Spring 5-15-2020

Joint Estimation of Attenuation and Scatter for Tomographic Imaging with the Broken Ray Transform

Michael Ray Walker
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Department of Electrical & Systems Engineering

Dissertation Examination Committee:
Dr. Joseph A. O'Sullivan
Dr. R. Martin Arthur
Dr. Ulugbek S. Kamilov
Dr. David G. Politte
Dr. Yuan-Chuan Tai

Joint Estimation of Attenuation and Scatter for Tomographic Imaging with the Broken Ray Transform

by

Michael R. Walker II

A dissertation presented to the
McKelvey School of Engineering
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Science

May 2020
Saint Louis, Missouri

Contents

List of Figures

List of Tables

List of Algorithms

Acknowledgments

I would like to extend my deepest gratitude to my advisor, Professor Joseph A. O'Sullivan, for his sustained investment in my success. This extends far beyond publications. Through changes to my family and professional life, he has taken time to listen to me, encourage me, and find a path forward. It is a privilege to benefit from his breadth of knowledge and experience avoiding the minutiae and identifying areas for fruitful research.

I would like to thank Professor R. Martin Arthur for many enlightening conversations. Working as a teaching assistant was a valuable experience. I appreciate the opportunity and support.

Thank you to my committee members Professor Ulugbek S. Kamilov, Professor David G. Politte, and Professor Yuan-Chuan Tai. Constructive criticism and insightful comments have significantly improved the quality of this research.

Thank you to Professor John C Schotland for bringing the broken ray transform to our attention. His perspective and early feedback on our approach were particularly helpful.

I had many good conversations with Dr. Sergey A. Komarov about both the theory and potential experimental implementations of the measurement system.

I appreciate the mentorship of Dr. Michael P. Hurst and Dr. David W. Rieken, applied researchers from industry who have helped me envision the benefits of indulging intellectual curiosity. I would also like to acknowledge the encouragement of Dr. James M. Bornholdt, Mr. Nathan Kastelein, Dr. Paul F. Rebillot III, Dr. İlker Tunay, and Mr. Peter Wittenberg.

I cannot begin to express my thanks to my wife, Erin. She has tackled many challenges while leaving space for me to reach this accomplishment. Without her encouragement and support this would not have been possible. Also, her love of language has helped me become a better communicator.

Finally, I would like to thank my children, Oliver and Charlotte, for tolerating me spending so much time "staring at that alien face." Now, let's play.

Michael R. Walker II

Washington University in Saint Louis

May 2020

Dedicated to my family: Erin, Oliver & Charlotte.

ABSTRACT OF THE DISSERTATION

Joint Estimation of Attenuation and Scatter for Tomographic Imaging with the Broken Ray Transform

by

Michael R. Walker II

Doctor of Science in Electrical Engineering

Washington University in St. Louis, 2020

Professor Dr. Joseph A. O'Sullivan, Chair

The single-scatter approximation is fundamental for many tomographic imaging problems. This class broadly includes x-ray scattering imaging and optical scatter imaging for certain media. In all cases, noisy measurements are affected by both local events and nonlocal attenuation. Related applications typically focus on reconstructing one of two images: scatter density or total attenuation. However, both images are media specific. Both images are useful for object identification. Knowledge of one image aides estimation of the other, especially when estimating images from noisy data.

Joint image recovery has been demonstrated analytically in the context of the broken ray transform (BRT) for attenuation and scatter-density images. The BRT summarizes the nonlocal affects of attenuation in single-scatter measurement geometries. We find BRT analysis particularly interesting as joint image recovery has been demonstrated analytically using only two scatter angles. Limiting observations to two scatter angles is significant because it supports joint reconstruction in two dimensions for anisotropic scatter modalities

(e.g. Bragg, Compton). However, all analytic inversion strategies share two fundamental assumptions limiting their utility: nonzero scatter everywhere, and a deterministic data model.

There are two themes to our work. First, we consider the BRT in a purely deterministic setting. We are the first to recognize the BRT as a linear shift-invariant operator. This linear-systems perspective motivates frequency-domain analysis both of the data and operator. Frequency-domain representations provide new insights on the operator and a common framework for contrasting recent inversion formulas. New algorithms are presented for regularized inversion of the BRT in addition to fast forward and adjoint operators. Second, we incorporate the BRT in a stochastic data model. Approximating the detectors as photon counting processes, we model the data as Poisson distributed. Our iterative algorithm, alternating scatter and attenuation image updates, guarantees monotonic reduction of the regularized log-likelihood function of the data. We are the first to consider joint image estimation from noisy data. Our results demonstrate a significant improvement over analytic methods for data sets with missing data (regions with zero scatter). In addition to joint image estimation, our approach can be specialized for single image estimation. With known attenuation, we can improve the quality of scatter image estimates. Similarly, with known scatter, we can improve the quality of attenuation image estimates.

Through analysis and simulations, we highlight challenges for attenuation image estimation from BRT data, and ambiguity in the joint image recovery problem. Performance will vary with scaling of the problem. Total attenuation, detected counts, and scatter angle all affect the quality of image estimates. We are the first to incorporate both scatter density and attenuation in noisy data models. Our results demonstrate the benefits of accounting for both images, and should inform design of future measurement systems.

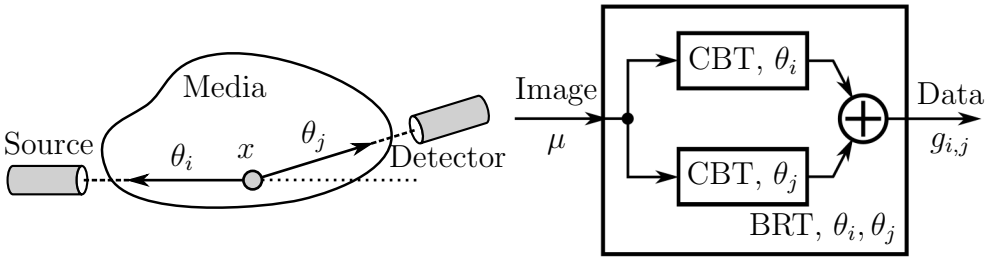
Chapter 1

Introduction

1.1 Background and Motivation

The broken ray transform (BRT) appears in the forward model of a number of imaging modalities and measurement geometries. It was first considered in the context of optical scatter imaging [2], later applied to x-ray scatter imaging [3], and appears whenever the single-scatter approximation holds. Under the single-scatter approximation, the mean photon path between a source and detector includes a single scatter event. Geometrically this path can be described by two rays sharing a common vertex. The integral along this path describes the nonlocal effects of the attenuation image. This path integral is referred to as the BRT. The BRT has been considered for both translation-only measurement geometries [1–8] and rotate-shift measurement geometries [7, 9, 10].

To our knowledge, all applications associated with the single-scatter BRT represent joint reconstruction problems. Two spatially-varying images must be resolved. For example, it may be necessary to recover attenuation despite nonuniform scatter density [3, 4], or recover two attenuation images at distinct energy levels [11]. Variations in the joint reconstruction problem have motivated several novel contributions related to the BRT. These contributions are not strictly academic. In application, the forward models must be tailored to the joint reconstruction problem. For example, scatter-density images that are not strictly positive will limit available or useful data. The joint reconstruction problem determines available BRT inversion strategies.



(a) Measurement gemoetry (b) System model

Figure 1.1: The generalized measurement geometry is depicted in Figure 1.1a. The location of the scattering event is indicated by x. The directions θ_i, and θ_j represent the direction of the source and detector relative to x, respectively. Indexing the data by the scatter location, x, both the BRT(1.5) and CBT(1.4) are linear shift-invariant. In Figure 1.1b we depict the forward operator as a linear system. The system operates on an image, μ, and returns data $g_{i,j}$.

Before contrasting prior work, we first define a notional measurement geometry. We will use the geometry to establish some notation and define a joint reconstruction problem involving the BRT. We generalize the joint reconstruction problem to cover coherent-scatter x-ray imaging. This modality has received renewed interest recently; however joint reconstruction of scatter density and attenuation has not yet been addressed [12–14]. In coherent-scatter x-ray imaging, scatter density is highly sensitive to scatter angle. For this reason, we focus on BRT inversion using only two scatter angles.

As a simplification we assume a monochromatic x-ray pencil-beam incident upon some media of interest. At point $x \in \mathbb{R}^2$ the beam interacts with the media and scatters coherently. We use $\theta_i \in S^1$ to represent the direction of the source relative to the scatter location. The direction of the scattered photon is $\theta_j \in S^1$. We assume the scatter direction is known due to a collimated detector. The combination of a pencil-beam and collimated detector determine the scatter location x, which we assume is known precisely. This measurement geometry is depicted in Figure 1.1a.

The intensity signals measured at the detector largely depends on two media-specific attributes: the attenuation image, and the scatter-density image. The incident path is a straight line defined by θ_i, x. The loss in intensity along this path due to attenuation is

governed by Beer's law,

$$\frac{I}{I_0} = \exp\left(-\int_0^\infty \mu(x - t\theta_i)dt\right). \tag{1.1}$$

We use I_0 to represent the source intensity. Here I represents the incident beam intensity at the scatter location x before scattering occurs. We use $\mu(x) : \mathbb{R}^2 \to \mathbb{R}^+$ as the attenuation image representing both scatter and absorption. Intensity loss along the scatter path due to attenuation has a similar form and combines multiplicatively. For incoherent-scatter applications (e.g. fluorescence imaging) it may be necessary to distinguish the energy levels of the attenuation image before and after the scatter event. This has been investigated recently [11].

Even in homogeneous media, the intensity observed at the detector may vary with respect to scatter angle (e.g. $\theta_i \cdot \theta_j$) and energy level. For coherent-scatter imaging, the scatter density does not depend on these terms independently, but rather through Bragg's law [15]. This relationship is summarized by the so-called momentum transfer $q(s, E) : (-1, 1) \times \mathbb{R}^+ \to \mathbb{R}^+$

$$q(s, E) = 2\frac{E}{hc}\sqrt{\frac{1-s}{2}}. \tag{1.2}$$

Here h and c are the Planck's constant and the speed of light, respectively. This definition is unconventional as we have chosen to define it over the cosine of the scatter angle, s, rather than the scatter angle directly. Our definition of momentum transfer is related to other definitions in the literature as shown in Appendix A. Scatter intensity for inhomogeneous media varies both spatially and with respect to momentum transfer. We use $f(x, q) : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^+$ to represent the scatter-density image.

Combining the effects of attenuation and scatter density we arrive at the measurement function

$$p(x, \theta_i, \theta_j, E) = f\left(x, q\left(-\theta_i \cdot \theta_j, E\right)\right) \exp\left(-\int_0^\infty \mu(x + t\theta_i) + \mu(x + t\theta_j)dt\right). \tag{1.3}$$

In this expression we have omitted a number of terms necessary for accurate models of measured data. However, we assume the remaining terms are known multiplicative factors. Measured data can then be scaled to achieve this generalized form.

To simplify the notation we will make use of three common transforms. Borrowing the notation of Natterer [16], we define the cone beam transform (CBT) B of μ as

$$(B\mu)(x,\theta) := \int_0^\infty \mu(x+t\theta)dt. \tag{1.4}$$

This transform appears in (1.3). In particular, the generalized model includes the linear combination of two cone beam transforms sharing a common vertex. The combination is commonly referred to as the broken ray transform

$$(G\mu)(x,\theta_i,\theta_j) := (B\mu)(x,\theta_i) + (B\mu)(x,\theta_j). \tag{1.5}$$

Denoting the left side of (1.5) by data $g_{i,j}$, Figure 1.1b represents (1.5) as a linear system with component CBT operators (1.4).

In subsequent sections we will also make use of the 2D Radon transform

$$(R\mu)(v,\theta) := \int_{-\infty}^\infty \mu(v\theta^\perp + t\theta)dt. \tag{1.6}$$

Here $v \in \mathbb{R}^1$, and $\theta \in S^1$ represent the shift and rotate coordinates of the transform. We assume $\theta^\perp$ is uniquely defined by rotating θ counter-clockwise by $\pi/2$.

Using these transforms we can express the log of the measured data

$$\ln p(x,\theta_i,\theta_j,E) = \ln f\left(x, q\left(\theta_i\cdot\theta_j,E\right)\right) - (B\mu)(x,\theta_i) - (B\mu)(x,\theta_j) \tag{1.7a}$$

$$= \ln f\left(x, q\left(\theta_i\cdot\theta_j,E\right)\right) - (G\mu)(x,\theta_i,\theta_j). \tag{1.7b}$$

The BRT is not directly available in (1.7b). However, the term f can be canceled with differential measurements [4] even for inhomogenous media.

Given three scatter angles $\theta_i,\theta_j,\theta_k$ such that

$$\theta_i\cdot\theta_k = \theta_j\cdot\theta_k, \tag{1.8}$$

we have

$$\ln p(x,\theta_i,\theta_k,E) - \ln p(x,\theta_j,\theta_k,E) = -(B\mu)(x,\theta_i) + (B\mu)(x,\theta_j). \tag{1.9}$$

The condition (1.8) is only required when the scatter density is a function of momentum transfer. For some modalities, scatter density varies with respect to scatter angle according to a known function (e.g. Klein–Nishina). In such cases the data can be corrected and momentum transfer removed from (1.7b).

For clarification, we will refer to the right-hand side of (1.9) as the signed broken ray transform (SBRT) due to the sign change between CBTs. The SBRT is equivalent to the signed V-line transform [8]. Some authors have reserved their definition of the BRT for this later expression [3]. While either definition of the BRT assumes a linear combination of two CBTs sharing a common vertex, the distinction is important. Also, while positive images yield positive BRT data, SBRT data may be negative.

For tomographic imaging applications it is common to index the data according to the source and detector locations. In contrast our indexing is somewhat unconventional. In the context of the BRT, Katsevich and Krylov were the first to demonstrate the benefits of indexing the data by the scatter location [3]. Under this indexing schema, both the CBT and BRT are linear and shift-invariant (LSI). Linear systems analysis is therefore applicable to the CBT and BRT. Their relationship is depicted in Figure 1.1b. The linear-systems perspective is a central theme of our contribution as we are the first to consider the two-dimensional Fourier transform of the BRT. This perspective has benefits which we will demonstrate in subsequent sections.

Our focus is limited to 2D single-scatter imaging problems where scatter events are observed throughout the media of interest. This distinction is important because the terms broken ray transform and V-line transform have been used to describe a number of related problems. We distinguish BRT problems integrating over multiple reflections [17, 18] or integrating over multiple vertices [19, 20]. Some constrain the vertex locations along the perimeter of the measurement geometry [17, 18, 21]. This is generally motivated by the use of Compton cameras. In three dimensions this results in the cone transform [22], which we distinguish from the cone beam transform (1.4) [16] applicable to our measurement geometry.

The first analytic inversion formula for the BRT is due to Florescu et al. [1]. The global inversion formula requires only two scatter angles to recover the attenuation image in the presence of spatially varying scatter density. The inversion technique can be summarized as a three-step process. First, obtain the one-dimensional Fourier transform of the data. For

the second step, each frequency is considered independently. Solve the resulting complex, one-dimensional, bounded differential equation. Third, obtain the inverse one-dimensional Fourier transform across the solutions. Florescu's global inversion formula yields an exact reconstruction of images with bounded support. The coordinates used to index the data in the original derivation were not linear-shift invariant. An inversion formula was later derived using data indexed by the scatter location and generalized for higher dimensions [5].

The number of available scatter angles is a discriminating factor in selecting a BRT inversion strategy. A local inversion formula was discovered by Katsevich and Krylov requiring three unique scatter angles [3]. In contrast with prior results [1], their reconstructions demonstrated a significant reduction in artifacts. The three scatter-angle approach was later generalized for additional scatter angles and source locations [6]. While the attenuation map can be recovered locally, the recovery of the scatter-density image still requires global reconstruction of the attenuation image. Due to our interest in scatter image recovery, and the requirement (1.8) for coherent-scatter imaging, we focus on global 2D BRT inversion techniques using only two scatter angles.

The initial results by Florescu et al. contained significant artifacts even for trivial phantoms [1]. These artifacts were broadly attributed to the nonlocal effects of integration. Artifacts in initial results exhibited striations at three distinct angles. Two of these angles are associated with the incident and scatter directions (θ_i and θ_j). However, the general explanation did not directly address the third direction. This was later explored using micro-local analysis [7]. Sherson was the first to recognize $\theta_i + \theta_j$ as the direction of integration required for inversion.

Most recently a new inversion technique was developed by Ambartsoumian and Jebelli [8]. They used linear shift-invariant indexing of the data but did not employ the Fourier transform. They thoroughly and eloquently derived a new inversion technique by extending the Fundamental Theorem of Calculus to higher dimensions under a linear change of variables. They consider V-Line transformed (VLT) data defined by the linear combination of CBTs along multiple directions $\{\theta_i\}$. Integrating VLT data along the direction $\sum_i \theta_i$ yields the integral of the image over the faceted cone defined by $\{\theta_i\}$ and the common vertex x. This unbounded volume can be reduced to a parallelepiped by linearly combining samples of this integral. Weighting the results, one obtains a reconstruction of the image averaged over this volume. This approach leads to a wonderfully concise inversion formula. In two dimensions

they replace differentiation along the directions θ_i and θ_j with sample differences. Image recovery then requires integration along the direction $\theta_i + \theta_j$. The consequence is potential blurring over the resulting parallelogram. The blurring can be arbitrarily small for noise-free environments with high resolution data. For noisy data, the size of the parallelogram must be larger, which effects blurring in the reconstruction. Additionally, artifacts appear along the direction of integration $(\theta_i + \theta_j)$ [8].

1.2 Main Contributions

Our contributions in Chapter 2 focus on the forward model in a purely deterministic setting. This is a best-case scenario for image separation in the absence of noise. Analysis, in a deterministic setting, is also useful for exposing numerical challenges due to conditioning and sampling that are only exacerbated by noise. We are the first to recognize the BRT as a LSI operator when the data are sampled by the scatter location. While Katsevich and Krylov were the first to index the data by the scatter location [3], we are the first to recognize the shift invariance of the operator. Perhaps for this reason previous authors limited Fourier analysis to one of two dimensions [4, 5, 7, 9, 23]. We are the first to describe this operator purely in the frequency domain. Our linear systems perspective leads to several contributions.

- For positive images with bounded support, BRT data has unbounded support. We clarify the extent of samples required to recover images with bounded support, and present methods for extending sufficient BRT data indefinitely.
- Filtering can be used to bound support of the data for arbitrary bounded images. Bounding the data is critical for obtaining frequency-domain representations of sampled signals.
- The forward operator has a nontrivial nullspace due to zeros along a line in the Fourier space. The existence of this nullspace complicates reconstruction. Mitigation strategies include enforcing continuity in the frequency domain, or equivalently, enforcing boundary conditions on the reconstruction.

- A new BRT inversion formula is presented based on the two-dimensional Fourier representation of the BRT.
- The frequency-domain representation of the BRT operator provides a common framework for contrasting prior inversion formulas. Both the integral-differential inversion formula of Sherson [7] and blurred reconstruction formula based on cumulative distributions of Ambartsumian [8] can be described as specializations of our approach.
- Fast algorithms for the forward and adjoint BRT operators are presented. Our novel filtering technique ensures periodicity in the data space. Periodicity enables discrete Fourier representations of the data. The forward and adjoint transform can then be implemented in the frequency domain. We demonstrate a significant reduction in computation time over a direct (matrix product) implementation.

Analytic inversion strategies share two fundamental assumptions which limit their utility: nonzero scatter everywhere, and a deterministic data model. In general, any missing samples are assumed to be obscured by regions with high attenuation. Assuming high attenuation is not a reasonable approach for data comprising large regions with zero scatter. Regions with zero scatter are particularly problematic when their locations are not known *a priori*. While analysis of the BRT suggests joint recovery is possible, prior algorithms are insufficient. The contributions in Chapter 3 include the following.

- We are the first to propose an algorithm for joint attenuation and scatter image recovery with missing data from a single-scatter measurement geometry. The algorithm is based on a photon counting model for the data: measurements are Poisson distributed.
- Our iterative algorithm guarantees monotonic improvement of the regularized data log-likelihood function while alternating updates to the scatter and attenuation images.
- Our generalized joint reconstruction algorithm can be specialized for single image recovery. Known attenuation can improve scatter-density image estimation, and known scatter can improve attenuation image estimation.

Chapter 2

Analysis of the Broken Ray Transform

2.1 Background and Motivation

In the following we take a fresh look at the BRT as a linear shift-invariant operator. A linear systems perspective provides new insights on the transforms and tools for contrasting prior inversion formulas. More specifically, we demonstrate images with bounded support do not guarantee data with bounded support. We are the first to consider the minimum data required for reconstruction and techniques for bounding support of the data. The two-dimensional Fourier transform of the BRT operator highlights numerical challenges. It exhibits zeros along direction of integration $(\theta_i + \theta_j)$ [8]. The ensuing ambiguity can be resolved analytically with boundary conditions on the reconstructed image (e.g. 0 after subtracting the background level). However, boundary conditions do not address numerical sensitivity near zeros in the frequency domain. The poles in the forward operator, along the directions θ_i and θ_j, also present numerical challenges. We contrast recent work [8] against prior inversion techniques [1,5,7] as different strategies for addressing the poles in the forward operator. To mitigate numerical issues associated with the forward operator, we incorporate regularization in our frequency-domain inversion formula. Due to rotational invariance of the two-dimensional Fourier transform, our inversion formula supports arbitrary source and scatter directions (θ_i, θ_j). In contrast, prior analytic inversion strategies require computation of integrals and sample differences in the spatial domain that are only convenient when the directions are aligned with sampling axes (e.g. [8]).

To exploit the benefits of linear system analysis, we first derive the two-dimensional Fourier transform of the BRT data. Since the BRT is LSI, we expect the result to have a specific

form. We can decompose the Fourier transform of the data into the product of two terms: the Fourier transform of the image and the Fourier transform of the system function. From the Fourier transform of the system function, several insights are directly available.

We consider an absolutely integrable image with bounded support. We define a closed, bounded, convex set $C \subset \mathbb{R}^2$, which we use to window the image according to

$$\mu_C(x) = \begin{cases} \mu(x), & \text{for } x \in C \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

We use $\hat{\mu}_C(w)$ to represent the two-dimensional Fourier transform of the image. Since the BRT is simply a linear combination of CBTs, we first define the two-dimensional Fourier transform of the CBT data

$$\hat{b}_\theta(w) = \mathcal{F}^2\left\{(B\mu_C)(x,\theta)\right\} \tag{2.2a}$$

$$= \hat{\mu}_C(w)\left[\frac{-1}{i2\pi w\cdot\theta} + \frac{1}{2}\delta(w\cdot\theta)\right]. \tag{2.2b}$$

The details of this derivation are in Appendix C.1. The two-dimensional Fourier transform of the BRT data is therefore

$$\hat{g}_{i,j}(w) = \mathcal{F}^2\left\{(G\mu_C)(x,\theta_i,\theta_j)\right\} \tag{2.3a}$$

$$= \mathcal{F}^2\left\{(B\mu_C)(x,\theta_i) + (B\mu_C)(x,\theta_j)\right\} \tag{2.3b}$$

$$= \hat{\mu}_C(w)\left[\frac{-w\cdot(\theta_i+\theta_j)}{i2\pi\,(w\cdot\theta_i)\,(w\cdot\theta_j)} + \frac{1}{2}\delta\,(w\cdot\theta_i) + \frac{1}{2}\delta\,(w\cdot\theta_j)\right]. \tag{2.3c}$$

Indeed, this result can be decomposed into the product of two terms. The bracketed term represents the two-dimensional Fourier transform of the BRT system function. For convenience, we will frequently reference a portion of this term

$$\hat{h}_{i,j}(w) \coloneqq \frac{-w\cdot(\theta_i+\theta_j)}{i2\pi\,(w\cdot\theta_i)\,(w\cdot\theta_j)}. \tag{2.4}$$

We emphasize $\hat{h}_{i,j}(w)$ is not the transform of the BRT forward operator as the delta functions have been excluded.

The expression (2.3c) highlights some challenges with BRT inversion. We observe singularities at $w \cdot \theta_i = 0$ and $w \cdot \theta_j = 0$. At these frequencies, finite $\hat{\mu}_C(w)$ does not guarantee finite $\hat{g}_{i,j}(w)$. As a consequence, the data may have unbounded support.

Additionally (2.3c) demonstrates zeros in the forward operator. We define the set $\Theta_{i,j} \subset \mathbb{R}^2$ as

$$\Theta_{i,j} \coloneqq \{w : w \cdot (\theta_i + \theta_j) = 0,\, w \cdot w > 0\}\,. \tag{2.5}$$

For all $w \in \Theta_{i,j}$, we have $\hat{g}_{i,j}(w) = 0$ for all $\hat{\mu}_C(w)$. In this way the BRT has a nontrivial nullspace. The zeros are limited to a line, and so the nullspace does not include images with bounded support. This does not preclude exact analytic reconstruction of images with bounded support. However, this is problematic for numeric reconstruction. We have arrived at these observations from a linear systems perspective. Similar observations were previously made applying microlocal analysis to the BRT [7].

A Fourier representation of the image is found by multiplying both sides of (2.3c) by the inverse of (2.4)

$$\hat{\mu}_C(w) = \hat{g}_{i,j}(w) \frac{-i2\pi\,(w \cdot \theta_i)\,(w \cdot \theta_j)}{w \cdot (\theta_i + \theta_j)}, \quad \forall\, w \notin \Theta_{i,j}. \tag{2.6}$$

Justification for removing the delta functions is given in Appendix C.2. Using (2.6) alone, we cannot recover $\hat{\mu}_C(w)$ for $w \in \Theta_{i,j}$. According to (2.3c), $\hat{g}_{i,j}(w) = 0$, for all $w \in \Theta_{i,j}$, which leaves (2.6) indeterminate. Instead, $\hat{\mu}_C(w)$ for $w \in \Theta_{i,j}$ must be recovered by imposing boundary conditions on $\mu_C(x)$ or, equivalently, continuity of $\hat{g}_{i,j}(w)$ (i.e. applying L'Hôpital's rule).

In the parlance of linear systems analysis, this inversion formula comprises two lines of zeros and one line of poles. The zeros are associated with directional derivatives, and the poles are associated with integration. Inverting this process leads to the reconstruction formulas

$$\mu_C(x) = \frac{1}{\|\theta_i + \theta_j\|} \frac{d}{d\theta_i} \frac{d}{d\theta_j} \int_0^\infty g_{i,j}\left(x + s\frac{(\theta_i + \theta_j)}{\|\theta_i + \theta_j\|}\right) ds \tag{2.7a}$$

$$= \frac{-1}{\|\theta_i + \theta_j\|} \frac{d}{d\theta_i} \frac{d}{d\theta_j} \int_{-\infty}^0 g_{i,j}\left(x + s\frac{(\theta_i + \theta_j)}{\|\theta_i + \theta_j\|}\right) ds. \tag{2.7b}$$

A detailed derivation is given in Appendix C.2. This is a generalization of previous inversion formulas derived by other means [1, 5, 7, 8]. Sherson was the first to recognize the symmetry

in the expressions (2.7a) and (2.7b) [7]. This redundancy is useful for numeric reconstructions from noisy data. For finite data the integrals are applied over different lengths and different noise realizations. The two results can be combined to minimize variance in the reconstruction.

The delta functions in (2.3c) present some obvious challenges. Delta functions imply images with bounded support do not guarantee data with bounded support. Unbounded support begs the question: what extent of data is necessary for reconstruction? We address this in Section 2.2. Additionally, these delta functions present numerical challenges when computing the Fourier transform from sampled data. We will address this in Section 2.3. Combining these results we obtain Fourier-based inversion formulas in Section 2.4.

2.2 Complete Representation of Data with Infinite Support

It is helpful to distinguish segments of the boundary of C with respect to the orthogonal basis θ, $\theta^\perp$. For this we define the scalar values

$$v_\theta^- \coloneqq \min_{x\in C} x\cdot\theta^\perp \tag{2.8}$$

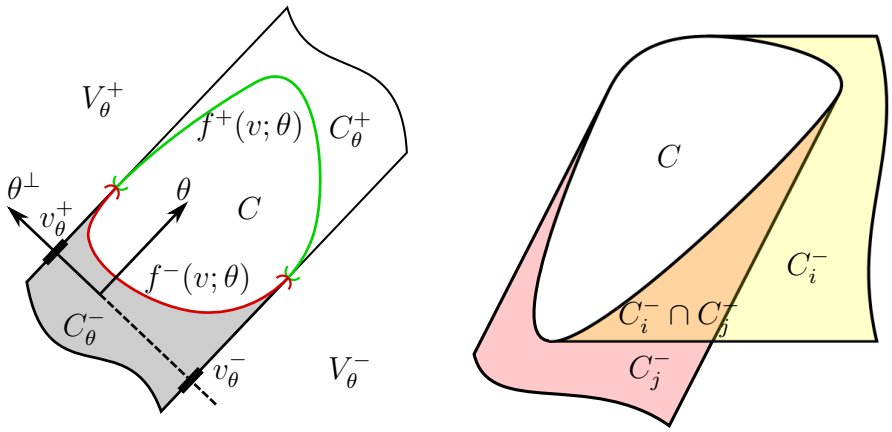
$$v_\theta^+ \coloneqq \max_{x\in C} x\cdot\theta^\perp \tag{2.9}$$

$$v_\theta \coloneqq v_\theta^+ - v_\theta^-. \tag{2.10}$$

Additionally we define the auxiliary functions $u_\theta^-, u_\theta^+ : \left[v_\theta^-, v_\theta^+\right] \to \mathbb{R}$

$$u_\theta^-(v) \coloneqq \min t, \text{ s.t. } t\theta + v\theta^\perp \in C \tag{2.11}$$

$$u_\theta^+(v) \coloneqq \max t, \text{ s.t. } t\theta + v\theta^\perp \in C. \tag{2.12}$$



(a) CBT data regions (b) BRT data regions

Figure 2.1: Images with bounded support, C, do not guarantee data with bounded support for either the CBT or BRT. For the CBT the support of the data is extended indefinitely in one direction, $-\theta$, over the region C_θ^- as depicted in Figure 2.1a. For the BRT the support of the data is extended in two directions. Depending on the shape of C, the resulting regions C_i^- and C_j^- may intersect.

We define non-overlapping line segments along the boundary of C as functions, $f^-, f^+ : \left[v_\theta^-, v_\theta^+\right] \to C$

$$f_\theta^-(v) \coloneqq u_\theta^-(v)\theta + v\theta^\perp \tag{2.13}$$
$$f_\theta^+(v) \coloneqq u_\theta^+(v)\theta + v\theta^\perp. \tag{2.14}$$

Using these functions, we define the mutually exclusive regions $C_\theta^-, C_\theta^+, V_\theta^-, V_\theta^+ \subset \mathbb{R}^2$:

$$C_\theta^- \coloneqq \left\{x : x\cdot\theta < u_\theta^-(x\cdot\theta^\perp;\theta), x\cdot\theta^\perp \in \left[v_\theta^-, v_\theta^+\right]\right\} \tag{2.15}$$
$$C_\theta^+ \coloneqq \left\{x : x\cdot\theta > u_\theta^+(x\cdot\theta^\perp;\theta), x\cdot\theta^\perp \in \left[v_\theta^-, v_\theta^+\right]\right\} \tag{2.16}$$
$$V_\theta^- \coloneqq \left\{x : x\cdot\theta^\perp < v_\theta^-\right\} \tag{2.17}$$
$$V_\theta^+ \coloneqq \left\{x : x\cdot\theta^\perp > v_\theta^+\right\}. \tag{2.18}$$

These definitions are illustrated in Figure 2.1a.

With definitions in place, we make some observations regarding the support of the CBT data. We state them as three theorems. First, we limit support of the data. The CBT data are zero for all x outside the support of the image, C, and the shadow region C_θ^-.

Theorem 2.1. $(B\mu_C)(x,\theta) = 0$ *for all* $x \in \overline{C \cup C_\theta^-}$.

Proof. The set $\overline{C \cup C_\theta^-}$ can be partitioned into three regions C_θ^+, V_θ^+, and V_θ^-, where

$$\overline{C \cup C_\theta^-} = C_\theta^+ \cup V_\theta^+ \cup V_\theta^-. \tag{2.19}$$

Since $C_\theta^+ \cap C = \varnothing$, we have $\mu_C(x) = 0$ for all $x \in C_\theta^+$. Additionally, if $x \in C_\theta^+$, then according to the definition (2.16), $x + s\theta \in C_\theta^+$ for all $s \geq 0$. Therefore $(B\mu_C)(x,\theta) = 0$ for all $x \in C_\theta^+$. The same is true for V_θ^- and V_θ^+. □

Next we observe the CBT, over the region C_θ^-, is constant along the direction θ. The values are determined by the Radon transform at θ.

Theorem 2.2. $(B\mu_C)(x,\theta) = (R\mu_C)(x \cdot \theta^\perp, \theta)$ *for all* $x \in C_\theta^-$.

Proof. For all $x \in C_\theta^-$, we can extend the integral of the CBT

$$(B\mu_C)(x,\theta) = \int_{-\infty}^{0} \mu_C(x + t\theta) + \int_0^\infty \mu_C(x + t\theta)dt, \tag{2.20}$$

where the first term is 0 due to the bounded support of μ_C. Combining these integrals and expanding x along the orthogonal basis vectors θ and $\theta^\perp$, we have

$$(B\mu_C)(x,\theta) = \int_{-\infty}^{\infty} \mu_C\left(\left(x \cdot \theta^\perp\right)\theta^\perp + t\theta\right) dt \tag{2.21a}$$

$$= (R\mu_C)(x \cdot \theta^\perp, \theta). \tag{2.21b}$$

□

Finally, the Radon transform, for fixed direction θ, is given by the CBT along the boundary of C.

Theorem 2.3. $(R\mu_C)(v,\theta) = (B\mu_C)(f_\theta^-(v),\theta)$ *for all* $v \in \left[v_\theta^-, v_\theta^+\right]$ *and* $(R\mu_C)(v,\theta) = 0$ *for all* $v \notin \left[v_\theta^-, v_\theta^+\right]$.

Proof. For $v \in \left[v_\theta^-, v_\theta^+\right]$, we can expand

$$(R\mu_C)(v,\theta) = \int_{-\infty}^{u^-(v)} \mu_C(v\theta^\perp + t\theta)dt + \int_{u^-(v)}^{\infty} \mu_C(v\theta^\perp + t\theta)dt \tag{2.22a}$$

$$= \int_0^\infty \mu_C(v\theta^\perp + u_\theta^-(v)\theta + t\theta)dt \tag{2.22b}$$

$$= (B\mu_C)(f_\theta^-(v),\theta). \tag{2.22c}$$

For $v < v^-$, we have $\mu_C(v\theta^\perp + t\theta) = 0$, since $v\theta^\perp + t\theta \in V_\theta^-$ for all t. Therefore $(R\mu_C)(v,\theta) = 0$ for $v < v_\theta^-$. The same can be shown for $v > v_\theta^+$. □

Corollary 2.1. $(B\mu_C)(x,\theta) = (R\mu_C)(x\cdot\theta^\perp,\theta)$ *for all* $x \in V_\theta^- \cup C_\theta^- \cup V_\theta^+$.

Proof. Theorem 2.2 demonstrates equality for $x \in C_\theta^-$. For $x \in V_\theta^- \cup V_\theta^+$, we have $(B\mu_C)(x,\theta) = 0$ according to Theorem 2.1. For $x \in V_\theta^-$, we have $x\cdot\theta^\perp < v_\theta^-$. Therefore, $(R\mu_C)(x\cdot\theta,\theta) = 0$ according to Theorem 2.3. Similarly, $(R\mu_C)(x\cdot\theta,\theta) = 0$ for $x \in V_\theta^+$. □

Theorem 2.2 demonstrates images with bounded support do not guarantee CBT data with bounded support since C_θ^- is unbounded. Aperiodic unbounded support is problematic for discrete Fourier analysis. However, data outside the support of the original image is redundant. If $(B\mu_C)(x,\theta)$ is known for all $x \in C$ including its boundary, $(R\mu_C)(v,\theta)$ is available. Combining Theorem 2.1 and Theorem 2.2, the CBT is then known for all $x \in \mathbb{R}^2$. This is significant as there may be problems for which data are not available outside the support of the original image. This demonstrates samples along the boundary, or alternatively direct-path (ballistic) measurements, are sufficient. Once this minimum extent of data are available, CBT data can be extended arbitrarily.

For our problems of interest CBT data are not available directly. The BRT is a linear combination of two CBTs sharing a common vertex. Similar to CBT data, bounded support of the image does not guarantee bounded support of the BRT data. The previous analysis of CBT data informs the sampling requirements on BRT data. Using the definition of the BRT (1.5), we distinguish two directions $\theta_i \neq \theta_j$. In addition to knowing $(G\mu_C)(x,\theta_i,\theta_j)$ for all $x \in C$, we additionally require the Radon transform in two directions: $(R\mu_C)(v,\theta_i)$, and $(R\mu_C)(v,\theta_j)$. The complication lies in the partitions of the BRT data. BRT data requires

additional partitions which may overlap. Resolving the Radon transform, with respect to two directions, from the BRT data is more challenging.

Following the previous work, our definitions for C and μ_C need not change. However, we use i and j to distinguish the directions the subscripts of the definitions (2.8)-(2.18). These indices are used only in subscripts to avoid confusion with the imaginary unit $i \coloneqq \sqrt{-1}$. Depending on C, θ_i, and θ_j, the set $C_i^- \cap C_j^-$ may be nontrivial. The BRT data can be partitioned as

$$(G\mu_C)(x,\theta_i,\theta_j) = \begin{cases} (B\mu_C)(x,\theta_i) + (B\mu_C)(x,\theta_j) & \text{for } x \in C \\ (R\mu_C)(x\cdot\theta_i^\perp,\theta_i) + (R\mu_C)(x\cdot\theta_j^\perp,\theta_j) & \text{for } x \in C_i^- \cap C_j^- \\ (R\mu_C)(x\cdot\theta_i^\perp,\theta_i) & \text{for } x \in C_i^- \setminus C_j^- \\ (R\mu_C)(x\cdot\theta_j^\perp,\theta_j) & \text{for } x \in C_j^- \setminus C_i^- \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

These regions are depicted in Figure 2.1b.

In contrast to the CBT, we must distinguish $(R\mu_C)(v,\theta_i)$ from $(R\mu_C)(v,\theta_j)$. Over C alone, they may not be directly available. We consider two scenarios. First, for some regions C and scatter angles θ_i, θ_j, the set $C_i^- \cap C_j^-$ is empty. For example, this is true for rectangular C, when θ_i is parallel to a boundary of C, and $\theta_i \cdot \theta_j \leq 0$. In such cases, $(R\mu_C)(v,\theta_i)$ from $(R\mu_C)(v,\theta_j)$ can be distinguished along the boundary of C. As a second scenario, forward scatter (ballistic) measurements at the two angles can be used to measure the Radon transforms directly. Forward scatter measurements are distinct from the BRT measurements we have considered. However, this may be useful for some modalities if measurements over the boundary of C are not available.

The notation introduced in this section is also useful for simplifying the assumed support of the image. Due to shift-invariance of the BRT, we can assume the image is centered about the origin without loss of generality.

Definition 2.1. *Let C represent a closed and bounded region in $\mathbb{R}^2$, and let θ_i and θ_j represent unique directions such that $|\theta_i \cdot \theta_j| < 1$. We define v_i^-, v_i^+, v_j^-, v_j^+ using (2.8) and (2.9). Then, C is centered with respect to θ_i and θ_j when both $v_i^+ = -v_i^-$ and $v_j^+ = -v_j^-$.*

Parallelograms are an important geometric shape in the context of the BRT. This was first recognized by Ambartsoumian and Jebelli [8]. It is often convenient to extend C to the circumscribed parallelogram.

Definition 2.2. *Let C represent a closed and bounded region centered with respect to θ_i and θ_j. The circumscribed parallelogram, with edges parallel to θ_i and θ_j, is given by*

$$P \coloneqq \left\{ x : \Pi_{v_i}\left(x \cdot \theta_i^{\perp}\right) \Pi_{v_j}\left(x \cdot \theta_j^{\perp}\right) > 0 \right\}, \tag{2.24}$$

where v_i, v_j are defined according to (2.8)-(2.10)*. Clearly, $C \subseteq P \subset \mathbb{R}^2$.*

In (2.24), P is expressed in terms of the orthogonal distance between parallel sides. Alternatively, we obtain the edge lengths α_i and α_j for the edges parallel to θ_i and θ_j, respectively

$$\alpha_i \coloneqq v_j / \left|\det\left(\theta_i, \theta_j\right)\right| \tag{2.25}$$

$$\alpha_j \coloneqq v_i / \left|\det\left(\theta_i, \theta_j\right)\right|. \tag{2.26}$$

We can equivalently express P in terms of the edge lengths

$$P = \left\{ s_i\theta_i + s_j\theta_j ; |s_i| \leq \alpha_i/2, |s_j| \leq \alpha_j/2 \right\}. \tag{2.27}$$

Related to P, we define a parallelogram indicator function in Appendix C.3 and derive its two-dimensional Fourier transform. The results will be referenced frequently in subsequent sections.

2.3 Filtering Unbounded Support of the Data

When the Fourier transform must be determined numerically, unbounded support of the BRT is problematic. Simply truncating BRT data effects blurring in the frequency domain. Windowing corrupts the spectral representation and invalidates the previous Fourier reconstruction methods. Alternatively, we consider a filtered representation of the data comprising shifted and negated copies. We define a generalized point-spread function (PSF) such that the shifted copies of the data combine destructively outside a bounded region of support.

We consider the PSF

$$m_{i,j}(x;a_i,a_j) = \delta\left(x+\frac{a_i}{2}\theta_i+\frac{a_j}{2}\theta_j\right)-\delta\left(x-\frac{a_i}{2}\theta_i+\frac{a_j}{2}\theta_j\right) \\ -\delta\left(x+\frac{a_i}{2}\theta_i-\frac{a_j}{2}\theta_j\right)+\delta\left(x-\frac{a_i}{2}\theta_i-\frac{a_j}{2}\theta_j\right). \tag{2.28}$$

Here $a_i, a_j > 0$, determine the shift lengths. To illustrate the effects of this function, its convolution with an image, $\mu(x)$, are demonstrated in Figure 2.5a. To be clear, however, we are motivated by the benefits convolving $m_{i,j}(x;a_i,a_j)$ and the data $g_{i,j}(x)$ as demonstrated in Figure 2.5b.

The expression (2.28) has the Fourier transform

$$\hat{m}_{i,j}(w;a_i,a_j) = -4\sin(\pi a_i w\cdot\theta_i)\sin(\pi a_j w\cdot\theta_j). \tag{2.29}$$

To reduce the number of variables defined we introduce new notation to distinguish signals, which support expansion using the PSF function (2.28). We define

$$\hat{g}^m_{i,j}(w;a_i,a_j) \coloneqq \hat{g}_{i,j}(w)\hat{m}_\theta(w;a_i,a_j) \tag{2.30}$$

$$\hat{\mu}^m_{i,j}(w;a_i,a_j) \coloneqq \hat{\mu}_C(w)\hat{m}_\theta(w;a_i,a_j). \tag{2.31}$$

The same superscript m will be subsequently applied to continuous signals in the spatial domain and to sampled signals. Plugging (2.3c) and (2.28) into (2.30) we have

$$\hat{g}^m_{i,j}(w;a_i,a_j) = \hat{\mu}_C(w)\left[-4\sin(\pi a_i w\cdot\theta_i)\sin(\pi a_j w\cdot\theta_j)\frac{-w\cdot(\theta_i+\theta_j)}{i2\pi(w\cdot\theta_i)(w\cdot\theta_j)}\right. \\ -4\sin(\pi a_i w\cdot\theta_i)\sin(\pi a_j w\cdot\theta_j)\frac{1}{2}\delta(w\cdot\theta_i) \\ \left.-4\sin(\pi a_i w\cdot\theta_i)\sin(\pi a_j w\cdot\theta_j)\frac{1}{2}\delta(w\cdot\theta_j)\right]. \tag{2.32}$$

The inverse two-dimensional Fourier transform of this expression involves integration over w. Due to the sampling property of the delta function, and since $\sin(0) = 0$, the final two bracketed terms vanish under integration. By the uniqueness of the Fourier transform, we

have

$$\hat{g}^m_{i,j}(w;a_i,a_j) = \hat{\mu}_C(w)\hat{m}_{i,j}(w;a_i,a_j)\frac{-w\cdot(\theta_i+\theta_j)}{i2\pi\,(w\cdot\theta_i)\,(w\cdot\theta_j)} \tag{2.33a}$$

$$= \hat{\mu}^m_{i,j}\,(w;a_i,a_j)\,\hat{h}_{i,h}(w). \tag{2.33b}$$

In (2.33b) we make use of both (2.31), and (2.4). Since the BRT is LSI, this result is expected. Filtering the input to an LSI system is equivalent to filtering the output. The significance is that the delta functions vanish when we filter the data using (2.28).

We obtain another useful form by expanding (2.33a) using (2.29). We reappropriate the denominator of $\hat{h}_{i,j}(w)$ to find

$$\hat{g}^m_{i,j}(w;a_i,a_j) = -i2\pi w\cdot(\theta_i+\theta_j)\,\hat{\mu}_C(w)a_ia_j\,\mathrm{sinc}\,(a_iw\cdot\theta_i)\,\mathrm{sinc}\,(a_jw\cdot\theta_j)\,. \tag{2.34}$$

The product of sinc functions in (2.34) is associated with a parallelogram window function as demonstrated in Appendix C.3. This motivates the definition

$$\hat{\mu}^p_{i,j}\,(w;a_i,a_j) \coloneqq \frac{\hat{\mu}_C(w)\hat{p}_{i,j}\,(w;a_i,a_j)}{a_ia_j\,|\det\,(\theta_i,\theta_j)|}, \tag{2.35}$$

where $\hat{p}_{i,j}\,(w;a_i,a_j)$ is defined according to (C.38). The scaling is motivated by (C.29). Using (2.35) in (2.34), we have

$$\hat{g}^m_{i,j}(w;a_i,a_j) = -i2\pi w\cdot(\theta_i+\theta_j)\,a_ia_j\hat{\mu}^p_{i,j}\,(w;a_i,a_j)\,. \tag{2.36}$$

Taking the inverse two-dimensional Fourier transform of (2.36) we find

$$g^m_{i,j}(x;a_i,a_j) = -\frac{d}{d\,(\theta_i+\theta_j)}a_ia_j\mu^p_{i,j}\,(x;a_i,a_j)\,. \tag{2.37}$$

Here the first factor represents the directional derivative in the direction $\theta_i+\theta_j$. This is clearly not a unit vector. In this form we observe $g^m_{i,j}(x;a_i,a_j)$ has bounded support.

Theorem 2.4. *For an absolutely integrable image with bounded support, filtering the BRT data with the PSF* (2.28) *bounds support of the data for all* a_i, $a_j \in (0,\infty)$. *Additionally, the data are finite everywhere.*

Proof. Without loss of generality, we assume the support of the image $\mu_C(x)$ is bounded by the circumscribed parallelogram, P, according to Definition 2.2. We first observe $\mu^p_{i,j}(x; a_i, a_j)$ has bounded support. We define

$$f(x) \coloneqq \mu^p_{i,j}(x; a_i, a_j)\, a_i a_j \left|\det(\theta_i, \theta_j)\right| \tag{2.38a}$$

$$= \mu_C(x) * p_{i,j}(x; a_i, a_j). \tag{2.38b}$$

The indicator function $p_{i,j}(x; a_i, a_j)$, defined by (C.28), has bounded support over a parallelogram similar to P in (2.24). Taking the convolution of two functions defined over similar parallelograms, the support of the result is also bounded by a similar parallelogram. The support of $f(x)$ is therefore limited to a parallelogram with sides parallel to θ_i and θ_j with perpendicular distances $v_i + b_i$ and $v_j + b_j$, respectively. The variables b_i and b_j are related to a_j and a_i according to (C.25) and (C.26), respectively.

Using (2.38b) in (2.37), we have

$$\left|g^m_{i,j}(x; a_i, a_j)\right| = \frac{1}{\left|\det(\theta_i, \theta_j)\right|} \left|\frac{d}{d(\theta_i + \theta_j)} f(x)\right| \tag{2.39}$$

Outside the the region of support of $f(x)$, its directional derivative is also zero. Therefore $g^m_{i,j}(x; a_i, a_j)$ has bounded support.

To show $g^m_{i,j}(x; a_i, a_j)$ is finite everywhere, we consider

$$\left|g^m_{i,j}(x; a_i, a_j)\right| = \left|g_{i,j}(x) * m_{i,j}(x; a_i, a_j)\right| \tag{2.40a}$$

$$= \Big|g_{i,j}\left(x + \frac{a_i}{2}\theta_i + \frac{a_j}{2}\theta_j\right) - g_{i,j}\left(x - \frac{a_i}{2}\theta_i + \frac{a_j}{2}\theta_j\right)$$
$$- g_{i,j}\left(x + \frac{a_i}{2}\theta_i - \frac{a_j}{2}\theta_j\right) + g_{i,j}\left(x - \frac{a_i}{2}\theta_i - \frac{a_j}{2}\theta_j\right)\Big| \tag{2.40b}$$

$$\leq 4\|g_{i,j}(x)\|_\infty. \tag{2.40c}$$

This is finite due to the assumption $\mu_C(x)$ is integrable. □

2.4 Image Reconstruction from BRT Data with Bounded Support

Bounded BRT data facilitates numeric inversion in the frequency domain. We consider two inversion strategies. The size of the available spreading parameters a_i and a_j in (2.28) plays an important role in selecting an inversion strategy. In both cases we reconstruct a version of the desired image subject to convolution. However, the PSFs associated with the reconstructed images are different.

Multiplying both sides of (2.33b) by the inverse of (2.4) we have the relationship

$$\hat{\mu}^m_{i,j}(w; a_i, a_j) = \hat{g}^m_{i,j}(w; a_i, a_j) \frac{-i2\pi (w \cdot \theta_i)(w \cdot \theta_j)}{w \cdot (\theta_i + \theta_j)}, \quad \forall\, w \notin \Theta_{i,j}. \tag{2.41}$$

This is similar to (2.6). However, the reconstruction is subject to multiplication with the PSF (2.29). Analytically, we can recover $\hat{\mu}^m_{i,j}(w; a_i, a_j)$ from $\hat{g}^m_{i,j}(w; a_i, a_j)$ using (2.41) and continuity assumptions or, equivalently, boundary conditions on $\mu^m_C(x)$.

We find a representation of the left-hand side of (2.41) in the spatial domain by taking the inverse two-dimensional Fourier transform of (2.31)

$$\begin{aligned} \mu^m_{i,j}(x; a_i, a_j) \quad &= \mu_C\left(x + \tfrac{a_i}{2}\theta_i + \tfrac{a_j}{2}\theta_j\right) - \mu_C\left(x - \tfrac{a_i}{2}\theta_i + \tfrac{a_j}{2}\theta_j\right) \\ &\quad - \mu_C\left(x + \tfrac{a_i}{2}\theta_i - \tfrac{a_j}{2}\theta_j\right) + \mu_C\left(x - \tfrac{a_i}{2}\theta_i - \tfrac{a_j}{2}\theta_j\right). \end{aligned} \tag{2.42}$$

For small a_i and a_j, the image copies will overlap. As a_i and a_j increase we can reconstruct $\mu_C(x)$ from segments without overlap.

Theorem 2.5. *An image with bounded support, $\mu_C(x)$, can be recovered from filtered BRT data $g^m_{i,j}(x; a_i, a_j)$ when $a_i > v_j / |2 \det(\theta_i, \theta_j)|$ and $a_j > v_i / |2 \det(\theta_i, \theta_j)|$ for v_i, v_j defined according to* (2.10)*.*

Proof. A portion of the image $\mu_C(x)$, without overlap, is associated with each shifted copy in (2.42). When the shifts are sufficiently large, the partial images can be combined to reconstruct the original image. To demonstrate this it is useful to first extend C to the

circumscribed parallelogram P in (2.27). The edge lengths of this geometric region bound the minimum shift lengths for image recovery.

To emphasize P as the assumed region of support, we use $\mu_P(x)$. Since $C \subseteq P \subset \mathbb{R}^2$, we have $\mu_P(x) = \mu_C(x)$, for all $x \in \mathbb{R}^2$. For $a_i > \alpha_i/2$ and $a_j > \alpha_j/2$ we can recover $\mu_P(x)$ from $\mu^m_{i,j}(x)$ using

$$\mu_P\left(s_i\theta_i + s_j\theta_j\right) = \begin{cases} \mu^m_{i,j}\left(\left(s_i - \frac{a_i}{2}\right)\theta_i + \left(s_j - \frac{a_j}{2}\right)\theta_j\right), & s_i, s_j \leq 0 \\ -\mu^m_{i,j}\left(\left(s_i - \frac{a_i}{2}\right)\theta_i + \left(s_j + \frac{a_j}{2}\right)\theta_j\right), & s_i \leq 0, s_j \geq 0 \\ -\mu^m_{i,j}\left(\left(s_i + \frac{a_i}{2}\right)\theta_i + \left(s_j - \frac{a_j}{2}\right)\theta_j\right), & s_i \geq 0, s_j \leq 0 \\ \mu^m_{i,j}\left(\left(s_i + \frac{a_i}{2}\right)\theta_i + \left(s_j + \frac{a_j}{2}\right)\theta_j\right), & s_i, s_j \geq 0. \end{cases} \tag{2.43}$$

Each case can be expanded as a series of four terms using (2.42). However, three of these terms are zero due to the support of P in (2.27). Combining the four cases we recover $\mu_P(x)$, and therefore $\mu_C(x)$, for all $x \in \mathbb{R}^2$. Expanding $a_i > \alpha_i/2$ and $a_j > \alpha_j/2$ using (2.25), (2.26) we obtain the boundary in the stated form. □

In general this approach requires large data sets to obtain $\hat{g}^m_{i,j}(w; a_i, a_j)$. For many cases we can extend BRT data using the techniques in Section 2.2. However, when we are limited to small a_i and a_j, another approach is necessary.

Alternatively, we can simply recover $\hat{\mu}^p_{i,j}(w; a_i, a_j)$. From (2.36), we have

$$\hat{\mu}^p_{i,j}(w; a_i, a_j) = \frac{-\hat{g}^m_{i,j}(w; a_i, a_j)}{i2\pi a_i a_j w \cdot (\theta_i + \theta_j)}, \quad \forall\, w \notin \Theta_{i,j}. \tag{2.44}$$

Taking the inverse two-dimensional Fourier transform of (2.44) we have

$$\mu^p_{i,j}(x; a_i, a_j) = \frac{-1}{\alpha_i\alpha_j\|\theta_i + \theta_j\|} \int_{-\infty}^{0} g^m_{i,j}\left(x - t\frac{\theta_i + \theta_j}{\|\theta_i + \theta_j\|}\right) dt. \tag{2.45}$$

This is equivalent to the inversion formula of Ambartsoumian and Jebelli [8]. For sampled data this formula can be implemented easily whenever the direction of integration is aligned with a sampling axis. For other cases, the frequency-domain representation (2.44) is useful.

We emphasize $\mu^p_{i,j}(x; a_i, a_j) \neq \mu_C(x)$. Taking the inverse two-dimensional Fourier transform of (2.35), we have

$$\mu^p_{i,j}(x; a_i, a_j) = \frac{\mu(x) * p_{i,j}(x; a_i, a_j)}{a_i a_j \left|\det(\theta_i, \theta_j)\right|}. \tag{2.46}$$

This demonstrates the recovered image as a blurring of the original image with a parallelogram window function. For high resolution, noise-free, data the size of this window can be made arbitrarily small. The recovery (2.45) only approaches $\mu_C(x)$ in a limiting sense [8].

2.5 Numeric Algorithms

In application we must reconstruct images from sampled data. Our analysis of the BRT from a linear systems perspective extends easily to sampled data. We demonstrate this with two new algorithms. First we provide an algorithm for extending CBT data motivated by the work in Section 2.2. For a broad class of problems this can be applied to BRT data and therefore facilitates use of the filtering methods of Section 2.3. Numeric implementation of the filtering methods have been included in Appendix E. Second, we present a numeric inversion algorithm for bounded BRT data. Leveraging the rotational invariance of the two-dimensional Fourier transform, the directions θ_i, θ_j are unconstrained in our algorithm. We also include regularization to address poor conditioning of the forward operator.

2.5.1 Extending Truncated Data

We consider CBT data sampled uniformly over a rectangular region. For consistency with previous definitions, we expand x along two scalar axes $x = (t, y)$. For the two axes we use subscripts to distinguish the number of samples N_t, N_y and the sample spacing Δ_t, Δ_y. We collect the available data in the $N_y \times N_x$ matrix B. The elements represent samples of the CBT data

$$[B]_{n,m} = (B\mu_C)\left((t_B + (m-1)\Delta_t, y_B + (n-1)\Delta_y), (\cos\xi, \sin\xi)\right) \tag{2.47}$$

for $n \in \{1, \ldots, N_y\}$, and $m \in \{1, \ldots, N_t\}$. We expand in the direction $\theta = (\cos\xi, \sin\xi)$. The spatial location associated with sample $B_{1,1}$ is $x_B = (t_B, y_B)$. In this configuration

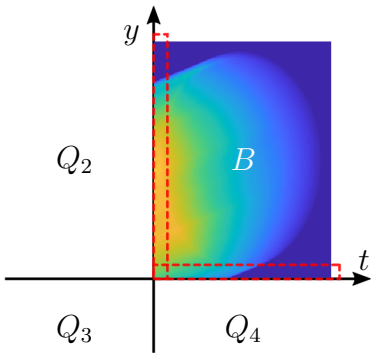


Figure 2.2: Truncated CBT data and relevant regions for extending the data. To extend the sampled CBT data, B, we use only the first row and first column as indicated by the dashed boxes. We first extend the data in the direction $-t$ to synthesize data in the second quadrant (indicated by Q_2). We then extend the combined data in the direction $-y$. The second extension synthesizes data in the third and fourth quadrants (indicated by Q_3 and Q_4). The process is detailed in Algorithm 1.

the y coordinate increases with the row index n, and the t coordinate increases with the column index m. It is not necessary to distinguish the terms Δ_t, Δ_y, and ξ for most of the computations related to sampled CBT data. For convenience we define

$$\lambda \coloneqq \frac{\Delta_t}{\Delta_y} \tan \xi, \tag{2.48}$$

which is a sufficient input for algorithms on uniformly sampled data.

Extending CBT data in the direction θ is trivial. For this we need only consider $x \in V_\theta^- \cup C_\theta^+ \cup V_\theta^+$, where $(B\mu_C)(x, \theta) = 0$ according to Theorem 2.1. Zero padding is sufficient.

Extending the data in the direction $-\theta$ is nontrivial. For simplicity we first consider only $\xi \in (0, \pi/2)$. Figure 2.2 illustrates the problem of extending the data, B, into the quadrants $Q2$, $Q3$, and $Q4$. The Radon transform serves as a proxy for extending the data according to Corollary 2.1. We assume the first row and column comprise no samples interior to C such that these data are samples of $(R\mu_C)(x \cdot \theta^\perp, \theta)$. We can then extend the data using $(B\mu_C)(x, \theta) = (R\mu_C)(x \cdot \theta^\perp, \theta)$. A brute-force approach would be to resample the Radon transform for each new data point. A computationally efficient approach is to extend the CBT data by shifting samples along the boundaries. This process is detailed in Algorithm 1. For $\xi \notin (0, \pi/2)$, we can still use Algorithm 1 by suitably flipping the inputs and outputs.

Algorithm 1 CbtExtend: Extend CBT data from a rectangular, uniformly sampled region. We assume the direction of integration is positive. Interpreting the available data as occupying the first quadrant, we extend the data into quadrants 2-4 as depicted in Figure 2.2. The inputs $\mathbf{b}_y$ and $\mathbf{b}_t$ represent the first column and row of the data matrix B, respectively. Input λ is given by (2.48). The inputs M_t and M_y indicate the number of requested samples in the direction $-t$ and $-y$, respectively. The input p indicates desired padding when using Algorithm 12 presented in E.5. We use vertcat and flipud to vertically concatenate and vertically flip matrices, respectively.

Require: $\mathbf{b}_y \in \mathbb{R}^{N_y}$, $\mathbf{b}_t \in \mathbb{R}^{N_t}$, $\lambda \in \mathbb{R}^+$, $M_t, M_y, p \in \mathbb{Z}^+$
Ensure: $Q_2 \in \mathbb{R}^{N_y \times M_t}$, $Q_3 \in \mathbb{R}^{M_y \times M_t}$, $Q_4 \in \mathbb{R}^{M_y \times N_t}$

$\mathbf{x}_R = \mathbf{b}_y$ ▷ Expand B in direction $-t$
$\mathbf{x}_L = \text{flipud}(\mathbf{b}_t(2 : 1 + p))$
$\mathbf{s} = \lambda \begin{bmatrix} -M_t & -M_t + 1 & \cdots & -1 \end{bmatrix}^T$
$p_W = p + M_y$
$W = \text{NonIntShift}(\mathbf{x}_R, \mathbf{s}, p_W, \mathbf{x}_L)$ ▷ Algorithm 12
$Q_2 = W(1 : N_y, :)$
$\mathbf{x}_R = \text{vertcat}(Q_2(1, :)^T, \mathbf{b}_t)$ ▷ Expand $[Q_2\, B]$ in direction $-y$
$\mathbf{x}_L = \text{flipud}(Q_2(2 : 1 + p, 1))$
$\mathbf{s} = \lambda^{-1} \begin{bmatrix} -M_y & -M_y + 1 & \cdots & -1 \end{bmatrix}^T$
$p_W = p + \lceil \lambda^{-1} M_y \rceil$
$W = \text{NonIntShift}(\mathbf{x}_R, \mathbf{s}, p_W, \mathbf{x}_L)$
$Q_3 = W(1 : M_t, :)^T$
$Q_4 = W(M_t + 1 : N_t + M_t, :)^T$

We can adapt this process to an important class of BRT problems. We consider the incident direction aligned with the t-axis and $\theta_i \cdot \theta_j < 0$. Specifically, we use $\theta_i = (-1, 0)$ and expand along $\theta_j = (\cos\xi, \sin\xi)$. We construct the BRT data matrix $G \in \mathbb{R}^{N_y \times N_t}$ with elements

$$[G]_{n,m} = (G\mu_C)\left((t_G + (m-1)\Delta_t, y_G + (n-1)\Delta_y), (-1, 0), (\cos\xi, \sin\xi)\right). \tag{2.49}$$

The previous definitions for Δ_t, Δ_y, and λ remain applicable.

Extending BRT data requires knowledge of both $(R\mu_C)(v, \theta_i)$ and $(R\mu_C)(v, \theta_j)$. We assume BRT data are sampled beyond the support of the image, such that no boundary samples of G correspond to points within C. For $\theta_i = (-1, 0)$, and $|\xi| < \pi/2$ this implies $[G]_{1,N_t} = [G]_{N_y,N_t} = 0$. In this case $(R\mu_C)(v, \theta_i)$ can be recovered from the last column of G. We can extend BRT data in the direction $-\theta_i$ simply by repeating the last column. For $\xi > 0$, the last row (maximum y) of G is then zero. The function $(R\mu_C)(v, \theta_j)$ can be recovered from the first column of G and the first row. The BRT data can be extended in the direction $-\theta_j$ using Algorithm 1. Alternatively, for $\xi < 0$, $(R\mu_C)(v, \theta_j)$ can be recovered from the first column of G and the last row. The BRT data can still be extended in the direction $-\theta_j$ using Algorithm 1. However, the inputs and outputs must be flipped accordingly.

Extending data does not address asymmetric boundaries precluding discrete Fourier representations. Rather, extended data are useful for synthesizing filtered data without overlap in the image space. Example algorithms for synthesizing filtered data are presented in Appendixes E.1 and E.2 for truncated CBT data and truncated BRT data, respectively. Note, these implementations operate on truncated data directly, and the necessary data extensions are performed internally.

2.5.2 Inversion of BRT Data with Regularization

Filtering ensures bounded support of $g^m_{i,j}(x; a_i, a_j)$. However, recovery of $\hat{\mu}^m_{i,j}(w; a_i, a_j)$ is still ill-posed due to conditioning of $\hat{h}_{i,j}(w)$. For this we use Tikhonov regularization which can be applied sample-wise in the frequency domain.

We restate (2.4) as an expression of scalar values by expanding $w = (w_t, w_y)$, $\theta_i = (\cos\xi_i, \sin\xi_i)$, and $\theta_j = (\cos\xi_j, \sin\xi_j)$, which leads to

$$\hat{h}_{i,j}((w_t, w_y)) = \frac{-w_t\left(\cos\xi_i + \cos\xi_j\right) - w_y\left(\sin\xi_i + \sin\xi_j\right)}{i2\pi\left(w_t\cos\xi_i + w_y\sin\xi_i\right)\left(w_t\cos\xi_j + w_y\sin\xi_j\right)}. \tag{2.50}$$

Notice this expression is commutative with respect to ξ_i and ξ_j. We define the system matrix $\hat{H}$ by sampling (2.50) uniformly

$$[\hat{H}]_{n,m} = \hat{h}_{i,j}\left(\frac{m}{N_t\Delta_t}, \frac{n}{N_y\Delta_y}\right). \tag{2.51}$$

The discrete analog of (2.33b) is then

$$\hat{G}^m = \hat{\Psi}^m \odot \hat{H}. \tag{2.52}$$

Here we have used $\hat{G}^m$ to represent the two-dimensional discrete Fourier transform of G^m, the filtered analog of (2.49). We use $\hat{\Psi}^m$ to represent samples of $\hat{\mu}^m_{i,j}(w)$. The symbol $\odot$ represents element-wise multiplication.

Zeros in the denominator of (2.50) are problematic for numeric analysis. We define the auxiliary function

$$d(w_t, w_y) = \left(w_t\cos\xi_i + w_y\sin\xi_i\right)\left(w_t\cos\xi_j + w_y\sin\xi_j\right). \tag{2.53}$$

However, filtering ensures $\hat{G}^m$ and $\hat{\Psi}^m$ are also zero when $d(w_t, w_y) = 0$.

Zeros in the numerator of (2.50) affect conditioning of the problem. In the frequency domain representation of the data, the magnitued of samples near these zeros will be small. In practice, these samples will be dominated by measurement noise and sampling errors. As the reciprocal of the forward operator is large at these samples, direct inversion of the forward operator will amplify these errors. Reducing the magnitude of the inverse at these frequencies mitigates the effects of noise and sampling errors in reconstructed images. We use Tikhonov regularization to approximate the inverse of $\hat{H}$ as

$$[K]_{n,m} \coloneqq \begin{cases} \frac{[\hat{H}^*]_{n,m}}{|[\hat{H}]_{n,m}|^2+\epsilon} & d(\frac{m}{N_t\Delta_t}, \frac{n}{N_y\Delta_y}) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.54}$$

Here * indicates complex conjugation, and ϵ is a small hyperparameter. For $|[\hat{H}]^{-1}_{n,m}| \gg \epsilon$, we have $[K]_{n,m} \approx [\hat{H}]^{-1}_{n,m}$. Conversely, for small $\epsilon = |\hat{H}_{n,m}|$, regularization forces $[K]_{n,m} \approx 1$, such that $|[K]_{n,m}| \ll |[\hat{H}]_{n,m}|^{-1}$. This mitigates sensitivity to sampling and measurement errors for samples where signal is attenuated by the forward operator.

The approximate inverse is applied element-wise to the Fourier transform of the data. The inversion formula reads

$$\hat{\Psi}^m \approx \hat{G}^m \odot K. \tag{2.55}$$

Applying the 2D inverse discrete Fourier transform to the result, we obtain a reconstruction of the filtered attenuation image. This process is described in Algorithm 2. The smoothing parameter ϵ, in (2.54), can be adjusted for measurement noise and numerical errors.

Algorithm 2 BrtInvertFiltered: Invert BRT data with bounded support. In this algorithm ComputeK refers to the computation of K using equations (2.50), (2.51), (2.53), and (2.54). Here we use DFT^2 and DFT^{-2} to represent the 2D discrete Fourier transform and its inverse, respectively.

Require: $G \in \mathbb{R}^{N_y \times N_t}$; $\Delta_t, \Delta_y, \epsilon \in \mathbb{R}^+$; $\xi_i, \xi_j \in \mathbb{R}$
Ensure: $\Psi \in \mathbb{R}^{N_y \times N_t}$
1: $K \leftarrow \mathrm{ComputeK}(N_t, N_y, \Delta_t, \Delta_y, \xi_i, \xi_j, \epsilon)$
2: $\hat{G} = \mathrm{DFT}^2\{G\}$
3: $\hat{\Psi} = \hat{G} \odot K$
4: $\Psi = \mathrm{DFT}^{-2}\left\{\hat{\Psi}\right\}$

Tikhonov regularization is generic and does not impose boundary conditions. For arbitrary angles ξ_i, and ξ_j, few samples of $\hat{\Psi}^m$ lie in the nullspace of the forward operator and it is sufficient to zero the results at these samples. Otherwise, it may be necessary to impose boundary conditions. For example, to ensure Ψ^m is zero along the $t = 0$ and $y = 0$ boundary, all columns and all rows of $\hat{\Psi}^m$ must sum to 0.

2.5.3 Fast BRT Operators

As a linear operator, it is natural to look to frequency-domain representations of the BRT for computationally efficient algorithms. BRT data represent numeric challenges, however, as the data exhibit unbounded aperiodic support for most images. Previously it was demonstrated

that unbounded support of BRT data can be canceled coherently through filtering [24]. By bounding support of the data, we can leverage the computationally efficient discrete Fourier transform (DFT).

The broken ray transform is the superposition of two improper line integrals sharing a common origin. For an analytic treatment we interpret $x \in \mathbb{R}^2$, and $\mu(x) : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$. We define the BRT of μ

$$(B\mu)(x, \boldsymbol{\theta}) \coloneqq \sum_{\theta \in \boldsymbol{\theta}} \int_0^\infty \mu(x + t\theta) dt. \tag{2.56}$$

The summation is over elements of the ordered pair $\boldsymbol{\theta} = (\theta_s, \theta_d)$ where we interpret the elements as unit vectors: $\theta_s, \theta_d \in S^1$. We define the two-dimensional Fourier transform of the BRT

$$\begin{aligned}
\tilde{b}_{\boldsymbol{\theta}}(w) &= \int_{\mathbb{R}^2} (B\mu)(x, \boldsymbol{\theta})\, e^{-j2\pi w \cdot x} d^2x && \text{(2.57a)} \\
&= \int_{\mathbb{R}^2} \sum_{\theta \in \boldsymbol{\theta}} \int_0^\infty \mu(x + t\theta) dt\, e^{-j2\pi w \cdot x} d^2x && \text{(2.57b)} \\
&= \sum_{\theta \in \boldsymbol{\theta}} \int_0^\infty \int_{\mathbb{R}^2} \mu(y) e^{-j2\pi w \cdot y} d^2y\, e^{j2\pi t w \cdot \theta} dt && \text{(2.57c)} \\
&= \tilde{\mu}(w) \sum_{\theta \in \boldsymbol{\theta}} \int_0^\infty e^{j2\pi t w \cdot \theta} dt && \text{(2.57d)} \\
&= \tilde{\mu}(w) \sum_{\theta \in \boldsymbol{\theta}} \left[\frac{-1}{j2\pi w \cdot \theta} + \frac{1}{2} \delta(w \cdot \theta) \right] && \text{(2.57e)} \\
&= \tilde{\mu}(w) \left[\frac{jw \cdot (\theta_s + \theta_d)}{2\pi (w \cdot \theta_s)(w \cdot \theta_d)} + \sum_{\theta \in \boldsymbol{\theta}} \frac{1}{2} \delta(w \cdot \theta) \right]. && \text{(2.57f)}
\end{aligned}$$

The delta function is associated with the unbounded support of the data and is problematic numerically as it is only defined by its integral. To address this in the frequency domain, we multiply $\tilde{b}$ with a function $\tilde{m}(w) = 0$ for all $w \cdot \theta = 0$, $\theta \in \boldsymbol{\theta}$. Specifically, we choose

$$m_{\boldsymbol{\theta}}(x; a_s, a_d) = \delta(x) + \delta(x + a_s\theta_s + a_d\theta_d) - \delta(x + a_d\theta_d) - \delta(x + a_s\theta_s), \tag{2.58}$$

which adds three shifted copies. This has the Fourier transform

$$\tilde{m}_{\boldsymbol{\theta}}(w; a_s, a_d) = -4 \sin(\pi a_s w \cdot \theta_s) \sin(\pi a_d w \cdot \theta_d) \exp\left(j\pi \left(a_s (w \cdot \theta_s) + a_d (w \cdot \theta_d)\right)\right). \quad (2.59)$$

Since the BRT is a linear operator, $\tilde{\mu}$ appears as a multiplicative term in $\tilde{b}$. It follows that multiplying $\tilde{\mu}$ by $\tilde{m}$ is equivalent to multiplying $\tilde{b}$ by $\tilde{m}$. Equivalently, adding shifted copies of the data is equivalent to adding shifted copies of the image. We define the BRT of the filtered image

$$\tilde{b}_{\boldsymbol{\theta}}(w)\tilde{m}_{\boldsymbol{\theta}}(w; a_s, a_d) = -j2\pi a_s a_d \big(w \cdot (\theta_s + \theta_d)\big) \operatorname{sinc}(\pi a_s w \cdot \theta_s) \operatorname{sinc}(\pi a_d w \cdot \theta_d) \times \exp\left(j\pi \left(a_s (w \cdot \theta_s) + a_d (w \cdot \theta_d)\right)\right) \tilde{\mu}(w). \quad (2.60)$$

In contrast to the right-hand side of (2.57f), the right-hand side of (2.60) is well defined for all w. Taking the inverse Fourier transform of (2.60) we obtain the BRT of the filtered image. The left-hand side of (2.60) represents the convolution of the desired data $b_{\boldsymbol{\theta}}(x)$ with (2.58). This convolution can be described as the superposition of four shifted copies of the data of interest. As long as one copy is retained without overlap, we can isolate the deisred data by computing the inverse Fourier transform of (2.60) and truncating the result. Conditions for avoiding overlap in the data copies are not obvious, however, due to the unbounded support of the data.

Support of the BRT data extends in the directions $-\theta_s$ and $-\theta_d$ indefinitely from all points in the support of the image [24]. The infinite regions of support in the data are analogous to shadows. Overlap between the original image and the shadows of the shifted copies will corrupt the forward transform. In application, this requires a_s, and a_d to be sufficiently large. The minimum distances will depend on θ_s, θ_d, and the assumed support of μ. A sufficient condition is the requirement that the support of μ lies in a parallelogram with edges parallel to θ_s, θ_d with lengths a_s, a_d, respectively (see Theorem 5 in [24]).

Motivated by (2.60), we define the analytic forward operator

$$\tilde{h}_{\boldsymbol{\theta}}(w) \coloneqq -j2\pi a_s a_d \big(w \cdot (\theta_s + \theta_d)\big) \operatorname{sinc}(\pi a_s w \cdot \theta_s) \times \operatorname{sinc}(\pi a_d w \cdot \theta_d) \times \exp\left(j\pi \left(a_s (w \cdot \theta_s) + a_d (w \cdot \theta_d)\right)\right). \quad (2.61)$$

In contrast to (2.57f), (2.61) is well defined for all $w \in \mathbb{R}^2$.

To discretize (2.61), we consider the matrix $X \in \mathbb{R}^{L_2 \times L_1}$ comprising samples of μ with sample spacing Δ_2, Δ_1. To avoid aliasing with filtering, the DFT must be computed with zero-padding. For $i \in \{1, 2\}$, the padded lengths are given by

$$N_i = 2\lceil 0.5 \left(L_i + \left(a_s \left[\theta_s \right]_i + a_d \left[\theta_d \right]_i \right) / \Delta_i \right) \rceil. \tag{2.62}$$

The ceiling operation is indicated using $\lceil \cdot \rceil$, and the scale factors ensure N_i are even.

Since w appears as a multiplicative factor in (2.61), there are numerical benefits to considering signed w (as opposed to a naive nonnegative w). For this we define the column vectors $\boldsymbol{w}_i \in \mathbb{R}^{N_i}$

$$\left[\boldsymbol{w}_i \right]_k \coloneqq \begin{cases} \dfrac{k-1}{\Delta_i N_i}, & k \in \left\{ 1, 2, \ldots, \dfrac{N_i}{2} \right\} \\ \dfrac{k-1-N_i}{\Delta_i N_i}, & \text{otherwise.} \end{cases} \tag{2.63}$$

Using these vectors we define $N_2 \times N_1$ matrices

$$W_s \coloneqq \left[\theta_s \right]_2 \boldsymbol{w}_2 + \left[\theta_s \right]_1 \boldsymbol{w}_1^T \tag{2.64}$$

$$W_d \coloneqq \left[\theta_d \right]_2 \boldsymbol{w}_2 + \left[\theta_d \right]_1 \boldsymbol{w}_1^T, \tag{2.65}$$

which we used to sample (2.61). We store the results in the matrix

$$\widetilde{H}_{\boldsymbol{\theta}} \coloneqq -j 2\pi a_s a_d \big(W_s + W_d \big) \operatorname{sinc} \left(\pi a_s W_s \right) \operatorname{sinc} \left(\pi a_d W_d \right) \exp \left(j\pi \left(a_s W_s + a_d W_d \right) \right). \tag{2.66}$$

In summary, $\widetilde{H}_{\boldsymbol{\theta}}$ is determined by $L_1, L_2 \in \mathbb{Z}_{>0}$, $\Delta_1, \Delta_2, a_s, a_d \in \mathbb{R}_{>0}$, and $\boldsymbol{\theta} = (\theta_s, \theta_d)$ (where $\theta_s, \theta_d \in S^1$) using (2.62)-(2.66). The size of $\widetilde{H}_{\boldsymbol{\theta}} \in \mathbb{C}^{N_2 \times N_1}$ is determined automatically from (2.62).

Given $\widetilde{H}_{\boldsymbol{\theta}}$, the forward BRT of X is implemented in the frequency domain using element-wise multiplication. This process is described in Algorithm 3. Since the DFT and inverse DFT can be scaled as a unitary operations, the adjoint BRT admits a convenient form. The adjoint BRT can be computed supplying the complex conjugate, $\widetilde{H}^*$, when calling Algorithm 3.

Algorithm 3 FBRT: Fourier BRT for uniformly sampled images. Here X is the sampled image, and $\widetilde{H}$ is a frequency-domain representation of the forward operator (2.66). We use DFT^2 and DFT^{-2} to represent the 2D discrete Fourier transform and its inverse, respectively. Additional arguments in DFT^2 specify the size after zero padding. Element-wise multiplication is indicated as $\odot$, and * indicates complex conjugation.

Require: $X \in \mathbb{R}^{L_2 \times L_1}$; $\widetilde{H} \in \mathbb{C}^{N_2 \times N_1}$; $L_2 \leq N_2, L_1 \leq N_1$
Ensure: $Y \in \mathbb{R}^{L_2 \times L_1}$
$\widetilde{X} = \mathrm{DFT}^2(X, N_2, N_1)$ ▷ Zero pad input
if ADJOINT **then**
    $\widetilde{Y} = \widetilde{X} \odot \widetilde{H}^*$ ▷ backward BRT
else
    $\widetilde{Y} = \widetilde{X} \odot \widetilde{H}$ ▷ forward BRT
end if
$Y = \mathrm{DFT}^{-2}\left(\widetilde{Y}\right)$
$Y = Y(1 : L_2, 1 : L_1)$ ▷ Truncate result

Bounding support of the data is possible for arbitrary directions θ_s, θ_d. However, for small scatter angles (e.g. $\|\theta_s \times \theta_d\| < 0.2$), the operator (2.61) requires significant padding of the DFT to cover support of the filtered data without aliasing. In some cases, other filter functions support more efficient spatial representations of the filtered data while preserving one copy without interference.

We propose a spatially efficient filtering strategy with periodic, nonzero, boundaries. Our approach requires at least one sample axis to be aligned with a BRT direction. Without loss of generality, we assume θ_s is aligned with the horizontal sampling axis. We refer to the original sampled image width $a_s = L_1\Delta_1$. We then select

$$a_d = \frac{a_s}{|\theta_s \cdot \theta_d|}. \tag{2.67}$$

To accomodate the increased extent of filtered data, we pad the original image size

$$N_1 = 3L_1 \tag{2.68}$$

$$N_2 = L_2 + \lceil a_d \|\theta_s \times \theta_d\| / \Delta_2 \rceil. \tag{2.69}$$

The horizontal padding is selected precisely to support nonzero symmetric boundaries. The vertical padding is only included to avoid aliasing. These suggested values ensure horizontal boundaries will be zero.

We employ two filter functions, one for each CBT

$$m_s(x) = \delta(x) - \delta\left(x + 3a_s\theta_s + a_d\theta_d\right) - \delta\left(x + 2a_s\theta_s\right) + \delta\left(x + 2a_s\theta_s + a_d\theta_d\right) \tag{2.70}$$

$$m_d(x) = \delta(x) - \delta\left(x + a_d\theta_d\right) - \delta\left(x + 2a_s\theta_s\right) + \delta\left(x + 2a_s\theta_s + a_d\theta_d\right). \tag{2.71}$$

These functions only differ by the second right-hand terms. The two-dimensional Fourier transforms of these functions read

$$\widetilde{m}_s(w) = -j2\sin\left(2\pi a_s w\cdot\theta_s\right)e^{j2\pi a_s w\cdot\theta_s} - j2\sin\left(\pi a_s w\cdot\theta_s\right)e^{j5\pi a_s w\cdot\theta_s}e^{j2\pi a_d w\cdot\theta_d} \tag{2.72}$$

$$\widetilde{m}_d(w) = -4\sin\left(2\pi a_s w\cdot\theta_s\right)\sin\left(\pi a_d w\cdot\theta_d\right)e^{j2\pi a_s w\cdot\theta_s}e^{j\pi a_d w\cdot\theta_d}. \tag{2.73}$$

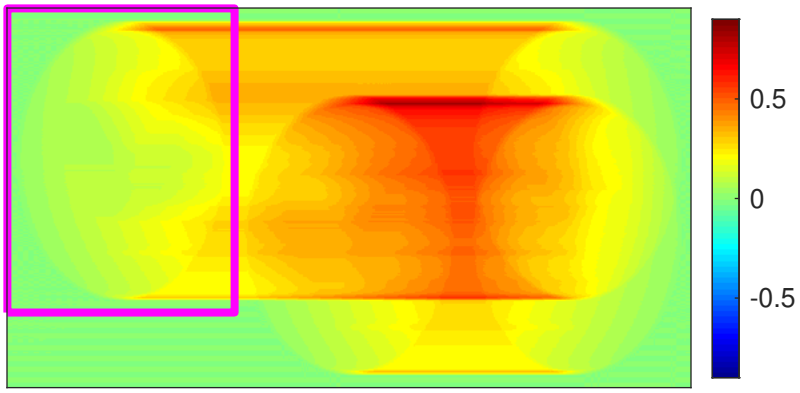
To obtain the filtered BRT, we must apply the corresponding CBT to each filter function. In this context, we omit the delta functions in the CBT (see (C.10)), and define

$$\tilde{h}_B(w;\theta) = \frac{-1}{i2\pi w\cdot\theta}. \tag{2.74}$$

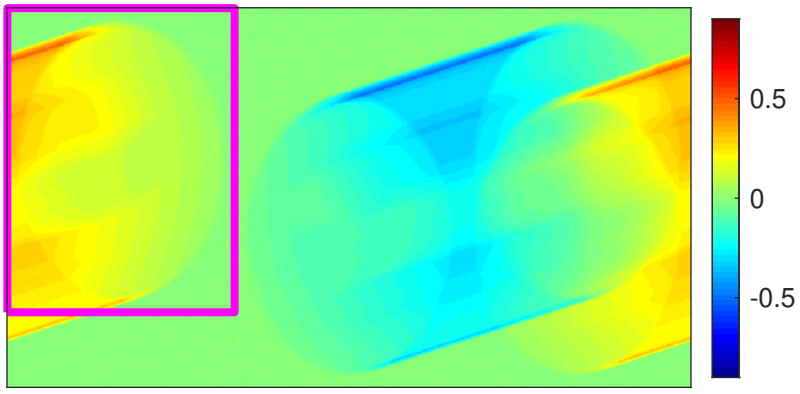
Superimposing the filtered results for θ_s and θ_d, we have

$$\tilde{h}_{\mathrm{FBRT}}(w) = \widetilde{m}_s(w)\tilde{h}_B(w;\theta_s) + \widetilde{m}_d(w)\tilde{h}_B(w;\theta_d). \tag{2.75}$$

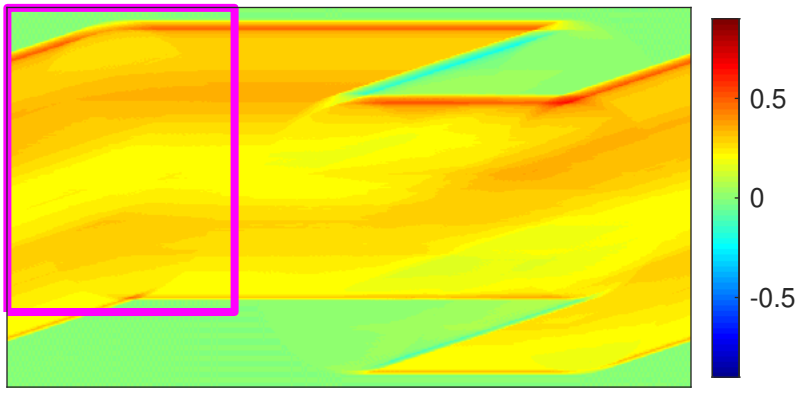
Sampling this continuous function at the frequencies dictated by the padded DFT yields filtered data with symmetric boundaries. For the operator (2.75), the required DFT padding decreases with $\|\theta_s \times \theta_d\|$ as indicated by (2.69).



(a) CBT, θ_s

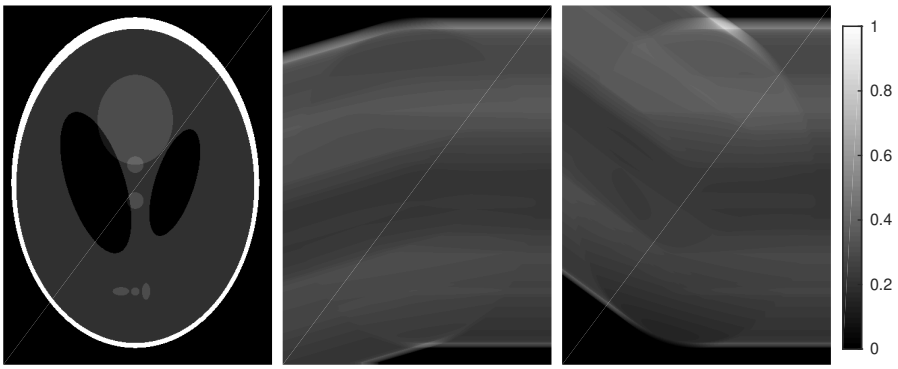


(b) CBT, θ_d



(c) BRT, θ_s, θ_d

Figure 2.3: Filtered CBT data effecting periodic BRT data. CBT data associated with the directions θ_s and θ_d are shown in Figures 2.3a and 2.3b, respectively. Summing these results, we obtain the periodic BRT data shown in Figure 2.3c. The magenta rectangle indicates the support of the original image. In Figure 2.3c, the rectangle also indicates cropping boundaries to obtain the truncated BRT data of interest.



(a) Reference image (b) BRT data, $\xi = \frac{\pi}{11}$ (c) BRT data, $\xi = -\frac{\pi}{5}$

Figure 2.4: Reference image and associated BRT data. Figure 2.4a depicts the Shepp-Logan phantom as a reference image. Figure 2.4b and 2.4c depict BRT data with different scatter angles. The BRT data were determined analytically and sampled at the scatter points associated with the pixel centers of Figure 2.4a.

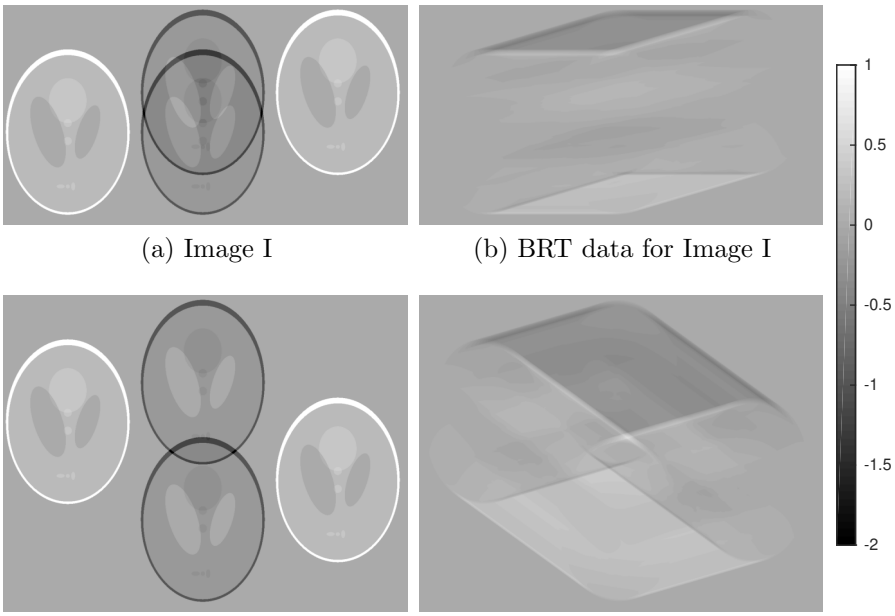
2.6 Numerical Simulations

2.6.1 Analytic BRT Inversion of Sampled Data

We provide results of numerical simulations to demonstrate the utility of this analysis. We use the modified Shepp-Logan phantom [25, 26] in most of our simulations as depicted in Figure 2.4. This phantom is reasonably challenging and the BRT data can be determined analytically as demonstrated in Appendix D.1. For Figure 2.4 we sample the image and data space uniformly in y and t. For y we use $N_y = 600$ sampling over $[-1, 1]$. For t we use $N_t = 400$ sampling over $[-0.75, 0.75]$. This effects different sampling rates in t and y. Limiting the extent of available BRT data in this way truncates the data both in y and t as shown in Figure 2.4b and Figure 2.4c.

We first demonstrate filtering bounds support of the data. Results are shown for both the BRT and SBRT in Figure 2.5. In this case the filtered image and filtered data were all obtained analytically and then sampled.

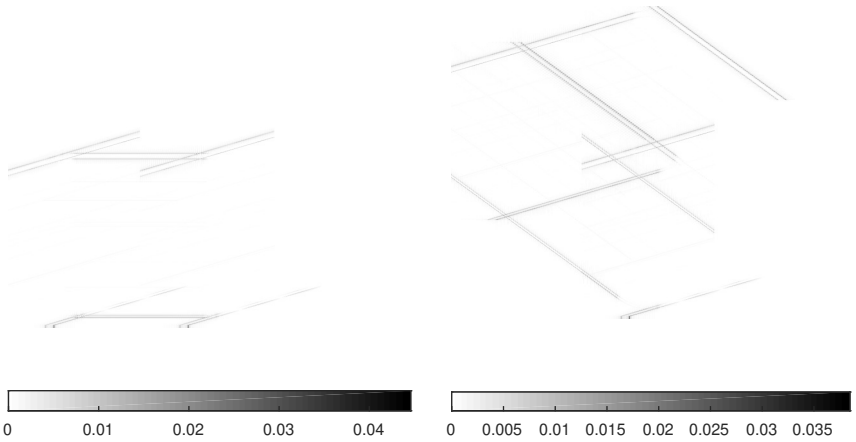
Filtering can also be applied to sampled BRT data directly. For sampled data this effects small errors which we quantify against the reference data of Figure 2.5. Results are shown



(a) Image I (b) BRT data for Image I

(c) Image II (d) SBRT data for Image II

Figure 2.5: Image filtering effects BRT data with bounded support. Figure 2.5a depicts a notional phantom defined by filtering the image of Figure 2.4a analytically using the PSF (2.28). In this case $\xi_i = \pi$ and $\xi_j = \pi/11$, where the subscripts distinguish the directions $\theta_i = (\cos\xi_i, \sin\xi_i)$. The associated analytic BRT data are shown in Figure 2.5b and indicate bounded support. To bound support of SBRT data, filtering need only address the unique scatter directions associated with the two BRT data sets. Figure 2.5c and Figure 2.5d show the filtered image and filtered SBRT data, respectively. Here the scatter angles for the BRT data composing the SBRT data are $\xi_j \in \{\pi/11, -\pi/5\}$.



(a) Absolute BRT filtering error (b) Absolute SBRT filtering error

Figure 2.6: Error extending and filtering sampled, truncated, BRT data. Figure 2.6a depicts the error due to extending and filtering the data of Figure 2.4b. The reference data are shown in Figure 2.5b. Similarly, Figure 2.6b depicts the error due to extending and filtering the data of Figure 2.4b and Figure 2.4c. In this case the reference data are shown in Figure 2.5d.

in Figure 2.6. Artifacts are observed at scatter points for which resulting rays are tangent to large transitions in the image. These artifacts are a consequence of sampling. For both BRT and SBRT filtering the peak absolute error is less than 5% the peak image value.

Further analysis of $h(w)$ provides insights on BRT inversion. We can express (2.4) in polar coordinates with the change of variables

$$w = \rho\left(\cos\phi, \sin\phi\right) \quad \theta_i = \left(\cos\xi_i, \sin\xi_i\right) \quad \theta_j = \left(\cos\xi_j, \sin\xi_j\right), \tag{2.76}$$

such that

$$\hat{h}((\rho,\phi)) = \frac{-\cos\left(\phi - \frac{1}{2}\left(\xi_i + \xi_j\right)\right)\cos\left(\frac{1}{2}\left(\xi_i - \xi_j\right)\right)}{i\pi\rho\cos\left(\phi - \xi_i\right)\cos\left(\phi - \xi_j\right)}. \tag{2.77}$$

We make a few observations. First, ρ in the denominator of (2.77) implies the BRT attenuates high frequency content. Reconstruction will be sensitive to noise at high frequencies. Second, there are singularities at $\phi = \xi_i \pm \pi/2$ and $\phi = \xi_j \pm \pi/2$. Filtering ensures the image and data are zero at these frequencies. Finally, (2.77) is zero at $\phi = \frac{1}{2}\left(\xi_i + \xi_j\right) \pm \pi/2$. These zeros do not appear in the CBT, but arise in the combination of two CBTs.

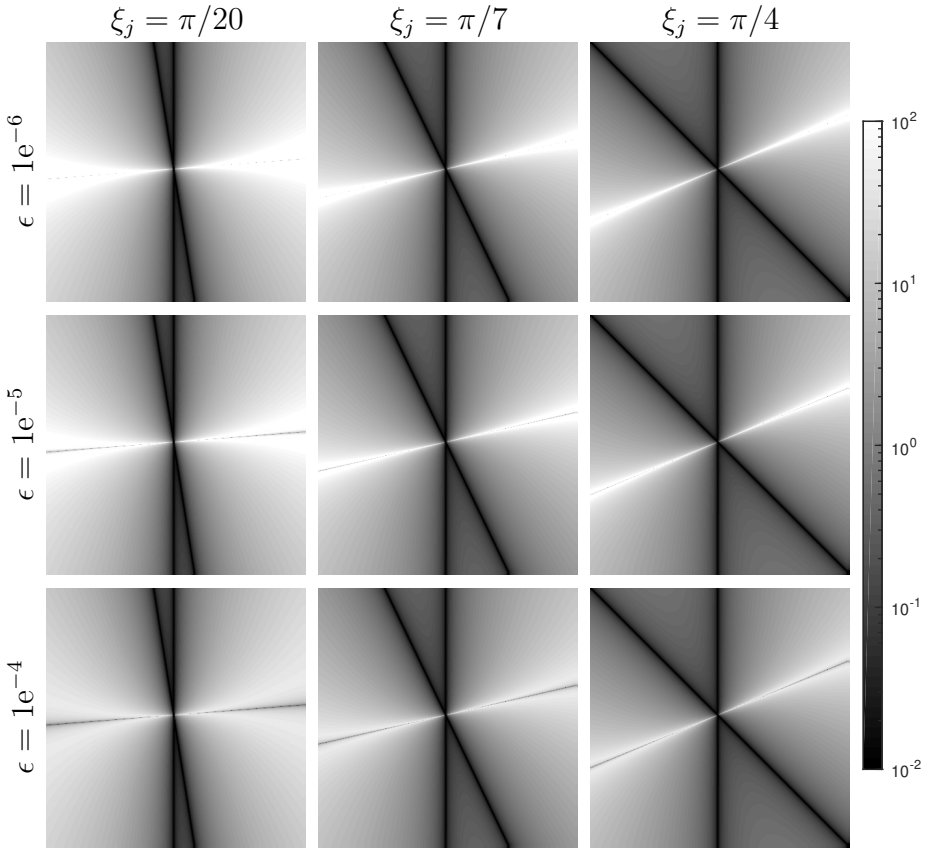


Figure 2.7: Changes in $|K|$ (2.54) with respect to regularization ϵ and angle ξ_j. The first column of images corresponds to scatter direction $\xi_j = \pi/20$, the second column to $\xi = \pi/7$, and the third column to $\xi = \pi/4$. For all images we fix $\xi_i = \pi$. Each row of images uses a different ϵ; the first row uses $1e^{-6}$, the second $1e^{-5}$, and the third $1e^{-4}$. For all images the zero-frequency content is centered for both axes. Further, the same display scale is used as shown in the colorbar.

The matrix K plays a critical role in BRT reconstruction (2.55). This incorporates changes to $h(w)$ due to ξ_j, and the regularization term ϵ. Changes to $|K|$ with respect to these terms is shown in Figure 2.7. Here we fix $\xi_i = \pi$ without loss of generality. The lines indicating strong attenuation are due to singularities of (2.77) at $\phi = \pm\pi/2$, and $\phi = \xi_j \pm \pi/2$. Zeros in (2.77) effect large amplitudes in $|K|$ along $\phi = \xi_j/2$. However, this amplitude is curtailed through regularization as ϵ increases. Large regularization will attenuate spectral content at these frequencies and effect blurring in reconstructed images. Without regularization, we would expect artifacts along this spectral line. Selection of ϵ will be application specific and will depend on the desired performance metric.

The original global BRT inversion formula is due to Florescu et al. [1]. We will refer to their inversion formula the FMS formula. Specifically contrasting with our algorithm, we analyze the same square phantom in Figure 2.8. The original work assumed data available over an infinite strip with no additional insights on limiting the data. The data of Figure 2.8b violates this assumption. Directly applying the FMS formula to this data yields poor results as shown in Figure 2.8c. However, we can simulate additional data using Algorithm 1. Applying the FMS formula to the extended BRT data yields results consistent with those previously published [1]. In this way, Algorithm 1 can be used as a preprocessing step to reduce the extent of sampling required for reconstruction using the FMS formula. The direction of the artifacts is explained by the nullspace of the forward operator (2.4). Striations are observed in the direction $\xi/2 + \pi/2$. Employing the regularizied approximation of the inverse (2.54) further improves reconstruction as demonstrated in Figure 2.8f.

Inversion results for the Shepp-Logan phantom on noisy data are shown in Figure 2.9. The BRT data were obtained analytically, sampled, and corrupted with additive Gaussian noise. For small ϵ, we see artifacts where the direction ξ is tangent to high frequency edges of the image. This is a consequence of sampling errors and extending the BRT data. Additionally, edges perpendicular to the direction $\xi/2$ are not well resolved. This blurring appears along the direction $\xi/2$. Increasing ϵ increases the angular extent of blurring. The effect is reduced as ξ increases.

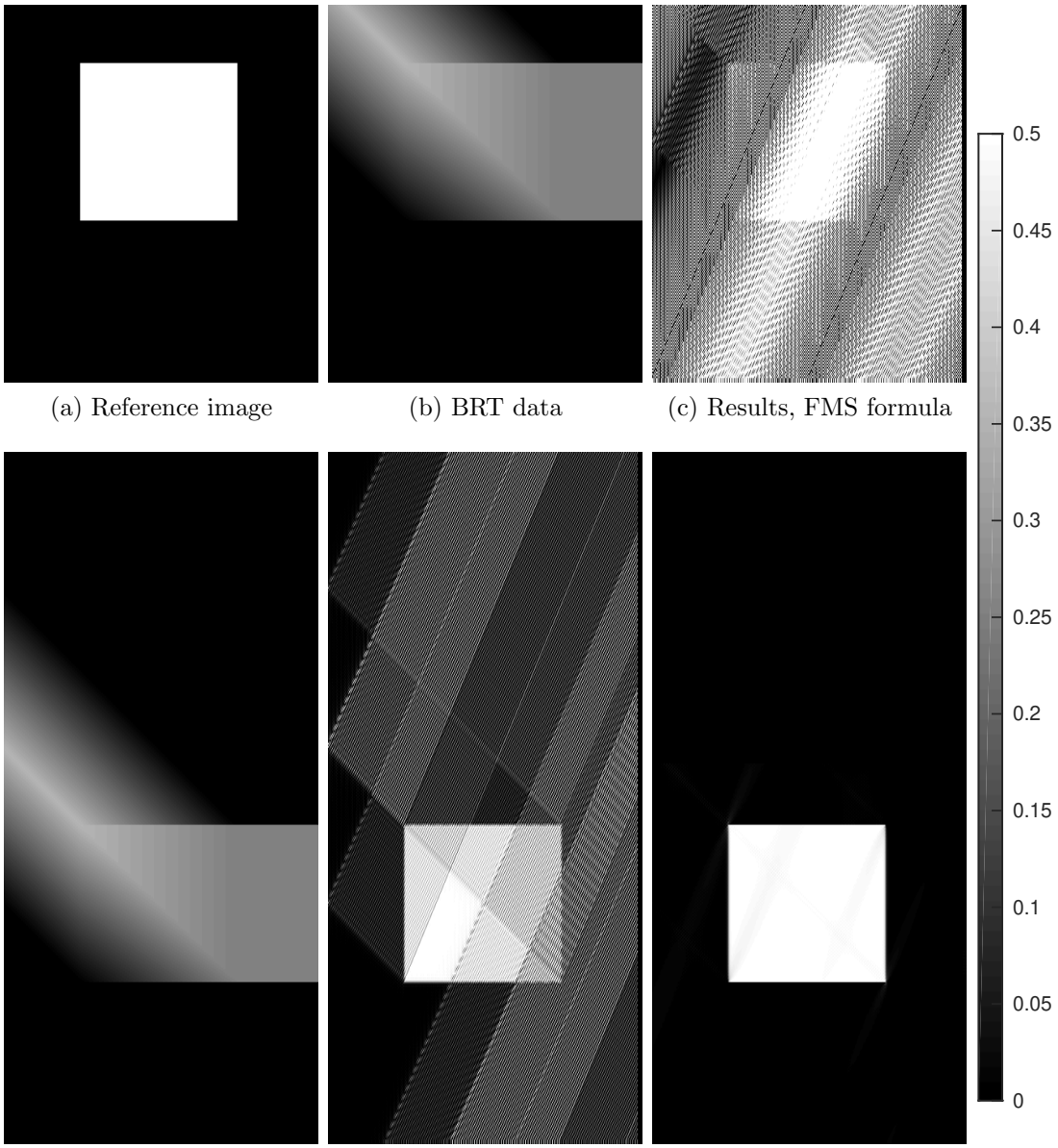
2.6.2 Contrasting BRT Operator Implementations

Iterative algorithms make frequent use of forward and backward (adjoint operators). For this reason it is important to ensure implementations are both accurate and computationally efficient. In the following we contrast performance of three implementations.

Linear forward and backward operators, on sampled data, can be summarized as

$$(G\mu)(y) = \sum_{x \in \mathcal{X}} h(y|x)\mu(x), \quad \forall\, y \in \mathcal{Y} \tag{2.78}$$

$$(G^\dagger g)(x) = \sum_{y \in \mathcal{Y}} h(y|x)g(y), \quad \forall\, x \in \mathcal{X}, \tag{2.79}$$



(a) Reference image (b) BRT data (c) Results, FMS formula

(d) Extended BRT data (e) Results, FMS formula (f) Results, Algorithm 2

Figure 2.8: Noise-free reconstruction from limited data. The reference image is shown in Figure 2.8a, and we limit the available BRT data as shown in Figure 2.8b with $\xi_j = -\pi/4$. FMS [1] reconstruction, using limited data, is shown in Figure 2.8c. The limited BRT data of Figure 2.8b can be extended using Algorithm 1 as shown in Figure 2.8d. Figure 2.8e depicts results applying the FMS formula to the extended data of Figure 2.8d. Similarly, Figure 2.8f depicts results applying Algorithm 2 to the extended data of Figure 2.8d. All images use the same display scale shown in the colorbar.

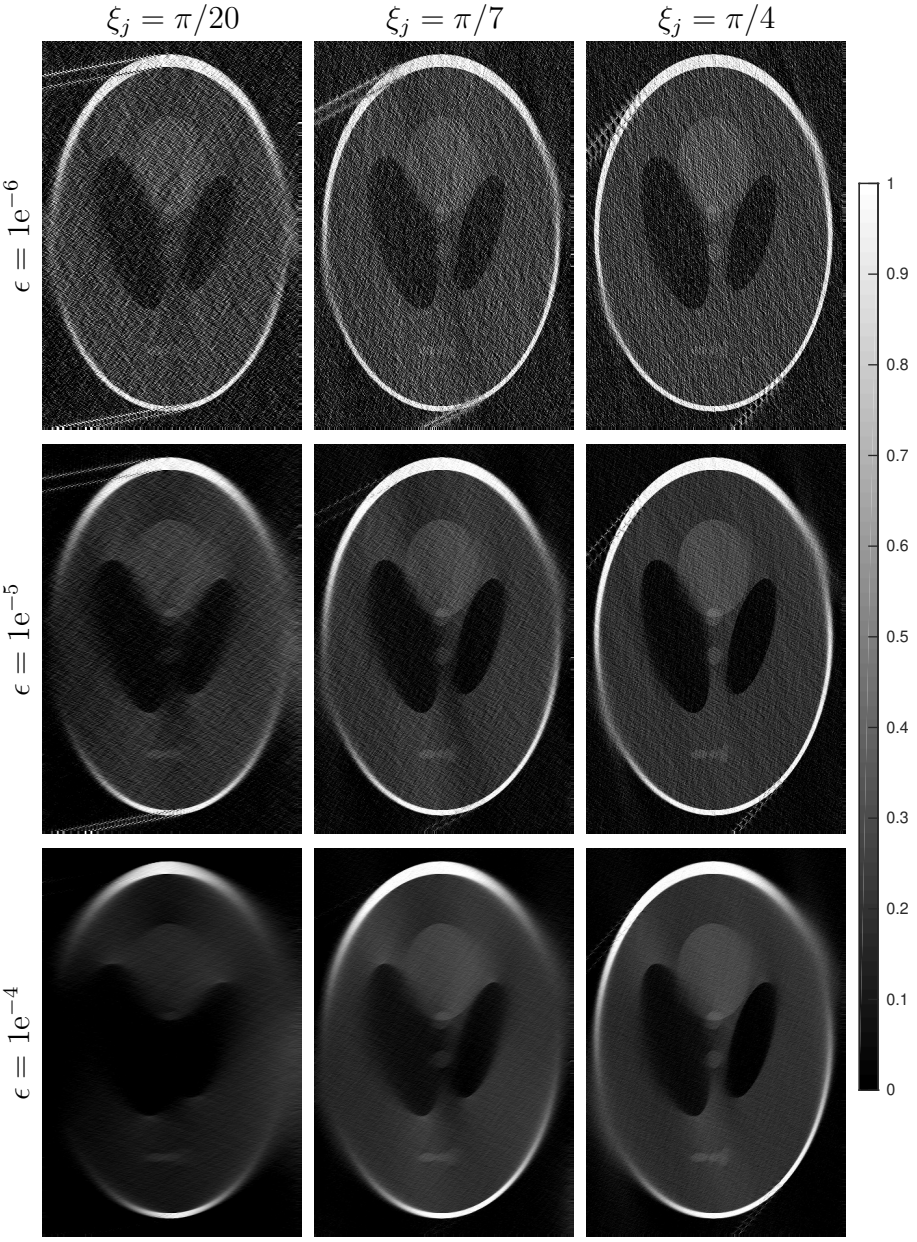


Figure 2.9: Reconstruction of noisy, truncated, BRT data using Algorithm 2. The first column of images corresponds to scatter direction $\xi_j = \pi/20$, the second column to $\xi_j = \pi/7$, and the third column to $\xi_j = \pi/4$. Each row of images uses a different ϵ in (2.54) which appears in the reconstruction formula (2.55); the first row uses $1e^{-6}$, the second $1e^{-5}$, and the third $1e^{-4}$. All images use the same display scale shown in the colorbar. The same realization of Gaussian noise was added to each data set. The standard deviation of the noise was 10^{-3} times the peak amplitude of the image.

respectively. Here we have omitted indices distinguishing the ray directions which we assume consistent between operators.

A direct implementation of (2.78) is achieved by explicitly computing the $|\mathcal{Y}| \times |\mathcal{X}|$ matrix H such that $[H]_{y,x} = h(y|x)$. This matrix represent numeric approximations of line integral through a sampled image. One approach is presented in Appendix D.2. For the broken ray transform this matrix is sparse as each broken ray path intersects few image voxels. Once H is available, the forward and adjoint operators are simply the inner product with H and its transpose, respectively.

When the forward operator is linear and shift-invariant (LSI), direct computation of H can be avoided using $h(y|x) = h(y - x)$. For LSI systems, (2.78) represents convolution. Representing μ and h in the frequency domain, the convolution in (2.78) becomes an elemnt-wise product. This process is described in Algorithm 3. However, aperiodic and unbounded support of the data complicate frequency-domain representations. Addressing these issues comes at costs both in computational complexity and fidelity.

Filtering has been demonstrated to bound the support of BRT data [24] and is convenient to implement in the frequency domain. However, padding is required to ensure filtered copies do not overlap. Judicious choice of filter functions can reduce the required padding. Padding can have a significant impact on computational efficiency of two-dimensional Fourier transforms.

Since we are concerned with real images, we can reduce the required DFT order by exploiting conjugate symmetry of real signals. For two-dimensional DFTs, this can be used to reduce the DFT order of one dimension by a factor of two. This approach is described in Algorithm 13.

Computation times associated with three methods are shown in Table 2.1. We contrast time required to initialize the operators and employ them for both forward and backward transforms. Additionally, we contrast time with the scatter angle. In both cases, the source direction is π, and the images are 300×400 pixels. All times are in seconds and were performed on a single laptop (MacBook Pro, late 2016. 2.9 GHz Quad-Core Intel Core i7).

Table 2.1: BRT Operator Processing Times in Seconds

	$\theta_j = \pi/4$			$\theta_j = \pi/20$		
Method	Setup	Forward	Backward	Setup	Forward	Backward
Direct	63.958	0.068	0.028	37.518	0.062	0.026
Fourier	0.138	0.016	0.016	0.076	0.009	0.010
Real Fourier	0.102	0.017	0.017	0.060	0.009	0.009

Setup time for the direct implementation includes computation of H, which is significant. Long setup times may be prohibitively expensive for approaches requiring frequent changes between operators (e.g. ordered subsets). It should be noted, however, that this task is highly paralellizable. The time required could be reduced significantly on some architectures. There is a small reduction in the setup time for the smaller scatter angle. This difference is due to asymmetry of the narrow image size. For our narrow image, $\theta_j = \pi/20$ effects an H with fewer nonzero entries.

The Fourier methods offer a significant reduction in processing time. The difference between the Fourier and Real Fourier results seem to be within the measurement error. For the direct operator, the backward transform is significantly faster than the forward transform. This is due to the column-major storage in Matlab.

To quantify consistency of the forward and adjoint operators, we consider the following relation

$$\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(y)h(y|x)\mu(x) = \sum_{y\in\mathcal{Y}} g(y)(G\mu)(y) \tag{2.80}$$

$$= \sum_{x\in\mathcal{X}} \mu(x)(G^\dagger g)(x). \tag{2.81}$$

In application, sampling precludes equality of these expressions. Errors are implementation-specific and vary with choice of μ, and g. We consider the the metric

$$\tau = \frac{2\left|\sum_{y\in\mathcal{Y}} g(y)(G\mu)(y) - \sum_{x\in\mathcal{X}} \mu(x)(G^\dagger g)(x)\right|}{\sum_{y\in\mathcal{Y}} g(y)(G\mu)(y) + \sum_{x\in\mathcal{X}} \mu(x)(G^\dagger g)(x)}, \tag{2.82}$$

Table 2.2: Percentage Error for BRT Forward and Backward Operators

	$\theta_j = \pi/4$		$\theta_j = \pi/20$	
Method	Smooth	Impulsive	Smooth	Impulsive
Direct	1.73e-15	1.53e-15	2.62e-15	1.15e-15
Fourier	4.80e-16	5.13e-07	4.37e-16	5.68e-06
Real Fourier	3.60e-09	5.34e-07	5.66e-09	5.57e-06

which represents a percentage error between the right-hand side of (2.80) and right-hand side of (2.81).

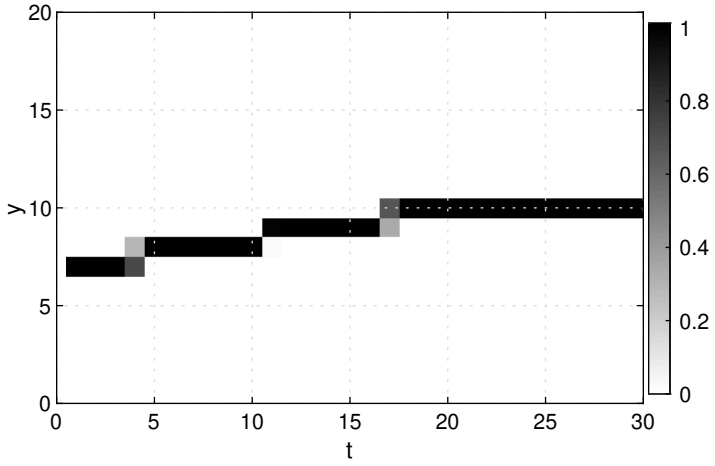
To contrast operator consistency, we consider the metric (2.82) for two datasets and scatter angles. For the smooth case, $\mu(x)$ is the modified Shepp-Logan phantom and $g(y)$ is the corresponding BRT which was generated analytically then sampled. For the impulsive case we set $\mu(x) = 1$ and $g(y)$ is a sparse binary image with 600 randomly spaced nonzero pixels. Resulting percentage error, using (2.82), are depicted in Table 2.2.

Table 2.2 demonstrates consistent performance using the direct implementation over both smooth and non-smooth datasets and scatter angles. The resulting percentage error is on the order of double-precision quantization errors. For smooth datasets, the Fourier method performs similarly. However, this percentage error increases for datasets characterized by impulsive noise. The percentage error increases significantly for the real Fourier method, even for the smooth dataset.

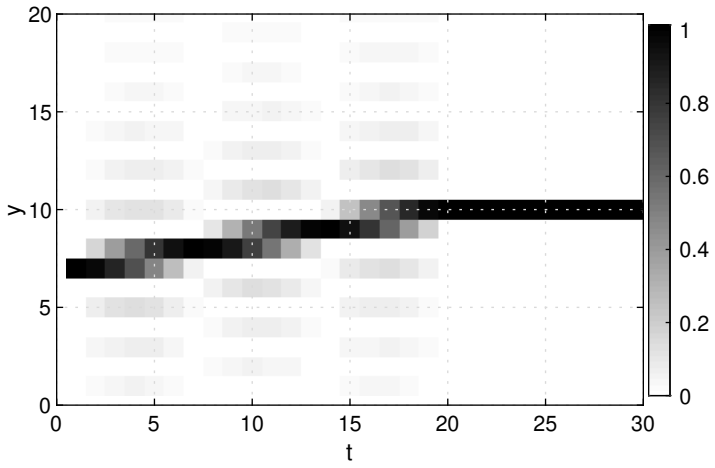
To further investigate the difference between the direct and Fourier operator implementations, we consider the transform of an image with a single nonzero pixel (20,10) with a weight of 1. We assume square pixels with an edge length of 1. Again we use $\theta_i = \pi$ and $\theta_j = \pi/20$. Transform results for both the direct and Fourier method are shown in Figure (2.10).

Discrepancies between Figures 2.10a and 2.10b are explained by different analytic interpretations of the sampled image. For the direct operator, images are interpreted as piecewise continuous. In this case, the analytic interpretation of the image is a two-dimensional rectangular function. For the Fourier operator, images are interpreted as band-limited by the sampling rate. In this case, the analytic interpretation of the image is a two-dimensional sinc function. Obviously the support of a two-dimensional rectangular function differs from the support of a two-dimensional sinc function. The small nonzero values in Figure 2.10b

are associated with ray integrals through sidelobes of the sinc function. In Figure 2.10a the corresponding samples are zero because a rectangular function has no sidelobes.



(a) Direct



(b) Fourier

Figure 2.10: BRT data computed using different forward operator implementations on the same unit-impulse image. Results computed using direct and Fourier implementations are shown in Figures 2.10a and 2.10b, respectively. Small nonzero samples in 2.10b are due to a bandlimited interpretation of the sampled data.

Chapter 3

Joint Image Estimation from Incomplete Data

X-ray scatter imaging has a long history [27] with many potential applications (see [12, 28] and references therein). The joint image estimation problem is important as both scatter density and attenuation are media dependent. Solving the joint problem could improve object identification [13] and avoid separate measurement systems. This problem has remained largely avoided as authors focus on scatter density alone. For example, the attenuation image has been trivialized as constant [28, 29] or negligible [12–14]. To our knowledge, we are the first to address joint estimation of scatter and attenuation from noisy measurements while accounting for regions with zero scatter.

Three measurement geometries have received attention recently. We focus on the selected volume tomography (SVT) [30] which has been referred to as a translation-only measurement geometry [24] and the BRT measurement system [3]. SVT is distinguished from rotational measurement geometries [7, 9, 10, 21, 23] which are related to secondary radiation tomography [30]. We briefly extend our notation to a rotational measurement geometry in Appendix G.4. Additionally, we distinguish the SVT measurement system from coded apertures [12–14]. Coded aperture measurement geometries share commonality with SVT in that both avoid rotation, and the single scatter approximation is utilized. However, coded aperture measurements do not distinguish a unique broken ray path. Further, the source-detector paths do not all share a common plane and elude two-dimensional analysis. Each

detector pixel in a coded aperture measurement geometry can be summarized as integrating sparse samples in a high-dimensional space comprising SVT measurements at multiple scatter angles.

The BRT was introduced in the context of single-scatter optical tomography [2] simulating a measurement geometry similar to SVT under an appropriate change of variables [1]. Early work relied on two significant assumptions: isotropic scatter, and strictly positive scatter everywhere. The first assumption simplifies aggregation of of data collected from multiple scatter angles. Isotropic scatter applies to x-ray fluorescence imaging, but additional conditions are required for Bragg and Compton scatter imaging [11, 24]. The second assumption ensures the log of the data is finite everywhere. This assumption is convenient since the attenuation and scatter terms separate in the log of the data. Linear combinations of the log-data, using multiple scatter angles, can be used to isolate the attenuation terms [3, 4]. The remaining log-residue of the data is not a sample of the attenuation image. Rather, it comprises integrals of the image along the single-scatter path. The BRT is this integral along the single-scatter path.

BRT data are typically not available directly. Instead, the BRT appears in the exponent in the data model according to Beer's law. Prior analytic BRT inversion strategies have focused on recovering the attenuation image from the log-residue of the data. Solutions are well-defined for analytic images with bounded support [1, 7, 8, 10] and with nonzero scatter (at least over the support of the attenuation image [3]). Under these constraints, joint image recovery has been simulated [1, 4].

Analytic BRT inversion presents several numerical challenges exacerbated by noise, sampling, and missing data (e.g. regions of zero scatter). Effects on analytic attenuation image reconstruction are shown in Figure 3.1. Analytic inversion of noise-free data yields images with few sampling artifacts (Figure 3.1 (b)). Recovery requires boundary conditions to be enforced [1]. Small sampling artifacts are a consequence of both truncated data and the nontrivial nullspace of the forward operator [24]. Enforcing boundary conditions is challenging for noisy data; removing them degrades reconstruction quality (see Figure 3.1 (c)). Additionally, some practical applications require imaging objects with regions of zero scatter (e.g. luggage scanning). Without accounting for such cases, the only explanation is highly attenuating regions obscuring regions with missing data. Invalidating the positive

scatter assumption considerably degrades reconstruction quality as depicted in 3.1 (d). As a final complication we consider Poisson-distributed data. Even with positive scatter everywhere, performance suffers using analytic techniques (see Figure 3.1 (e)). Results for Poisson-distributed data with nonnegative scatter are shown in Figure 3.1 (f).

3.1 Problem Formulation

3.1.1 Measurement Geometry

We consider a two-dimensional, single-scatter measurement geometry where the data are indexed by scatter location $y \in \mathcal{Y}$. Additionally, for each scatter location, we assume data are available for multiple scatter angles. An example measurement geometry is shown in Figure 3.2. In this case, a pencil-beam source is directed through the media of interest. Along the beam, scattering is observed in two direction $\{\theta_1, \theta_2\}$. Along the incident beam, multiple scatter locations are resolved simultaneously using a collimated array of detectors. Translating the source and detector arrays as a system allows scanning over all sample points $\mathcal{Y}$. This schema could be extended to three dimensions using fan beam illumination and two-dimensional collimated detector panels.

As a matter of notation, we use $\boldsymbol{\theta} = (\theta_s, \theta_d)$ to represent an ordered pair of source and detector directions, respectively. Let $\mathcal{I} \coloneqq \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\}$ represent the collection of source-detector pairs for which data are available. For the measurement system depicted in Figure 3.2, we have $\mathcal{I} = \{(\theta_0, \theta_1), (\theta_0, \theta_2)\}$. Here we assume only one source-direction, θ_0, is utilized. To simplify the notation, we will simply refer to broken ray directions using the index i which implies $\boldsymbol{\theta}_i = (\theta_0, \theta_i)$. Each pair $(i, y) \in \mathcal{I} \times \mathcal{Y}$ uniquely define a path, from source to detector, through the scatter location y.

3.1.2 Data Model

Our objective is the recovery of two images, attenuation and scatter density, from a single dataset. For clarity we use separate discretizations of the images using $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

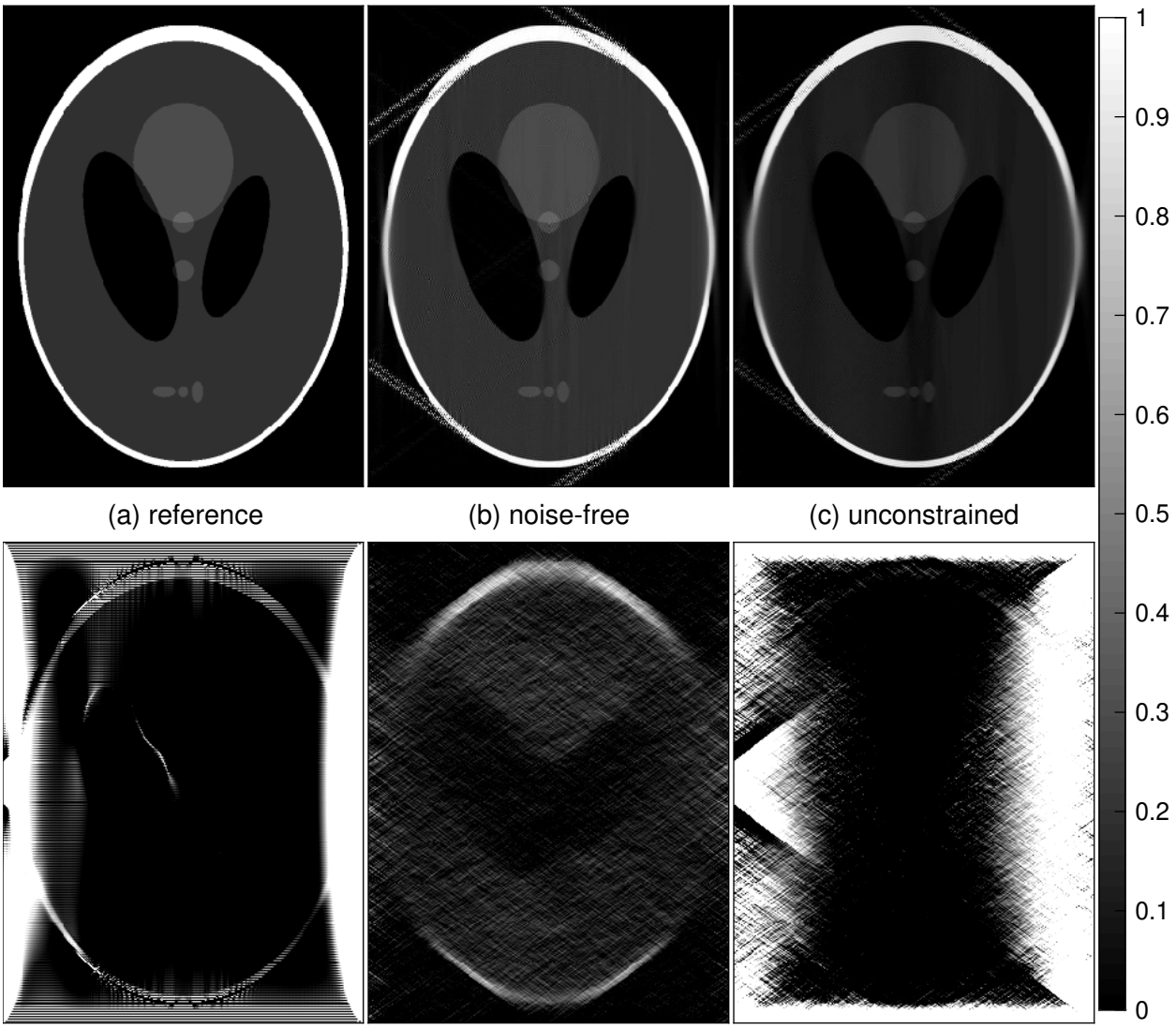


Figure 3.1: Degradation of analytic attenuation image reconstruction quality. The reference attenuation image $\mu(x)$ is shown in Figure 3.1(a). The remaining figures contain reconstructions from: (b), noise-free data; (c), noise-free data without enforcing boundary conditions; (d), noise-free data with missing samples (regions of zero scatter); (e) Poisson-distributed data with uniform scatter; and (f) Poisson-distributed data with missing samples.

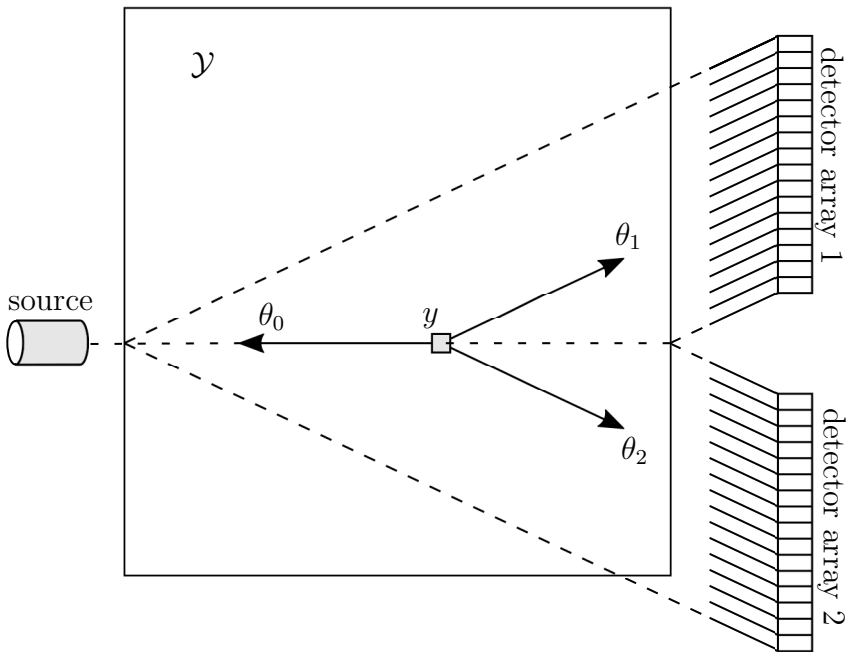


Figure 3.2: Measurement geometry with pencil-beam source and two detector arrays. We use $y \in \mathcal{Y}$ to indicate the location of the scatter event. The direction θ_0 indicates the direction of the source from the scatter location y. The directions θ_i, $i > 0$, indicate the observed scatter direction using the i^{th} detector array.

to index the set of attenuation and scatter image voxels, respectively. Later we will enforce $\mathcal{X} = \mathcal{Y}$, enabling computationally efficient algorithms. Let $\mathcal{A} \coloneqq \{\boldsymbol{\mu} : \mu(x) \in \mathbb{R}_{\geq 0}\}$ and $\mathcal{S} \coloneqq \{\boldsymbol{\alpha} : \alpha(y) \in [0,1]\}$ represent the set of possible attenuation and scatter images, respectively. Let $\mathbf{d}$ represent the available data where $d_i(y) \in \mathbb{R}_{\geq 0}$ for each $(i,y) \in \mathcal{I} \times \mathcal{Y}$.

We model the data $d_i(y)$ as Poisson distributed with mean

$$g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) \coloneqq \beta_i(y) + I_0(y)\alpha(y) \exp\left(- \sum_{x \in \mathcal{X}} h_i\left(y|x\right) \mu(x) \right). \tag{3.1}$$

Here we use $I_0(y) > 0$ and $\beta(y) \geq 0$ to represent the known source intensity and background counts, respectively. The exponential term is a numeric approximation of Beer's law along our broken ray path through the attenuation image. We explicitly refer to the summation as the discrete broken ray transform. We assume both the image and forward transform are finite everywhere.

The log-likelihood function of the data, parameterized by $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$, is

$$l(\mathbf{d} : \boldsymbol{\alpha}, \boldsymbol{\mu}) \coloneqq \sum_{i \in \mathcal{I}} \sum_{y \in \mathcal{Y}} d_i(y) \ln g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) - g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) \tag{3.2}$$

excluding constant terms of the data alone. Maximizing the log-likelihood is equivalent to minimizing the Csiszár I-divergence between the data and the mean

$$I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right)\big) \coloneqq \sum_{i \in \mathcal{I}} \sum_{y \in \mathcal{Y}} d_i(y) \ln \frac{d_i(y)}{g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu})} - d_i(y) + g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}). \tag{3.3}$$

Here we use $\mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right) \coloneqq \{g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) : i \in \mathcal{I}\}$. This form is a generalization of the Kullback-Leibler divergence [31].

3.1.3 Objective Functions and Surrogates

Joint image recovery from (3.3) is ill-posed due to conditioning of the BRT forward operator [24] in the exponent (3.1), noise, and scaling. To improve conditioning of this problem we

incorporate two regularization terms in the objective function

$$J(\boldsymbol{\alpha}, \boldsymbol{\mu}) \coloneqq I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right)\big) + \lambda_\alpha R(\boldsymbol{\alpha}) + \lambda_\mu R(\boldsymbol{\mu}). \tag{3.4}$$

Here R is a convex regularization function (further conditions given in Appendix F.3). The scalars λ_α and λ_μ emphasize regularization of the corresponding images.

Direct minimization of (3.4) remains difficult due to its high dimensionality and interdependence of the image pixels. To make the problem tractable, we employ two techniques. First, we use separable surrogate functions for the terms in (3.4). Separability here means the gradients separate as functions of single image samples. The ensuing algorithm is highly parallelizable, allowing each pixel update to be computed in parallel. The second technique we employ is alternating updates between the scatter and attenuation images. The use of surrogate functions guarantees monotonic reduction in the objective while alternating image updates.

We use a surrogate for the data fidelity term

$$\overline{D}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}\right) \geq I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\hat{\boldsymbol{\alpha}}, \boldsymbol{\mu}\right)\big), \quad \forall \boldsymbol{\mu} \in \mathcal{A} \tag{3.5}$$

$$\overline{D}\left(\hat{\boldsymbol{\mu}} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}\right) = I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}\right)\big), \tag{3.6}$$

which is given by (F.30) and derived in Appendix F.2.

Additionally, we consider a surrogate for the regularization term

$$\overline{R}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right) \geq R\left(\boldsymbol{\mu}\right), \quad \forall \boldsymbol{\mu} \in \mathcal{A} \tag{3.7}$$

$$\overline{R}\left(\hat{\boldsymbol{\mu}} : \hat{\boldsymbol{\mu}}\right) = R\left(\hat{\boldsymbol{\mu}}\right)., \tag{3.8}$$

which is given by (F.45) and derived in Appendix F.3.

Making use of these surrogate functions, we define two objectives

$$\overline{J}_\alpha(\boldsymbol{\alpha} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}) \coloneqq I\Big(\mathbf{d} \,\Big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \hat{\boldsymbol{\mu}}\right)\Big) + \lambda_\alpha \overline{R}(\boldsymbol{\alpha} : \hat{\boldsymbol{\alpha}}) \tag{3.9}$$

$$\overline{J}_\mu(\boldsymbol{\mu} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}) \coloneqq \overline{D}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}\right) + \lambda_\mu \overline{R}(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}), \tag{3.10}$$

which lead to an iterative update algorithm

$$\boldsymbol{\alpha}^{(k+1)} = \arg\min_{\boldsymbol{\alpha}\in\mathcal{S}} \overline{J}_{\alpha}(\boldsymbol{\alpha} : \boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)}) \tag{3.11}$$

$$\boldsymbol{\mu}^{(k+1)} = \arg\min_{\boldsymbol{\mu}\in\mathcal{S}} \overline{J}_{\mu}(\boldsymbol{\mu} : \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)}). \tag{3.12}$$

This approach guarantees monotonic reduction of the regularized objective and convergence to a local minimum.

Lemma 3.1. *Monotonic reduction of $J(\boldsymbol{\alpha}, \boldsymbol{\mu})$ is guaranteed whenever a local objective,* (3.9) *or* (3.10)*, is reduced.*

Proof. Combining the definitions (3.4) and (3.9), and making use of (3.7), we have

$$J(\boldsymbol{\alpha}, \hat{\boldsymbol{\mu}}) - \lambda_{\mu} R\,(\hat{\boldsymbol{\mu}}) \leq \overline{J}_{\alpha}(\boldsymbol{\alpha} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}). \tag{3.13}$$

According to (3.8), we have equality when $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$, such that

$$J(\boldsymbol{\alpha}, \hat{\boldsymbol{\mu}}) - J(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}) \leq \overline{J}_{\alpha}(\boldsymbol{\alpha} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}) - \overline{J}_{\alpha}(\hat{\boldsymbol{\alpha}} : \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}). \tag{3.14}$$

Therefore, any $\boldsymbol{\alpha}$ reducing $\overline{J}_{\alpha}$ guarantees a reduction in J. Further, the improvement in the objective is bounded by the improvement to the local surrogate. The same can be shown for any $\boldsymbol{\mu}$ reducing $\overline{J}_{\mu}$. □

Applying Lemma 3.1, alternating updates ensure

$$J(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k+1)}) \leq J(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)}) \leq J(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)}). \tag{3.15}$$

Iterative updates result in a sequence of costs that are monotonically decreasing and bounded from below, since $J(\boldsymbol{\alpha}, \boldsymbol{\mu}) \geq 0$. Convergence of this sequence is guaranteed.

Lemma 3.2. *Convergence of* (3.11) *and* (3.12) *implies a fixed point.*

Proof. The divergence (3.3) is convex with respect to $\boldsymbol{\alpha}$ (see Appendix F.1), and $\overline{R}$ is strictly convex (see Appendix F.3). Therefore, the local surrogate $\overline{J}_{\alpha}$ is strictly convex over $\boldsymbol{\alpha} \in \mathcal{S}$,

with a unique minimizer $\boldsymbol{\alpha}^*$. When (3.11) does not improve the objective, we have

$$\overline{J}_\alpha(\boldsymbol{\alpha}^* : \boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)}) = \overline{J}_\alpha(\boldsymbol{\alpha}^{(k)} : \boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)}) \implies \boldsymbol{\alpha}^* = \boldsymbol{\alpha}^{(k)} \tag{3.16}$$

such that $\boldsymbol{\alpha}^{(k)}$ is a fixed point. The same can be shown for $\overline{J}_\mu$. □

Applying Lemma 3.2, equality in (3.15) implies $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)}$ and $\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)}$.

3.2 Algorithms

Our iterative joint estimation approach is a two-step process summarized by Algorithm 4. The first step requires computation of the forward BRT, and the second requires computation of the backward BRT.

Algorithm 4 JointEstimate: Iterative algorithm for joint image estimation

Require: $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\mu}^{(0)}$
1: **for** $k = 0, 1, 2, \ldots$ **do**
2: $\quad \boldsymbol{\alpha}^{(k+1)} = \textsc{ScatterUpdate}\left(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)}\right)$
3: $\quad \boldsymbol{\mu}^{(k+1)} = \textsc{AttenuationUpdate}\left(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)}\right)$
4: **end for**

3.2.1 Regularized Scatter Update

The objective function (3.9) is strictly convex with respect to $\boldsymbol{\alpha}$ over $\mathcal{S}$ (see Appendices F.1 and F.3). Therefore, each $\alpha(y)$ is either 0 or the solution to an unconstrained minimization problem. Nonzero voxels are then determined by setting the gradient of (3.9) equal to zero and solving for $\alpha(y)$. Expanding (3.9), we have

$$\frac{\partial \overline{J}_\alpha(\boldsymbol{\alpha} : \boldsymbol{\alpha}^{(k)}, \boldsymbol{\mu}^{(k)})}{\partial \alpha(y)} = \frac{\partial I\left(\mathbf{d} \,\middle\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}^{(k)}\right)\right)}{\partial \alpha(y)} + \lambda_\alpha \frac{\partial \overline{R}(\boldsymbol{\alpha} : \boldsymbol{\alpha}^{(k)})}{\partial \alpha(y)}. \tag{3.17}$$

The first and second terms on the right-hand side are given by (F.3) and (F.46), respectively. Expanding these terms, we find the gradient is separable with respect to $\alpha(y)$. Each $\alpha^{(k+1)}(y)$

can be determined in parallel solving

$$0 = \sum_{i \in \mathcal{I}} \left(\dot{g}_i(y : \boldsymbol{\mu}^{(k)}) - \frac{d_i(y)\dot{g}_i(y : \boldsymbol{\mu}^{(k)})}{\alpha(y)\dot{g}_i(y : \boldsymbol{\mu}^{(k)}) + \beta_i(y)} \right) + \lambda_\alpha \left(c_1(y : \boldsymbol{\alpha}^{(k)}) + 2c_2(y : \boldsymbol{\alpha}^{(k)}) \left(\alpha(y) - \alpha^{(k)}(y) \right) \right). \quad (3.18)$$

Here $\dot{g}$ is given by (F.2). For each $i \in \mathcal{I}$, computing $\dot{g}_i$ requires computing the forward BRT of $\boldsymbol{\mu}^{(k)}$. The functions c_1 and c_2 refer to (F.43) and (F.44), respectively. These have been re-appropriated for use with the scatter image and depend on the previous estimate $\boldsymbol{\alpha}^{(k)}$. This process is described in Algorithm 5. We emphasize (3.18) has at most one positive solution since (3.9) is strictly convex over $\boldsymbol{\alpha} \in \mathcal{A}$.

Algorithm 5 ScatterUpdate: Single update of the scatter image.

Require: $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}$
Ensure: $\boldsymbol{\alpha}$
1: **for each** scatter angle i **do**
2: $\quad b_i(y) = \sum_{x \in \mathcal{X}} h_i(y|x)\hat{\boldsymbol{\mu}}(x)$ ▷ Forward BRT
3: $\quad \dot{g}_i(y) = I_0(y) \exp\left(-b_i(y)\right)$
4: **end for**
5: Compute $c_1\left(y : \hat{\boldsymbol{\alpha}}\right)$ using Eq. (F.43)
6: Compute $c_2\left(y : \hat{\boldsymbol{\alpha}}\right)$ using Eq. (F.44)
7: **for each** point y **do**
8: $\quad$ Compute $\alpha(y)$ by solving Eq. (3.18)
9: **end for**

3.2.2 Regularized Attenuation Update

The objective function (3.10) is strictly convex with respect to $\boldsymbol{\mu}$ over $\mathcal{A}$ (see Appendices F.2 and F.3). Therefore, each $\mu(x)$ is either 0 or the solution to an unconstrained minimization problem. Nonzero voxels are then determined by setting the gradient of (3.10) equal to zero and solving for $\mu(x)$. Expanding (3.10), we have

$$\frac{\partial \overline{J}_\mu(\boldsymbol{\mu} : \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)})}{\partial \mu(x)} = \frac{\partial \overline{D}(\boldsymbol{\mu} : \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)})}{\partial \mu(x)} + \lambda_\mu \frac{\partial \overline{R}\left(\boldsymbol{\mu} : \boldsymbol{\mu}^{(k)}\right)}{\partial \mu(x)}. \quad (3.19)$$

The first and second terms on the right-hand side are given by (F.31) and (F.46), respectively. Expanding these terms, we find the gradient is separable with respect to $\mu(x)$. Each $\mu^{(k+1)}(x)$ can be determined in parallel solving

$$0 = b_1(x : \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)}) - b_2(x : \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\mu}^{(k)}) \exp\Big(- Z_0\big(\mu(x) - \mu^{(k)}(x)\big)\Big) + \lambda_\mu \left(c_1(x : \boldsymbol{\mu}^{(k)}) + 2c_2(x : \boldsymbol{\mu}^{(k)}) \left(\mu(x) - \mu^{(k)}(x)\right)\right). \quad (3.20)$$

The functions b_1 and b_2 are given by (F.28) and (F.29), respectively. Here we emphasize their dependence on prior estimates $\boldsymbol{\alpha}^{(k+1)}$ and $\boldsymbol{\mu}^{(k)}$. Prior estimates are used to compute $\hat{q}_i(y, 1)$ using (F.6), which determine $\hat{p}_i(y, 1)$ using (F.9b). The adjoint BRT is required to compute b_1 and b_2 from $\hat{p}_i(y, 1)$ and $\hat{q}_i(y, 1)$, respectively. This process is described in Algorithm 6.

Algorithm 6 AttenuationUpdate: Single update of the attenuation image.

Require: $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}$
Ensure: $\boldsymbol{\mu}$
$\mathbf{b}_1, \mathbf{b}_2 \leftarrow \mathbf{0}$
for each scatter angle i **do**
  Compute $\hat{q}_i(y)$ using Eq. (F.6)
  Compute $\hat{p}_i(y)$ using Eq. (F.9b)
  $b_1(x)$ += $\sum_{y \in \mathcal{Y}} h_i(y|x)\hat{p}_i(y)$ ▷ Backward BRT
  $b_2(x)$ += $\sum_{y \in \mathcal{Y}} h_i(y|x)\hat{q}_i(y)$
end for
Compute $c_1\left(x : \hat{\boldsymbol{\mu}}\right)$ using Eq. (F.43)
Compute $c_2\left(x : \hat{\boldsymbol{\mu}}\right)$ using Eq. (F.44)
for each point x **do**
  Compute $\mu(x)$ by solving Eq. (3.20)
end for

3.3 Simulations

In the following we quantify performance of our algorithm using simulated data. First we demonstrate performance on a simple rectangular phantom from noise-free data. This case highlights challenges which manifest in more complicated datasets. Subsequently we demonstrate performance on the Shepp-Logan phantom. For the Shepp-Logan phantom we demonstrate performance on noisy data for both single image and joint image estimation.

3.3.1 Noise-Free Reconstruction for A Rectangular Phantom

Image reconstruction from noise-free data for simple phantoms helps characterize the performance of our algorithm. Here we consider a simple rectangular phantom with a binary attenuation image ($\mu(x) \in 0, 1$). We consider two cases distinguished by the scatter image. In the first case, we assume nonzero scatter everywhere

$$\alpha_{>0}(y) \coloneqq \sqrt{0.1 + 0.2\mu(y)}. \tag{3.21}$$

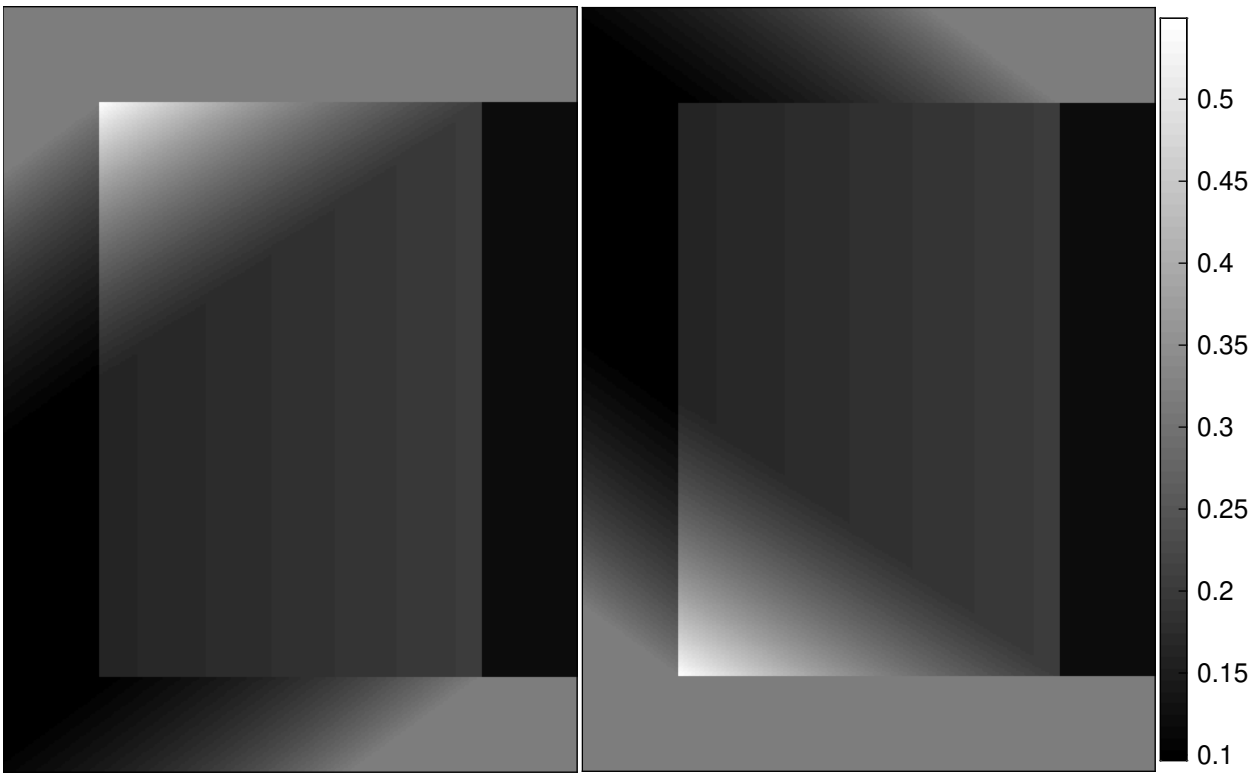
For the second case, we allow the scatter to be zero outside support of the attenuation image

$$\alpha_{\geq 0}(y) \coloneqq \sqrt{0.15\mu(y)}, \tag{3.22}$$

which implies missing data. This transform ensures spatial discontinuities occur in the same places for both images. Here we use source intensity I_0 =1e4, and background counts $b = 50$. To preserve scaling, we divide the data by I_0. Data are shown for the nonzero scatter and missing data case in Figure 3.3, and Figure 3.4, respectively.

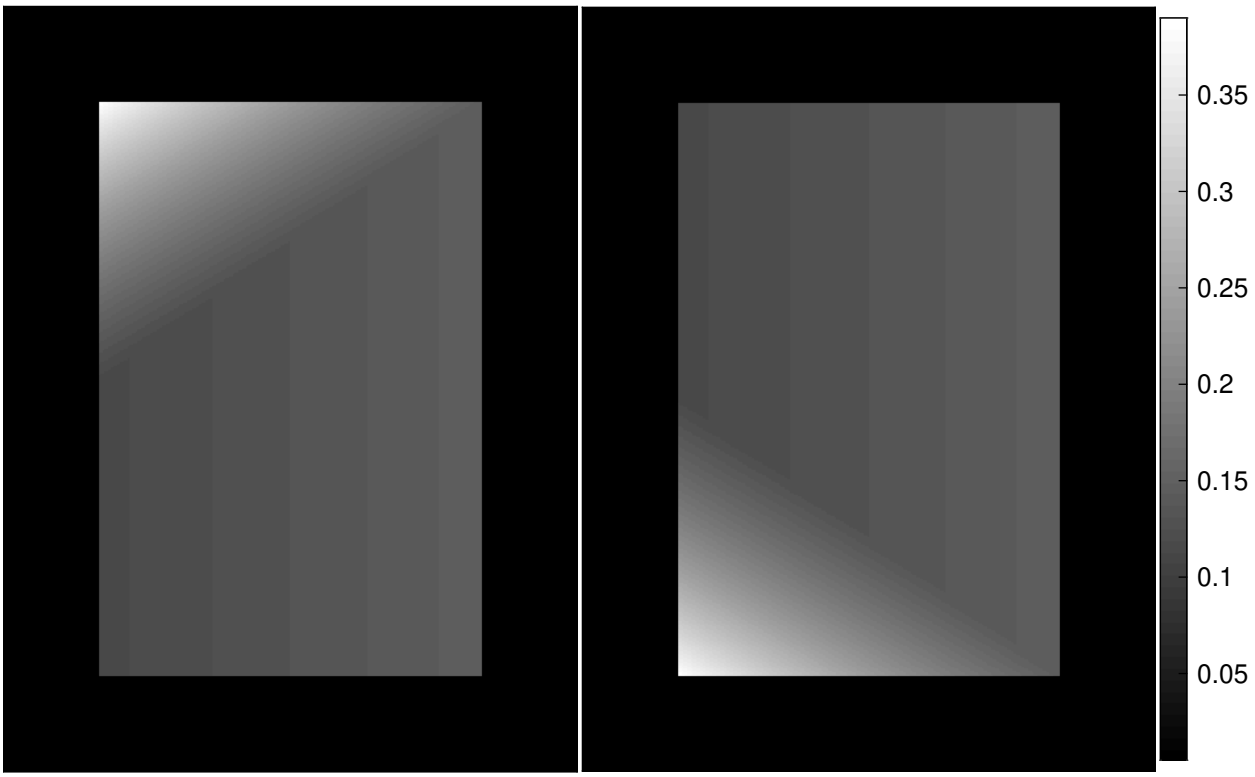
In Figure 3.3, the data are nonzero everywhere. Data associated with the positive scatter angle, Figure 3.3 (a), have the most signal in the top left corner of the rectangle. For this phantom, both scatter and attenuation are higher within the rectangle. Along the top of the rectangle, the scatter direction is oriented away from the centroid. These broken ray paths suffer less attenuation. Similar structures arise in Figure 3.4 over the rectangle indicating support of the scatter image. Data outside the support of the scatter image are nonzero due to b alone. However, they are not affected by attenuation and the gradations depicted in Figure 3.3 do not appear in 3.4 outside the rectangular region. This loss of information complicates reconstruction.

For the case $\alpha_{>0}$, reconstructed scatter and attenuation images are shown in Figure 3.5 and Figure 3.6, respectively. Reconstruction errors are shown in Figure 3.7. The scatter image is recovered reasonably well with small artifacts along the top and bottom of the rectangle. For the attenuation image, transitions orthogonal to the source direction are challenging to resolve spatially.



(a) Data, $\theta = +\pi/5$ (b) Data, $\theta = -\pi/5$

Figure 3.3: Simulated data for the rectangular phantom with strictly positive scatter.



(a) Data, $\theta = +\pi/5$ (b) Data, $\theta = -\pi/5$

Figure 3.4: Simulated data for the rectangular phantom with nonnegative scatter (missing data).

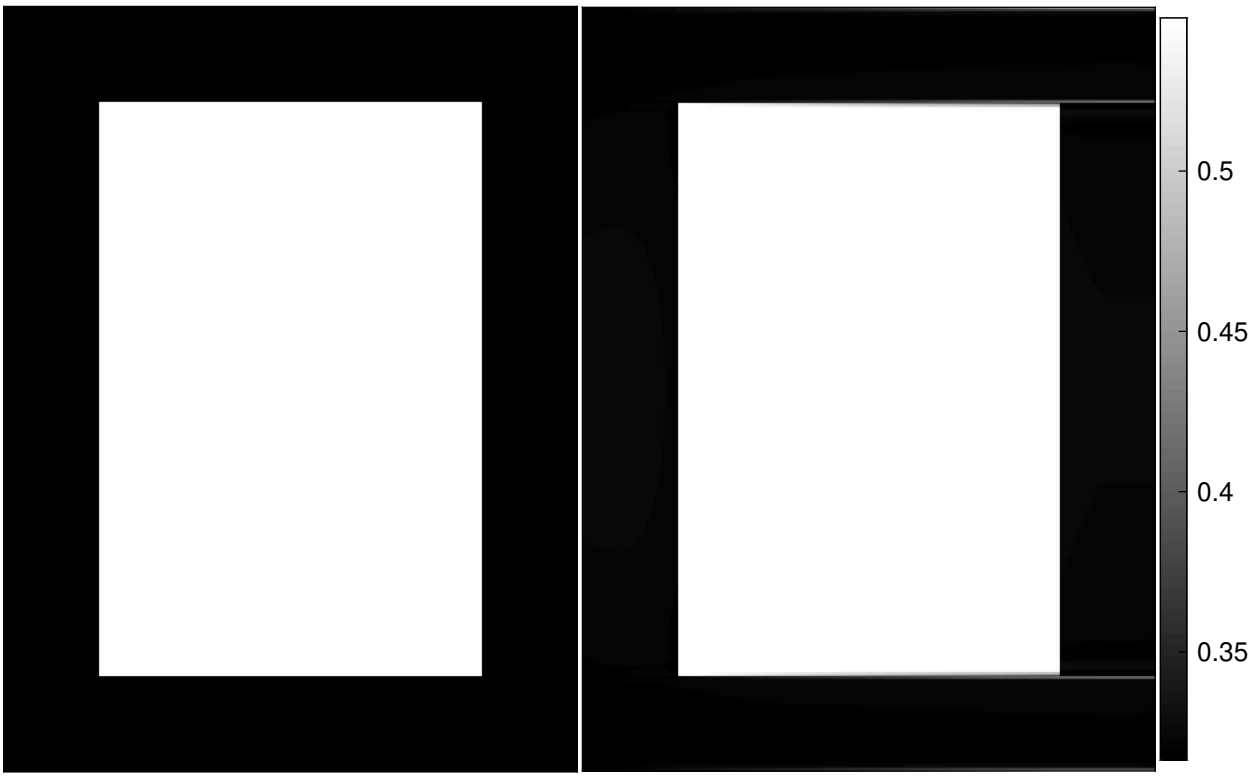


Figure 3.5: Recovered scatter image for the rectangular phantom with strictly positive scatter.

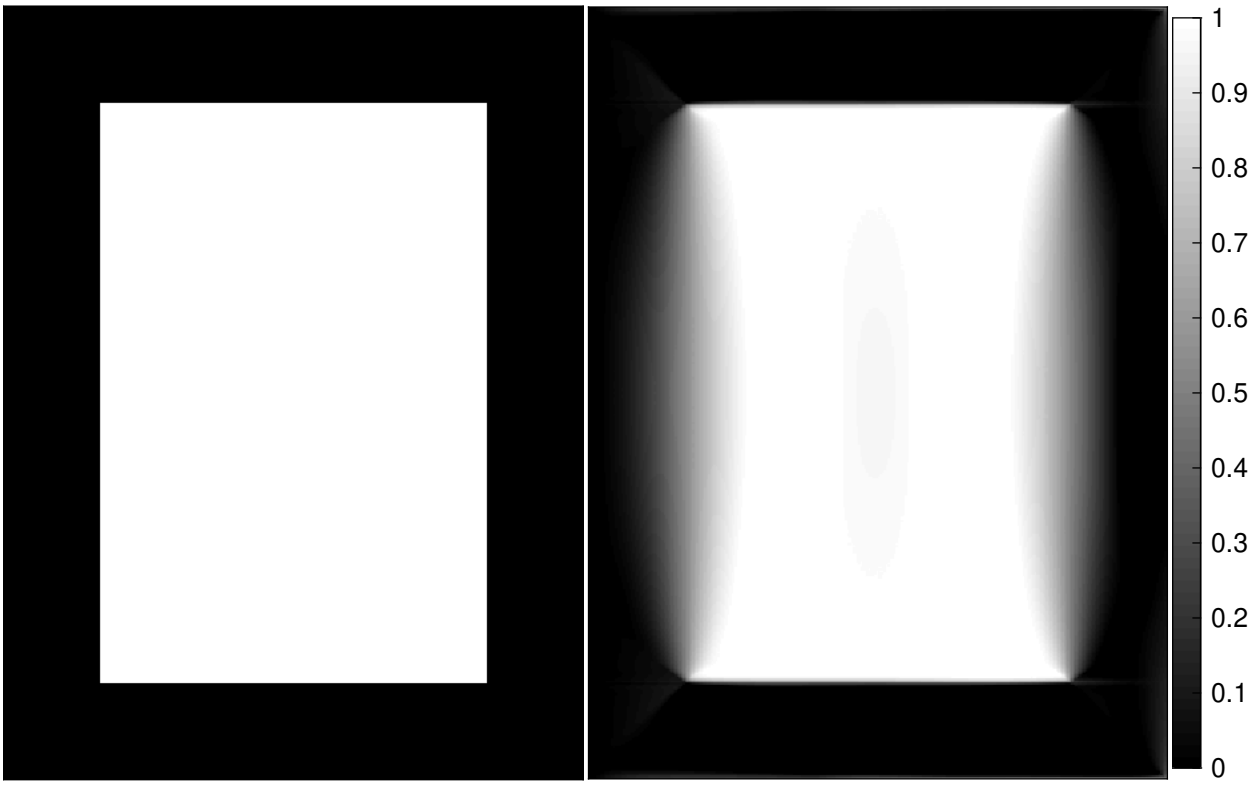
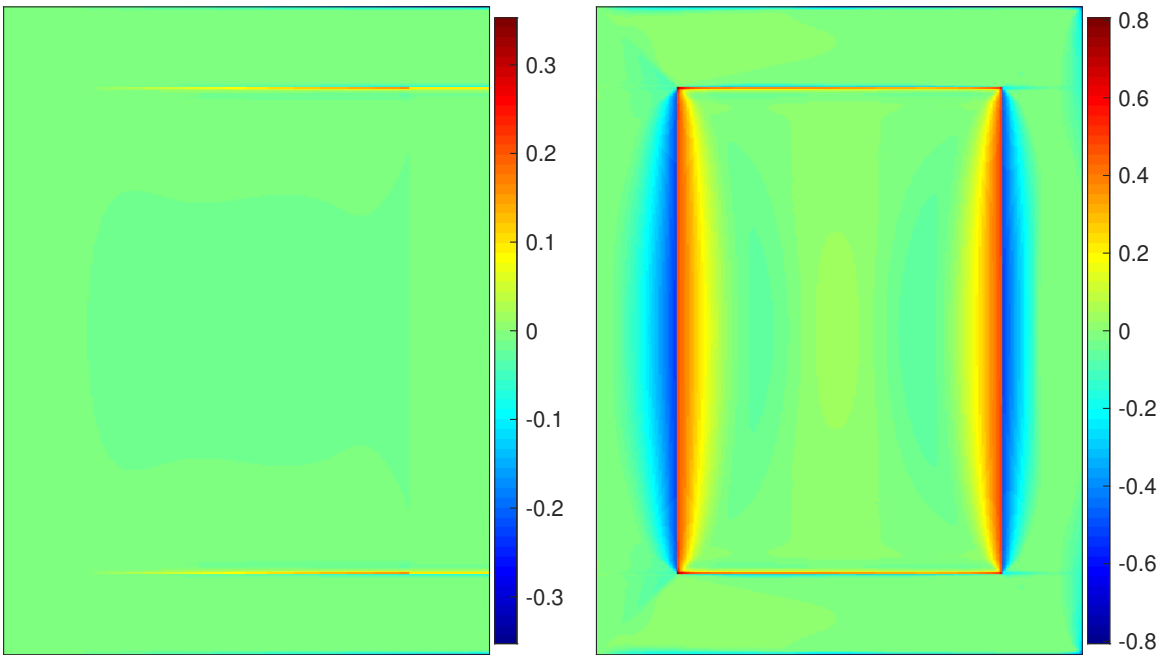


Figure 3.6: Recovered attenuation image for the rectangular phantom with strictly positive scatter.



(a) Scatter Error (b) Attenuation Error

Figure 3.7: Error in recovered images for the rectangular phantom with strictly positive scatter. Figure 3.7a and Figure 3.7b depict results for the the scatter and attenuation images as labeled. Poor resolution of vertical edges in the attenuation image estimate are due to strong attenuation of the BRT operator.

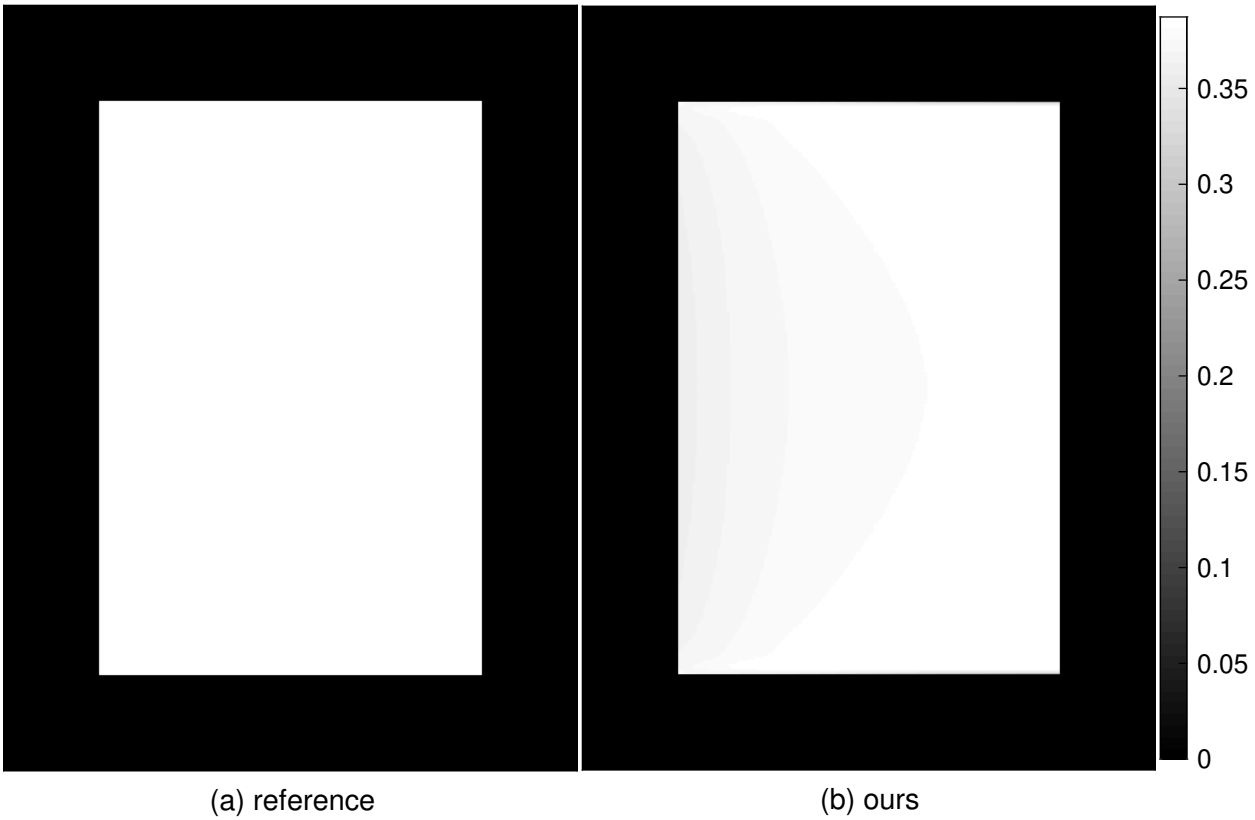


Figure 3.8: Recovered scatter image for phantom with missing data.

For the missing-data case $\alpha_{\geq 0}$, reconstructed scatter and attenuation images are shown in Figure 3.8 and Figure 3.9, respectively. Reconstruction errors are shown in Figure 3.10. Again, the scatter image is recovered reasonably well, although there is a smoothing of the leading edge. In contrast, the attenuation image suffers significant errors. There is some blurring of the leading edge, and the trailing edge is not resolved.

A vertical slice, bisecting the rectangular phantom is shown in Figure 3.11. Even for this noise-free case, we observe errors in the scatter image estimates. In Figure 3.11a, the errors are largest toward the middle where the broken ray paths suffer the most attenuation. The attenuation image, shown in Figure 3.11b demonstrates significant error.

So far we have depicted attenuation estimates only in the image space. It is also helpful to consider attenuation in the data space. In Figure 3.12, we contrast the BRT of the true attenuation image with the BRT of our estimated attenuation image. Differences in the data

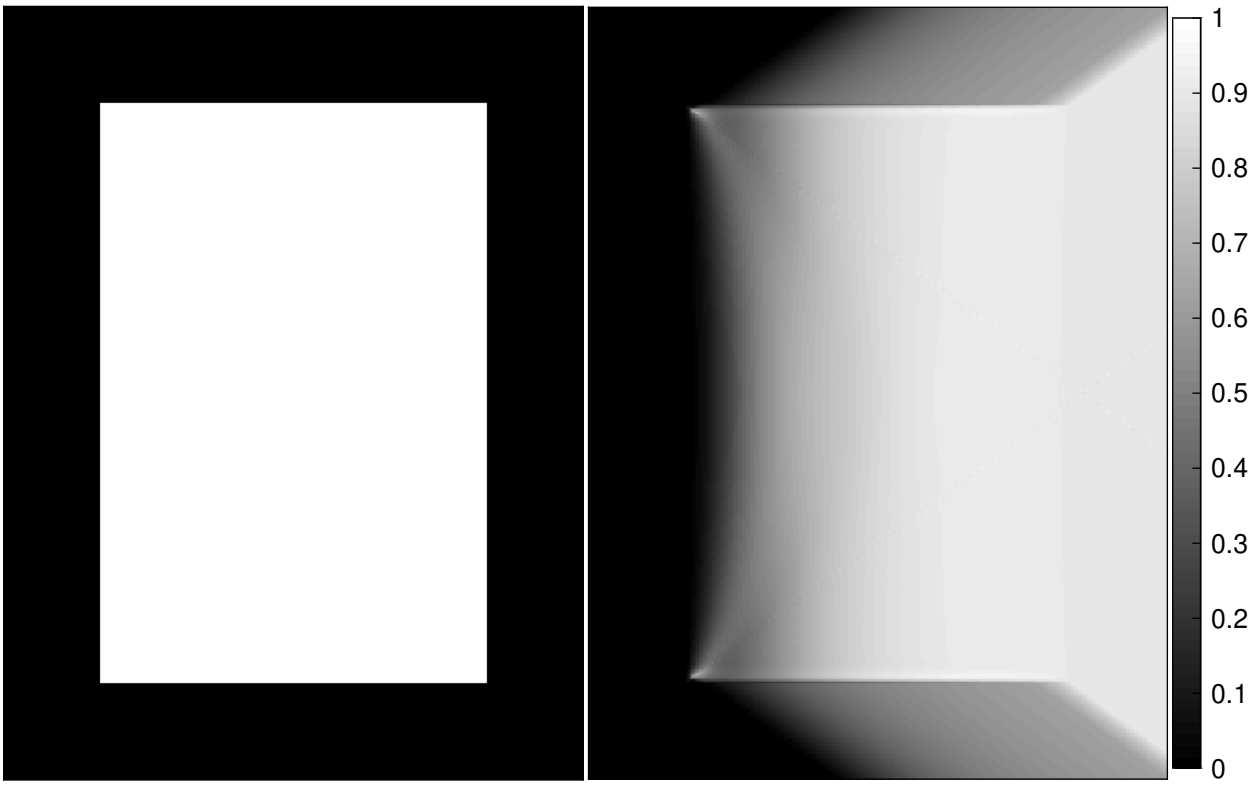
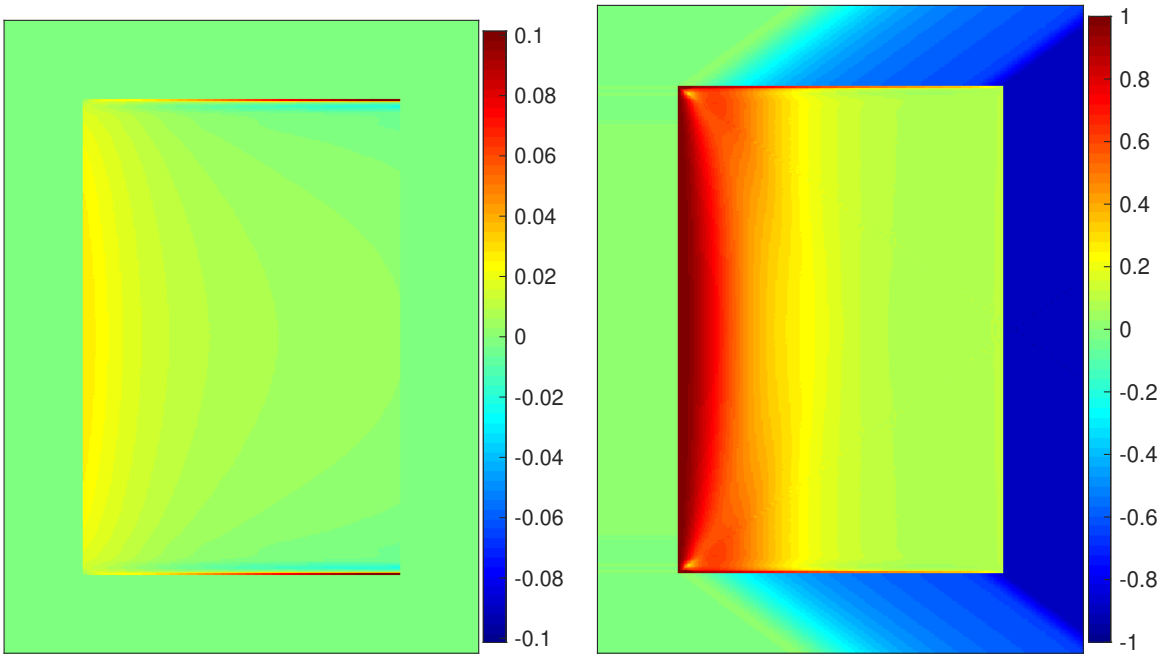
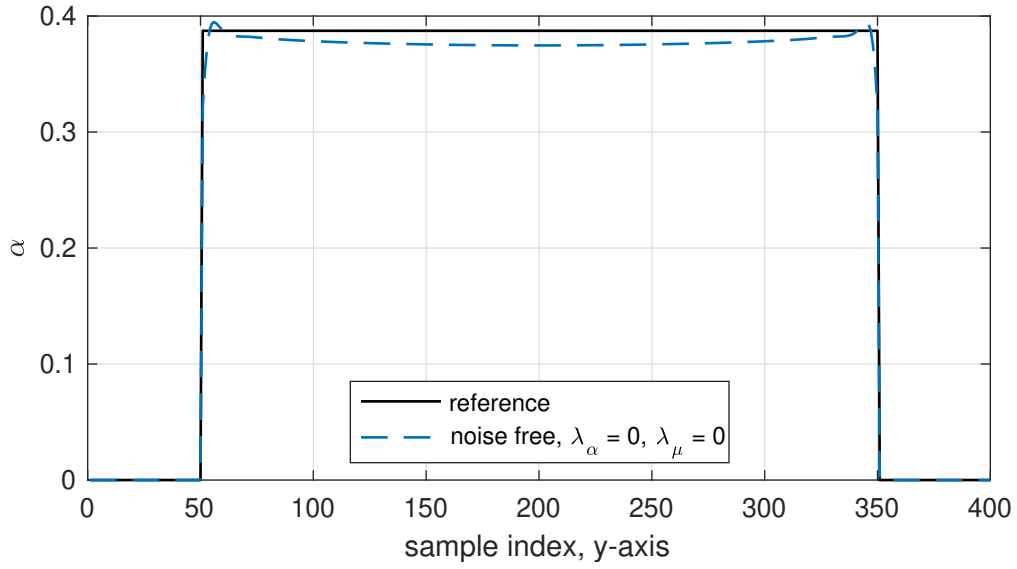


Figure 3.9: Recovered attenuation image for phantom with missing data.

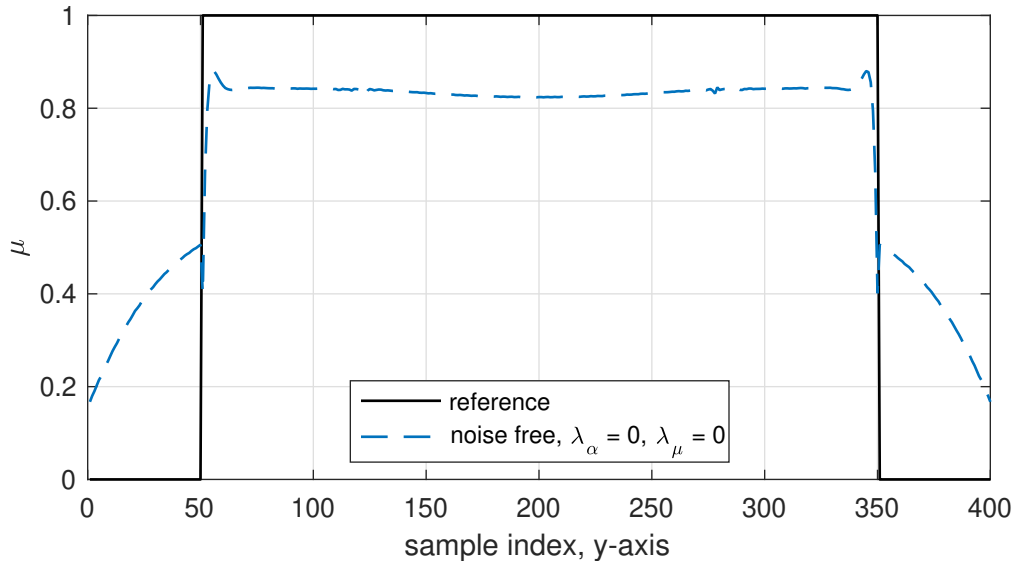


(a) Scatter Error (b) Attenuation Error

Figure 3.10: Error in recovered images for phantom with missing data. Figure 3.10a and Figure 3.10b depict results for the the scatter and attenuation images as labeled.

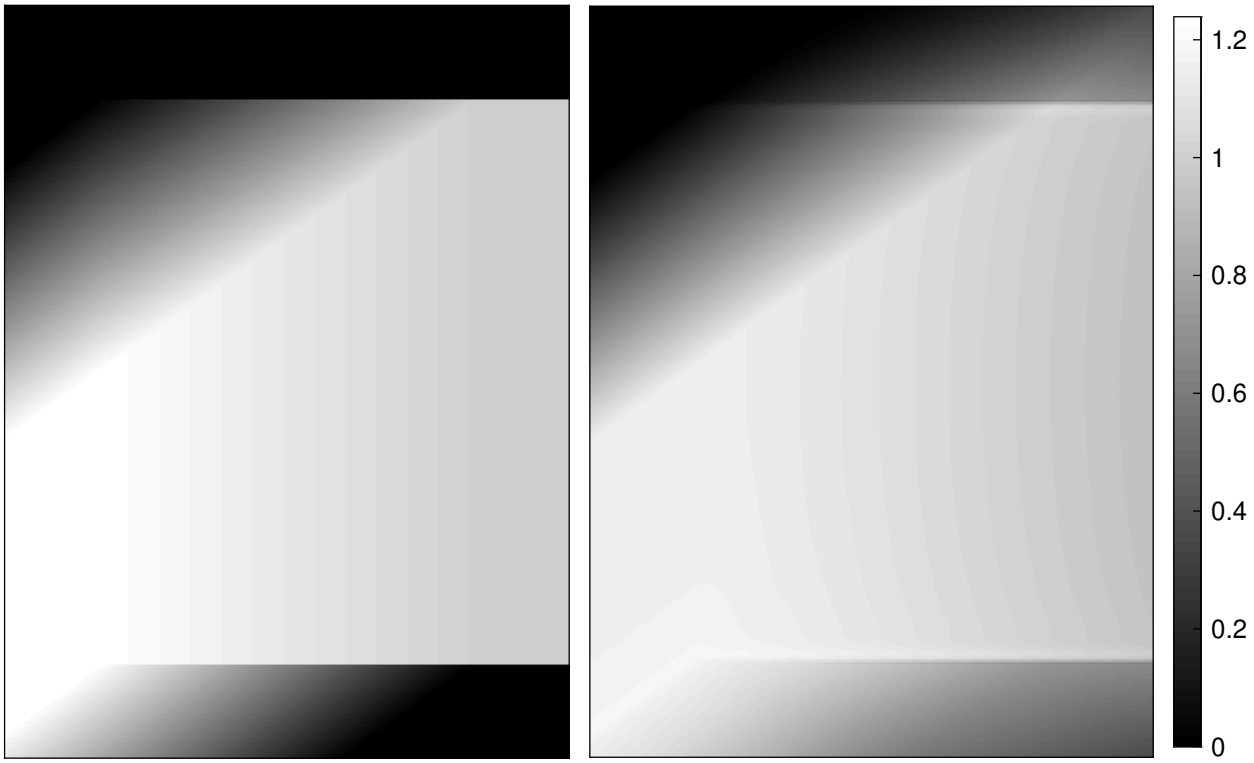


(a) scatter



(b) attenuation

Figure 3.11: Vertical slice of reconstructed images taken at the midpoint. Reconstructed from the positive scatter data shown in Figure 3.4.

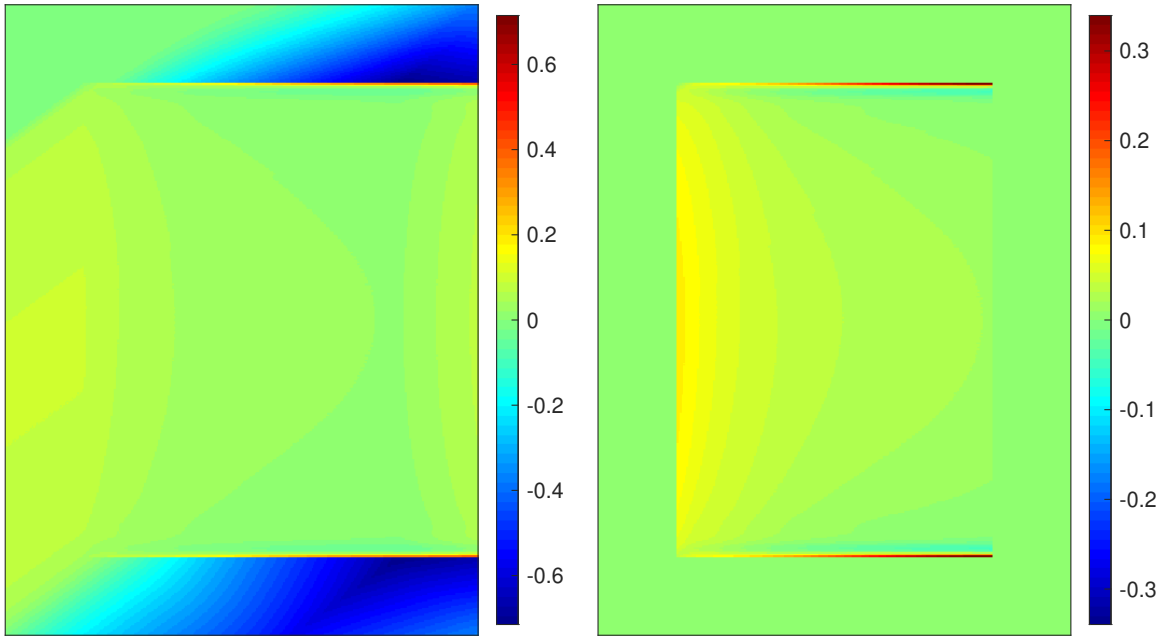


(a) BRT of true μ (b) BRT of estimated μ

Figure 3.12: BRT of true and estimated attenuation images.

space are small relative to the differences in the image space (c.f. Figure 3.9). This is one illustration of the poor conditioning of the inverse problem for attenuation image recovery.

It is also interesting to note attenuation errors in the data space are predominantly outside the support of the scatter image. Since the scatter image applies multiplicatively in the data model, errors outside this support have no effect on the data fidelity term. In Figure 3.13a we show errors in the data space. In Figure 3.13b we set all errors outside the support of the scatter image to 0. Over the support of the scatter image, the data space attenuation errors have a similar structure to the scatter image errors shown in Figure 3.10a. This similarity demonstrates ambiguity in the joint image estimation problem.



(a) Error (b) Error with mask

Figure 3.13: Error in BRT dataspace between true attenuation, $\boldsymbol{\mu}$, and our result, $\hat{\boldsymbol{\mu}}$, estimated from noise-free data. Figure 3.13a depicts the error $(G\boldsymbol{\mu})(x) - (G\hat{\boldsymbol{\mu}})$. Figure 3.13b depicts the errors subject to a mask, where all values outside the support of $\boldsymbol{\alpha}$ are set to 0.

Table 3.1: Simulation parameters and reconstruction hyperparameters

case	α	c_μ	β	I_0	θ_1	λ_μ	λ_α	δ
1	$\alpha_{\geq 0}$	1	50	1e+03	$\pi/5$	1e-03	1e-03	1e-02
2	$\alpha_{\geq 0}$	10	50	1e+04	$\pi/5$	4e-04	4e-04	1e-02

3.3.2 Shepp-Logan Phantom

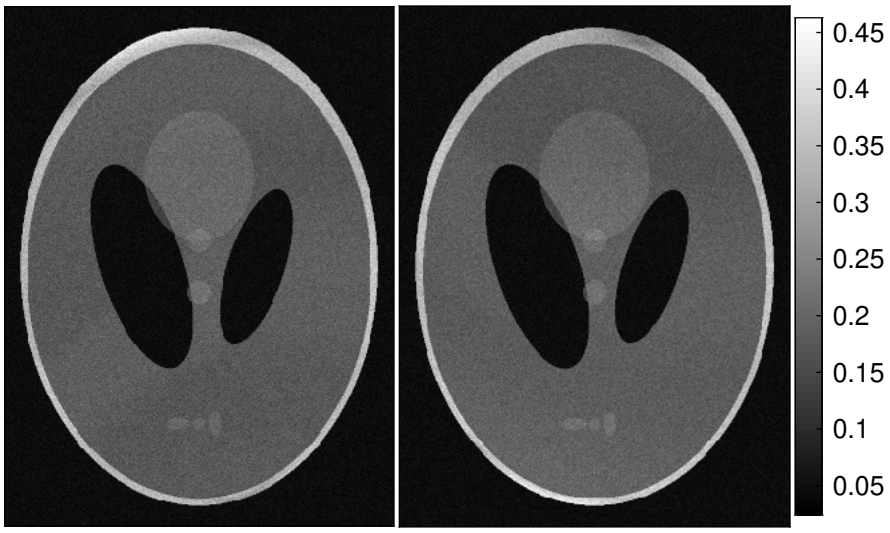
To assess performance of our algorithm in a more realistic setting, we consider simulated data using the Shepp-Logan phantom. In general, reconstruction quality will vary greatly depending on the problem parameters. Here we demonstrate results for a few cases intended to demonstrate some of this variability. The simulation parameters and reconstruction hyperparameters are listed in Table 3.1.

We simulate data at two symmetric scatter angles. For the attenuation image we use the modified Shepp logan phantom [25, 26]. From this analytic phantom, the BRT is computed analytically (see Appendix D.1) and sampled. For the scatter image, we use a nonlinear transform of the attenuation image using (3.22). The data are sampled uniformly as 400×300 square pixels pixels with edge lengths $5e-3$. The same sample locations are used for each scatter angle. Resulting data are shown in Figure 3.14 and Figure 3.15 for cases 1 and 2, respectively. In contrast, case 2 demonstrates stronger attenuation.

We first consider recovery of the attenuation image using analytic BRT inversion formulas. These formulas do not operate on detector counts directly. Rather, they assume availability of BRT data directly. For this we consider the crude estimate of the BRT data

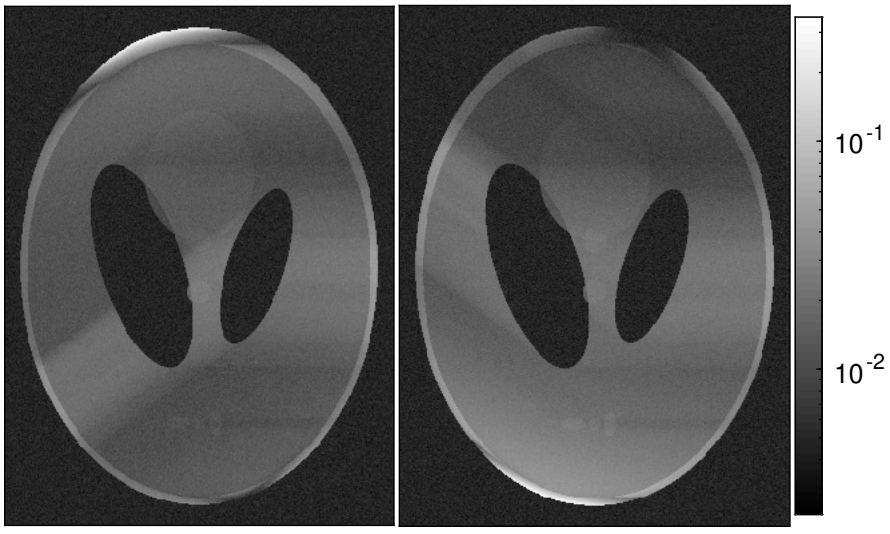
$$g_i(x) \approx \ln I_0 - \ln\left(\max(d_i(x), 1)\right). \tag{3.23}$$

The max operation is necessary as the data may be zero for any given pixel. The expression (3.23) highlights a significant limitation of the analytic inversion strategy: samples with no scatter are associated with high-attenuating broken ray paths. For analytic reconstruction of the attenuation image in the presence of unknown scatter, we require difference of the log data. This residue approximates the modified [24] (or signed [8]) broken ray transform of the attenuation image. Results are shown in Figure 3.1(f), for the case 1 data. The poor performance is largely due to the missing data and not simply the stochastic nature of the



(a) data, θ_1 (b) data, θ_2

Figure 3.14: Simulated data. Simulation parameters correspond to case 1 in Table 3.1. Here the data counts have been normalized by I_0^{-1}.



(a) data, θ_1 (b) data, θ_2

Figure 3.15: Simulated data. Simulation parameters correspond to case 2 in Table 3.1. Here the data counts have been normalized by I_0^{-1}, and the color scale is logarithmic.

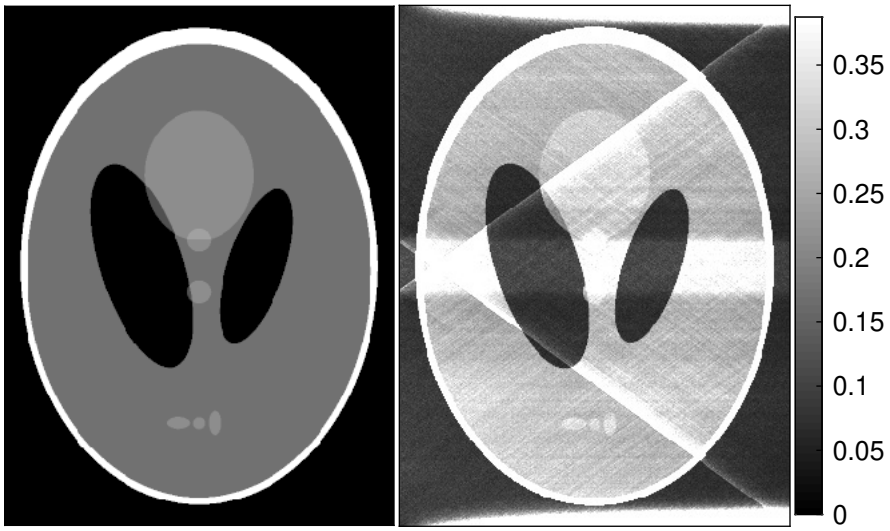


Figure 3.16: Scatter image and reconstructions using the analytic BRT inversion formula. From left to right: (a) the reference scatter image $\alpha(y)$; and (b) analytic reconstruction using case 1 data.

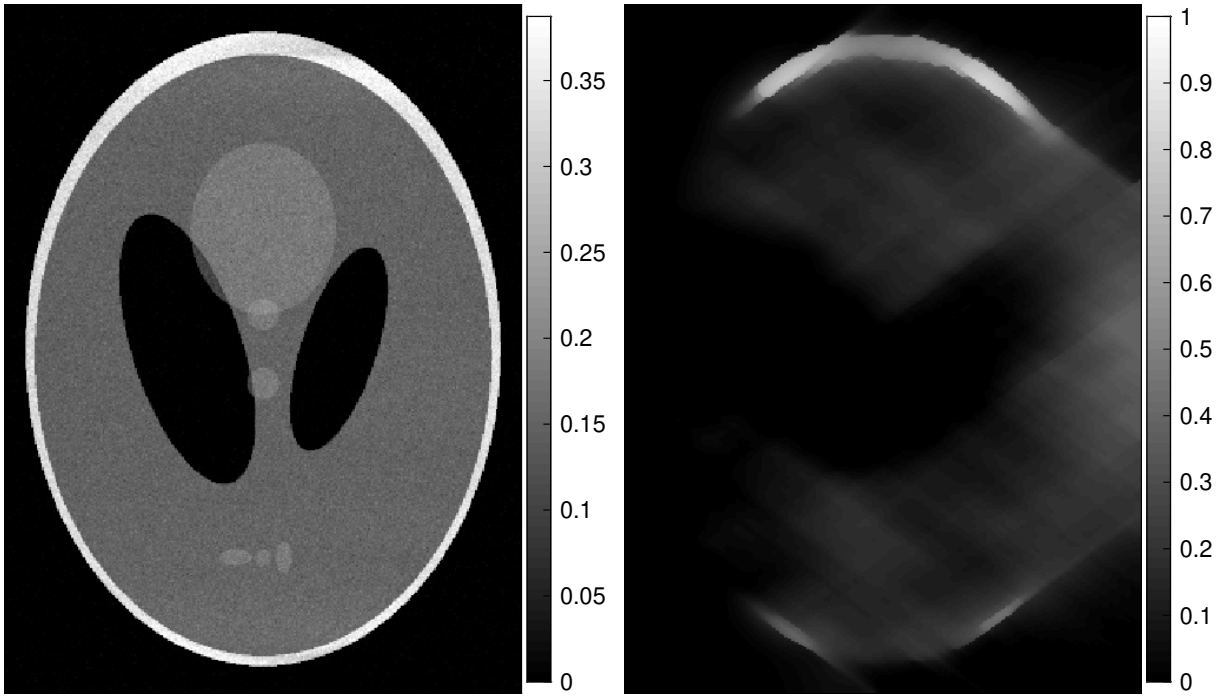
data. In contrast, Figure 3.1(d) comprises results on noise-free data. For Poisson distributed data characterized by uniform scatter, analytic inversion formulas yield a modest estimate as depicted in Figure 3.1(e).

Analytic reconstruction of the scatter image has received less attention (exceptions include [4]). Analytically, scatter images can be recovered from single-angle observations when the scatter is strictly positive and the attenuation image is finite everywhere. A more robust approach is to estimate the scatter image by averaging the corrected data based on the attenuation image, $\hat{\mu}$. Given an analytically reconstructed attenuation image estimate, we reconstruct the scatter image analytically

$$\hat{\alpha}(y) \coloneqq \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left[\frac{d_i(y)}{I_0(y)} \exp\left(\sum_{x \in \mathcal{X}} h_i(y|x) \hat{\mu}(x) \right) \right]. \tag{3.24}$$

Results are shown in Figure 3.16.

Our iterative reconstruction algorithm represents a generalized approach to the joint image estimation problem. It can also be specialized for single-image estimates. Here we consider four reconstruction problems: (1) scatter estimation with known attenuation; (2) attenuation

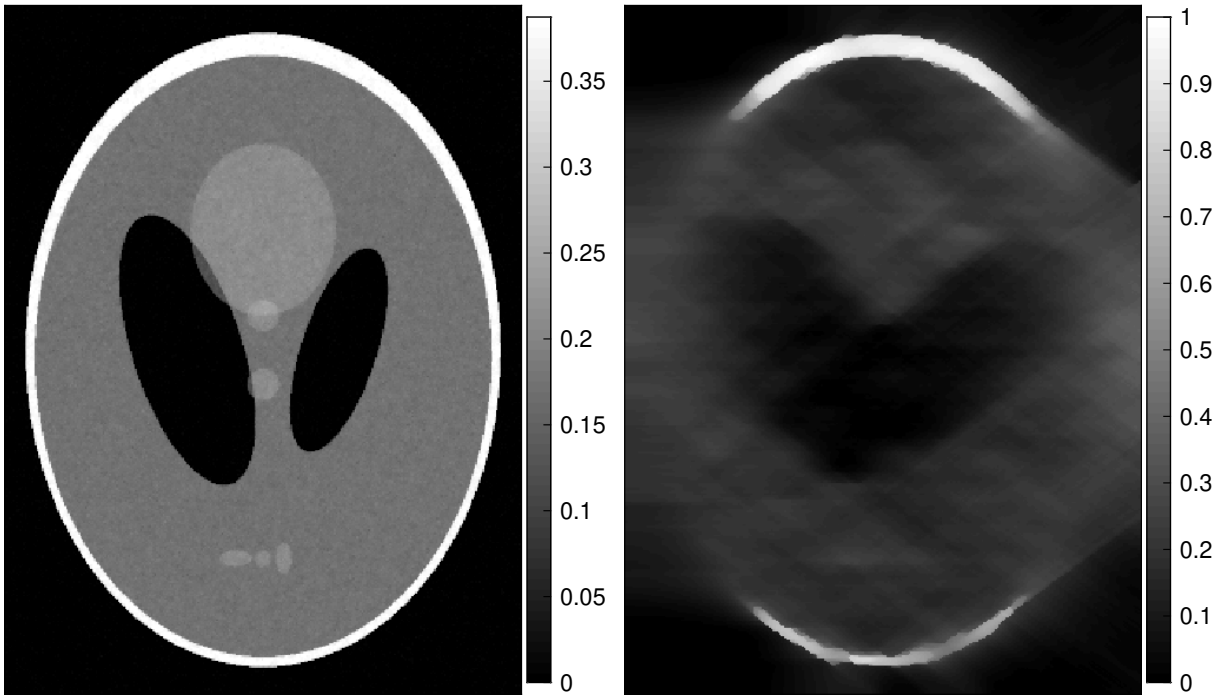


(a) scatter (b) attenuation

Figure 3.17: Joint image reconstructions from Poisson data (case 1). Scatter and attenuation reconstructions are shown in Figures 3.17a and 3.17b, respectively.

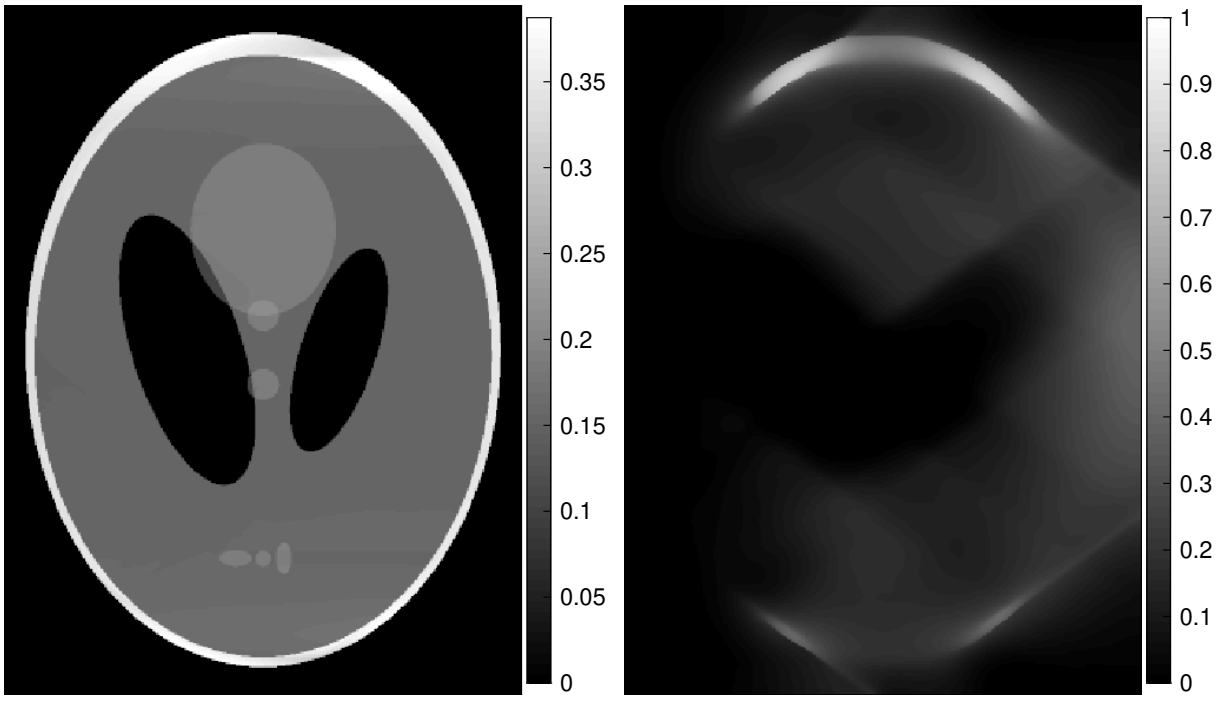
estimation with known scatter; (3) joint image estimation from Poisson data; and (4) joint image estimation from noise-free data. For case 1 data, image reconstructions are shown in Figures 3.17a and 3.17b for the single-image estimation problems (1) and (2), respectively. Image reconstructions for the joint image estimation problems (3) and (4) are shown in Figures 3.18 and 3.19, respectively.

The results shown in Figure 3.17 are not particularly impressive alone. The scatter estimate was recovered well, but this image was featured prominently in the data (c.f. Figure 3.14). The attenuation estimate was not resolved spatially, and most features of the phantom were lost. However, the joint image estimates of Figure 3.17 were obtained from the same data (case 1) yielding the analytic attenuation reconstruction of Figure 3.1(f) and the scatter reconstruction of Figure 3.16(b). Our results demonstrate significant improvements over analytic reconstructions.



(a) scatter (b) attenuation

Figure 3.18: Single-image reconstructions using known images and case 1 data. Figure 3.18a depicts reconstructed scatter given known attenuation. Conversely, Figure 3.18b depicts reconstructed attenuation given known scatter. These images were reconstructed independently from the same realization of Poisson-distributed data.



(a) scatter (b) attenuation

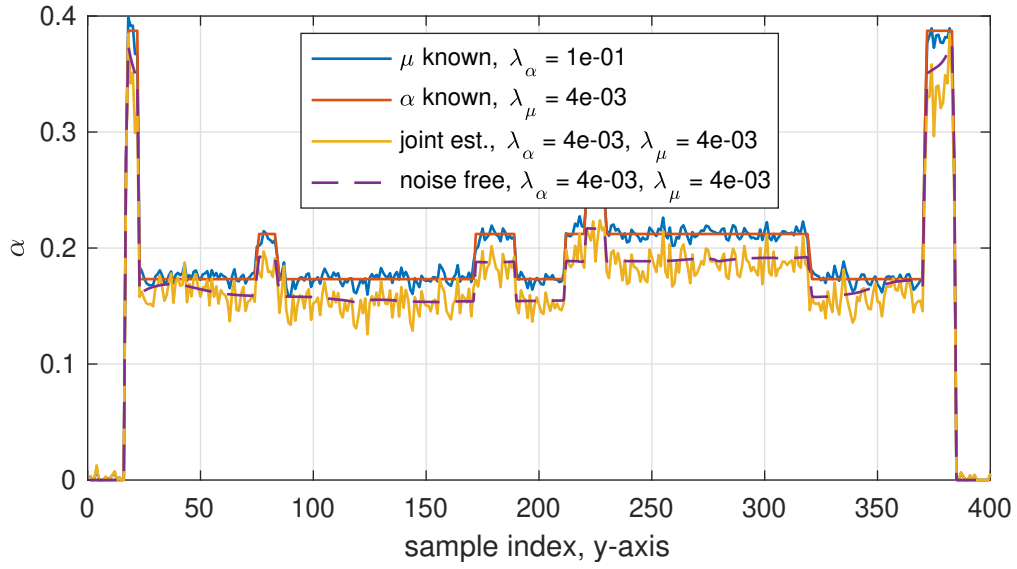
Figure 3.19: Joint image reconstructions from noise-free data (case 1). Scatter and attenuation are shown in Figures 3.19a and 3.19b, respectively.

Plotting slices through the reconstructed images provides another assessment of image quality. For each problem we obtain a slice through the reconstructed image and overlay the results on the same plot. The results will vary depending on which slice is extracted. We contrast three slices depicted in Figure 3.20, Figure 3.21, and Figure 3.22.

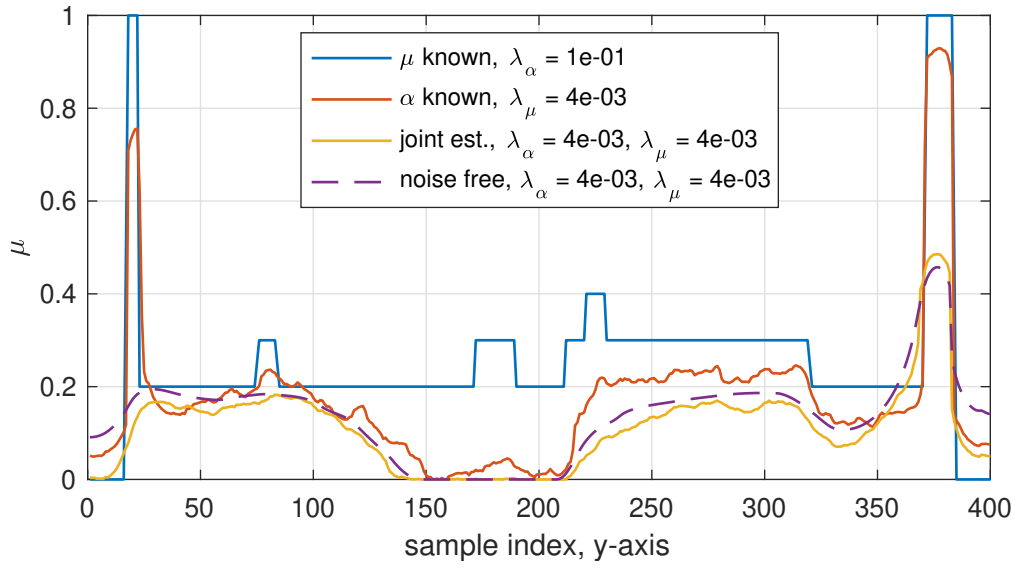
A vertical slice, bisecting the phantom is shown in Figure 3.20. In Figure 3.20a, the line "α known" represents truth. The line "μ known" overlays nicely subject to some noise. We observe a small trend in the joint estimates (3) and (4), which are negatively biased. In Figure 3.20b, the line "μ known" represents truth. Both α and μ estimates for both joint estimation problems exhibit significant deviations from truth, particularly through the midpoint. This artifact is less significant for the weak-attenuation media (case 1) and will be displayed more prominently in results for the strong-attenuation media (case 2).

Figure 3.21 and Figure 3.22 both represent horizontal slices through the images. Figure 3.21 bisects the phantom, while Figure 3.22 represents a cut through the small detail ellipses in the Shepp Logan phantom. The horizontal slice bisecting the attenuation image demonstrates poor agreement as shown in Figure 3.21b. Off the midpoint, agreement improves somewhat as in Figure 3.22b. However, agreement remains poor on the edges (e.g. sample indices 50 and 250). This blurring is consistent with the poor spatial resolution of edges orthogonal to the common direction θ_0 as shown in Figure 3.9.

As a second example, we consider data from media with stronger attenuation as shown in Figure 3.15. Results of joint image reconstruction are shown in Figure 3.23. In this case the attenuation image is better resolved, however the scatter image shows some cupping. We believe this is due to a bias in the estimated attenuation and the longer path lengths through the midpoint of the image. This cupping is particularly evident in vertical slices as shown in Figure 3.24. This is also demonstrated contrasting Figure 3.25 and 3.26 which represent medial and inferior horizontal slices, respectively.

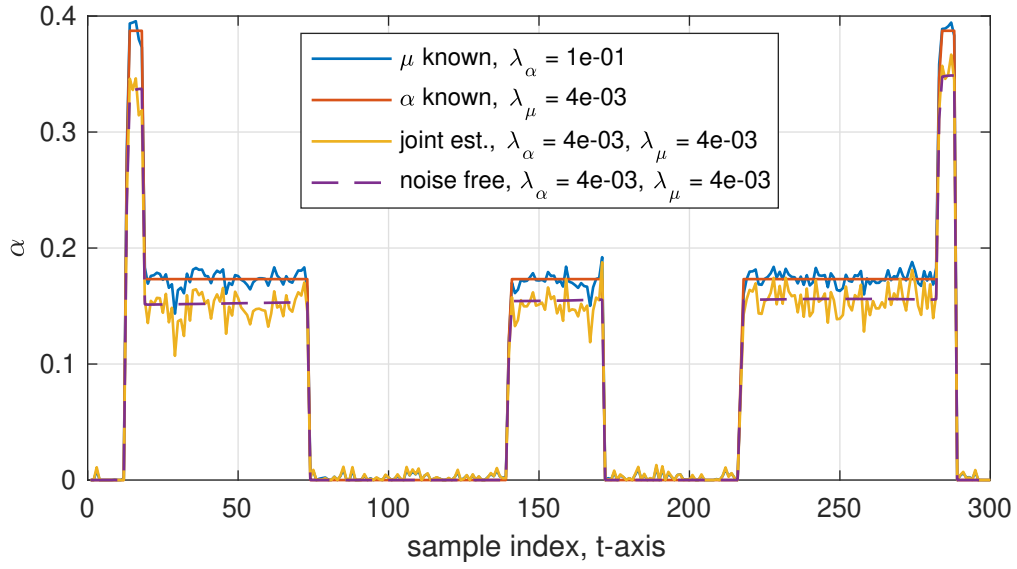


(a) scatter

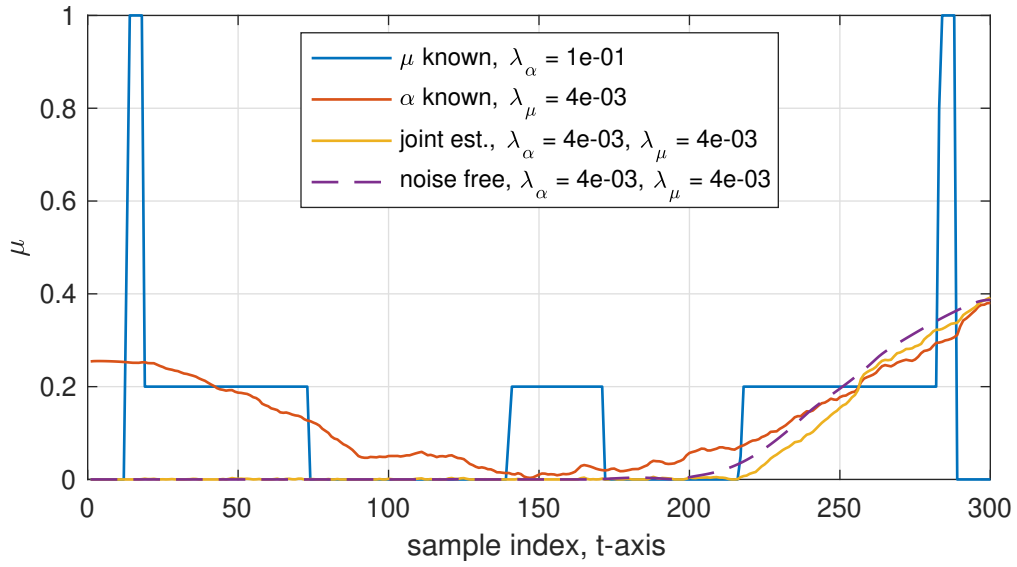


(b) attenuation

Figure 3.20: Vertical slice of reconstructed images taken at the midpoint. Image estimates from case 1 data.

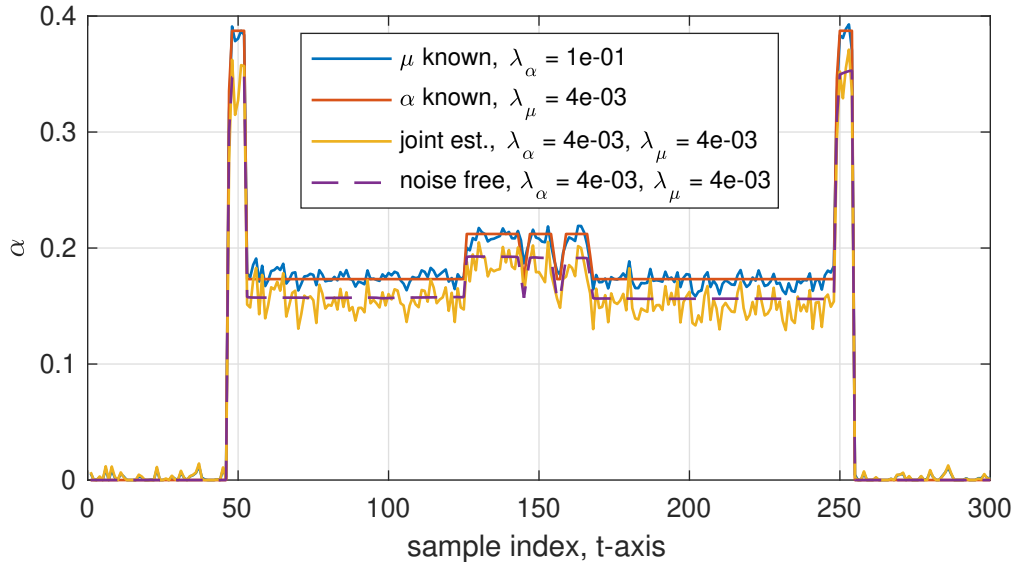


(a) scatter

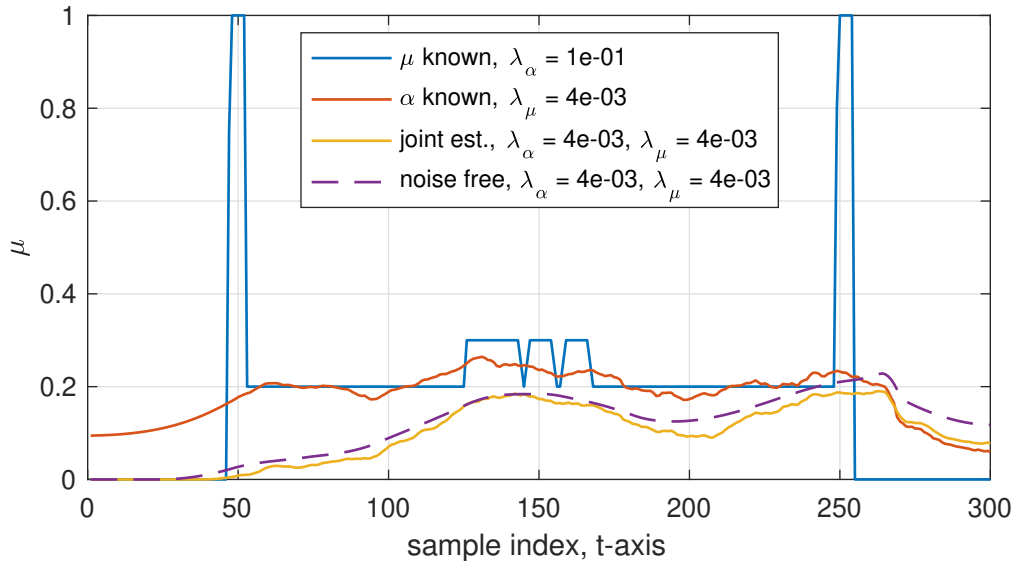


(b) attenuation

Figure 3.21: Horizontal slice of reconstructed images taken at the midpoint. Image estimates from case 1 data.

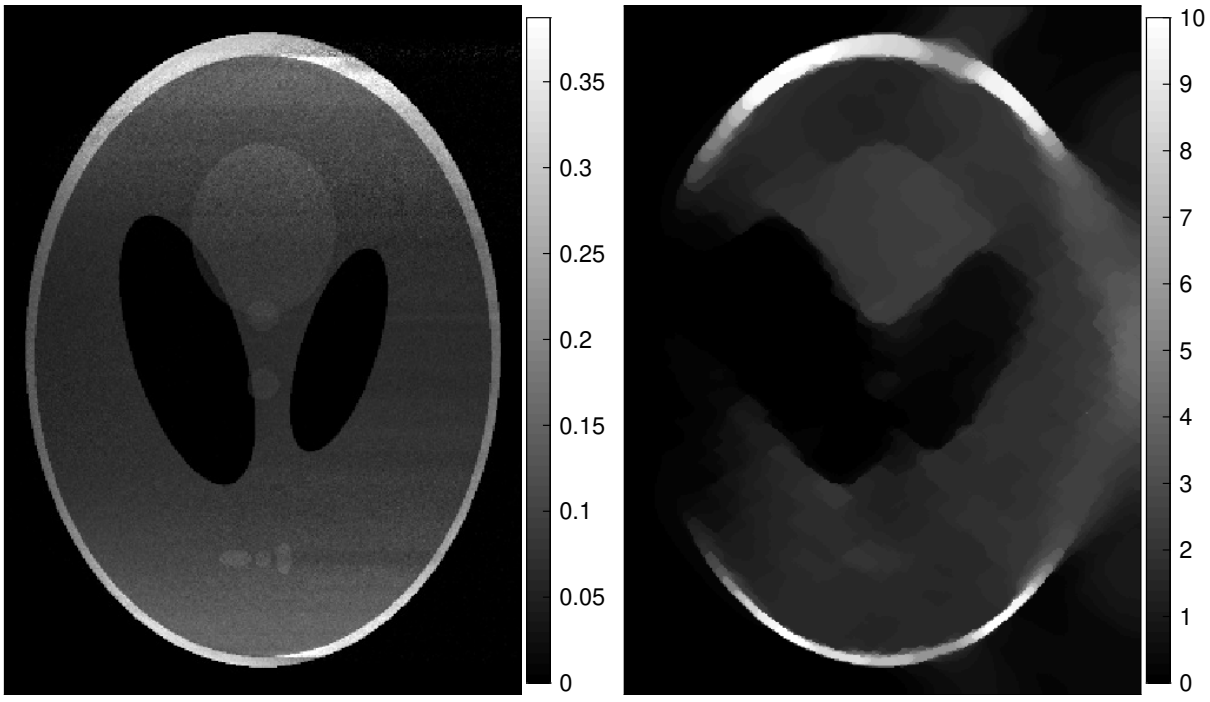


(a) scatter



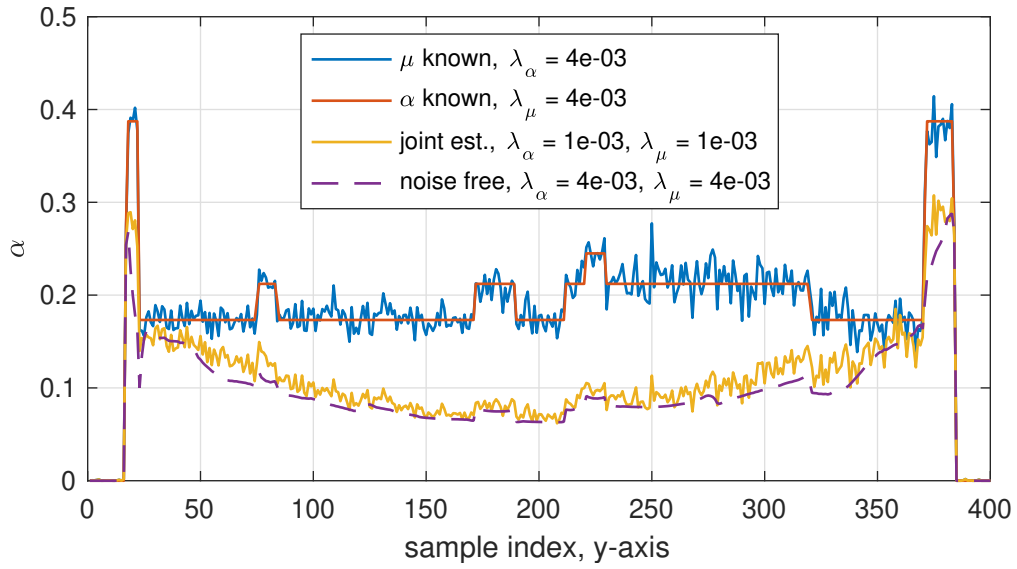
(b) attenuation

Figure 3.22: Horizontal slice of reconstructed images at $y = -0.8$ through the small detail ellipses in the Shepp Logan phantom. Image estimates from case 1 data.

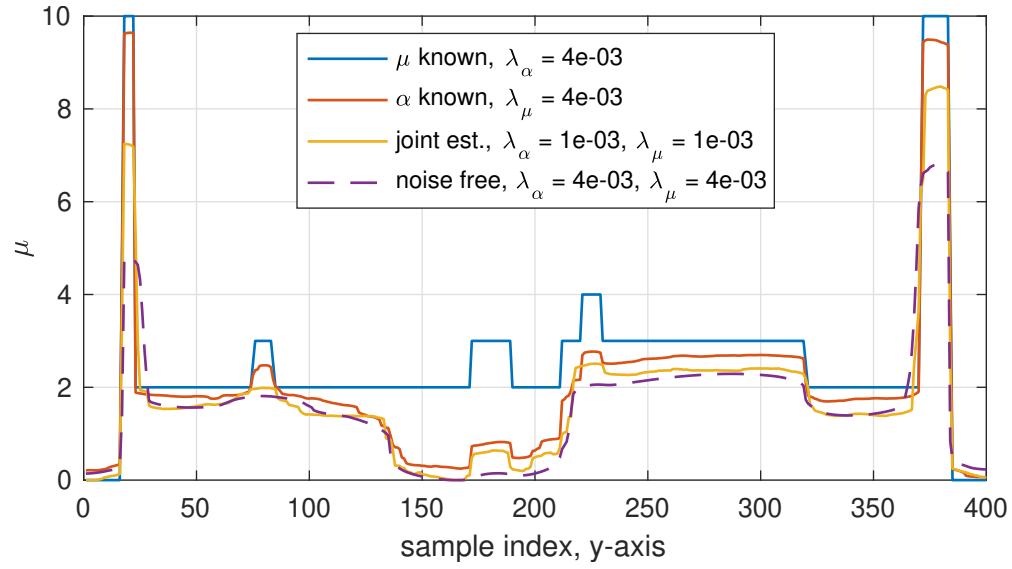


(a) scatter (b) attenuation

Figure 3.23: Joint image reconstructions from Poisson data (case 2). Scatter and attenuation are shown in Figures 3.23a and 3.23b, respectively.

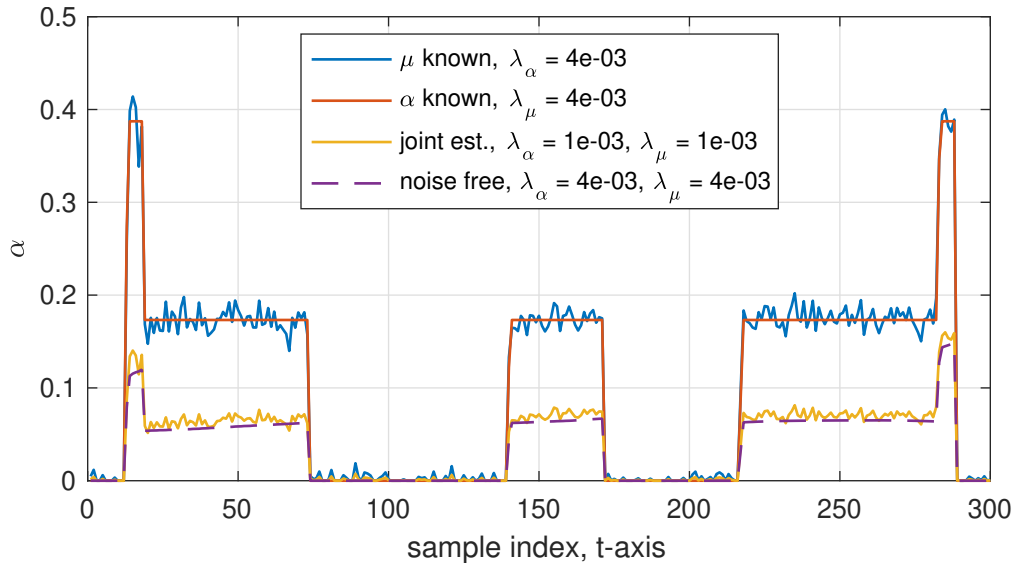


(a) scatter

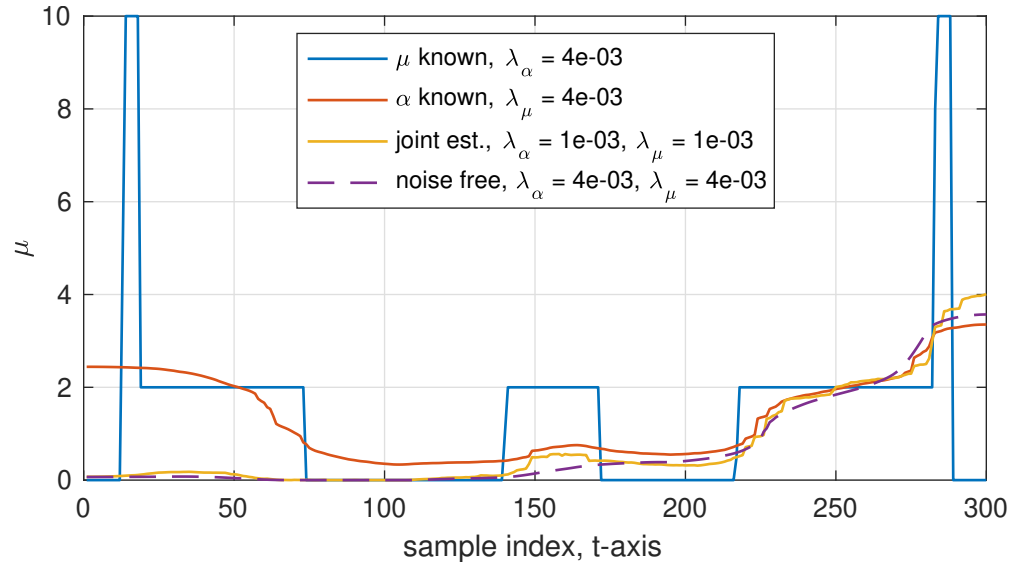


(b) attenuation

Figure 3.24: Vertical slice of reconstructed images taken at the midpoint. Image estimates from case 2 data.

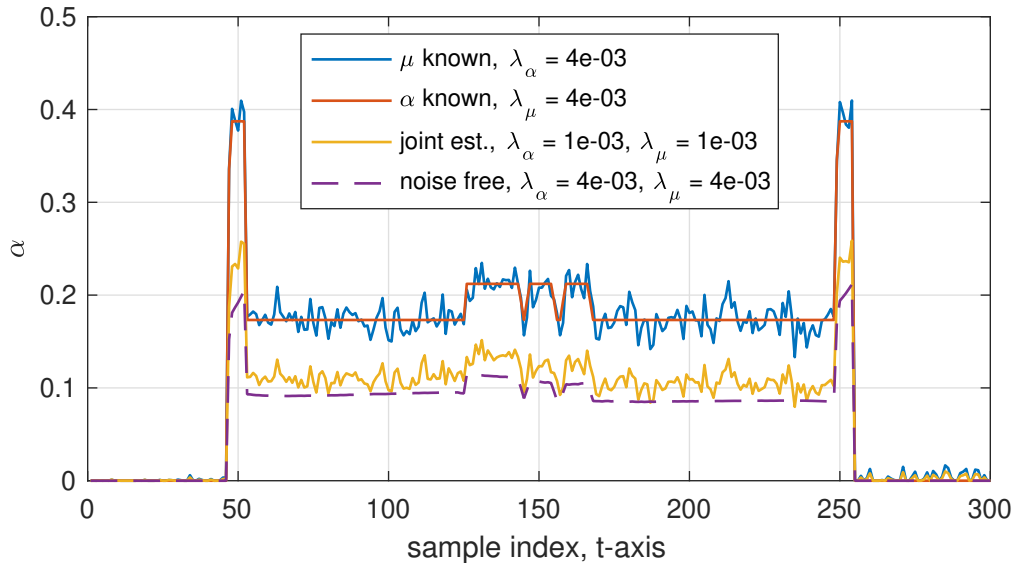


(a) scatter

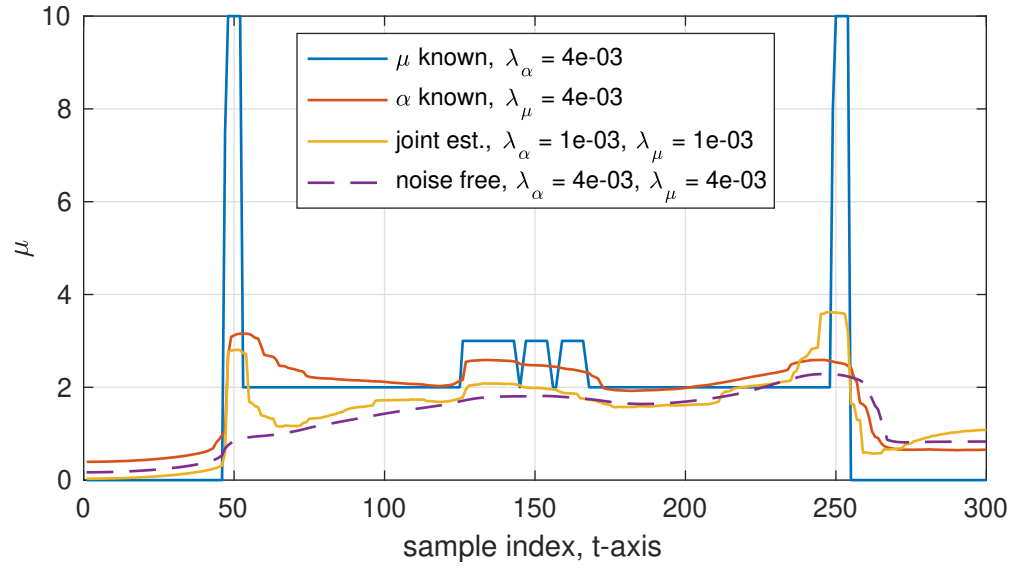


(b) attenuation

Figure 3.25: Horizontal slice of reconstructed images taken at the midpoint. Image estimates from case 2 data.



(a) scatter



(b) attenuation

Figure 3.26: Horizontal slice of reconstructed images at $y = -0.8$ through the small detail ellipses in the Shepp Logan phantom. Image estimates from case 2 data.

Chapter 4

Conclusions and Future Work

4.1 Summary and Conclusions

The broken ray transform is an important operator for a broad class of imaging problems characterized by the single scatter approximation. Analysis of the operator has suggested joint image separation is possible in this setting. Howver, prior inversion formulas perform poorly for practical cases with missing data. Our analysis of the BRT has highlighted sampling issues and identified solutions necessary for numeric applications. We then developed computationally efficient inversion formulas and operators for sampled data. Computationally efficient operators are particularly useful for iterative algorithms. Our choice of an iterative algorithm was motivated by a noisy model for the data. A major benefit of our approach is image reconstruction with missing data. Our generalized algorithm can be specialized for single image recovery. Estimates of the attenuation image improve is the scatter-density image is known. Estimates of the scatter-density image improve if the attenuation image is known. One known image is particularly useful for resolving ambiguity in the joint problem at low frequencies.

4.2 Future Directions

Anticipating use in problems with photon counting processes, we assume Poisson models. The approach can be altered for Gaussian data models which may be a convenient approximation, employing the central limit theorem, when the number of counts for each detector

is high. However, in scattering applications the number of counts per detector is characteristically small (e.g. ~ 100) for which the Poisson model is appropriate.

The choices for the simulations in this paper were made to demonstrate the potential performance improvement in disparate cases. The exact statistical model would need to be determined by the physics of that system. In particular, scaling of both the scatter and attenuation images affect the convergence rate and quality of the reconstructed images.

Our simulation results demonstrate ambiguity resolving low frequency content for joint image estimation. Often both images were under estimated, and the severity depends on the total path attenuation. Providing one image clarifies ambiguity at low frequencies. A similar ambiguity appears when attempting to recover both activity and attenuation images from TOF PET data [32]. Similarities between the BRT and TOF PET are discussed further in Appendix G.1. It is also interesting to note the puerly local inversion strategy of Katsevich [3], discussed in Appendix G.2, does not apply to TOF PET. In Appendix G.3 we demonstrate Katsevich's local inversion strategy does not apply in the TOF PET measurement geometry because the TOF PET measurement geometry invalidates Zhao's conditions [6].

Here we focus on the SVT measurement geometry which we view as a necessary first step to addressing joint image reconstruction for coded aperture measurement geometries. The coded aperture data can be described as a sparse sampling across multiple SVT measurement geometries with differing scatter directions. Coded apertures simultaneously observe multiple scatter paths with each detector. Integrating over multiple paths is important for addressing the low signal counts associated with single-path single-scatter measurements. Previous work on coded apertures trivialized the effects of the attenuation image as constant [29], or negligible [13, 14]. Our results suggests joint attenuation estimation, with no additional data, may improve estimation of momentum transfer.

Joint image recovery for coded aperture measurement geometries will require computations in a high dimensional space comprising scattering from many angles. Ordered subsets is a modern range-decomposition approach for iterative image reconstruction in high dimensions. Our simplification of the BRT forward and adjoint operator is useful in this setting as it reduces the computational burden when both constructing and applying the operator. However, our forward operator assumes global transform of a bounded image. Transforms of subsets must account for boundary conditions of the cone beam transform [24].

Bibliography

[1] L. Florescu, V. A. Markel, and J. C. Schotland, "Inversion formulas for the broken-ray radon transform," *Inverse Problems*, vol. 27, no. 2, p. 025002, Jan. 2011. [Online]. Available: https://doi.org/10.1088%2F0266-5611%2F27%2F2%2F025002

[2] L. Florescu, J. C. Schotland, and V. A. Markel, "Single-scattering optical tomography," *Physical Review E*, vol. 79, p. 036607, Mar. 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.79.036607

[3] A. Katsevich and R. Krylov, "Broken ray transform: inversion and a range condition," *Inverse Problems*, vol. 29, no. 7, p. 075008, June 2013. [Online]. Available: https://doi.org/10.1088%2F0266-5611%2F29%2F7%2F075008

[4] L. Florescu, V. A. Markel, and J. C. Schotland, "Single-scattering optical tomography: Simultaneous reconstruction of scattering and absorption," *Physical Review E*, vol. 81, p. 016602, Jan. 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.81.016602

[5] R. Gouia-Zarrad and G. Ambartsoumian, "Exact inversion of the conical Radon transform with a fixed opening angle," *Inverse Problems*, vol. 30, no. 4, p. 045007, Mar. 2014. [Online]. Available: https://doi.org/10.1088%2F0266-5611%2F30%2F4%2F045007

[6] F. Zhao, J. C. Schotland, and V. A. Markel, "Inversion of the star transform," *Inverse Problems*, vol. 30, no. 10, p. 105001, Sept. 2014. [Online]. Available: https://doi.org/10.1088%2F0266-5611%2F30%2F10%2F105001

[7] B. Sherson, "Some results in single-scattering tomography," Ph.D. dissertation, Oregon State University, Dec. 2015.

[8] G. Ambartsoumian and M. J. L. Jebelli, "The V-line transform with some generalizations and cone differentiation," *Inverse Problems*, vol. 35, no. 3, p. 034003, Feb. 2019. [Online]. Available: https://doi.org/10.1088%2F1361-6420%2Faafcf3

[9] G. Ambartsoumian, "Inversion of the V-line Radon transform in a disc and its applications in imaging," *Computers & Mathematics with Applications*, vol. 64, no. 3, pp. 260 – 265, Aug. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0898122112000806

[10] G. Ambartsoumian and S. Roy, "Numerical inversion of a broken ray transform arising in single scattering optical tomography," *IEEE Transactions on Computational Imaging*, vol. 2, no. 2, pp. 166–173, June 2016.

[11] L. Florescu, V. A. Markel, and J. C. Schotland, "Nonreciprocal broken ray transforms with applications to fluorescence imaging," *Inverse Problems*, vol. 34, no. 9, p. 094002, July 2018. [Online]. Available: https://doi.org/10.1088%2F1361-6420%2Faacec7

[12] K. MacCabe, K. Krishnamurthy, A. Chawla, D. Marks, E. Samei, and D. Brady, "Pencil beam coded aperture x-ray scatter imaging," *Optics Express*, vol. 20, no. 15, pp. 16 310–16 320, July 2012. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-20-15-16310

[13] D. J. Brady, D. L. Marks, K. P. MacCabe, and J. A. O'Sullivan, "Coded apertures for x-ray scatter imaging," *Applied Optics*, vol. 52, no. 32, pp. 7745–7754, Nov. 2013. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-52-32-7745

[14] I. Odinaka, J. A. O'Sullivan, D. G. Politte, K. P. MacCabe, Y. Kaganovsky, J. A. Greenberg, M. Lakshmanan, K. Krishnamurthy, A. J. Kapadia, L. Carin, and D. J. Brady, "Joint system and algorithm design for computationally efficient fan beam coded aperture x-ray coherent scatter imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 506–521, Dec. 2017.

[15] G. Harding and J. Kosanetzky, "Status and outlook of coherent-x-ray scatter imaging," *Journal of the Optical Society of America A*, vol. 4, no. 5, pp. 933–944, May 1987. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-4-5-933

[16] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction.* Society for Industrial and Applied Mathematics, 2001.

[17] M. Hubenthal, "The broken ray transform on the square," *Journal of Fourier Analysis and Applications*, vol. 20, no. 5, pp. 1050–1082, Oct. 2014. [Online]. Available: https://doi.org/10.1007/s00041-014-9344-3

[18] M. V. de Hoop and J. Ilmavirta, "Abel transforms with low regularity with applications to x-ray tomography on spherically symmetric manifolds," *Inverse Problems*, vol. 33, no. 12, p. 124003, Nov. 2017. [Online]. Available: https://doi.org/10.1088%2F1361-6420%2Faa9423

[19] M. Morvidone, M. K. Nguyen, T. T. Truong, and H. Zaidi, "On the V-line Radon transform and its imaging applications," *International Journal of Biomedical Imaging*, vol. 2010, p. 6, 2010, article ID 208179. [Online]. Available: https://www.hindawi.com/journals/ijbi/2010/208179/

[20] T. T. Truong and M. K. Nguyen, "On new $\mathfrak{V}$-line Radon transforms in $\mathbb{R}^2$ and their inversion," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 7, p. 075206, Jan. 2011. [Online]. Available: https://doi.org/10.1088%2F1751-8113%2F44%2F7%2F075206

[21] M. Haltmeier, S. Moon, and D. Schiefeneder, "Inversion of the attenuated v-line transform with vertices on the circle," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 853–863, Dec. 2017.

[22] F. Terzioglu, P. Kuchment, and L. Kunyansky, "Compton camera imaging and the cone transform: a brief overview," *Inverse Problems*, vol. 34, no. 5, p. 054002, Apr. 2018. [Online]. Available: https://doi.org/10.1088%2F1361-6420%2Faab0ab

[23] G. Ambartsoumian and S. Moon, "A series formula for inversion of the V-line Radon transform in a disc," *Computers & Mathematics with Applications*, vol. 66, no. 9, pp. 1567 – 1572, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0898122113000692

[24] M. R. Walker and J. A. O'Sullivan, "The broken ray transform: additional properties and new inversion formula," *Inverse Problems*, vol. 35, no. 11, p. 115003, Oct. 2019. [Online]. Available: https://doi.org/10.1088%2F1361-6420%2Fab355f

[25] P. Toft, "The Radon transform - theory and implementation," Ph.D. dissertation, Technical University of Denmark, Department of Mathematical Modelling, June 1996.

[26] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Transactions on Nuclear Science*, vol. 21, no. 3, pp. 21–43, June 1974.

[27] P. G. Lale, "The examination of internal tissues, using gamma-ray scatter with a possible extension to megavoltage radiography," *Physics in Medicine and Biology*, vol. 4, no. 2, pp. 159–167, Oct. 1959. [Online]. Available: https://doi.org/10.1088%2F0031-9155%2F4%2F2%2F305

[28] G. Redler, K. C. Jones, A. Templeton, D. Bernard, J. Turian, and J. C. H. Chu, "Compton scatter imaging: A promising modality for image guidance in lung stereotactic body radiation therapy," *Medical Physics*, vol. 45, no. 3, pp. 1233–1240, 2018. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.12755

[29] J. A. Greenberg, K. Krishnamurthy, and D. Brady, "Snapshot molecular imaging using coded energy-sensitive detection," *Optics Express*, vol. 21, no. 21, pp. 25 480–25 491, Oct. 2013. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-21-21-25480

[30] C. A. Carlsson, "Imaging modalities in x-ray computerized tomography and in selected volume tomography," *Physics in Medicine and Biology*, vol. 44, no. 3, pp. R23–R56, Jan. 1999. [Online]. Available: https://doi.org/10.1088%2F0031-9155%2F44%2F3%2F011

[31] I. Csiszár, "Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems," *Ann. Statist.*, vol. 19, no. 4, pp. 2032–2066, Dec. 1991. [Online]. Available: https://doi.org/10.1214/aos/1176348385

[32] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Physics in Medicine and Biology*, vol. 57, no. 4, pp. 885–899, Jan. 2012. [Online]. Available: https://doi.org/10.1088%2F0031-9155%2F57%2F4%2F885

[33] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, Oct. 1982.

[34] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson, and J. Nuyts, "Simultaneous reconstruction of activity and attenuation in time-of-flight PET," *IEEE Transactions on Medical Imaging*, vol. 31, no. 12, pp. 2224–2233, Dec. 2012.

[35] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 393–403, May 1999.

[36] P. Grangeat, "Analyse d'un systeme d'imagerie 3d par reconstruction a partir de radiographies x en geometrie conique," Ph.D. dissertation, École Normale Supérieure des Télécommunications, 1987.

[37] D. B. Keesing, A. Mathews, S. Komarov, H. Wu, T. Y. Song, J. A. O'Sullivan, and Y.-C. Tai, "Image reconstruction and system modeling techniques for virtual-pinhole PET insert systems," *Physics in Medicine & Biology*, vol. 57, no. 9, p. 2517, May 2012. [Online]. Available: http://stacks.iop.org/0031-9155/57/i=9/a=2517

[38] M. Defrise, M. E. Casey, C. Michel, and M. Conti, "Fourier rebinning of time-of-flight PET data," *Physics in Medicine & Biology*, vol. 50, no. 12, p. 2749, June 2005. [Online]. Available: http://stacks.iop.org/0031-9155/50/i=12/a=002

[39] D. L. Snyder, L. J. Thomas, and M. M. Ter-Pogossian, "A matheematical model for positron-emission tomography systems having time-of-flight measurements," *IEEE Transactions on Nuclear Science*, vol. 28, no. 3, pp. 3575–3583, June 1981.

[40] D. L. Snyder and D. G. Politte, "Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements," *IEEE Transactions on Nuclear Science*, vol. 30, no. 3, pp. 1843–1849, June 1983.

[41] J. Ilmavirta, "Broken ray tomography in the disc," *Inverse Problems*, vol. 29, no. 3, p. 035008, Mar. 2013. [Online]. Available: https://doi.org/10.1088%2F0266-5611%2F29%2F3%2F035008

[42] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing (4th Edition).* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[43] J. A. O'Sullivan, *Alternating Minimization Algorithms: From Blahut-Arimoto to Expectation-Maximization.* Boston, MA: Springer US, 1998, pp. 173–192. [Online]. Available: https://doi.org/10.1007/978-1-4615-5121-8_13

[44] J. A. O'Sullivan and J. Benac, "Alternating minimization algorithms for transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 283–297, Mar. 2007.

[45] A. R. De 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, Mar. 1995.

[46] P. Huber, *Robust Statistics*, ser. Wiley Series in Probability and Statistics. Wiley, 1981.

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Appendix A

Momentum Transfer and Bragg's Law

Differing conventions for momentum transfer have been used in the literature. All known conventions are consistent up to a scale factor. However, scaling discrepancies add unnecessary complication when relating fundamental concepts. In this section we derive Bragg's law from momentum transfer using traditional wavevector notation.

A scattering event can be described as a change in wavevectors. Let $\mathbf{k}$ and $\mathbf{k}'$ represent the incident and scattered wavevectors respectively. The result, $\mathbf{q} = \mathbf{k}' - \mathbf{k}$, is commonly referred to as the momentum transfer. These vectors are depicted in Fig. A.1.

Elastic, or coherent, scattering by definition preserves the incident wave energy. This forces $|\mathbf{k}| = |\mathbf{k}'|$. In such cases, we can express $q \coloneqq |\mathbf{q}|$ in terms of $|\mathbf{k}|$ and θ

$$q = 2|\mathbf{k}| \sin\left(\frac{\theta}{2}\right). \tag{A.1}$$

This quantity is generally given in units of Å^{-1}.

We assume the convention $|\mathbf{k}| = 1/\lambda$. Recall the Planck-Einstein relation for photon energy, $E = hf$, where h is Planck's constant and f is the photon frequency. Assuming wave velocity is constant in all media of interest, wavelength and frequency are related according to the

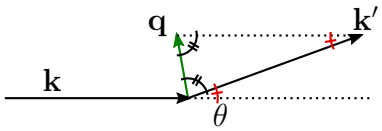


Figure A.1: Momentum transfer and related vectors for coherent scattering.

speed of light using $\lambda = c/f$. These relations lead to equivalent expressions

$$q = 2\frac{1}{\lambda}\sin\left(\frac{\theta}{2}\right) \tag{A.2}$$

$$= 2\frac{f}{c}\sin\left(\frac{\theta}{2}\right) \tag{A.3}$$

$$= 2\frac{E}{hc}\sin\left(\frac{\theta}{2}\right). \tag{A.4}$$

Several authors omit the factor of 2 in (A.1). This omission appears to have started with expression (3) in [15] which is similar to our (A.2). Similarly, (10) in [14] and (1) in [29] varies from our (A.4) by a factor of 2. However, in [29] the authors include an additional factor of 2 in the lattice spacing. A different variant was used in [12] similar to (A.3) without the factor c. Finally, physicists, by convention, include a factor of 2π in wavecectors. Therefore,

$$q_{\text{phys}} = \frac{4\pi}{\lambda}\sin\left(\frac{\theta}{2}\right) \tag{A.5}$$

$$= 2\frac{E}{\hbar c}\sin\left(\frac{\theta}{2}\right) \tag{A.6}$$

are also common.

Regardless of the variant, all of these expressions are within a scale factor of Bragg's Law

$$\frac{n}{d} = 2\frac{1}{\lambda}\sin\left(\frac{\theta}{2}\right). \tag{A.7}$$

This expression is often used in x-ray crystalography to relate crystalline lattice spacing (d) to the strongest scattering angle (θ) and incident wavelength (λ). Integer-valued n extends the model to cover harmonics. Measurements of x-ray scatter intensity will be negligible for incident angles and energies which do not satisfy (A.7) for the media of interest.

In some cases it is convenient to define the momentum transfer as a function of $a = \cos(\theta)$. This leads to

$$q(a, E) \coloneqq \frac{2E}{hc}\sqrt{\frac{1-a}{2}}. \tag{A.8}$$

Clearly, we have

$$D_1 q(a, E) = -\frac{E}{hc}\sqrt{\frac{1}{2(1-a)}} \tag{A.9}$$

$$D_2 q(a, E) = q(a, 1). \tag{A.10}$$

Appendix B

Differential Analysis

B.1 Derivatives of the Measured Data

Our interest is in distinguishing the terms in the log data using multiple measurements. In this section we generalize the notion of multiple measurements using differential equations. For convenience, we restate (1.7a)

$$l(x, \alpha, \beta, E) = \ln f\left(x, q\left(\alpha \cdot \beta, E\right)\right) - (B\mu)(x, -\alpha) - (B\mu)(x, \beta). \tag{B.1}$$

Here we present notation for the first derivative of (1.7) with respect to each input parameter. These differential equations provide a framework for contrasting prior work.

The inputs α, β to (B.1) are drawn from S^2. The notion of a derivative is not directly available over S^2 as it is not a vector space. However, the set S^2 is a two-dimensional manifold embedded in three-dimensional space. Since a manifold locally approximates Euclidean space, we may locally approximate S^2 as the tangent plane in $\mathbb{R}^3$. This approximation supports directional derivatives for any direction in the tangent plane. This concept is discussed further in Appendix B.3.

For some differentiable function $f : S^2 \to \mathbb{R}^3$, we use $\langle Df(\theta), u\rangle$ for the directional derivative of f in the direction u. We emphasize $\langle Df(\theta), u\rangle$ is scalar-valued. For an orthogonal basis set θ, u, v we have

$$Df(\theta) = \langle Df(\theta), u\rangle u + \langle Df(\theta), v\rangle v. \tag{B.2}$$

Here u, v are a basis for the plane orthogonal to θ. While selection of u, v are not unique, $Df(\theta) \in \mathbb{R}^3$ is. By definition $\langle Df(\theta), \theta \rangle = 0$.

Differentiating (B.1) with respect to each of its four arguments we obtain

$$D_1 l(x,\alpha,\beta,E) = \frac{D_1 f\left(x, q\left(\alpha\cdot\beta, E\right)\right)}{f\left(x, q\left(\alpha\cdot\beta, E\right)\right)} - D_1(B\mu)(x,-\alpha) - D_1(B\mu)(x,\beta) \tag{B.3}$$

$$D_2 l(x,\alpha,\beta,E) = \frac{D_2 f\left(x, q\left(\alpha\cdot\beta, E\right)\right)}{f\left(x, q\left(\alpha\cdot\beta, E\right)\right)} D_1 q(\alpha\cdot\beta, E)\beta - D_2(B\mu)(x,-\alpha) \tag{B.4}$$

$$D_3 l(x,\alpha,\beta,E) = \frac{D_2 f\left(x, q\left(\alpha\cdot\beta, E\right)\right)}{f\left(x, q\left(\alpha\cdot\beta, E\right)\right)} D_1 q(\alpha\cdot\beta, E)\alpha - D_2(B\mu)(x,\beta) \tag{B.5}$$

$$D_4 l(x,\alpha,\beta,E) = \frac{D_2 f\left(x, q\left(\alpha\cdot\beta, E\right)\right)}{f\left(x, q\left(\alpha\cdot\beta, E\right)\right)} D_2 q(\alpha\cdot\beta, E). \tag{B.6}$$

Here we use Euler's notation for the differential operator where the subscript identifies the argument acted upon. Since $f : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$, $D_1 f \in \mathbb{R}^3$, and $D_2 f \in \mathbb{R}$. Differentiation over S^2 results in an element of $\mathbb{R}^3$ as discussed previously.

Expanding (B.4),(B.5) using (B.2) requires an orthogonal basis in $\mathbb{R}^3$. We can expand (B.4),(B.5) using (B.2) as

$$D_2 l(x,\alpha,\beta,E) = \langle D_2 l(x,\alpha,\beta,E), \psi\rangle\psi + \langle D_2 l(x,\alpha,\beta,E), \phi\rangle\phi \tag{B.7}$$

$$D_3 l(x,\alpha,\beta,E) = \langle D_3 l(x,\alpha,\beta,E), \theta\rangle\theta + \langle D_3 l(x,\alpha,\beta,E), \phi\rangle\phi. \tag{B.8}$$

The four directional derivatives are

$$\langle D_2 l(x,\alpha,\beta,E), \phi\rangle = -\langle D_2(B\mu)(x,-\alpha), \phi\rangle \tag{B.9}$$

$$\langle D_2 l(x,\alpha,\beta,E), \psi\rangle = \frac{D_2 q(x,\alpha\cdot\beta,E)}{q(x,\alpha\cdot\beta,E)}\langle\beta,\psi\rangle - \langle D_2(B\mu)(x,-\alpha), \psi\rangle \tag{B.10}$$

$$\langle D_3 l(x,\alpha,\beta,E), \phi\rangle = -\langle D_2(B\mu)(x,\beta), \phi\rangle \tag{B.11}$$

$$\langle D_3 l(x,\alpha,\beta,E), \theta\rangle = \frac{D_2 q(x,\alpha\cdot\beta,E)}{q(x,\alpha\cdot\beta,E)}\langle\alpha,\theta\rangle - \langle D_2(B\mu)(x,\beta), \theta\rangle. \tag{B.12}$$

These expressions are significant in that (B.9) and (B.11) are not influenced by $q(x,\alpha\cdot\beta,E) > 0$. This suggests μ could be recovered, up to an additive constant, given sufficiently sampled data. One could obtain (B.9) and (B.10) by perturbing the transmitter location about the

focal point x. Similarly, perturbing the receiver location could reveal (B.11) and (B.12). However, such measurements are insufficient to distinguish the two terms in (B.10) and (B.12).

B.2 Angular Differential Measurements

Using subscripts to reference the initial angle β_1, we define a second scatter direction

$$\beta_2 = \alpha\,(\alpha \cdot \beta_1) + \psi_1\,(\theta_1' \cdot \beta_1)\,. \tag{B.13}$$

This represents a 90-degree rotation of β about α with the following relations

$$\beta_1 \cdot \alpha = \beta_2 \cdot \alpha \tag{B.14}$$

$$\theta_2' = \phi_1 \tag{B.15}$$

$$\phi_2 = -\theta_1' \tag{B.16}$$

$$\langle D_2(l)(x, \alpha, \beta_2, E), \theta_1' \rangle = -\langle D_2(B\mu)(x, -\alpha), \theta_1' \rangle \tag{B.17}$$

$$\frac{D_2 q(x, \alpha \cdot \beta, E)}{q(x, \alpha \cdot \beta, E)} = \frac{1}{\beta_1 \cdot \theta_1'}\left(\langle D_2(l)(x, \alpha, \beta_1, E), \theta_1' \rangle - \langle D_2(l)(x, \alpha, \beta_2, E), \theta_1' \rangle\right) \tag{B.18}$$

$$D_2(B\mu)(x, \alpha) = -\phi_1 \langle D_2(l)(x, \alpha, \beta_1, E), \phi_1 \rangle - \theta_1' \langle D_2(l)(x, \alpha, \beta_2, E), \theta_1' \rangle. \tag{B.19}$$

A similar approach can be employed to obtain $D_2(B\mu)(x, \beta)$. Obiously 90-degree rotations are not required. These quantities could be found for arbitrary rotations subject to the conditioning of the associated linear system. While this expansion is concise, obtaining these measurements is not convenient using current acquisition systems.

B.3 Directional Derivative on S^2

Practical examples of directional derivatives may avoid some confusion. The directional derivative on S^2 implies a rotational derivative. However, the derivative is only relevant in certain directions. In this section we expand the vectors in Cartesian coordinates. Additionally, we will re-appropriate θ and ϕ to reference spherical coordinates.

Any element of S^2 can be described with two spherical coordinates, θ and ϕ, with the implication $r = 1$. For example

$$d(\theta, \phi) = \begin{bmatrix} \cos\theta \sin\phi \\ \sin\theta \sin\phi \\ \cos\phi \end{bmatrix}. \tag{B.20}$$

Clearly

$$D_2 d(\theta, \phi) = \begin{bmatrix} \cos\theta \cos\phi \\ \sin\theta \cos\phi \\ -\sin\phi \end{bmatrix}. \tag{B.21}$$

For the special case $\phi_1 = \phi_2 = \frac{\pi}{2}$

$$\frac{\partial}{\partial\phi} \left(d(\theta_1, \phi_1) \cdot d(\theta_2, \phi_2) \right) \Big|_{\phi_1, \phi_2 = \frac{\pi}{2}} = 0. \tag{B.22}$$

This special case can be useful. For any two unique vectors, $d_1 \neq d_2$, $d_1, d_2 \in \mathbb{R}^3$ there exists one unique plane containing these two vectors. Without loss of generality we may assume a rotation such that d_1 and d_2 lie in the x-y plane. Therefore, (B.22) applies whenever the derivative is taken in the direction orthogonal to both vectors. This explains the absence of the momentum transfer term in (B.9) and (B.11).

We may extend the definition of the cone beam transform to include θ, ϕ as

$$(Bf)(s, \theta, \phi) = \int_0^\infty f \left(s + t \begin{bmatrix} \cos\theta \sin\phi \\ \sin\theta \sin\phi \\ \cos\phi \end{bmatrix} \right) dt. \tag{B.23}$$

In this expression we use $s \in \mathbb{R}^3$ to avoid confusion with the coordinate x. Again, a special case simplifies the result

$$-\frac{\partial}{\partial\phi}(Df)(s,\theta,\phi)\bigg|_{\phi=\frac{\pi}{2}} = -\int_0^\infty \begin{bmatrix}\cos\theta\cos\phi\\ \sin\theta\cos\phi\\ -\sin\phi\end{bmatrix} \cdot \nabla f\left(s + t\begin{bmatrix}\cos\theta\sin\phi\\ \sin\theta\sin\phi\\ \cos\phi\end{bmatrix}\right)\Bigg|_{\phi=\frac{\pi}{2}} dt \tag{B.24}$$

$$= \int_0^\infty \begin{bmatrix}0\\0\\1\end{bmatrix} \cdot \nabla f\left(s + t\begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix}\right) dt \tag{B.25}$$

$$= \int_0^\infty \frac{\partial}{\partial z} f\left(s + t\begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix}\right) dt. \tag{B.26}$$

This may provide a more intuitive expansion of Grangeat's formula. In this case, the negative sign accounts for the difference in the orientation of ϕ. This follows one of two proofs given by Natterer [16].

Appendix C

Fourier Analysis

C.1 Derivation of the Fourier Transform of the CBT

We define the Fourier transform as a function of frequency to avoid scaling the inverse. For a one-dimensional function we define the one-dimensional Fourier transform and its inverse

$$\mathcal{F}^1\{f(x)\} := \int_{-\infty}^{+\infty} f(x)e^{-i2\pi wx}dx \tag{C.1}$$

$$\mathcal{F}^{-1}\{\hat{f}(w)\} := \int_{-\infty}^{+\infty} \hat{f}(w)e^{i2\pi wx}dw. \tag{C.2}$$

For two-dimensional functions we define the two-dimensional Fourier transform and its inverse

$$\mathcal{F}^2\{f(x)\} := \int_{\mathbb{R}^2} f(x)e^{-i2\pi w\cdot x}d^2x \tag{C.3}$$

$$\mathcal{F}^{-2}\{\hat{f}(w)\} := \int_{\mathbb{R}^2} \hat{f}(w)e^{i2\pi w\cdot x}d^2w. \tag{C.4}$$

In this form, we have

$$\mathcal{F}^1\{u(x)\} = \frac{1}{i2\pi w} + \frac{1}{2}\delta(w), \tag{C.5}$$

where $\delta(x)$ and $u(x)$ represent the Dirac delta function and the unit step function, respectively.

For CBT data associated with a fixed direction, θ, we define the two-dimensional Fourier transform

$$\begin{aligned}
\hat{b}_\theta(w) &= \mathcal{F}^2\{(B\mu_C)(x,\theta)\} && \text{(C.6)}\\
&= \int_{\mathbb{R}^2}\int_0^\infty \mu_C(x+t\theta)dt\, e^{-i2\pi w\cdot x}d^2x && \text{(C.7)}\\
&= \int_0^\infty \int_{\mathbb{R}^2} \mu_C(y)e^{-i2\pi w\cdot y}d^2y\, e^{i2\pi tw\cdot\theta}dt && \text{(C.8)}\\
&= \hat{\mu}_C(w)\int_0^\infty e^{i2\pi tw\cdot\theta}dt && \text{(C.9)}\\
&= \hat{\mu}_C(w)\left[\frac{-1}{i2\pi w\cdot\theta}+\frac{1}{2}\delta(w\cdot\theta)\right]. && \text{(C.10)}
\end{aligned}$$

In (C.8) we changed the order of integration and substituted $y = x + t\theta$. In (C.9) we substituted $\hat{\mu}_C(w) = \mathcal{F}^2\{\mu_C(x)\}$. Finally, in (C.10) we made use of (C.5).

C.2 BRT Inversion by Fourier Analysis

To invert the BRT, we start by multiplying both sides of (2.3c) by the reciprocal of (2.4). Rearranging terms, we have

$$\begin{aligned}
\hat{\mu}_C(w) = \;& i\pi\hat{\mu}_C(w)\tfrac{(w\cdot\theta_i)(w\cdot\theta_j)}{w\cdot(\theta_i+\theta_j)}\delta\left(w\cdot\theta_i\right) + i\pi\hat{\mu}_C(w)\tfrac{(w\cdot\theta_i)(w\cdot\theta_j)}{w\cdot(\theta_i+\theta_j)}\delta\left(w\cdot\theta_j\right)\\
&+\hat{g}_{i,j}(w)\tfrac{-i2\pi(w\cdot\theta_i)(w\cdot\theta_j)}{w\cdot(\theta_i+\theta_j)}, \quad \forall\, w\notin\Theta_{i,j}. \qquad \text{(C.11)}
\end{aligned}$$

The first two terms on the right-hand side vanish under integration. We note the inverse two-dimensional Fourier transform

$$\int_{\mathbb{R}^2}(w\cdot\theta)\,\delta\left(w\cdot\theta\right)e^{i2\pi w\cdot x}d^2w = 0, \quad \forall\, x\in\mathbb{R}^2. \qquad \text{(C.12)}$$

Incorporating multiplicative functions does not change this result as long as they are finite for all $w\cdot\theta = 0$. For the first term in (C.11), we expand $w = s\theta_i + t\theta_i^\perp$ as an orthonormal

basis and set $s = 0$. This leads to

$$\hat{\mu}_C(w) \left. \frac{w \cdot \theta_j}{w \cdot (\theta_i + \theta_j)} \right|_{w = t\theta_i^{\perp}} = \hat{\mu}_C(t\theta_i^{\perp}), \quad \forall\, \theta_i \neq \theta_j, \tag{C.13}$$

which is finite for all t by our assumptions on $\mu_C(x)$. Applying a similar process for the second term, we find

$$\mathcal{F}^{-2} \left\{ i\pi \hat{\mu}_C(w) \frac{(w \cdot \theta_i)\,(w \cdot \theta_j)}{w \cdot (\theta_i + \theta_j)} \left[\delta\,(w \cdot \theta_i) + \delta\,(w \cdot \theta_j) \right] \right\} = 0. \tag{C.14}$$

To address the third term in (C.11) we make use of the derivative and integral properties of the Fourier transform. Now we expand $x = s\theta + t\theta^{\perp}$ and consider the directional derivative

$$\left. \frac{d}{d\theta} f(x) \right|_{x = s\theta + t\theta^{\perp}} = \frac{d}{ds} f(s\theta + t\theta^{\perp}). \tag{C.15}$$

Applying this to the inverse two-dimensional Fourier transform, we find

$$\begin{aligned} \frac{d}{d\theta} f(x) &= \frac{d}{d\theta} \mathcal{F}^{-2} \left\{ \hat{f}(w) \right\} && \text{(C.16)} \\ &= \int_{\mathbb{R}^2} \frac{d}{ds} \hat{f}(w) e^{i2\pi \left(sw \cdot \theta + tw \cdot \theta^{\perp} \right)} d^2 w && \text{(C.17)} \\ &= \mathcal{F}^{-2} \left\{ i2\pi\,(w \cdot \theta)\, \hat{f}(w) \right\}. && \text{(C.18)} \end{aligned}$$

We previously derived the integration property of the two-dimensional Fourier transform in the context of the CBT. From (C.10), we have

$$\mathcal{F}^2 \left\{ \int_0^{\infty} f\,(x + s\theta)\, ds \right\} = \hat{f}(w) \left[\frac{-1}{i2\pi w \cdot \theta} + \frac{1}{2} \delta(w \cdot \theta) \right]. \tag{C.19}$$

When $\hat{f}(w \cdot \theta) = 0$ for all $w \in \mathbb{R}^2$, substituting $-\theta$ in (C.19) we also find

$$\begin{aligned} \int_0^{\infty} f\,(x + s\theta)\, ds &= -\int_0^{\infty} f\,(x + s(-\theta))\, ds && \text{(C.20)} \\ &= -\int_{-\infty}^{0} f\,(x + s\theta)\, ds. && \text{(C.21)} \end{aligned}$$

From (2.3c), $\hat{g}_{i,j}(0)$ is not guaranteed to be finite, much less zero. However, from (2.3c) we have

$$(w \cdot \theta_i)\,(w \cdot \theta_j)\,\hat{g}_{i,j}(w) = 0, \quad \forall\, w \cdot (\theta_i + \theta_j) = 0. \tag{C.22}$$

Putting this all together, we have equivalent expressions

$$\mu_C(x) = \frac{1}{\|\theta_i + \theta_j\|} \int_0^\infty \frac{d}{d\theta_i}\frac{d}{d\theta_j} g_{i,j}\left(x + s\frac{\theta_i + \theta_j}{\|\theta_i + \theta_j\|}\right) ds \tag{C.23}$$

$$= \frac{-1}{\|\theta_i + \theta_j\|} \int_{-\infty}^0 \frac{d}{d\theta_i}\frac{d}{d\theta_j} g_{i,j}\left(x + s\frac{\theta_i + \theta_j}{\|\theta_i + \theta_j\|}\right) ds. \tag{C.24}$$

We emphasize equality only holds for images, $\mu_C(x)$, with bounded support. This assumption is necessary due to the nullspace of the forward operator.

C.3 Two-Dimensional Fourier Transform of a Parallelogram

Parallelograms are often expressed in terms of the edge directions and edge lengths. We consider the directions θ_i, θ_j and associated edge lengths a_i, a_j, respectively. The total area of the resulting parallelogram is $a_i a_j |\det(\theta_i, \theta_j)|$. As an alternative to edge lengths, we also consider the orthogonal distance between parallel sides. We define b_i and b_j as the extent (height) of the parallelogram in the orthogonal directions $\theta_i^\perp$ and $\theta_j^\perp$, respectively. These distances are related to the edge lengths through the change of variables

$$b_i \coloneqq a_j \left|\det(\theta_i, \theta_j)\right| \tag{C.25}$$

$$b_j \coloneqq a_i \left|\det(\theta_i, \theta_j)\right|. \tag{C.26}$$

Additionally, we define the one-dimensional rectangular function

$$\Pi_T(t) \coloneqq \begin{cases} 1, & |t| \le T/2 \\ 0, & \text{otherwise.} \end{cases} \tag{C.27}$$

We define the two-dimensional parallelogram indicator function, centered at $x = 0$,

$$p_{i,j}(x; a_i, a_j) \coloneqq \Pi_{b_i}\left(x \cdot \theta_i^{\perp}\right) \Pi_{b_j}\left(x \cdot \theta_j^{\perp}\right). \tag{C.28}$$

Here b_i, b_j are determined by a_j, a_i using (C.25) and (C.26), respectively. The area of this function is equivalently

$$\int_{R^2} p_{i,j}(x; a_i, a_j) d^2x = a_i a_j |\det(\theta_i, \theta_j)| = \frac{b_i b_j}{|\det(\theta_i, \theta_j)|}. \tag{C.29}$$

To determine the two-dimensional Fourier transform of (C.28), we exploit the convolution property of the Fourier transform. We transform the two rectangular functions separately, then convolve the results in the frequency domain. The one-dimensional Fourier transform of (C.27) is

$$\mathcal{F}\{\Pi_T(t)\} = T \operatorname{sinc}(wT). \tag{C.30}$$

Extending this to two dimensions, we have the relation

$$\mathcal{F}^2\left\{\Pi_T(x \cdot \theta)\delta\left(x \cdot \theta^{\perp}\right)\right\} = T \operatorname{sinc}(Tw \cdot \theta) \tag{C.31}$$

$$\mathcal{F}^2\{\Pi_T(x \cdot \theta)\} = T \operatorname{sinc}(Tw \cdot \theta)\delta\left(w \cdot \theta^{\perp}\right). \tag{C.32}$$

We derive the two-dimensional Fourier transform of (C.28) as

$$\hat{p}_{i,j}(w; a_i, a_j) = b_i b_j \operatorname{sinc}(b_i w \cdot \theta_i^{\perp})\delta(w \cdot \theta_i) * \operatorname{sinc}(b_j w \cdot \theta_j^{\perp})\delta(w \cdot \theta_j) \tag{C.33}$$

$$= b_i b_j \int_{\mathbb{R}^2} \operatorname{sinc}(b_i y \cdot \theta_i^{\perp})\delta(y \cdot \theta_i) \operatorname{sinc}(b_j (w - y) \cdot \theta_j^{\perp}) \times \delta((w - y) \cdot \theta_j) d^2y \tag{C.34}$$

$$= b_i b_j \int_{\mathbb{R}} \operatorname{sinc}(b_i t) \operatorname{sinc}(b_j \left(w - t\theta_i^{\perp}\right) \cdot \theta_j^{\perp})\delta\left(\left(w - t\theta_i^{\perp}\right) \cdot \theta_j\right) dt \tag{C.35}$$

$$= \frac{b_i b_j}{|\det(\theta_i, \theta_j)|} \operatorname{sinc}\left(b_i \frac{w \cdot \theta_j}{\theta_i^{\perp} \cdot \theta_j}\right) \operatorname{sinc}\left(b_j w \cdot \left(\theta_j^{\perp} - \theta_j \frac{\theta_i \cdot \theta_j}{\theta_i^{\perp} \cdot \theta_j}\right)\right) \tag{C.36}$$

$$= \frac{b_i b_j}{|\det(\theta_i, \theta_j)|} \operatorname{sinc}\left(b_i \frac{w \cdot \theta_j}{\theta_i^{\perp} \cdot \theta_j}\right) \operatorname{sinc}\left(b_j \frac{w \cdot \theta_i}{\theta_i \cdot \theta_j^{\perp}}\right) \tag{C.37}$$

$$= a_i a_j |\det(\theta_i, \theta_j)| \operatorname{sinc}(a_j w \cdot \theta_j) \operatorname{sinc}(a_i w \cdot \theta_i). \tag{C.38}$$

For (C.35), we expand the integration variable in (C.34) using the orthonormal basis, $y = s\theta_i + t\theta_i^{\perp}$, and integrate over s. Changing the variable of integration again effects a change in scaling in (C.36). The second sinc function of (C.36) comprises expansion of θ_i using the orthonormal basis θ_j, $\theta_j^{\perp}$. Restoring θ_i, we obtain (C.37). Restoring a_j, a_i using (C.25), (C.26), we obtain (C.38).

Appendix D

Geometric Problems

The BRT, as a forward operator, can be decomposed as a series of geometric problems. Solving these problems analytically improves precision and computational efficiency. Our work here is useful both for simulation, and for inversion algorithms requiring a forward model.

D.1 Ray Intersection with Ellipsoid

Ellipsoids are a convenient geometric shape for specifying bounded regions within a larger volume. These shapes can be used to define nontrivial phantoms analytically such as the Shepp-Logan phantom [26]. BRT data from such phantoms can also be determined analytically. For this we must determine whether the ray intersects the ellipsoid, and if so, where along its path.

We choose to represent the ellipsoid in point-matrix form. Let $a \in \mathbb{R}^n$ represent the centroid of the ellipsoid, and symmetric positive-definite matrix $A \in \mathbb{R}^{n \times n}$ determine the scaling and principle axes. Using

$$h_e(x) = (x-a)^T A (x-a) - 1 \tag{D.1}$$

we define the surface of the ellipsoid as the set $S \coloneqq \{x \in \mathbb{R}^n : h_e(x) = 0\}$.

We consider the ray, originating from $x_0 \in \mathbb{R}^n$, in the direction $\theta \in S^2$. Intersections with the ellipsoid are determined by roots of the polynomial

$$h_e(x + s\theta) = s^2\theta^T A\theta + s2x^T A\theta + x^T Ax - 1. \tag{D.2}$$

This is a real quadratic polynomial with scalar coefficients. As such, the roots must be real or represent conjugate pairs. There are three cases of interest: complex roots, repeated real roots, and distinct real roots.

Complex roots are the result of a line which does not intersect the ellipsoid. Repeated real roots imply the line only intersects the ellipsoid once. In such cases the line lies within the tangent plane at the intersection point. Two distinct real roots imply the line travels through the ellipsoid with two distinct intersection points.

It is also interesting to note the case of one positive and one negative root of (D.2). Such results imply x lies within the ellipse. Conversely two positive or two negative roots imply x lies before or after the ellipse with respect to the direction θ.

The tangent plane associated with a point on the ellipse can be determined by differentiating (D.2) which yields

$$\nabla h_e(x) = \left(A + A^T\right) x. \tag{D.3}$$

This is the outward-facing normal vector at x. From the constraint $h_e(x) = 0$, using (D.1) and (D.3), we have $x^T \nabla h_e(x) = 2$. We can define the tangent plane at x as the set

$$P_e = \{y : y \cdot \nabla h_e(x) = 2\}. \tag{D.4}$$

.

D.2 Ray Discretization through Uniformly Sampled Image

Many iterative algorithms require a forward operator for sampled images. For the BRT, the forward operator involves integrals along two rays. For each point in the sampled data, we

must resolve the contribution of each voxel in the image. This operator is sparse in that each broken ray intersects few voxels. To represent these line integrals concisely, we first consider a related problem: the discretization of a single ray through a sampled image.

Along a ray, through a sampled image, the associated sample index changes as a function of length. We approximate the image as constant-valued over voxels. The sample index along the ray only changes as the ray passes through voxel boundaries. This piecewise-constant function can be summarized by the location of discontinuities and associated change in sample indices. In the following we justify this representation and present an efficient algorithm for determining its parameters. This is straightforward for sampling over rectangular pixels. However, generalizing the problem to higher dimensions, and sampling over nonorthogonal basis vectors [8], requires additional care.

Here we consider a ray passing through an image which has been discretized over a uniformly sampled lattice. We seek a a right-contiguous discretization of a ray. The discontinuities occur at ray lengths where the ray traverses boundaries of the sampling lattice. Ultimately, we will define a one-dimensional discretization of the ray $l(m) : \mathbb{Z} \to \mathbb{R}$ and associated lattice coordinates $k(m) : \mathbb{Z} \to \mathbb{Z}^n$ for segment index m and lattice dimension n.

We consider a mapping of coordinates, k, to spatial locations using $f : \mathbb{Z}^n \to \mathbb{R}^n$

$$f(k) = x_0 + Vk. \tag{D.5}$$

Here x_0 represents the spatial location of the point $k = 0$. The matrix V represents a set of linearly independent basis vectors. We will refer to the i^{th} column of V using v_i.

Let $X \subset \mathbb{R}^n$ represent a point lattice

$$X = \{f(k) : k \in \mathbb{Z}^n\}. \tag{D.6}$$

We partition $\mathbb{R}^n$ into voxels, indexed using coordinates $k \in \mathbb{Z}^n$, with centroids $f(k) \in X$. When $V = I$ (and $n = 2$) the resulting voxels are simply square unit pixels. In general the voxels are parallelepipeds with edges parallel to the columns of V.

To map each spatial location to a voxel coordinate we consider an inverse mapping from $x \in \mathbb{R}^n$ to $k \in \mathbb{Z}^n$. For this we approximate the sampled image as constant-valued over each

voxel. The voxels are bounded by a series of planes. The normal vectors to these planes are given by the basis vectors for the reciprocal lattice. We introduce the notation * to indicate reciprocal basis vectors

$$V^{-1} = \begin{bmatrix} v_1^* & v_2^* & \cdots & v_n^* \end{bmatrix}. \tag{D.7}$$

Making use of these vectors, we form a mapping $g : \mathbb{R}^n \to \mathbb{Z}^n$

$$g(x) = \lfloor V^{-1}(x - x_0) \rceil. \tag{D.8}$$

Here we use $\lfloor \cdot \rceil$ to represent the nearest-integer function. Clearly, $(g \circ f)(k) = k$ for all $k \in \mathbb{Z}^n$.

We now turn our attention to the ray. Let $r_0 \in \mathbb{R}^n$ represent the ray origin, and $\theta \in S^{n-1}$ represent the ray direction. Locations along the ray are specified by the function $r(l) : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$,

$$r(l) = r_0 + l\theta. \tag{D.9}$$

We can now map ray length to a lattice coordinate using (D.8) and (D.9)

$$(g \circ r)(l) = \lfloor V^{-1}(r_0 + l\theta - x_0) \rceil. \tag{D.10}$$

We decompose this expression as n scalar functions $h_i : \mathbb{R}_{\geq 0} \to \mathbb{Z}$

$$h_i(l) = \lfloor b_i + l w_i \rceil, \tag{D.11}$$

using the change of variables $w_i, b_i \in \mathbb{R}$

$$w_i \coloneqq v_i^* \cdot \theta, \qquad\qquad b_i \coloneqq v_i^* \cdot (r_0 - x_0). \tag{D.12}$$

We recognize (D.11) as the quantization of an affine function.

Immediately we have $h_i(0) = \lfloor b_i \rceil$. For $l > 0$, we find $h_i(l)$ is piecewise-constant, and discontinuities occur at frequency w_i. At each discontinuity, $h_i(l)$ monotonically changes by $\operatorname{sgn}(w_i)$. Let ℓ_i indicate the location of the first discontinuity satisfying

$$b_i + \ell_i w_i = \lfloor b_i \rceil + \frac{1}{2}\operatorname{sgn}(w_i). \tag{D.13}$$

We then specify $h_i(l)$ as the left-continuous function with segments indexed by $m \in \mathbb{Z}_{\geq 0}$ according to

$$l_i(m) = \ell_i + \frac{m}{|w_i|} \tag{D.14}$$

$$(h_i \circ l_i)(m) = \lfloor b_i \rceil + \operatorname{sgn}(w_i)\, m. \tag{D.15}$$

Combining these definitions for $i \in \{1, \ldots, n\}$, we obtain a discretization of the ray. When $|w_i| \approx 0$, we assume the i^{th} lattice coordinate does not change (no discontinuities occur).

In application we are limited to a finite collection of samples. This implies a maximum length along the ray before the ray exits the sampled volume. We consider a finite number of samples along the i^{th} direction indicated by K_i. Ensuring $(h_i \circ l_i)(m) \in [0, K_i - 1]$ implies an upper bound on m. We refer to this bound as

$$\widehat{M_i} = \begin{cases} \infty, & w_i = 0 \\ K_i - 1 - \lfloor b_i \rceil, & w_i > 0 \\ \lfloor b_i \rceil & w_i < 0. \end{cases} \tag{D.16}$$

The maximum ray length associated with the sampled volume is therefore

$$\ell_{\text{MAX}} \coloneqq \min_i l_i\left(\widehat{M_i}\right), \tag{D.17}$$

which we use to correct $M_i \leq \widehat{M_i}$ using

$$M_i \coloneqq \lfloor (\ell_{\text{MAX}} - \ell_i)\, |w_i| \rfloor. \tag{D.18}$$

For $|w_i| \approx 0$, we assign $M_i = -1$ indicating no discontinuities occur. All that remains is to interleave the discontinuities across all basis vectors.

We form the array of unordered discontinuities

$$t = \begin{bmatrix} l_1(0) & l_1(1) & \cdots & l_1(M_1) & l_2(0) & \cdots & l_n(M_n) \end{bmatrix}^T. \tag{D.19}$$

At each discontinuity, the lattice coordinates update monotonically. To determine the coordinates, we must cumulatively sum the discretization boundaries along each sampling basis.

Associated with t, we define a binary matrix indicating updates to lattice coordinates for each basis $i \in \{1, \ldots, n\}$. We store these in the block-diagonal matrix S with n blocks each comprising a column vector of length $(M_i + 1)$,

$$[S]_i = |w_i| \, \mathbf{1}_{M_i+1}. \tag{D.20}$$

Here $\mathbf{1}_N$ is the N-length column vector of all ones. In the event $|w_i| \approx 0$, we assign $M_i = -1$, effectively setting the i^{th} column of S to all zeros with the inclusion of no additional rows. Finally, we sort the rows of S using t. Let P represent the permutation matrix such that the vector Pt is monotonically increasing.

Putting this together, we partition the ray into right-contiguous segments associated with segment index $m > 0$. The length associated with the start of each ray segment is therefore

$$l(m) = [Pt]_m \tag{D.21}$$

and is associated with the lattice coordinates

$$[k(m)]_i = \lfloor b_i \rceil + \sum_{j=0}^{m-1} [PS]_{j,i} \, . \tag{D.22}$$

The origin of the ray is associated with $l(0) = 0$ and $[k(0)]_i = \lfloor b_i \rceil$. The length segment m is therefore $l(m+1) - l(m)$ which is, in general, nonuniform.

Appendix E

Additional Numeric Algorithms

E.1 Filtering Sampled Cone Beam Transform Data

With extended CBT data available, filtered data can be synthesized by superimposing a shifted copy of the original data with the extended data. This process is described in Algorithm 7.

Algorithm 7 CbtFilter: Bound support of the data by adding a negated shifted copy. The original data are first extended. The shift distance is determined automatically such that the horizontal translation is exactly N_t samples (width of B). Here we use horzcat to horizontally concatenate matrices.

Require: $B \in \mathbb{R}^{N_y \times N_t}$, $\alpha \in \mathbb{R}^+$, $p \in \mathbb{Z}^+$
Ensure: $B^m \in \mathbb{R}^{N_y+N_\xi \times 2N_t}$, $\mathbf{b}_s \in \mathbb{R}^{N_y+1}$
1: $N_\xi = \lceil \alpha N_t \rceil$
2: $Q_2, Q_3, Q_4 \leftarrow \text{CbtExtend}(B(:,1), B(1,:)^T, |\alpha|, p)$ ▷ Algorithm 1
3: $B_1 = \text{vertcat}(Q3, Q2)$
4: $B_2 = \text{vertcat}(Q4, B)$
5: $W = \text{NonIntShift}(B_2, -\alpha N_t, p+1, [])$ ▷ Algorithm 12
6: $\mathbf{b}_s = W(1 : N_y + 1, N_t)$
7: $B_1(1 : N_y + 1, 1 : N_t) \mathrel{+}= W(1 : N_y + 1, :)$
8: $B^m = \text{horzcat}(B_1, B_2)$

E.2 Filtering Sampled Broken Ray Transform Data

Numeric filtering of BRT data is simplified when at least one direction (θ_i or θ_j) is axis aligned. The process is simplified because integer-sample translations along axis-aligned directions require no interpolation. Using the definitions in Section 2.5.1, we extend the data in the direction $+t$ (or $-\theta_i$) by simply replicating the column $G(:, N_t)$. However, we must extend the data after translation in the direction $-y$, as well. It is for this reason Algorithm 7 returns the additional term $\mathbf{b}_s$. The complete algorithm for filtering BRT data is detailed in Algorithm 8. Notice here we expand support of the scatter angle to $\xi \in (-\pi/2, \pi/2)$ by flipping G on input and G^m on return based on the sign of the extension length, α.

Algorithm 8 BrtFilter: Filter truncated BRT data.

Require: $G \in \mathbb{R}^{N_y \times N_t}$, $\alpha \in \mathbb{R}$, $p \in \mathbb{Z}^+$
Ensure: $G^m \in \mathbb{R}^{(N_y+N_\xi)\times 3N_t}$

if $\alpha < 0$ **then**
    $G = \text{flipud}(G)$
end if
$N_\xi = \lceil |\alpha| N_t \rceil$
$B^m, \mathbf{b}_s \leftarrow$ CbtFilter(G, $|\alpha|$, p) ▷ Algorithm 7
$G^m = \text{horzcat}(-B^m(:, 1 : N_t), B^m)$
$G^m \mathrel{+}= -B^m(:, N_t + 1 : 2N_t)$
$G^m(1 : N_y + 1, N_t + 1 : 2 * N_t) \mathrel{+}= \mathbf{b}_s$
$G^m(N_\xi + 1 : N_\xi + N_y, 2 * Nt + 1 : 2N_t) \mathrel{+}= -G(:, N_t)$
if $\alpha < 0$ **then**
    $G^m = \text{flipud}(G^m)$
end if

E.3 Inversion of BRT Data with Regularization

In Section E.2 we presented a filtering algorithm for BRT data, which assumed $\xi \in (-\pi/2, \pi/2)$. For such data, a complete inversion algorithm is listed in Algorithm 9. In addition to filtering the data, this algorithm truncates the result. We crop the result such that the output image has the same dimensions as the input and reconstructs μ devoid of shifted copies.

Algorithm 9 BrtInvert: Invert truncated BRT data. This implementation accepts only one angle ξ. We assume the other direction is aligned with the direction $-t$ (i.e. π as indicated in Line 3).

Require: $G \in \mathbb{R}^{N_y \times N_t}$; $\Delta_t, \Delta_y, \epsilon \in \mathbb{R}^+$; $\xi \in (-\pi/2, \pi/2)$; $p \in \mathbb{Z}^+$
Ensure: $\Psi \in \mathbb{R}^{N_y \times N_t}$

1: $\alpha = \frac{\Delta_t}{\Delta_y} \tan \xi$
2: $G^m \leftarrow \text{BrtFilter}(G, \alpha, p)$ ▷ Algorithm 8
3: $\Psi^m =\leftarrow \text{BrtInvertFiltered}(G^m, \Delta_t, \Delta_y, \epsilon, \xi, \pi)$ ▷ Algorithm 2
4: $N_0 = \max(0, \lceil \alpha N_t \rceil)$ ▷ Account for filtering shift
5: $\Psi = \Psi^m(N_0 + 1 : N_0 + N_y, 2N_t + 1 : 3N_t)$ ▷ Truncate filtered copies

E.4 Modified Broken Ray Transform Data and Filtering

The MBRT is the difference between two BRT data sets. The motivation for this is to cancel the spatially varying scatter density term using two BRTs sharing one common direction. In addition, this operation cancels the attenuation effects along the common direction. Filtering need not address the common direction. This motivates a subtly different approach to filtering for MBRT data.

We distinguish the two BRT data sets by the unique scatter angles ξ_i and ξ_j. We will use the same subscript to distinguish the data G_i from G_j. We consider the MBRT of $G_i - G_j$.

Since the MBRT represents a linear combination, as does filtering, there is some flexibility in the order of operations. Filtering could be applied to the modified BRT data $(G_i - G_j)$ or applied to G_i and G_j separately. Again, the challenge lies in the data extension. For MBRT data, $|\xi_i - \xi_j|$ may be small. It may be difficult to distinguish $(R\mu_C)(x \cdot \theta_i^\perp, \theta_i)$ from $(R\mu_C)(x \cdot \theta_j^\perp, \theta_j)$. Instead we assume a known background scatter density that is constant along the perimeter of the data. This term can be removed as a correction to the data. In this case, $(R\mu_C)(x \cdot \theta_i^\perp, \theta_i)$ and $(R\mu_C)(x \cdot \theta_j^\perp, \theta_j)$ can be recovered from G_i and G_j separately. Under this assumption G_i and G_j can be extended and filtered independently.

Algorithm 10 operates on BRT data independently. It does not bound support of the data as we do not address the incident direction. However, results from separate BRT data, at

different scatter angles, can be superimposed to form filtered MBRT data. This process is demonstrated in Algorithm 11 lines (3)-(5).

Algorithm 10 MbrtFilter: Filtering truncated BRT data for MBRT. Generate 4 shifted copies of the data to ensure bounded support in two scatter directions. Here we ignore the incident direction. We distinguish the direction α_i associated with the input BRT data. The input α_j identifies the scatter direction associated with the other BRT data externally used to form the MBRT.

Require: $G \in \mathbb{R}^{N_y \times N_t}$; $\alpha_i, \alpha_j \in \mathbb{R}$; $p \in \mathbb{Z}^+$
Ensure: $G^m \in \mathbb{R}^{(N_y+N_{\xi_i}+N_{\xi_j})\times 3N_t}$
1: $N_{\xi_i} = \lceil |\alpha_i| N_t \rceil$
2: $N_{\xi_j} = \lceil |\alpha_j| N_t \rceil$
3: **if** $\alpha_i < 0$ **then**
4: $\quad G = \text{flipud}(G)$
5: **end if**
6: $B^m \leftarrow$ CbtFilter(G, $|\alpha_i|$, p) ▷ Algorithm 7
7: **if** $\alpha_i < 0$ **then**
8: $\quad B^m = \text{flipud}(B^m)$
9: **end if**
10: **if** $\alpha_j < 0$ **then**
11: $\quad B^m = \text{vertcat}(B^m, \mathbf{0}^{N_{\xi_j} \times 2N_t})$
12: **else**
13: $\quad B^m = \text{vertcat}(\mathbf{0}^{N_{\xi_j} \times 2N_t}, B^m)$
14: **end if**
15: $W \leftarrow$ NonIntShift(B^m, $-N_t\alpha_j$, $p+1$) ▷ Algorithm 12
16: $B_s = W(1 : N_{\xi_i} + N_{\xi_j} + N_y, :)$
17: $G^m = \text{horzcat}(-B^s(:, 1 : N_t), B^m)$
18: $G^m(:, N_t + 1 : 2N_t) \mathrel{+}= -B_s(:, N_t + 1 : 2N_t)$

For MBRT the forward operator has a subtly different form. Due to the difference of CBTs, the forward operator (2.4) is not directly applicable. Instead we have

$$\frac{-1}{j2\pi\,(w \cdot \theta_i)} + \frac{1}{j2\pi\,(w \cdot \theta_j)} = \frac{-w \cdot (\theta_i + (-\theta_j))}{j2\pi\,(w \cdot \theta_i)\,(w \cdot (-\theta_j))}. \tag{E.1}$$

Notice the right-hand side of (E.1) is equivalent to (2.4) excepting the sign change for θ_j. Here we lose commutativity with respect to θ_i and θ_j. Of course negating θ_j is equivalent to adding π to ξ_j. To invert filtered MBRT data, we can reuse Algorithm 2 by adding π to the input ξ_j. Algorithm 11 describes an approach for filtering truncated BRT data sets, sharing a common direction, which is robust against spatially varying scatter density.

Algorithm 11 MbrtInvert: Invert MBRT from two BRT data sets sharing one common direction. The unique directions ξ_i and ξ_j distinguish the data G_i from G_j.

Require: $G_i, G_j \in \mathbb{R}^{N_y \times N_t}$; $\Delta_t, \Delta_y, \epsilon \in \mathbb{R}^+$; $\xi_i, \xi_j \in (-\pi/2, \pi/2)$; $p \in \mathbb{Z}^+$
Ensure: $\Psi \in \mathbb{R}^{N_y \times N_t}$

$\alpha_i = \frac{\Delta_t}{\Delta_y} \tan \xi_i$
$\alpha_j = \frac{\Delta_t}{\Delta_y} \tan \xi_j$
$G_i^m \leftarrow \text{MbrtFilter}(G_i,\, \alpha_i,\, \alpha_j,\, p)$ ▷ Algorithm 10
$G_j^m \leftarrow \text{MbrtFilter}(G_j,\, \alpha_j,\, \alpha_i,\, p)$
$G^m = G_i^m - G_j^m$ ▷ Filtered MBRT data
$\Psi^m =\leftarrow \text{BrtInvertFiltered}(G^m, \Delta_t,\, \Delta_y,\, \epsilon,\, \xi_i,\, \pi + \xi_j)$ ▷ Algorithm 2
$N_0 = \max(0, \lceil \alpha_i N_t \rceil)$
$N_0 \mathrel{+}= \max(0, \lceil \alpha_j N_t \rceil)$
$\Psi = \Psi^m(N_0 + 1 : N_0 + N_y, 2N_t + 1 : 3N_t)$

E.5 Non-Integer Shifts of Sampled Signals

Non-integer shifts of sampled signals requires interpolation. A shift in the spatial domain represents a phase ramp in the frequency domain. Fast implementation of the discrete Fourier transform can be leveraged to implement a computationally efficient shifting algorithm for sampled signals. This approach is exact for periodic, band-limited signals. However, application for aperiodic signals requires additional considerations.

For continuous signal $x(t)$, and uniform sample spacing Δ, we define the sampled signal

$$x[n] \coloneqq x(\Delta n), \quad \forall\, n = \{0, \ldots, N-1\}. \tag{E.2}$$

We seek an approximation to

$$z[n] = x(\Delta n - \Delta s), \quad \forall n = \{0, \ldots, N-1\} \tag{E.3}$$

$$= x[n - s]. \tag{E.4}$$

from the discrete samples $x[n]$. The problem is that when s is non-integer valued, $x[n-s]$ is not available directly. For this we make use of the DFT.

For an N-length signal $x[n]$, we define the Fourier coefficients using the DFT

$$y[m] = \sum_{n=0}^{N-1} x[n] \exp\left(-j2\pi nm/N\right) \tag{E.5}$$

for $m \in \{0, \ldots, N-1\}$. The corresponding IDFT is

$$x[n] = \mathrm{DFT}^{-1}\left\{y[m]\right\} \tag{E.6}$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} y[m] \exp\left(j2\pi nm/N\right) \tag{E.7}$$

To approximate $z[n]$, from $y[m]$ we simply plug (E.4) into (E.7)

$$z[n] = \frac{1}{N} \sum_{m=0}^{N-1} y[m] \exp\left(j2\pi(n-s)m/N\right) \tag{E.8}$$

$$= \mathrm{DFT}^{-1}\left\{y[m] \exp\left(-j2\pi sm/N\right)\right\}. \tag{E.9}$$

Since s is represented in samples (E.9) is independent of sampling rate. This is particularly efficient when multiple shifted copies of the same signal are required. In such cases $y[m]$ need only be computed once. Additional savings are realized computing the IDFT in (E.9) for all signals at once.

This approach is exact for periodic, band-limited, signals. In application the input signals do not satisfy this requirement. There are two distinct consequences due to aperiodicity. First, if $x[0] \neq x[N-1]$, the periodicity assumption of the DFT will effect aliasing artifacts. Small shifts may effect large oscillations around $z[0]$ and $z[N-1]$. The mitigation strategy for this is to add samples to X smoothing the transition. Reducing the transition reduces the effects of sidelobes. Secondly, shifting an aperiodic requires extrapolation. The assumed values of this extrapolated signal (e.g. 0) should be incorporated in the input signal.

For application of (E.9) to aperiodic signals, we define three vectors

$$\mathbf{x}_L \approx \begin{bmatrix} x(-N_L\Delta) & x(-(N_L-1)\Delta) & \cdots & x(-\Delta) \end{bmatrix}^T \tag{E.10}$$

$$\mathbf{x}_C = \begin{bmatrix} x(0) & x(\Delta) & \cdots & x((N_C-1)\Delta) \end{bmatrix}^T \tag{E.11}$$

$$\mathbf{x}_R \approx \begin{bmatrix} x(N_C\Delta) & x((N_C+1)\Delta) & \cdots & x((N_C+N_R-1)\Delta) \end{bmatrix}^T. \tag{E.12}$$

Here we use $\mathbf{x}_C$ to represent an N_C-length vector of the available data. We use vectors $\mathbf{x}_L$ and $\mathbf{x}_R$ to approximate samples to the left and right of $\mathbf{x}_C$ respectively. The intention is to minimize sidelobe artifacts within a window of output samples and precision is not required. In the absence of additional information, $\mathbf{x}_L$ and $\mathbf{x}_R$ could be filled with values $x(0)$ and $x((N_C-1)\Delta)$ respectively. We then construct the input vector

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_C^T & \mathbf{x}_R^T & 0^p & \mathbf{x}_L^T \end{bmatrix}^T. \tag{E.13}$$

where 0^p is a p-length row-vector of all zeros. The length of $\mathbf{x}$ is therefore

$$N = N_C + N_R + p + N_L. \tag{E.14}$$

The selection of p, N_L, and N_R will depend on $x(t)$ and s. For $s > 0$, it is expected $K + N_L > s$ to avoid overlap with $\mathbf{x}_R$. Alternatively for $s < 0$, we expect $K + N_R > |s|$ to avoid overlap with $\mathbf{x}_L$.

Algorithm 12 describes this processes for a vector of shift values $\mathbf{s}$. In the resulting matrix the first and final rows approximately represent

$$Z_{n,m} \approx x(\Delta(n - s[m])) \tag{E.15}$$

$$Z_{N-n,m} \approx x(\Delta(-n - s[m])) \tag{E.16}$$

respectively. Some applications may require repartitioning the result based on the sign of s.

Algorithm 12 NonIntShift: Non-integer shifting of a sampled signal. Without loss of generality, we use $\mathbf{x}_R$ to refer to the vertical concatenation of (E.11) and (E.12). We use vertcat$\{\cdot\}$ to vertically concatenate matrices, and $\odot$ represents element-wise multiplication with assumed expansion along singleton dimensions.

Require: $\mathbf{x}_R$, $\mathbf{s}$, p, $\mathbf{x}_L$

Ensure: Z

1: $\mathbf{x} = \text{vertcat}(\mathbf{x}_R,\, 0^p,\, \mathbf{x}_L)$
2: $\mathbf{n} = \begin{bmatrix} 0 & 1 & \cdots & \dim(\mathbf{x}) - 1 \end{bmatrix}^T$
3: $\mathbf{y} = \text{DFT}(\mathbf{x})$
4: $W = \exp\left(-j2\pi \left(\mathbf{n} \odot \mathbf{s}^T\right)/N\right)$
5: $Z = \text{DFT}^{-1}\{\mathbf{y} \odot W\}$

E.6 Real BRT Operator

While many DFT implementations operate on complex signals, we are interested in transforms on real signals only. Real signals exhibit conjugate symmetry in the frequency domain. This symmetry can be exploited to transform a length 2N real signal as a length-N complex signal. For multi-dimensional images this trick can only be exploited in one dimension only. However, this may be particularly useful for BRT computations of large images. In the following we generalize this as filtering an $N_2 \times 2N_1$ image.

For the real Fourier BRT implementation, we start by decimating the image. Assuming $X \in \mathbf{R}^{N_2 \times 2N_1}$, we define $F \in \mathbf{C}^{N_2 \times N_1}$

$$[F]_{n,m} \coloneqq [X]_{n,2m} + j\,[X]_{n,2m+1} \tag{E.17}$$

From the Fourier BRT system matrix (2.66), we compose two matrices

$$\left[\tilde{H}_A\right]_{n,m} = \left[\tilde{H}_\theta\right]_{n,m} (1 + \sin(-2\pi m/2N_1)) + \left[\tilde{H}_\theta\right]_{n,m+N_1} (1 - \sin(-2\pi m/2N_1)) \tag{E.18}$$

$$\left[\tilde{H}_B\right]_{n,m} = \left[\tilde{H}_\theta\right]_{n,m+N_1} \cos(2\pi m/2N_1) \tag{E.19}$$

Taking the two-dimensional Fourier transform of F, we apply the filter

$$\left[\tilde{G}\right]_{n,m} = \frac{1}{2}\tilde{F}(n,m)\left[\tilde{H}_A\right]_{n,m} + j\frac{1}{2}\tilde{F}^*(N_2-n, N_1-m)\left[\tilde{H}_B\right]_{n,m}. \tag{E.20}$$

Taking the inverse two-dimensional Fourier transform of G we approximate the filtered image

$$[Y]_{n,2m} = \Re\left\{[G]_{n,m}\right\} \tag{E.21}$$

$$[Y]_{n,2m+1} = \Im\left\{[G]_{n,m}\right\}. \tag{E.22}$$

For the adjoint BRT operator, we simply replace $\tilde{H}_A$ and $\tilde{H}_B$ in (E.20) with their complex conjugates. This process is described in Algorithm 13.

Algorithm 13 RealFBRT: Fourier BRT on real images employing decimation by two.

Require: $X \in \mathbb{R}^{L_2 \times L_1}$; $\widetilde{H} \in \mathbb{C}^{N_2 \times N_1}$; $L_2 \leq N_2, L_1 \leq N_1$
Ensure: $Y \in \mathbb{R}^{L_2 \times L_1}$

Compute $F(n,m) = X(n,2m) + jX(n,2m+1)$ ▷ Decimation by 2
$\widetilde{F} = \mathrm{DFT}^2(F, N_2, N_1)$ ▷ Zero pad input
Compute $\widetilde{F}_B(n,m) = \widetilde{F}_B(N_2-n, N_1-m)$
if ADJOINT **then**
    $\widetilde{G} = \frac{1}{2}\widetilde{F} \odot \widetilde{H}_A^* + j\frac{1}{2}\widetilde{F}_B \odot \widetilde{H}_B^*$ ▷ backward BRT
else
    $\widetilde{G} = \frac{1}{2}\widetilde{F} \odot \widetilde{H}_A + j\frac{1}{2}\widetilde{F}_B \odot \widetilde{H}_B$ ▷ forward BRT
end if
$G = \mathrm{DFT}^{-2}\left(\widetilde{G}\right)$
$Y(n,2m) = \Re\{G(n,m)\}$ ▷ Truncate result
$Y(n,2m+1) = \Im\{G(n,m)\}$ ▷ Truncate result

Appendix F

Data Fidelity and Regularization Terms

F.1 Scatter Image Fidelity

Our choice of $\mathcal{Y}$ for indexing both $\mathbf{d}$ and $\boldsymbol{\alpha}$ aides separability when updating scatter image estimates. Differentiating (3.3) with respect to $\alpha(y)$ we obtain separable functions for each $y \in \mathcal{Y}$. Further, (3.3) is convex with respect to $\boldsymbol{\alpha}$ for $\boldsymbol{\mu}$ fixed. Surrogate approximations are not necessary to update $\boldsymbol{\alpha}$. This is demonstrated through the following Lemma.

Lemma F.1. *The KL-divergence* (3.3) *is convex over* $\boldsymbol{\alpha} \in \mathcal{S}$ *when there exists at least one* $i \in \mathcal{I}$ *such that* $d_i(y) > 0$.

Proof. For convenience we define

$$\dot{g}_i(y : \boldsymbol{\mu}) \coloneqq \frac{\partial}{\partial \alpha(y)} g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) \tag{F.1}$$

$$= I_0(y) \exp\left(- \sum_{x \in \mathcal{X}} h_i\left(y|x\right) \mu(x) \right) \tag{F.2}$$

We emphasize $\dot{g}_i(y : \boldsymbol{\mu})$ is independent of $\boldsymbol{\alpha}$. Since the BRT of the image is finite, and $I_0(y) > 0$, we have $\dot{g}_i(y : \boldsymbol{\mu}) > 0$ for all y, i.

Taking the derivative of (3.3) with respect to $\alpha(y)$, we find

$$\frac{\partial I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right)\big)}{\partial \alpha(y)} = \sum_{i \in \mathcal{I}} \left(\dot{g}_i(y) - \frac{d_i(y) \dot{g}_i(y : \boldsymbol{\mu})}{\alpha(y) \dot{g}_i(y : \boldsymbol{\mu}) + \beta_i(y)} \right). \tag{F.3}$$

The second derivative is then

$$\frac{\partial^2 I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right)\big)}{\partial \alpha(y)^2} = \sum_{i \in \mathcal{I}} \frac{d_i(y) \dot{g}_i(y : \boldsymbol{\mu})}{\left(\alpha(y) \dot{g}_i(y : \boldsymbol{\mu}) + \beta_i(y)\right)^2}. \tag{F.4}$$

If $d_i(y)$ is positive for at least one $i \in \mathcal{I}$, then (F.4) is also positive. Therefore (3.3) is strictly convex over $\boldsymbol{\alpha} \in \mathcal{S}$. □

F.2 Attenuation Fidelity Surrogate

Direct minimization of (3.3), with respect to $\boldsymbol{\mu}$, is complicated by $h_i(y|x)$ and $\beta_i(y) > 0$. Instead, local surrogate approximations lead to computationally efficient updates guaranteeing monotonic reduction of (3.3). Here we adopt the approach of O'Sullivan and Benac: recasting the problem as joint estimation over members of a linear family and an exponential family [44].

The first surrogate is found expanding both $\mathbf{g}$ and $\mathbf{d}$ as linear combinations. For this purpose, two families of functions are introduced. Let $\mathcal{L}(\mathbf{d})$ define a linearly family whose marginals equal the data

$$\mathcal{L}(\mathbf{d}) = \left\{ \mathbf{p} : p_i(y, E) \geq 0, \quad \sum_E p_i(y, E) = d_i(y) \right\}. \tag{F.5}$$

Let $\mathcal{E}(\boldsymbol{\alpha})$ define an exponential family, associated with $\boldsymbol{\alpha}$, and parameterized by $\boldsymbol{\mu}$

$$\mathcal{E}(\boldsymbol{\alpha}) = \left\{ \mathbf{q} : q_i(y, 0 : \boldsymbol{\mu}) = \beta_i(y), q_i(y, 1 : \boldsymbol{\mu}) = I_0(y) \alpha(y) \exp\left(- \sum_{x \in \mathcal{X}} h_i\left(y|x\right) \mu(x) \right) \right\} \tag{F.6}$$

Using a single element of the exponential family, we composes the data model

$$g_i(y : \boldsymbol{\alpha}, \boldsymbol{\mu}) = \sum_E q_i(y, E : \boldsymbol{\mu}). \tag{F.7}$$

In [44], E was used to distinguish spectral measurements. However, the index remains useful for mono-energetic measurements when $\beta(y) > 0$.

The divergence between elements of the linear and exponential families reads

$$I\big(\mathbf{p} \,\big\|\, \mathbf{q}\big) = \sum_{i\in\mathcal{I}} \sum_{y\in\mathcal{Y}} \sum_{E} \left(p_i(y,E) \ln \frac{p_i(y,E)}{q_i(y,E:\boldsymbol{\mu})} - p_i(y,E) + q_i(y,E:\boldsymbol{\mu}) \right). \tag{F.8}$$

Let $\hat{\mathbf{q}} \in \mathcal{E}(\boldsymbol{\alpha})$ indicate the element of the exponential family associated with $\hat{\boldsymbol{\mu}}$. Fixing $\hat{\mathbf{q}}$ in (F.8), we consider the minimizer $\mathbf{p} \in \mathcal{L}(\mathbf{d})$ which is subject to the linear constraint (F.5). The result is available in closed form

$$\hat{\mathbf{p}} = \operatorname*{arg\,min}_{\mathbf{p}\in\mathcal{L}(\mathbf{d})} I\big(\mathbf{p} \,\big\|\, \hat{\mathbf{q}}\big) \tag{F.9a}$$

$$= \left\{ p_i(y,E) = d_i(y) \frac{q_i(y,E:\hat{\boldsymbol{\mu}})}{\sum_{E'} q_i(y,E':\hat{\boldsymbol{\mu}})} \right\}. \tag{F.9b}$$

Plugging these result back into (F.8), we find a variational form of (3.3),

$$I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha},\boldsymbol{\mu}\right)\big) = \min_{\mathbf{p}\in\mathcal{L}(\mathbf{d})} I\big(\mathbf{p} \,\big\|\, \mathbf{q}\big), \tag{F.10}$$

due to O'Sullivan and Benac [44]. This motivates the surrogate function

$$D(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) \coloneqq I\big(\hat{\mathbf{p}} \,\big\|\, \mathbf{q}\big), \tag{F.11}$$

where $\hat{\mathbf{p}}$, parameterized by $\hat{\boldsymbol{\mu}}$, is given by (F.9b). That (F.11) serves as a surrogate for (3.3) can be summarized as

$$D\left(\hat{\boldsymbol{\mu}} : \hat{\boldsymbol{\mu}}\right) = I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha},\hat{\boldsymbol{\mu}}\right)\big) \tag{F.12}$$

$$D\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right) \geq I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha},\boldsymbol{\mu}\right)\big), \quad \forall \boldsymbol{\mu} \in \mathcal{A}. \tag{F.13}$$

The equality in (F.12) is a restatement of (F.10). The inequality in (F.13) is an application of the convex decomposition lemma [44]

$$f\left(\sum_x t(x)\right) \leq \sum_x r(x) f\left(\frac{t(x)}{r(x)}\right). \tag{F.14}$$

This holds for all $\mathbf{r} \in \{\mathbf{r} : r(x) \geq 0, \sum_x r(x) = 1\}$ and follows from Jensen's inequality [44].

Equivalently, we restate

$$D(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) = d_0\left(\hat{\boldsymbol{\mu}}\right) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} \hat{p}_i(y,1) \sum_{x\in\mathcal{X}} h_i(y|x)\mu(x) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} I_0(y)\alpha(y) \exp\left(-\sum_{x\in\mathcal{X}} h_i(y|x)\mu(x)\right), \quad \text{(F.15)}$$

where we have summarized all of the terms which do not depend on $\boldsymbol{\mu}$ with the single additive scalar

$$d_0\left(\hat{\boldsymbol{\mu}}\right) = \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} \left(\sum_E \hat{p}_i(y,E) \ln\left(\hat{p}_i(y,E)\right)\right) - d_i(y) + \beta_i(y) - \hat{p}_i(y,0)\ln\beta_i(y) - \hat{p}_i(y,1)\ln I_0(y)\alpha(y). \quad \text{(F.16)}$$

Introducing the auxiliary function

$$\psi_i(x|y) \coloneqq h_i(y|x)\left(\mu(x) - \hat{\mu}(x)\right), \quad \text{(F.17)}$$

we restate

$$D(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) = \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} \hat{p}_i(y,1) \sum_{x\in\mathcal{X}} \left(\psi_i(x|y) + h_i(y|x)\hat{\mu}(x)\right) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} \hat{q}_i(y,1) \exp\left(-\sum_{x\in\mathcal{X}} \psi_i(x|y)\right) + d_0\left(\hat{\boldsymbol{\mu}}\right). \quad \text{(F.18)}$$

Recognizing the function

$$f(y,E,t) = t\hat{p}_i(y,1) + \hat{q}_i(y,1)\exp(-t) \quad \text{(F.19)}$$

as convex over t, we again make use of the convex decomposition lemma. This yields

$$D(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) \leq d_0\left(\hat{\boldsymbol{\mu}}\right) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}\Bigg[\hat{p}_i(y,1)h_i(y|x)\mu(x) \\ + r_i(x|y)\hat{q}_i(y,1)\exp\left(-\frac{h_i(y|x)}{r_i(x|y)}\left(\mu(x)-\hat{\mu}(x)\right)\right)\Bigg] \quad \text{(F.20)}$$

for all $r_i(x|y) > 0$ such that

$$\sum_{x\in\mathcal{X}} r_i(x|y) = 1. \quad \text{(F.21)}$$

This constraint can be mitigated with the addition of a dummy $x = 0$ such that $h_i(y|0) = 0$, $\forall y, i$. This has no effect on (F.15), but $x = 0$ contributes to the right-hand side of (F.20). This bias is independent of $\boldsymbol{\mu}$ but varies with $\hat{\mathbf{q}}$. As such it will not affect the current update but obscures reduction in the objective $D(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}})$.

We select

$$r_i(x|y) = \begin{cases} \dfrac{h_i(y|x)}{Z_i(x)}, & x \neq 0 \\ 1 - \displaystyle\sum_{x\in\mathcal{X}\setminus\{0\}} \dfrac{h_i(y|x)}{Z_i(x)}, & x = 0. \end{cases} \quad \text{(F.22)}$$

In general, $Z_i(x)$ must be sufficiently large such that $r_i(0|y) \geq 0$. This motivates the decoupled objective function

$$\overline{D}(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) \coloneqq \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}\setminus\{0\}}\Bigg[\mu(x)\hat{p}_i(y,1)h_i(y|x) \\ + \hat{q}_i(y,1)\frac{h_i(y|x)}{Z_i(x)}\left(\exp\Big(-Z_i(x)\left(\mu(x)-\hat{\mu}(x)\right)\Big)-1\right)\Bigg] \\ + d_0\left(\hat{\boldsymbol{\mu}}\right) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}}\hat{q}_i(y,1). \quad \text{(F.23)}$$

Here we incorporate additional terms due to $x = 0$ such that

$$\overline{D}(\hat{\boldsymbol{\mu}} : \hat{\boldsymbol{\mu}}) = D\left(\hat{\boldsymbol{\mu}} : \hat{\boldsymbol{\mu}}\right) \quad \text{(F.24)}$$

$$\overline{D}(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) \geq D\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right), \quad \forall \boldsymbol{\mu} \in \mathcal{A}. \quad \text{(F.25)}$$

Combining these with (F.13) and (F.12), we find $\overline{D}$ is also a surrogate for $I\big(\mathbf{d} \,\big\|\, \mathbf{g}\left(\boldsymbol{\alpha}, \boldsymbol{\mu}\right)\big)$.

The expression (F.23) can be simplified when $Z_i(x)$ is constant over i and x. For this purpose, we define

$$Z_0 \coloneqq \max_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} h_i(y|x). \tag{F.26}$$

Further simplifying the notation, we define

$$b_0 \coloneqq d_0\left(\hat{\boldsymbol{\mu}}\right) + \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} \hat{q}_i(y,1) \tag{F.27}$$

$$b_1(x) \coloneqq \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} h_i(y|x)\hat{p}_i(y,1) \tag{F.28}$$

$$b_2(x) \coloneqq \sum_{i\in\mathcal{I}}\sum_{y\in\mathcal{Y}} h_i(y|x)\hat{q}_i(y,1). \tag{F.29}$$

The expressions (F.28) and (F.29) comprise adjoint broken ray transforms of $\hat{p}_i(y,1)$ and $\hat{q}_i(y,1)$, respectively. Putting this all together, we have

$$\overline{D}(\boldsymbol{\mu}:\hat{\boldsymbol{\mu}}) = b_0 + \sum_{x\in\mathcal{X}\setminus\{0\}} \left[\mu(x)b_1(x) + b_2(x)\frac{1}{Z_0}\left(\exp\Big(-Z_0\big(\mu(x)-\hat{\mu}(x)\big)\Big)-1\right)\right]. \tag{F.30}$$

The gradient separates as

$$\frac{\partial\overline{D}(\boldsymbol{\mu}:\hat{\boldsymbol{\mu}})}{\partial\mu(x)} = b_1(x) - b_2(x)\exp\Big(-Z_0\big(\mu(x)-\hat{\mu}(x)\big)\Big). \tag{F.31}$$

The second derivative is nonnegative for all $\mu(x)$ since $b_2(x) \geq 0$. Therefore, $\overline{D}$ is convex with respect to $\boldsymbol{\mu}$.

F.3 Regularization Surrogate

We consider the general regularization term for image $\boldsymbol{\mu}$ and sample indices $\mathcal{X}$

$$R(\boldsymbol{\mu}) \coloneqq \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{N}_x} w\left(x,z\right)\phi_\delta\left(\mu(x)-\mu(z)\right). \tag{F.32}$$

Here $\mathcal{N}_x \subset \mathcal{X}$ indicates the collection of voxels within a neighborhood of $x \in \mathcal{X}$, and $\phi_\delta : \mathbb{R} \to \mathbb{R}$ is an edge-preserving potential function. Specifically, we assume ϕ_δ is strictly

convex, even, and $\dot{\phi}_\delta(t)/t$ is monotone decreasing for $t > 0$. Therefore, $R(\boldsymbol{\mu}) \geq 0$ with equality for any constant image $\mu(x) = \mu_0$.

Minimizing (F.32) is problematic for a few reasons. Differentiating this function directly with respect to $\mu(x)$ is not separable as each element of the gradient references $\mu(z)$ for all $z \in \mathcal{N}_x$. Additionally, nonlinear ϕ_δ precludes closed-form solutions. Computationally efficient surrgotates are useful for iterative solvers. A common approach involves two approximations. First, we define a separable surrogate making use of Jensen's inequality. Second, we employ a quadratic approximation to ϕ_δ.

Using a constant image $\hat{\boldsymbol{\mu}}$, we expand $\phi_\delta(\mu(x) - \mu(z))$ using Jensen's inequality

$$\phi_\delta(\mu(x) - \mu(z)) \leq \frac{1}{2}\left[\phi_\delta(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)) + \phi_\delta(2\mu(z) - \hat{\mu}(x) - \hat{\mu}(z))\right]. \tag{F.33}$$

Here we make use of both the convexity and symmetry of ϕ. This motivates the separable surrogate regularization function

$$\breve{R}(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) \coloneqq \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{N}_x} w(x,z)\left[\phi_\delta(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)) + \phi_\delta(2\mu(z) - \hat{\mu}(x) - \hat{\mu}(z))\right] \tag{F.34}$$

due to De Pierro [45]. Separability is emphasized by restating

$$\breve{R}(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}) = \sum_{x\in\mathcal{X}} \breve{R}_x(\mu(x)), \tag{F.35}$$

where

$$\begin{aligned}\breve{R}_x(t) \coloneqq \frac{1}{2}\sum_{z\in\mathcal{N}_x} w(x,z)\,\phi_\delta(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)) \\ + \frac{1}{2}\sum_{z\in\mathcal{N}_x^b} w(z,x)\,\phi_\delta(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)).\end{aligned} \tag{F.36}$$

Here we use $\mathcal{N}_x^b \coloneqq \{y : \mathcal{N}_y \ni x\}$ to represent the set of voxels with x as a neighbor. For symmetric problems, when $\mathcal{N}_x = \mathcal{N}_x^b$ and $w(x,z) = w(z,x)$, the two terms in (F.36) are equivalent.

As a second approximation, we replace ϕ in (F.34) with the quadratic surrogate $\overline{\phi} : \mathbb{R} \to \mathbb{R}$,

$$\overline{\phi}(t) \coloneqq \phi\left(\hat{t}\right) + \dot{\phi}\left(\hat{t}\right)\left(t - \hat{t}\right) + \frac{1}{2}\frac{\dot{\phi}\left(\hat{t}\right)}{\hat{t}}\left(t - \hat{t}\right)^2 \tag{F.37}$$

$$\frac{\partial\overline{\phi}(t)}{\partial t} = \dot{\phi}\left(\hat{t}\right) + \frac{\dot{\phi}\left(\hat{t}\right)}{\hat{t}}\left(t - \hat{t}\right). \tag{F.38}$$

This represents an upper bound on ϕ under the requirements ϕ is convex, symmetric, and when $\dot{\phi}\left(t\right)/t$ is monotone decreasing for $t > 0$. Equality is achieved at $t = \hat{t}$ such that $\overline{\phi}(\hat{t}) - \phi(\hat{t}) = 0$. Further, it can be shown $t = \hat{t}$ minimizes this difference (see Lemma 8.3 in [46]).

Specifying the expansion point in (F.37) as

$$\hat{t} = \hat{\mu}(x) - \hat{\mu}(z) \tag{F.39}$$

$$t - \hat{t} = 2\left(\mu(x) - \hat{\mu}(x)\right), \tag{F.40}$$

we define the separable quadratic surrogate

$$\overline{R}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right) \coloneqq \frac{1}{2}\sum_{x\in\mathcal{X}}\left[\sum_{z\in\mathcal{N}_x} w\left(x,z\right)\overline{\phi}_\delta\left(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)\right) + \sum_{z\in\mathcal{N}_x^b} w\left(z,x\right)\overline{\phi}_\delta\left(2\mu(x) - \hat{\mu}(x) - \hat{\mu}(z)\right)\right]. \tag{F.41}$$

Therefore, $\overline{R}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right) \geq R(f)$ with equality when $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$. For convenience, the following definitions are parameterized by $\hat{\boldsymbol{\mu}}$ and independent of $\boldsymbol{\mu}$:

$$c_0 \coloneqq \frac{1}{2}\sum_{x\in\mathcal{X}}\left[\sum_{z\in\mathcal{N}_x} w\left(x,z\right)\phi_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right) + \sum_{z\in\mathcal{N}_x^b} w\left(z,x\right)\phi_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right)\right] \tag{F.42}$$

$$c_1(x) \coloneqq \sum_{z\in\mathcal{N}_x} w\left(x,z\right)\dot{\phi}_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right) + \sum_{z\in\mathcal{N}_x^b} w\left(z,x\right)\dot{\phi}_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right) \tag{F.43}$$

$$c_2(x) \coloneqq \sum_{z\in\mathcal{N}_x} w\left(x,z\right)\frac{\dot{\phi}_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right)}{\hat{\mu}(x) - \hat{\mu}(z)} + \sum_{z\in\mathcal{N}_x^b} w\left(z,x\right)\frac{\dot{\phi}_\delta\left(\hat{\mu}(x) - \hat{\mu}(z)\right)}{\hat{\mu}(x) - \hat{\mu}(z)}. \tag{F.44}$$

Using these definitions in (F.41), we obtain

$$\overline{R}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right) = c_0 + \sum_{x \in \mathcal{X}} \left(c_1(x) \left(\mu(x) - \hat{\mu}(x)\right) + c_2(x) \left(\mu(x) - \hat{\mu}(x)\right)^2 \right). \tag{F.45}$$

Taking the derivative with respect to $\mu(x)$, we find

$$\frac{\partial \overline{R}\left(\boldsymbol{\mu} : \hat{\boldsymbol{\mu}}\right)}{\partial \mu(x)} = c_1(x) + 2c_2(x) \left(\mu(x) - \hat{\mu}(x)\right). \tag{F.46}$$

Observing $c_2(x) > 0$ for all $\hat{\boldsymbol{\mu}}$, $\overline{R}$ is strictly convex.

Here we derived a surrogate for the regularization image $\boldsymbol{\mu}$. Regularization can be applied to the scatter image by replacing $\boldsymbol{\mu}, \mathcal{X}$ with $\boldsymbol{\alpha}, \mathcal{Y}$, respectively. The weights w, δ, and neighborhood $\mathcal{N}$ may also be tailored to differences in sampling and image scaling.

Appendix G

Related Problems

G.1 Small-Angle Scatter and the Time of Flight Analog

The problem of joint estimation of activity and attenuation arises in positron emission tomography (PET) [35]. Time-of-flight information improves the conditioning of the separation problem for PET as demonstrated by Defrise et al. [32, 34]. In this section we summarize prior results using notation presented in previous sections.

In sinogram coordinates the attenuation-corrected activity distribution for PET in two-dimensions is given by

$$p(s,t,\xi) = \int_{-\infty}^{+\infty} f\left(s\tilde{\beta}^{\perp}(\xi) + l\tilde{\beta}(\xi)\right) w(t-l)dl. \tag{G.1}$$

Here $s \in \mathbb{R}^1$ and $\xi \in [0,\pi)$ are the sinogram coordinates. The direction vector $\tilde{\beta}(\xi) \in S^1$ is parallel to the line of response (LOR). Here we use the $\tilde{}$ decoration to distinguish $\beta \in S^2$ which will be utilized below. The time of flight (TOF) measurement isolates the event location along the LOR with some ambiguity. This is modeled using a Gaussian weighting function

$$w(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-t^2/2\sigma^2\right). \tag{G.2}$$

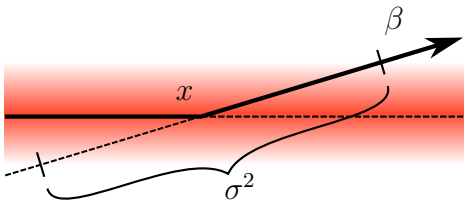


Figure G.1: Finite beam width and small-angle scattering lead to positional ambiguity in scatter source location.

Without loss of generality we assume $\tilde{\beta}(\xi)$, and $\tilde{\beta}^{\perp}(\xi)$ are defined such that $d\tilde{\beta}/d\xi = -\tilde{\beta}^{\perp}(\xi)$, and $d\tilde{\beta}^{\perp}/d\xi = \tilde{\beta}(\xi)$. The function (G.1) then satisfies the differential equation

$$\mathcal{D}p\,(s,t,\xi) = t\frac{dp}{ds} + \frac{dp}{d\xi} - s\frac{dp}{dt} + \sigma^2\frac{d^2p}{dsdt} = 0. \tag{G.3}$$

This differential equation applies when s, t are coordinates in the rotated measurement space. This expression is fundamental to the separation of the attenuation and scatter images as identified by Defrise et al. [32].

In the broader context of the BRT, the scatter location is the break-point of the ray. In all prior work this point is assumed known within quantization errors of the measurement system. Scatter point localization accuracy depends on scatter angle and incident beam width as shown in Fig. G.1. Here we consider positional ambiguity along the scatter direction only.

Transitioning to our notation, we define the attenuation-corrected scatter image over (x, β),

$$p(x,\beta) = \int_{-\infty}^{+\infty} f\,(x+\beta l)\,w(l)dl. \tag{G.4}$$

Here we maintain $x \in \mathbb{R}^3$, and $\beta \in S^2$. We define the directional derivative with respect to the first argument, in the direction β

$$\langle D_1 p(x,\beta), \beta\rangle = \int_{-\infty}^{+\infty} w(l)\beta\cdot\nabla f(x+\beta l)dl \tag{G.5}$$

$$= \int_{-\infty}^{+\infty} w(l) f(x+\beta l)\frac{l}{\sigma^2}dl. \tag{G.6}$$

The second equality is found using integration by parts and depends on the Gaussian weighting function (G.2). Similarly, we have

$$\langle D_1 p(x,\beta),\theta\rangle = \int_{-\infty}^{+\infty} w(l)\theta\cdot\nabla f(x+\beta l)dl. \tag{G.7}$$

Here we make use of an orthogonal direction θ such that $\theta\cdot\beta = 0$. Taking the derivative of (G.6) with respect to x again, in the orthogonal direction θ, we have

$$\langle D_1\langle D_1 p(x,\beta),\beta\rangle,\theta\rangle = \int_{-\infty}^{+\infty} w(l)\theta\cdot\nabla f(x+\beta l)\frac{l}{\sigma^2}dl. \tag{G.8}$$

Taking the derivative with respect to the second argument, in the direction θ, we have

$$\langle D_2 p(x,\beta),\theta\rangle = \int_{-\infty}^{+\infty} w(l)\theta\cdot\nabla f(x+\beta l)ldl. \tag{G.9}$$

This leads to the second-order partial differential equation

$$\mathcal{D}p(x,\beta) = \sigma^2\langle D_1\langle D_1 p(x,\beta),\beta\rangle,\theta\rangle - \langle D_2 p(x,\beta),\theta\rangle = 0. \tag{G.10}$$

This is analogous to (G.3) where we have extended the differential equation to three dimensions using our notation.

G.2 Pointwise Inversion Leveraging Additional Detectors

We consider a measurement geometry in which the incident beam direction, α, is fixed. Let β_i represent the scattering direction associated with the i^{th} detector. Let x represent the intersection of all paths defined by the incident beam and collimated detectors. To simplify notation we define

$$c_{\alpha i} \coloneqq \alpha\cdot\beta_i, \quad c_{ij} \coloneqq \beta_i\cdot\beta_j. \tag{G.11}$$

We choose the log of the observed scattering intensity as our data function

$$g_i(x) = \ln f(x, c_{\alpha i}) - \int_0^\infty \mu(x + t\alpha(x))dt - \int_0^\infty \mu(x + t\beta_i(x))dt. \tag{G.12}$$

Here $\mu(x)$ represents the total attenuation (scattering plus absorption) and $f(x, c_{\alpha i})$ represents the scatter density function. To preserve the scatter density between detectors we fix $c_{\alpha i} = c_{\alpha j}$ for all i, j. In the subsequent expressions we omit the second argument in q without ambiguity.

It will be useful to define a unique orthonormal basis for each detector. For scatter direction β_j, we define ϕ_j and ψ_j which together form an orthonormal basis. We will use

$$c_{ij}^{\phi} :- \beta_i \cdot \phi_j, \tag{G.13}$$

where the superscript is always associated with the second subscript. Without loss of generality we can define ϕ_j in the plane containing β_j and α where $c_{\alpha j}^{\phi} > 0$. We can then define $\psi_j = \beta_j \times \phi_j$. When the collimated detectors lie in a plane orthogonal to α, they must be placed in a circular configuration to enforce $c_{\alpha i} = c_{\alpha j}$. In this description ψ_j lie in the detector plane, tangent to the ring, at the j^{th} detector.

Building on the previous notation we can expand the directional derivative

$$\nabla_i^{\beta} = c_{ij}^{\beta}\nabla_j^{\beta} + c_{ij}^{\phi}\nabla_j^{\phi} + c_{ij}^{\psi}\nabla_j^{\psi}. \tag{G.14}$$

In this expression ∇_i^{β} represents the directional derivative operator in the direction β_i.

Observing the change in the data function as the target scatter location is shifted along the incident beam, we have

$$\nabla_\alpha g_i(x) = \nabla_\alpha \ln f(x) + \mu(x) - \nabla_\alpha \int_0^\infty \mu(x + t\beta_i(x))dt \tag{G.15}$$

$$= \nabla_\alpha \ln f(x) + \left(1 + c_{\alpha i}^{\beta}\right)\mu(x) - \left(c_{\alpha i}^{\phi}\nabla_i^{\phi} + c_{\alpha i}^{\psi}\nabla_i^{\psi}\right)\int_0^\infty \mu(x + t\beta_i(x))dt \tag{G.16}$$

$$= \nabla_\alpha \ln f(x) + \left(1 + c_{\alpha i}^{\beta}\right)\mu(x) - c_{\alpha i}^{\phi}J_i^{\phi} - c_{\alpha i}^{\psi}J_i^{\psi}, \tag{G.17}$$

where

$$J_i^{\phi} = \nabla_i^{\phi} \int_0^{\infty} \mu(x + t\beta_i(x))dt \tag{G.18}$$

$$J_i^{\psi} = \nabla_i^{\psi} \int_0^{\infty} \mu(x + t\beta_i(x))dt, \tag{G.19}$$

which we will solve for in the following.

Contrasting the observations of separate detectors, we can remove the scatter density term. We define the differential data function as a difference of two observations sharing a common transmitter and scattering angle

$$g_{ij}(x) = \int_0^{\infty} \mu(x + t\beta_i(x))dt - \int_0^{\infty} \mu(x + t\beta_j(x))dt. \tag{G.20}$$

The directional derivative (G.20) yields a similar form,

$$\nabla_i g_{ij}(x) = -\mu(x) + c_{ij}^{\beta}\mu(x) - c_{ij}^{\phi}\nabla_j^{\phi} \int_0^{\infty} \mu(x + t\beta_i(x))dt - c_{ij}^{\psi}\nabla_j^{\psi} \int_0^{\infty} \mu(x + t\beta_j(x))dt \tag{G.21}$$

$$= (-1 + c_{ij}^{\beta})\mu(x) - c_{ij}^{\phi} J_j^{\phi}(x) - c_{ij}^{\psi} J_j^{\psi}(x). \tag{G.22}$$

We can collect multiple differential measurements as a system of equations

$$\begin{bmatrix} \nabla_2 g_{21}(x) \\ \nabla_3 g_{31}(x) \\ \nabla_4 g_{41}(x) \end{bmatrix} = \begin{bmatrix} c_{21}^{\beta} - 1 & -c_{21}^{\phi} & -c_{21}^{\psi} \\ c_{31}^{\beta} - 1 & -c_{31}^{\phi} & -c_{31}^{\psi} \\ c_{41}^{\beta} - 1 & -c_{41}^{\phi} & -c_{41}^{\psi} \end{bmatrix} \begin{bmatrix} \mu(x) \\ J_1^{\phi}(x) \\ J_1^{\psi}(x) \end{bmatrix}. \tag{G.23}$$

When the first matrix on the right-hand side is nonsingular, we can solve for $\mu(x)$. Solving for $\mu(x), J_1^{\phi}(x), J_1^{\psi}(x)$, the results can be plugged into (G.17) to obtain $\nabla_{\alpha} \ln f(x)$.

The solutions to $J_1^{\phi}(x)$ and $J_1^{\psi}(x)$ must satisfy

$$\nabla_1 J_1^{\phi} = -\nabla_1^{\phi}\mu(x) \tag{G.24}$$

$$\nabla_1 J_1^{\psi} = -\nabla_1^{\psi}\mu(x). \tag{G.25}$$

This is equivalent to (2.12) in [3] for flat detectors in three dimensions. Using (G.23) to solve for $\mu(x), J_1^{\phi}(x), J_1^{\psi}(x)$ and plugging the results back into (G.25) extends the range condition of Katsevich and Krylov [3].

G.3 Infeasibility of Pointwise Inversion for TOF PET

Local inversion methods for the broken ray transform were developed in the context of single-scatter optical tomography. In that context the incident and scatter angles were not colinear. Here we demonstrate that the TOF PET data does not satisfy the local inversion criteria of the BRT.

Local inversion methods were initially developed by Katsevich and Krylov [3]. Subsequently Zhao et al. provided additional mathematical formalism including an inversion criterion [6]. This provides a more natural starting point for the TOF PET extension.

If we could measure a single cone beam transform $(B\mu)(x,\alpha)$, then a simple, local, reconstruction method is provided through a first-order, directional derivative

$$-\langle D_1(B\mu)(x,\beta),\beta\rangle = \mu(x). \tag{G.26}$$

We consider a notional data function superimposing multiple cone beam transforms sharing a common origin

$$\Phi(x) = \sum_j w_j \beta_j (B\mu)(x,\beta). \tag{G.27}$$

In this expression we have introduced scalar-valued w_k which are assumed known weighting coefficients. Once $\Phi(x)$ is available, inversion is achieved through the divergence

$$\mu(x) = -\frac{1}{\sum_j w_j}\nabla\cdot\Phi(x). \tag{G.28}$$

When scatter density varies spatially, an approach to obtaining $\Phi(x)$ is not obvious. A major contribution by Zhao et al. was providing criteria for obtaining this data function [6]. Zhao's

approach began with expanding $\Phi(x)$ as a combination of the measured data

$$\Phi(x) = \sum_j \sum_k \mathbf{c}_{jk} l(x, -\beta_j, \beta_k) \tag{G.29}$$

$$= \sum_j \sum_k \mathbf{c}_{jk} \left(\ln f\left(x\right) - (B\mu)(x, \beta_j) - (B\mu)(x, \beta_k) \right). \tag{G.30}$$

Here we ignore momentum transfer and omit the E term in our measured data function. The vector-valued $\mathbf{c}_{jk}$ must be selected such that (G.30) equals (G.27) while also taking into account the available data. This form suggests a transmitter and receiver are colocated at every direction β_j. The following conditions on $\mathbf{c}$ are due to Zhao:

(i) $\sum_{jk} \mathbf{c}_{jk} = 0$

(ii) $\mathbf{c}_{kk} = 0$

(iii) $\mathbf{c}_{jk} = \mathbf{c}_{kj}$

(iv) $\mathbf{s}_k = \sum_j \mathbf{c}_{jk} = w_k \beta_k$.

Enforcing (ii) alleviates the need for backscatter measurements. Since $l(x, -\beta_j, \beta_k) = l(x, -\beta_k, \beta_j)$, it is sufficient to enforce (iii). Enforcing (i) cancels the $f(x)$ term in $\Phi(x)$. Combining (i) with (iv) ensures G.30 and (G.27) are consistent.

In general there remain many solutions to $\mathbf{c}$ that satisfy the conditions (i)-(iv). This accommodates considerable flexibility in measurement systems such as the number of unique incident and scatter angles. The general solution for arbitrary dimension, satisfying (i)-(iv), is to select $\mathbf{c}$ according to

$$\mathbf{c}_{jk} = \left(w_j \beta_j + w_k \beta_k\right) \left(1 - \delta(j - k)\right) \tag{G.31}$$

such that

$$\sum_k w_k \beta_k = 0. \tag{G.32}$$

For 3 unique scatter angles, we can represent $\mathbf{c}$ with the matrix

$$\begin{array}{ccc|c} 0 & w_1\beta_1 + w_2\beta_2 & w_1\beta_1 + w_3\beta_3 & w_1\beta_1 \\ w_2\beta_2 + w_1\beta_1 & 0 & w_2\beta_2 + w_3\beta_3 & w_2\beta_2 \\ w_3\beta_3 + w_1\beta_1 & w_3\beta_3 + w_2\beta_2 & 0 & w_3\beta_3 \\ \hline w_1\beta_1 & w_2\beta_2 & w_3\beta_3 & 0 \end{array} \tag{G.33}$$

Here the final row and final column represent the column and row sums respectively.

Applying this formulation to TOF PET data requires additional constraints. While we can consider multiple rays through x, the incident and scatter angles must be colinear. In contrast to the traditional BRT, this limits the available data to few combinations of transmitter-receiver pairs. Let us index the scatter angles such that

$$\beta_{2j-1} = -\beta_{2j}. \tag{G.34}$$

This implies $\mathbf{c}$ is zero except for the first off-diagonal row. Since β_{2j-1} and β_{2j} are colinear, we can set $w_j = 0$ for j-odd. This limits $\mathbf{c}$ to

$$\mathbf{c}_{jk} = \begin{cases} w_k\beta_k, & j \text{ odd}, j = k-1 \\ 0, & \text{otherwise}, \end{cases} \tag{G.35}$$

where w_k must be selected such that

$$\sum_j w_{2j}\beta_{2j} = 0. \tag{G.36}$$

The expression (G.36) implies a minimum of three coincident detector pairs are required which represent 6 unique directions. We can represent this as a matrix

$$\begin{array}{cccccc|c} 0 & w_2\beta_2 & 0 & 0 & 0 & 0 & w_2\beta_2 \\ w_2\beta_2 & 0 & 0 & 0 & 0 & 0 & w_2\beta_2 \\ 0 & 0 & 0 & w_4\beta_4 & 0 & 0 & w_4\beta_4 \\ 0 & 0 & w_4\beta_4 & 0 & 0 & 0 & w_4\beta_4 \\ 0 & 0 & 0 & 0 & 0 & w_6\beta_6 & w_6\beta_6 \\ 0 & 0 & 0 & 0 & w_6\beta_6 & 0 & w_6\beta_6 \\ \hline w_2\beta_2 & w_2\beta_2 & w_4\beta_4 & w_4\beta_4 & w_6\beta_6 & w_6\beta_6 & 0 \end{array} \tag{G.37}$$

Plugging (G.35) into (G.30), we have

$$\Phi(x) = \sum_{j=1}^{K/2} w_{2j}\beta_{2j} \left(-(B\mu)(x, \beta_{2j-1}) - (B\mu)(x, \beta_{2j})\right) \tag{G.38}$$

$$= \sum_{j=1}^{K/2} w_{2j} \left(\beta_{2j-1}(B\mu)(x, \beta_{2j-1}) - \beta_{2j}(B\mu)(x, \beta_{2j})\right) \tag{G.39}$$

$$= \sum_{j=1}^{K} \bar{w}_j \beta_j (B\mu)(x, \beta_j). \tag{G.40}$$

For the last expression we have defined

$$\bar{w}_j \coloneqq \begin{cases} w_{j+1} & j \text{ odd} \\ -w_j, & \text{otherwise,} \end{cases} \tag{G.41}$$

in which case we have

$$\sum_j \bar{w}_j = 0, \tag{G.42}$$

which is incompatible with (G.28). Therefore, this inversion approach is not applicable for TOF PET.

There is an intuitive explanation for this as well. The break-point in the BRT is analogous to the TOF coordinate in TOF PET. The measured data in TOF PET comprises both

local scattering and a line integral of the attenuation map through the complete data. This integral simply does not change with the TOF coordinate.

G.4 Rotational BRT

Rotational measurement geometries for the BRT have received considerable attention recently [10, 21]. This is largely motivated by applications of SPECT with Compton cameras.

The rotational measurement geometry for BRT is considerably different from translational measurement geometries used in single-scatter optical tomography and measurement geometries associated with the Radon transform. In particular, the source is always directed through the origin. Accordingly, a separate coordinate system and inversion formulas have been developed.

The Fourier slice theorem is a well-known equality between the one-dimensional Fourier transform of image projections (e.g. Radon transform) and a slice through the two-dimensional Fourer transform of the image. Here we derive an analog for projections from the rotational BRT. The BRT for rotational measurement geometries has received considerable attention in the literature [9, 10, 23, 41]. Our contribution is unique in that we maintain consistency with the previously defined coordinate system.

In previous sections we defined the incident beam and scatter directions $\alpha, \beta \in S^2$. Additionally we defined a vector, θ, orthogonal to β but within the span of two vectors $\{\alpha, \beta\}$. We emphasize α, β, θ are coplanar.

In this section we limit our imaging system to a 2D plane. In this context it is convenient to consider $x \in \mathbb{R}^2$, and $\alpha, \beta, \theta \in S^1$. This is otherwise consistent with previous definitions, as β, θ remain orthonormal.

For S^1, we only require a scalar angle to specify elements of the set. It is convenient to define the scatter direction as a rotation from the incident direction. For this we introduce

the vector-valued function $\hat{u}(a) : \mathbb{R} \to S^1$

$$\hat{u}(a) \coloneqq \begin{bmatrix} \cos a \\ \sin a \end{bmatrix}. \tag{G.43}$$

The incident direction can be parameterized by $a \in [-\pi, \pi)$ such that

$$\alpha = \hat{u}(a). \tag{G.44}$$

We also define the rotation matrix as a function $R(a) : \mathbb{R} \to SO(2)$

$$R(a) \coloneqq \begin{bmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{bmatrix}. \tag{G.45}$$

The scatter direction can be specified as

$$\begin{aligned} \beta &= R(b)\alpha && \text{(G.46)} \\ &= \hat{u}(b + a). && \text{(G.47)} \end{aligned}$$

For $b \in [0, \pi/2)$, we additionally define

$$\begin{aligned} \theta &= R\left(\pi/2\right)\beta && \text{(G.48)} \\ &= \hat{u}\left(b + a + \pi/2\right), && \text{(G.49)} \end{aligned}$$

which is entirely consistent with previous definitions. Negative values of b would require a rotation of $-\pi/2$.

We consider a rotational measurement system in which the source is always directed through the origin. A collimated detector is focused at a point along the incident beam. This point can be specified using the radial distance t from the origin, along α, toward the source. The scatter location is uniquely defined using

$$x = -\alpha t. \tag{G.50}$$

We assume multiple detector locations are available to measure scatter in the direction β at varying distances $t \geq 0$ along the incident path for all $a \in [0, 2\pi)$. We define the 2D

rotational BRT $g(t,a) : \mathbb{R}^+ \times [-\pi, \pi) \to \mathbb{R}$

$$g(t,a) = g_\alpha(t,a) + g_\beta(t,a), \tag{G.51}$$

where

$$g_\alpha(t,a) = (B\mu)\left(-t\hat{u}(a), -\hat{u}(a)\right) \tag{G.52}$$

$$g_\beta(t,a) = (B\mu)\left(-t\hat{u}(a), \hat{u}(a+b)\right). \tag{G.53}$$

The conebeam transform in (G.52) is simplified since the incident path pases through the origin, so

$$(Bf)(-t\alpha, -\alpha) = \int_{-\infty}^{-t} f(r\alpha)dr. \tag{G.54}$$

We can expand

$$g_\alpha(t,a) = \int_t^{\infty} f(r\hat{u}(a+\pi))dr. \tag{G.55}$$

In this section we use the radial frequency for the Fourier transform. The Fourier transform of (G.55) with respect to a is

$$G_\alpha(t,w) = \int_{-\pi}^{\pi} g_\alpha(t,a)e^{-jwa}da \tag{G.56}$$

$$= \int_{-\pi}^{\pi}\int_{-\infty}^{-t} \mu(r\hat{u}(a))e^{-jwa}drda \tag{G.57}$$

$$= \int_t^{\infty} \hat{\mu}_a(r,w)e^{-jw\pi}dr. \tag{G.58}$$

In the final expression we use

$$\hat{\mu}_a(r,w) \coloneqq \int_{-\pi}^{\pi} \mu(r\hat{u}_a)e^{-jwa}da, \tag{G.59}$$

which represents a Fourier transform along a circular path in the image centered about the origin.

We would like to express (G.51) in terms of (G.59) as well. However, expanding the line integral in polar coordinates is challenging as the radial index is not unique. Consider the

point $y \in \mathbb{R}^3$

$$y \coloneqq t\sin(b)\theta. \tag{G.60}$$

This represents the point along the scatter path closest to the origin. In other words, the radial coordinate is minimized at this point. We can expand (G.51) as the line integral

$$g_\beta(t,a) = (B\mu)(-t\alpha,\beta) \tag{G.61}$$

$$= \int_{-t\cos b}^{\infty} \mu(t\sin(b)\theta + l\beta)dl, \tag{G.62}$$

where we have made use of the orthonormal basis β, θ. We make a change of variables replacing l with an angular measure v, defined implicitly as

$$\tan v = -l/t\sin b \tag{G.63}$$

$$dl = -t\sin b\sec^2 v dv. \tag{G.64}$$

The radial distance as a function of v is $t\sin b\sec v$. Putting this together we have

$$g_\beta(t,a) = \int_{-\pi/2}^{\pi/2-b} t\sin b\sec^2 v\,\mu\left(t\sin b\sec v\,\hat{u}\left(v+a+b+\pi/2\right)\right)dv. \tag{G.65}$$

The Fourier transform with respect to a is

$$G_\beta(t,w) = \int_{-\pi}^{\pi} g_\beta(t,a)e^{-jwa}da \tag{G.66}$$

$$= \int_{-\pi/2}^{\pi/2-b} t\sin b\sec^2 v\,\hat{\mu}_a\left(t\sin b\sec v, w\right)e^{iw(v+b+\pi/2)}dv. \tag{G.67}$$

Therefore we can decompose the Fourier transform of the BRT for rotational geometries

$$G(t,w) = \int_0^{2\pi} g(t,a)e^{-jwa}da \tag{G.68}$$

$$= G_\alpha(t,w) + G_\beta(t,w), \tag{G.69}$$

where $g(t,a)$ is the brokenray transform according to (G.51). The remaining terms $G_\alpha(t,w)$, $G_\beta(t,w)$ are given by (G.58), (G.67) respectively. Notice that the Fourier transform is taken

with respect to a single dimension in both the image and data. In this way, the result is analogous to the Fourier slice theorem for the Radon transform.

Vita

Michael R. Walker II

Degrees

D.Sc., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2020
M.S., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2011
B.S., Electrical Engineering, Cedarville University, Cedarville, Ohio, May 2004

Professional Experience

Stereotaxis, Saint Louis, MO 2016 - current
Research Scientist - Clinical Engineering & Advanced Concepts

The Boeing Company, Saint Louis, MO 2012 - 2016
Electrophysics Engineer - Phantom Works

Aclara Power-Line Systems, Inc., Saint Louis, MO 2004 - 2012
Senior Communication System Engineer - Advanced Systems R&D

Publications

Journal Publications:

M. R. Walker II, and J. A. O'Sullivan *"The broken ray transform: additional properties and new inversion formula,"* in *Inverse Problems,* vol. 35, no.11, October 3, 2019.

D. W. Rieken and **M. R. Walker II** *"Ultra Low Frequency Power-Line Communications Using a Resonator Circuit,"* in *IEEE Transactions on Smart Grid,* vol. 2, no. 1, pp. 41-50, March 2011.

Invited Presentations:

M. R. Walker II, "The Broken Ray Transform: From Linear Systems Analysis to Iterative Algorithms," *Inverse Problems: Modeling and Simulation (IPMS)*, Malta, May 24–30, 2020 (Postponed to May 2021).

Conference Publications:

M. R. Walker II, "Local Parametric Surface Approximation With Automatic Order Selection From Position Data," to appear in *Medical Robotics, 2020 IEEE Int. Symposium on*, April 2020.

D. W. Rieken and **M. R. Walker II** *"Distance Effects in Low-Frequency Power Line Communications,"* in *Power Line Comm. and Its Applications, 2010 IEEE Int. Symposium on,* 2010, pp. 22-27.

May 2020