principle has been applied to create the widely used separable paraboloid surrogate (SPS) algorithm for attenuation CT [49, 35, 37, 54]. This algorithm has

a spatially even convergence rate and produces update steps that are ensured to monotonically decrease the cost function. Applying this principle to the IBSIR optimization problem will be the topic of this section.

As explained for attenuation CT in Sec. 2.4, it is necessary to construct a voxel-wise separable paraboloid surrogate function for the IBSIR log-likelihood term,

$$L(\mu, \epsilon, \delta) = \sum_i \left(y_i - I_i^0 e^{-\sum_j a_{ij} \mu_j} \left(1 + V_i^0 e^{-c_D \sum_j a_{ij} \epsilon_j} \cos(\Phi_i^0 + c_\Phi \sum_j a_{ij}^D \delta_j) \right) \right)^2 w_i, \quad (5.10)$$

where a_{ij} are the elements of a system matrix describing the forward projection and a_{ij}^D are the elements of a system matrix describing the *differential* forward projection.

To construct a voxel-wise separable surrogate function L has to be written as a function of the line integrals \mathbf{l}^μ and \mathbf{l}^ϵ and the differential line integrals \mathbf{l}^{δ^D} :

$$l_i^\mu = \sum_j a_{ij} \mu_j, \quad l_i^\epsilon = c_D \sum_j a_{ij} \epsilon_j, \quad l_i^{\delta^D} = -c_\Phi \sum_j a_{ij}^D \delta_j, \quad (5.11)$$

$$L(\mathbf{l}^\mu, \mathbf{l}^\epsilon, \mathbf{l}^{\delta^D}) = \sum_i \left(y_i - I_i^0 e^{-l_i^\mu} \left(1 + V_i^0 e^{-l_i^\epsilon} \cos(\Phi_i^0 + l_i^{\delta^D}) \right) \right)^2 w_i. \quad (5.12)$$

Using this notation, the cost function can be written as a sum of pixel-wise (but not material-wise) separate functions $h_i(\mathbf{l}_i)$

$$L(\mathbf{l}^\mu, \mathbf{l}^\epsilon, \mathbf{l}^{\delta^D}) = \sum_i h_i(\mathbf{l}_i), \quad (5.13)$$

where \mathbf{l}_i is a vector containing the relevant line integrals for pixel i . Using a suited quadratic surrogate function q_i a convex paraboloid surrogate function $Q_{PS,1}$ can be constructed:

$$Q_{PS,1}(\mathbf{l}; \mathbf{l}^{(n)}) = \sum_i q_i(\mathbf{l}_i; \mathbf{l}_i^{(n)}) \quad (5.14)$$

with

$$\begin{aligned}
q_i(\mathbf{l}_i; \mathbf{l}_i^{(n)}) &= h_i(\mathbf{l}_i^{(n)}) \\
&+ \sum_k \frac{\partial h_i(\mathbf{l}_i)}{\partial l_i^k} \Big|_{l_i^k=l_i^{k(n)}} (l_i^k - l_i^{k(n)}) \\
&+ \sum_k \sum_m c_i^{km(n)} (l_i^k - l_i^{k(n)})(l_i^m - l_i^{m(n)}) ,
\end{aligned} \tag{5.15}$$

where k and m sum over μ , ϵ , and δ^D . The curvature $c_i \in \mathbb{R}^{3 \times 3}$ needs to be positive definite and it has to ensure that $q_i(\mathbf{l}_i; \mathbf{l}_i^{(n)}) \geq h_i(\mathbf{l}_i)$ is valid for all \mathbf{l}_i . Possible choices for c_i will be discussed later but first a voxel-wise separable paraboloid surrogate function Q_{SPS} will be constructed on basis of the convex paraboloid surrogate function $Q_{PS,1}$.

Finding a separable surrogate is usually achieved by invoking Jensen's inequality, see Sec. 2.4,

$$f\left(\sum_j \lambda_j \theta_j\right) \leq \sum_j \lambda_j f(\theta_j) \tag{5.16}$$

for a convex function f and positive weights λ_j with $\sum_j \lambda_j = 1$. This equation is valid for scalar and vector valued θ_j . It is possible to identify f with q_i and θ with a vector containing $\boldsymbol{\mu}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\delta}$. Unfortunately λ_j cannot be directly identified with a_{ij} because the sum over a_{ij} is only connecting $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ with their line integral but $\boldsymbol{\delta}$ is connected to its *differential* line integrals by a sum over a_{ij}^D . To be able to invoke Jensen's inequality a surrogate function $Q_{PS,2}$ will be constructed based on $Q_{PS,1}$, which does not use any differential projections but only conventional projections. The first step for this effort is to rewrite the differential forward projection $A^D \mathbf{x}$ as the differentiated conventional forward projection $DA \mathbf{x}$, where D is a suitable matrix that differentiates the projections in direction perpendicular to the grating bars:

$$l^{\delta^D} = A^D \boldsymbol{\delta} = DA \boldsymbol{\delta} = D l^\delta . \tag{5.17}$$

Expanding the terms of $Q_{PS,1}$,

$$\begin{aligned}
Q_{PS,1}(l; l^{(n)}) &= \sum_i q_i(l_i^\mu, l_i^\epsilon, l_i^{\delta^D}; l_i^{\mu(n)}, l_i^{\epsilon(n)}, l_i^{\delta^D(n)}) \\
&= \sum_i h_i^{(n)} \\
&+ \sum_i \dot{h}_i^\mu (l_i^\mu - l_i^{\mu(n)}) + \sum_i \dot{h}_i^\epsilon (l_i^\epsilon - l_i^{\epsilon(n)}) + \sum_i \dot{h}_i^{\delta^D} (l_i^{\delta^D} - l_i^{\delta^D(n)}) \\
&+ \sum_i c_i^{\mu\mu(n)} (l_i^\mu - l_i^{\mu(n)})(l_i^\mu - l_i^{\mu(n)}) + \sum_i c_i^{\epsilon\epsilon(n)} (l_i^\epsilon - l_i^{\epsilon(n)})(l_i^\epsilon - l_i^{\epsilon(n)}) \\
&+ \sum_i c_i^{\delta^D\delta^D(n)} (l_i^{\delta^D} - l_i^{\delta^D(n)})(l_i^{\delta^D} - l_i^{\delta^D(n)}) + \sum_i c_i^{\mu\epsilon(n)} (l_i^\mu - l_i^{\mu(n)})(l_i^\epsilon - l_i^{\epsilon(n)}) \\
&+ \sum_i c_i^{\mu\delta^D(n)} (l_i^\mu - l_i^{\mu(n)})(l_i^{\delta^D} - l_i^{\delta^D(n)}) + \sum_i c_i^{\epsilon\delta^D(n)} (l_i^\epsilon - l_i^{\epsilon(n)})(l_i^{\delta^D} - l_i^{\delta^D(n)})
\end{aligned} \tag{5.18}$$

and recalling that the surrogate function should have the same value for $\mathbf{l} = \mathbf{l}^{(n)}$ and the same gradient at $\mathbf{l} = \mathbf{l}^{(n)}$, allows to see that there is no need to change the construction of the constant and linear part in \mathbf{l} . So only the sums which are quadratic in \mathbf{l} need attention (and here only

the ones with $l_i^{\delta^D}$ contributions). For the term containing $l_i^{\delta^D}$ squared one can write

$$\begin{aligned}
& \sum_i c_i^{\delta^D \delta^D} \left(l_i^{\delta^D} - l_i^{\delta^D(n)} \right)^2 \\
&= \sum_i c_i^{\delta^D \delta^D} \left(\sum_{i'} d_{ii'} l_{i'}^\delta - l_i^{\delta^D(n)} \right)^2 \\
&= \sum_i c_i^{\delta^D \delta^D} \left(\sum_{i'} \left[d_{ii'} l_{i'}^\delta - d_{ii'} l_{i'}^{\delta(n)} + d_{ii'} l_{i'}^{\delta(n)} \right] - l_i^{\delta^D(n)} \right)^2 \\
&= \sum_i c_i^{\delta^D \delta^D} \left(\sum_{i'} \left[d_{ii'} (l_{i'}^\delta - l_{i'}^{\delta(n)}) \right] + \sum_{i'} d_{ii'} l_{i'}^{\delta(n)} - l_i^{\delta^D(n)} \right)^2 \\
&= \sum_i c_i^{\delta^D \delta^D} \left(\sum_{i'} \left[d_{ii'} (l_{i'}^\delta - l_{i'}^{\delta(n)}) \right] \right)^2 \\
&= \sum_i c_i^{\delta^D \delta^D} \left(\sum_{i'} \left[\beta_{ii'} \frac{d_{ii'}}{\beta_{ii'}} (l_{i'}^\delta - l_{i'}^{\delta(n)}) \right] \right)^2 \\
&\leq \sum_i c_i^{\delta^D \delta^D} \sum_{i'} \beta_{ii'} \left(\frac{d_{ii'}}{\beta_{ii'}} (l_{i'}^\delta - l_{i'}^{\delta(n)}) \right)^2 \\
&\leq \sum_i c_i^{\delta^D \delta^D} \sum_{i'} \sum_j \beta_{ii'} \alpha_{i'j} \left(\frac{d_{ii'}}{\beta_{ii'}} \frac{\alpha_{i'j}}{\alpha_{i'j}} (\delta_j - \delta_j^{(n)}) \right)^2
\end{aligned} \tag{5.19}$$

If this term is differentiated two times one will be left with

$$\begin{aligned}
& \frac{\partial}{\partial \delta_k} \frac{\partial}{\partial \delta_k} \sum_i c_i^{\delta^D \delta^D} \sum_{i'} \sum_j \beta_{ii'} \alpha_{i'j} \left(\frac{d_{ii'}}{\beta_{ii'}} \frac{a_{i'j}}{\alpha_{i'j}} (\delta_j - \delta_j^{(n)}) \right)^2 \\
&= \frac{\partial}{\partial \delta_k} \sum_i \sum_{i'} 2c_i^{\delta^D \delta^D} d_{ii'} a_{i'k} \left(\frac{d_{ii'}}{\beta_{ii'}} \frac{a_{i'k}}{\alpha_{i'k}} (\delta_k - \delta_k^{(n)}) \right) \\
&= \sum_i \sum_{i'} 2c_i^{\delta^D \delta^D} \frac{d_{ii'}^2 a_{i'k}^2}{\beta_{ii'} \alpha_{i'k}} \\
&= \sum_i \sum_{i'} 2c_i^{\delta^D \delta^D} \frac{a_{i'k}^2}{\alpha_{i'k}} \frac{d_{ii'}^2}{|d_{ii'}|} \sum_{i''} |d_{ii''}| \\
&= \sum_i \sum_{i'} 2c_i^{\delta^D \delta^D} \frac{a_{i'k}^2}{\alpha_{i'k}} |d_{ii'}| \\
&= \sum_i 2c_i^{\delta^D \delta^D} \left(\frac{1}{2} \frac{a_{i+1k}^2}{\alpha_{i+1k}} + \frac{1}{2} \frac{a_{i-1k}^2}{\alpha_{i-1k}} \right) \\
&= \sum_i c_{i+1}^{\delta^D \delta^D} \frac{a_{ik}^2}{\alpha_{ik}} + \sum_i c_{i-1}^{\delta^D \delta^D} \frac{a_{ik}^2}{\alpha_{ik}} \\
&= \sum_i c_{i+1}^{\delta^D \delta^D} |a_{ik}| \sum_j |a_{ij}| + \sum_i c_{i-1}^{\delta^D \delta^D} |a_{ik}| \sum_j |a_{ij}| \\
&= \sum_i |a_{ik}| (c_{i+1}^{\delta^D \delta^D} + c_{i-1}^{\delta^D \delta^D}) \sum_j |a_{ij}| \\
&\approx \sum_i 2c_i^{\delta^D \delta^D} |a_{ik}| \sum_j |a_{ij}| .
\end{aligned} \tag{5.20}$$

This means that the curvature $c_i^{\delta^D \delta^D}$ derived for a surrogate problem with a conventional projection of $\boldsymbol{\delta}$ instead of the differential projection needs to be multiplied by two to get a suitable curvature for the original problem with the differential projection. The terms containing the off-diagonal elements $c_i^{\delta^D \mu}$ and $c_i^{\delta^D \epsilon}$ are majorized by terms containing the absolute value of the respective entry of c_i and the conventional line integrals of $\boldsymbol{\delta}$.

These two results lead to the new surrogate function $Q_{PS,2}$:

$$\begin{aligned}
Q_{PS,2}(l; l^{(n)}) &= \sum_i q_i(l_i^\mu, l_i^\epsilon, l_i^{\delta^D}; l_i^{\mu(n)}, l_i^{\epsilon(n)}, l_i^{\delta^D(n)}) \\
&= \sum_i h_i^{(n)} \\
&+ \sum_i \dot{h}_i^\mu (l_i^\mu - l_i^{\mu(n)}) + \sum_i \dot{h}_i^\epsilon (l_i^\epsilon - l_i^{\epsilon(n)}) + \sum_i \dot{h}_i^{\delta^D} (l_i^{\delta^D} - l_i^{\delta^D(n)}) \\
&+ \sum_i c_i^{\mu\mu(n)} (l_i^\mu - l_i^{\mu(n)})(l_i^\mu - l_i^{\mu(n)}) + \sum_i c_i^{\epsilon\epsilon(n)} (l_i^\epsilon - l_i^{\epsilon(n)})(l_i^\epsilon - l_i^{\epsilon(n)}) \\
&+ \sum_i 2c_i^{\delta^D\delta^D(n)} (l_i^\delta - l_i^{\delta(n)})(l_i^\delta - l_i^{\delta(n)}) + \sum_i c_i^{\mu\epsilon(n)} (l_i^\mu - l_i^{\mu(n)})(l_i^\epsilon - l_i^{\epsilon(n)}) \\
&+ \sum_i |c_i^{\mu\delta^D(n)}| (l_i^\mu - l_i^{\mu(n)})(l_i^\delta - l_i^{\delta(n)}) + \sum_i |c_i^{\epsilon\delta^D(n)}| (l_i^\epsilon - l_i^{\epsilon(n)})(l_i^\delta - l_i^{\delta(n)}) .
\end{aligned} \tag{5.21}$$

Now the separation trick (multiplying by $\frac{\alpha_{ij}}{\alpha_{ij}}$ and evoking Jensen's inequality to get the sum over j and one α_{ij} to the front) can finally be performed to get the following voxel-wise separable function.

$$\begin{aligned}
Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i h_i^{(n)} \\
&+ \sum_i \dot{h}_i^\mu (l_i^\mu - l_i^{\mu(n)}) + \sum_i \dot{h}_i^\epsilon (l_i^\epsilon - l_i^{\epsilon(n)}) + \sum_i \dot{h}_i^{\delta^D} (l_i^{\delta^D} - l_i^{\delta^D(n)}) \\
&+ \sum_i \sum_j c_i^{\mu\mu(n)} \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^{(n)}) \right) \\
&+ \sum_i \sum_j c_i^{\epsilon\epsilon(n)} \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\epsilon_j - \epsilon_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\epsilon_j - \epsilon_j^{(n)}) \right) \\
&+ \sum_i \sum_j 2c_i^{\delta^D\delta^D(n)} \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\delta_j - \delta_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\delta_j - \delta_j^{(n)}) \right) \\
&+ \sum_i \sum_j c_i^{\mu\epsilon(n)} \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\epsilon_j - \epsilon_j^{(n)}) \right) \\
&+ \sum_i \sum_j |c_i^{\mu\delta^D(n)}| \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\delta_j - \delta_j^{(n)}) \right) \\
&+ \sum_i \sum_j |c_i^{\epsilon\delta^D(n)}| \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\epsilon_j - \epsilon_j^{(n)}) \right) \left(\frac{a_{ij}}{\alpha_{ij}} (\delta_j - \delta_j^{(n)}) \right) .
\end{aligned} \tag{5.22}$$

As in Sec. 2.4 α_{ij} is chosen as

$$\alpha_{ij} = \frac{a_{ij}}{\sum_{j'} a_{ij'}} . \quad (5.23)$$

Taking the second derivative towards two different voxels k and l (and arbitrary material properties θ and θ') reveals that this surrogate function is indeed voxel-wise separated:

$$\begin{aligned} & \frac{\partial^2}{\partial\theta_k \partial\theta'_l} Q_{SPS}(\theta; \theta^{(n)}) \\ &= \frac{\partial}{\partial\theta'_l} \sum_i \dot{h}_i^\theta a_{ik} \\ &+ \frac{\partial}{\partial\theta'_l} \sum_i c_i^{\theta\mu^{(n)}} (\mu_k - \mu_k^{(n)}) \sum_{k'} a_{ik'} \\ &+ \frac{\partial}{\partial\theta'_l} \sum_i c_i^{\theta\epsilon^{(n)}} (\epsilon_k - \epsilon_k^{(n)}) \sum_{k'} a_{ik'} \\ &+ \frac{\partial}{\partial\theta'_l} \sum_i c_i^{\theta\delta^{(n)}} (\delta_k - \delta_k^{(n)}) \sum_{k'} a_{ik'} \\ &= 0 . \end{aligned} \quad (5.24)$$

But taking the second derivative towards a single voxel k reveals that it is not separated material property-wise:

$$\begin{aligned} & \frac{\partial^2}{\partial\theta_k \partial\theta'_k} Q_{SPS}(\theta; \theta^{(n)}) \\ &= \frac{\partial}{\partial\theta'_k} \sum_i \dot{h}_i^\theta a_{ik} \\ &+ \frac{\partial}{\partial\theta'_k} \sum_i c_i^{\theta\mu^{(n)}} (\mu_k - \mu_k^{(n)}) \sum_{k'} a_{ik'} \\ &+ \frac{\partial}{\partial\theta'_k} \sum_i c_i^{\theta\epsilon^{(n)}} (\epsilon_k - \epsilon_k^{(n)}) \sum_{k'} a_{ik'} \\ &+ \frac{\partial}{\partial\theta'_k} \sum_i c_i^{\theta\delta^{(n)}} (\delta_k - \delta_k^{(n)}) \sum_{k'} a_{ik'} \\ &= \sum_i c_i^{\theta\theta'^{(n)}} a_{ik} \sum_{k'} a_{ik'} . \end{aligned} \quad (5.25)$$

This quadratic surrogate function can be exactly minimized by a Newton step, which leads to the following image update:

$$x_{n+1} = x_n - C^{-1} \nabla Q , \quad (5.26)$$

where C is the Hessian matrix of Q_{SPS} .

To calculate the inverse of C one can use the fact that the inverse of a block diagonal matrix is simply the block matrix consisting of the inverted blocks of the original matrix. As the curvature matrix C is a block matrix consisting of n , where n is the number of voxels, 3×3 blocks one now simply needs to invert n 3×3 matrices and, for each voxel j , multiply the corresponding inverted 3×3 matrix C_j onto the gradient vector in this voxel,

$$\begin{pmatrix} \mu_j \\ \delta_j \\ \epsilon_j \end{pmatrix}^{(n+1)} = \begin{pmatrix} \mu_j \\ \delta_j \\ \epsilon_j \end{pmatrix}^{(n)} - C_j^{-1} \begin{pmatrix} \frac{\partial Q_{SPS}^n}{\partial \mu_j} \\ \frac{\partial Q_{SPS}^n}{\partial \delta_j} \\ \frac{\partial Q_{SPS}^n}{\partial \epsilon_j} \end{pmatrix}. \quad (5.27)$$

5.3.2 Choice of curvature

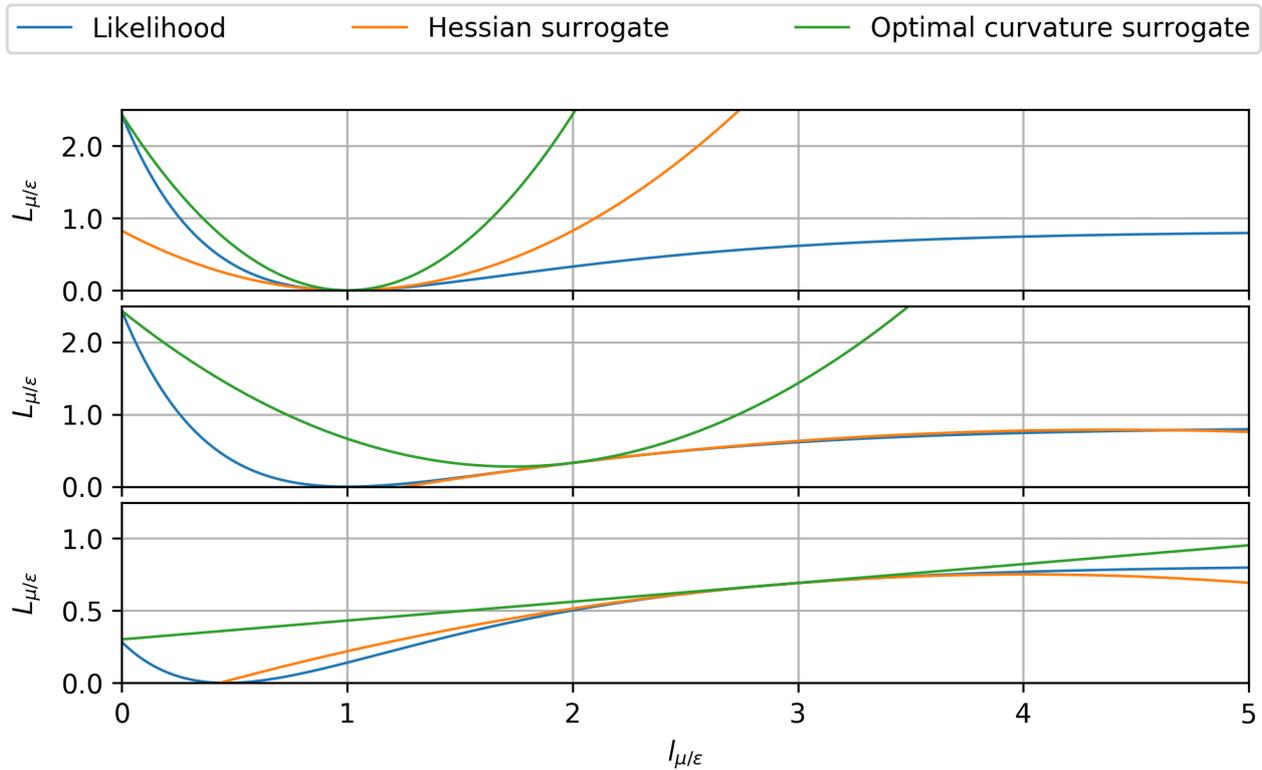


Figure 5.1: The blue lines show three different realizations of h_i^μ or h_i^ϵ . In green optimal quadratic surrogates are given and in orange surrogates with curvatures equal to the second derivatives are given.

To utilize the update step derived in the last section, it is necessary to choose a viable curvature

for the surrogate functions. As there is no general way of constructing a sharp curvature for a multivariate quadratic surrogate function of a non-convex function [52], two pseudo-curvatures will be proposed in the following: One based on the sharp curvature calculatable under the assumption that updates to $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, $\boldsymbol{\epsilon}$ do not directly influence each other in one step and the other based on the Hessian matrix in each voxel. In addition, a pre-computed curvature based on the curvature derived from the Hessian matrix will be presented.

5.3.2.1 Optimal diagonal curvature

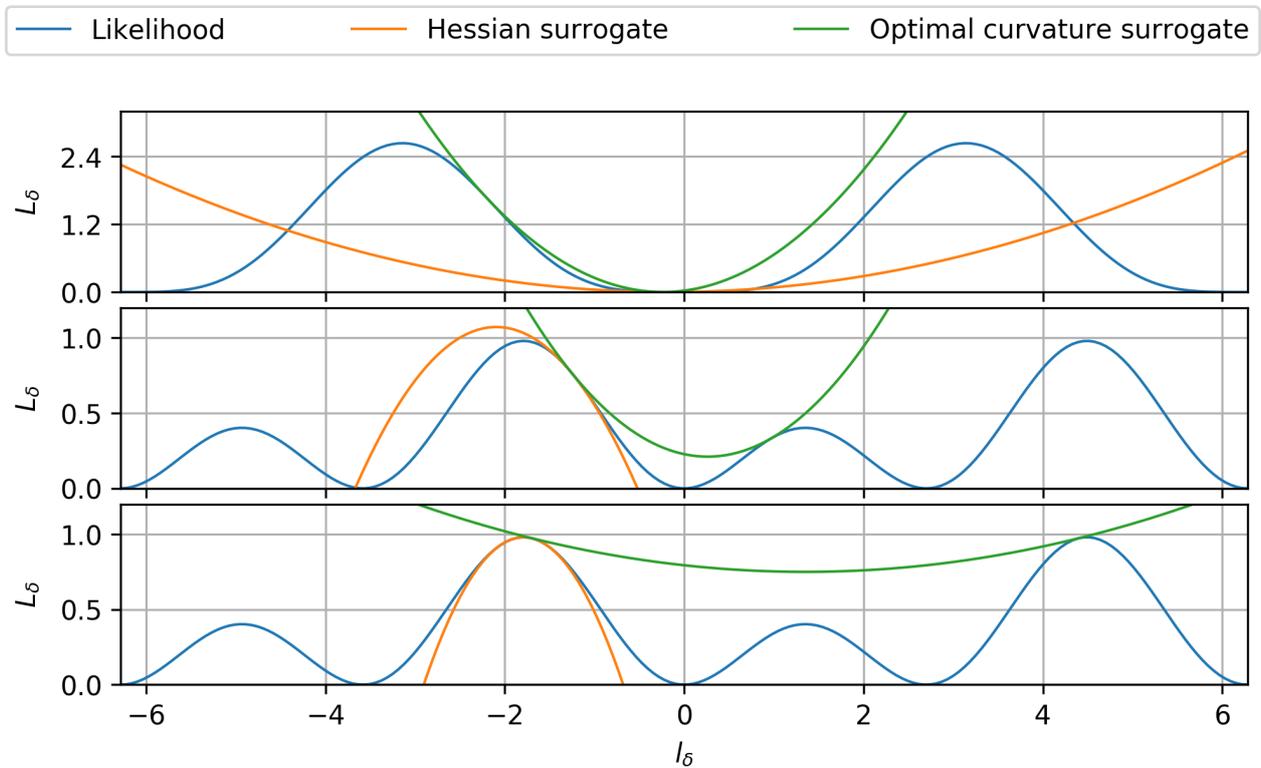


Figure 5.2: The blue lines show three different realizations of h_i^δ . In green the optimal quadratic surrogates are given and in orange surrogates with curvatures equal to the second derivatives are given.

For a 3D function, like $h_i(\mathbf{l}_i)$, it is hard to design an optimal curvature, once because there is no general way to do it and once because even deciding what optimal means is not even clearly defined. But if the coupling between the dimensions is not too strong it might be viable to reduce the 3D problem into three separate 1D problems where for each dimension the smallest possible curvature is clearly the best choice.

To design optimal curvatures for the separate 1D problems, it helps to rewrite the function $h_i(l_i^\mu, l_i^\epsilon, l_i^{\delta^D})$ into three simplified functions $h_i^\mu(l_i^\mu)$, $h_i^\epsilon(l_i^\epsilon)$, and $h_i^\delta(l_i^\delta)$

$$h_i^\mu = (\exp(-l_i^\mu) - y_i^\mu)^2 w_i^\mu \quad (5.28)$$

$$h_i^\epsilon = (\exp(-l_i^\epsilon) - y_i^\epsilon)^2 w_i^\epsilon \quad (5.29)$$

$$h_i^\delta = (\cos(\Phi_i^0 + l_i^{\delta^D}) - y_i^\delta)^2 w_i^\delta \quad (5.30)$$

with

$$y_i^\mu = y_i / (I_i^0 (1 + V_i^0 \exp(-l_i^\epsilon) \cos(\Phi_i^0 + l_i^{\delta^D}))) \quad (5.31)$$

$$w_i^\mu = w_i (I_i^0 (1 + V_i^0 \exp(-l_i^\epsilon) \cos(\Phi_i^0 + l_i^{\delta^D})))^2 \quad (5.32)$$

$$y_i^\epsilon = (y_i - I_i^0 \exp(-l_i^\mu)) / (I_i^0 \exp(-l_i^\mu) V_i^0 \cos(\Phi_i^0 + l_i^{\delta^D})) \quad (5.33)$$

$$w_i^\epsilon = w_i (I_i^0 \exp(-l_i^\mu) V_i^0 \cos(\Phi_i^0 + l_i^{\delta^D}))^2 \quad (5.34)$$

$$y_i^\delta = (y_i - I_i^0 \exp(-l_i^\mu)) / (I_i^0 \exp(-l_i^\mu) V_i^0 \exp(-l_i^\epsilon)) \quad (5.35)$$

$$w_i^\delta = w_i (I_i^0 \exp(-l_i^\mu) V_i^0 \exp(-l_i^\epsilon))^2 \quad (5.36)$$

The shape of h_i^μ and h_i^ϵ is now equivalent to the per pixel cost function of the non-linearized CT reconstruction problem. The optimal curvature for this problem with $l_i^\mu \geq 0$, respectively $l_i^\epsilon \geq 0$, has been published in [35]:

$$c_i^\mu(l^\mu) = \begin{cases} \left[2 \frac{h_i^\mu(0) - h_i^\mu(l_i^\mu) + \dot{h}_i^\mu(l_i^\mu) l_i^\mu}{l_i^{\mu 2}} \right]_+ & \text{if } l_i^\mu \geq 0 \\ \left[\ddot{h}_i^\mu(0) \right]_+ & \text{otherwise.} \end{cases} \quad (5.37)$$

It uses the knowledge that l_i^μ and l_i^ϵ are bound to be positive and the second derivative of h_i is bounded on the interval $[0, \infty]$ and monotonically decreasing while $\ddot{h}_i \geq 0$. Several examples of the shape of h_i^μ or h_i^ϵ and the paraboloid surrogate function resulting from the optimal curvature for different values of y_i^μ , w_i^μ , and l_i^μ respectively y_i^ϵ , w_i^ϵ , and l_i^ϵ are given in Fig. 5.1.

Finding the optimal curvature for h_i^δ is much harder as the space of viable values of $l_i^{\delta^D}$ is not limited and the second derivative is non-monotonic. If one cannot find an analytic way to

calculate the optimal curvature, it is always possible to calculate the optimal curvature by brute force: Simply start with $c_i^\delta = 0$, numerically check if $q_i^\delta(l_i^{\delta D}) \geq h_i^\delta(l_i^{\delta D})$ for all viable values of $l_i^{\delta D}$ and increase c_i^δ by a small amount until the domination condition is not violated anymore. Of course, this process is computationally expensive and this cost prohibits calculating the optimal curvature individually for each pixel during reconstruction. But c_i^δ can be treated as a function of y_i^δ and Φ_i^0 and subsequently be approximated by interpolating between values previously calculated and saved into a lookup table before the reconstruction. Calculating such a lookup table is simple at a first glance but as y_i^δ and Φ_i^0 can both take all values from $-\infty$ to ∞ it would require careful consideration where the points evaluated for the lookup table should be placed. Fortunately, the function $c_i^\delta(y_i^\delta, \Phi_i^0)$ is 2π -periodic in Φ_i^0 (as h_i^δ is 2π -periodic in Φ_i^0), thus only input values between $-\pi$ and π must be considered. Through careful analysis of different possible input value ranges an analytic formula for $|y_i^\delta| \geq 4$ can be found:

$$c_i^\delta(l_i^\delta) = \begin{cases} \frac{(y_i^\delta - \cos \Phi_i^0) \sin \Phi_i^0}{\Phi_i^0} & \text{if } l_i^\delta \geq 4 \wedge \Phi_i^0 \neq 0 \\ y_i^\delta - 1 & \text{if } l_i^\delta \geq 4 \wedge \Phi_i^0 = 0 \\ \frac{(-y_i^\delta - \cos(\pi - \Phi_i^0)) \sin(\pi - \Phi_i^0)}{\pi - \Phi_i^0} & \text{if } l_i^\delta \leq -4 \wedge 0 \leq \Phi_i^0 \leq \pi \\ -y_i^\delta - 1 & \text{if } l_i^\delta \leq -4 \wedge \Phi_i^0 = \pi \\ \frac{(-y_i^\delta - \cos(\pi - \Phi_i^0)) \sin(\pi - \Phi_i^0)}{-\pi - \Phi_i^0} & \text{if } l_i^\delta \leq -4 \wedge -\pi \leq \Phi_i^0 \leq 0 \\ -y_i^\delta - 1 & \text{if } l_i^\delta \leq -4 \wedge \Phi_i^0 = -\pi \\ c_i^\delta(l_i^\delta) & \text{else.} \end{cases} \quad (5.38)$$

These analytic solutions remove the need to calculate lookup table values for data points with $|y_i^\delta| \geq 4$. With these two restrictions on the range of the lookup table it is only necessary to calculate and store points in the ranges $y_i^\delta \in [-4, 4]$ and $\Phi_i^0 \in [-\pi, \pi]$. A visual inspection of the values of c_i^δ calculated by brute force, see Fig. 5.3, reveals that the function is smoothly varying with y_i^δ and Φ_i^0 and thus an approximation with interpolation between points saved in a lookup table is a viable approximation.

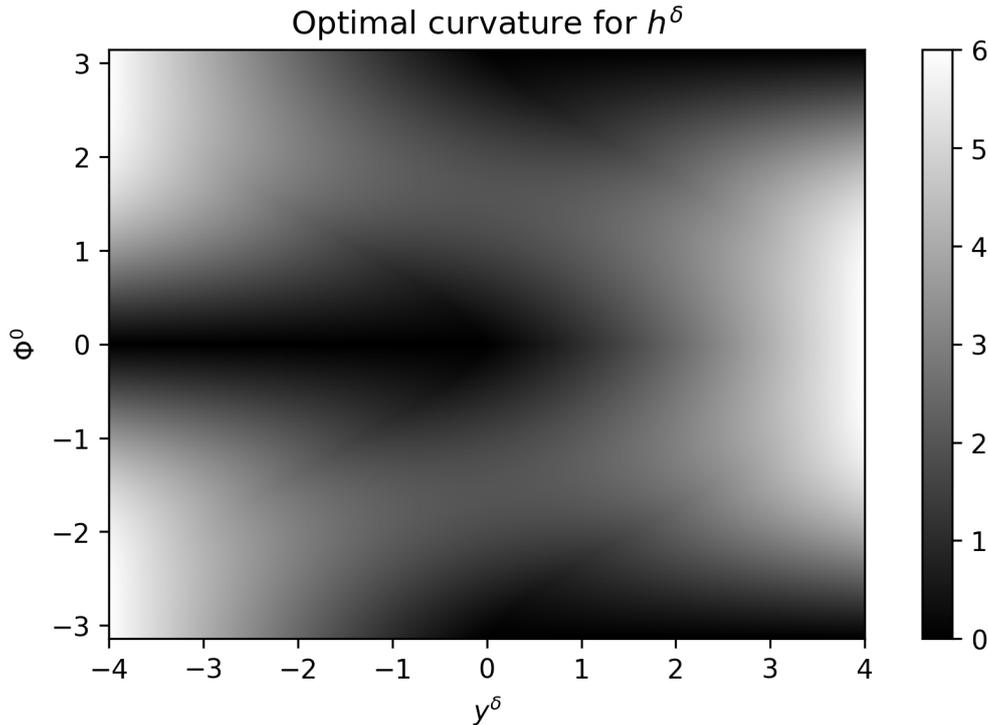


Figure 5.3: Visualization of the lookup table for the optimal curvature of h_i^δ calculated in the value range not accessible with the presented analytic solutions.

5.3.2.2 Hessian curvature

Another starting point in designing a curvature is to look at the Hessian matrix of the problem at hand. For a quadratic problem the Hessian would be constant and equal to the ideal curvature and would consider dependencies between updates of the different variables. In the case of non-quadratic problems using the Hessian as curvature will not result in a proper surrogate function, as there is no guarantee that the surrogate function will lie on top of the original cost function for all estimates, see Fig. 5.1 and Fig. 5.2. But if the current estimate of the reconstruction is already near the minimum and if the original cost function is nearly quadratic near its minimum the use of the Hessian as curvature should result in a quick optimization (but without any convergence guarantee) [74]. The equations necessary to calculate the Hessian matrix for the IBSIR problem can be found in App. B.1. Unfortunately, calculating and inverting the full Hessian in each voxel at every iteration is quite computationally expensive (six backprojections are necessary to calculate the entries of the hessian matrix and for each voxel a 3x3 matrix has to be inverted). It would be very beneficial if we could find a curvature that does not need

recalculation at every iteration but could be precalculated once before the reconstruction.

5.3.2.3 Precalculated Hessian curvature

If one takes a look at the second derivatives of h_i , see App. B.1, two ways of simplification may come to mind: If the reconstruction is sufficiently near to the minimum of the cost function the residual at each pixel becomes small and the mean of the residuals approaches zero, as in convergence the residuals are Gaussian distributed around zero. This means the backprojection (which acts as a strong low-pass filter) of terms containing the residual of the pixels is approximately zero for all voxels. Another simplification we can use is that the phase p_i is approximately equally distributed across all projections and pixels. Actually, this is a necessary condition for an artifact free reconstruction of a GBCT dataset, see Sec. 5.4.1. This means the backprojection (which acts as a low-pass filter) of terms containing $\sin(p_i)$, $\cos(p_i)$ or $\sin(p_i)\cos(p_i)$ is zero and terms containing $\sin^2(p_i)$ or $\cos^2(p_i)$ become equal to 0.5. This leads to very simple approximations of the entries, which can be found in App. B.1. Here only I^A needs to be estimated in some way: If the first estimate of μ and ϵ is good the value of I^A before the first iteration is a good estimate or, if no good estimate of μ and ϵ is available, I^0V^0 will be a viable estimate, which will result in a looser curvature.

This curvature does not require recalculation in every iteration and using it could result in a speed-up compared to the curvature based on the Hessian if the optimization is already near its minimum.

5.3.3 Ordered subsets

Calculating the gradient is one of the most expensive parts in the reconstruction algorithm and any possibility to speed up this part of the algorithm will greatly increase the overall speed of the optimization. One method to achieve this has been proposed: The ordered subsets method. This method has been used extensively in CT reconstruction in conjunction with the SPS algorithm under the name OS-SPS [35].

Taking subsets increases the speed of calculating the gradient of the likelihood function ∇L by not calculating the gradient using the full available information from all projections but by using only the information of a $\frac{1}{M}$ -fraction of all projections. The gradient calculated from this sub-sampled dataset ∇L_m will still approximate the gradient of the full dataset quite well (after scaling with the factor M) but is roughly M -times faster to calculate than the full gradient. For the best performance the subsets should be ordered, meaning that the k 'th subset contains the projections $k, k + M, k + 2M$, and so on, and each subsequent sub-iteration should use a subset that is as orthogonal as possible to the previous subset.

This principle can be used for the IBSIR optimization problem in the same way as it has been used for attenuation CT, except that care should be taken in selecting a viable number of subsets that still ensures a good sampling of the enhanced Radon space in each subset. If, for example, a sliding window phase stepping scan with five different equidistant grating positions is sub-sampled into five subsets, each subset will only contain projections of the same grating position, resulting in an extremely uneven sampling of the enhanced Radon space in each subset and possibly resulting in very uneven and unstable image updates.

It should be noted that the approximation

$$\nabla L = M \nabla L_m \quad (5.39)$$

becomes less viable with increasing convergence because the overall structure of the object becomes more and more fixed and only subtle details remain to be reconstructed and the information about these details is not evenly distributed across all projections. It is thus recommended to use a large number of subsets for a strong initial acceleration and decrease the number of subsets in subsequent iterations.

5.3.4 Fourier space preconditioning

As discussed in the beginning of this section, the optimization problem at hand is not only badly conditioned in the image-domain and across the channels, which are both addressed by the choice of curvature, but also in the frequency domain. This can be well observed by look-

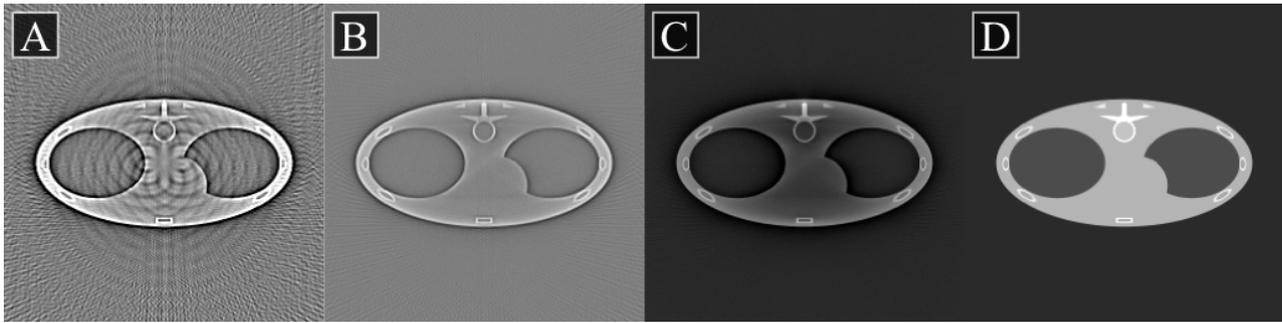


Figure 5.4: Intermediate results of the reconstruction of δ using IBSIR with a gradient descent solver after 16 (A), 128 (B), and 1000 (C) iterations. The ground truth is shown in (D). The color windows are $[-0.0003, 0.0003]$ (A), $[-0.01, 0.01]$ (B), and $[-0.01, 0.05]$ (C and D).

ing at intermediate results of a reconstruction of the refractive index using gradient descent, see Fig. 5.4. In the first iteration only high frequency features are visible and it takes even a considerable time for mid- frequency features to appear. Low frequency features like in the central area of the phantom are even not visible after many hundreds of iterations. For the other channels the frequency convergence is comparable to the frequency behavior observed for simple attenuation CT reconstruction algorithms [72]: Low frequency image components are favored over high frequency features.

A possible way to address this challenge is to calculate the PSF caused by the forward- and backward projection in the reconstruction algorithm and use its inverse as a Fourier space filter on the image updates. A rough spatially invariant approximation of the PSF can be calculated by simply forward and backward projecting a δ -peak function placed in the center of the image space [72]. The image updates generated by filtering the gradient with this filter may reduce the cost function slightly slower than using the unfiltered image update, but the low frequency components will improve much faster leading to a far faster convergence of the image impression.

5.3.5 Momentum acceleration

Momentum techniques have been widely used in CT reconstruction to speed up the convergence of first-order optimization algorithms, like the OS-SPS algorithm [54, 37], and there is no reason not to use them in the case of the IBSIR cost function. Momentum techniques use information

about previous descent directions to increase the speed of convergence. It was shown [75] that such an accelerated gradient method has a rate of convergence of the order $\mathcal{O}(1/n^2)$ as compared to conventional gradient methods with a rate of convergence of $\mathcal{O}(1/n)$. A possible choice of momentum acceleration is based on the work of Y. Nesterov [75] and an algorithm for use in CT reconstruction based on this choice has been proposed [54]:

$$\begin{aligned} t^{(n+1)} &= (1 + \sqrt{1 + 4t^{(n)2}})/2 \\ \boldsymbol{\theta}^{(n+1)} &= (\mathbf{m}^{(n)} - D^{-1}\nabla C) \\ \mathbf{m}^{(n+1)} &= \boldsymbol{\theta}^{(n+1)} + \frac{t^{(n)} - 1}{t^{(n+1)}}(\boldsymbol{\theta}^{(n+1)} - \boldsymbol{\theta}^{(n)}) \end{aligned} \tag{5.40}$$

with $\mathbf{m}^{(0)} = \boldsymbol{\theta}^{(0)}$ and $t^{(0)} = 1$.

If the initial step size is small, it will take the momentum term several iterations to approach its optimal value. This slow initial acceleration is well suited for combination with the ordered subset acceleration that provides a strong acceleration during the first iterations.

5.3.6 Simulations

The choice of the proposed acceleration techniques has a tremendous effect on the convergence of the reconstruction algorithm. The effect of the different options was compared by performing various reconstructions of the simulated axial scan of a phantom. The scan consisted of 1203 equidistantly spaced projections taken in parallel beam geometry. The simulated detector is a photon-counting detector with perfect quantum efficiency consisting of 16 lines with 256 pixels in each line. The simulated interferometer has a visibility of 0.4 and 10^5 monochromatic photons per pixel were emitted from the X-ray source. The Moiré fringes visible on the detector were simulated to have a period of 10.1 pixels in both detector dimensions. After acquiring a projection the grating position was moved by $\frac{1}{2.618}$ of the grating pitch. The attenuation and linear diffusion coefficients of the phantom were chosen such that the lowest transmission and dark-field signal was around 0.4. The refractive index was chosen such that the highest differential phase contrast was about 0.6. All reconstructions were started from an empty initial guess and Huber regularization with a low regularization strength was used for each channel.

5.3.6.1 Choice of curvature

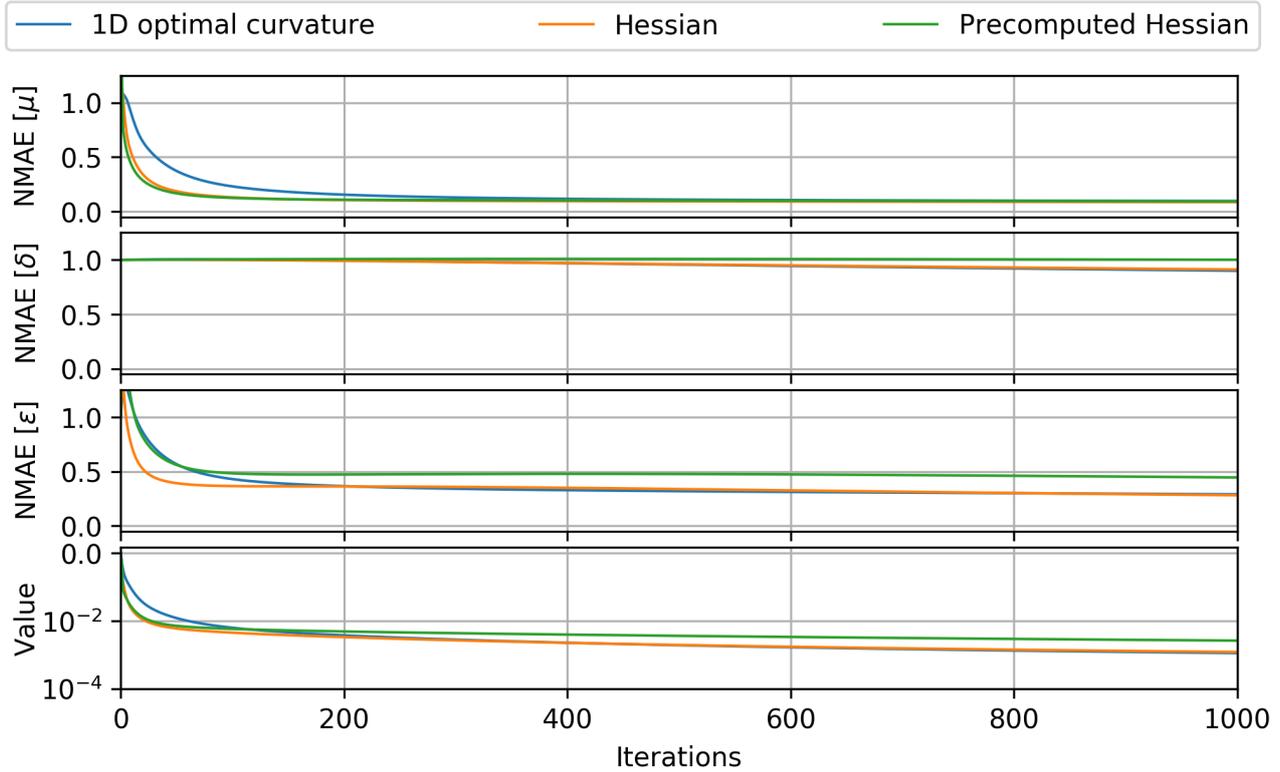


Figure 5.5: Change of the cost function value and the normalized mean absolute error (NMAE) in each image channel with the number of iterations for three different choices of curvature for the surrogate function. No momentum acceleration, subset acceleration, or frequency space preconditioning was used.

First a reconstruction without any acceleration was performed and the three different proposed methods to calculate a step size are compared. The pre-computed curvature based on the Hessian C^{prec} was only calculated once at the beginning of the reconstruction and the 1D optimal curvature C^{opt} as well as the curvature based on the Hessian C^{hess} were recomputed after each iteration.

As C^{opt} is always larger than the corresponding entry of C^{hess} , it is to be expected that the initial convergence of the reconstruction is slower using C^{opt} than using C^{hess} . In Fig. 5.5 one can see that this is true in the simulation as the normalized mean absolute error (NMAE) in all three channels and the value of the cost function initially decreases quicker for C^{hess} but after several hundreds of iterations the distance becomes marginal. The reconstruction using the pre-computed C^{prec} is decreasing the NMAE in the attenuation channel as quick as C^{hess} per iteration but does not work as well as the other choices for the reconstruction of ϵ and δ .

Designing a pre-computed curvature that works well for the reconstruction of ϵ and δ would be highly beneficial because an image update using a pre-computed curvature requires only half as many forward- and backprojections and thus has the potential to double the reconstruction speed.

A quick solution to ease the problem of calculating the curvature newly in each iteration is to update the curvature only every few iterations. This will result in a slightly non-optimal curvature at most iterations but greatly reduces the computational cost. This results only in a minor decrease in convergence per iteration but decreases the number of forward- and backprojections by a factor of $\frac{1+n}{2n}$, where n is the number of iterations after which the curvature is recalculated.

5.3.6.2 Fourier space preconditioning of δ

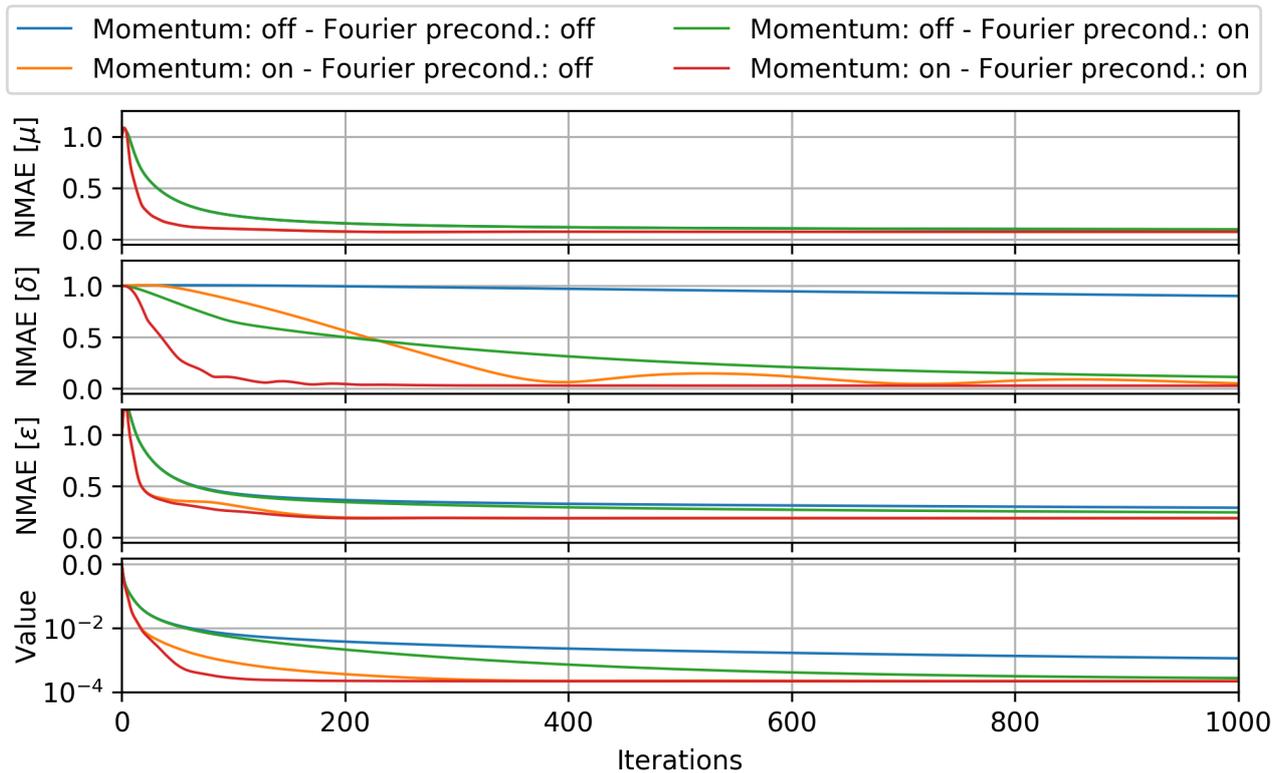


Figure 5.6: Change of the cost function value and the normalized mean absolute error (NMAE) in each image channel with the number of iterations for reconstructions using different combinations of the Nesterov momentum acceleration and/or the frequency space preconditioning of the refractive index channel. A single subset was used for all reconstructions.

As could be seen in the convergence of the un-accelerated reconstructions in Fig. 5.5, the con-

vergence of the MAE of the refraction index channel is extremely slow, regardless of the choice of curvature, albeit the value of the cost function is readily decreasing. This is caused by the fact that only the highest frequencies of δ are correctly reconstructed, these frequencies are important for the cost function, but the overall image of the refractive index reconstruction does not converge in a reasonable time. It can be seen in Fig. 5.6 that filtering the update of δ with the inverse of the Fourier transform of the PSF caused by differential forward- and backprojection greatly reduces this focus of the optimization on the high frequencies and provides a consistently decreasing NMAE of δ . The convergence of the other channels is largely unchanged by this change to the updates of δ .

5.3.6.3 Momentum acceleration

As expected, the momentum acceleration provides a significant overall boost in convergence speed. The cost function and the NMAE in each channel drop much quicker than for the unaccelerated reconstruction. Only for the refractive index channel the momentum factor needs many iterations to pick up speed and accelerate the reconstruction. But the acceleration in this channel seems to show a ringing behavior and the MAE increases again. This is caused by the fact that, if no Fourier preconditioning is used on δ , still only the very high frequencies are reconstructed and even the momentum acceleration cannot efficiently converge the low frequencies. By far the best results are obtained when Fourier space preconditioning and momentum acceleration are combined, even quickly converging the refractive index channel.

5.3.6.4 Subset acceleration

Last but not least, the effect of subset acceleration on the convergence of the preconditioned and momentum accelerated reconstruction algorithm was studied. In Fig. 5.7 it can be seen that the reconstructions using different numbers of subsets show very similar convergence speed per iteration as a reconstruction using only a single subset. Only for large numbers of subsets the convergence speed per iteration starts to slightly decrease and for even higher numbers of subsets the reconstructions diverge. The results are not shown here, but without any acceleration the

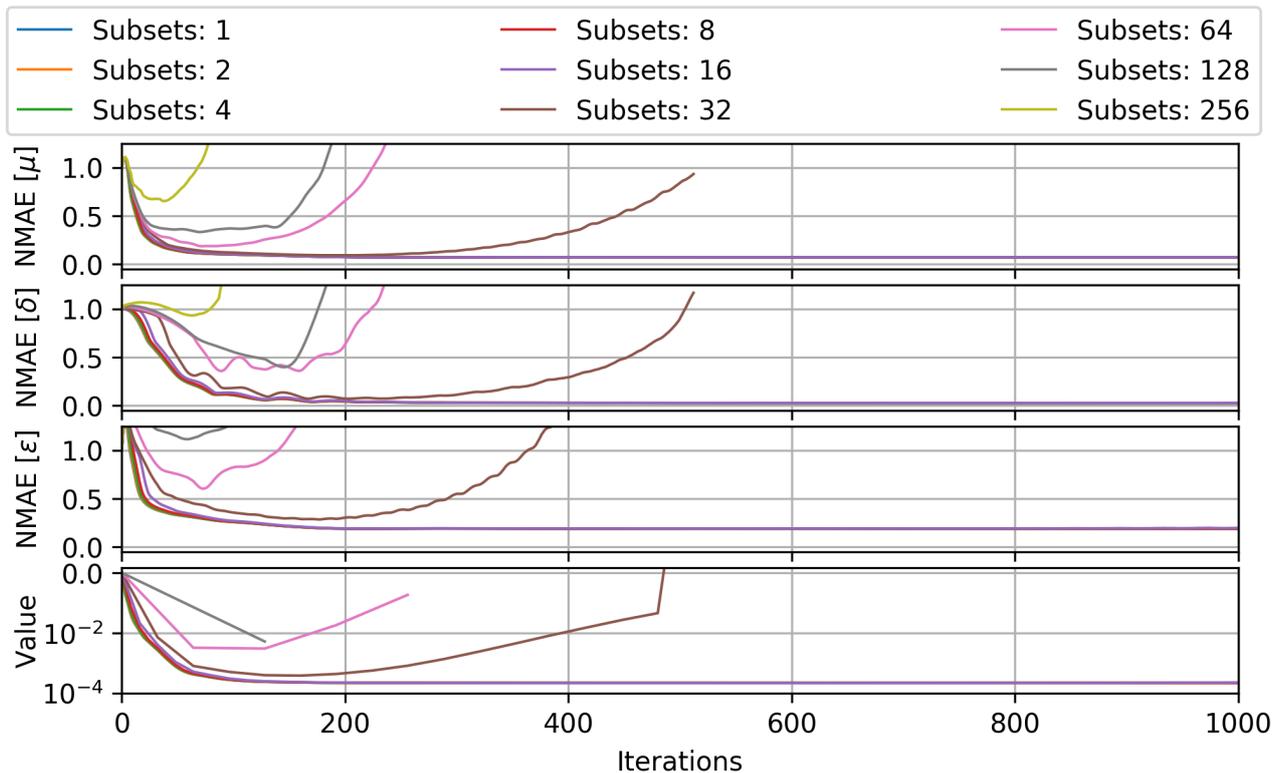


Figure 5.7: Change of the cost function value and the normalized mean absolute error (NMAE) in each image channel with the number of iterations for reconstructions using different numbers of subsets. All reconstructions use Nesterov momentum acceleration and frequency space preconditioning of the refractive index channel.

number of subsets can be increased far higher before divergence starts to occur. But the increase in convergence speed offered by frequency space preconditioning and momentum acceleration clearly out-weights the acceleration of using an extreme amount of subsets.

5.4 Single shot acquisition

One of the main challenges for the successful translation of GBCT from laboratory setups to a clinical full body CT scanner is the continuous rotation and signal acquisition of a modern CT scanner. The conventionally used phase-stepping procedure, see Ch. 3, is not usable in this scenario. Several solutions have been proposed to overcome this challenge, but those solutions all come with shortcomings. For example, the reverse projection method [76] combines two interferograms taken exactly 180 degrees apart. This method poses very strict requirements on the stability of the grating alignment and gives only access to attenuation and differential

phase-contrast projection images. Additionally, it is unclear if this technique can be adapted to systems with significant cone-beam geometry and helical acquisition protocols. The Moiré analysis method [77, 13] is a true single-shot technique that uses Fourier analysis to detect subtle changes caused by the measured object on a high-frequency Moiré fringe pattern created from slight detuning of the interferometer, see Sec. 2.2.3. This approach does not require grating movements but suffers from a significant loss in resolution in one dimension of the detector. A different approach, which is a direct successor of traditional phase-stepping, is the sliding window method [61], see Ch. 4. This technique requires only a single interferogram per projection angle but changes the grating position at each angular position. To mimic a classical phase-stepping scan, the missing interferograms are simply interpolated from adjacent projection angles. This works well for scans with high angular sampling but if the number of projection angles is reduced, the interpolation step causes blurring and artifacts. In addition to techniques that require only a single interferogram per projection angle, a solution that overcomes the need for precise mechanical grating movements (but still requires multiple measurements per angular position) has been presented: Electromagnetic phase-stepping (EPS) [78] employs electrical beam steering to move the object's projection slightly across the detector while a deliberately created Moiré fringe pattern is fixed on the detector. Several of these shifted images can be combined to form a pseudo phase-stepping. All these solutions, including traditional phase-stepping, have in common that they calculate sets of attenuation, differential phase-contrast and dark-field projections for each projection angle from the interferograms.

Using the IBSIR reconstruction method allows to circumvent this intermediate signal retrieval step and should allow for high quality, artifact free reconstructions from just single measurements per projection angle. In the following, considerations about the necessary sampling required for artifact-free reconstructions are made and then simulations and measurements are presented for several successful single-shot imaging scenarios.

Many of the following results have already been published by the author in [68]

5.4.1 Sampling

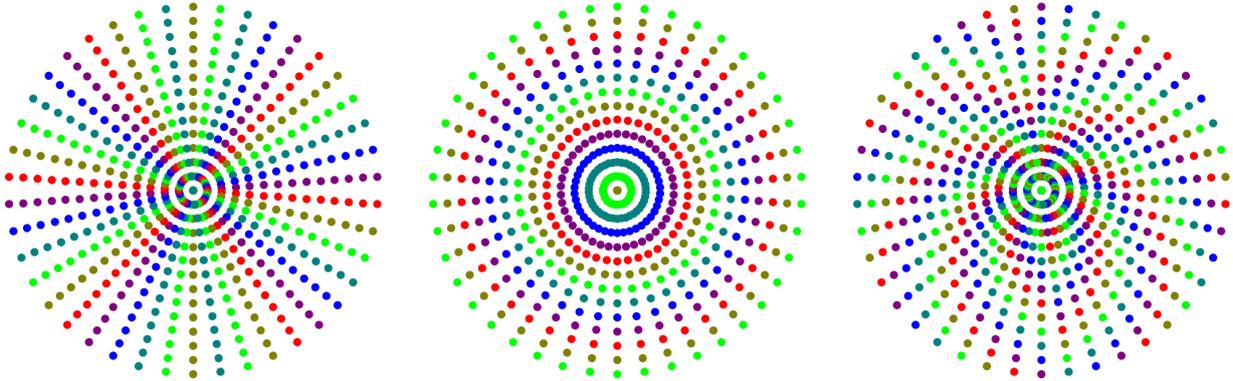


Figure 5.8: Sampling of expanded Radon space for three different acquisition protocols. The left image shows the sampling for a sliding window scan. The central image represents a scan without grating movements and high frequency Moiré fringes on the detector and the right image visualizes the sampling of a scan with high frequency Moiré fringes and SSEPS acquisition. The color represents the phase of the measurement on the phase stepping curve.

The formulation of the GBCT reconstruction problem in the previous section provides an extremely versatile reconstruction method that will produce results even for measurement data, that is acquired in patterns very different to conventional phase-stepping GBCT. The ability of the IBSIR algorithm to calculate a result for all possible input information does, of course, not guarantee that the result is similar to the true object. There must be requirements on the information content necessary for an artifact-free reconstruction. In conventional CT one would analyze this information content by looking at the positions of the measurements in the space of possible measurements, the Radon space [28]: Every single measured pixel of a conventional CT scan represents information at one point of the Radon transform of the measured object and the task of a CT reconstruction algorithm is to reconstruct the measured object by somehow calculating the inverse Radon transform using this information. In the case of GBCT, the Radon space representation of the measurements is not sufficient as each measured pixel is not only taken at a specific angular position and a specific distance from the center of the detector, but also at a specific phase position of the interferometer stepping curve. This expands the measurement space by an additional phase dimension in which sampling can be influenced by various effects: by shifting the gratings, by the presence and position of Moiré fringes or by the phase shift induced by the object itself. The task of GBCT reconstruction can

now be understood as the inversion of this expanded radon space (consisting of the conventional spatial dimensions and the additional phase dimension) back to the object space (consisting of the spatial distributions of the linear attenuation coefficient, the refractive index, and the linear diffusion coefficient). All previous GBCT reconstruction algorithms solve this problem by first collapsing the phase dimension of the expanded Radon space into three separate conventional Radon spaces, then inverting them separately. For a conventional phase stepping scan this is easily done by the signal estimation step, see Sec. 3.1, as the whole phase dimension is sufficiently well sampled (with at least three points along the phase dimension) at each point in the spatial dimensions. Single shot methods can be understood as a collapse onto coarser grids or as interpolation of missing values in the expanded Radon space. For example, the Fourier analysis method [77, 13] collapses onto a coarser grid along the detector pixel dimension of the Radon space and the sliding window interpolation method [61] uses nearest neighbor interpolation to fill missing values in the phase dimension with values taken from along the angular dimension. IBSIR on the other hand takes a very different approach because it directly inverts the expanded Radon space without first collapsing it. In situations without a complete phase stepping, this has the potential to be highly beneficial to the image quality of the results because collapsing the expanded Radon space will always lose information and thus necessarily reduce the possible accuracy of the results. From the theory of compressed sensing it is known that the highest probability to correctly recover a signal in an unknown under-determined system from noisy measurements can be achieved if the signal is sampled as non-uniform as possible. The distribution of the measurements along the conventional axes of the expanded Radon is given by the setup geometry, but the distribution along the phase axis can be influenced by the acquisition protocol. A way to visualize this expanded Radon space is to use the conventional Radon space and add color as an additional dimension to show measurement phase positions between 0 and 2π .

The expanded Radon space for three different acquisition protocols can be found in Fig 5.8. The first protocol shows the sampling in a scan with a sliding window grating movement. This scan is quite well sampled in the phase dimension in the central region of the spatial sampling space, but the sampling becomes worse in regions further out from the center when the spatial

sampling in angular direction is getting thinner. This will cause under-sampling like artifacts in regions far from the rotation axis for situations with a sparse angular sampling. The second figure shows the sampling in an acquisition with constant gratings but high frequency Moiré fringes across the detector. This provides a quickly changing phase in the radial direction of the sampling space but a constant phase along the angular direction. This results in a bad phase sampling in the center of the spatial region of the Radon space and will subsequently result in ring-like artifacts in the center of the reconstruction, see Fig. 5.9. The last expanded Radon space is representative for a scan with fixed grating positions, high frequency Moiré fringes and a sample shift of the object by one pixel per projection angle. This movement can be provided by changing the position of the focal spot (electromagnetic phase stepping) or by shifting the object. This pattern is diversely sampled along angular and radial direction and should result in artifact-free reconstructions. It should be noted that this is a very rough way to analyze the necessary sampling of a GBCT scan, but it allows for plausible explanations of artifact-prone and artifact-free acquisition protocols.

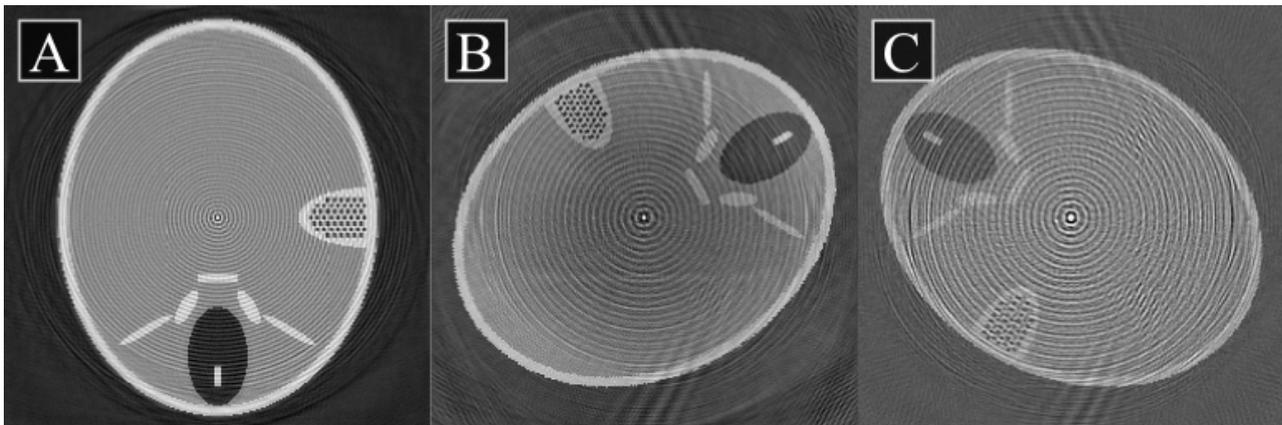


Figure 5.9: Reconstruction of attenuation coefficient (A), refractive index decrement (B), and linear diffusion coefficient (C) from a generic simulated scan with Moiré fringes but without phase stepping or sample movement.

5.4.2 Single-shot simulations

The performance of using IBSIR to reconstruct single-shot GBCT scans is first demonstrated on numerical simulations. The simulated phantom exhibits X-ray attenuation, refraction, and

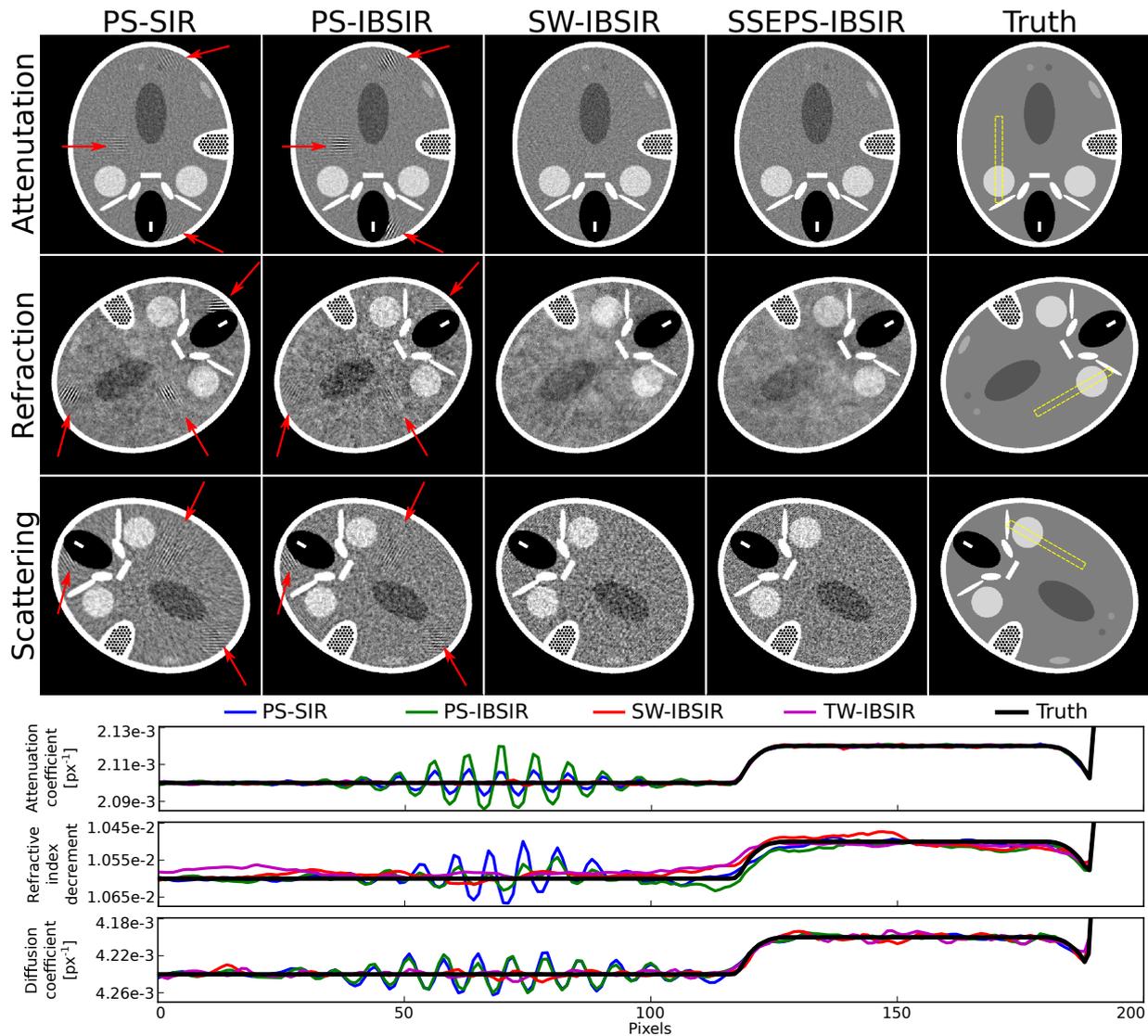


Figure 5.10: Comparison of different reconstruction methods for several acquisition patterns. The rows from top to bottom show reconstruction of a phantom’s attenuation, refractive index, and linear diffusion coefficient. (PS-SIR) shows the reconstruction of a phase-stepping scan composed of 101 angular positions with three phase-steps each. (PS-IBSIR) shows reconstructions of the same scan using direct reconstruction. Both reconstructions show distinct artifacts, marked by the arrows, caused by angular undersampling. (SW-IBSIR) uses the IBSIR algorithm to reconstruct a scan acquired with a sliding window acquisition pattern. (SSEPS-IBSIR) is the reconstruction of a single-shot electromagnetic phase-stepping acquisition pattern using a single fixed grating position. Both methods strongly reduce the undersampling artifacts. Below the images line plots along the yellow box for all channels are given. Figure adapted from [68].

ultra-small-angle scattering. To create this phantom the FORBILD head phantom [65] was extended to three sub-phantoms with a well-defined linear attenuation coefficient μ , refractive index decrements δ , and linear diffusion coefficient ϵ for each material property. The phantom contains high-contrast elements, which are small in size and whose reconstruction is prone to undersampling artifacts, and large low-contrast elements that are challenging to reconstruct

correctly at the same time. Rotating the sub-phantoms with respect to each other (see ground truth in Fig. 5.10) allowed us to better attribute which property of the phantom was causing image artifacts. This destroys any anthropomorphology of the phantom but, for example, if all elements were perfectly co-aligned, artifacts from photon starvation caused by high attenuation would be visible in all three volumes, but would not be unambiguously attributable to the attenuation property because at the same location the phantom also exhibits high ultra-small-angle scattering and refraction. Scans with three different acquisition patterns (phase stepping (PS), sliding window (SW), and single-shot electro-magnetic phase stepping (SSEPS)) were simulated. Each simulation used the same total exposure time and number of detector readouts. The PS scan used three different grating positions at each of 101 angular positions equally distributed over 360 degrees. The SW scan took measurements at each of 303 angular positions equally distributed over 360 degrees and after each measurement the grating position was shifted by one third of the grating period. For the SSEPS scan, 303 measurements equally distributed over 360 degrees were simulated and the focal spot position of the X-ray tube was shifted by one pixel after each measurement. The reference phase of the interferograms was set to mimic vertical Moiré fringes with a frequency of 0.05 px^{-1} for the PS and the SW scan and a frequency of 0.38 px^{-1} for the SSEPS scan. A uniform visibility of 0.75 was simulated. This unrealistically high visibility was chosen to achieve similar CNR levels in the attenuation and dark-field image. The geometry constants c_D and c_ϕ were set to one. A 500×5 pixel ideal photon-counting detector and parallel beam geometry were simulated. Afterwards Poisson noise corresponding to a total of 10^{13} emitted photons, evenly distributed over all angles and pixels, was added to the projections. The values of the material properties of the phantoms were re-scaled to achieve a similar CNR across all images. All simulations were performed on a $500 \times 500 \times 5$ voxel volume. The simulations were then reconstructed using several methods: The PS scan was processed into attenuation, differential phase and dark-field projections. Then the different sets of projections were reconstructed using FBP and SIR. Additionally, this scan was directly reconstructed using the IBSIR method. The PS-SIR and the PS-IBSIR reconstructions are shown in Fig. 5.10. The PS-FBP reconstruction produced vastly inferior results with strong undersampling artifacts (the image quality metrics can be found in

	NRMSE in %			NMAE in %			MSSIM		
	μ	δ	ϵ	μ	δ	ϵ	μ	δ	ϵ
PS-FBP	26.0	24.9	26.6	19.1	18.0	19.9	.556	.557	.558
PS-SIR	.411	.742	3.98	.251	.393	.793	.731	.687	.649
PS-IBSIR	.521	.632	3.62	.291	.400	.767	.729	.661	.642
SW-FBP	25.1	35.1	57.3	16.0	25.8	42.8	.543	.537	.472
SW-SIR	17.1	22.4	22.7	3.90	9.77	8.32	.609	.610	.518
SW-IBSIR	.632	.637	3.96	.272	.404	.889	.731	.679	.628
EPS-IBSIR	.372	.564	3.74	.239	.356	.830	.730	.700	.632

Table 5.1: Normalized root-mean-squared error (NRMSE), mean absolute error (NMAE), and mean structural similarity (MSSIM) were separately calculated for μ , δ , and ϵ . PS stands for the phase stepping simulation, SW for the sliding window simulation and SSEPS for the single-shot electromagnetic phase stepping simulation. Reconstructions were performed using filtered backprojection (FBP), statistical iterative reconstruction (SIR) and interferogram-based SIR (IBSIR). The SIR and IBSIR approaches perform better than the FBP method for all simulations, whereas superior behavior of IBSIR over SIR can only be claimed for the SW simulation. The SSEPS simulation, that could only be reconstructed using IBSIR, shows similar image quality as the IBSIR reconstruction of the SW scan. Table adapted from [68].

Table 5.1). PS-SIR and PS-IBSIR reveal very comparable reconstruction results. Both show slight but clearly visible undersampling artifacts originating at the high-frequency high-contrast areas of the phantom in all modalities. Similar image quality was expected as both scans are based on the same underlying information. In this specific case, adequate photon statistics and sufficient sampling of the stepping curve enabled the signal estimation step to correctly extract the underlying signals and correctly propagate the statistical weights for the SIR from the raw measurements. The SW scan was also reconstructed using FBP, SIR, and IBSIR. Before the reconstruction using FBP and SIR, a sliding window interpolation between neighboring views had to be performed. The interpolated data was then processed the same way as the PS scan. The SW-FBP and SW-SIR reconstructions show severe artifacts, which could be expected as interpolation between neighboring views is only viable if the information content in one pixel changes slowly between neighboring views. These reconstructions are not shown, but image quality metrics can be found in Table 5.1. The SW-IBSIR reconstruction, see Fig. 5.10, does not require the intermediate interpolation step, but works directly on the interferograms and thus does not suffer from interpolation artifacts. The increase in angular sampling, compared to PS-IBSIR, results in reconstructions that do not show angular under-sampling artifacts. On this note, very weak streak artifacts can be seen in the attenuation reconstruction that

seem to be originating from the location of the high-frequency high-contrast objects in the scattering volume. In this instance, the algorithm was not able to attribute the changes in measured intensity to the correct signal. The SSEPS simulation was only reconstructed using IBSIR as no other algorithm is able to reconstruct measurements acquired this way. The resulting SSEPS-IBSIR reconstructions show no artifacts and provide a significant improved image quality compared to the other results, see Table 5.1.

5.4.3 Single-shot measurements

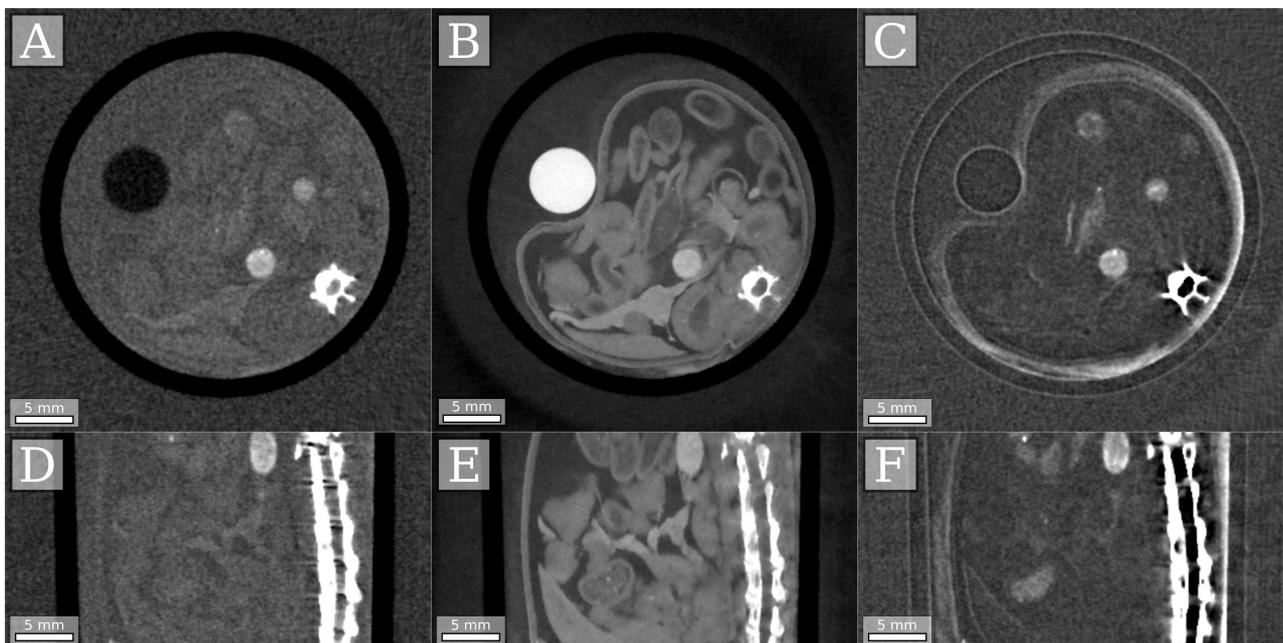


Figure 5.11: Axial (A-C) and sagittal (D-F) slices of the reconstructed attenuation coefficient (A, D), refractive index (B, E), and linear diffusion coefficient (C, F) of in-vitro mouse intestines scanned using the described single-shot electromagnetic phase stepping equivalent acquisition and the direct reconstruction algorithm. Figure adapted from [68].

As simulations can never reproduce all the effects present in a real measurement, experimental verification is necessary to fully illustrate the feasibility of single-shot GBCT without grating movement. Thus, an experiment was performed using a Talbot-Lau interferometer with a rotating anode X-ray tube as radiation source. Instead of shifting the focal spot position by applying a magnetic field to the X-ray tubes electron beam, we physically shifted the sample by a known distance after each measurement. This is readily possible at most X-ray microscopy setups and (for small opening angles) it produces virtually the same measurements as a focal

spot shift. For the experiment the intestines of a mouse fixated in formalin and placed in a falcon tube, were used as a sample, exhibiting a complex combination of attenuation, refraction, and scattering. To avoid extreme refraction at the falcon tube-air interface the whole sample was placed in a water bath. For this scan 900 measurements were taken, equally distributed over 360 degrees, and the sample stage was cyclically stepped between five different positions, resulting in object shifts of 0, 4, 8, 2 and 6 pixels in the direction perpendicular to the rotational axis. The average period of the Moiré fringes in this direction was about 10 pixels. The setup used was a symmetric Talbot-Lau interferometer [79] with grating distances $G_0G_1 = G_1G_2 = 85.7$ cm. The three gratings G_0 , G_1 , and G_2 are all made of gold on a silicon substrate with periods of $5.4 \mu\text{m}$. The absorption gratings G_0 and G_2 have a height of $70 \mu\text{m}$ and $65 \mu\text{m}$ respectively. The phase grating G_1 is designed to give a phase shift of π at the system design energy of 27 keV . The source was an ENRAF Nonius FR 591 rotating anode X-ray tube with a Molybdenum target operated at 40 kVp and 70 mA . The detector was a photon-counting PILATUS 100k with 487×195 pixels. It uses an 1 mm thick silicon sensor with a quadratic pixel size of $172 \times 172 \mu\text{m}^2$. The effective pixel size is $93 \times 93 \mu\text{m}^2$. The source to detector distance was 2.56 m , resulting in a fan angle of 3.7° . The exposure time for every single interferogram was 6 s . The mean visibility of the interferometer was at 0.27 and the mean frequency of the Moiré fringes parallel to the rotational axis was 0.1 px^{-1} .

The result of the IBSIR reconstruction of this scan is shown in Fig. 5.11. It shows no image artifacts and excellent image quality.

5.5 Internal parameter estimation during reconstruction

The second huge challenge of moving GBCT from highly controlled laboratory setups to a large continuously rotating CT gantry in an only mildly controlled environment is the presence of vibrations and drifts in such a setup.

Estimating the transmission, dark-field signal and differential phase shift caused by an object in the beam path requires very precise knowledge about the characteristics of the interferometer (or more precisely about the mean intensity, visibility and phase of all relevant stepping

curves) without any objects present in the beam path. Of course, the same is true for the direct estimation of the attenuation coefficient, the refractive index decrement, and the linear diffusion coefficient using IBSIR.

The required system characterization is usually achieved by performing an as identical as possible reference scan before or after the sample scan. Unfortunately, the characteristics of every real setup will change over time due to mechanical vibrations, thermal drifts, or grating re-positioning inaccuracies [80]. Thus, there is a difference between the characteristics measured during the reference scan and the real characteristics of the setup during the sample scan.

The effect of these drifts can greatly reduce the image quality of the estimated projections or reconstructions. Consequently, various methods for the correction of these drifts have been proposed for the projection-wise signal estimation problem [80, 81, 82, 83] but they are not usable for the direct reconstruction method using IBSIR as the corrections are applied in the projection domain after the signal estimation, which does not exist for the IBSIR method.

In the following three different variants of a possible solution for this problem will be presented: The vibrations and drifts are incorporated into the forward model used for the reconstruction and the amplitude of the vibrations and drifts is estimated from the measurements simultaneously to the reconstruction. The first variant is only concerned with constant phase shifts that can arise from erroneous grating positioning in direction perpendicular to the grating bars, the second variant is expanding this concept to slowly spatially varying phase shifts that can be caused by rotations, shifts or bending of the gratings during the scan, and the third variant will estimate flux variations and visibility variations in addition to constant phase shifts.

The estimation of drifts of the relative grating position has already been published by the author in [69].

5.5.1 Extension of the forward model

The forward model proposed at the beginning of this chapter,

$$I_i(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) = I_i^0 T_i [1 + V_i^0 D_i \cos(\phi_i + \Phi_i^0)] \quad , \quad (5.41)$$

can be slightly rewritten to differentiate at which physical detector pixel p and during which projection θ (and subsequently at which time and projection angle) a measurement i was taken:

$$I_{p,\theta}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}) = I_{p,\theta}^0 T_{p,\theta} [1 + V_{p,\theta}^0 D_{p,\theta} \cos(\phi_{p,\theta} + \Phi_{p,\theta}^0)] . \quad (5.42)$$

We can also slightly rewrite $I_{p,\theta}^0$, $V_{p,\theta}^0$, and $\Phi_{p,\theta}^0$, which represent the stepping curve of a hypothetical scan without an object in the beam path at exactly the same time as the actual object scan, into two components each:

$$I_{p,\theta}^0 = I_{p,\theta}^R \Delta I_{p,\theta}^0 , \quad (5.43)$$

$$V_{p,\theta}^0 = V_{p,\theta}^R \Delta V_{p,\theta}^0 , \quad (5.44)$$

$$\Phi_{p,\theta}^0 = \Phi_{p,\theta}^R + \Delta \Phi_{p,\theta}^0 . \quad (5.45)$$

The first components $I_{p,\theta}^R$, $V_{p,\theta}^R$, and $\Phi_{p,\theta}^R$ are the values acquired using a reference scan shortly before (or after) the scan and the second components are $\Delta I_{p,\theta}^0$, $\Delta V_{p,\theta}^0$, and $\Delta \Phi_{p,\theta}^0$ are the difference between the acquired reference scan and the hypothetical true value.

The task of the internal parameter estimation is now to estimate all $\Delta I_{p,\theta}^0$, $\Delta V_{p,\theta}^0$, and $\Delta \Phi_{p,\theta}^0$. In the form written above, there are three times as many parameters as measurements were taken, so it is impossible to estimate them all directly and approximations must be made to reduce the number of parameters to estimate.

5.5.2 Grating position estimation

Precisely known grating positions are of highest importance during a GBCT scan, as the positioning of the gratings perpendicular to the grating bars strongly influences the phase of the measured interferograms.

Generally, the reference scan phase $\Phi_{p,\theta}^0$ can be separated into a spatially varying contribution that describes the shape of the Moiré fringe pattern on the detector $\Phi_{p,\theta}^M$ and a component Φ_{θ}^G that is constant per projection and directly dependent on the positioning of the gratings.

Φ_{θ}^G is directly proportional to the position of each grating in direction perpendicular to the

grating bars relative to an arbitrary reference point (the choice of this reference point influences $\Phi_{\theta,p}^M$). For each grating g the contribution to Φ_{θ}^G is given as $\frac{\Delta x_{\theta}^g}{p_g} 2\pi$, where Δx_{θ}^g is the relative position of the grating g to its reference point at time θ and p_g is the period of this grating. Combining these contributions leads to

$$\Phi_{\theta}^G = \Delta x_{\theta} 2\pi , \quad (5.46)$$

with

$$\Delta x_{\theta} = \sum_g \frac{\Delta x_{\theta}^g}{p_g} . \quad (5.47)$$

Δx_{θ} depends on the exact positioning of the gratings and even errors in the positioning on a scale significantly below the grating pitch, will cause deviations in Φ_{θ}^G . This required level of exact and repeated positioning of the gratings is not only hard to consistently achieve but also thermal expansion of the gratings and vibrations can easily cause significant changes in the relative positioning of the gratings [80].

If we assume that the shape of the Moiré fringe pattern ($\Phi_{p,\theta}^{0,M}$) does not change over time, there is only a grating drift term $\Delta\Phi_{\theta}^{0,G}$ left to estimate. The true reference phase can thus be written as follows:

$$\Phi_{p,\theta}^0 = \Phi_{p,\theta}^{R,M} + \underbrace{\Delta\Phi_{p,\theta}^{0,M}}_{=0} + \underbrace{\Phi_{\theta}^{R,G} + \Delta\Phi_{\theta}^{0,G}}_{=\Delta\Phi_{\theta}^{0,G'}} . \quad (5.48)$$

5.5.2.1 Algorithm

Using the notation above, the optimization problem for the simultaneous estimation of the reconstruction and the grating positions can be written as

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{c}} = \arg \min_{\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \boldsymbol{c}} \sum_{p,\theta} (y_{p,\theta} - I_{p,\theta}^0 T_{p,\theta} (1 + V_{p,\theta}^0 D_{p,\theta} \cos(\phi_{p,\theta} + \Phi_{p,\theta}^R + c_{\theta})))^2 w_{p,\theta} , \quad (5.49)$$

with $c_{\theta} = \Delta\Phi_{\theta}^{0,G}$.

Modifying the solver developed for IBSIR in Sec. 5.3 for this new objective might be possible but the whole estimation problem is non-linear and especially the optimization along the

c_θ -dimensions is difficult as the terms are encompassed in a cosine and, if c_θ is large, a gradient-based optimization (as employed by the OS-SPS algorithm) could easily converge to a local minimum and not the global minimum. A quick solution that is easy to implement and circumvents the problem of a gradient-based optimizer converging to an only-local minimum is presented in the following.

Each iteration of the optimization algorithm is divided into two steps: In the first step only $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ are updated and in the second step only \mathbf{c} is updated. These two steps are repeated until convergence.

For the first step an iteration of the OS-SPS algorithm for IBSIR is used. The second step uses an analytic solution to the optimization problem, which will be derived in the following.

For fixed $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ the optimization of \mathbf{c} can be written as

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \sum_{p,\theta} (y'_{p,\theta} - \cos(\Phi'_{p,\theta} + c_\theta))^2 w'_{p,\theta} , \quad (5.50)$$

with $y'_{p,\theta} = \frac{y_{p,\theta} - I_{p,\theta}^0 T_{p,\theta}}{I_{p,\theta}^0 T_{p,\theta} V_{p,\theta}^0 D_{p,\theta}}$, $w'_{p,\theta} = w_{p,\theta} (I_{p,\theta}^0 T_{p,\theta} V_{p,\theta}^0 D_{p,\theta})^2$, and $\Phi'_{p,\theta} = \phi_{p,\theta} + \Phi_{p,\theta}^R$.

This optimization problem can be solved separately for each θ and, after dropping the θ index and the superscript $'$ for simplicity, using trigonometric identities it can be rewritten for a single projection as

$$\hat{c} = \arg \min_c \sum_p (y_p - \cos(c) \cos(\Phi_p) + \sin(c) \sin(\Phi_p))^2 w_p . \quad (5.51)$$

Explicitly calculating out the square and factoring out $\cos(c)$ and $\sin(c)$ leaves

$$\begin{aligned} \hat{c} = \arg \min_c & \underbrace{\sum_p y_p^2 w_p}_{=k_{yy}} + \underbrace{\cos^2(c) \sum_p -\cos^2(\Phi_p) w_p}_{k_{cc}} + \underbrace{\sin^2(c) \sum_p \sin^2(\Phi_p) w_p}_{=k_{ss}} \\ & + \underbrace{\cos(c) \sum_p -2y_p \cos(\Phi_p) w_p}_{=k_c} + \underbrace{\sin(c) \sum_p 2y_p \sin(\Phi_p) w_p}_{=k_s} \\ & + \underbrace{\cos(c) \sin(c) \sum_p -2 \cos(\Phi_p) \sin(\Phi_p) w_p}_{=k_{cs}} . \end{aligned} \quad (5.52)$$

Finding the extrema of this second order trigonometric polynomial can be achieved by first taking the derivative and setting it to zero, which results in a second order trigonometric polynomial equation,

$$\cos^2(c) k_{cs} - \sin^2(c) k_{cs} + \cos(c) \sin(c)(2k_{ss} - 2k_{cc}) + \cos(c) k_s - \sin(c) k_c = 0 . \quad (5.53)$$

Any trigonometric polynomial equation can be transformed into a two-dimensional polynomial equation of the same order in C and S under the constraint $C^2 + S^2 = 1$ by simply substituting $C = \cos(c)$ and $S = \sin(c)$. Directly solving this system of equations is not possible but after squaring the first equation one can insert the constraint into the equation and arrive at the following fourth order polynomial equation in C .

$$\begin{aligned} & C^4(2k_{cs}^2 + 4(k_{ss} - k_{cc})^2) \\ & + C^3(4k_c(k_{ss} - k_{cc}) + 4k_{cs}k_s) \\ & + C^2(k_c^2 + k_s^2 - 4k_{cs}^2 - 4(k_{ss} - k_{cc})^2) \\ & + C(4k_c(k_{ss} - k_{cc}) - 2k_{cs}k_s) \\ & + k_{cs}^2 - k_c^2 = 0 \end{aligned} \quad (5.54)$$

Finding the roots of a fourth order polynomial equation can be solved analytically but it is much faster to use a standard numerical algorithm for finding the roots of polynomials [84]. Finding the global minimum of the original minimization problem is now just a matter of comparing the likelihood values all real roots (there may be two or four real roots).

5.5.2.2 Results

5.5.2.2.1 Simulation To assess the viability of the proposed approach first a GBCT scan was simulated. As scanned object a phantom exhibiting different X-ray attenuation, refractive index, and linear diffusion coefficient was used, see top row of Fig. 5.12. The GBCT scan consisted of 403 equidistant angular positions and at each angular position three measurements at nominally equidistant grating positions were performed. To simulate imprecise grating po-

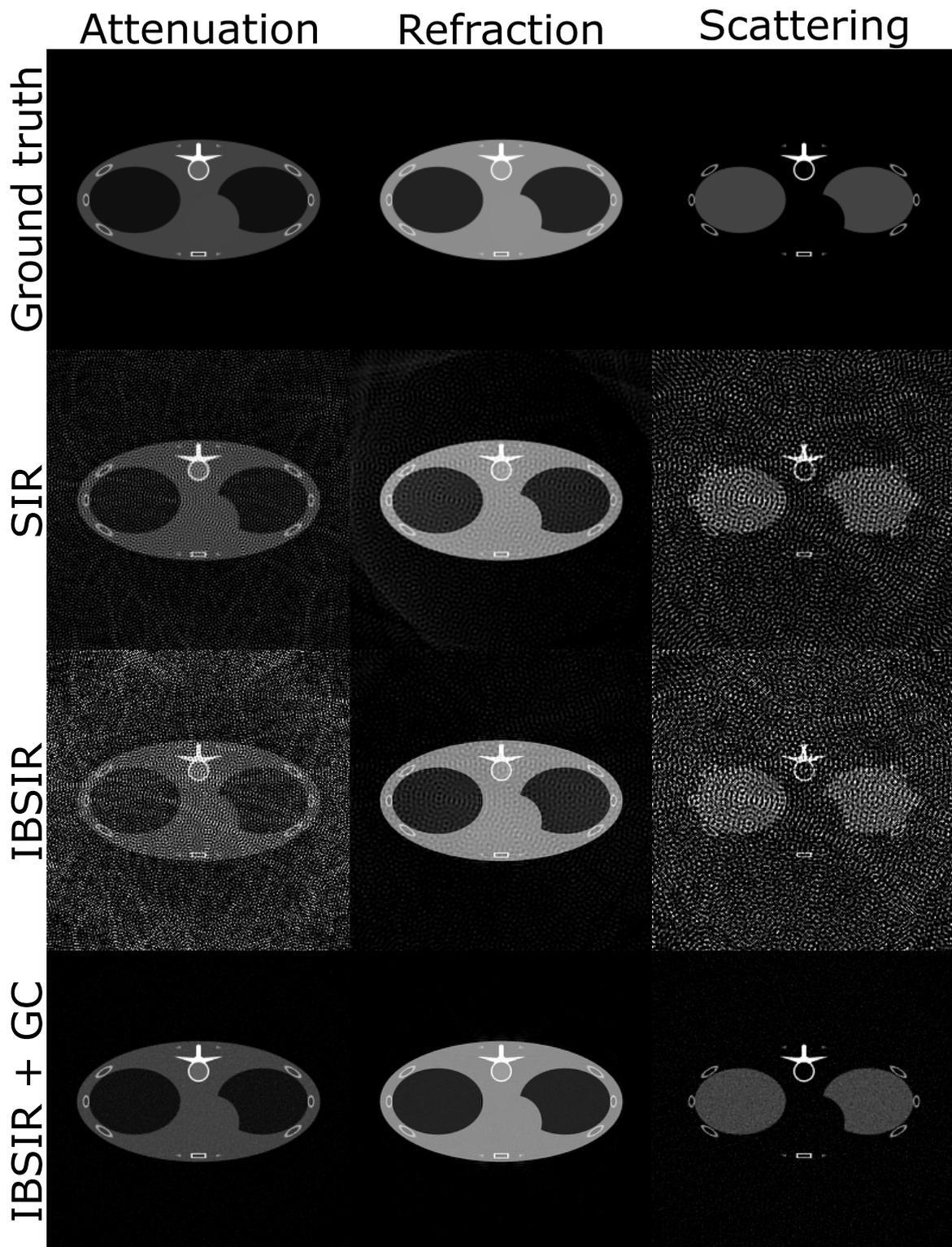


Figure 5.12: Different reconstruction techniques were applied to a scan with errors in the, nominally equidistant, phase stepping positions. SIR and IBSIR show reconstructions corrupted by high frequency artifact patterns. The artifact appearance is similar but slightly different for both reconstructions, which can be attributed to cross-talk between channels in the IBSIR reconstruction. A reconstruction using IBSIR and the proposed grating position correction (GC) provided reconstruction results comparable to the ground truth. All images are shown on the same level and window. Figure adapted from [69].

sitioning or vibrations the actual grating positions were shifted by a random number between -5% and +5% of the grating period perpendicular to the grating bars separately for each measurement. The visibility of the setup was chosen to be 0.4 and 3400 monochromatic photons emitted from the X-ray tube per pixel were simulated. The simulated detector was a noise-free photon counting detector in a parallel beam geometry.

The results of the simulation, see Fig. 5.12, were reconstructed using various reconstruction methods: First the measurements were analyzed using the maximum-likelihood signal estimation, see Sec. 3.1, to extract attenuation, differential phase, and dark-field sinograms. These sinograms were then reconstructed using SIR with the projections weights as extracted from the maximum-likelihood signal estimation. The reconstructions result in volumes that are strongly corrupted by artifacts. All three images show high frequency artifacts evenly distributed over the whole volume and the phase image additionally shows a slight gradient in the background of the reconstructed volume. The shape of these artifacts can be convincingly explained as every projection angle has slightly different errors from the incorrect grating positions and sometimes the errors add up in the reconstructed volumes and sometimes, they equal out, resulting in the observed wavy pattern.

Additionally, the raw measurements were directly reconstructed using the IBSIR algorithm without internal parameter estimation. Here the results are quite similar, the artifacts seem somewhat stronger in the attenuation and scattering volumes but the gradient in the phase image is gone. It is not clearly evident where the differences to the SIR reconstruction come from but as in the IBSIR model one channel influences the others and it might be that the regularization forcing a smooth background in the phase image is causing additional errors in the two other images.

As a last reconstruction technique the IBSIR method with internal parameter estimation for the grating positions was used. The resulting volumes are artifact free and show a high resolution close to the ground truth. From these results one can conclude that the errors in the grating position could be corrected for.

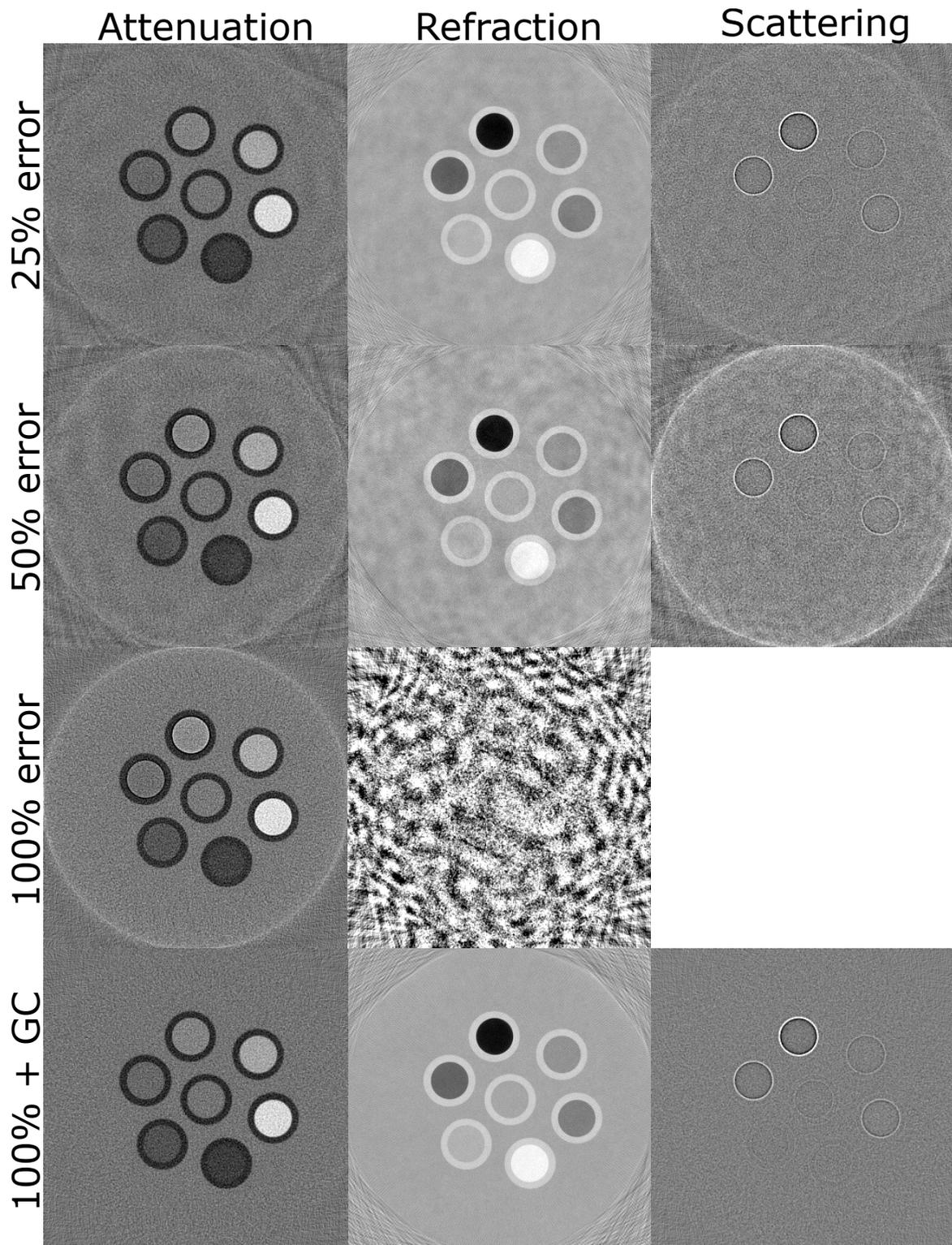


Figure 5.13: Reconstructions using IBSIR with faulty knowledge about the grating positions. The reconstructions show foggy artifacts which worsen with less precise knowledge of the grating position until a meaningful reconstruction is impossible. Using grating position correction (GC) a reconstruct of the image is possible even with unknown grating positions. The same level and window were chosen for all reconstructions. Figure adapted from [69].

5.5.2.2.2 Experiment The grating position optimization was also tested on a real measurement: Several plastic tubes, which were submerged in a water bath and filled with different liquids, were scanned and reconstructed. The scan was performed using a symmetric Talbot-Lau interferometer with grating distances $\overline{G_0 G_1} = \overline{G_1 G_2} = 85.7$ cm. The three gratings G_0 , G_1 , and G_2 are all made of gold on a silicon substrate with periods of $5.4 \mu\text{m}$. The absorption gratings G_0 and G_2 have a height of $70 \mu\text{m}$ and $65 \mu\text{m}$ respectively. The phase grating G_1 is designed to give a phase shift of π at the system design energy of 27 keV. The source is an ENRAF Nonius FR 591 rotating anode X-ray tube with a Molybdenum target operated at 40 kVp and 70 mA. The detector used is a single photon counting PILATUS 100k module with 487×195 pixels and a pixel size of $172 \times 172 \mu\text{m}^2$. The effective resolution is $100 \times 100 \mu\text{m}^2$ due to a magnification of 1.7 . 800 projection angles evenly spaced over 360 degrees were measured with 11 different grating positions at each angular position. All reconstructions were carried out using the IBSIR algorithm. Prior to reconstruction a random error, uniformly distributed between $\pm\frac{\pi}{4}$, $\pm\frac{\pi}{2}$ or $\pm\pi$, was added to the phase reference term $\Phi_\theta^{R,G}$ for every interferogram. This equals to an uncertainty in the position of the grating of $\pm\frac{p}{8}$, $\pm\frac{p}{4}$, or $\pm\frac{p}{2}$, where p is the grating period. Thus the last scenario does not contain any knowledge about the grating positions at all.

With increasing grating position uncertainty IBSIR reconstructions without grating position estimation show increasing artifacts (low-frequency cloud-like distortions in all three images, ring artifacts at the border of the FOV, and cross-talk between images), see Fig. 5.13. For completely unknown grating positions, the reconstruction of δ and ϵ failed completely, stopping after one iteration at a noise like structure for δ or at a very high level for ϵ . The IBSIR with grating position estimation converges to the same reconstruction as with complete knowledge about the grating positions, even for the case of no knowledge about the grating position at all, compare Fig. 5.13.

5.5.3 Higher order phase drift estimation

The estimation of the grating position, respectively of a constant phase shift per projection, is definitely the most important phase correction but the phase drifts encountered in real

systems can often be more complex than the spatially uniform drifts described in the previous subsection [80, 82]. These non-uniform phase drifts can be caused by various different effects: e.g. rotation of the gratings with respect to each other, bending of the gratings, or movement of the gratings in the beam direction.

5.5.3.1 Algorithm

To solve this problem one can stick to the general formulation employed for the grating position estimation but expand its model by replacing $\Delta\Phi_{\theta}^{0,G}$ with $\Delta\Phi_{p,\theta}^0$, which now also encompasses changes of the shape of the Moiré fringes per projection:

$$C = \sum_{p,\theta} (y_{p,\theta} - I_{p,\theta}^0 T_{p,\theta} (1 + V_{p,\theta}^0 D_{p,\theta} \cos(\phi_{p,\theta} + \Phi_{p,\theta}^R + \Delta\Phi_{p,\theta}^0)))^2 w_{p,\theta} . \quad (5.55)$$

Treating each $\Delta\Phi_{p,\theta}^0$ as a variable that has to be directly estimated is, of course, infeasible as it would increase the number of unknowns by the number of measurements $y_{p,\theta}$. But as all continuous functions can be approximated up to an arbitrary degree by polynomials, it is proposed to approximate each $\Delta\Phi_{p,\theta}^0$ with two-dimensional polynomials $\mathbf{P}(\mathbf{c}_{\theta})$ with different coefficients \mathbf{c}_{θ} for each projection θ . Many different basis functions for two-dimensional polynomials exist and the optimal choice is not obvious. In this work only Cartesian polynomials were used, as they are very simple and the shape of the indeterminates can be linked to different real physical situations that would change the phase of the interferometer and should thus be well capable of approximating the total phase shift.

The optimization problem can now be written as

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\epsilon}}, \hat{\mathbf{c}} = \sum_{p,\theta} \left(y_{p,\theta} - I_{p,\theta}^0 T_{p,\theta} \left(1 + V_{p,\theta}^0 D_{p,\theta} \cos(\phi_{p,\theta} + \Phi_{p,\theta}^R + \sum_{k=0}^n c_{k,\theta} P_{k,p}) \right) \right)^2 w_{p,\theta} . \quad (5.56)$$

Solving this non-linear, non-convex problem is challenging and not guaranteed to converge to the global minimum.

Again, it is proposed to utilize a block-wise optimization approach that alternates between updating $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ and \mathbf{c} . If \mathbf{c} is kept fixed, the update to $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ can be handled by the

OS-SPS solver for IBSIR. If $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ are fixed, calculating an update to \mathbf{c} can be performed in parallel for each θ . Unfortunately, no analytic solution for arbitrary \mathbf{P} could be found for this step, so it is necessary to resort to ordinary gradient-based optimization methods for this sub-optimization problem. As \mathbf{c}_θ is usually relatively low dimensional and because there are simple analytic expressions for the second derivatives of the cost function with respect to the coefficients \mathbf{c} for arbitrary \mathbf{P} , it is proposed to use a simple Newton update step as it is not overly expensive to calculate the full inverse hessian matrix:

$$\mathbf{c}_\theta^{(n+1)} = \mathbf{c}_\theta^{(n)} - H_{\mathbf{c}_\theta}^{-1} \nabla_{\mathbf{c}_\theta} L . \quad (5.57)$$

5.5.3.1.1 Error correction Finding the global minimum of the cost function with higher order polynomial phase terms is not trivial as the minimization can easily get stuck in local minima. Generally, this behavior is hard to track as there is no easy way to check if a minimum is global or local. But in the specific case at hand, a handy way to verify if the optimization of the phase term was successful exists: The likelihood in the global minimum per projection at projection θ' is extremely similar to the likelihood per projection in the global minimum at neighboring projections θ , as the values of the transmission and dark-field averaged per projection do not change quickly (as the object only changes slowly). During the optimization process we can use this to identify a projection θ' where the coefficients $\mathbf{c}_{\theta'}$ are not at their global minima by simply comparing the value of the likelihood per projection for projection θ' ,

$$l_{\theta'} = \sum_p \left(y_{p,\theta'} - I_{p,\theta'}^0 T_{p,\theta'} (1 + V_{p,\theta'}^0 D_{p,\theta'} \cos(\phi_{p,\theta'} + \Phi_{p,\theta'}^R + \sum_{k=0}^n c_{k,\theta'} P_{k,p})) \right)^2 w_{p,\theta'} . \quad (5.58)$$

with the likelihood per projection at neighboring projections θ . If the value of θ' is much higher than the lowest value for any of the other θ in the neighborhood of θ' we can safely assume that the optimization is stuck in a local minimum.

In most scenarios we can assume that the phase drift between two projections taken right after each other is smaller than the difference between projections with a longer temporal distance. This motivates restarting the optimization of $\mathbf{c}_{\theta'}$ with the values of $\mathbf{c}_{\theta'+1}$ or $\mathbf{c}_{\theta'-1}$, whichever

has a lower value and is thus more probable at its global minimum. If the grating position was deliberately changed between θ' and $\theta' + 1$ respectively $\theta' - 1$ (e.g. due to phase stepping), it is highly advised to perform an analytic update of $c_{0,\theta'}$ first to estimate the correct grating position and then commence the gradient-based optimization of all variables in $\mathbf{c}_{\theta'}$.

This correction was not used in the following results.

5.5.3.2 Results

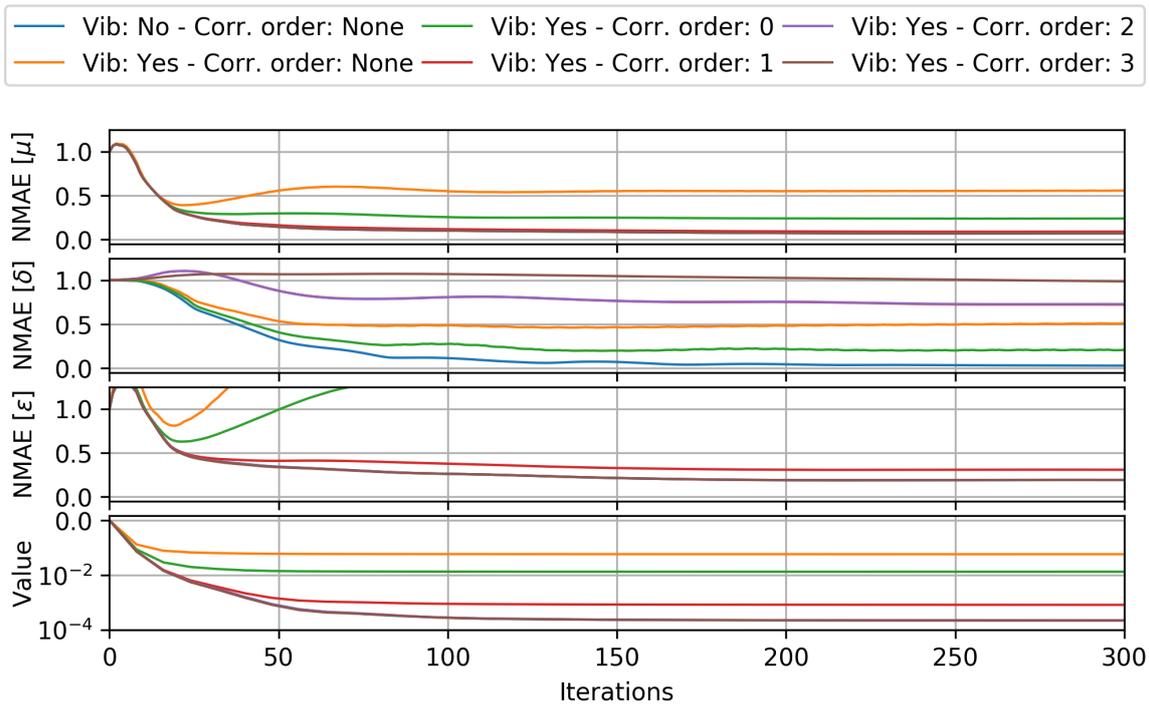


Figure 5.14: Change of the cost function value and the normalized mean absolute error (NMAE) in each image channel with the number of iterations for six reconstructions. One reconstruction was performed with perfect knowledge of the vibrations (“Vib: None”), for the other reconstructions the phase shift caused by the vibrations was unknown. One of these reconstructions (“Corr. order: None”) was performed without correction and for the others the proposed phase shift estimation was used with 2D Cartesian polynomials of different orders as model functions.

To test the algorithm in the presence of phase shifts with a non-constant spatial shape a simulation study was performed: The simulated scan was in parallel beam geometry with a photon-counting detector with perfect quantum efficiency consisting of 16 lines with 256 pixels in each line. 1203 equidistantly spaced projections were taken and between each projection the grating position was moved by $\frac{1}{2.618}$ of the grating pitch. Moiré fringes with a period of 10.1 pixels in both detector dimensions were simulated. The simulated interferometer has a visibil-

ity of 0.4 and 10^5 monochromatic photons per pixel were emitted from the X-ray source. The attenuation and linear diffusion coefficients of the phantom were chosen such that the lowest transmission and dark-field signal was around 0.4. The refractive index was chosen such that the highest differential phase contrast was about 0.6. To simulate an unknown spatially non-uniform phase shift, 2D Cartesian Polynomials with random coefficients were created for each projection and added to $\Delta\Phi_{p,\theta}^0$. The value of each random coefficient was chosen from random uniform distributions of values. The maximum (respectively minimum) value for each of these distributions was chosen such that this one coefficient alone would cause a phase change of at most π (respectively $-\pi$) per projection.

Five different reconstructions were performed: Without phase estimation and with phase estimation using 2D Cartesian Polynomials of order zero to three. All reconstructions were started from an empty initial guess. The progress of the reconstructions can be seen in Fig. 5.14. The distributions of μ and ϵ are perfectly recovered for corrections of order two and three and the value of the cost function is also reduced to the same level as for a reconstruction with perfect knowledge of the phase. Just for the reconstruction of δ we see a behavior that is different: For correction order zero the NMAE is reduced (compared to the uncorrected reconstruction) but for a reconstruction with correction order of one or higher the NMAE is larger than for the uncorrected reconstruction! If we look at the actual reconstruction result of δ , see Fig. 5.15, we can observe that the reconstruction using corrections of order two successfully removed the high frequency artifacts but exhibits low frequency artifacts. These artifacts are created because on the one hand, the cost function is not really increased by them (as low frequency structures have very little impact on the DPC signal and thus on the cost function), and on the other hand, the DPC signal (if the reconstruction is not yet perfectly converged) is very prone to be slightly fitted by the correction polynomial as is not evenly distributed per projection but is generally negative on the left side and positive on the right side (or vice versa).

Fortunately, these low frequency artifacts can be easily corrected: We can generate a simulation of the DPC signal implied by the reconstruction by simply forward projecting δ and we know that the true DPC signal is distributed around zero in regions without sample (these regions can be easily distinguished by, for example, looking at the forward projection of μ) and that

the part of the polynomial that was fitted wrongly is again a polynomial of the same order. This means we simply have to look for the surface described by a polynomial of the same order as the correction that best fits the DPC signal at pixels outside of the object and subtract it from the DPC signal to generate a corrected DPC signal. In Fig. 5.15C we can see a simple FBP reconstruction from the DPC signal corrected this way. All low frequencies artifacts are well removed by this correction.

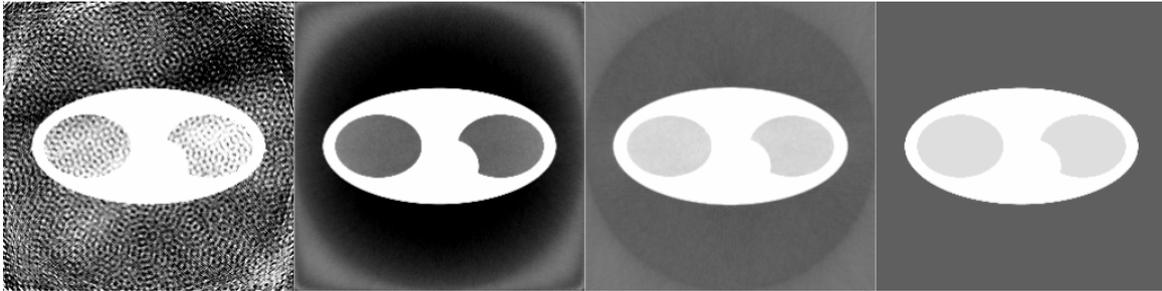


Figure 5.15: Slices from reconstructions of the refractive index decrement. (A) From a reconstruction without phase shift estimation. (B) from a reconstruction with phase shift estimation using a second order Cartesian 2D polynomial. (C) from an FBP reconstruction based on the low-frequency corrected DPC signal generated from (B). (D) Ground truth. All reconstructions are shown on the same window.

5.5.4 Flux and visibility drifts

5.5.4.1 Flux drift model

Changing the X-ray photon flux per time is a well-established method to reduce the dose applied to patients and widely used in all modern clinical CTs [85]. But the nominally applied flux and the real flux emitted are not exactly the same, thus the real flux emitted is often measured by utilization of a specialized reference detector [86, 87] or by monitoring the mean intensity in an detector area not covered by the sample. Not all setups utilize a reference detector and the later solution is not viable for a GBI setup, as the interference pattern changes over time and thus the change of the measured intensity in a sample-free region is not only depending on changes of tubes fluence but also on changes of the interference pattern. So having an algorithm to estimate the fluence from the measurements would be beneficial for GBI.

If we assume that the shape of the X-ray illumination and the specific sensitivity of each

pixel stays constant over time, the mean intensity $I_{p,\theta}^R$ acquired at a reference scan is simply modulated by a time dependent contribution ΔI_θ^0 that has the same value for all pixels in a projection. The model function can thus be written as

$$I_{p,\theta} = \Delta I_\theta^0 I_{p,\theta}^R T_{p,\theta} (1 + V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta})) . \quad (5.59)$$

5.5.4.2 Visibility drift model

Drifts in the relative positioning of the gratings do not only change the phase of the interferometer but can also affect the perceived visibility of the setups: If we assume the nominal visibility of the setup to be constant over time and we assume that the relative position of one of the gratings g changes with a constant velocity v_t during the integration period of the detector P centered around time t the measured intensity $y_{p,t}$ can be calculated as follows:

$$\begin{aligned} I_{p,\theta} &= \int_{-P/2}^{P/2} \frac{I_{p,\theta}^R T_{p,\theta}}{P} (1 + V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta} + \frac{v_t \tau}{p_g} 2\pi)) d\tau \\ &= I_{p,\theta}^R T_{p,\theta} + \frac{I_{p,\theta}^R T_{p,\theta}}{P} V_{p,\theta}^R D_{p,\theta} \int_{-P/2}^{P/2} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta} + \frac{v_t \tau}{p_g} 2\pi) d\tau \\ &= I_{p,\theta}^R T_{p,\theta} + \frac{I_{p,\theta}^R T_{p,\theta}}{P} V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta}) \int_{-P/2}^{P/2} \cos(\frac{v_t \tau}{p_g} 2\pi) d\tau \\ &\quad - \frac{I_{p,\theta}^R T_{p,\theta}}{P} V_{p,\theta}^R D_{p,\theta} \sin(\Phi_{p,\theta}^R + \phi_{p,\theta}) \underbrace{\int_{-P/2}^{P/2} \sin(\frac{v_t \tau}{p_g} 2\pi) d\tau}_{=0} \\ &= I_{p,\theta}^R T_{p,\theta} + I_{p,\theta}^R T_{p,\theta} V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta}) \sin(\frac{P v}{2 p} 2\pi) \frac{2 p}{P v_t} \frac{1}{2\pi} \\ &= I_{p,\theta}^R T_{p,\theta} + I_{p,\theta}^R T_{p,\theta} V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta}) \operatorname{sinc}(\frac{v_t P}{p_g} \pi) \\ &= I_{p,\theta}^R T_{p,\theta} (1 + \Delta V_\theta^0 V_{p,\theta}^R D_{p,\theta} \cos(\Phi_{p,\theta}^R + \phi_{p,\theta})) \end{aligned} \quad (5.60)$$

From this calculation one can see that phase drifts during the integration period reduce the perceived visibility of a GBI setup and, as the velocity of these drifts is seldomly constant over longer periods of time, the reduction of visibility is possibly different for each measurement.

5.5.4.3 Algorithm

Combining the flux drift and the visibility drift with a constant phase drift per projection leads to the following forward model:

$$I_{p,\theta}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \Delta I_\theta^0, \Delta V_\theta^0, \Delta \Phi_\theta^0) = \Delta I_\theta^0 I_{p,\theta}^{R'} T_{p,\theta} (1 + \Delta V_\theta^0 V_{p,\theta}^{R'} D_{p,\theta} \cos(\Delta \Phi_\theta^0 + \Phi_{p,\theta}^0 + \phi_{p,\theta})) . \quad (5.61)$$

The signal estimation in the maximum-likelihood sense can consequently be written as follows:

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\epsilon}}, \hat{\Delta \mathbf{I}}^0, \hat{\Delta \mathbf{V}}^0, \hat{\Delta \Phi}^0 = \arg \min_{\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \Delta \mathbf{I}^0, \Delta \mathbf{V}^0, \Delta \Phi^0} \sum_{p,\theta} (y_{p,\theta} - I_{p,\theta}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, \Delta I_\theta^0, \Delta V_\theta^0, \Delta \Phi_\theta^0))^2 , \quad (5.62)$$

where the bold symbols are vectors containing the values for all projections θ .

Solving this non-linear, non-convex optimization problem in a straight-forward fashion with a gradient-based optimization algorithm is again not guaranteed to be fast nor does it necessarily converge to the global minimum, as it is not guaranteed to have a single minimum. Again the optimization can be divided into alternating updates of the reconstruction and of the drifts. The optimization of the reconstruction problem can be solved using the OS-SPS solver for IBSIR. The optimization problem for the estimation of the drifts,

$$\hat{\Delta \mathbf{I}}^0, \hat{\Delta \mathbf{V}}^0, \hat{\Delta \Phi}^0 = \arg \min_{\Delta \mathbf{I}^0, \Delta \mathbf{V}^0, \Delta \Phi^0} \sum_{p,\theta} (y_{p,\theta} - I_{p,\theta}(\Delta I_\theta^0, \Delta V_\theta^0, \Delta \Phi_\theta^0))^2 , \quad (5.63)$$

can be solved easily as it can be processed separately for each projection and rewritten as follows:

$$\hat{\Delta I}^0, \hat{\Delta V}^0, \hat{\Delta \Phi}^0 = \arg \min_{\Delta I^0, \Delta V^0, \Delta \Phi^0} \sum_p (y_p - \Delta I^0 I_p^{R'} (1 + \Delta V^0 V_p^{R'} \cos(\Delta \Phi^0 + \Phi_p^{R'})))^2 . \quad (5.64)$$

This equation has the same shape as the maximum-likelihood signal estimation presented in Sec. 3.1 and is thus quickly solved using the same linearization approach.

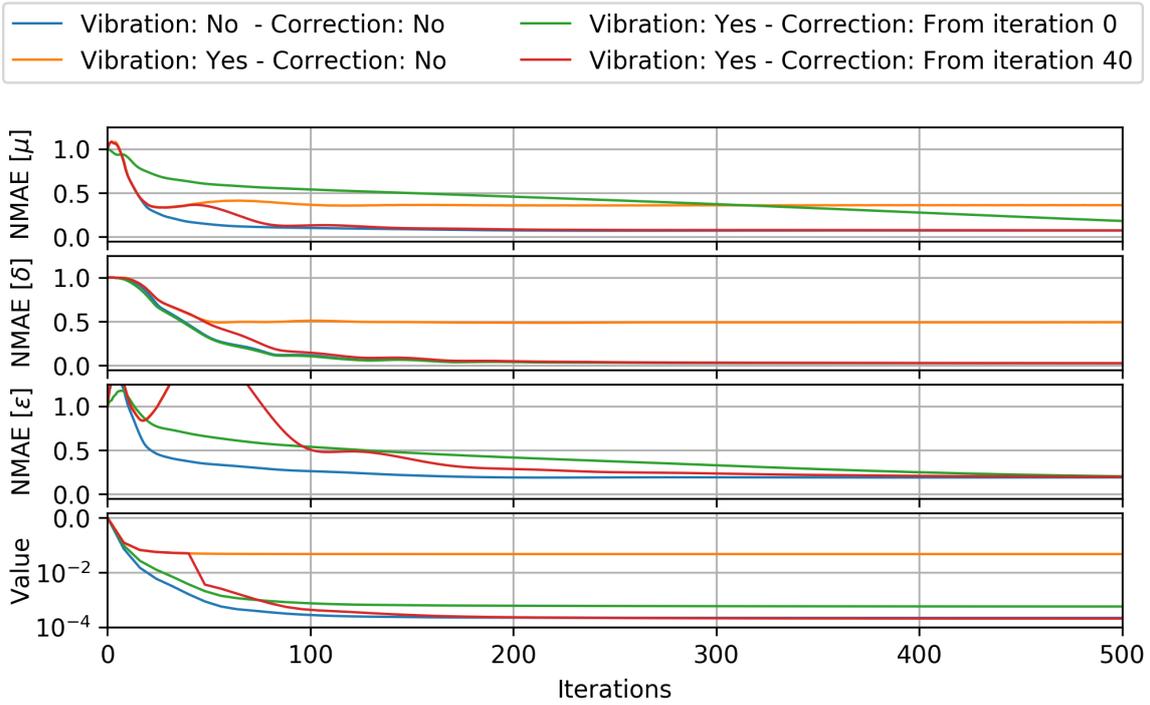


Figure 5.16: Change of the cost function value and the normalized mean absolute error (NMAE) in each image channel with the number of iterations for four reconstructions. One reconstruction was performed with perfect knowledge of the vibrations ("Vib: No"), for the other reconstructions the effect of the vibrations was initially unknown. One of these reconstructions was performed without correction, one with the correction active from the first iteration and one with the correction activated after 40 iterations.

5.5.4.4 Results

The effect of this first order drift correction was evaluated on a set of simulations with nominal scan parameters as for the simulation in Sec. 5.5.3.2. For one simulation the nominal values were not changed and for one simulation the nominal values of I^0 , V^0 , and Φ^0 were changed by a random uniform distributed factor per projection between 0.9 and 1.1, respectively -0.1π and 0.1π , to simulate strong vibrations. These measurements were reconstructed once using no correction, once using the correction after each IBSIR iteration, starting from the first iteration on, and once using the correction after each IBSIR iteration but only starting after 40 initial ordinary IBSIR iterations. The reconstruction of the corrupted scan using no parameter estimation results in an artifact ridden reconstruction in each channel, which results in a high value of the cost function and high NMAEs in each channel after convergence, see Fig. 5.16. Using the drift correction after each OS-SPS iteration right from the start of the reconstruction

results in a much better final reconstruction that is converging to the reference solution, but the convergence is extremely slow. This is caused by the fact that $\Delta \mathbf{I}^0$ (and to a smaller extent $\Delta \mathbf{V}^0$) is highly correlated with $\boldsymbol{\mu}$ (respectively $\boldsymbol{\epsilon}$) and in the beginning the flux estimation is wrongly trying to fit the reduced measured intensity caused by attenuation in the object by simply reducing the overall flux of each projection. Over many iterations this over-fitting is corrected but this takes a long time as in each iteration only a small correction between $\Delta \mathbf{I}^0$ and $\boldsymbol{\mu}$ can be performed.

A simple solution that circumvents this problem is to first create a low quality starting image by simply not using the parameter estimation step for the first 40 iterations. This way the over-fitting does not take place and the reconstruction converges much faster.

5.6 Discussion

In this chapter I proposed a novel direct reconstruction method for GBCT that is suitable for all acquisition schemes and developed a suitable algorithm to efficiently use this reconstruction method. I showed that the proposed algorithm can produce artifact-free reconstruction even from single shot GBCT scans and proposed a loose sampling criterion, that is not a rigorous explanation but nevertheless helps with designing suitable acquisition protocols for GBCT.

One such acquisition protocol is the combination of single shot acquisition with focal spot sweeping. This protocol is perfectly suited for the application within a continuously rotating clinical CT system equipped with a grating interferometer as these systems are generally equipped with the required focal spot sweeping technique.

In addition, I showed that the IBSIR model can be expanded to incorporate various possible drifts of the setup characteristics. These corrections work extremely well for the estimation of the grating positions, up to the point, that no knowledge of the grating positions is required anymore. This might allow for the construction of much simplified GBCT systems with respect to the required phase stepping accuracy.

The estimation of moderate non-uniform phase drifts is also possible, but it is not guaranteed to perfectly correct arbitrarily strong phase shifts. There is much room for improved optimization

algorithms that could provide an improved level of convergence to the global minimum.

The estimation of constant flux and visibility variations is also possible, but the convergence speed of the algorithm proposed in this thesis can certainly be greatly improved.

Chapter 6

Conclusion and outlook

In this chapter the main results of this PhD thesis are summarized and a brief outlook about possible worthwhile research areas is given.

Bias reduced signal estimation

The main result of chapter 3 was the introduction of a novel estimator for the dark-field signal. It was shown that the regularized bias-corrected maximum-likelihood estimator (BCMLE) has a reduced estimation bias compared to the ordinary maximum-likelihood (MLE) estimator. The BCMLE may have an increased variance compared to the MLE but, through proper choice of the regularization parameter, it was possible to create an estimator that is strictly superior to the MLE. This result proved that the estimation bias in the dark-field signal can be substantially reduced compared to the MLE by the choice of a proper estimator.

It should be noted that often not the dark-field signal is the final quantity that is of interest, but the dark-field signal is only used as an intermediate result for the CT reconstruction of the linear diffusion coefficient. The input for standard reconstruction algorithms (like filtered backprojection or a statistical iterative reconstruction with a linear forward model) for the linear diffusion coefficient is not directly the dark-field signal but the negative logarithm of the dark-field signal. Unfortunately, a good estimator for the dark-field signal is not necessarily a good estimator for the logarithm of the dark-field signal, as the logarithm is a non-linear transformation. So, if the final quantity of interest is the linear diffusion coefficient it might be

advantageous to design an improved, bias-reduced estimator directly for the negative logarithm of the dark-field signal.

Alternatively, the reconstruction of the linear diffusion coefficient could be performed using the BCMLE estimate of the dark-field signal but then statistical iterative reconstruction using a non-linear forward model (analog to the non-linearized forward model for attenuation CT) should be employed. This approach would benefit from an estimate of the variance of the BCMLE of the dark-field signal.

The signal estimation break-down of the DPC signal was found to be not directly attributable to signal estimation bias but more to the fact that the estimation of the DPC signal is the estimate of a direction and not of a scalar value. Developing proper filter algorithms for the DPC signal that consider this fact should result in greatly improved results compared to conventional filter methods.

The development of a signal estimation algorithm that is regularized by taking into account the estimates in neighboring pixels could prove extremely useful in reducing both the variance and the bias in the signal estimation for the dark-field signal and the differential phase signal.

Sliding window signal estimation

The estimation of proper values for the attenuation, dark-field and differential phase signal from a sliding window scan could be slightly improved on by using a method of robust estimation. But the signal estimation still does not always present satisfactory results at sharp object edges with strong contrast. The artifacts caused by this can be reduced by subsequent application of suitable filters but overall it does not seem that sliding window interpolation is able to create GBCT reconstructions with the highest possible resolution.

A regularized signal estimation that considers the values of neighboring projections and/or pixels could possibly provide improved results but may be very expensive to calculate.

Intensity-based statistical iterative reconstruction

In the main chapter of the thesis various aspects of the intensity-based statistical iterative reconstruction (IBSIR) algorithm were evaluated.

First a fast optimization algorithm for the optimization problem was proposed and evaluated. Many different accelerations were tested and could provide a combined speed-up of more than

two magnitudes compared to the un-accelerated optimization algorithm.

Many different aspects of the optimization algorithm could still be improved on: the curvature derived from the optimal 1-d curvatures could be expanded to take into account the correlation between attenuation, dark-field and phase signal, the Fourier preconditioning approach could be expanded to the attenuation and dark-field channels (e.g. see [72] and [88]), and other momentum acceleration methods could be explored (see [89]).

The quality of the reconstructions using IBSIR was evaluated for several different acquisition protocols and compared to reconstructions using projection-based signal estimation and conventional reconstruction algorithms. IBSIR showed an improved reconstruction quality and was able to perfectly reconstruct a phantom scan using only a single measurement per projection angle.

A rigorous comparison of the quality of IBSIR compared to conventional reconstruction especially in the case of low dose scans could provide further insights, especially in the context of bias in the estimation of the dark-field signal and signal break-down in the differential phase signal.

The reconstruction using IBSIR can be made far more robust to the presence of vibrations and drifts by including possible vibration effects into the forward model of the cost function and estimating the strength of these effects simultaneously to the reconstruction. The simulation and experimental results proved that this method works extremely well and fast for the estimation of drifts in the grating positioning. The estimation of spatially uniform flux variations and of visibility reductions worked equally well but the convergence speed of the algorithm could still be greatly improved on. The correction of moderate spatially varying drifts of the reference phase was also introduced but required a subsequent correction step after the reconstruction. Better suited optimization approaches and an optimized choice of vibration basis functions may greatly increase the stability and speed of this approach.

Overall, I believe that the results presented in this thesis represent a major step forwards in the path of grating-based computed tomography from an experimental laboratory tool to a robust and highly usable diagnostic imaging technique that may significantly improve clinical diagnostics in the future.

Appendix A

Signal estimation

A.1 Linearized signal estimation

To solve the linearized signal estimation problem,

$$\hat{M}, \hat{C}, \hat{S} = \arg \min_{M, C, S} L(M, C, S) = \sum_i (Mm_i + Cc_i + Ss_i - y_i)^2 w_i, \quad (\text{A.1})$$

we first rewriting the likelihood function as follows

$$\begin{aligned} L(M, C, S) &= M^2 \sum_i m_i^2 w_i + C^2 \sum_i c_i^2 w_i + S^2 \sum_i s_i^2 w_i \\ &+ MC \sum_i 2m_i c_i w_i + MS \sum_i 2m_i s_i w_i + CS \sum_i 2c_i s_i w_i \\ &- M \sum_i 2m_i y_i w_i - C \sum_i 2c_i y_i w_i - S \sum_i 2s_i y_i w_i + \sum_i y_i^2 w_i \end{aligned} \quad (\text{A.2})$$

and simplifying it using

$$\begin{aligned} \Sigma_{mm} &= \sum_i m_i^2 w_i, \Sigma_{cc} = \sum_i c_i^2 w_i, \Sigma_{ss} = \sum_i s_i^2 w_i \\ \Sigma_{mc} &= \sum_i m_i c_i w_i, \Sigma_{ms} = \sum_i m_i s_i w_i, \Sigma_{cs} = \sum_i c_i s_i w_i \end{aligned}$$

$$\Sigma_{my} = \sum_i m_i y_i w_i, \Sigma_{cy} = \sum_i c_i y_i w_i, \Sigma_{sy} = \sum_i s_i y_i w_i, \Sigma_{yy} = \sum_i y_i^2 w_i$$

into

$$\begin{aligned} L(M, C, S) &= M^2 \Sigma_{mm} + C^2 \Sigma_{cc} + S^2 \Sigma_{ss} \\ &\quad + 2MC \Sigma_{mc} + 2MS \Sigma_{ms} + 2CS \Sigma_{cs} \\ &\quad - 2M \Sigma_{my} - 2C \Sigma_{cy} - 2S \Sigma_{sy} + \Sigma_{yy}. \end{aligned} \tag{A.3}$$

Now the minimization of this quadratic function with respect to M , C , and S is equivalent to finding the vector \mathbf{x} that solves the system of linear equations with system matrix A and measurement vector \mathbf{b} :

$$\begin{aligned} \frac{\partial L}{\partial M} = 0 &\Rightarrow M \Sigma_{mm} + C \Sigma_{mc} + S \Sigma_{ms} = \Sigma_{my} \\ \frac{\partial L}{\partial C} = 0 &\Rightarrow M \Sigma_{mc} + C \Sigma_{cc} + S \Sigma_{cs} = \Sigma_{cy} \\ \frac{\partial L}{\partial S} = 0 &\Rightarrow M \Sigma_{ms} + C \Sigma_{cs} + S \Sigma_{ss} = \Sigma_{sy} \\ \Leftrightarrow \underbrace{\begin{pmatrix} \Sigma_{mm} & \Sigma_{mc} & \Sigma_{ms} \\ \Sigma_{mc} & \Sigma_{cc} & \Sigma_{cs} \\ \Sigma_{ms} & \Sigma_{cs} & \Sigma_{ss} \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} M \\ C \\ S \end{pmatrix}}_{=\mathbf{x}} &= \underbrace{\begin{pmatrix} \Sigma_{my} \\ \Sigma_{cy} \\ \Sigma_{sy} \end{pmatrix}}_{=\mathbf{b}} \end{aligned} \tag{A.4}$$

. The estimates of M , C , and S are now the solution of $A^{-1}\mathbf{b}$ with

$$A^{-1} = \frac{1}{\Delta} \begin{pmatrix} \Sigma_{cc}\Sigma_{ss} - \Sigma_{cs}^2 & \Sigma_{cs}\Sigma_{ms} - \Sigma_{mc}\Sigma_{ss} & \Sigma_{cs}\Sigma_{mc} - \Sigma_{cc}\Sigma_{ms} \\ \Sigma_{cs}\Sigma_{ms} - \Sigma_{mc}\Sigma_{ss} & \Sigma_{mm}\Sigma_{ss} - \Sigma_{ms}^2 & \Sigma_{mc}\Sigma_{ms} - \Sigma_{mm}\Sigma_{cs} \\ \Sigma_{cs}\Sigma_{mc} - \Sigma_{cc}\Sigma_{ms} & \Sigma_{mc}\Sigma_{ms} - \Sigma_{cs}\Sigma_{mm} & \Sigma_{cc}\Sigma_{mm} - \Sigma_{mc}^2 \end{pmatrix} \tag{A.5}$$

using the determinant

$$\Delta = \det(A) = (\Sigma_{cc}\Sigma_{ss} - \Sigma_{cs}^2)\Sigma_{mm} + 2\Sigma_{cs}\Sigma_{mc}\Sigma_{ms} - \Sigma_{mc}^2\Sigma_{ss} - \Sigma_{cc}\Sigma_{ms}^2.$$

The corresponding covariance matrix K with variances and covariances of $\hat{\mathbf{x}}$ can be calculated by error propagation, leading to the following entries:

$$\begin{aligned} \hat{K}_{kl} &= \sum_i \left(\frac{\partial x_k}{\partial y_i} \right) \left(\frac{\partial x_l}{\partial y_i} \right) \sigma_i^2 \\ &= \sum_i a_{ki}^{-1} a_{li}^{-1} a_{ii} \\ &\quad + \sum_i \sum_{j \neq i} (a_{ki}^{-1} a_{lj}^{-1} + a_{kj}^{-1} a_{li}^{-1}) a_{ij} \end{aligned} \tag{A.6}$$

A.2 Bias estimation of the linearized signal estimation

A.2.1 Derivatives

A.2.1.1 Transmission

$$\hat{T} = \hat{M} \tag{A.7}$$

$$\frac{\partial \hat{T}}{\partial \hat{M}} = 1 \tag{A.8}$$

$$\frac{\partial^2 \hat{T}}{\partial \hat{M}^2} = 0 \tag{A.9}$$

A.2.1.2 Dark-field

$$\hat{D} = (\hat{C}^2 + \hat{S}^2)^{0.5} \hat{M}^{-1}$$

$$\frac{\partial \hat{D}}{\partial \hat{M}} = -(\hat{C}^2 + \hat{S}^2)^{0.5} \hat{M}^{-2} \tag{A.10}$$

$$\frac{\partial \hat{D}}{\partial \hat{C}} = (\hat{C}^2 + \hat{S}^2)^{-0.5} \hat{M}^{-1} C \tag{A.11}$$

$$\frac{\partial \hat{D}}{\partial \hat{S}} = (\hat{C}^2 + \hat{S}^2)^{-0.5} \hat{M}^{-1} S \tag{A.12}$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{M}^2} = 2(\hat{C}^2 + \hat{S}^2)^{0.5} \hat{M}^{-3} \tag{A.13}$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{C}^2} = \hat{S}^2 (\hat{C}^2 + \hat{S}^2)^{-1.5} \hat{M}^{-1} \quad (\text{A.14})$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{S}^2} = \hat{C}^2 (\hat{C}^2 + \hat{S}^2)^{-1.5} \hat{M}^{-1} \quad (\text{A.15})$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{M} \partial \hat{C}} = -(\hat{C}^2 + \hat{S}^2)^{-0.5} \hat{M}^{-2} C \quad (\text{A.16})$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{M} \partial \hat{S}} = -(\hat{C}^2 + \hat{S}^2)^{-0.5} \hat{M}^{-2} S \quad (\text{A.17})$$

$$\frac{\partial^2 \hat{D}}{\partial \hat{C} \partial \hat{S}} = -(\hat{C}^2 + \hat{S}^2)^{-1.5} \hat{M}^{-1} CS \quad (\text{A.18})$$

A.2.1.3 Phase

$$\hat{\phi} = \arctan \frac{\hat{C}}{\hat{S}} \quad (\text{A.19})$$

$$\frac{\partial \hat{\phi}}{\partial \hat{C}} = \hat{S} (\hat{C}^2 + \hat{S}^2)^{-1} \quad (\text{A.20})$$

$$\frac{\partial \hat{\phi}}{\partial \hat{S}} = -\hat{C} (\hat{C}^2 + \hat{S}^2)^{-1} \quad (\text{A.21})$$

$$\frac{\partial^2 \hat{\phi}}{\partial \hat{C}^2} = -2\hat{C}\hat{S}(\hat{C}^2 + \hat{S}^2)^{-2} \quad (\text{A.22})$$

$$\frac{\partial^2 \hat{\phi}}{\partial \hat{S}^2} = 2\hat{C}\hat{S}(\hat{C}^2 + \hat{S}^2)^{-2} \quad (\text{A.23})$$

$$\frac{\partial^2 \hat{\phi}}{\partial \hat{C} \partial \hat{S}} = (\hat{C}^2 - \hat{S}^2)(\hat{C}^2 + \hat{S}^2)^{-2} \quad (\text{A.24})$$

A.3 Bias corrected maximum likelihood estimator

To calculate the Cramer-Rao lower bound and the first-order bias corrected maximum-likelihood estimator the Fisher information and various other derivatives and functions of the maximum-

likelihood must be calculated in the following.

$$\begin{aligned} L(T, D, \phi) &= \frac{1}{2} \sum_i (Tm_i + TD \cos(\phi)c_i + TD \sin(\phi)s_i - y_i)^2 w_i \\ &= \frac{1}{2} \sum_i r_i^2 w_i \end{aligned} \quad (\text{A.25})$$

$$\frac{\partial L(T, D, \phi)}{\partial T} = \sum_i r_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) w_i \quad (\text{A.26})$$

$$\frac{\partial L(T, D, \phi)}{\partial D} = \sum_i r_i (T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \quad (\text{A.27})$$

$$\frac{\partial L(T, D, \phi)}{\partial \phi} = \sum_i r_i (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \quad (\text{A.28})$$

$$\frac{\partial^2 L(T, D, \phi)}{\partial T^2} = \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)^2 w_i \quad (\text{A.29})$$

$$\frac{\partial^2 L(T, D, \phi)}{\partial D^2} = \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)^2 w_i \quad (\text{A.30})$$

$$\begin{aligned} \frac{\partial^2 L(T, D, \phi)}{\partial \phi^2} &= \sum_i (-TD \sin(\phi) c_i + TD \cos(\phi) s_i)^2 w_i \\ &\quad + r_i (-TD \cos(\phi) c_i + TD \sin(\phi) s_i) w_i \end{aligned} \quad (\text{A.31})$$

$$\begin{aligned} \frac{\partial^2 L(T, D, \phi)}{\partial T \partial D} &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \\ &\quad + r_i (\cos(\phi) c_i + \sin(\phi) s_i) w_i \end{aligned} \quad (\text{A.32})$$

$$\begin{aligned} \frac{\partial^2 L(T, D, \phi)}{\partial T \partial \phi} &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \\ &\quad + r_i (-D \sin(\phi) c_i + D \cos(\phi) s_i) w_i \end{aligned} \quad (\text{A.33})$$

$$\begin{aligned} \frac{\partial^2 L(T, D, \phi)}{\partial D \partial \phi} &= \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \\ &\quad + r_i (-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \end{aligned} \quad (\text{A.34})$$

$$\frac{\partial^3 L}{\partial T^3} = 0 \quad (\text{A.35})$$

$$\frac{\partial^3 L}{\partial D^3} = 0 \quad (\text{A.36})$$

$$\begin{aligned}
\frac{\partial^3 L}{\partial \phi^3} &= \sum_i 2 (-TD \sin(\phi) c_i - TD \cos(\phi) s_i) (-TD \cos(\phi) c_i + TD \sin(\phi) s_i) w_i \\
&\quad + (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \\
&\quad + r_i (TD \sin(\phi) c_i - TD \cos(\phi) s_i) w_i
\end{aligned} \tag{A.37}$$

$$\frac{\partial^3 L}{\partial T^2 \partial D} = \sum_i 2 (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(\cos(\phi) c_i + \sin(\phi) s_i) w_i \tag{A.38}$$

$$\frac{\partial^3 L}{\partial T^2 \partial \phi} = \sum_i 2(m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-D \sin(\phi) c_i + D \cos(\phi) s_i) w_i \tag{A.39}$$

$$\frac{\partial^3 L}{\partial D^2 \partial T} = \sum_i 2(T \cos(\phi) c_i + T \sin(\phi) s_i)(\cos(\phi) c_i + \sin(\phi) s_i) w_i \tag{A.40}$$

$$\frac{\partial^3 L}{\partial D^2 \partial \phi} = \sum_i 2(T \cos(\phi) c_i + T \sin(\phi) s_i)(-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \tag{A.41}$$

$$\begin{aligned}
\frac{\partial^3 L}{\partial \phi^2 \partial T} &= \sum_i 2(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (-D \sin(\phi) c_i + D \cos(\phi) s_i) w_i \\
&\quad + \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \\
&\quad + \sum_i r_i (-D \cos(\phi) c_i - D \sin(\phi) s_i) w_i
\end{aligned} \tag{A.42}$$

$$\begin{aligned}
\frac{\partial^3 L}{\partial \phi^2 \partial D} &= \sum_i 2(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \\
&\quad + \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i) (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \\
&\quad + \sum_i r_i (-T \cos(\phi) c_i - T \sin(\phi) s_i) w_i
\end{aligned} \tag{A.43}$$

$$\begin{aligned}
\frac{\partial^3 L}{\partial T \partial D \partial \phi} &= \sum_i (-D \sin(\phi) c_i + D \cos(\phi) s_i)(T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \\
&\quad + \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \\
&\quad + \sum_i (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (\cos(\phi) c_i + \sin(\phi) s_i) w_i \\
&\quad + \sum_i r_i (-\sin(\phi) c_i + \cos(\phi) s_i) w_i
\end{aligned} \tag{A.44}$$

Expected values

For the analytical calculation of the Fisher Information Matrix and the bias-corrected maximum likelihood estimator the expected values of various derivative terms are necessary. They can be easily calculated by considering the following trivial results, which are valid for consistent estimators and uncorrelated measurements:

$$\mathbb{E}[r_i] = 0 \quad (\text{A.45})$$

$$\mathbb{E}[r_i r_j] = 0 \quad (\text{A.46})$$

$$\mathbb{E}[r_i r_i] = \text{var}_i = \frac{1}{w_i}. \quad (\text{A.47})$$

With these follows for the expected values of the derivatives

$$\begin{aligned} \mathbb{E}\left[\frac{\partial L(T, D, \phi)}{\partial T}\right] &= 0 \\ \mathbb{E}\left[\frac{\partial L(T, D, \phi)}{\partial D}\right] &= 0 \\ \mathbb{E}\left[\frac{\partial L(T, D, \phi)}{\partial \phi}\right] &= 0 \end{aligned} \quad (\text{A.48})$$

$$\begin{aligned} \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial T^2}\right] &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)^2 w_i \\ \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial D^2}\right] &= \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)^2 w_i \\ \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial \phi^2}\right] &= \sum_i (-TD \sin(\phi) c_i + TD \cos(\phi) s_i)^2 w_i \\ \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial T \partial D}\right] &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \\ \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial T \partial \phi}\right] &= \sum_i (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \\ \mathbb{E}\left[\frac{\partial^2 L(T, D, \phi)}{\partial D \partial \phi}\right] &= \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)(-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \end{aligned} \quad (\text{A.49})$$

$$\begin{aligned}
\mathbb{E} \left[\frac{\partial^3 L(T, D, \phi)}{\partial T^3} \right] &= 0 \\
\mathbb{E} \left[\frac{\partial^3 L(T, D, \phi)}{\partial D^3} \right] &= 0 \\
\mathbb{E} \left[\frac{\partial^3 L(T, D, \phi)}{\partial \phi^3} \right] &= \sum_i 2 (-TD \sin(\phi) c_i - TD \cos(\phi) s_i)(-TD \cos(\phi)c_i + TD \sin(\phi)s_i) w_i \\
&\quad + (-TD \sin(\phi) c_i + TD \cos(\phi) s_i)(-TD \cos(\phi)c_i - TD \sin(\phi)s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial T^2 \partial D} \right] &= \sum_i 2 (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(\cos(\phi) c_i + \sin(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial T^2 \partial \phi} \right] &= \sum_i 2 (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-D \sin(\phi) c_i + D \cos(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial D^2 \partial T} \right] &= \sum_i 2 (T \cos(\phi) c_i + T \sin(\phi) s_i)(\cos(\phi) c_i + \sin(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial D^2 \partial \phi} \right] &= \sum_i 2 (T \cos(\phi) c_i + T \sin(\phi) s_i)(-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial \phi^2 \partial T} \right] &= \sum_i 2 (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (-D \sin(\phi) c_i + D \cos(\phi) s_i) w_i \\
&\quad + (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial \phi^2 \partial D} \right] &= \sum_i 2 (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \\
&\quad + (T \cos(\phi) c_i + T \sin(\phi) s_i) (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \\
\mathbb{E} \left[\frac{\partial^3 L}{\partial T \partial D \partial \phi} \right] &= \sum_i (-D \sin(\phi) c_i + D \cos(\phi) s_i)(T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \\
&\quad + (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i)(-T \sin(\phi) c_i + T \cos(\phi) s_i) w_i \\
&\quad + (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) (\cos(\phi) c_i + \sin(\phi) s_i) w_i
\end{aligned} \tag{A.50}$$

Several expectation values of multiplications of second and first derivatives are also necessary for the bias correction:

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T^2} \frac{\partial L}{\partial T} \right] = \mathbb{E} \left[\frac{\partial^2 L}{\partial T^2} \frac{\partial L}{\partial D} \right] = \mathbb{E} \left[\frac{\partial^2 L}{\partial T^2} \frac{\partial L}{\partial \phi} \right] = 0 \tag{A.51}$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial D^2} \frac{\partial L}{\partial T} \right] = \mathbb{E} \left[\frac{\partial^2 L}{\partial D^2} \frac{\partial L}{\partial D} \right] = \mathbb{E} \left[\frac{\partial^2 L}{\partial D^2} \frac{\partial L}{\partial \phi} \right] = 0 \tag{A.52}$$

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 L}{\partial \phi^2} \frac{\partial L}{\partial T} \right] &= \mathbb{E} \left[\sum_{i,j} r_i r_j (-TD \cos(\phi) c_i - TD \sin(\phi) s_i)(m_j + D \cos(\phi) c_j + D \sin(\phi) s_j) w_i w_j \right] \\ &= \sum_i (-TD \cos(\phi) c_i - TD \sin(\phi) s_i)(m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) w_i \end{aligned} \quad (\text{A.53})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial \phi^2} \frac{\partial L}{\partial D} \right] = \sum_i (T \cos(\phi) c_i + T \sin(\phi) s_i)(-TD \cos(\phi) c_i - TD \sin(\phi) s_i) w_i \quad (\text{A.54})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial \phi^2} \frac{\partial L}{\partial \phi} \right] = \sum_i (-TD \cos(\phi) c_i - TD \sin(\phi) s_i) (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \quad (\text{A.55})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial D} \frac{\partial L}{\partial T} \right] = \sum_i (\cos(\phi) c_i + \sin(\phi) s_i) (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) w_i \quad (\text{A.56})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial D} \frac{\partial L}{\partial D} \right] = \sum_i (\cos(\phi) c_i + \sin(\phi) s_i) (T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \quad (\text{A.57})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial D} \frac{\partial L}{\partial \phi} \right] = \sum_i (\cos(\phi) c_i + \sin(\phi) s_i) (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \quad (\text{A.58})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial \phi} \frac{\partial L}{\partial T} \right] = \sum_i (-D \sin(\phi) c_i + D \cos(\phi) s_i) (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) w_i \quad (\text{A.59})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial \phi} \frac{\partial L}{\partial D} \right] = \sum_i (-D \sin(\phi) c_i + D \cos(\phi) s_i) (T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \quad (\text{A.60})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial T \partial \phi} \frac{\partial L}{\partial \phi} \right] = \sum_i (-D \sin(\phi) c_i + D \cos(\phi) s_i) (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \quad (\text{A.61})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial D \partial \phi} \frac{\partial L}{\partial T} \right] = \sum_i (-T \sin(\phi) c_i + T \cos(\phi) s_i) (m_i + D \cos(\phi) c_i + D \sin(\phi) s_i) w_i \quad (\text{A.62})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial D \partial \phi} \frac{\partial L}{\partial D} \right] = \sum_i (-T \sin(\phi) c_i + T \cos(\phi) s_i) (T \cos(\phi) c_i + T \sin(\phi) s_i) w_i \quad (\text{A.63})$$

$$\mathbb{E} \left[\frac{\partial^2 L}{\partial D \partial \phi} \frac{\partial L}{\partial \phi} \right] = \sum_i (-T \sin(\phi) c_i + T \cos(\phi) s_i) (-TD \sin(\phi) c_i + TD \cos(\phi) s_i) w_i \quad (\text{A.64})$$

Appendix B

Intensity-based statistical iterative reconstruction

B.1 Curvatures

B.1.1 Hessian matrix

In the following the derivatives of the per pixel cost functions h_i , necessary to calculate the Hessian matrix in each voxel will be given:

$$h_i = \frac{1}{2} \left(\underbrace{I_i - \underbrace{I_i^r e^{-l_i^\mu}}_{I_i^M} - \underbrace{I_i^r e^{-l_i^\mu} V_i^r e^{-l_i^\epsilon}}_{I_i^A} \cos(\underbrace{\phi_i^r + l_i^{\delta D}}_{p_i})}_{res_i} \right)^2 w_i \quad (\text{B.1})$$

$$\frac{\partial h_i}{\partial l_i^\mu} = res_i (I_i^M + I_i^A \cos(p_i)) w_i \quad (\text{B.2})$$

$$\frac{\partial h_i}{\partial l_i^\epsilon} = res_i I_i^A \cos(p_i) w_i \quad (\text{B.3})$$

$$\frac{\partial h_i}{\partial l_i^{\delta^D}} = res_i I_i^A \sin(p_i) w_i \quad (\text{B.4})$$

$$\frac{\partial^2 h_i}{\partial l_i^{\mu^2}} = (I_i^M + I_i^A \cos(p_i))^2 w_i - res_i (I_i^M + I_i^A \cos(p_i)) w_i \quad (\text{B.5})$$

$$\frac{\partial^2 h_i}{\partial l_i^{\epsilon^2}} = I_i^{A^2} \cos^2(p_i) w_i - res_i I_i^A \cos(p_i) w_i \quad (\text{B.6})$$

$$\frac{\partial^2 h_i}{\partial l_i^{\delta^{D^2}}} = I_i^{A^2} \sin^2(p_i) w_i + res_i I_i^A \cos(p_i) w_i \quad (\text{B.7})$$

$$\frac{\partial^2 h_i}{\partial l_i^\mu \partial l_i^\epsilon} = (I_i^M + I_i^A \cos(p_i)) I_i^A \cos(p_i) w_i - res_i I_i^A \cos(p_i) w_i \quad (\text{B.8})$$

$$\frac{\partial^2 h_i}{\partial l_i^\mu \partial l_i^{\delta^D}} = (I_i^M + I_i^A \cos(p_i)) I_i^A \sin(p_i) w_i - res_i I_i^A \sin(p_i) w_i \quad (\text{B.9})$$

$$\frac{\partial^2 h_i}{\partial l_i^\epsilon \partial l_i^{\delta^D}} = I_i^{A^2} \cos(p_i) \sin(p_i) w_i - res_i I_i^A \sin(p_i) w_i \quad (\text{B.10})$$

B.1.2 Precomputed curvature

In the following the per voxel terms for the pre-computed curvature will be given:

$$\begin{aligned} \frac{\partial^2}{\partial \mu_k^2} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \mu_k^2} a_{ik} \sum_j a_{ij} \\ &\approx \sum_i (I_i^M + I_i^A \cos(p_i))^2 w_i a_{ik} \sum_j a_{ij} = \sum_i y_i^2 w_i a_{ik} \sum_j a_{ij} \end{aligned} \quad (\text{B.11})$$

$$\begin{aligned} \frac{\partial^2}{\partial \epsilon_k^2} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \epsilon_k^2} a_{ik} \sum_j a_{ij} \\ &\approx \sum_i 0.5 I_i^{A^2} w_i a_{ik} \sum_j a_{ij} \end{aligned} \quad (\text{B.12})$$

$$\begin{aligned} \frac{\partial^2}{\partial \delta_k^2} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \delta_k^2} a_{ik} \sum_j a_{ij} \\ &\approx \sum_i 0.5 I_i^{A^2} w_i a_{ik} \sum_j a_{ij} \end{aligned} \quad (\text{B.13})$$

$$\begin{aligned} \frac{\partial^2}{\partial \mu_k \partial \epsilon_k} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \mu_k \partial \epsilon_k} a_{ik} \sum_j a_{ij} \\ &\approx \sum_i 0.5 I_i^{A^2} w_i a_{ik} \sum_j a_{ij} \end{aligned} \quad (\text{B.14})$$

$$\begin{aligned} \frac{\partial^2}{\partial \mu_k \partial \delta_k} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \mu_k \partial \delta_k} a_{ik} \sum_j a_{ij} \\ &\approx 0 \end{aligned} \quad (\text{B.15})$$

$$\begin{aligned} \frac{\partial^2}{\partial \epsilon_k \partial \delta_k} Q_{SPS}(\theta; \theta^{(n)}) &= \sum_i \frac{\partial^2 h_i^{(n)}}{\partial \epsilon_k \partial \delta_k} a_{ik} \sum_j a_{ij} \\ &\approx 0 \end{aligned} \quad (\text{B.16})$$

Bibliography

- [1] W. C. Rontgen, “On a New Kind of Rays,” *Science*, vol. 3, no. 59, pp. 227–231, 1896.
- [2] G. N. Hounsfield, “Computerized transverse axial scanning (tomography): Part 1. Description of system,” *British Journal of Radiology*, vol. 46, no. 552, pp. 1016–1022, 1973.
- [3] M. Beister, D. Kolditz, and W. A. Kalender, “Iterative reconstruction methods in X-ray CT,” *Physica medica*, vol. 28, pp. 94–108, 4 2012.
- [4] A. Bravin, P. Coan, and P. Suortti, “X-ray phase-contrast imaging: from pre-clinical applications towards clinics,” *Physics in medicine and biology*, vol. 58, no. 1, pp. 1–35, 2013.
- [5] T. Köhler, B. Brendel, and E. Roessl, “Iterative reconstruction for differential phase contrast imaging using spherically symmetric basis functions,” *Medical Physics*, vol. 38, no. 8, p. 4542, 2011.
- [6] X. Li and S. Luo, “A compressed sensing-based iterative algorithm for CT reconstruction and its possible application to phase contrast imaging,” *BioMedical Engineering OnLine*, vol. 10, p. 73, 1 2011.
- [7] Q. Xu, E. Sidky, and X. Pan, “Investigation of discrete imaging models and iterative image reconstruction in differential X-ray phase-contrast tomography,” *Optics Express*, vol. 20, no. 10, pp. 278–289, 2012.

- [8] T. Gaass, G. Potdevin, M. Bech, P. B. Noël, M. Willner, A. Tapfer, F. Pfeiffer, and A. Haase, “Iterative reconstruction for few-view grating-based phase-contrast CT - An in vitro mouse model,” *Europhysics Letters*, vol. 102, p. 48001, 5 2013.
- [9] J. Fu, S. Schleede, R. Tan, L. Chen, M. Bech, K. Achterhold, M. Gifford, R. Loewen, R. Ruth, and F. Pfeiffer, “An algebraic iterative reconstruction technique for differential X-ray phase-contrast computed tomography,” *Zeitschrift fuer Medizinische Physik*, 2013.
- [10] A. Ritter, F. Bayer, J. Durst, K. Gödel, W. Haas, T. Michel, J. Rieger, T. Weber, L. Wucherer, and G. Anton, “Simultaneous maximum-likelihood reconstruction for x-ray grating based phase-contrast tomography avoiding intermediate phase retrieval,” *arXiv preprint arXiv:1307.7912*, 7 2013.
- [11] D. Hahn, *Statistical Iterative Reconstruction for X-ray Phase-Contrast Computed Tomography*. PhD thesis, Technische Universität München, 2014.
- [12] D. Hahn, P. Thibault, A. Fehringer, M. Bech, T. Koehler, F. Pfeiffer, and P. B. Noël, “Statistical iterative reconstruction algorithm for X-ray phase-contrast CT,” *Scientific Reports*, vol. 5, p. 10452, 2015.
- [13] N. Bevins, J. Zambelli, K. Li, Z. Qi, and G. Chen, “Multicontrast x-ray computed tomography imaging using Talbot-Lau interferometry without phase stepping,” *Medical physics*, vol. 39, pp. 424–8, 1 2012.
- [14] H. F. Talbot, “LXXVI. Facts relating to optical science. No. IV.,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 9, no. 56, pp. 401–407, 1836.
- [15] R. F. Edgar, “The Fresnel Diffraction Images of Periodic Structures,” *Optica Acta*, vol. 16, no. 3, pp. 281–287, 1969.
- [16] T. J. Suleski, “Generation of Lohmann images from binary-phase Talbot array illuminators,” *Applied Optics*, vol. 36, no. 20, p. 4686, 1997.

- [17] H. H. Wen, E. E. Bennett, R. Kopace, A. F. Stein, and V. Pai, “Single-shot x-ray differential phase-contrast and diffraction imaging using two-dimensional transmission gratings,” *Optics Letters*, vol. 35, no. 12, p. 1932, 2010.
- [18] K. S. Morgan, D. M. Paganin, and K. K. W. Siu, “Quantitative single-exposure x-ray phase contrast imaging using a single attenuation grid,” *Optics Express*, vol. 19, no. 20, p. 19781, 2011.
- [19] M. Kagias, Z. Wang, P. Villanueva-Perez, K. Jefimovs, and M. Stampanoni, “2D-Omnidirectional Hard-X-Ray Scattering Sensitivity in a Single Shot,” *Physical Review Letters*, vol. 116, no. 9, pp. 1–5, 2016.
- [20] M. Bech, *X-ray imaging with a grating interferometer*. PhD thesis, University of Copenhagen, 2009.
- [21] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, “Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources,” *Nature physics*, vol. 2, pp. 258–261, 3 2006.
- [22] W. Yashiro, Y. Terui, K. Kawabata, and A. Momose, “On the origin of visibility contrast in x-ray Talbot interferometry,” *Optics Express*, vol. 18, no. 16, pp. 16890–16901, 2010.
- [23] W. Yashiro and A. Momose, “Effects of unresolvable edges in grating-based X-ray differential phase imaging,” *Optics Express*, vol. 23, no. 7, p. 9233, 2015.
- [24] OECD, “Computed tomography (CT) exams (indicator),” 2019.
- [25] UNEP, *Radiation Effects and Sources*. United Nations Environment Programme, 2016.
- [26] J. Radon, “Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten,” *Classic papers in modern diagnostic radiology*, vol. 5, p. 21, 2005.
- [27] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer, 2008.

- [28] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. IEEE Press, 1988.
- [29] C. E. Shannon, “Communication in the Presence of Noise,” *Proceedings of the IRE*, 1949.
- [30] E. Jaynes, *Probability: The Logic of Science*. Cambridge University Press, 2003.
- [31] K. Mechlem, *Advanced Statistical Iterative Reconstruction for X-ray Computed Tomography*. PhD thesis, Technical University of Munich, 2015.
- [32] H. Zhang, J. Wang, D. Zeng, X. Tao, and J. Ma, “Regularization strategies in statistical image reconstruction of low-dose x-ray CT: A review,” *Medical Physics*, vol. 45, no. 10, pp. e886–e907, 2018.
- [33] J. Fessler, “Analytical approach to regularization design for isotropic spatial resolution,” *2003 IEEE Nuclear Science Symposium, Conference Record*, vol. 3, pp. 2022–2026, 2003.
- [34] C. Bouman and K. Sauer, “A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation,” *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, 1993.
- [35] H. Erdogan, *Statistical Image Reconstruction Algorithms Using Paraboloidal Surrogates for PET Transmission Scans by Surrogates for PET Transmission Scans*. PhD thesis, University of Michigan, 1999.
- [36] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, “A three-dimensional statistical approach to improved image quality for multislice helical CT,” *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, 2007.
- [37] D. Kim, S. Ramani, and J. A. Fessler, “Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 167–178, 2015.
- [38] M. Bech, O. Bunk, T. Donath, R. Feidenhans'l, C. David, and F. Pfeiffer, “Quantitative x-ray dark-field computed tomography,” *Physics in Medicine and Biology*, vol. 55, no. 18, pp. 5529–5539, 2010.

- [39] Z. T. Wang, K. J. Kang, Z. F. Huang, and Z. Q. Chen, “Quantitative grating-based x-ray dark-field computed tomography,” *Applied Physics Letters*, vol. 95, no. 9, pp. 3–6, 2009.
- [40] P. Modregger, R. Bellotti, C. David, M. Stampanoni, B. R. Pinzer, and F. Scattarella, “Imaging the Ultrasmall-Angle X-Ray Scattering Distribution with Grating Interferometry,” *Physical Review Letters*, vol. 108, no. 4, pp. 2–5, 2012.
- [41] P. Modregger, S. Rutishauser, J. Meiser, C. David, and M. Stampanoni, “Two-dimensional ultra-small angle X-ray scattering with grating interferometry,” *Applied Physics Letters*, vol. 105, no. 2, 2014.
- [42] A. Malecki, G. Potdevin, T. Biernath, E. Eggl, E. Grande Garcia, T. Baum, P. B. Noël, J. S. Bauer, and F. Pfeiffer, “Coherent Superposition in Grating-Based Directional Dark-Field Imaging,” *PLoS ONE*, vol. 8, no. 4, p. e61268, 2013.
- [43] A. Malecki, G. Potdevin, T. Biernath, E. Eggl, K. Willer, T. Lasser, J. Maisenbacher, J. Gibmeier, A. Wanner, and F. Pfeiffer, “X-ray tensor tomography,” *EPL*, vol. 105, no. 3, 2014.
- [44] M. Strobl, “General solution for quantitative dark-field contrast imaging with grating interferometers,” *Scientific reports*, vol. 4, p. 7243, 1 2014.
- [45] F. Pfeiffer, C. Kottler, O. Bunk, and C. David, “Hard X-Ray Phase Tomography with Low-Brilliance Sources,” *Physical Review Letters*, vol. 98, p. 108105, 3 2007.
- [46] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
- [47] A. De Pierro, “On the Convergence of the Iterative Image Space Reconstruction Algorithm for Volume ECT,” *IEEE Transactions on Medical Imaging*, vol. 6, no. 2, pp. 174–175, 1987.
- [48] K. Lange and J. A. Fessler, “Globally Convergent Algorithms for Maximum a Posteriori Transmission Tomography,” *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1430–38, 1995.

- [49] J. Fessler and H. Erdogan, “A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction,” *1998 IEEE Nuclear Science Symposium Conference Record*, vol. 2, 1998.
- [50] K. Lange, *MM Optimization Algorithms*. Society for Industrial and Applied Mathematics, 2016.
- [51] M. W. Jacobson and J. A. Fessler, “An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms,” *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2411–2422, 2007.
- [52] J. de Leeuw and K. Lange, “Sharp quadratic majorization in one dimension,” *Computational Statistics and Data Analysis*, vol. 53, no. 7, pp. 2471–2484, 2009.
- [53] A. D. Pierro, “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,” *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1995.
- [54] D. Kim, D. Pal, J. B. Thibault, and J. A. Fessler, “Accelerating ordered subsets image reconstruction for X-ray CT using spatially nonuniform optimization transfer,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 11, pp. 1965–1978, 2013.
- [55] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. Brönnimann, C. Grünzweig, and C. David, “Hard-X-ray dark-field imaging using a grating interferometer,” *Nature materials*, vol. 7, pp. 134–7, 2 2008.
- [56] F. Pfeiffer, M. Bech, O. Bunk, T. Donath, B. Henrich, P. Kraft, and C. David, “X-ray dark-field and phase-contrast imaging using a grating interferometer,” *Journal of Applied Physics*, vol. 105, no. 10, 2009.
- [57] C. J. Morgan, “Least-squares estimation in phase-measurement interferometry,” *Optics Letters*, 1982.

- [58] X. Y. Ji, Y. Ge, R. Zhang, K. Li, and G.-H. Chen, “Studies of signal estimation bias in grating-based x-ray multicontrast imaging,” *Medical physics*, vol. 44, no. 6, pp. 2453–2465, 2017.
- [59] D. R. Cox and E. J. Snell, “A General Definition of Residuals,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 248–275, 1968.
- [60] I. Zanette, M. Bech, F. Pfeiffer, and T. Weitkamp, “Interlaced phase stepping in phase-contrast x-ray tomography,” *Applied Physics Letters*, vol. 98, no. 9, p. 094101, 2011.
- [61] I. Zanette, M. Bech, and A. Rack, “Trimodal low-dose X-ray tomography,” *Proceedings of the National Academy of Sciences*, vol. 109, pp. 10199–204, 6 2012.
- [62] P. Green, “Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternative,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 2, pp. 149–192, 1984.
- [63] P. J. Huber, *Robust statistics*. John Wiley and Sons Ltd, 1981.
- [64] A. Bovik, *Handbook of Image and Video Processing*. Elsevier Science, 2010.
- [65] [Http://www.imp.uni-erlangen.de/phantoms](http://www.imp.uni-erlangen.de/phantoms), “Forbild.”
- [66] B. Brendel, M. von Teuffenbach, P. B. Noël, F. Pfeiffer, and T. Koehler, “Penalized maximum likelihood reconstruction for x-ray differential phase-contrast tomography,” *Medical Physics*, vol. 43, no. 1, pp. 188–194, 2016.
- [67] A. Ritter, G. Anton, and T. Weber, “Simultaneous maximum-likelihood reconstruction of absorption coefficient, refractive index and dark-field scattering coefficient in x-ray talbot-lau tomography,” *PLoS ONE*, vol. 11, no. 10, pp. 1–15, 2016.
- [68] M. v. Teuffenbach, T. Koehler, A. Fehringer, M. Viermetz, B. Brendel, J. Herzen, R. Proksa, E. J. Rummeny, F. Pfeiffer, and P. B. Noël, “Grating-based phase-contrast and dark-field computed tomography: A single-shot method,” *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

- [69] M. von Teuffenbach, W. Noichl, B. Brendel, J. Herzen, F. Pfeiffer, T. Koehler, and P. B. Noël, “Grating Position Estimation for Grating-based Computed Tomography,” in *Fully3D 2017: Proceedings of the International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2017.
- [70] M. von Teuffenbach, B. Brendel, A. Fehringer, P. B. Noël, and T. Koehler, “Iterative reconstruction of grating-based PCCT without phase-stepping,” in *Proceedings of the 4th International Meeting on Image Formation in X-ray CT*, pp. 367–70, 2016.
- [71] B. Brendel, M. von Teuffenbach, P. B. Noël, F. Pfeiffer, and T. Koehler, “Intensity Based Iterative Reconstruction for Differential Phase Contrast Imaging with Reconstruction Parameter Estimation,” in *Fully3D 2015*, 2015.
- [72] S. Booth and J. Fessler, “Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography,” in *International Conference on Image Processing*, vol. 2, pp. 441–444, 1995.
- [73] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, “Preconditioning Methods for Improved Convergence Rates in Iterative Reconstructions,” *IEEE Transactions on Medical Imaging*, vol. 12, no. 1, pp. 78–83, 1993.
- [74] T. Weidinger, T. M. Buzug, T. Flohr, S. Kappler, and K. Stierstorfer, “Polychromatic Iterative Statistical Material Image Reconstruction for Photon-Counting Computed Tomography,” *International Journal of Biomedical Imaging*, 2016.
- [75] Y. Nesterov, “A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$,” *Proceedings of the USSR Academy of Sciences*, 1983.
- [76] P. Zhu, K. Zhang, Z. Wang, and Y. Liu, “Low-dose, simple, and fast grating-based X-ray phase-contrast imaging,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 31, pp. 13576–81, 2010.
- [77] A. Momose, W. Yashiro, S. Harasse, and H. Kuwabara, “Four-dimensional X-ray phase tomography with Talbot interferometry and white synchrotron radiation: dynamic observation of a living worm,” *Optics express*, vol. 19, no. 9, pp. 43–47, 2011.

- [78] H. Miao, L. Chen, E. E. Bennett, N. M. Adamo, A. A. Gomella, A. M. Deluca, A. Patel, N. Y. Morgan, and H. Wen, “Motionless phase stepping in X-ray phase contrast imaging with a compact source,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 48, pp. 19268–72, 2013.
- [79] L. Birnbacher, M. Willner, A. Velroyen, M. Marschner, A. Hipp, J. Meiser, F. Koch, T. Schröter, D. Kunka, J. Mohr, F. Pfeiffer, and J. Herzen, “Experimental Realisation of High-sensitivity Laboratory X-ray Grating-based Phase-contrast Computed Tomography,” *Scientific reports*, vol. 6, no. October 2015, p. 24022, 2016.
- [80] A. Tapfer, M. Bech, A. Velroyen, J. Meiser, J. Mohr, M. Walter, J. Schulz, B. Pauwels, P. Bruyndonckx, X. Liu, A. Sasov, and F. Pfeiffer, “Experimental results from a preclinical X-ray phase-contrast CT scanner,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 39, pp. 15691–6, 2012.
- [81] G. Pelzer, J. Rieger, C. Hauke, F. Horn, T. Michel, M. Seifert, and G. Anton, “Reconstruction method for grating-based x-ray phase-contrast images without knowledge of the grating positions,” *Journal of Instrumentation*, vol. 10, no. 12, 2015.
- [82] M. Seifert, S. Kaeppler, C. Hauke, F. Horn, G. Pelzer, J. Rieger, T. Michel, C. Riess, and G. Anton, “Optimisation of image reconstruction for phase-contrast x-ray Talbot-Lau imaging with regard to mechanical robustness,” *Physics in Medicine and Biology*, vol. 61, pp. 6441–6464, 8 2016.
- [83] P. Zhang, Q. Shao, R. Hu, S. Wang, G. Zan, S. Wang, Y. Bao, F. Wali, Y. Tian, G. Liu, Z. Wang, and K. Gao, “Correcting Intensity Drift in X-ray Grating-based Phase Contrast Imaging,” *Journal of Instrumentation*, vol. 11, no. 03, pp. C03055–C03055, 2016.
- [84] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing (2nd ed.)*. Cambridge University Press, 1992.
- [85] S. P. Raman, P. T. Johnson, S. Deshmukh, M. Mahesh, K. L. Grant, and E. K. Fishman, “CT dose reduction applications: Available tools on the latest generation of CT scanners,” *Journal of the American College of Radiology*, vol. 10, no. 1, pp. 37–41, 2013.

- [86] T. Resnick and W. Dupuis, “X-Ray tube current control, US4311913A,” 1982.
- [87] T. Resnick and R. Mattson, “Reduction of temporal variations in X-ray radiation, US6215842B1,” 1998.
- [88] L. Fu, J. Wang, X. Rui, J. B. Thibault, and B. De Man, “Modeling and estimation of detector response and focal spot profile for high-resolution iterative CT reconstruction,” *IEEE Nuclear Science Symposium Conference Record*, no. 1, 2013.
- [89] D. Kim and J. A. Fessler, “Optimized first-order methods for smooth convex minimization,” *Mathematical Programming*, vol. 159, pp. 81–107, 2016.

List of publications and scientific presentations

First-authored publications

M. von Teuffenbach, T. Koehler, A. Fehringer, M. Viermetz, B. Brendel, J. Herzen, R. Proksa, E. J. Rummeny, F. Pfeiffer, P. B. Noël, "Grating-based phase-contrast and dark-field computed tomography: a single-shot method," Scientific Reports, vol. 7 (1), 2017.

Co-authored publications

S. Umkehrer, L. Birnbacher, R. Burkhardt, M. von Teuffenbach, A. Ö. Yildirim, D. Pfeiffer, J. Herzen, F. Pfeiffer, "Optimization of in vivo murine X-ray dark-field computed tomography," Review of Scientific Instruments, vol. 90 (10), 2019.

L. Birnbacher, M. Viermetz, W. Noichl, S. Allner, A. Fehringer, M. Marschner, M. von Teuffenbach, M. Willner, K. Achterhold, P. B. Noël, T. Koehler, J. Herzen, F. Pfeiffer, "Tilted grating phase-contrast computed tomography using statistical iterative reconstruction," Scientific Reports, vol. 8 , 2018

B. Brendel, M. von Teuffenbach, P. B. Noël, F. Pfeiffer, T. Koehler, "Penalized maximum likelihood reconstruction for x-ray differential phase-contrast tomography," Medical Physics, vol. 43, 2015

Patent applications

R. Proksa, A. Fingerle, D. Muenzel, F. Pfeiffer, T. Koehler, P. Noel, M. von Teuffenbach, J. Herzen, K. Willer, "Dark-field enhanced virtual x-ray colonoscopy," WO/2017/046670 A1, 2017

T. Koehler, G. Martens, R. Proksa, H. I. Maack, U. van Stevendaal, F. Pfeiffer, P. Noël, M. von Teuffenbach, "Scanning x-ray apparatus with full-field detector," WO/2017/001294 A1, 2017

T. Koehler, B. J. Brendel, P. Noël, F. Pfeiffer, M. von Teuffenbach, "Robust reconstruction for dark-field and phase contrast CT," WO/2016/207423 A1, 2016

Oral presentations

"Iterative Reconstruction of Sliding Window PCCT", XNPIG 2015 (The 3rd Meeting on X-Ray and Neutron Phase Imaging with Gratings), Washington, D.C., USA, September 2015

"Iterative Reconstruction of Grating-based PCCT Without Phase-Stepping", CTMeeting 2016 (The 4th International Conference on Image Formation in X-Ray Computed Tomography), Bamberg, Bavaria, Germany, July 2016

"Internal Parameter Estimation for Grating-based Computed Tomography", Fully3D 2017 (The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine), Xi'an, Shaanxi, China, June 2017

"Grating Position Estimation for Grating-based Computed Tomography", XNPIG 2017 (The 4th Meeting on X-Ray and Neutron Phase Imaging with Gratings), Zürich, Switzerland, September 2017

Acknowledgements

This thesis would not have been possible without the help of many people. For all the personal and scientific support I got would like to thank

- my supervisor Prof. Franz Pfeiffer for initially sparking my interest in the field of X-ray imaging with his lectures, than having me in his research group and giving me the opportunity to do exciting work with grating-based X-ray imaging! The team you build up at E17 is great and makes scientific work as pleasant as possible! Thank you especially for providing such excellent technical equipment and always sufficient computing resources for my work. Additionally, I want to express my deep gratitude for the opportunity to visit so many great conferences and workshops all around the world!
- Julia Herzen and Peter Noël for providing much needed feedback during our regular meetings and many opportunities to test and optimize my algorithms on many different setups. Special thanks also for sharing so many insights into academic and healthcare politics.
- Tobias Lasser for fruitful discussions, advice on choosing the right algorithms, and for serving as a second examiner of my dissertation.
- Thomas Köhler for supporting my research since the beginning and keeping my work on the right track over all the years. Your scientific experience and our regular discussions always helped me identify the next step to the right solution for my problems.
- Roland Proksa for his trust in our work and his great support with all matters relating our collaboration with Philips.
- Martin Dierolf and Wolfgang Noichl for keeping such good care of our IT infrastructure and our computational servers. You are saving others so much time with your work, thank you!

- Andreas Fehringer for guiding me through the first months of my work with CT reconstruction and providing great CT projectors without which a lot of my work would not have been possible.
- Pierre Thibault for holding great lectures about signal processing and statistical methods in image processing and thus exciting me for this great field of research!
- Nelly de Leiris and Klaus Achterhold for never getting tired of my stupid questions regarding administrative work and organizational problems. You make the life of everyone in the chair so much easier!
- all my colleagues in the Reco-Crew, especially Mathias Marschner, Lorenz Birnbacher, Sebastian Allner, Wolfgang Noichl, Lorenz Hehn, and Korbinian Mechlem. You provided a great atmosphere in our office and without all the exciting in-depth discussions about our research ideas I would never have been able to build up so much knowledge about CT reconstruction and signal estimation. Thank you!
- Manuel Viermetz, Nicolai Gustschin, and Clemens Schmidt for the great work at the iCT. You are doing a great job and I am thrilled to see the results of your future work!
- Dieter and Ulrike Renker for the great dinners in your flat and gatherings in the beer-garden. Attending these meetings was always a very special moment for the social life at the chair!
- All the other people in E17 who made being part of this group such a great experience. Especially to Fabio de Marco, Simone Ferstl, Manuela Frank, Regine Gradl, Benedikt Günther, Alex Gustschin, Eva Prexl, Manuel Schultheiss, Thorstan Sellerer, Stephan Umkehrer, Theresa Urban, Konstantin Willer, Jonathan Schock, Mark Müller, Sebastian Ehn, Kai Scherer, Marian Willner, and Lukas Gromann. It was such a great time working with you, going to conferences together, or just having fun as friends.

Finally, I want to thank my ever-supporting dearest friends, my family, and my girlfriend: Lukas, Moritz, Jonathan, Jakob, Mami, Papi, Valerie, and Kati. Without you, this work would never have been completed. Thank you for this and for everything else!