MODEL-BASED ITERATIVE RECONSTRUCTION FOR MICRO-SCALE AND NANO-SCALE IMAGING

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Singanallur V. Venkatakrishnan

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2014

Purdue University

West Lafayette, Indiana

This dissertation is dedicated to my father, S. V. Vaidyanathan (1952-2013).

ACKNOWLEDGMENTS

I thank my advisor, Prof. Charles Bouman, without whose inputs and support this dissertation would not have happened. I believe that his high standards for the quality of work, emphasis on thinking ideas through and the pursuit of perfection have made me a better engineer and moulded my views on what it means to do research. His incredible enthusiasm for research is infectious and the energy with which he pursues it is an inspiration to me. He has also given me the opportunity to teach classes, mentor students, and present my work at different forums which has helped me immensely. Finally, I have always enjoyed all our "non-research" conversations whether at work or during research related travel and it has made being part of the lab a lot more fun!

Thanks to Dr. Jeff Simmons and Dr. Larry Drummy from the Air-Force Research Lab for supporting my research and working closely with me throughout the course of my PhD. It was a far-sighted initiative from them to bring together researchers from the material sciences and signal processing to work on some cool problems; and I feel fortunate to be a part of that team. I also acknowledge Larry's part in educating me about electron microscopy and giving me the opportunity to observe one in use to peer into the word at a scale I had never seen before.

I thank Mike Jackson for the collaboration on turning the electron-tomography project into a usable GUI based software. Mike has been incredibly patient in answering my questions and helping me with various software related aspects of the work. Special thanks to Prof. Marc De Graef at CMU for his inputs on the physics of electron microscopy and for his prompt and insightful comments on all our submitted manuscripts and papers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Venkatakrishnan, Singanallur V. Ph.D., Purdue University, August 2014. Model-Based Iterative Reconstruction For Micro-scale and Nano-scale Imaging. Major Professor: Charles A. Bouman.

Transmission electron microscopes (TEM) and synchrotron X-ray (SX) sources are widely being used to characterize materials at the nano-scale and micron-scale in two/three dimensions. While there has been significant progress in enhancing the hardware in these instruments to improve image quality, the algorithms used for image reconstruction have not fully exploited the statistical information in the data and the properties of the material being imaged to enhance the quality of the images. Model-based iterative reconstruction (MBIR) is an emerging theme for image reconstruction that combines a probabilistic model for the measurement system (forward model) with a probabilistic model for the image (prior model) to formulate the reconstruction as a high-dimensional estimation problem. In this dissertation, we propose MBIR algorithms for different imaging modalities used in a TEM and in SX imaging.

First, we propose an MBIR algorithm for high angle annular dark field - scanning TEM (HAADF-STEM) tomography. Next, we present an MBIR algorithm for handling anomalous measurements encountered in bright field - electron tomography (BF-ET) of crystalline samples. Results on simulated as well as real data show significant improvements over the typical reconstruction approaches used for HAADF-STEM tomography and BF-ET. Furthermore, the proposed MBIR for BF-ET is also useful for SX tomography as it can handle anomalous measurements from saturated detector pixels.

Finally, we propose a flexible optimization framework, termed Plug-and-Play priors, that allows state-of-the-art forward models of imaging systems to be matched

with state-of-the-art denoising algorithms for MBIR. We will demonstrate how the Plug-and-Play priors can be used to mix and match a wide variety of denoising algorithms based on advanced image models with forward models encountered in TEM tomography, SX tomography and in sparse image reconstruction from STEM data, thus greatly expanding the range of possible problem solutions.

1. INTRODUCTION

The past decade has witnessed an unprecedented interest in the development of new materials to solve key problems in areas ranging from energy to medicine. The recent announcement of the Materials Genome Initiative [1] for advanced manufacturing further emphasizes the interest in the rapid development of novel materials for various applications. A key component in the development of advanced materials is to have the ability to quantitatively characterize samples at the nano-meter scale and micron-scale. Transmission electron microscopy (TEM) is a very popular imaging method used in the physical and biological sciences for characterizing samples at the nano-meter scale in 2D and 3D [2, 3]. Similarly, synchrotron X-ray (SX) tomography is widely used for characterizing materials at the micron scale in 3D [4, 5]. While there has been significant progress in enhancing the hardware in these instruments to improve image quality, the algorithms used for image reconstruction have not fully exploited the statistical information in the data and the properties of the material being imaged to enhance the quality of the images. It is evident that the design of advanced algorithms that exploit knowledge about the physics of acquisition, noise characteristics of the detectors as well as knowledge about the materials being imaged, can help improve the quality of reconstructions for the data acquired or alternately reduce the amount of data acquired for a fixed reconstruction quality. Thus, the design of novel imaging algorithms can play a key role in the characterization of materials at the micro-meter and nano-meter length scale.

Model-based iterative reconstruction (MBIR) is a systematic approach for solving problems in image reconstruction. MBIR approaches have dramatically helped to improve image quality in various applications including denoising [6, 7], medical X-ray CT [8, 9], positron emission tomography [10], dual energy CT [11], optical diffusion tomography [12, 13], and MRI [14, 15]. Model-based approaches to imaging combine

the physics of data formation with a noise model for the sensors and a probabilistic model for the underlying image to formulate a posterior probability distribution for the image given the measurements. Using this description we can reconstruct the unknown image in several ways, the most popular being maximum *a posteriori* probability (MAP) estimation. A key benefit of MBIR approaches are that we can include the estimation of missing calibration parameters associated with the measurement as a part of the reconstruction. A common method of formulating the reconstruction problem in the presence of unknown calibration parameters is [16]

$$(\hat{f}, \hat{\phi}) \leftarrow \underset{f, \phi}{\operatorname{argmin}} \left\{ -\log p(g|f, \phi) - \log p(f) \right\}$$

where g is the data, f is the image/volume to be reconstructed, ϕ refers to unknown calibration parameters, $p(g|f, \phi)$ is the probability density function (pdf) of the measurements given the unknowns and $p(f)$ is the pdf for the image/volume. The main challenges in MBIR are the *formulation of tractable models* and the subsequent *development of convergent optimization algorithms for very large data sets*.

In this dissertation, we propose MBIR algorithms for imaging modalities used in a TEM and in SX tomography (SXT). First, we will present MBIR algorithms for two modalities used in TEM tomography. Finally, we will present an algorithmic framework which can enable sophisticated prior models of images to be incorporated into MBIR in a simple manner. We will apply this framework to tomography and to the reconstruction of an image from a sparse set of measured pixels.

1.1 MBIR for Electron Tomography

When high energy electrons interact with a sample a variety of signals are produced such as the direct transmission, elastic scatter, inelastic scatter, back-scatter electrons and secondary electrons (see Fig. 1.1). Each of these signals can be used for imaging a sample and reveal different properties of the material. For example, a scanning electron microscope (SEM) can collect secondary electrons and back-scattered
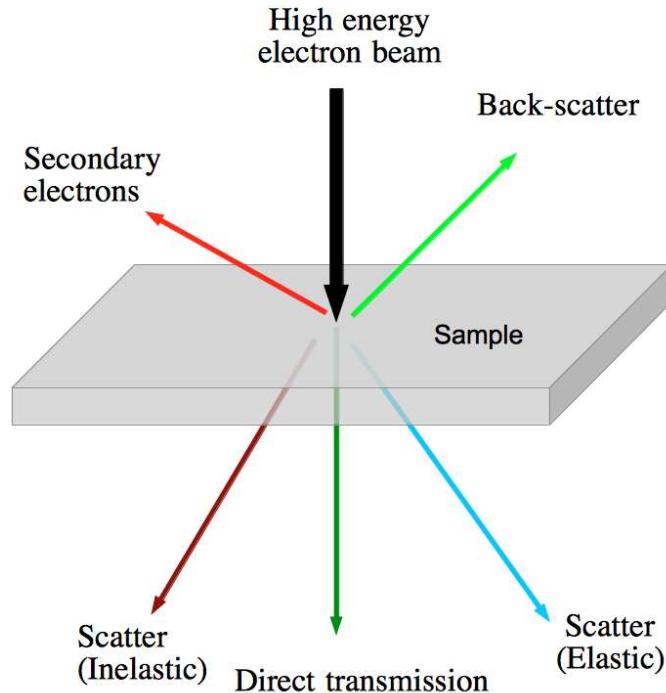
Fig. 1.1.: Some signals generated when high energy electrons interact with a sample. These signals can be analyzed using an electron microscope. A scanning electron microscope forms images of secondary electrons and back-scattered electrons while a transmission electron microscope can form images of the direct beam and the forward scattered electrons.

electrons (BSE) and is used for 2-D imaging of the surface of a sample [17]. A transmission electron microscope (TEM), on the the other hand, can collect electrons that are transmitted through the sample and can therefore be used to probe depth information in addition to the 2-D surface.

TEMs are widely used for 2-D and 3-D characterization of materials at the nanometer scale in the physical and biological sciences. The two dominant 2-D imaging modalities in a TEM/scanning-TEM (STEM) are - bright field (BF) and high angle annular dark field (HAADF). BF images are formed by collecting electrons transmitted through the sample (or scattered through small angles), while in the HAADF case, electrons scattered by the sample through high angles are collected using an
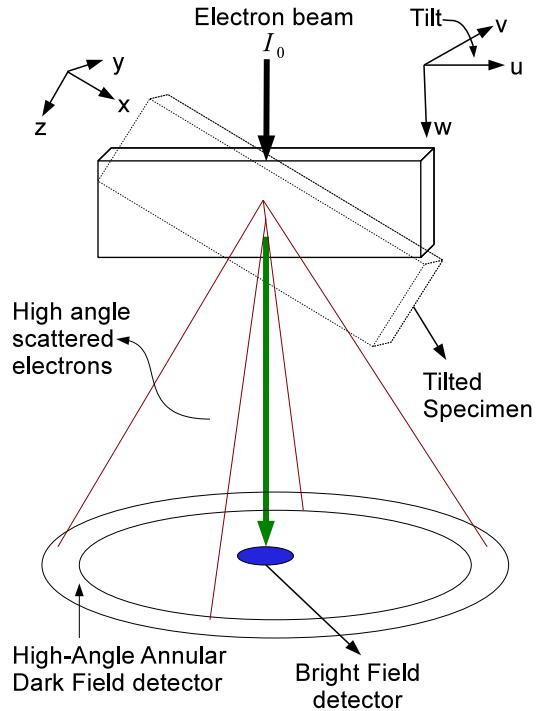
Fig. 1.2.: Illustration of a typical set-up for STEM tomography. The electron beam is rastered across the sample to form a single image. The sample is then tilted and the process is repeated. The two dominant modalities for STEM tomography are HAADF and BF.

annular detector (see Fig. 1.2). Fig. 1.3 shows a BF and HAADF image acquired of the same sample. Since the BF detector collects electrons transmitted (or scattered through very small angles), the images are "bright" in the absence of any material. In contrast, the HAADF detector captures electrons scattered by the sample through high angles (in the range of 50-250 mrad) and hence the image is "dark" in regions where there is no material present. 3-D characterization of samples in a (S)TEM is carried out using electron tomography (ET). For ET, the sample is tilted and imaged in a limited angular range (see Fig. 1.2) to acquire a "tilt-series" data set to which an algorithm is applied reconstruct the object.

In the first part of this dissertation, we propose an MBIR algorithm for HAADF-STEM tomography. HAADF-STEM data is increasingly being used in the physical

BF image   HAADF image
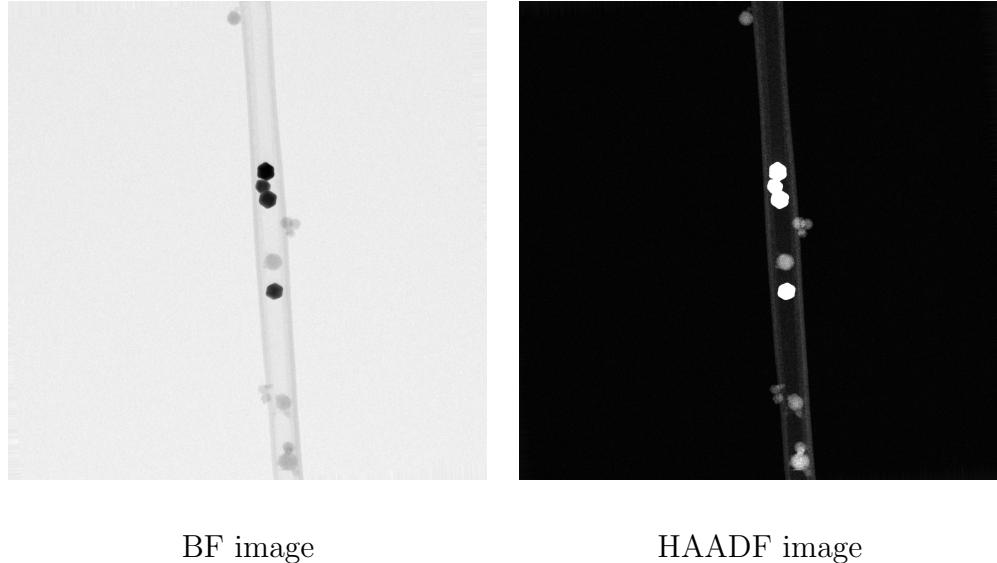
Fig. 1.3.: Example of BF and HAADF-STEM image acquired from a sample containing gold and aluminum nano-particles in a carbon support. HAADF images are characterized by low counts in regions where there is no sample present while BF images are characterized by high counts in region where there is no sample present.

sciences because it reduces the effects of Bragg diffraction [18] seen in BF TEM data. Typically, tomographic reconstructions are performed by directly applying either filtered back projection (FBP) or the simultaneous iterative reconstruction technique (SIRT) to the data [19–22]. Since HAADF-STEM tomography data typically has a low signal-to-noise ratio, a limited set of views and certain missing calibration parameters associated with the measurement, methods such as FBP and SIRT can result in significant artifacts in the reconstructed volume. We propose a novel likelihood function, $p(g|f, \phi)$, for HAADF-STEM imaging that models the data formation process while accounting for calibration parameters such as detector gains, offsets and noise variance. We then combine it with a model for the object to formulate the MBIR cost function. Since there are missing calibration parameters, we include their estimation as a part of the reconstruction. Minimizing the MBIR cost function is computationally expensive; hence we propose a fast, parallel, multi-resolution method that makes the algorithm usable in practice.

Next, we propose an MBIR algorithm for BF-ET. BF data has generally been avoided [2, 23] for tomography in the physical sciences because of sudden changes in intensity in certain regions of the acquired images due to Bragg diffraction from crystalline samples [24], leading to a measurement uncharacteristic of attenuation due to thickness alone. As a result of these anomalous measurements, methods such as FBP or SIRT fail to reconstruct the samples effectively. We propose a new likelihood function, $p(g|f, \phi)$, for the data using a pdf based on a generalized Huber function [25] which models the Bragg scatter events as outliers. One challenge with using the generalized Huber function for MBIR is that the resulting cost function may not be differentiable and hence is difficult to minimize using standard gradient-based techniques. Hence, we propose an iterative optimization strategy based on majorization-minimization [26, 27] that results in simple closed form updates while monotonically decreasing the cost. Finally, we note that the forward model for SXT shares similarities with the case of BF-ET; hence the proposed algorithm has also been shown to be useful for SXT [28], where there are anomalous measurements due to saturated detector pixels.

1.2 Towards Advanced Prior Models for MBIR

Most MBIR approaches for applications with complex forward models as in tomography, use conventional Markov random field (MRF) based prior models, $p(f)$, for the underlying image to be reconstructed [8, 29–35]. These models result in tractable cost functions and give reasonable results in several applications. However, these models result in some artifacts and the question of which prior models best describe images is an open problem. Recent advances in image denoising based on sophisticated patch-based non-local image models (for example [6, 36, 37]), have shown that it is possible to dramatically improve upon what was previously possible. However incorporating these denoising algorithms into problems like tomography is challeng-

ing because it can result in cost functions that are difficult to optimize; or in some cases the denoising methods do not correspond to an explicit prior model.

In the last part of this dissertation, we propose a flexible framework for MBIR, termed Plug-and-Play priors, that allows state-of-the-art forward models of imaging systems to be matched with state-of-the-art priors via denoising algorithms. We will demonstrate how the Plug-and-Play priors can be used to mix and match a wide variety of existing denoising algorithms with forward models encountered in TEM tomography, SXT and in sparse image reconstruction of STEM data, thus greatly expanding the range of possible problem solutions.

1.3   Organization of Dissertation

The organization of the rest of this dissertation is as follows. In Chapter 2 we present an MBIR algorithm for HAADF-STEM tomography. In Chapter 3 we present an MBIR algorithm that can handle anomalous data in BF-ET. Finally, in Chapter 4 we present details of the Plug-and-Play priors framework for combining advanced prior models with forward models in tomography and image reconstruction from a sparse set of measurements.

2. MODEL-BASED ITERATIVE RECONSTRUCTION FOR HIGH-ANGLE ANNULAR DARK FIELD - SCANNING TRANSMISSION ELECTRON MICROSCOPE (HAADF-STEM) TOMOGRAPHY

2.1 Introduction

Electron tomography has been widely used in the life sciences to study biological structures in 3D [38]. In the life sciences, electron tomography is dominated by Bright Field (BF) tomography which is a transmission modality. However, direct extension of the BF technique to the material sciences has proved challenging because of diffraction effects in some regions of the acquired projection images (sinogram) caused by crystalline samples [2]. This has led to a growing interest in the use of HAADF-STEM tomography to study materials in 3D (e.g. [19–22, 39, 40]). This modality, which is based on measuring electrons scattered by the material, is relatively free from diffraction effects observed in BF TEM data. The HAADF-STEM signal contains information about the atomic number of the region being imaged and hence is useful for extracting chemical information about the material. As a result HAADF-STEM tomography is also referred to as Z-contrast tomography [2].

A typical HAADF-STEM acquisition involves focusing an electron probe at a point on the material for a short duration and measuring the electrons scattered into an annular detector as shown in Fig. 2.1(a). The electron beam is raster scanned and at each point a measurement is made to obtain a projection image of the object. The object is then tilted along a single axis and the process is repeated. At the end of the acquisition, a set of projection images is obtained corresponding to each tilt of the object. In most cases, due to mechanical constraints the object can only be tilted in the range of approximately $\pm 70°$. Therefore, HAADF-STEM tomography

can be classified as a parallel beam, limited angle tomography modality. More details of HAADF-STEM acquisition and pre-processing can be found in [2, 41].

Most approaches (e.g. [19–22]) for HAADF-STEM tomography have directly used the measured HAADF data for tomographic inversion using FBP or SIRT [42]. The reconstruction quality for these algorithms is limited by the fact that they do not incorporate noise statistics of the measurement, and do not incorporate a prior model. Furthermore, algorithms such as SIRT require the selection of an ad-hoc stopping criterion; otherwise the reconstructed image will typically diverge or become excessively noisy. Recently, there have been efforts on using compressed sensing (i.e. total variation prior model) based methods [39, 40] and discrete tomography [43] for HAADF-STEM tomography. However, these examples have not explicitly taken into account specific noise models and do not deal with the case when there are missing calibration parameters associated with the measurement.

Model Based Iterative Reconstruction (MBIR) algorithms have enabled significant qualitative and quantitative improvements in tomography applications like X-Ray CT [8, 29, 30, 32–34], positron emission tomography (PET) [10, 31, 35, 44], bright field electron tomography [45] , optical diffusion tomography (ODT) [12, 13], and atomic resolution tomography [46]. These methods typically require the specification of a model for the measurement system and its associated noise (i.e., a forward model); a model for the unknown image volume being reconstructed (i.e., a prior model); an estimation criteria for finding the reconstruction (typically the MAP estimate); and an algorithm to compute the estimate. However, computing the estimate requires solving a computationally demanding optimization problem, and typically an iterative approach is employed to solve it. A variety of iterative algorithms have been explored for tomographic reconstruction such as Expectation Maximization [47], Ordered Subsets (OS) [48, 49], Conjugate Gradient [50, 51], graph cuts [52] and ICD [53]. It has been shown that ICD has relatively faster convergence than other popular methods especially at higher spatial frequencies in the X-ray CT case [54] when the algorithms

were initialized with a FBP reconstruction. Furthermore positivity constraints can be easily incorporated in ICD compared to other gradient based methods.

In this chapter, we present a MBIR algorithm for quantitative HAADF-STEM tomography [55]. We begin by specifying a quantitative forward model for the HAADF-STEM acquisition system which depends on the 3D volume of scatter coefficients along with other parameters which can be either estimated or measured. Next, we incorporate knowledge of either sharp or diffuse interfaces between materials by using a q-Generalized Gaussian Markov Random Field (q-GGMRF) [56] prior for the scatter coefficients. If the interfaces between materials are sharp, then the parameters of the model can be adjusted to produce an edge preserving reconstruction; and if they are diffuse, then the parameters can be adjusted to produce more smooth boundaries. Finally, we formulate the reconstruction as a maximum *a posteriori* probability (MAP) estimation problem and develop a fast multiresolution iterative algorithm to minimize the corresponding cost function.

Our algorithm to minimize the MAP cost function is based on the ICD [53] algorithm. In HAADF-STEM tomography, it is also necessary to estimate certain unknown calibration parameters, in addition to the actual 3D volume of scatter coefficients. In every iteration of the optimization, the voxels are updated, followed by the update of other unknown parameters to lower the value of the MAP cost function. To speed convergence of the algorithm we update voxels in a random order [8], use a substitute function approach and parallelize the updates. The convergence is further sped up by using a multi-resolution initialization for the parameters, which also eliminates the need to compute an analytical reconstruction as an initial condition. We evaluate our method by comparing the reconstructions with those produced by FBP and SIRT for a simulated as well as a real data set. Reconstructions of a simulated data set as well as a real data set show that MBIR produces superior reconstructions by suppressing artifacts and enhancing contrast.

The organization of this chapter is as follows. In Section 2.2 we discuss the model used for the measurement process in HAADF-STEM tomography. In Section 2.3
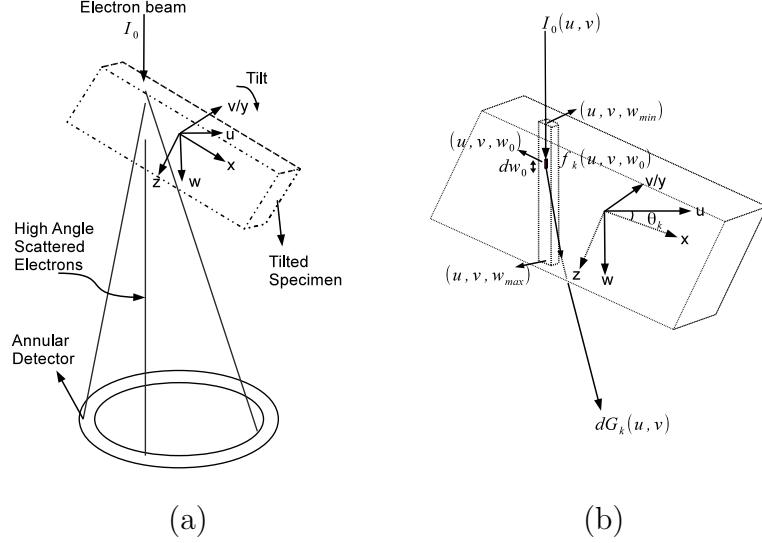
Fig. 2.1.: Illustration of the measurement process in a HAADF-STEM acquisition system. (a) A single measurement is obtained by focusing the electron beam I_0 on the material and measuring electrons scattered into the annular detector. (b) Equivalent model for the measurement process. The scattered beam from location w_0, $dG_k(u,v)$, which is deflected by angles in the range of $50-300$ mrad, is detected by the annular detector.

we present details of the q-GGMRF prior model. In Section 2.4 we use the image formation model along with the prior to formulate the MAP estimation problem and describe our algorithm to find the MAP estimate. In Section 2.5 we present results from a simulated data set and follow it up with results on a real data set. Finally in Section 2.6 we draw our conclusions.

2.2 Measurement Model

In HAADF-STEM tomography, a single measurement is obtained by focusing an electron beam at a point on the surface of the material of interest. As the beam propagates through the material, the electrons which are scattered through high angles, approximately in the range $50-300$ mrad, are detected by an annular detector (see Fig. 2.1). The number of electrons detected depends on the value of the HAADF

scatter coefficients along the region being probed which is related to the type and number of atoms per unit volume at each location, the beam energy used as well as the inner and outer angle of the detector. The total number of electrons detected by the annular detector is typically assumed to be proportional to the projected value of the scatter coefficients through the region being probed. The electron beam is raster scanned and at each point we get a single measurement. The measurements resulting from a single raster scan of the material constitute a single electron microscope image. The sample is then tilted around a single axis (y axis in Fig. 2.1) and the process is repeated. Thus we get 2-D parallel beam tomography data which we need to invert in order to reconstruct the HAADF scatter coefficients.

The goal of HAADF-STEM tomography is to reconstruct the HAADF scatter coefficients (units of nm^{-1}) denoted by $f(x, y, z)$ at every point in space. If (x, y, z) is the frame of reference of the object and (u, v, w) is the reference frame for the electron source (see Fig. 2.1), then any function of space can be reparameterized so that $f_k(u, v, w) = f([u, v, w]R_{\theta_k})$ where R_{θ_k} is an orthonormal rotation of the spatial coordinates by an angle θ_k. For tomographic reconstruction we require measurements of the projection integral $\int_{-\infty}^{+\infty} f_k(u, v, w)dw$ through the object for every tilt θ_k and every point (u, v). We begin by describing how this measurement can be obtained from the HAADF-STEM signal. Let $I_0(u, v)$ be the source electron flux in units of electrons per nm^2 corresponding to the beam at location (u, v). Under the assumptions that attenuation effects are minimal and the sample is relatively thin, the total HAADF scatter is given by [57]

$$G_k(u, v) \;\; = \;\; I_0(u, v) \int_{w_{min}}^{w_{max}} f_k(u, v, w)\, dw. \tag{2.1}$$

Thus we get

$$
\begin{aligned}
\frac{G_k(u, v)}{I_0(u, v)} \;\; &= \;\; \int_{w_{min}}^{w_{max}} f_k(u, v, w)dw \\
&= \;\; \int_{w_{min}}^{w_{max}} f\left([u, v, w]R_{\theta_k}\right) dw.
\end{aligned} \tag{2.2}
$$

So the normalized quantity $G_k(u,v)/I_0(u,v)$ is an estimate of the tomographic projection at angle θ_k and at position (u,v).

We note that the above model, though widely used for HAADF-STEM tomography, does not account for the attenuation effects observed during the imaging of thick specimens of heavy elements [58]. While Van den Broek at al. have addressed some of these non-linear effects [59], the topic of developing simple tractable models that describe all the experimentally observed characteristics of the HAADF-STEM signal including contrast reversal [60] remains an open problem. We use the linear model because it is still widely applicable for a range of materials and thicknesses studied in STEM tomography including all the cases considered in this chapter.

Next we model the process by which the detector converts the incident electron flux to a measured signal. The HAADF-STEM detector typically consists of a photomultiplier tube (PMT) which converts the detected electron flux to a current which is then converted to a voltage using a preamplifier and read out using an A/D converter [61]. Since the HAADF-STEM detector has a limited dynamic range, the gain associated with the PMT (referred to as contrast) and offset associated with the preamplifier [61, 62] (referred to as brightness) can be adjusted so a wide range of materials and thicknesses can be imaged. We denote the PMT gain associated with the measurements at tilt k by α_k and the additive offset associated with the measurements at tilt k by d_k. Therefore, we model the ith measurement (corresponding to the electron source at (u_i, v_i)) at tilt k by a Gaussian random variable $g_{k,i}$, with mean

$$\mathbb{E}[g_{k,i}] \;\; = \;\; \alpha_k \int \int G_k(u,v) h_i(u,v) \, du dv + d_k \qquad (2.3)$$

where $h_i(u,v)$ is a kernel which averages the electron flux over the area corresponding to the ith measurement. Let f be a discretized version of $f(x,y,z)$ organized as a $N \times 1$ vector, where N is the total number of voxels, A_k is a $M \times N$ projection matrix

for tilt k, and M is the number of measurements per tilt. Then substituting (2.1) in (2.3) we get

$$
\begin{aligned}
\mathbb{E}[g_{k,i}] &\approx I_k \int\int \left(\int_{w_{min}}^{w_{max}} f_k(u,v,w)dw \right) h_i(u,v)dudv + d_k \\
&= I_k \left[A_k f \right]_i + d_k
\end{aligned}
\tag{2.4}
$$

where $[A_k f]_i$ is the ith entry of the vector $A_k f$, $I_k = I_0 \alpha_k$ is the product of the source electron dose (counts) and the PMT gain at tilt k. For simplicity we will call I_k the gain associated with the measurements at tilt k. The variance of each measurement is modeled by

$$
\text{Var}[g_{k,i}] = \sigma_k^2 \mathbb{E}[g_{k,i}]
\tag{2.5}
$$

where σ_k^2 is a parameter used to model the noise variance at tilt k and $\mathbb{E}[g_{k,i}]$ accounts for the Poisson characteristics of the measurement. We assume that all the measurements are conditionally independent. If $\mathbb{E}[g_{k,i}] \approx g_{k,i}$, $\Lambda_k = \text{diag}\left(\frac{1}{g_{k,1}}, \cdots, \frac{1}{g_{k,M}} \right)$, $g_k = [g_{k,1}, \cdots, g_{k,M}]^t$, $g = [g_1^t \cdots g_K^t]^t$, $I = [I_1, \cdots, I_K]$, $d = [d_1, \cdots, d_K]$ and $\sigma^2 = [\sigma_1^2, \cdots \sigma_K^2]$ then using (2.4) and (2.5) we get

$$
\begin{aligned}
p(g|f, I, d, \sigma^2) &= \left(\prod_{k=1}^{K} \frac{1}{(2\pi\sigma_k^2)^{\frac{M}{2}} |\Lambda_k|^{-\frac{1}{2}}} \right) \\
&\exp\left\{ -\frac{1}{2} \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \|g_k - I_k A_k f - d_k \mathbb{1}\|_{\Lambda_k}^2 \right\}
\end{aligned}
\tag{2.6}
$$

where K is the total number of tilts.

2.3 Prior Model

We use a q-GGMRF [56] model for the probability density of f. The parameters of this model can be adjusted to account for sharp or diffuse interfaces between materials. Moreover since it has a bounded second derivative, the substitute function method of

[8] can be used, significantly speeding up the subsequent MAP estimation algorithm. If $\mathcal{N}$ is the set of all pairs of neighboring voxels (e.g. a 26 point neighborhood in 3D), w_{ij} is a weighting kernel which is inversely proportional to the distance between voxel i and voxel j, normalized so that $\sum\limits_{j \in \mathcal{N}_i} w_{ij} = 1$, $\mathcal{N}_i$ is the set of all neighbors of voxel i, then the density function corresponding to the q-GGMRF prior is given by

$$p(f) \;\; = \;\; \frac{1}{Z} \exp \left\{ - \sum_{\{i,j\} \in \mathcal{N}} w_{ij} \rho(f_i - f_j) \right\} \tag{2.7}$$

$$\rho(f_i - f_j) \;\; = \;\; \frac{\left| \frac{f_i - f_j}{\sigma_f} \right|^q}{c + \left| \frac{f_i - f_j}{\sigma_f} \right|^{q-p}}$$

where Z is a normalizing constant and p, q, c and σ_f are q-GGMRF parameters. Typically $1 \leq p \leq q \leq 2$ is used to ensure convexity of the function $\rho(.)$, thereby simplifying the subsequent MAP optimization. The value of σ_f is typically set to achieve a balance between resolution and noise in the reconstructions. In this chapter we fix $q = 2$. When $p = 1$ the prior model corresponds to strong edge preserving reconstructions while $p = 2$ corresponds to smooth reconstructions (bearing some similarities to the Cahn-Hilliard phase field model [63]). We note that when c is zero, the $p = 1$ case of the q-GGMRF corresponds to the total variation prior [29]. Thus the q-GGMRF provides a flexible prior model framework enabling us to model a range of possible materials from those with very sharp interfaces to those with smooth interfaces.

2.4   MAP Estimation And MBIR Algorithm

We use the MAP estimate to reconstruct the values of the HAADF scatter coefficients. Since the gains (I), offsets (d) and variance parameters (σ^2) are typically not

measured we treat them as nuisance parameters in the MAP estimation framework. The MAP estimate of the parameters is given by

$$(\hat{f}, \hat{I}, \hat{d}, \hat{\sigma}^2) = \underset{f \geq 0, I \in \Omega, d, \sigma^2}{\operatorname{argmin}} \left\{ -\log p(g|f, I, d, \sigma^2) - \log p(f) \right\}$$

where $\Omega = \left\{ I \in \mathbb{R}^K : \frac{1}{K} \sum_{k=1}^{K} I_k = \bar{I} \right\}$, $p(I, d, \sigma^2)$ is uniformly distributed, f is conditionally independent of (I, d, σ^2). We impose positivity constraints on the voxels $(f \geq 0)$ as it is physically meaningful to have positive values of the HAADF scatter coefficients. The constraint Ω forces the average value of the gains to be equal to a value $\bar{I}$ to prevent our algorithm from diverging to unreasonable values of the HAADF scatter coefficients. The choice of $\bar{I}$ is arbitrary but affects the scaling of f and hence the choice of the prior model parameter, σ_f. Hence if $\bar{I}$ is set to the product of detector gain and source electron dose, then the reconstructions will be quantitative.

Using (3.4) and (3.10) we obtain the MAP estimate by minimizing the cost

$$c(f, I, d, \sigma^2) = \frac{1}{2} \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \| g_k - I_k A_k f - d_k \mathbb{1} \|_{\Lambda_k}^2$$

$$+ \frac{1}{2} \sum_{k=1}^{K} \log \left((2\pi\sigma_k^2)^M |\Lambda_k|^{-1} \right) + \sum_{\{i,j\} \in \mathcal{N}} w_{ij} \rho(f_i - f_j) \tag{2.8}$$

In general the cost function $c(f, I, d, \sigma^2)$ is convex in f but not jointly convex in (f, I, d, σ^2). We adapt the ICD algorithm [30] to minimize the cost function (2.8). In ICD the parameters are updated one at a time such that each update results in a lower value of the cost function. The basic structure of our algorithm is to repeatedly perform the following steps until convergence is achieved.

(i) For each voxel j, $\hat{f}_j \leftarrow \underset{f_j \geq 0}{\operatorname{argmin}} c(f, I, d, \sigma^2)$

(ii) $(\hat{I}, \hat{d}) \leftarrow \underset{I \in \Omega, d}{\operatorname{argmin}} c(\hat{f}, I, d, \sigma^2)$

(iii) $\hat{\sigma}^2 \leftarrow \underset{\sigma^2}{\operatorname{argmin}} \, c(\hat{f}, \hat{I}, \hat{d}, \sigma^2)$

In Step (i) we update each voxel to lower the original cost function using a substitute function approach [8]. The form of the substitute function is chosen so that it results in simple closed form updates for the voxels, speeding up the implementation of our method. It has also been shown in [8] that the substitute function approach speeds up the overall convergence of the algorithm to the minimum. Minimizing the substitute function does not minimize the original cost but it results in voxel updates that lower the original cost. In Step (ii) we find the minimum of the cost function with respect to the gain I and offset parameters d by turning the constrained optimization to an unconstrained problem by using a Lagrange multiplier. Finally in (iii) we minimize the cost function with respect to σ^2. Thus each of the above updates lowers the value of the original cost function. The algorithm is terminated if the relative change in the total magnitude of the reconstruction is less than a preset threshold. Next, we derive the update equations for the three steps of our algorithm.

2.4.1 Voxel Update

We adopt a strategy similar to [8] based on a substitute function approach and random order update of the voxels to minimize the cost with respect to each voxel. Sets of voxels having the same (x, z) coordinate constitute a voxel line. Voxels in the same voxel line share geometry computation and therefore it is computationally efficient to update a single line of voxels together. Updating the voxel lines in random order and using a substitute function approach has also been shown to speed up the convergence of the ICD algorithm [8]. Further the voxel updates are implemented in parallel (on multicore machines) by allocating sets of contiguous $x - z$ slices to different cores similar to [64].

```
UpdateVoxel(j,e) {
```

$$\tilde{\theta}_2 \leftarrow \sum_{k=1}^{K} (I_k A_{k,*,j})^t \tilde{\Lambda}_k (I_k A_{k,*,j})$$

$$\tilde{\theta}_1 \leftarrow -\sum_{k=1}^{K} e_k^t \tilde{\Lambda}_k (I_k A_{k,*,j})$$

for each $i \in \mathcal{N}_j$ { Compute substitute function parameter a_{ji} using (3.11)}

$$u^* \leftarrow \frac{\sum_{i\in\mathcal{N}_j} w_{ji} a_{ji} f_i + \tilde{\theta}_2 \tilde{f}_j - \tilde{\theta}_1}{\sum_{i\in\mathcal{N}_j} w_{ji} a_{ji} + \tilde{\theta}_2}$$

$$f_j \leftarrow \max(u^*, 0)$$

$$e_k \leftarrow e_k - (f_j - \tilde{f}_j) I_k A_{k,*,j}, k = 1, 2, \cdots K$$

```
}
```

Fig. 2.2.: Pseudocode for updating a single voxel j. The parameters for the 1-D optimization problem are computed based on a substitute function approach and a new value for the voxel is evaluated. The error sinogram e is then updated based on this new voxel value.

In order to find the optimal update for the jth voxel we begin by rewriting the cost function (2.8) by ignoring terms not involving voxel j as

$$\tilde{c}(u) = \tilde{\theta}_1 u + \frac{\tilde{\theta}_2}{2}\left(u - \tilde{f}_j\right)^2 + \sum_{i\in\mathcal{N}_j} w_{ji}\rho(u - f_i) \tag{2.9}$$

where $\tilde{\Lambda}_k = \frac{1}{\sigma_k^2}\Lambda_k$, $\tilde{\theta}_1 = -\sum_{k=1}^{K} e_k^t \tilde{\Lambda}_k (I_k A_{k,*,j})$, $\tilde{\theta}_2 = \sum_{k=1}^{K}(I_k A_{k,*,j})^t \tilde{\Lambda}_k (I_k A_{k,*,j})$, $A_{k,*,j}$ is the jth column of the forward projection matrix A_k at tilt k, $e_k = g_k - I_k A_k f - d_k \mathbb{1}$, $\tilde{f}_j$ is the present value of voxel j, and $\mathcal{N}_j$ is the set of all neighbors of voxel j. We must minimize this function with respect to u to find the optimal update for voxel j. Taking the derivative with respect to u and setting it to zero does not give a closed form update for u due to the complicated form of the potential function $\rho(\cdot)$. Therefore we use a substitute function approach in which we find a function which bounds (2.9) from above, such that minimizing the new function results in a lower

value of the original cost function. Typically the substitute function is chosen so that it can be minimized in a computationally efficient manner. We find a substitute function for each $\rho(u - f_i)$ of the form

$$\rho(u - f_i; \tilde{f}_j - f_i) \quad = \quad \frac{a_{ji}}{2}(u - f_i)^2 + b_{ji}. \tag{2.10}$$

Using such a form results in a simple closed form update for a given voxel. The values of a_{ji} and b_{ji} can be derived as shown in Appendix A and are given by

$$a_{ji} \quad = \quad \begin{cases} \frac{\rho'(\tilde{f}_j - f_i)}{(\tilde{f}_j - f_i)} & \tilde{f}_j \neq f_i \\ \rho''(0) & \tilde{f}_j = f_i \end{cases} \tag{2.11}$$

$$b_{ji} \quad = \quad \rho(\tilde{f}_j - f_i) - \frac{a_{ji}}{2}(\tilde{f}_j - f_i)^2 \tag{2.12}$$

Given these values, the new cost function we need to minimize is

$$\tilde{c}_{\text{sub}}(u) = \tilde{\theta}_1 u + \frac{\tilde{\theta}_2}{2}\left(u - \tilde{f}_j\right)^2 + \sum_{i \in \mathcal{N}_j} w_{ji}\rho(u - f_i; \tilde{f}_j - f_i).$$

Since $\rho(u - f_i; \tilde{f}_j - f_i)$ is quadratic in u, the minimum of $\tilde{c}_{\text{sub}}(u)$ has a closed form and is given by

$$u^* \quad = \quad \frac{\displaystyle\sum_{i \in \mathcal{N}_j} w_{ji}a_{ji}f_i + \tilde{\theta}_2\tilde{f}_j - \tilde{\theta}_1}{\displaystyle\sum_{i \in \mathcal{N}_j} w_{ji}a_{ji} + \tilde{\theta}_2}. \tag{2.13}$$

Enforcing the positivity constraint, the final update for voxel j is given by

$$\hat{f}_j \quad \leftarrow \quad \max(u^*, 0) \tag{2.14}$$

The pseudocode for updating a single voxel is given in Fig. 2.2.

```
UpdateGainOffset(f, e) {
      for each tilt k {
            A_k f ← (g_k − d_k 𝟙 − e_k) / I_k
            Compute Q_k, b_k
      }
      Compute λ̂ using (2.15)
      for each tilt k {
            Compute Î_k and d̂_k using (2.16)
            e_k ← g_k − Î_k(A_k f) − d̂_k 𝟙
      }
}
```

Fig. 2.3.: Pseudocode for updating the gain and offset parameters. The parameters for the constrained optimization problem are computed efficiently using the present error sinogram and used to evaluate the optimal gains and offsets. Based on these new values, the error sinogram is updated.

2.4.2 Gain and Offset Update

To minimize the function (2.8) with respect to the gains and offsets subject to the constraint, we turn the constrained optimization problem into an unconstrained one by using a Lagrange multiplier λ. If $Q_k = \begin{bmatrix} (A_k\hat{f})^t\tilde{\Lambda}_k(A_k\hat{f}) & (A_k\hat{f})^t\tilde{\Lambda}_k\mathbb{1} \\ (A_k\hat{f})^t\tilde{\Lambda}_k\mathbb{1} & \mathbb{1}^t\tilde{\Lambda}_k\mathbb{1} \end{bmatrix}$,

$b_k = \begin{bmatrix} b_{k,1} \\ b_{k,2} \end{bmatrix} = \begin{bmatrix} g_k^t\tilde{\Lambda}_k A_k\hat{f} \\ g_k^t\tilde{\Lambda}_k\mathbb{1} \end{bmatrix}$, $\tilde{\Lambda}_k = \frac{1}{\sigma_k^2}\Lambda_k$, $I = [I_1,\cdots,I_K]$, $Q_k^{-1} = \begin{bmatrix} \tilde{q}_{k,11} & \tilde{q}_{k,12} \\ \tilde{q}_{k,21} & \tilde{q}_{k,22} \end{bmatrix}$,

the optimal update can be derived as shown in Appendix B and is given by

$$\hat{\lambda} = \frac{\sum_{k=1}^{K}(\tilde{q}_{k,11}b_{k,1} + \tilde{q}_{k,12}b_{k,2}) - K\bar{I}}{\frac{1}{K}\sum_{k=1}^{K}\tilde{q}_{k,11}} \qquad (2.15)$$

and

$$\begin{bmatrix} \hat{I}_k \\ \hat{d}_k \end{bmatrix} = Q_k^{-1}\left(b_k - \frac{1}{K}\begin{bmatrix} \hat{\lambda} \\ 0 \end{bmatrix}\right). \qquad (2.16)$$

The pseudocode for updating the gains and offsets efficiently is given in Fig. 2.3.

2.4.3   Variance Parameter Update

To find the optimal value of the variance parameter, we can take the gradient of (2.8) with respect to σ^2 and set it to zero. This gives us the optimal update for each σ_k^2 as

$$\hat{\sigma}_k^2 \quad \leftarrow \quad \frac{\|e_k\|_{\Lambda_k}^2}{M} \tag{2.17}$$

where $e_k = g_k - \hat{I}_k A_k \hat{f} - \hat{d}_k \mathbb{1}$ and M is the number of measurements for each tilt.

2.4.4   Multi-resolution Initialization

In model based X-Ray CT it is common to initialize the reconstruction using FBP. In HAADF-STEM tomography initializing with FBP is challenging because there are several nuisance parameters for which we have no initial estimates. Moreover since the data sets are large and the optimization tend to be computationally intensive, the overall convergence can be sped up by using a multi-resolution algorithm to initialize the parameters. Multiresolution methods have proved to be useful in initializing parameters in other tomographic modalities [13, 35, 66]. These methods transfer the computational load to the coarser scales where the optimization algorithm is faster because the dimensionality of the problem is effectively reduced. In our application we set the parameter c in the prior model to be small, and $q = 2$ so our prior behaves similar to a GGMRF [67]. Therefore we adapt the scaling parameter of the prior model, σ_f, for different resolutions using Eq. 28 in [65]. Next we detail a method to initialize the parameters at the coarsest resolution.

```
UpdateParameters(f, I, d, σ², N_InnerIter, T){
    Λ̃_k ← (1/σ²_k)Λ_k, k = 1, ⋯, K
    OuterIter ← 0 //holds the number of iterations
    do {
        e_k ← g_k − I_k(A_k f) − d_k𝟙, k = 1, ⋯, K
        for Iter=1 to N_InnerIter //inner loop
            for each voxel j in random order
                (f̂_j, e) ← UpdateVoxel(j, e)
        (Î, d̂, e) ← UpdateGainOffset(f̂, e)
        for each tilt k
            σ̂²_k ← ‖e_k‖²_Λ_k / M
            Λ̃_k ← (1/σ̂²_k)Λ_k
        if (OuterIter == 0)
            N_InnerIter ← 1
        OuterIter ← OuterIter+1
    } while (Stopping Threshold ≥ T or OuterIter ≤ 1)
}
```

Fig. 2.4.: Pseudocode to update all the parameters and find the optimal value of the voxels. First, all the voxels are updated in random order, followed by the gains, offsets and noise variance. The value $N_{\text{InnerIter}}$ determines the number of inner iterations over the voxels. At the very first outer iteration at the coarsest resolution this parameter is set to a fixed value greater than 1. In all subsequent iterations this value is set to 1. The algorithm is terminated if the relative change in the magnitude of the reconstruction is less than a threshold T and the number of outer iterations is greater than one. Keeping track of the error sinogram (e_k) at each tilt k results in a computationally efficient algorithm.

Parameter Initialization At Coarsest Scale

In HAADF-STEM tomography, we observe that the signal measured at each view is offset by an additive constant due to the detector. In order to estimate these offsets in the optimization framework we require a reasonable initial estimate. To do so we assume that the material is homogenous. Thus the measured signal at each view k will be proportional to $\frac{1}{\cos(\theta_k)}$ plus an unknown offset. At the coarsest resolution the offset parameter vector d is initialized by a least squares fit of the average count in each view k to $\frac{1}{\cos(\theta_k)}$. More specifically, if G is a $K \times 1$ vector containing the

MultiResolutionReconstruction(g, σ_f, $\bar{I}$, N_{Res}, N_{Inner}, T){
    Set $d^{(0)}$ according to (2.18) and (2.19)
    $I^{(0)} = \bar{I}\mathbb{1}$, $\sigma^{2(0)} = \mathbb{1}$, $f^{(0)} = 0$
    for $k = 1$ to N_{Res}
    if($k == 1$)
        $\left(f^{(k)}, I^{(k)}, d^{(k)}, \sigma^{2(k)}\right) \leftarrow$
UpdateParameters $\left(f^{(k-1)}, I^{(k-1)}, d^{(k-1)}, \sigma^{2(k-1)}, N_{\text{Inner}}, T\right)$
    else
            $f^{(k-1)} \leftarrow$ UpSample $\left(f^{(k-1)}\right)$
            $\sigma_f^{(k-1)} \leftarrow$ Update σ_f using Eq. 28 in [65]
            $\left(f^{(k)}, I^{(k)}, d^{(k)}, \sigma^{2(k)}\right) \leftarrow$
UpdateParameters $\left(f^{(k-1)}, I^{(k-1)}, d^{(k-1)}, \sigma^{2(k-1)}, 1, T\right)$
}

Fig. 2.5.: Pseudocode to perform multiresolution initialized reconstruction. N_{Res} is the number of resolutions and N_{Inner} is the number of inner iterations at the coarsest resolution. The superscript attached to each unknown parameter indicates the resolution index. A superscript of 0 corresponds to the coarsest resolution with higher values corresponding to finer resolutions. At the coarsest resolution the parameters are initialized as shown. At subsequent resolutions the algorithm is initialized using the output of the previous resolution. The function Upsample produces a finer resolution object from a coarser resolution object by using a suitable up-sampling technique.

average counts of the data at each tilt, and $D = \begin{bmatrix} \frac{1}{\cos(\theta_1)} & \cdots & \frac{1}{\cos(\theta_K)} \\ 1 & \cdots & 1 \end{bmatrix}^t$ then the least squares estimate $\hat{\phi} = [\hat{\phi}_1 \, \hat{\phi}_2]^t$ is given by

$$\hat{\phi} = (D^t D)^{-1}(D^t G) \tag{2.18}$$

and initial value of the offset is set to

$$d = \hat{\phi}_2 \mathbb{1}. \tag{2.19}$$

The initial value of the voxels (f) is set to 0. Since the source electron dose and detector gains are typically unknown, the gains I, are set to $\bar{I}\mathbb{1}$. Finally the noise variance parameter σ^2 is set to $\mathbb{1}$.

Iterations Over Voxels At The Coarsest Scale

At the very first iteration at the coarsest scale, given the initial values of the parameters, we update all the voxels iteratively multiple times ($N_{\text{InnerIter}}$ in Fig. 2.4) so that we do not converge to a local minimum with unreasonable values for the HAADF scatter coefficients. Beyond this stage the parameters are all updated once per iteration.

The algorithm can be efficiently implemented by keeping track of the error vector in the course of the iterations as shown in Fig. 2.4. We note that the stopping criteria for our method is that the relative change in the magnitude of the reconstruction is less than a preset threshold and the number of iterations is greater than one. The pseudocode for the full multiresolution initialization based reconstruction is shown in Fig. 2.5.

2.5 Experimental Results

2.5.1 Simulated Data Set

We begin by studying the performance of the MBIR algorithm on a simulated data set produced from a phantom consisting of spheres in vacuum. Our objective is to compare the results of our method to the most widely used algorithms for HAADF-STEM tomography, FBP and SIRT. The FBP and SIRT reconstructions are performed using a popular electron microscopy package, IMOD [68]. Since the scaling is arbitrary and no positivity is enforced in IMOD [68], we clip the reconstructions to be positive and perform a least squares fit to scale the reconstructions to a similar range as the true phantom. We use the visual quality as well as the root mean square
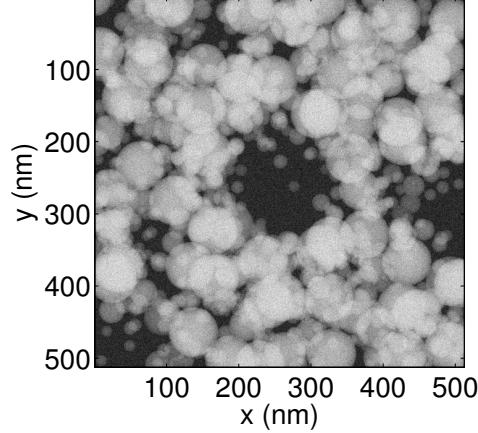
Fig. 2.6.: Simulated HAADF microscope image of aluminum spheres in vacuum when the object is not tilted.

Table 2.1: Comparison of the Root Mean Square Error of the reconstruction with respect to the original phantom for various scenarios. SIRT and FBP have higher RMSE than MBIR indicating that MBIR can produce quantitatively accurate reconstructions.

Method	RMSE (nm^{-1})
FBP	1.278×10^{-4}
SIRT	0.988×10^{-4}
MBIR ($p = 1$)	0.284×10^{-4}
MBIR ($p = 1.2$)	0.324×10^{-4}
MBIR ($p = 2$)	0.496×10^{-4}

error (RMSE) between the reconstruction and the original phantom to evaluate the reconstructions.

Fig. 2.6 show the simulated HAADF data corresponding to the phantom at zero tilt. The images have been displayed by scaling them using the minimum and maximum count in the data. The spheres in the phantom have a scattering coefficient of 4.132×10^{-4} nm^{-1} corresponding to Aluminum at 300 kV with detection angles $50 - 250$ mrad (value obtained using the Monte-Carlo simulator, CASINO [69]). The sphere diameters vary (up to 100 nm) and the sample thickness is about 128 nm.

Phantom   FBP   SIRT

(a)   (c)   (e)

(b)   (d)   (f)

(g)   (i)   (k)

(h)   (j)   (l)
MBIR $(p = 1)$ MBIR $(p = 1.2)$ MBIR $(p = 2)$
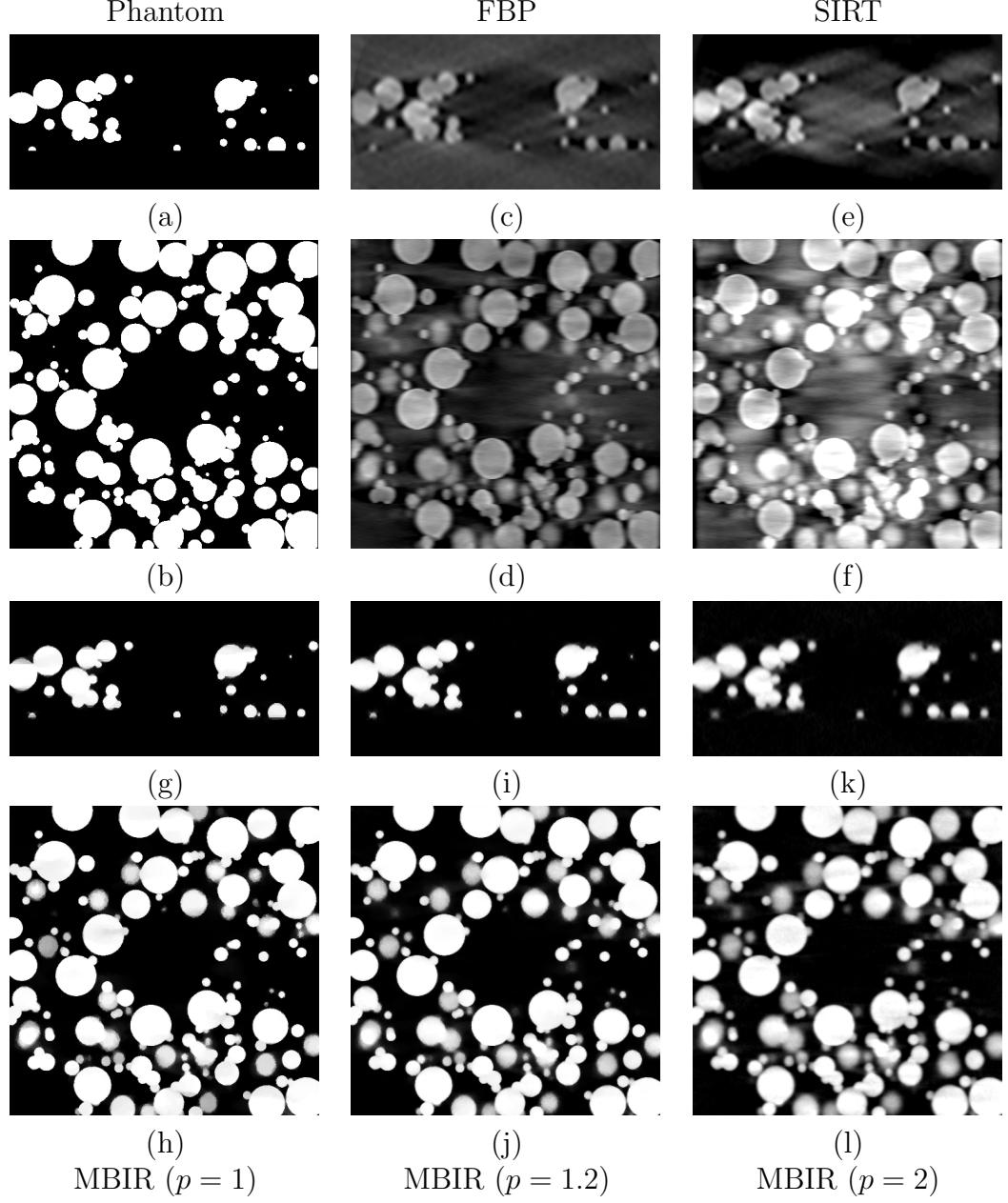
Fig. 2.7.: Reconstructions using FBP, SIRT and MBIR with different prior models. (a), (c), (e), (g), (i), and (k) show a $x - z$ slice and (b), (d), (f), (h), (j), and (l) show a $x - y$ slice. (a) and (b) show ground truth corresponding to a single slice. (c) and (d) show reconstructions from FBP, (e) and (f) show SIRT reconstructions using IMOD. (g) - (l) show MBIR reconstructions with different values of the prior model parameter p. FBP and SIRT reconstructions are noisy and have streaking artifacts in the $x - z$ plane while MBIR significantly suppresses these artifacts and can produce sharper reconstructions. Increasing the value of the prior model parameter p from 1 to 2 produces smoother MBIR reconstructions. Thus the value of p can be chosen to best match the type of material being imaged.
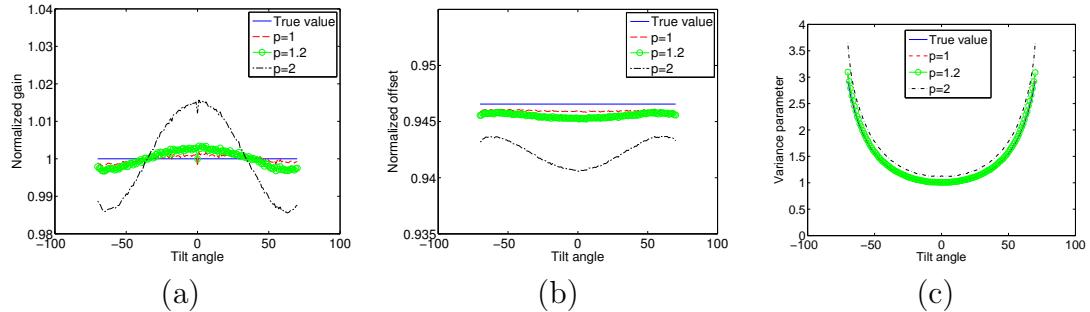
Fig. 2.8.: Final value of the gains, offset and variance parameters upon termination of the algorithm. The gains are normalized by the dosage used and the offset are normalized to the mean value of the counts in the data set. We observe that the MBIR algorithm can estimate these unknown parameters and hence they need not be explicitly measured.

The object is tilted from $+70°$ to $-70°$ in steps of $1°$ and the projection images are obtained using (2.1) with an electron flux of 50000 counts per nm^2 for every measurement and the detector gain set to 1. The offset d_k is set to 9000 counts for each view. To each HAADF projection measurement Gaussian noise is added with variance parameter σ^2 set so that the noise variance increases with tilt. The values of σ_k^2 corresponds to a minimum SNR $\left(\min_{k,i} 10 \log \left(\frac{g_{k,i}}{\sigma_k^2} \right) \right)$ of 34.471 dB. The projection images are acquired at a pixel size of 1nm $\times$ 1nm. All the reconstructions are performed with $\sigma_f = 4.1 \times 10^{-5}$ nm^{-1}, $q = 2$, and $c = 0.01$ using a 3 stage multi-resolution initialization. At the coarsest resolution the initial value of the image is set to $f = 0$, the variance parameter is set to $\mathbb{1}$ and the number of inner iterations, $N_{\text{InnerIter}}$ (see Fig. 2.5), is set to 10. The interpolation between resolutions is performed using pixel replication. The value of $\bar{I}$ is set to 50000, the value of the source electron dose - detector gain product, so that the reconstructions are quantitative. The stopping threshold is set to 0.1%.

Fig. 2.7 (a) and (b) show a single $x - z$ and $x - y$ slice from the original phantom. Fig. 2.7 (c) and (d) shows the corresponding slices from the reconstruction obtained using the FBP algorithm implemented in IMOD [68]. The algorithm results in blurry reconstructions with significant streaking artifacts in the $x - z$ plane. In the $x - y$

plane the reconstructions are noisy. Fig. 2.7 (e) and (f) show the reconstructions using the SIRT algorithm. While some of the noise appears suppressed compared to FBP we still observe the streaking artifacts in the $x - z$ plane and noise in the $x - y$ plane. Moreover in regions where the phantom has no material, we observe that the reconstructions still contain material with non zero scatter coefficients. The RMSE of the SIRT reconstructions are lower than those of the FBP reconstructions as shown in Table. 2.1.

Fig. 2.7 (g) - (l) shows the reconstructions when we apply MBIR algorithm to the simulated data set with different values of the q-GGMRF parameter p. This phantom has discontinuous boundaries so as we would expect the total-variation prior ($p = 1$ in Fig. 2.7 (g) and (h)) is well matched to its behavior and produces the lowest RMSE (see Table. 2.1) reconstruction. Fig. 2.7 (i) and (j) shows the MBIR reconstruction when we set $p = 1.2$. This produces results with slightly more smooth edges than the $p = 1$ case. This prior been found to be useful [56] since it provides a good balance between preserving edges and modeling the smooth regions in the reconstruction. Finally Fig. 2.7 (k) and (l) show the MBIR reconstruction when $p = 2$. In this case the edges are most diffuse and hence the RMSE is higher than the $p = 1$ and $p = 1.2$ case for this phantom. In all cases the streaking artifacts in the $x - z$ plane are significantly suppressed and noise in the $x - y$ slices is effectively reduced compared to FBP and SIRT. The RMSE of the MBIR reconstructions (see Table. 2.1) are lower than those of the SIRT and FBP suggesting that the MBIR reconstructions are quantitatively more accurate than FBP and SIRT. Thus this experiment illustrates that MBIR is superior to FBP and SIRT and furthermore the parameters of the algorithm can be chosen to model a range of interfaces from very sharp to diffuse.

Fig. 2.8 (a), (b), and (c) show the final estimated values of the normalized gain, $\frac{I_k}{I}$, normalized offset, $\frac{d_k}{\bar{g}}$, (where $\bar{g}$ is the average count in the data set) and variance parameters, σ_k^2. We observe that the parameters are accurately estimated demonstrating that the MBIR method can accurately account for missing calibration parameters without explicitly measuring them.

2.5.2   Experimental Data Set
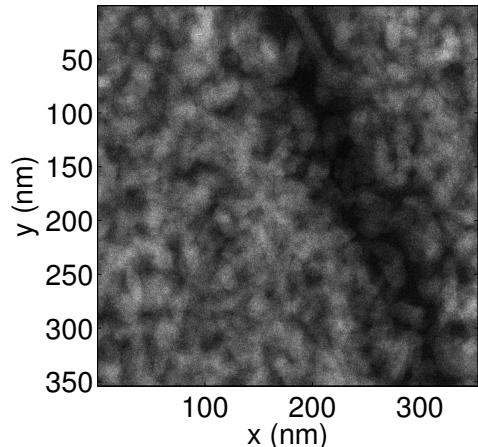
Fig. 2.9.:   Acquired HAADF data for a Titanium dioxide nanoparticle data set at zero tilt. The dark regions represent a void in the material.

In order to evaluate our approach on real data, we compare our algorithm with FBP and SIRT from IMOD [68]. The data acquired is of a ≈ 150 nm thick sample of polystyrene functionalized Titanium dioxide nano particle assembly [70]. The TEM used was a FEI Titan operating in STEM mode with 300 kV accelerating voltage, spot size 7, +/- 70° with 2° increments for +/- 54° and 1° increments for 54° to 70° and $-54°$ to $-70°$. The exposure time was 12.6 seconds, magnification was set to 225 kX, the frame size set to 2048×2048, with 0.34 nm pixel size, and STEM dynamic focus activated. A Fischione model 3000 HAADF photomultiplier tube detector was used at camera length of 130 mm. We use a ≈ 350 nm $\times 350$ nm section of the projection images for reconstruction. Fig. 2.9 shows a single projection image acquired when the object is at zero tilt, displayed by scaling it to the range of the data.

The FBP and SIRT reconstructions are performed with voxels of size 0.343 nm $\times$ 0.343 nm $\times$ 0.343 nm. The filter parameters for FBP are chosen to produce the most visually appealing results. The particles of interest in this data set are approximately cylindrical with a diameter of 18 nm and a height of 40 nm [70]. Thus in order to reduce computation, MBIR is performed with voxels of size (3×0.343) nm $\times (3 \times$

$0.343)$ nm $\times$ (3×0.343) nm. The parameter σ_f is chosen for the best visual quality of reconstruction. The value of c is set to 0.01. We use a three stage multiresolution initialization for the reconstruction. The interpolation between resolutions is performed using pixel replication. At the coarsest resolution the number of inner iterations ($N_{\text{InnerIter}}$ in Fig. 2.5) is set to 10. Since the source electron dose and detector gains are unknown in this case, we set $\bar{I} = 20000$ and jointly estimate the gains along with the offsets and variance parameters as a part of the reconstruction. The stopping threshold is set to 0.9%. The dimensions of the reconstructed volume are set so as to account for all the voxels contributing to the projection data. In presenting the results we only show voxels that can be reliably reconstructed from the projection data i.e. at every tilt there is a projection measurement corresponding to those voxels. Additionally while displaying the results, we use a scaling window ranging from the minimum to the maximum value in the reconstructed volume.

Fig. 2.10 shows a single $x - z$ and $x - y$ slice from FBP, SIRT and the MBIR reconstruction. We observe that in SIRT and FBP there are streaking artifacts in the $x - z$ plane of reconstruction while MBIR significantly suppresses these artifacts. In the $x - y$ plane the effects of noise are effectively suppressed in MBIR clearly showing the Titanium dioxide nano particles against the background support material. This demonstrates the effectiveness of the method even for this particularly limited tilt data set.

Finally we study the effect of varying the parameter p of the q-GGMRF prior on the quality of reconstruction. Fig. 2.11 shows a reconstructed $x - z$ and $x - y$ slice for $p = 1$ and $p = 2$. Values of p close to 1 represent strong edge preserving reconstructions values close to 2 result in smoother reconstructions. This is an additional flexibility which our model based approach offers compared to SIRT and FBP i.e. the ability to incorporate our prior knowledge of the nature of the interfaces of the material. Knowledge of smooth interfaces can be incorporated by setting $p = 2$ and sharp interfaces can be represented by the setting p close to 1. Fig. 2.12 shows the final values of the gains, offsets and variance parameters upon termination of the

FBP   SIRT   MBIR ($p = 1.2$)

(a)   (c)   (e)

(b)   (d)   (f)
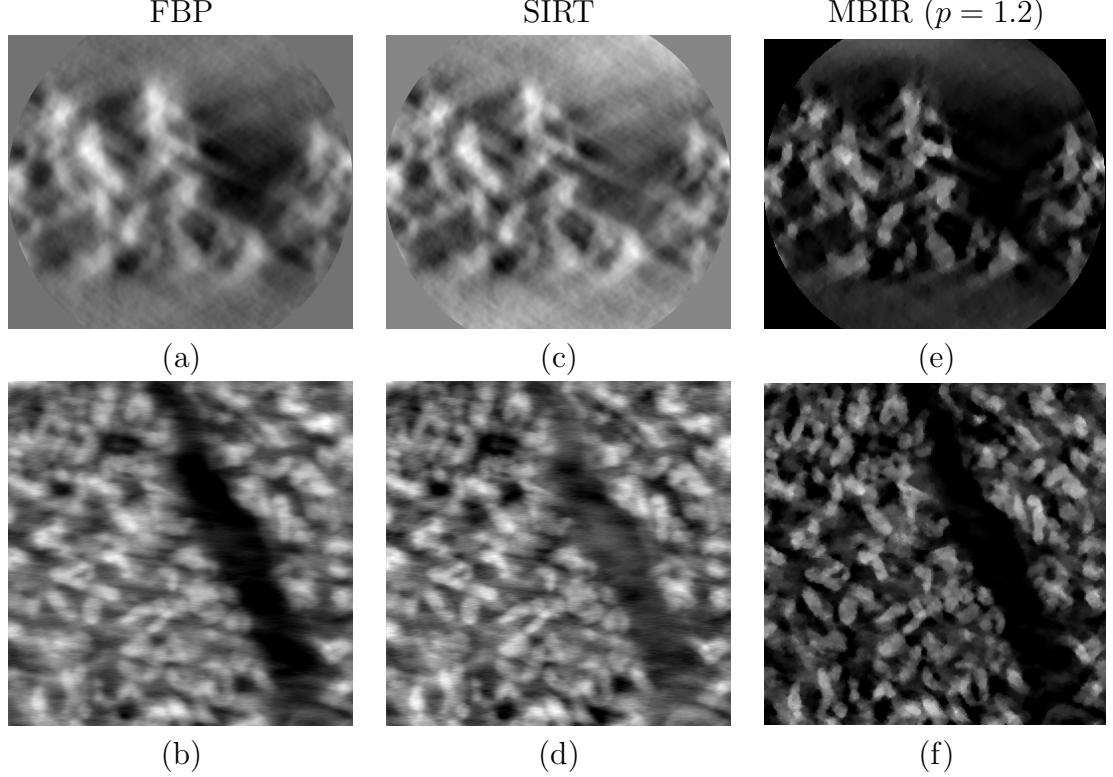
Fig. 2.10.:   Comparison of MBIR with FBP and SIRT on a real HAADF-STEM data set. The top row shows a $x - z$ reconstructed slice and the bottom shows a $x - y$ reconstructed slice. (a) and (b) FBP reconstruction, (c) and (d) SIRT with 20 iterations, (e) and (f) MBIR with $p = 1.2$. MBIR produces images with no streaks in the $x - z$ plane and significantly suppresses noise in the $x - y$ plane.

algorithm for the case when $p = 1$, $p = 1.2$ and $p = 2$. However, we do not have knowledge of the true value of the parameters to compare with those estimated by our algorithm.

2.6   Conclusions

In this chapter, we presented a model based iterative reconstruction algorithm for HAADF-STEM tomography. We used a model for the measurement process and combined it with a prior model of the 3D volume to formulate the reconstruction as a MAP estimation problem. We presented a fast algorithm for quantitative recon-

MBIR ($p = 1$)   MBIR ($p = 2$)

(a)   (c)

(b)   (d)
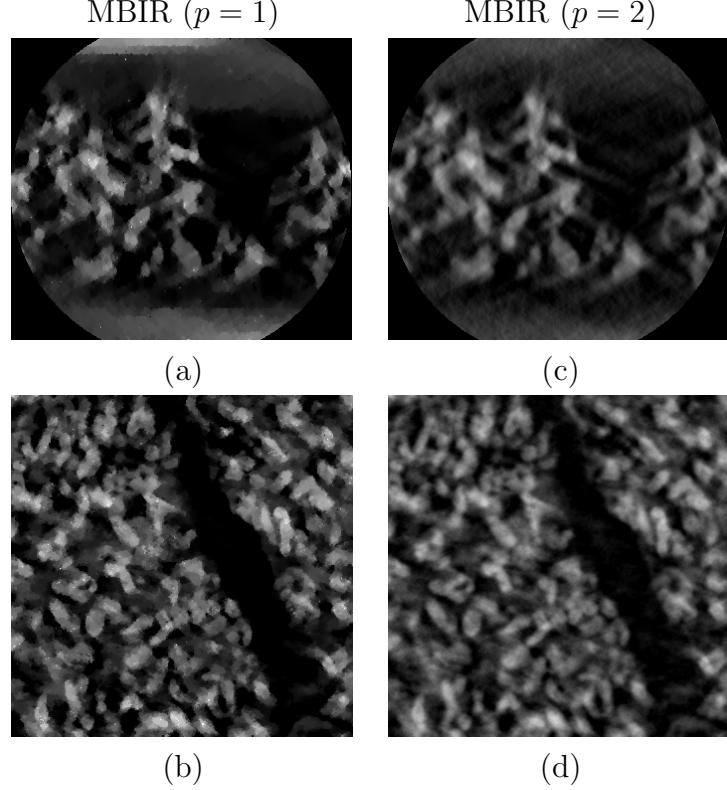
Fig. 2.11.: Effect of varying q-GGMRF shape parameter p on the reconstruction. The top row shows a $x - z$ reconstructed slice and the bottom row shows a $x - y$ reconstructed slice. (a) and (b) $p = 1$, (c) and (d) $p = 2$. As the value of p increases the edges in the reconstructions become smoother.

(a)   (b)   (c)
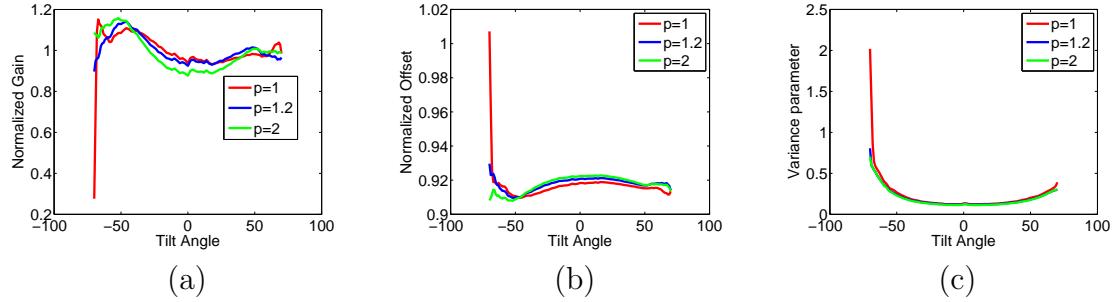
Fig. 2.12.: Final estimates for normalized gain $\left(\frac{I_k}{\bar{I}}\right)$, normalized offset $\left(\frac{d_k}{\bar{g}}\right)$ and variance parameter (σ_k^2) with respect to tilt θ_k for $p = 1$, $p = 1.2$ and $p = 2$.

struction of the local HAADF scatter rate per unit distance which also accounted for unknown calibration parameters. Furthermore, the algorithm is flexible to incorporate knowledge of the type of interfaces present in the material via the prior model. We showed that our algorithm produces superior reconstructions compared to FBP and SIRT on a simulated and a real data set.

3. MODEL-BASED ITERATIVE RECONSTRUCTION FOR BRIGHT FIELD ELECTRON TOMOGRAPHY

3.1 Introduction

Bright Field (BF) electron tomography (ET) has been widely used in the life sciences to characterize biological specimens in 3D [38] using either a transmission electron microscope (TEM) or a scanning transmission electron microscope (STEM) [71]. BF-ET typically involves focusing an electron beam on a sample, acquiring images of transmitted electrons corresponding to various sample tilts, and using an algorithm on the acquired "tilt-series" to reconstruct the object. In most cases due to the geometry of the acquisition and mechanical limitations of the tilting stages, BF-ET is a limited angle parallel beam transmission tomography modality. Further details of the ET acquisition and preprocessing are discussed in [3].

While there are a few instances where BF-ET has been used in the physical sciences [72–74], it has generally been avoided [2,23], due to the occurrence of contrast reversals [24] from dynamical diffraction effects such as Bragg scatter [18]. Bragg scatter occurs when the crystal lattice is oriented in such a manner that the incident electrons are elastically scattered away from the direct path [18] leading to a measurement uncharacteristic of attenuation due to thickness alone. We refer to measurements which are not well modeled by attenuation due to thickness alone as anomalous measurements. These anomalies make interpretation of the individual BF images complicated and result in strong artifacts if the BF tilt-series is used for tomographic reconstruction using standard reconstruction algorithms such as FBP [23]. Thus BF-ET has generally been avoided in the physical sciences due to the complicated nature of the data and the inability of the standard reconstruction algorithms like FBP to handle such data.
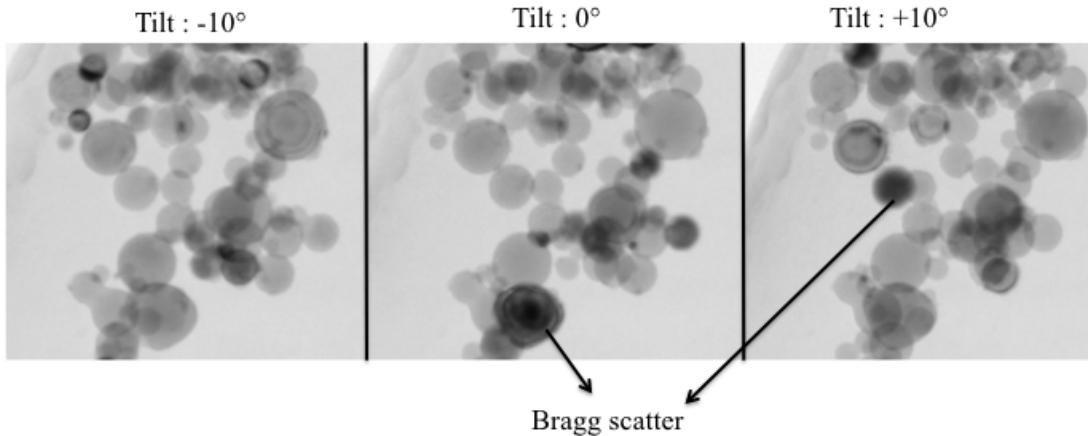
Fig. 3.1.: Illustration of the "anomalies" present in a real BF-TEM data set of Aluminum nanoparticles. The figure shows BF images corresponding to three different tilts of the specimen. Note that certain spheres turn dark (fewer counts) and then again turn bright due to Bragg scatter (contrast reversal). These effects make it challenging to directly apply standard analytic tomographic reconstruction algorithms to the data.

Model-based iterative reconstruction (MBIR) provides a powerful framework for tomographic reconstruction that uses a probabilistic model for the measurement (forward model) and a probabilistic model for the object (prior model) to obtain reconstructions that are qualitatively superior and quantitatively accurate for a variety of applications [8, 10, 12, 31, 33, 75]. Typically MBIR involves the design and minimization of a cost function corresponding to the maximum *a posteriori* probability (MAP) estimate with two sets of terms - one corresponding to a likelihood for the data and the other corresponding to a prior model for the object. While most efforts for BF-ET have used the FBP algorithm [23, 71–74] Levine [45] has developed a MBIR algorithm for BF-ET in the case of thick specimens. However his work deals with amorphous samples, for which there are no anomalies due to dynamical diffraction in the measurement.

In this paper, we present an MBIR algorithm for accurate reconstruction of BF-ET data [76] containing anomalous measurements that typically result from crystalline

samples. Our approach is based on a novel generalized Huber function that is used in the forward model (i.e., log likelihood) to account for the anomalous measurements due to Bragg or other errors. The generalized Huber function is parameterized so that it can model the heavy tailed distribution of the errors that are present in anomalous measurements. Using this forward model, we derive an MBIR cost function which allows for joint estimation of both the unknown image, f, and a key parameter of the generalized Huber function. This approach allows for adaptive parameter estimation in the reconstruction process, which is important in practical applications.

We also propose an optimization algorithm that is based on majorization-minimization using a surrogate function [26, 27] together with a fast multiresolution iterative algorithm to find a minima of the MBIR cost function. The surrogate function that we propose bounds the MBIR cost as a function of both the unknown image, f, and the unknown model parameters of the generalized Huber function. Consequently, minimization of the surrogate function allows for efficient estimation of model parameters during the MBIR reconstruction process.

Importantly, the surrogate function requires the introduction of a binary auxiliary variable for each measurement. This auxiliary variable has the interpretation of forming a binary mask that classifies each measurement as either anomalous or normal. During the reconstruction process, the anomalous measurements are then attenuated relative to normal measurements. The resulting binary masks also contains potentially important information about the Bragg scatter and consequently the crystal orientation for each particle in the 3D reconstruction. Therefore, we also propose an algorithm that extracts a Bragg feature vector for each particle in the volume by correlating the projection of each segmented particle with the binary anomaly mask at each view angle. We conjecture that the extracted Bragg feature vector for each particle can provide useful information in applications.

We apply our method to simulated data containing Bragg scatter like anomalies as well as real TEM data from crystalline particles. Results from the simulated as well as the real data set show that MBIR with anomaly modeling can significantly improve

the reconstruction quality compared to FBP and conventional MBIR, suppressing the artifacts that arise due to the anomalous measurements. We also use our new method to extract a Bragg feature vector for each particle and demonstrate how this feature vector can potentially provide useful information about the crystal orientation for each particle in the 3D volume. The source code along with a GUI application implementing our method is publicly available at the website - *www.openmbir.org.*

The organization of the rest of the paper is as follows. In section 3.2 we introduce a new statistical model for the measurement system and formulate the MBIR cost function. In section 3.3 we propose an efficient algorithm to minimize the cost function. In section 3.4 we present results from a simulated data set, followed by results from a real data set. Finally, in section 3.5 we draw our conclusions.

3.2 Statistical Model and Cost Formulation
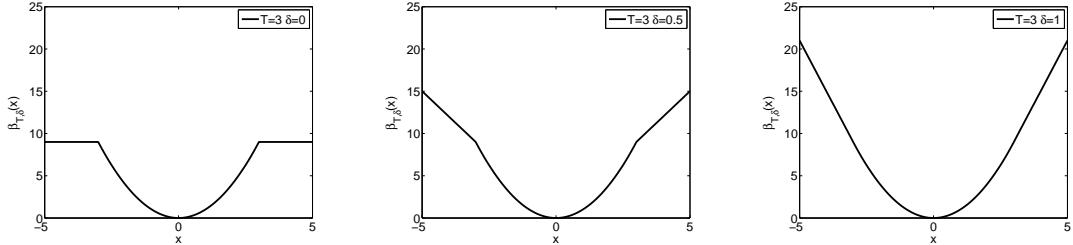
Fig. 3.2.: Illustration of the generalized Huber function $\beta_{T,\delta}$ used for the likelihood term with $T = 3$ and $\delta = 0, 0.5$ and 1. When $\delta = 1$ the function reduces to the Huber function. Large model mismatch errors are penalized by restricting their influence on the overall cost function.

The goal of BF-ET is to reconstruct an attenuation coefficient at every point in the sample. The attenuation coefficient is related to the ability of the material to scatter the incident beam away from the direct path which is dependent on the differential cross section, geometry of the detector, density of the material and incident electron energy. An electron beam is focused on the material and the electrons that are scattered by the sample through small angles are captured by a BF detector to

obtain a single image. The sample is then tilted along a fixed axis and the process is repeated. Thus, at the end of the acquisition, we obtain a collection of BF images that can be used for tomographic reconstruction of the attenuation coefficients.

In order to reconstruct the attenuation coefficients associated with the sample, we use an MBIR framework. The reconstruction in the MBIR framework is typically given by the joint-MAP [16] estimate

$$(\hat{f}, \hat{\phi}) = \underset{f,\phi}{\mathrm{argmin}} \left\{ -\log p(g|f,\phi) - \log p(f) \right\} \tag{3.1}$$

where g is the vector of measurements, f is the vector of unknown voxels (attenuation coefficients), ϕ is a vector of unknown calibration parameters (nuisance parameters), $p(g|f,\phi)$ is the likelihood (forward model) and $p(f)$ is the prior probability for the unknown voxels.

In order to develop a forward model for BF-ET that accounts for the anomalous effects, we start with the simple case when there are no anomalies. Let $\lambda_{k,i}$ be the electron counts corresponding to the ith measurement at the kth tilt and $\lambda_{D,k}$ be the counts that would be measured in the absence of the sample at that tilt (blank scan). We model the attenuation of the beam through the material using Beer's law. Thus, an estimate of the projection integral corresponding to the ith measurement at the kth tilt is given by $\log\left(\frac{\lambda_{D,k}}{\lambda_{k,i}}\right)$. Notice that unlike in high-angle annular dark field electron microscopy [75], the BF case requires a log operation to be applied to a normalized version of the measurement. There can be cases in which the blank scan, $\lambda_{D,k}$, is not measured, and we can include it as an unknown parameter in the MBIR framework and estimate it as a part of the reconstruction. If g_k is a $M \times 1$ vector with $g_{k,i} = -\log(\lambda_{k,i})$, f is an $N \times 1$ vector of unknown attenuation coefficients of the material and, $d_k = -\log(\lambda_{D,k})$, then, assuming $\lambda_{k,i}$'s are conditionally independent Poisson random variables it has been shown that [53] the conditional mean of $g_{k,i}$ can be approximated by $A_{k,i,*}f + d_k$ and the conditional variance is proportional to

$\frac{1}{\mathbb{E}[\lambda_{k,i}]}$, where A_k is a $M \times N$ forward projection matrix and $A_{k,i,*}$ is the ith row of A_k. Modeling the conditional density of the measurements as a Gaussian distribution [11] results in a probability density function (pdf),

$$p(g|f, d, \sigma) =$$
$$\frac{1}{Z_l} \exp \left\{ -\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} (g_{k,i} - A_{k,i,*}f - d_k)^2 \frac{\Lambda_{k,ii}}{\sigma^2} \right\}$$

$$(3.2)$$

where K is the total number of tilts, $g = [g_1^t \cdots g_K^t]^t$ is a $KM \times 1$ data vector, Λ_k is a diagonal matrix with entries set so that $\frac{\sigma^2}{\Lambda_{k,ii}}$ is the variance of the measurement $g_{k,i}$ with σ^2 being a proportionality constant, $d = [d_1 \cdots d_K]$ is a vector containing the offset parameters, and Z_l is a normalizing constant. For such a transmission tomography model it has been shown that $\Lambda_{k,ii} = \mathbb{E}[\lambda_{k,i}] \approx \lambda_{k,i}$ [29]. We note that our formulation can account for more sophisticated physics models as introduced in [77], but in this paper we focus on using Beer's law as it has been found to be accurate for a class of materials and thickness typically studied using BF-ET [77].

3.2.1 Generalized Huber Functions for Anomaly Modeling

Bragg scatter from crystalline material can cause the BF-ET measurements to vary substantially from the model of equation (3.2). Fig. 3.1 shows an example of three tilts from a BF tilt series with regions having significant anomalies.

A precise way of accounting for these anomalies would require identifying 3D regions of the object that consist of a single crystal, and modeling the associated crystal structure. While possible, this would be a highly ill-posed inverse problem to recover from a single 2D tilt series due to the unknown 3D orientation of the crystals. Furthermore, modeling other classes of anomalies such as Fresnel fringes [78] and extinction contours involve more complex physics making the data more diffi-

cult to invert. Therefore, instead of modeling the complicated physics of dynamical diffraction that leads to anomalies, we will use an alternate approach.

In order to account for anomalous effects like Bragg scatter, we propose a modified likelihood function that models the anomalies as outliers of a pdf

$$p(g|f, d, \sigma) =$$
$$\frac{1}{Z} \exp \left\{ -\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_{T,\delta} \left((g_{k,i} - A_{k,i,*}f - d_k) \frac{\sqrt{\Lambda_{k,ii}}}{\sigma} \right) \right\}$$
(3.3)

where $\beta_{T,\delta} : \mathbb{R} \to \mathbb{R}$ such that

$$\beta_{T,\delta}(x) = \begin{cases} x^2 & |x| < T \\ 2\delta T|x| + T^2(1 - 2\delta) & |x| \geq T \end{cases}$$

and Z is a normalizing constant. We call $\beta_{T,\delta}$ the generalized Huber function. Fig. 3.2 shows the generalized Huber function for three different values of δ. Notice that δ controls the tail behavior of the density function while T controls the threshold beyond which a measurement is considered anomalous. When $\delta = 0$, $\beta_{T,\delta}$ corresponds to the weak-spring potential [79] used for image modeling and results in a function with the heaviest tails among the three cases. However, when $\delta = 0$ we cannot jointly estimate the calibration parameters because the likelihood is not a valid density function since it does not integrate to 1. When $\delta = 1$, $\beta_{T,\delta}$ reduces to the Huber function [25] which is a convex function and corresponds to a pdf with the lightest tail among the three cases. When T is very large then $\beta_{T,\delta}$ is effectively a quadratic function and the likelihood reduces to the standard transmission tomography model in (3.2). Thus the generalized Huber function can be adjusted to have heavier tails than the density function in (3.2) to account for the various anomalies in the data set.

Restricting $0 < \delta \leq 1$ and using the fact that

$$\int p(g|f,d,\sigma)dg = 1$$

we can show that the normalizing constant has the form

$$Z = \sigma^{MK} \times \text{Constants}.$$

Hence, the modified log-likelihood function for the BF-ET case is given by

$$-\log p(g|f,d,\sigma) =$$

$$\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_{T,\delta} \left((g_{k,i} - A_{k,i,*}f - d_k) \frac{\sqrt{\Lambda_{k,ii}}}{\sigma} \right)$$

$$+ MK \log(\sigma) + \text{Constants} \quad (3.4)$$

Each term in the summation corresponds to a penalty on the ratio of the data mismatch error $(g_{k,i} - A_{k,i,*}f - d_k)$ to the noise standard deviation $\left(\frac{\sigma}{\sqrt{\Lambda_{k,ii}}} \right)$. Thus T has the interpretation that if the data fit error is greater than T times the noise standard deviation then that measurement is likely to be an anomaly. Notice that typically σ is not known and hence we will jointly estimate it as a part of the reconstruction.

3.2.2 MBIR Cost Formulation

Combining the statistical model in (3.4) with a prior model of the form

$$p(f) \quad = \quad \frac{1}{Z_f} \exp\left\{-s(f)\right\} \quad (3.5)$$

where Z_f is a normalizing constant, the reconstruction is obtained by minimizing the cost

$$c(f, d, \sigma) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_{T,\delta} \left((g_{k,i} - A_{k,i,*}f - d_k) \frac{\sqrt{\Lambda_{k,ii}}}{\sigma} \right)$$

$$+ MK \log(\sigma) + s(f). \quad (3.6)$$

Alternately, we can define $h_{k,i} : \mathbb{R}^{N+K+1} \to \mathbb{R}$ to be a function such that

$$h_{k,i}(f, d, \sigma) = (g_{k,i} - A_{k,i,*}f - d_k) \frac{\sqrt{\Lambda_{k,ii}}}{\sigma}.$$

The cost function can then be written as

$$c(f, d, \sigma) = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_{T,\delta}(h_{k,i}(f, d, \sigma)) + MK \log(\sigma) + s(f).$$

$$(3.7)$$

Additionally, we will constrain $f \geq 0$, as it is physically meaningful to have positive values of the attenuation coefficients. Thus, the MBIR BF-ET reconstruction is given by

$$\left(\hat{f}, \hat{d}, \hat{\sigma} \right) \leftarrow \underset{f \geq 0, d, \sigma}{\operatorname{argmin}} c(f, d, \sigma)$$

The cost function (3.7) is non-convex in general, and thus finding the global minimum is computationally expensive. Therefore we will present an algorithm to find a local minimum of the cost. Furthermore, since the likelihood term of (3.7), is not differentiable, gradient-based methods can not be directly applied. Hence, we use a majorization-minimization strategy [26, 27] combined with alternating minimization to find a minimum of the cost.

3.3 Optimization Algorithm

Our optimization approach is based on the repeated minimization of a differentiable surrogate function. The function $q(z; z')$ is a surrogate function for the function $t(z)$ at the point z' if the following two conditions hold.

$$
\begin{aligned}
q(z; z') &\geq t(z) \\
q(z'; z') &= t(z')
\end{aligned}
\tag{3.8}
$$

If $Q(f, d, \sigma; f', d', \sigma')$ is a surrogate function to $c(f, d, \sigma)$ at the point (f', d', σ'), our algorithm consists of repeating the following steps until convergence

(i) For each voxel j

$$
f'_j \leftarrow \underset{f_j \geq 0, f_k = f'_k \forall k \neq j}{\mathrm{argmin}} \ Q(f, d', \sigma'; f', d', \sigma')
$$

(ii) $d' \leftarrow \underset{d}{\mathrm{argmin}} \ Q(f', d, \sigma'; f', d', \sigma')$

(iii) $\sigma' \leftarrow \underset{\sigma}{\mathrm{argmin}} \ Q(f', d', \sigma; f', d', \sigma')$

The algorithm is terminated if the ratio of the average change in the magnitude of the reconstruction to the average magnitude of the reconstruction is less than a preset threshold. In addition we use a multiresolution initialization [35] to speed up the convergence of the algorithm and prevent the method from getting stuck in undesirable local minima. In multi-resolution initialization, we perform a reconstruction at a coarser resolution (larger voxel sizes) and use this result to initialize a finer resolution reconstruction. This transfers the computational load to the coarser scale
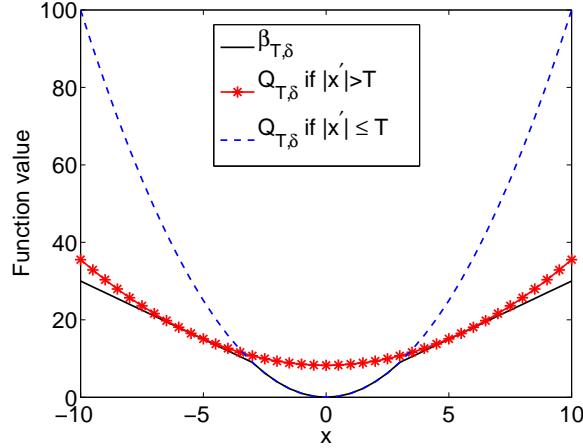
Fig. 3.3.: Illustration of the surrogate function, $Q_{T,\delta}$, to the generalized Huber function with $T = 3$ and $\delta = \frac{1}{2}$. The surrogate function is plotted for two cases : one when $x' = 2$ and the other when $x' = 5$. The surrogate function in each case is a quadratic function thereby simplifying the subsequent optimization.

where the optimization can be done quickly due to to the reduced dimensionality of the problem.

Note that the surrogate function approach ensures monotonic decrease of the original cost function (3.7) with each update; so the sequence of costs must be convergent. In addition we have empirically observed that the reconstructions also converge. While theoretical convergence proof exist for majorization techniques with alternating minimization, we have no formal proof of convergence in this case due to the complicated nature of the cost function.

To derive the exact updates for the above algorithm we will first design a surrogate function to the original cost assuming any general prior model $s(f)$. Next we will present a specific $s(f)$ and derive a surrogate for this case and use it to derive the update equations for each iteration.

3.3.1 Construction of Surrogate Function

We design surrogate functions for each function $\beta_{T,\delta}(h_{k,i}(f, d, \sigma))$ in (3.7) at a given point (f', d', σ') and sum them up to form a surrogate to the overall cost function. In order to design a surrogate function, note that each term in the summation in (3.7) is a composition between the generalized Huber function $\beta_{T,\delta}$ and the function $h_{k,i}$. Therefore, we first design a surrogate function to the generalized Huber function, $\beta_{T,\delta}$, and then use this function along with a composition property to design a surrogate function to the composition $\beta_{T,\delta} \circ h_{k,i}$.

In particular

$$
Q_{T,\delta}(x; x') = \begin{cases} x^2 & |x'| < T \\ \frac{\delta T}{|x'|}x^2 + \delta T|x'| + T^2(1 - 2\delta) & |x'| \geq T \end{cases}
$$

is a surrogate function to $\beta_{T,\delta}(x)$. Fig. 3.3 shows the construction of a surrogate function to the generalized Huber function for the case when $T = 3$ and $\delta = 0.5$. Notice that while the generalized Huber function is non-differentiable, the surrogate function is quadratic and hence differentiable in x.

Next, we will use the composition property of surrogate functions to design a surrogate function for each $\beta_{T,\delta} \circ h_{k,i}$ in (3.7). The composition property of surrogate function states that if $q(z; z')$ is a surrogate function to $t(z)$ at z' then $q(h(z); h(z'))$ is a surrogate function to $t(h(z))$ at z' (proof in Appendix C). Using the composition property of surrogate functions, the composition of $Q_{T,\delta}$ with $h_{k,i}$, $Q_{T,\delta}(h_{k,i}(f, d, \sigma); h_{k,i}(f', d', \sigma'))$ is a surrogate function to $\beta_{T,\delta}(h_{k,i}(f, d, \sigma))$ in (3.7).

Using the surrogate function for each $\beta_{T,\delta}(h_{k,i}(f,d,\sigma))$,

$$\tilde{Q}(f,d,\sigma;f',d',\sigma') =$$

$$\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{M}Q_{T,\delta}(h_{k,i}(f,d,\sigma);h_{k,i}(f',d',\sigma'))$$

$$+ MK\log(\sigma) + s(f) \quad (3.9)$$

is a surrogate function to the original cost (3.7). Hence, even though the terms corresponding to the generalized Huber function in the original cost function may be non-differentiable, we have constructed a surrogate function which overcomes this difficulty and makes the optimization tractable.

3.3.2 Prior Model and Surrogate Function

We use a special case of the q-generalized Gaussian Markov random field (qG-GMRF) [56] for the prior, resulting in

$$s(f) = \sum_{\{j,k\}\in\mathcal{N}} w_{jk}\rho(f_j - f_k)$$

$$\rho(f_j - f_k) = \frac{\left|\frac{f_j-f_k}{\sigma_f}\right|^2}{c+\left|\frac{f_j-f_k}{\sigma_f}\right|^{2-p}}$$

$\mathcal{N}$ is the set of pairs of neighboring voxels (e.g. a 26 point neighborhood), and p, c and σ_f are qGGMRF parameters. The weights w_{jk} are inversely proportional to the distance between voxels j and k, normalized to 1. We fix $c = 0.001$ and restrict $1 \leq p \leq 2$ similar to [75].

In order to simplify the optimization, we also introduce a surrogate function for the prior with the form

$$\rho(f_j - f_k; f'_j - f'_k) = \frac{a_{jk}}{2}(f_j - f_k)^2 + b_{jk}. \quad (3.10)$$

The values of a_{jk} and b_{jk} can be derived as shown in [75] and are given by

$$
a_{jk} = \begin{cases} \frac{\rho'(f'_j - f'_k)}{(f'_j - f'_k)} & f'_j \neq f'_k \\ \rho''(0) & f'_j = f'_k \end{cases} \tag{3.11}
$$

$$
b_{jk} = \rho(f'_j - f'_k) - \frac{a_{jk}}{2}(f'_j - f'_k)^2 \tag{3.12}
$$

Thus a surrogate function to $s(f)$ at $f = f'$ is given by

$$
s(f; f') = \sum_{\{j,k\} \in \mathcal{N}} w_{jk} \rho(f_j - f_k; f'_j - f'_k). \tag{3.13}
$$

Substituting (3.13) into (3.9) results in the final surrogate function given by

$$
Q(f, d, \sigma; f', d', \sigma') =
$$
$$
\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} Q_{T,\delta}(h_{k,i}(f, d, \sigma); h_{k,i}(f', d', \sigma'))
$$
$$
+ MK \log(\sigma) + \sum_{\{j,k\} \in \mathcal{N}} w_{jk} \rho(f_j - f_k; f'_j - f'_k). \tag{3.14}
$$

In order to simplify the subsequent updates, we define the following binary indicator variable,

$$
b'_{k,i} = \begin{cases} 1 & |(g_{k,i} - A_{k,i,*}f' - d'_k)\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}| < T \\ 0 & |(g_{k,i} - A_{k,i,*}f' - d'_k)\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}| \geq T \end{cases} \tag{3.15}
$$

---

function $[\hat{f}, \hat{d}, \hat{\sigma}] \leftarrow$ RECONSTRUCT(g, f', d', σ')
    %**Inputs**: Measurements g, Initial reconstruction f', Initial offset d', Initial variance parameter σ'
    %**Outputs**: Reconstruction $\hat{f}$ and nuisance parameters $(\hat{d}, \hat{\sigma})$
    $e' = g - Af' - d'\mathbb{1}$ ▷ Initialize error vector
    while Stopping criteria is not met **do**
        for each voxel j **do**   ▷ Voxel updates
            Compute θ_1 and θ_2 using (3.17)
            for $k \in \mathcal{N}_j$ **do**
                Compute surrogate function parameter a_{jk} using (3.11)
            end for
            Compute u^* using (3.18)
            $f_j \leftarrow \max(u^*, 0)$
            $e' \leftarrow e' - (f_j - f'_j)A_{*,j}$
            $f'_j \leftarrow f_j$
            Update b' using (3.15)
        end for
        $d' \leftarrow$ Update d using (3.20)   ▷ Offset parameter update
        Update e'
        Update b' using (3.15)
        $\sigma' \leftarrow$ Update σ using (3.21)   ▷ Variance parameter update
        Update b' using (3.15)
    end while
    $\left(\hat{f}, \hat{d}, \hat{\sigma}\right) \leftarrow (f', d', \sigma')$
end function

Fig. 3.4.: MBIR algorithm for BF data with anomalies. The algorithm works by constructing a surrogate to the original function based on the current values of the voxels and nuisance parameter and minimizing this surrogate function with respect to each variable. The process is then repeated. The algorithm can be efficiently implemented by keeping track of the error sinogram, e', and the anomaly classifier vector, b'.

Intuitively $b'_{k,i}$ indicates if a given measurement is classified as anomalous or not, based on the current state of the reconstruction. If we define the error $e_{k,i} = g_{k,i} - A_{k,i,*}f - d_k$ and $e'_{k,i} = g_{k,i} - A_{k,i,*}f' - d'_k$ we can rewrite (3.14) as

$$
\begin{aligned}
Q(f, d, \sigma; f', d', \sigma') = \\
\frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} e_{k,i}^2 \frac{\Lambda_{k,ii}}{\sigma^2} \left(b'_{k,i} + (1 - b'_{k,i}) \frac{\delta T \sigma'}{|e'_{k,i}| \sqrt{\Lambda_{k,ii}}} \right) \\
+ MK \log(\sigma) + \sum_{\{j,k\} \in \mathcal{N}} w_{jk} \rho(f_j - f_k; f'_j - f'_k) \\
+ \text{Terms not dependent on } (f, d, \sigma) \quad (3.16)
\end{aligned}
$$

Thus we observe that the surrogate function can be easily constructed by maintaining an error vector, e', based on the current values of the unknowns, and the indicator variable, b'.

3.3.3 Update Equations Corresponding to the Surrogate Function

Voxel Update

The voxels are updated in random order similar to [8] in order to speed up the overall convergence of the algorithm. To speed up the implementation of the algorithm the voxel updates are parallelized similar to [64]. To minimize the surrogate function with respect to voxel j, we can fix $f_k = f'_k \ \forall k \in \{1, \cdots, M\} \setminus \{j\}$, $d = d'$ and $\sigma = \sigma'$ in (3.16). The cost function to minimize is

$$
\tilde{c}_{\text{sub}}(u) = \theta_1 u + \frac{\theta_2}{2} \left(u - f'_j \right)^2 + \sum_{k \in \mathcal{N}_j} w_{jk} \rho(u - f'_k; f'_j - f'_k)
$$

where $\mathcal{N}_j$ is the set of all neighbors of voxel j and

$$\theta_1 = -\sum_{k=1}^{K}\sum_{i=1}^{M}A_{k,i,j}\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}\left[b'_{k,i}e'_{k,i}\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}+(1-b'_{k,i})\delta T\frac{e'_{k,i}}{|e'_{k,i}|}\right]$$

$$\theta_2 = \sum_{k=1}^{K}\sum_{i=1}^{M}A_{k,i,j}^2\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}\left[b'_{k,i}\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'}+(1-b'_{k,i})\frac{\delta T}{|e'_{k,i}|}\right].$$

$$(3.17)$$

Since $\rho(u - f'_k; f'_j - f'_k)$ is quadratic in u, the minimum of $\tilde{c}_{\mathrm{sub}}(u)$ has a closed form and is given by

$$u^* = \frac{\displaystyle\sum_{k\in\mathcal{N}_j}w_{jk}a_{jk}f'_k + \theta_2 f'_j - \theta_1}{\displaystyle\sum_{k\in\mathcal{N}_j}w_{jk}a_{jk} + \theta_2}. \qquad (3.18)$$

Enforcing the positivity constraint, the update for the voxel is

$$f'_j \leftarrow \max\left(u^*, 0\right) \qquad (3.19)$$

Offset Parameter Update

In order to minimize the surrogate function with respect to the offset parameter d, we take the gradient of the surrogate function (3.16) $Q(f', d, \sigma'; f', d', \sigma')$ with respect to d and set it to zero. This gives the optimal update as

$$d'_k \leftarrow d'_k + \frac{\displaystyle\sum_{k=1}^{K}\sqrt{\Lambda_{k,ii}}\left[e'_{k,i}b'_{k,i}\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'} + \delta T\frac{e'_{k,i}}{|e'_{k,i}|}(1-b'_{k,i})\right]}{\displaystyle\sum_{k=1}^{K}\sqrt{\Lambda_{k,ii}}\left[b'_{k,i}\frac{\sqrt{\Lambda_{k,ii}}}{\sigma'} + \frac{\delta T}{|e'_{k,i}|}(1-b'_{k,i})\right]}.$$

$$(3.20)$$

Variance Parameter Update

In order to update the variance parameter we minimize the surrogate function (3.16) with respect to σ setting $f = f'$ and $d = d'$. This gives the optimal update as

$$\sigma' \leftarrow \sqrt{\frac{\sum_{k=1}^{K}\sum_{i=1}^{M} e_{k,i}'^{2}\Lambda_{k,ii}b_{k,i}' + \sum_{k=1}^{K}\sum_{i=1}^{M}(1 - b_{k,i}')\delta T |e_{k,i}'|\sigma'\sqrt{\Lambda_{k,ii}}}{MK}}.$$

(3.21)

The MBIR BF-ET algorithm for a single resolution is summarized in Fig. 3.4.

3.3.4 Initialization

Since the MBIR cost function is non-convex, initializing the algorithm with a reasonable initial estimate is important. We use a multi-resolution initial condition to prevent the algorithm from becoming stuck in undesirable local minima. We initialize the values of f to 0 nm^{-1} at the coarsest scale. The value of d and σ are initialized from a region of the image where there is no sample present. Furthermore, at the coarsest scale we perform a few iterations (typically 10) over the voxels with the value of T set to be very large in order to obtain a reasonable initial condition for the overall multiresolution algorithm. Since the size of the voxels is large at the coarse scales, this initialization takes few iterations and is computationally inexpensive to perform.

3.4 Results

In this section we compare four algorithms for BF-ET: FBP, conventional model-based iterative reconstruction (MBIR), the proposed MBIR with anomaly modeling and known parameter values (MBIR-AM), and the proposed method with anomaly modeling and parameter estimation (MBIR-AM-PE). We apply the methods to simulated data as well as real data. For the simulated data we will compare results from
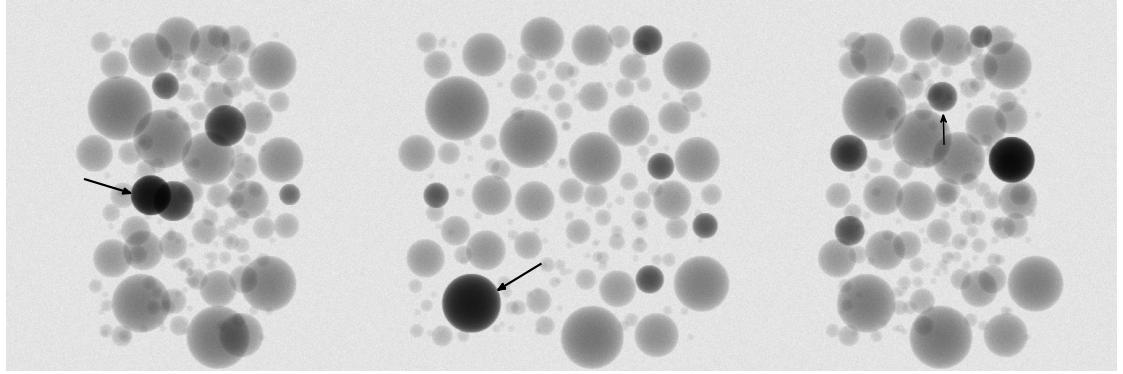
Fig. 3.5.: Simulated BF data corresponding to a 3D phantom of spheres for three successive tilts. The arrows in the figure show example locations with the simulated Bragg scatter obtained by increasing the attenuation coefficient of a few spheres in the phantom. We model these as anomalies in the projection data as they can cause artifacts in the reconstructions produced using the standard reconstruction techniques.

Table 3.1: Comparison of the Root Mean Square Error of the reconstruction with respect to the original phantom for various scenarios. MBIR with anomaly modeling produces quantitatively more accurate reconstructions.

Algorithm	RMSE ($\times 10^{-4}$ nm^{-1})
FBP	13.90
MBIR	4.95
MBIR-AM ($T = 3, \delta = 0.5$)	4.30
MBIR-AM-PE ($T = 3, \delta = 0.5$)	4.31

Table 3.2: Comparison of the Root Mean Square Error ($\times 10^{-4}$nm^{-1}) of the reconstruction with respect to the original phantom when varying T and δ. A value of $T = 3$ and $\delta = 0.5$ produces the lowest RMSE reconstruction.

T \ δ	0.1	0.5	1
1	4.50	4.42	4.44
3	4.40	**4.31**	4.60
5	4.33	5.09	5.14
20	5.06	5.06	5.06

Phantom | FBP | Conventional MBIR

(a) | (c) | (e)

(b) | (d) | (f)
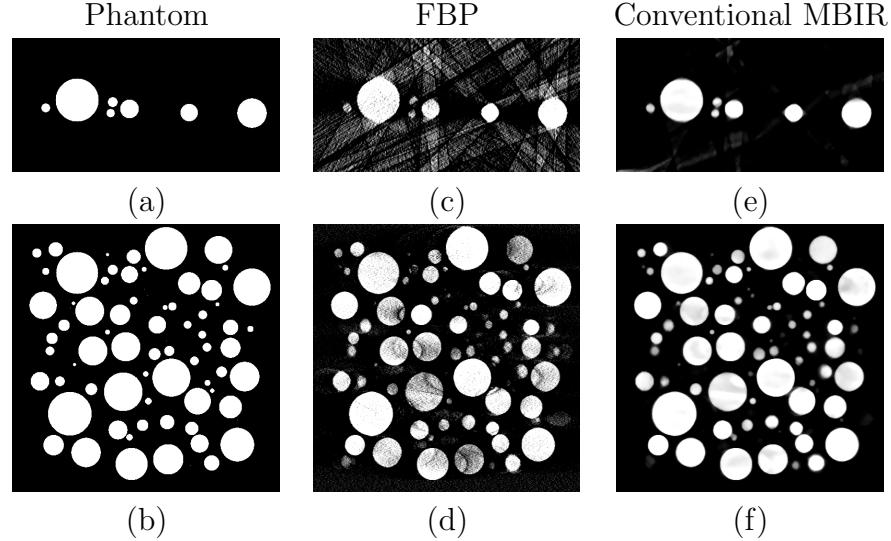
Proposed MBIR ($T = 3, \delta = 0.5$) Proposed MBIR with parameter estimation

(g) | (i)

(h) | (j)
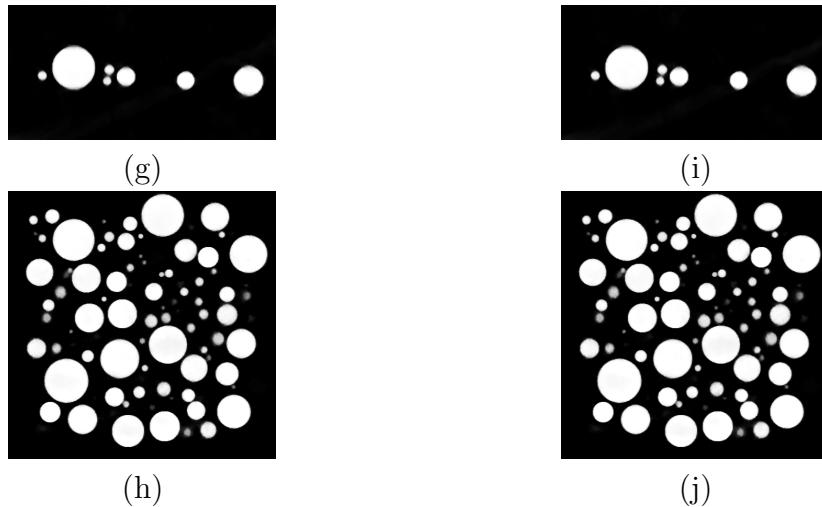
Fig. 3.6.: Comparison of BF reconstructions for a data set with Bragg scatter like anomalies. (a) and (b) show a single $x - z$ and $x - y$ cross-section from the phantom used. The horizontal direction represents the x axis. (c) and (d) show the corresponding cross sections from a FBP reconstruction. (e) and (f) show the conventional MBIR reconstruction. The reconstruction has streaks because of Bragg scatter but much lesser compared to FBP. (g) and (h) show the cross-section from MBIR with anomaly modeling ($T = 3$ and $\delta = 0.5$). The method effectively suppresses the artifacts in (c) - (f), and produces a more accurate reconstruction. Finally (i) and (j) show the reconstruction using MBIR with anomaly modeling and nuisance parameter estimation. The reconstructions are comparable to the MBIR-AM case showing that the algorithm can work well despite of the unknown parameters. All images are scaled in the range of $0 - 7.45 \times 10^{-3}$ nm^{-1}.

(a) $T = 1, \delta = 0.5$ (b) $T = 5, \delta = 0.5$ (c) $T = 20, \delta = 0.5$
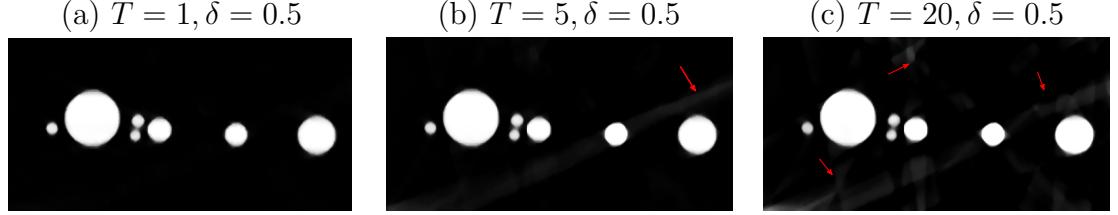
Fig. 3.7.:   Illustrates the impact of varying anomaly threshold T on the proposed MBIR reconstructions. (a) shows a $x - z$ cross section from the 3-D reconstruction when $T = 1$. (b) and (c) shows the corresponding slices when $T = 5$ and $T = 20$. Notice that for (b) and (c) there are visible streaking artifacts. A value of $T = 3$ as shown in Fig. 3.6 produces an accurate reconstruction.

all four methods while in the real data, since we do not know the parameters, we will not consider the MBIR-AM case. Finally, we will present a method for using the anomalies identified by our method to associate a Bragg feature vector for each particle in the reconstructed volume.

The FBP reconstructions are performed in Matlab using the *iradon* command and the output is clipped to be positive. For the MBIR reconstructions with anomaly modeling, we set $T = 3$, $\delta = 0.5$, and $p = 1.2$. The value of σ_f is chosen to obtain the lowest root mean square error (RMSE) for the simulated data set and is chosen to obtain the best visual quality of reconstruction for the real data set. Since our prior behaves similar to a GGMRF [67], we adapt the scaling parameter σ_f according to Eq.28 in [65] for the multi-resolution reconstructions. The offset parameter for each tilt, d_k, is initialized to the mean value of the log of the measurements from a void region in the sample. The variance parameter, σ^2, is initialized as the ratio of the mean value of the log measurements to the mean value of the measurements from a void region in the sample.

Data and anomaly classifier as T is varied
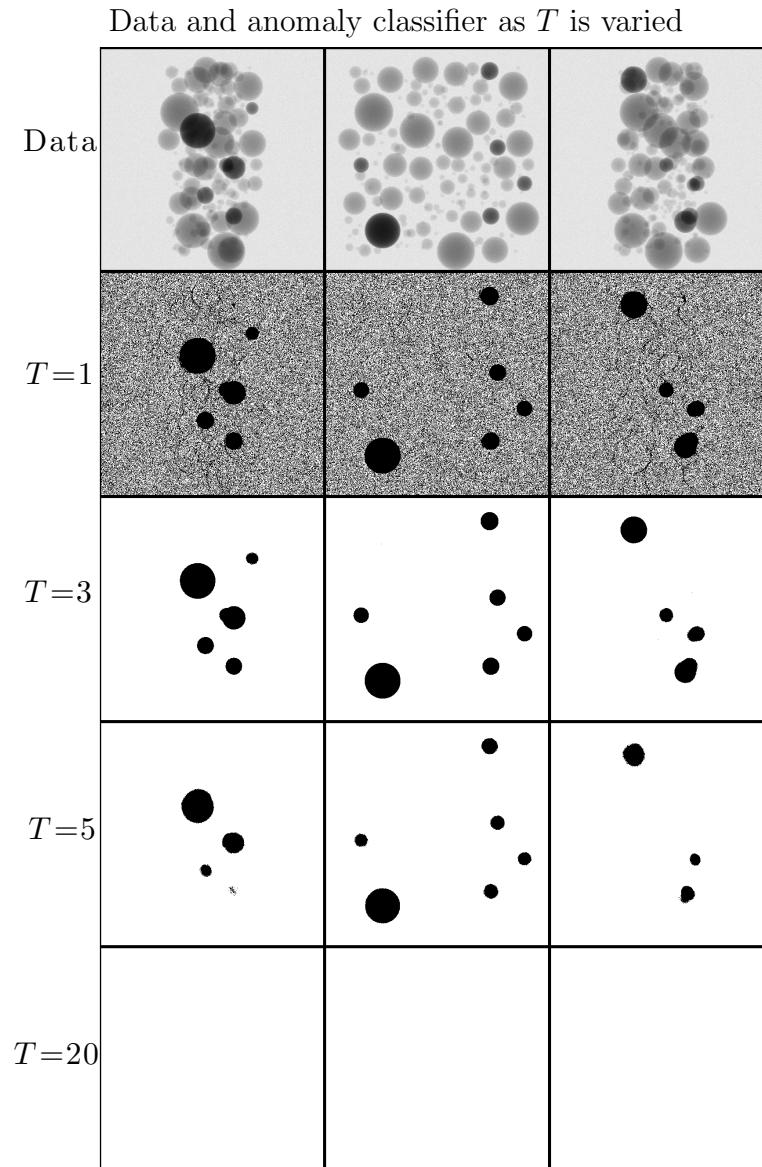
Fig. 3.8.: Data (top row) and corresponding anomaly classifier upon termination of the algorithm for three tilts from the phantom data set corresponding to different values of T. The white regions indicate areas classified as non-anomalous and the black regions correspond to the anomalies identified by the algorithm. As the value of T increases the algorithm starts to misclassify anomalies. A value of $T = 3$ provides a good trade off between the false alarms and missed detections.

3.4.1   Simulated Data Set

We use a 3-D cubic phantom containing spheres of varying radii with an attenuation coefficient of 7.45×10^{-3} nm^{-1} to generate the simulated data set. The phantom has a dimension of 256 nm $\times$ 512 nm $\times$ 512 nm ($z - x - y$ respectively). It is projected at 36 tilts in the range of $-70°$ to $+70°$ in steps of $4°$ about the y axis with a dosage $\lambda_{D,k} = 1865$ counts using the Beer's law model with Gaussian noise having variance equal to the mean of the signal. The value of σ is set to 1. At certain tilts the attenuation of a fraction of the spheres are adjusted to simulate Bragg scatter like effects (Fig. 3.5) as in a real data set.

Fig. 3.6 (a) and (b) shows a single $x - z$ and $x - y$ cross-section from the original phantom. Fig. 3.6 (c) - (j) shows the corresponding cross-section from the reconstructed volume using the different algorithms. The FBP reconstruction (Fig. 3.6(c), (d)) has strong streaking artifacts in the $x - z$ cross section and noise in the $x - y$ cross section. The conventional MBIR (Fig. 3.6 (e), (f)) shows prominent streaking artifacts in the $x - z$ cross section even though there are fewer artifacts than in FBP. Furthermore, there is some underestimation at the center of the spherical particles. However MBIR with anomaly modeling (MBIR-AM) (Fig. 3.6(g)-(h)) produces a reconstruction which effectively suppresses these artifacts. In the $x - y$ cross section, we notice that the MBIR reconstructions are less noisy as compared to FBP and that the anomaly modeling significantly improves the reconstruction. Next, we evaluate the performance of the proposed MBIR algorithm with parameter estimation (MBIR-AM-PE). Fig. 3.6 (i) and (j) show that the MBIR-AM-PE can accurately reconstruct the 3D volume suppressing the artifacts despite the unknown nuisance parameters. The value of σ upon termination of the algorithm is 1.770. We note that this value is not the final converged value of the parameter since our stopping criteria depends only on the relative change in voxel values. However we still get a good reconstruction at this termination point. In addition to the qualitative improvements shown in Fig. 3.6, Table 3.1 shows that MBIR with the anomaly modeling (MBIR-AM and

MBIR-AM-PE) significantly improves the quantitative accuracy of the reconstruction compared to FBP as well as conventional MBIR. The wall-clock time taken for the MBIR-AM-PE reconstruction ($256 \times 512 \times 512$ voxels) using a node with two Intel Xeon-E5s (total of 16 cores) was approximately 11 minutes.

Finally, we study the sensitivity of the MBIR reconstructions to the choice of parameters T and δ. Fig. 3.7 shows an $x - z$ cross section from the reconstructions for different values of T when $\delta = 0.5$. Notice that as T increases, streak artifacts start to appear in the reconstruction. This is because some of the anomalous measurements are misclassified. Fig. 3.8 shows the binary classifier mask, b', corresponding to three successive tilts from simulated data upon completion of the reconstruction. This variable indicates which measurements are classified as anomalous based on the generalized Huber function used for the reconstruction. Notice that when $T = 1$, several non-anomalous measurements are classified as anomalous (false alarms). When $T = 5$ the algorithm misses certain anomalies. When $T = 20$, all the measurements are classified as non-anomalous leading to large errors in the reconstruction. A value of $T = 3$ provides a good tradeoff and is intuitively appealing because this implies that if the data fit error for a measurement is less than 3 times the standard deviation of the noise, then that measurement is non-anomalous. Thus the trade off between false positives and missed detection of anomalies can be varied via the parameter T in the algorithm. Table. 3.2 shows the RMSE when we vary δ for the different values of T. The value of δ controls the influence of the anomalous measurements on the reconstruction. A value of δ close to 0 implies the anomalous measurements are weighted less in the cost function, while $\delta = 1$ implies the anomalies are weighted significantly. For this particular simulation, we get the lowest RMSE for the $T = 3$ and $\delta = 0.5$ case.

FBP   Conventional MBIR   MBIR with anomaly modeling

(a)   (c)   (e)

(b)   (d)   (f)
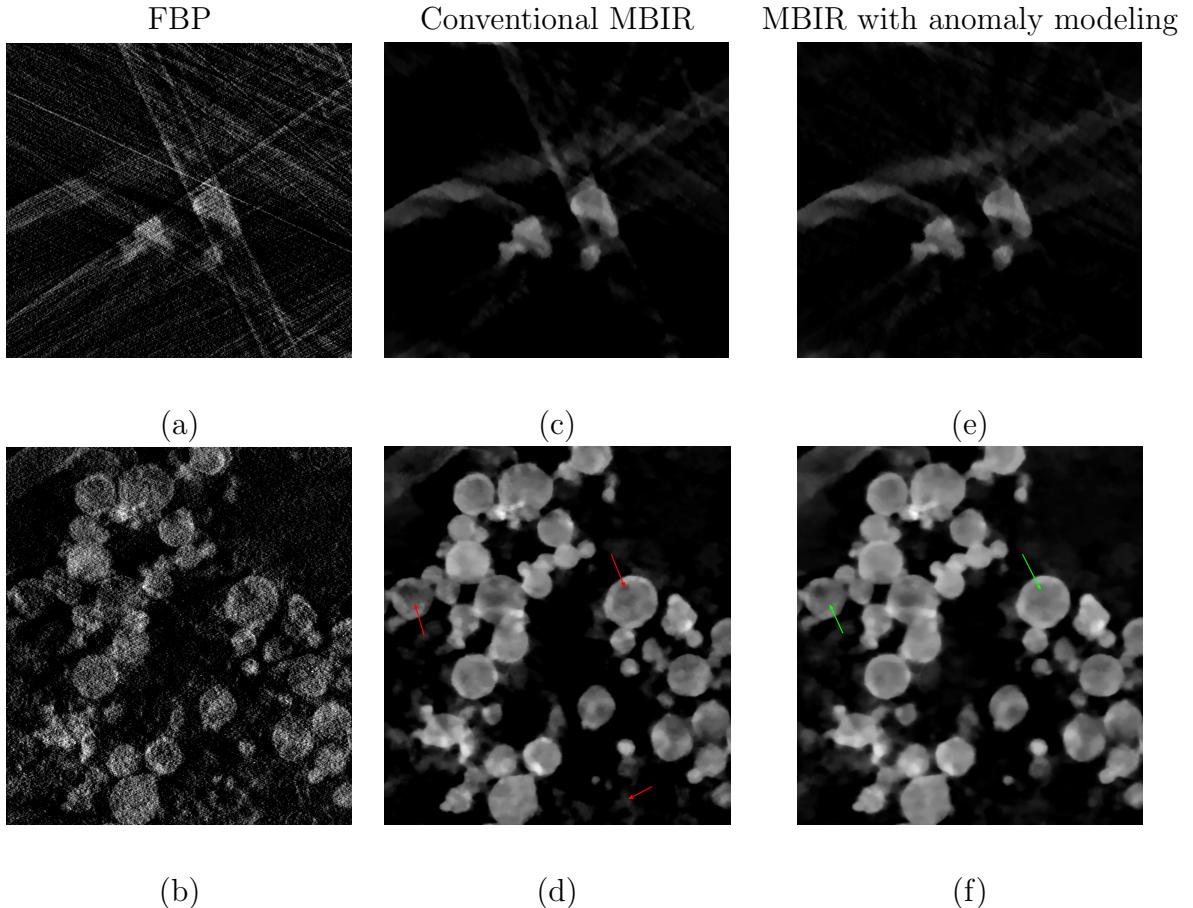
Fig. 3.9.: A single $x-z$ and $x-y$ cross-section reconstructed using different algorithms from a BF-TEM data set of Aluminum sphere nanoparticles. The horizontal direction represents the x axis. The FBP reconstruction (a)-(b) has very strong streaking artifacts in the $x - z$ cross-section, and noise in the $x - y$ cross-section suggesting why it has been avoided for BF-ET. The MBIR algorithm with the anomaly modeling and parameter estimation ($T = 3$ and $\delta = 0.5$) (e)-(f) is superior to the conventional MBIR (c)-(d), suppressing the streaking artifacts seen in (c). In the case of MBIR, the circular cross section of the spherical particles are clearly visible compared to FBP. All images are scaled in the range of $0 - 6.0 \times 10^{-3}$ nm^{-1}.

3.4.2 Real Data Set

In order to evaluate our approach on real data, we use a data set of approximately 700 nm thick crystalline aluminum nanoparticles in a carbon support. We used a FEI

Data and anomaly classifier (MBIR-AM-PE)
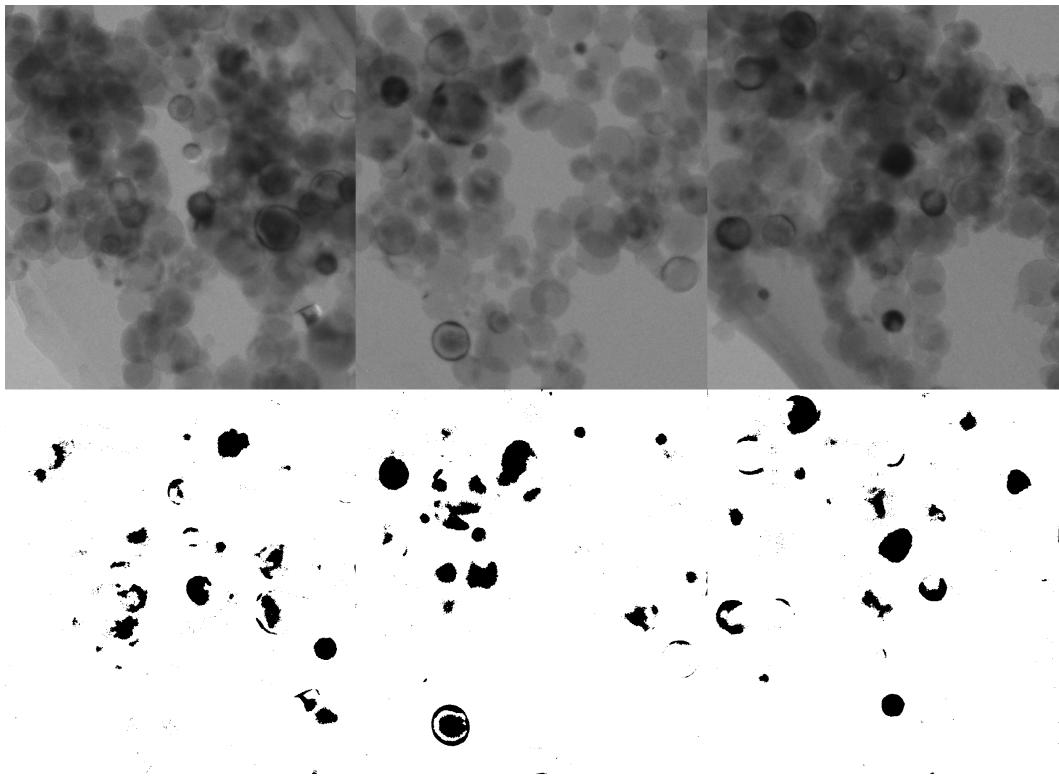
Fig. 3.10.: Data and corresponding anomaly classifier upon termination of the algorithm for the real data set corresponding to three different tilts. The white region in the classifier correspond to non-anomalous measurements and the black regions indicate an anomaly. While the classifier selects certain non-anomalous regions notice that the regions in the data with anomalies are accurately classified by the algorithm.

Titan TEM with a 300 kV accelerating voltage, and a spot size[1] of 5. The exposure time was set to 1 second, magnification was set to 100 kX, the frame size set to 2048×2048, with a pixel size of 0.83 nm $\times$ 0.83 nm. The detector used was a CCD with a 30 μm objective aperture resulting in a detector which captures electron scattered in the $0 - 15$ mrad range. The BF-TEM data consists of 33 tilts in the range of $-70°$ to $+70°$. We use a $\approx$ 580 nm $\times$580 nm section of the projection images for reconstruction. The dimensions of the reconstructed volume are set so as

---

[1]The spot size is a manufacturer dependent unit-less parameter that refers to the size of the condenser aperture and controls the electron flux on the sample.

to account for all the voxels contributing to the projection data. In presenting the results we only show voxels that can be reliably reconstructed from the projection data [75]. We reconstructed the data set using our algorithm (MBIR-AM-PE), FBP and conventional MBIR without anomaly modeling. All reconstructions are performed with voxels of size 0.83 nm × 0.83 nm × 0.83 nm.

Fig. 3.9 (a) and (b) show a $x - z$ and $x - y$ cross-section reconstructed from the data using FBP. The reconstruction has strong streaking artifacts in the $x - z$ plane and noise in the $x - y$ plane similar to the simulated data set. The reconstruction using the conventional MBIR algorithm (Fig. 3.9 (c)-(d)), also has streaking artifacts in the $x - z$ plane that are similar to those in the simulated data set of Fig. 3.6. However, the conventional MBIR result also significantly reduces streaking artifacts as compared to FBP. This is likely do to the fact that MBIR reduces the weighting of the highly attenuated projections corresponding to measurements with anomalous Bragg scatter. Fig. 3.9 (e) shows that using the anomaly modeling and parameter estimation reduces streaking in the $x - z$ plane. The arrows in Fig. 3.9 (d,f) indicate regions where the MBIR with anomaly modeling reduces the under-estimation as well as other artifacts in the $x - y$ cross-section compared to the conventional MBIR. The wall-clock time taken for the proposed MBIR reconstruction ($844 \times 4516 \times 1008$ voxels) using a node with two Intel Xeon-E5s (total of 16 cores) was approximately 9 hours and 40 minutes.

Fig. 3.10 shows the binary classifier mask b' along with 3 successive tilts from the real data upon termination of the reconstruction algorithm (MBIR-AM-PE). Notice that most of the anomalous measurements are successfully identified by the generalized Huber function at the end of the reconstruction. Similar to the simulated data set a few of the noisy measurements are also classified as anomalous but this does not effect the final quality of the reconstruction significantly.
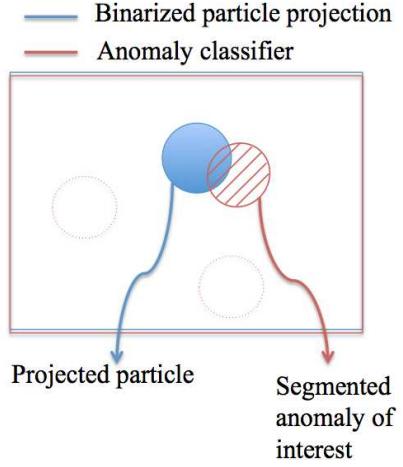
Fig. 3.11.: Illustration of calculation of the similarity score between a certain anomaly and the projection of a given segmented particle in the 3D volume. The projection is binarized and the score is then computed as the extent of overlap between the anomaly and the projection using (3.22).

3.4.3 Bragg Feature Extraction

While the particles that undergo Bragg diffraction in a given tilt result in anomalous measurements, the Bragg scatter event contains potentially useful information about the crystal structure and the orientation of the particles. For this reason it is advantageous to correlate the anomalous Bragg event identified in the acquired data with the particle in which it occurred, to produce a Bragg feature vector for each particle.

We will use the binary classifier mask, b', produced by our algorithm along with the reconstructed volume to associate a particle in the volume to an identified anomaly. In order to extract the Bragg feature vector for each particle we apply the following algorithm :

1) Segment the reconstructed volume into individual particles. We use a fixed threshold for segmentation followed by a watershed transformation [80] to separate the fused particles.

(a) Cross section from segmentation (MBIR-AM-PE)
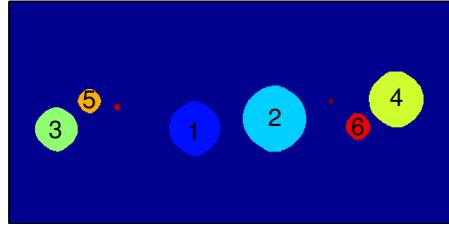
(b) Bragg feature vector extracted for particle (2)
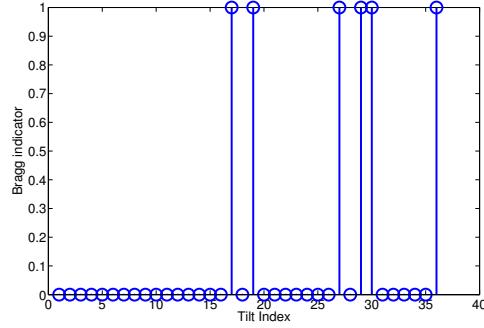
Fig. 3.12.: Illustration of Bragg feature identification for the simulated data set. (a) shows the output of segmentation from a single slice of the reconstruction. The particle labeled 2 was simulated to be in the Bragg condition at tilts indexed by 17, 19, 27, 29, 30 and 36. (b) shows the estimated Bragg feature vector for the particle labeled 2 using the proposed algorithm. In this case the estimated Bragg feature vector matches the ground truth indicating that the Bragg condition can be accurately identified.

2) Identify the connected components (CC) of the anomaly classifier.

3) For each identified CC in the anomaly classifier, project each particle, binarize the projection and find its similarity with that CC. Fig. 3.11 illustrates how to find the similarity between the binarized projection of a single particle and a particular CC anomaly at a given tilt. If p_k is a binarized version of the projection of a given particle at tilt k and b_k is the binary anomaly classifier with the relevant CC segmented out, we define the similarity score as

$$S_k = 1 - \frac{\|p_k^t \bar{b}_k\|_1 + \|\bar{p}_k^t b_k\|_1}{\|p_k\|_1 + \|b_k\|_1} \tag{3.22}$$

(a) Cross section from segmentation (MBIR-AM-PE)
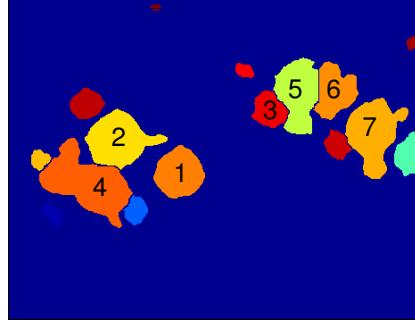
(b) Bragg feature vector extracted for particle (1)
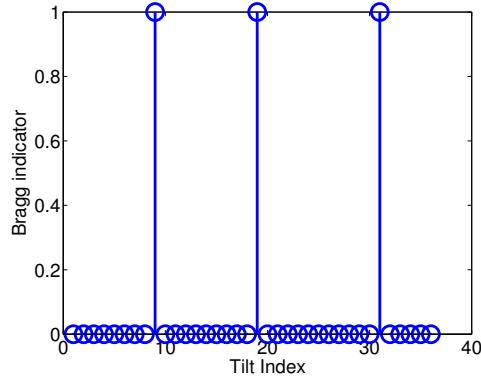
Fig. 3.13.: Illustration of Bragg feature identification for a single particle in a real data set. (a) shows the output of segmentation from a single slice of the reconstruction. (b) shows the Bragg feature vector for the particle labeled (1). While we do not have ground truth for this, we visually observed that the Bragg feature extracted matches what can be seen in the acquired tilt series.

where $\bar{b}_k$ and $\bar{p}_k$ refers to the binary complement operator and $\|.\|_1$ is the l_1 norm.

4) For a given CC anomaly (at tilt k), find the particle that has the maximum similarity score with it. If this score is higher than a threshold, we label that particle as being in the Bragg condition at tilt k.

In this manner we can associate a binary Bragg feature vector with each segmented particle from the MBIR reconstruction. In order to test our algorithm we apply it to the simulated data set as well as the real data presented earlier. Fig. 3.12 (a) shows the result of segmentation from a single reconstructed slice of the MBIR-AM-PE reconstruction in 3.4.1. In our simulation we had set the particle labeled 2 to

be in the Bragg condition at tilts indexed by $17, 19, 27, 29, 30$ and 36. Fig. 3.12 (b) shows the estimated Bragg feature vector for the particle labeled 2 by using the above algorithm. We observe that this matches the ground truth, illustrating the potential of the proposed technique.

Fig. 3.13 (a) and (b) show a similar result from the real data set. Note that in this case segmentation of the particles are very challenging. However for the particle labeled 1 we are still able to recover the Bragg feature vector and this matches our visual observation from the tilt series data. The results of Fig. 3.13 show that it is possible to extract potentially useful information about when a single particle is in the Bragg scattering condition. However, the method depends on the assumptions that the volume can be accurately segmented into individual particles corresponding to single crystal orientations.

3.5   Conclusion

In this paper we presented a MBIR algorithm for BF-ET which can significantly decrease the artifacts in the reconstruction due to anomalies such as Bragg scatter. Our method works by modeling the image formation and the sample being imaged to formulate a cost function that lowers the influence of measurements that do not fit the model accurately. Results on simulated and real data demonstrate that our method can effectively suppress the artifacts due to the anomalies, producing qualitatively and quantitatively accurate reconstructions. We also proposed a simple method for extracting a Bragg feature vector for each particle in a volume that contains potentially useful information about crystal orientation.

4. PLUG-AND-PLAY PRIORS FOR MODEL-BASED RECONSTRUCTION

4.1 Introduction

Model-based reconstruction is a powerful framework for solving a variety of inverse problems in imaging including denoising, deblurring, tomographic reconstruction, and MRI reconstruction. The method typically involves formulating a model for the noisy measurement system (i.e., a forward model) and a model for the image to be reconstructed (i.e., a prior model). The reconstruction is then computed by minimizing a cost function that balances a fit to these two models. For example, a typical approach is to compute the maximum *a posteriori* (MAP) estimate as the minimum of the sum of the log likelihood forward model and the log probability of the prior distribution.

In recent years, there have been enormous advances in the solution of a particular inverse problem generally referred to as image denoising [81, 82]. The problem of image denoising is to recover an image that has been corrupted by noise, the most commonly considered noise model being additive white Gaussian noise. Since image denoising is the simplest case of an inverse problem, the forward model being the identity operator, research in this field tends to provide a fertile environment for the creation of new prior models. Some denoising algorithms are based on an explicit regularized inversion formulation using, for example, a MAP estimate; but in other cases, the denoising algorithms are directly formulated as ad hoc non-linear estimates of the noiseless image. In fact, a number of very novel and effective approaches have recently emerged for image denoising. Examples of new methods include a wide variety of patch based non-local means approaches [83], generalizations of bi-lateral filtering approaches [81], patch-based dictionary learning methods such as K-SVD [6], block-matching with transform-based denoising such as BM3D [36], and a variety of

total-variation [82], and Markov random field (MRF) based approaches [7]. These new methods have demonstrated that it is possible to vastly improve on what was previously believed to be possible.

In parallel with these efforts, researchers have been pioneering ways to create forward models for a wide array of imaging and sensing systems from medical scanners [8] to microscopes [75]. Research in this field has demonstrated that model-based inverse methods can greatly improve the quality of reconstructed images [84]. However, since this research primarily deals with the challenges of accurately modeling large and complex forward models and solving the associated optimization problems, there has been much less emphasis on the incorporation of state-of-the-art prior models. Therefore, research in model-based inversion has tended to lag behind from the perspective of advanced prior modeling; and moreover, has not fully benefited from the recent progress in denoising methods.

In fact, recent progress has been made in incorporating advanced priors into general inverse problems. For example, patch based dictionary priors have been used in inverse imaging problems such as tomography [85, 86] and MRI [14]. Egizarian et al. [87] have proposed a method for using a general denoising algorithm for regularized inversion problems but their approach does not allow for the incorporation of noise models associated with the measurement system. Furthermore, while BM3D may not naturally lend itself to formulation as a prior, Danielyan et al. [88, 89] have adapted the BM3D [36] denoising for the inverse problem of image deblurring. However, this approach is not directly applicable to a general inverse problem. So, while some significant advances have been made in the integration of advanced prior and forward models, they tend to be highly customized to the problem and currently no simple turn-key approach exists to match denoising algorithms as regularizers for general inverse problems.

In this chapter, we propose a flexible framework for using denoising algorithms as regularizers for model-based inversion. This framework, which we term Plug-and-Play priors, has the advantage that it simplifies software integration, and moreover,

it allows state-of-the-art denoising methods that are not explicitly formulated as optimization problems to be used. For denoising algorithms based on well-behaved optimization criteria (e.g. - closed, proper and convex functions [90]), it can be easily shown that the Plug-and-Play framework is convergent to the MAP estimate of the reconstruction. In more general cases, we empirically demonstrate that the method converges robustly to a good solution. We show with some simple examples how Plug-and-Play priors can be used to mix and match a wide variety of existing denoising models with the type of forward model that is typically used in applications such as tomographic reconstruction and image reconstruction from a sparse set of samples. Consequently, this new approach can greatly expand the range of possible models used in model-based inversion.

Our proposed Plug-and-Play framework is based on a direct application of the alternating directions method of multipliers (ADMM) [90–93] that has recently become popular for the solution of a variety of MAP estimation/regularized inverse problems [9, 15, 94–97]. Our application of ADMM works by first splitting the state variable so as to decouple the prior and forward model terms of MAP estimation problem. The application of the ADMM technique to the resulting constrained minimization problem then results in two decoupled optimizations, one for the forward model and one for the prior model. We note that this allows for a completely decoupled software implementation with one module corresponding to a denoising algorithm only dependent on the prior, and a second module corresponding to a model-based inversion with l_2 regularization only dependent on the forward model. Importantly, this framework can be used to solve a regularized inversion problem even when the explicit function corresponding to the prior model is not known. Moreover, we also demonstrate empirically that the method can be used with denoising algorithms such as BM3D [36] and PLOW [37] that are not explicitly formulated as optimization problems.

In order to demonstrate the approach, we first apply a wide range of denoising algorithms as Plug-and-Play priors for a simple tomographic reconstruction problem

using the well-known Shepp-Logan phantom. The results indicate that methods such as K-SVD [6], BM3D [36], PLOW [37], q-GGMRF [56], TV [82], and discrete recon-struction (DR) [98] can all be easily applied as priors through direct use of software implementations of the denoising algorithms. Next, we apply the method to a real synchrotron X-ray tomography data set and compare the effect of different priors models/denoising algorithms on the reconstruction. Finally, we apply the algorithm to the problem of reconstructing a scanning TEM image from a sparse set of mea-surements to potentially reduce dose for beam sensitive specimens.

The organization of the rest of this chapter is as follows. In Section 4.2 we in-troduce the cost function corresponding to the MAP estimation for a general inverse problem. In Section 4.3 we briefly discuss the variable splitting and ADMM algo-rithm and use it to propose the Plug-and-Play priors method. In Section 4.4 we apply the algorithm on simulated and real data sets for various applications while ex-perimenting with different denoising algorithms (priors) and in Section 4.5 we draw our conclusions.

4.2 MAP cost function for solving inverse problems

Let g be a $M \times 1$ measurement vector from which we desire to estimate the unknown f, a $N \times 1$ vector. Let $p(g|f)$ be the conditional probability density function (pdf) of the measurements g given f, and $p(f)$ be the pdf of the unknown, then the MAP estimate of f is given by

$$\hat{f} \quad \leftarrow \quad \underset{f}{\operatorname{argmin}} \left\{ -\log p(g|f) - \log p(f) \right\}$$
$$\hat{f} \quad \leftarrow \quad \underset{f}{\operatorname{argmin}} \left\{ l(g; f) + s(f) \right\} \tag{4.1}$$

where $l(g; f) = -\log p(g|f)$ and $-\log p(f) = s(f) +$ terms independent of f. In the special case of $l(g; f) = \frac{1}{2\sigma_n^2} \|g - f\|_2^2 + \frac{M}{2} \log\left(2\pi\sigma_n^2\right)$ the MAP estimate corresponds to denoising designed to remove additive white Gaussian noise of variance σ_n^2. For this

special case, we define $H(g; \sigma_n^2)$ to be the operator that denoises the signal g when it has been corrupted by additive Gaussian noise of variance σ_n^2. This operator is then given by the solution to the following MAP optimization problem:

$$\mathbb{H}(g; \sigma_n^2) = \operatorname*{argmin}_{f} \left\{ \frac{1}{2\sigma_n^2} \|g - f\|_2^2 + s(f) \right\}. \tag{4.2}$$

Sometimes it is useful to have an additional regularization parameter to control the relative effect of the prior model on the reconstruction. To allow for this additional control, we can rewrite the estimation problem as

$$\hat{f} \quad \leftarrow \quad \operatorname*{argmin}_{f} \left\{ l(g; f) + \beta s(f) \right\}, \tag{4.3}$$

where β can be used to modulate the amount of regularization applied to the inversion. Notice that when $\beta = 1$ the problem is exactly the MAP estimation problem (4.1).

4.3   Variable Splitting and ADMM

In order to design an algorithm for (4.3) that decouples the forward and prior terms, we first split the variable f into two new variables f and v, and reformulate equation (4.3) as the following constrained optimization problem [94, 95].

$$(\hat{f}, \hat{v}) \quad \leftarrow \quad \operatorname*{argmin}_{f,v} \left\{ l(g; f) + \beta s(v) \right\}$$
$$\text{subject to } f = v. \tag{4.4}$$

We then solve (4.4) by forming the augmented Lagrangian function and using the ADMM technique [90–93]. The augmented Lagrangian for this problem is given by

$$L_\lambda(f, v, u) = l(g; f) + \beta s(v) + \frac{\lambda}{2} \|f - v + u\|_2^2 - \frac{\lambda}{2} \|u\|_2^2. \tag{4.5}$$

where u is a scaled dual variable and λ is the penalty parameter. The ADMM algorithm consists of repeatedly performing the following steps until convergence.

$$
\begin{aligned}
\hat{f} &\leftarrow \underset{f}{\operatorname{argmin}}\, L_\lambda(f, \hat{v}, u) \\
\hat{v} &\leftarrow \underset{v}{\operatorname{argmin}}\, L_\lambda(\hat{f}, v, u) \\
u &\leftarrow u + (\hat{f} - \hat{v})\,.
\end{aligned}
$$

Notice that for convex problems, λ does not effect the final result but controls the rate of convergence of the ADMM algorithm.

If $\tilde{f} = \hat{v} - u$ and $\tilde{v} = \hat{f} + u$ then each iteration of the algorithm can be written as

$$
\hat{f} \leftarrow \underset{f}{\operatorname{argmin}} \left\{ l(g; f) + \frac{\lambda}{2} \|f - \tilde{f}\|_2^2 \right\} \tag{4.6}
$$

$$
\hat{v} \leftarrow \underset{v}{\operatorname{argmin}} \left\{ \frac{\lambda}{2} \|\tilde{v} - v\|_2^2 + \beta s(v) \right\} \tag{4.7}
$$

$$
u \leftarrow u + (\hat{f} - \hat{v}). \tag{4.8}
$$

The first step only depends on the choice of forward model. The second step only depends on the choice of prior and can be interpreted as a denoising operation as in equation (4.2).

In order to emphasize the modular structure of the ADMM update, we define the operator $\mathbb{F}(g, \tilde{f}; \lambda)$ as

$$
\mathbb{F}(g, \tilde{f}; \lambda) = \underset{f}{\operatorname{argmin}} \left\{ l(g; f) + \frac{\lambda}{2} \|f - \tilde{f}\|_2^2 \right\}. \tag{4.9}
$$

This function returns the MAP estimate of f given the data g, using very simple quadratic regularization to a value, $\tilde{f}$. We call $\mathbb{F}$ a simplified reconstruction operator. Notice that $\mathbb{F}$ is also the proximal mapping [99] associated with the function $\frac{1}{\lambda} l(g; f)$. Using our definition of the simplified reconstruction operator $\mathbb{F}(g, \tilde{f}; \lambda)$ from (4.9), and

our definition of the denoising operator $\mathbb{H}(g; \sigma_n^2)$ from (4.2), we may now reformulate the ADMM iterations as the following three steps.

$$\hat{f} \leftarrow \mathbb{F}(g, \tilde{f}; \lambda) \tag{4.10}$$

$$\hat{v} \leftarrow \mathbb{H}(\tilde{v}; \frac{\beta}{\lambda}) \tag{4.11}$$

$$u \leftarrow u + (\hat{f} - \hat{v}). \tag{4.12}$$

The overall algorithm is summarized in Fig. 4.1. Importantly, using this Plug-and-Play framework, the minimization can now be written as two independent software modules - one for implementing the simplified reconstruction operator $\mathbb{F}(g, \tilde{f}; \lambda)$ and the other for implementing the denoising algorithm $\mathbb{H}(\tilde{v}; \sigma_n^2)$. Furthermore changing the prior model only involves changing the implementation of $\mathbb{H}(\tilde{v}; \sigma_n^2)$. Thus the Plug-and-Play priors framework can be used to mix and match different denoising algorithms (priors) with the forward model of interest. Notice that the minimization corresponding to the simplified reconstruction operator and the denoising operator need not be exact. Instead, they can be replaced by the approximate operators $\tilde{\mathbb{F}}$ and $\tilde{\mathbb{H}}$ that do not minimize the respective cost functions but instead decrease its value sufficiently. This is an important technique for speeding up the implementation of the ADMM [90] and making the algorithm useful in practical applications.

We note that the variable splitting approach discussed here has been exploited to solve a variety of inverse problems [94, 95, 97, 100]. However the main motivation of this research was to create better algorithms for solving the optimization problems resulting from regularized inversion. For example, this variable splitting/ADMM approach has been used to more effectively solve problems with l_1 norms, TV norms, and positivity constraints that can create difficulties in conventional gradient based optimization. In distinction to this earlier research, our primary goal is to use splitting strategies as a mechanism to create a flexible framework to easily match prior models (embodied in the form of denoising algorithms) with forward models.

```
function [f̂] ← RECONSTRUCT(g, β, λ)
    % Inputs: Measurements g, Regularization β, Augmented Lagrangian pa-
rameter λ
    Initialize f̂
    v̂ ← f̂, u ← 0                                          ▷ Initialize v̂ and u
    σ²_n ← β/λ                                    ▷ Variance for denoising algorithm
    while Stopping criteria are not met do
        f̃ ← v̂ − u
        f̂ ← 𝔽(g, f̃; λ)                          ▷ Only dependent on forward model
        ṽ ← f̂ + u
        v̂ ← ℍ(ṽ; σ²_n)                 ▷ Denoising operator only dependent on prior
        u ← u + (f̂ − v̂)                    ▷ Update the scaled dual variable u
    end while
end function
```

Fig. 4.1.: Pseudocode for Plug-and-Play priors framework. In each iteration an alternating minimization is done. The first minimization depends only on the likelihood function while the second minimization only requires the application of a denoising algorithm. Thus introducing a new prior only requires introducing a new denoising software module.

Finally, we note that in this chapter we do not discuss theoretical convergence properties of the Plug-and-Play framework. While the ADMM is guaranteed to converge if l and s are convex, closed and proper functions and L_0 has a saddle point [90], we observe via our numerical experiments that substituting $\mathbb{H}$ with denoising algorithms that do not explicitly correspond to a convex function s or even a strict optimization problem, still produces a stable result. Thus we rely on empirical evidence from our experiments to show that our framework produces a stable result.

4.4 Experimental Results

In this section, we will use the Plug-and-Play priors for two applications - tomography and image reconstruction from a sparse set of measured pixels. We will first present results on a simulated tomography dataset to demonstrate the empirical convergence properties of the method and how it compares to a direct implementation

in case of a prior model that corresponds to a well defined function. Next, we will apply the method to a real synchrotron X-ray tomography data set and compare the reconstructions using different prior models. Finally, we will apply the method to the problem of recovering a scanning TEM image from a sparse set of measured pixels.

4.4.1   Tomography

In the case when there are no missing calibration parameters or anomalous measurements, the log-likelihood for the electron tomography problem has the form $l(g; f) = \frac{1}{2}\|g - Af\|_\Lambda^2$, where A is a tomographic forward projector, and Λ is a diagonal weighting matrix with entries corresponding the to inverse variance of the measurements. This type of model also occurs in synchrotron X-ray tomography [28]. We will match this type of forward model with a variety of state-of-the-art denoising algorithms which may or may not explicitly be formulated as prior models in a regularized inversion framework.

Simulated Data

We evaluate the Plug-and-Play method on a 64×64 Shepp-Logan phantom with values scaled between $0-255$. The phantom is forward projected at 141 views between $-70°$ and $+70°$ and the noise is simulated to be Poisson. This type of forward model is widely used in electron tomography [55, 101], where due to mechanical constraints, the sample can only be tilted in a limited range. We compare reconstructions using the Plug-and-Play priors framework by experimenting with six different denoising techniques/priors - K-SVD [6], BM3D [36], PLOW [37], Total Variation (TV) [102], q-GGMRF [56] and discrete reconstruction (DR) [98]. The regularization parameter β is adjusted for achieving the minimum root mean square error (RMSE) between the reconstruction and phantom. Instead of using the simplified reconstruction operator $\mathbb{F}$ in the ADMM loop, we use an approximate operator $\tilde{\mathbb{F}}$, which lowers the value of the cost function corresponding to $\mathbb{F}$ using N_{Iter} number of iterations of iterative

coordinate descent (ICD) [53] with random order updates [8]. The algorithm is initialized with a filtered back projection reconstruction. The value of N_{Iter} is set to 1 for all algorithms except the DR prior in which case it is set to 20. The value of λ is set to 1/20 for all experiments. For non-convex priors like the K-SVD and the DR prior we observed that the value of λ effects the final solution. The algorithm is terminated after 150 iterations. Further details of the parameters used for different denoising algorithms are given in Appendix D.

Fig. 4.2 shows the reconstructions resulting from the use of the six denoising algorithms as regularizers, and Table. 4.1 shows the corresponding RMSE for each regularizer. For this very simple Shepp-Logan image, the DR prior results in the lowest RMSE. However, the other methods result in a comparable RMSE. Most importantly, each denoising algorithm was easily matched to the tomographic forward model and for each prior, the convergence in terms of RMSE to the original phantom is stable and robust (see Fig. 4.3). Another measure of convergence in the ADMM is that the $\|\hat{f} - \hat{v}\|$ (primal residual) and the change in $\hat{v}$ in subsequent iterations (dual residual) [90] must go to zero as the number of iterations increase. Fig. 4.4 shows the plots the primal and dual residual normalized by the number of pixels for each of the denoising algorithms. We observe that even for denoising algorithms that are not formulated as optimization problems these residuals decrease and hence provide empirical evidence that the framework can be useful. However, the error in the constraint being met (primal residual) is greater in the case of K-SVD, BM3D,PLOW and DR than in the TV and qGGMRF case. Interestingly, the BM3D [36] and PLOW [37] algorithms do not explicitly correspond to a prior model, so the Plug-and-Play methodology provides a simple and robust framework to use them for regularized inversion problems.

Finally we compare the convergence of the ADMM technique to the direct implementation of the MAP estimate using a q-GGMRF prior model that is tightly integrated into the cost function [56]. The traditional approach of tight integration allows for more flexibility in the design of optimization algorithms, so it might be

(a) Ground Truth

(b) K-SVD

(c) BM3D

(d) PLOW

(e) TV

(f) q-GGMRF

(g) Discrete reconstruction
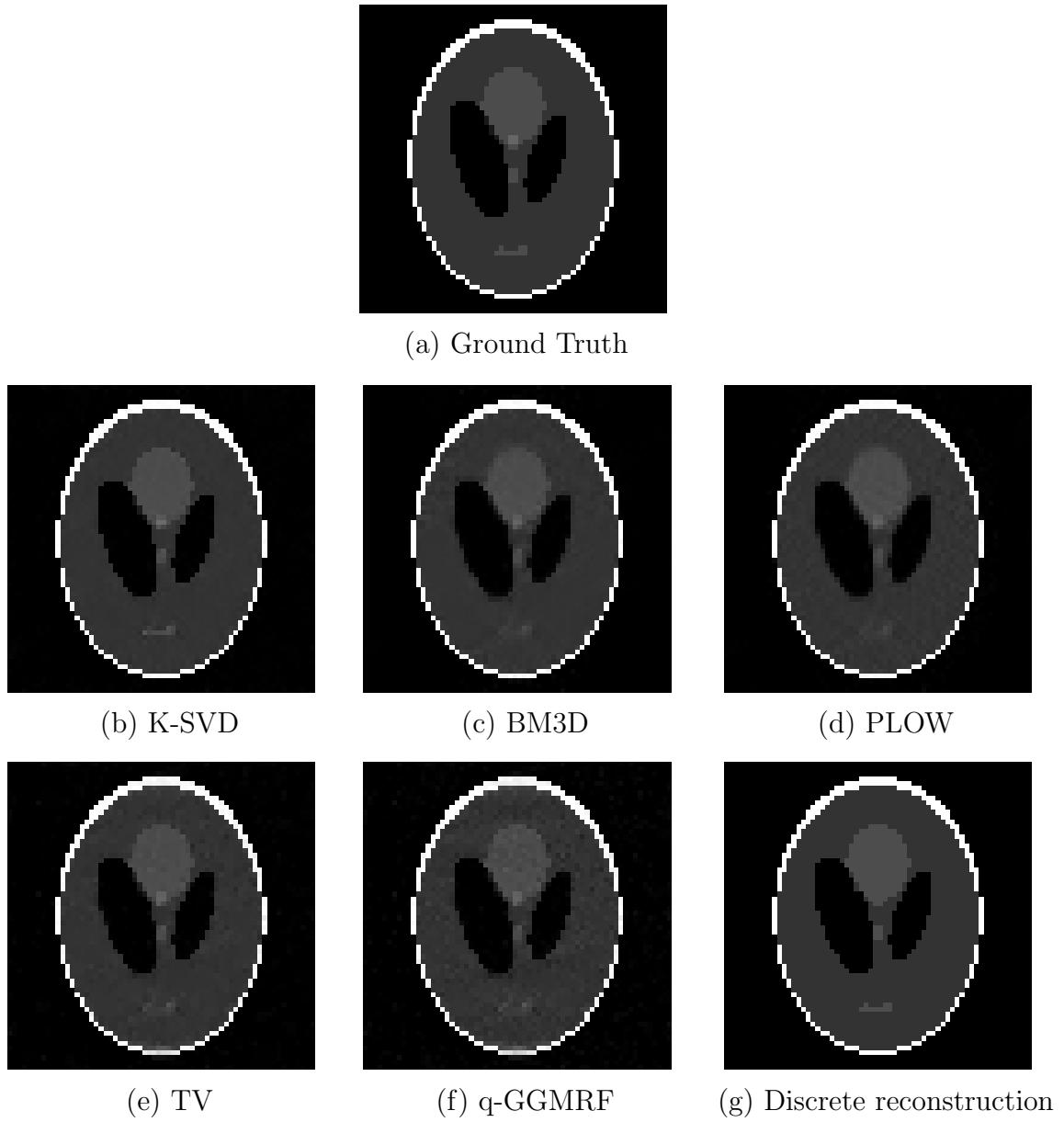
Fig. 4.2.: Comparison of the minimum RMSE reconstructions using different priors for the Shepp-Logan phantom projected in a limited angular range $(+/-70°)$. All images are displayed in the window $[0-255]$. (a) Phantom (b) K-SVD (c) BM3D (d) PLOW (e) TV (f) q-GGMRF (g) Discrete reconstruction. We observe that the patch based denoising algorithms (b) - (d) work well producing qualitatively comparable reconstructions to the typically used priors like TV and q-GGMRF. Some of the features in the phantom are not reconstructed accurately due to the limited angle nature of the projection data.
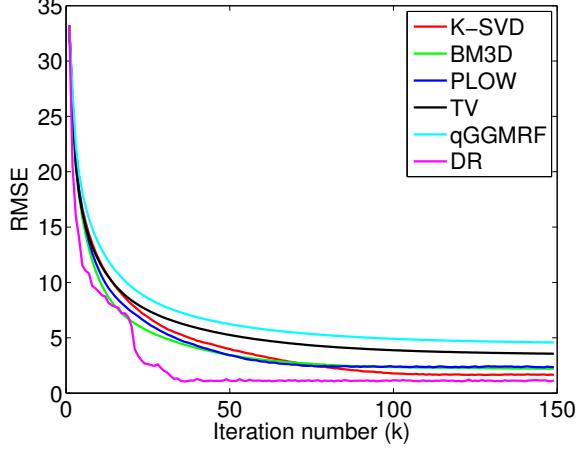
Fig. 4.3.: Comparison of the convergence (RMSE between the reconstruction and the original phantom) as a function of iteration number for the different denoising algorithms used. The original phantom has values in the range $0 - 255$. We note that the convergence for all algorithms is robust and stable. Furthermore the convergence rates across the different denoising algorithms are similar.
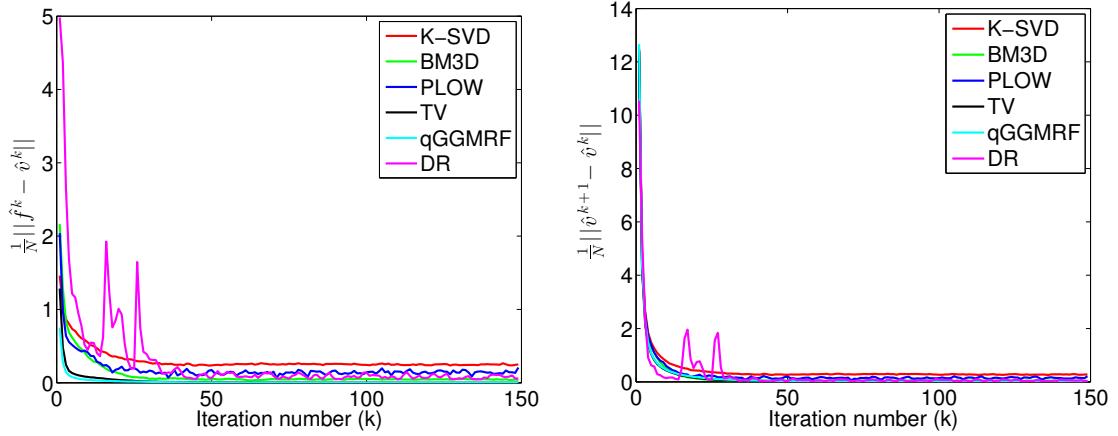
Fig. 4.4.: Comparison of the primal residual ($\frac{1}{N}\|\hat{f}^k - \hat{v}^k\|$) and dual residual ($\frac{1}{N}\|\hat{v}^{k+1} - \hat{v}^k\|$) as a function of iteration number for the different denoising algorithms used. We note that the convergence for all algorithms is robust and stable.

expected to have faster convergence; but this faster convergence is at the cost of a less modular and flexible design. Nonetheless, one would expect that both the Plug-and-Play and traditional formulation should yield the same result when run

Table 4.1: Comparison of the minimum Root Mean Square Error of the reconstruction with respect to the original phantom for various priors. The original phantom has values in the range $0-255$. We observe that the the patch based nonlocal denoising operators result in a low RMSE reconstruction.

Algorithm	RMSE
K-SVD [6]	1.67
BM3D [36]	2.20
PLOW [37]	2.34
TV [102]	3.56
q-GGMRF [56]	4.57
Discrete Recon [98]	1.10

to convergence. Moreover, for this case, the denoising algorithm corresponds to the minimization of a convex function, and hence the ADMM technique is known to be globally convergent provided that the inner minimizations are either exact or are run for a sufficiently large number of iterations [90]. Fig. 4.5 shows the RMSE as a function of iteration number for the tightly integrated approach where the ICD algorithm is used to solve the tomographic inversion compared to the two step ADMM approach with a q-GGMRF denoising prior. We set the number of inner iterations in the q-GGMRF denoising step in ADMM to 1 for a fair comparison. While the ADMM approach is slower than ICD for our implementation and requires greater memory to store the auxiliary variables [9, 97], it provides a more flexible framework for incorporating different priors. Furthermore, several techniques that have been proposed to speed up the ADMM algorithm [9, 90] can also be applied to the Plug-and-Play priors framework.

Real Data

In this section, we apply the Plug-and-Play priors to a real data from a synchrotron X-ray tomography (SXT) experiment. SXT corresponds to a parallel beam tomography modality and has the same form of likelihood function as the simulated
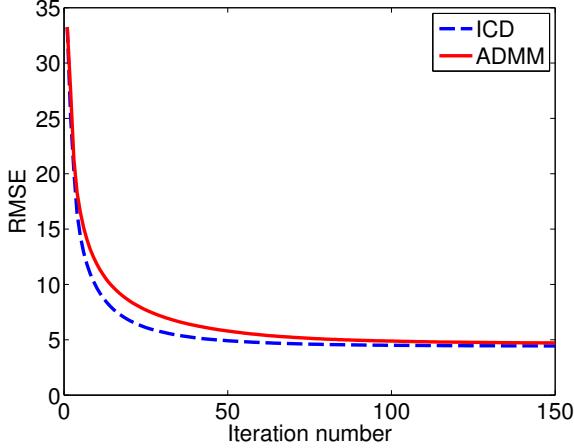
Fig. 4.5.: Comparison of the convergence of ICD versus ADMM for the q-GGMRF prior with $p = 2, q = 1.2, c = 1/100, \sigma_f = 0.594$. The convergence is measured by using the RMSE between the reconstruction and the original phantom. The number of inner iterations in the ADMM is set to 1 for a fair comparison. We observe that ICD converges faster but ADMM's speed is comparable. However the time taken per iteration in ADMM will be higher due to the two step minimization.

data set discussed above. The data acquired is of a coarsening experiment of an Aluminum-Copper alloy at the Swiss Light Source [103]. The CCD detector has 1024×1024 pixels. The data acquired consists of 1024 views between $0°$ and $180°$. We will reconstruct a single 2-D slice of the 3-D volume from the data set using the different prior models/denoising algorithms use for the simulated data set. In order to reduce the computation, the acquired data is subsampled in the view space, and in the detector direction resulting in a sinogram with 141 views and 512 detector entries. The algorithm is terminated after 100 iterations. The value of the regularization parameters are adjusted for achieving a good visual quality of reconstruction. Details of the parameters used for different denoising algorithms are given in Appendix D.

Fig. 4.6 (a) shows the FBP reconstruction from a single slice of the object. Fig. 4.6 (b)-(g) shows the MBIR reconstruction using different denoising algorithms. Notice that the MBIR reconstructions produce a less noisy and sharper reconstruction than the FBP case. Furthermore, all the priors produce a satisfactory reconstruction.

(a) Filtered Back Projection

(b) K-SVD   (c) BM3D   (d) PLOW

(e) TV   (f) q-GGMRF   (g) Discrete reconstruction
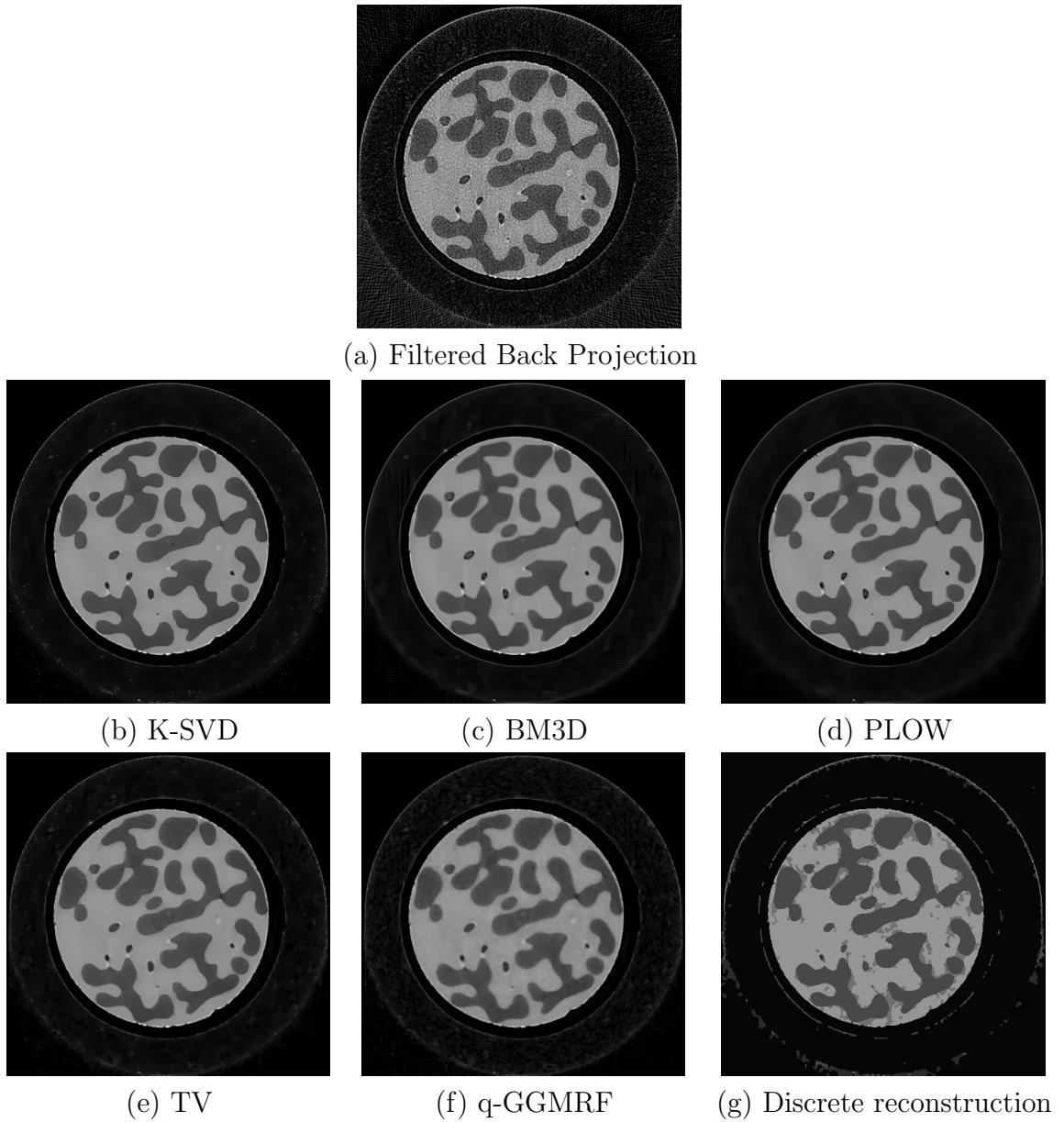
Fig. 4.6.: Comparison of tomographic reconstructions using different priors for a real data set acquired from a synchrotron X-ray set up. All images are displayed in the same viewing window. (a) FBP reconstruction (b) K-SVD (c) BM3D (d) PLOW (e) TV (f) q-GGMRF (g) Discrete reconstruction. We observe that the patch based denoising algorithms (b) - (d) work well producing qualitatively comparable reconstructions to the typically used priors like TV and q-GGMRF. The discrete prior (g) produces a reconstruction with some artifacts, most likely due to the algorithm getting stuck in a local minimum.

Table 4.2: Comparison of the minimum Root Mean Square Error of the reconstruction with respect to the original *cameraman* image for various priors. We observe that the the patch based nonlocal denoising operators result in a low RMSE reconstruction.

Algorithm	RMSE
K-SVD [6]	12.49
BM3D [36]	11.97
PLOW [37]	13.00
TV [102]	15.16
q-GGMRF [56]	15.53

However, note that in the DR case there are some artifacts because of the algorithm getting stuck in undesirable local minima. This experiment demonstrates that the Plug-and-Play framework can be used to incorporate different prior models for SXT reconstructions potentially enabling improvements in image quality or reduction in the amount of data acquired at the synchrotron facilities.

4.4.2 Sparse Image Reconstruction

Image interpolation/inpainting from a sparse set of measurements has been hypothesized to be useful for reducing dose in transmission electron microscopy [104] and for reducing the acquisition time in scanning electron microscopy [105]. Here we will apply the Plug-and-Play priors for the sparse image reconstruction problem for a simulated data set consisting of the *cameraman* image with only 40% of the pixels measured. We assume the measurements are corrupted by Gaussian noise with variance σ_w^2 set to approximately 10% of the maximum value in the original image. The likelihood function is given by $l(g; f) = \frac{1}{2\sigma_w^2}\|g - Af\|_2^2$, where A is a $M \times N$ matrix which selects a subset of pixels from the underlying image f. Thus the entries of A

(a) Ground Truth   (b) Sampled, noisy data

(c) K-SVD   (d) BM3D   (e) PLOW

(f) TV   (g) q-GGMRF
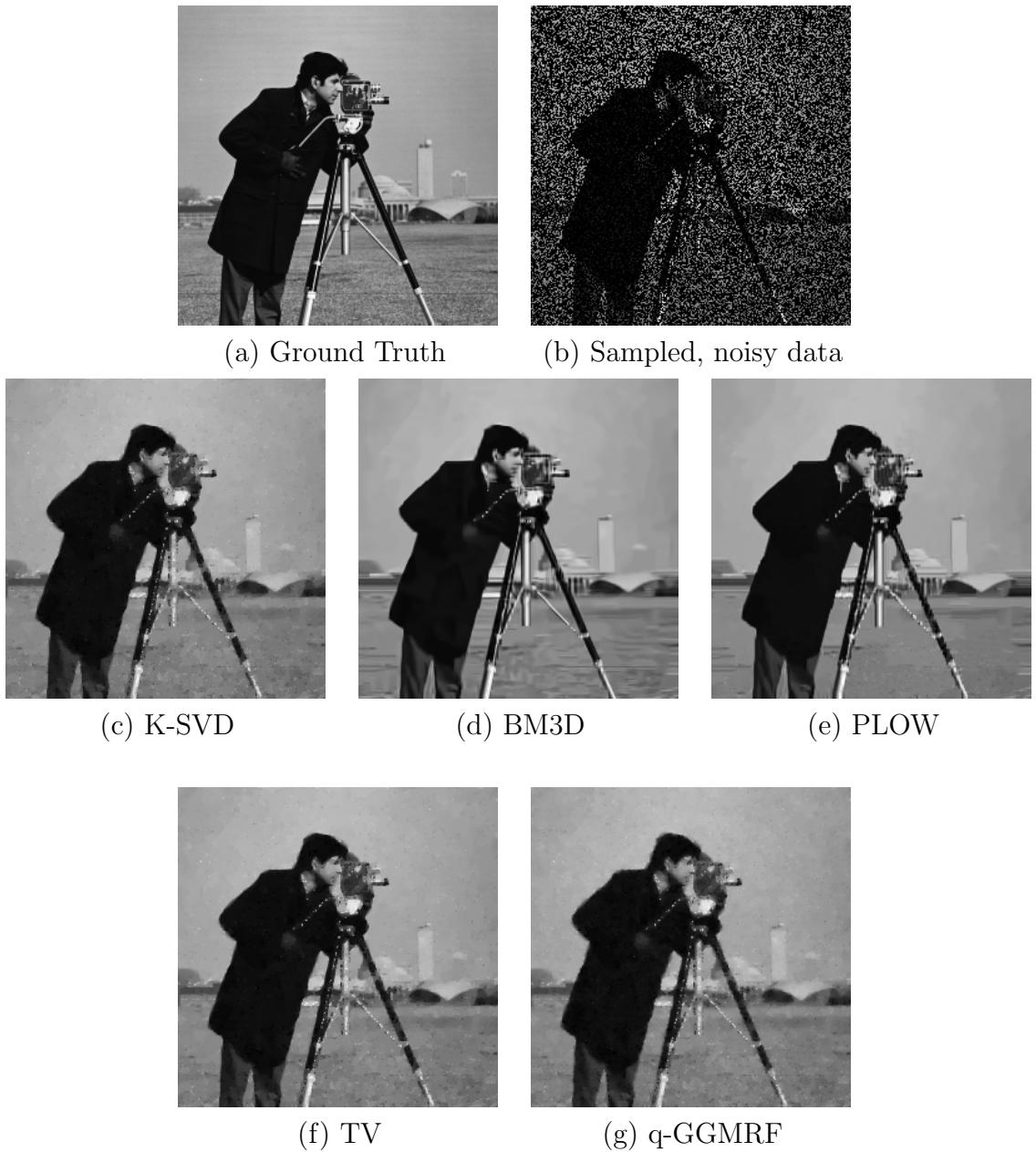
Fig. 4.7.: Comparison of the minimum RMSE reconstructions using different priors for the *Cameraman* image sampled at 40% of the pixels with Gaussian noise of standard deviation 10. All images are displayed in the window $[0 - 255]$. (a) Phantom (b) Measured data (c) K-SVD (d) BM3D (e) PLOW (f) TV (g) q-GGMRF reconstruction. We observe that the patch based denoising algorithms (c) - (e) produce qualitatively better reconstructions than priors like TV and q-GGMRF.
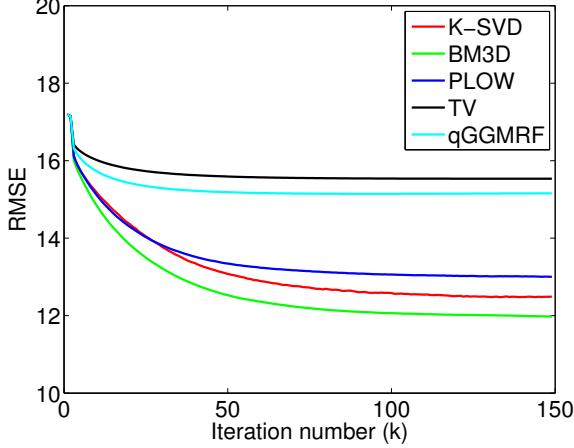
Fig. 4.8.: Comparison of the convergence (RMSE between the reconstruction and the original phantom) as a function of iteration number for the different denoising algorithms used. The original phantom has values in the range $0 - 255$. We note that the convergence for all algorithms is robust and stable. In this case the patch based denoising methods achieve a lower RMSE than the TV and qGGMRF models.
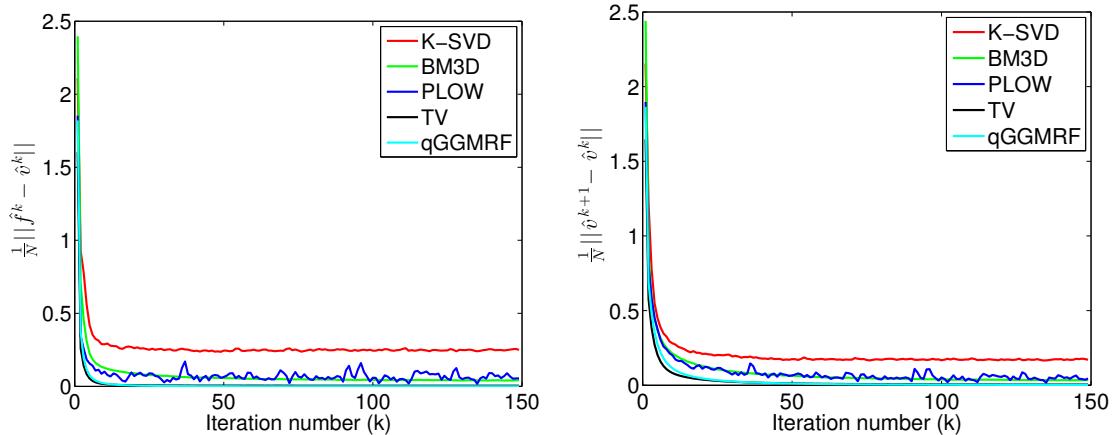
Fig. 4.9.: Comparison of the primal residual ($\frac{1}{N}\|\hat{f}^k - \hat{v}^k\|$) and dual residual ($\frac{1}{N}\|\hat{v}^{k+1} - \hat{v}^k\|$) as a function of iteration number for the different denoising algorithms used. We note that the convergence for all algorithms is robust and stable.

are either 1 or 0. Using this system matrix, the simplified reconstruction operator can be computed as

$$\mathbb{F}(g, \tilde{f}; \lambda) = \left[\frac{1}{\sigma_w^2} A^t A + \lambda I \right]^{-1} \left[A^t g + \lambda \tilde{f} \right] \tag{4.13}$$

where I is a $N \times N$ identity matrix. Due to the simple form of the A matrix this expression can be exactly evaluated during the reconstruction. We will reconstruct the image using five denoising algorithms - K-SVD [6], BM3D [36], PLOW [37], TV [102] and q-GGMRF [56]. The algorithm is terminated after 150 iterations. We initialize the algorithm using Shepard's method [106]. The value of the regularization parameters are adjusted for achieving a good visual quality of reconstruction. Further details of the parameters used for different denoising algorithms are given in Appendix D.

Fig. 4.7 shows the reconstruction using the different prior models. The patch-based non-local denoising algorithms (K-SVD, BM3D and PLOW) produce a qualitatively and quantitatively (see Table 4.2) more accurate reconstruction for this image, reinforcing that for natural images the patch based non-local models are good regularizers. Fig. 4.8 and Fig. 4.9 show the convergence of the different methods as a function of the number of iterations. The convergence across all methods is robust and stable. However, the error in the constraint being met (primal residual) is greater in the case of K-SVD, BM3D and PLOW than in the TV and qGGMRF case. This example demonstrates the utility of the Plug-and-Play framework to incorporate sophisticated image models for the sparse image reconstruction problem to enable high quality reconstructions via better prior models.

4.5 Conclusions

In this chapter, we proposed a flexible framework that allows state-of-the-art forward models of imaging systems to be matched with state-of-the-art prior or denoising models. The framework, which is based on variable splitting and use of the ADMM algorithm, simplifies the software architecture by decoupling the forward and prior models. Furthermore the framework enables state-of-the-art denoising algorithms, even those that have no known formulation as an optimization problem, to be used as priors/regularizers for model based inversion. We demonstrated the utility of the

method in two applications relevant to materials imaging - tomography and sparse image reconstruction.

LIST OF REFERENCES

LIST OF REFERENCES

[1] National Science and Technology Council, Office of the President of the United States, "Materials genome initiative for global competitiveness." `http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf`, June 2011.

[2] P. Midgley and M. Weyland, "3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography," *Ultramicroscopy*, vol. 96, no. 34, pp. 413 – 431, 2003.

[3] J. Frank, *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell.* Springer, 2008.

[4] R. Mizutani and Y. Suzuki, "X-ray microtomography in biology," *Micron*, vol. 43, no. 23, pp. 104 – 115, 2012.

[5] H. A. Bale, A. Haboub, A. A. MacDowell, J. R. Nasiatka, D. Y. Parkinson, B. N. Cox, D. B. Marshall, and R. O. Ritchie, "Real-time quantitative imaging of failure events in materials under load at temperatures above 1,600° C," *Nature Materials*, vol. 12, no. 1, 2013.

[6] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.

[7] E. Haneda and C. Bouman, "Implicit priors for model-based inversion," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 3917–3920, 2012.

[8] Z. Yu, J. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. on Image Processing*, vol. 20, pp. 161 –175, Jan. 2011.

[9] S. Ramani and J. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 3, pp. 677–688, 2012.

[10] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Physics in Medicine and Biology*, vol. 51, no. 15, p. R541, 2006.

[11] R. Zhang, J. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *Medical Imaging, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.

[12] D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Proc. Magazine*, vol. 18, pp. 57 –75, Nov. 2001.

[13] J. C. Ye, C. Bouman, K. Webb, and R. Millane, "Nonlinear multigrid algorithms for Bayesian optical diffusion tomography," *IEEE Trans. on Image Processing*, vol. 10, pp. 909 –922, June 2001.

[14] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly under-sampled k-space data by dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 30, pp. 1028 –1041, May 2011.

[15] S. Ramani and J. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 3, pp. 694–706, 2011.

[16] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a bayesian approach of solving inverse problems," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 1, pp. 473 –476 vol.2, Sept. 1996.

[17] K. Vernon-Parry, "Scanning electron microscopy: an introduction," *III-Vs Review*, vol. 13, no. 4, pp. 40 – 44, 2000.

[18] M. De Graef, *Introduction to Conventional Transmission Electron Microscopy.* Cambridge University Press, 2003.

[19] H. Li, H. L. Xin, D. A. Muller, and L. A. Estroff, "Visualizing the 3D internal structure of calcite single crystals grown in agarose hydrogels," *Science*, vol. 326, no. 5957, pp. 1244–1247, 2009.

[20] P. R. Buseck, R. E. Dunin-Borkowski, B. Devouard, R. B. Frankel, M. R. McCartney, P. A. Midgley, M. Psfai, and M. Weyland, "Magnetite morphology and life on Mars," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13490–13495, 2001.

[21] H. Wei, Z. Wang, J. Zhang, S. House, Y.-G. Gao, L. Yang, H. Robinson, L. H. Tan, H. Xing, C. Hou, I. M. Robertson, J.-M. Zuo, and Y. Lu, "Time-dependent, protein-directed growth of gold nanoparticles within a single crystal of lysozyme," *Nature Nanotechnology*, vol. 6, no. 2, 2011.

[22] A. E. Porter, M. Gass, K. Muller, J. N. Skepper, P. A. Midgley, and M. Welland, "Direct imaging of single-walled carbon nanotubes in cells," *Nature Nanotechnology*, vol. 2, pp. 713 – 717, 2007.

[23] S. Bals, C. F. Kisielowski, M. Croitoru, and G. V. Tendeloo, "Annular dark field tomography in TEM," *Microscopy and Microanalysis*, vol. 11, pp. 2118–2119, 8 2005.

[24] P. Ercius, M. Weyland, D. A. Muller, and L. M. Gignac, "Three-dimensional imaging of nanovoids in copper interconnects using incoherent bright field tomography," *Applied Physics Letters*, vol. 88, no. 24, p. 243116, 2006.

[25] R. L. Stevenson and E. J. Delp, "Fitting curves with discontinuities," in *Proc. of the first IEEE Int. Workshop on Robust Computer Vision*, pp. 127–136, Oct. 1990.

[26] J. Zheng, S. Saquib, K. Sauer, and C. Bouman, "Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence," *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1745–1759, 2000.

[27] H. Erdogan and J. Fessler, "Monotonic algorithms for transmission tomography," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 9, pp. 801–814, 1999.

[28] K. A. Mohan, S. V. Venkatakrishnan, L. F. Drummy, J. Simmons, D. Y. Parkinson, and C. A. Bouman, "Model-based iterative reconstruction for synchrotron X-ray tomography," accepted for IEEE ICASSP 2014.

[29] K. Sauer and C. Bouman, "Bayesian Estimation of Transmission Tomograms Using Segmentation Based Optimization," *IEEE Trans. on Nuclear Science*, vol. 39, pp. 1144–1152, 1992.

[30] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, pp. 534–548, Feb. 1993.

[31] J. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 290 –300, June 1994.

[32] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, pp. 4526–4544, 2007.

[33] I. Stojanovic, H. Pien, S. Do, and W. C. Karl, "Low-dose X-ray CT reconstruction based on joint sinogram smoothing and learned dictionary-based representation," in *9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1012 –1015, May 2012.

[34] S. Ramani and J. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. on Medical Imaging*, vol. 31, pp. 677 –688, Mar. 2012.

[35] M. Kamasak, C. Bouman, E. Morris, and K. Sauer, "Direct reconstruction of kinetic parameter images from dynamic PET data," *IEEE Trans. on Medical Imaging*, vol. 24, pp. 636 –650, May 2005.

[36] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[37] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1635–1649, 2012.

[38] M. Bárcena and A. J. Koster, "Electron tomography in life science," *Seminars in Cell & Developmental Biology*, vol. 20, no. 8, pp. 920 – 930, 2009.

[39] Z. Saghi, D. J. Holland, R. Leary, A. Falqui, G. Bertoni, A. J. Sederman, L. F. Gladden, and P. A. Midgley, "Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography (CS-ET) approach," *Nano Letters*, vol. 11, no. 11, pp. 4666–4673, 2011.

[40] B. Goris, W. Van den Broek, K. J. Batenburg, H. H. Mezerji, and S. Bals, "Electron tomography based on a total variation minimization reconstruction technique," *Ultramicroscopy*, vol. 113, pp. 120 – 130, 2012.

[41] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science.," *Nature Materials*, vol. 8, no. 4, pp. 271–280, 2009.

[42] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Philadephia, PA: Society for Industrial and Applied Mathematics, 2001.

[43] K. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. Midgley, J. Hernandez, U. Kaiser, E. Encina, E. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730 – 740, 2009.

[44] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. on Medical Imaging*, vol. 31, pp. 2194–2204, Dec. 2012.

[45] Z. H. Levine, A. J. Kearsley, and J. G. Hagedorn, "Bayesian tomography for projections with an arbitrary transmission function with an application in electron microscopy," *Journal of Research of the National Institute of Standards and Technology*, vol. 111, pp. 411 –417, Nov. 2006.

[46] W. Van den Broek, S. Van Aert, and D. Van Dyck, "A model based atomic resolution tomographic algorithm," *Ultramicroscopy*, vol. 109, no. 12, pp. 1485 – 1490, 2009.

[47] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans. on Medical Imaging*, vol. 1, pp. 113 –122, Oct. 1982.

[48] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 601 –609, Dec. 1994.

[49] H. Erdogan, G. Gualtieri, and J. A. Fessler, "An ordered subsets algorithm for transmission tomography," *Med. Biol*, vol. 44, pp. 2835–2851, 1998.

[50] E. Mumcuoglu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 687 –701, Dec. 1994.

[51] J. Fessler and S. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. on Image Processing*, vol. 8, pp. 688 –699, May 1999.

[52] A. Tuysuzoglu and W. C. Karl, "A graph-cut method for tomographic reconstruction," in *2012 IEEE Statistical Signal Processing Workshop (SSP) (SSP'12)*, (Ann Arbor, Michigan, USA), Aug. 2012.

[53] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, pp. 480 –492, Mar. 1996.

[54] B. De Man, S. Basu, J.-B. Thibault, J. Hsieh, J. Fessler, C. Bouman, and K. Sauer, "A study of four minimization approaches for iterative reconstruction in X-ray CT," in *2005 IEEE Nuclear Science Symposium Conference Record*, vol. 5, pp. 2708 –2710, Oct. 2005.

[55] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. Bouman, "Bayesian Tomographic Reconstruction for High Angle Annular Dark Field (HAADF) Scanning Transmission Electron Microscopy (STEM)," in *2012 IEEE Statistical Signal Processing Workshop (SSP) (SSP'12)*, (Ann Arbor, Michigan, USA), Aug. 2012.

[56] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[57] S.J.Pennycook, S.D.Berger, and R.J.Culbertson, "Elemental mapping with elastically scatterd electrons," *Journal of Microscopy*, vol. 144, pp. 229–249, 1986.

[58] Z. Saghi, X. Xu, Y. Peng, B. Inkson, and G. Mobus, "Three-dimensional chemical analysis of tungsten probes by energy dispersive X-ray nanotomography," *Applied Physics Letters*, vol. 91, no. 25, p. 251906, 2007.

[59] W. Van den Broek, A. Rosenauer, B. Goris, G. Martinez, S. Bals, S. Van Aert, and D. Van Dyck, "Correction of non-linear thickness effects in HAADF STEM electron tomography," *Ultramicroscopy*, vol. 116, pp. 8–12, 2012.

[60] P. Ercius, M. Weyland, D. A. Muller, and L. M. Gignac, "Three-dimensional imaging of nanovoids in copper interconnects using incoherent bright field tomography," *Applied Physics Letters*, vol. 88, no. 24, p. 243116, 2006.

[61] J. M. LeBeau and S. Stemmer, "Experimental quantification of annular dark-field images in scanning transmission electron microscopy," *Ultramicroscopy*, vol. 108, no. 12, pp. 1653 – 1658, 2008.

[62] A. Rosenauer, K. Gries, K. Müller, A. Pretorius, M. Schowalter, A. Avramescu, K. Engl, and S. Lutgen, "Measurement of specimen thickness and composition in $Al_xGa_{1-x}N/GaN$ using high-angle annular dark field images," *Ultramicroscopy*, vol. 109, no. 9, pp. 1171 – 1182, 2009.

[63] J. W. Cahn and J. E. Hilliard, "Free Energy of a Nonuniform System. I. Interfacial Free Energy," *The Journal of Chemical Physics*, vol. 28, no. 2, pp. 258–267, 1958.

[64] P. Jin, E. Haneda, K. Sauer, and C. A. Bouman, "A Model-Based 3D Multi-slice Helical CT Reconstruction Algorithm for Transportation Security Application," in *Second International Conference on Image Formation in X-Ray Computed Tomography*, (Salt Lake City, Utah, USA), June 2012.

[65] S. Oh, A. Milstein, C. Bouman, and K. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Trans. on Image Processing*, vol. 14, pp. 125 –140, Jan. 2005.

[66] S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb, "Source-detector calibration in three-dimensional Bayesian optical diffusion tomography," *J. Opt. Soc. Am. A*, vol. 19, pp. 1983–1993, Oct. 2002.

[67] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296 –310, July 1993.

[68] K. J. D. Mastronarde and J. McIntosh, "Computer visualization of three-dimensional image data using IMOD," *Journal Of Structural Biology*, vol. 116, pp. 71 –76, Aug. 1996.

[69] H. Demers, N. Poirier-Demers, A. R. Couture, D. Joly, M. Guilmain, N. de Jonge, and D. Drouin, "Three-dimensional electron microscopy simulation with the CASINO Monte Carlo software," *Scanning*, vol. 33, no. 3, pp. 135–146, 2011.

[70] M. N. Tchoul, S. P. Fillery, H. Koerner, L. F. Drummy, F. T. Oyerokun, P. A. Mirau, M. F. Durstock, and R. A. Vaia, "Assemblies of Titanium Dioxide-Polystyrene Hybrid Nanoparticles for Dielectric Applications," *Chemistry of Materials*, vol. 22, no. 5, pp. 1749–1759, 2010.

[71] A. Sousa, M. Hohmann-Marriott, A. Azari, G. Zhang, and R. Leapman, "BF STEM tomography for improved 3D imaging of thick biological sections," *Microscopy and Microanalysis*, vol. 15, pp. 572–573, 7 2009.

[72] R. J. Spontak, M. C. Williams, and D. A. Agard, "Three-dimensional study of cylindrical morphology in a styrene-butadiene-styrene block copolymer," *Polymer*, vol. 29, no. 3, pp. 387 – 395, 1988.

[73] A. Koster, U. Ziese, A. Verkleij, A. Janssen, J. de Graaf, J. Geus, and K. de Jong, "Development and application of 3-dimensional transmission electron microscopy (3D-TEM) for the characterization of metal-zeolite catalyst systems," in *12th International Congress on Catalysis Proceedings of the 12th ICC* (S. M. Avelino Corma, Francisco V. Melo and J. L. G. Fierro, eds.), vol. 130 of *Studies in Surface Science and Catalysis*, pp. 329 – 334, Elsevier, 2000.

[74] M. Bar Sadan, L. Houben, S. G. Wolf, A. Enyashin, G. Seifert, R. Tenne, and K. Urban, "Toward atomic-scale bright-field electron tomography for the study of fullerene-like nanostructures," *Nano Letters*, vol. 8, no. 3, pp. 891–896, 2008.

[75] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. Bouman, "A model based iterative reconstruction algorithm for high angle annular dark field - scanning transmission electron microscope (HAADF-STEM) tomography," *IEEE Trans. on Image Processing*, vol. 22, Nov. 2013.

[76] S. V. Venkatakrishnan, L. F. Drummy, M. De Graef, J. P. Simmons, and C. A. Bouman, "Model based iterative reconstruction for bright field electron tomography," *Proc. SPIE 8657, Computational Imaging XI, 86570A (February 14, 2013)*, pp. 86570A–86570A–12, 2013.

[77] Z. H. Levine, "Theory of bright-field scanning transmission electron microscopy for tomography," *Journal of Applied Physics*, vol. 97, no. 3, p. 033101, 2005.

[78] D. Williams and C. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science. Diffraction. II.* Transmission Electron Microscopy: A Textbook for Materials Science.

[79] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 2–12, Jan. 1989.

[80] F. Meyer, "Topographic distance and watershed lines," *Signal Process.*, vol. 38, pp. 113–125, July 1994.

[81] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 106–128, 2013.

[82] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, pp. 323–343, Apr. 2009.

[83] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 60–65 vol. 2, 2005.

[84] D. B. Husarik, D. Marin, E. Samei, S. Richard, B. Chen, T. A. Jaffe, M. R. Bashir, and R. C. Nelson, "Radiation dose reduction in abdominal computed tomography during the late hepatic arterial phase using a model-based iterative reconstruction algorithm: How low can we go?," *Invest Radiol.*, vol. 47, no. 8, pp. 468–474, 2012.

[85] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *IEEE International Symposium on Biomedical Imaging*, pp. 1375–1378, 2008.

[86] Q. Xu, H. Yu, X. Mou, D. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 9, pp. 1682–1697, 2012.

[87] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I – 549–I – 552, Sept 2007.

[88] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Image deblurring by augmented Langrangian with BM3D frame prior," in *Proc. 2010 Workshop on Information Theoretic Methods in Science and Engineering, WITMSE 2010, Tampere, Finland*, Aug. 2010.

[89] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1715–1728, 2012.

[90] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, July 2011.

[91] R. Glowinski and A. Marrocco, "Sur lapproximation, par elements finis dordre un, et la resolution, par penalisation-dualite, dune classe de problems de Dirichlet non lineares," *Revue Francaise dAutomatique, Informatique, et Recherche Operationelle*, vol. 9, pp. 41–76, 1975.

[92] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17 – 40, 1976.

[93] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[94] M. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681–695, 2011.

[95] M. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2345–2356, 2010.

[96] M. Nilchian and M. Unser, "Differential phase-contrast X-ray computed tomography: From model discretization to image reconstruction," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 90–93, 2012.

[97] M. G. McGaffin, S. Ramani, and J. A. Fessler, "Reduced memory augmented Lagrangian algorithm for 3D iterative X-ray CT image reconstruction," pp. 831327–831327–6, 2012.

[98] T. Frese, C. A. Bouman, and K. Sauer, "Multiscale Bayesian methods for discrete tomography," in *Discrete Tomography: Foundations, Algorithms and Applications* (G. T. Herman and A. Kuba, eds.), pp. 237–261, Birkhaüser Boston, Cambridge, MA, 1999.

[99] R. Rockafellar, "Moreaus proximal mappings and convexity in Hamilton-Jacobi theory," in *Nonsmooth Mechanics and Analysis* (P. Alart, O. Maisonneuve, and R. Rockafellar, eds.), vol. 12 of *Advances in Mechanics and Mathematics*, pp. 3–12, Springer US, 2006.

[100] Y. Huang, M. Ng, and Y. Wen, "A fast total variation minimization method for image restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 2, pp. 774–795, 2008.

[101] S. V. Venkatakrishnan, L. F. Drummy, M. De Graef, J. P. Simmons, and C. A. Bouman, "Model based iterative reconstruction for bright field electron tomography," *Proc. SPIE 8657, Computational Imaging XI, 86570A (February 14, 2013)*, pp. 86570A–86570A–12, 2013.

[102] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, Nov. 1992.

[103] "The dynamics of interfaces during coarsening in solidliquid systems," *Acta Materialia*, vol. 70, no. 0, pp. 66 – 78, 2014.

[104] A. Stevens, H. Yang, L. Carin, I. Arslan, and N. D. Browning, "The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution stem images," *Microscopy*, vol. 63, no. 1, pp. 41–51, 2014.

[105] H. S. Anderson, J. Ilic-Helms, B. Rohrer, J. Wheeler, and K. Larson, "Sparse imaging for fast electron microscopy," *Proc. SPIE 8657, Computational Imaging XI, 86570C (February 14, 2013)*, vol. 8657, pp. 86570C–86570C–12, 2013.

[106] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM National Conference*, ACM '68, pp. 517–524, 1968.

[107] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, 1979.

APPENDICES

A. SINGLE VOXEL UPDATE USING SUBSTITUTE FUNCTION

In order to find a suitable substitute function to the original cost (2.9), each of the potential functions $\rho(u - f_i)$ can be replaced by a function $\rho(u - f_i; f'_j - f_i)$ which satisfy the following properties [8]

$$\rho(u - f_i; f'_j - f_i) \;\geq\; \rho(u - f_i) \;\; \forall u \in \mathbb{R} \tag{A.1}$$

$$\rho'(f'_j - f_i; f'_j - f_i) \;=\; \rho'(f'_j - f_i) \tag{A.2}$$

where f'_j is the point of approximation. Intuitively (A.1) ensures that the substitute function upper bounds the original potential function and (A.2) ensures that the derivatives of the original function and the substitute function are matched at the point of approximation. We use a substitute function of the form

$$\rho(u - f_i; f'_j - f_i) \;=\; \frac{a_{ji}}{2}(u - f_i)^2 + b_{ji} \tag{A.3}$$

because it results in a simple closed form update for a given voxel. Thus we need to find the values of a_{ji} and b_{ji} which satisfies (A.1) and (A.2). Taking the derivative of the substitute function (A.3) and matching it to the derivative of the original potential function we get

$$a_{ji} \;=\; \begin{cases} \frac{\rho'(f'_j - f_i)}{(f'_j - f_i)} & f'_j \neq f_i \\ \rho''(0) & f'_j = f_i \end{cases}$$

To choose b_{ji} we set the value of the original potential function and substitute function to be the same at the point of approximation f'_j. This gives

$$b_{ji} = \rho(f'_j - f_i) - \frac{a_{ji}}{2}(f'_j - f_i)^2$$

B. GAIN AND OFFSET UPDATE WITH UNKNOWN TRANSMITTED ATTENUATION

Rewriting the first summation of the cost function in (2.8) as a sum of quadratics in $[I_k, d_k]$ and dropping terms which do not involve I_k, d_k gives us a new cost function corresponding to the unconstrained optimization problem,

$$\tilde{c}(I, d, \lambda) = \frac{1}{2} \sum_{k=1}^{K} \left([I_k \, d_k] \, Q_k \begin{bmatrix} I_k \\ d_k \end{bmatrix} - 2 \, [I_k \, d_k] \, b_k \right) + \lambda \left(\frac{1}{K} \sum_{k=1}^{K} I_k - \bar{I} \right)$$

where Q_k and b_k are defined as in chapter 2.4.2. To find the minimum of the cost function, we take gradient with respect to each $[I_k \, d_k]$ and set it to zero. This gives us

$$Q_k \begin{bmatrix} I_k \\ d_k \end{bmatrix} - b_k + \begin{bmatrix} \frac{\lambda}{K} \\ 0 \end{bmatrix} = 0 \tag{B.1}$$

and

$$\begin{bmatrix} I_k \\ d_k \end{bmatrix} = Q_k^{-1} \left(b_k - \begin{bmatrix} \frac{\lambda}{K} \\ 0 \end{bmatrix} \right). \tag{B.2}$$

If $\tilde{Q}_k = Q_k^{-1}$ then $I_k = \tilde{q}_{k,11}(b_{k,1} - \frac{\lambda}{K}) + \tilde{q}_{k,12} b_{k,2}$. Using this in the constraint $\left(\frac{1}{K} \sum_{k=1}^{K} I_k - \bar{I} \right)$ and solving for λ we get

$$\hat{\lambda} \;=\; \frac{\displaystyle\sum_{k=1}^{K}(\tilde{q}_{k,11}b_{k,1} + \tilde{q}_{k,12}b_{k,2}) - K\bar{I}}{\dfrac{1}{K}\displaystyle\sum_{k=1}^{K}\tilde{q}_{k,11}}. \tag{B.3}$$

Finally we can use this value of $\hat{\lambda}$ in (B.2) to solve for the optimal values of $[I_k \; d_k]$.

C. COMPOSITION PROPERTY OF SURROGATE FUNCTIONS

Theorem C.0.1 *Composition property - Let $q(z; z')$ be a surrogate function for the minimization of $t(z)$ on $\mathcal{A} \subset \mathbb{R}^N$; and let $h : \mathcal{A} \to \mathcal{A}$. Then define*

$$
\begin{aligned}
\tilde{t}(z) &\triangleq t(h(z)) \\
\tilde{q}(z; z') &\triangleq q(h(z); h(z'))
\end{aligned}
$$

Then $\tilde{q}(z; z')$ is a surrogate function for $\tilde{t}(z)$.

Proof   The theorem can be proved by verifying the sufficiency coditions for surrogate functions in (3.8). Notice that at $z = z'$, $\tilde{q}(z'; z') = \tilde{t}(z')$ because q is a surrogate function to t. Furthermore, for any $z \subset \mathcal{A}$, $\tilde{q}(z; z') \geq \tilde{t}(z)$ by the construction of q i.e., $q(z; z') \geq t(z)$ for $z \subset \mathcal{A}$. Therefore the compostion is still a surrogate function because it satisfies the sufficiency conditions in (3.8). ∎

D. PARAMETERS FOR DENOISING ALGORITHMS

The parameters of the algorithms are adjusted for the minimum RMSE reconstruction in the case of the simulated data sets and the best visual quality for the real data sets. In this appendix we specify the parameters used for the different denoising routines. We refer to the simulated tomography data set as - Shepp-Logan Tomo, the real synchrotron tomography data set as - SX Tomo and the simulated sparse image reconstruction data set as - Cameraman Interp.

- **K-SVD** : We use the K-SVD denoising code from `http://www.cs.technion.ac.il/~ronrubin/software.html`. The parameters used for the results presented in Chapter 4 are given in D.1.

Table D.1: K-SVD parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo	Cameraman Interp
Patch size	4×4	4×4	8×8
Dictionary size	3600	40000	40000
Number of iteration	10	1	10
σ_n	5.47	3.46	5.47

The other parameters are set to the default values in the software. The K-SVD dictionary is initialized with the default settings in the software. For each subsequent outer iteration of ADMM, the K-SVD dictionary is initialized with the final dictionary from the previous iteration.

- **BM3D** : We use the BM3D code from `http://www.cs.tut.fi/~foi/GCF-BM3D/index.html\#ref\_software`. The parameters used for the different applications are shown in Table D.2

The other parameters are set to the default values in the software.

Table D.2: BM3D parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo	Cameraman Interp
Patch size	4×4	4×4	4×4
σ_n	10.19	3.46	5.48

- **PLOW** : The code for PLOW was downloaded from `http://users.soe.ucsc.edu/~priyam/PLOW/`.

  The parameters used for the different applications are shown in Table D.3

Table D.3: PLOW parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo	Cameraman Interp
Patch size	5×5	5×5	5×5
σ_n	5.47	2	4.47

The other parameters are set to the default values in the software.

- **q-GGMRF** : The q-GGMRF denoising operator is given by

$$\hat{f} \leftarrow \underset{f}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \|g - f\|_2^2 + \frac{1}{p\sigma_f^p} \sum_{\{i,j\} \in \chi} w_{ij} \frac{|f_i - f_j|^p}{1 + |\frac{f_i - f_j}{c}|^{p-q}} \right\} \qquad (D.1)$$

where $p, q, c, \sigma_f, w_{ij}$ are the q-GGMRF parameters, σ_n^2 is the variance of the white noise in the data, and χ is the set of all neighboring pixels (8 point neighborhood). The weights are set to $\frac{1}{12}$ for diagonal neighbors and to $\frac{1}{6}$ for horizontal and vertical neighbors. The parameters of the denoising are set to $p = 2, q = 1.2, c = 1/100$ which was found to produce a reasonable result for images with white noise of variance σ_n^2. The random order ICD with surrogate functions [8] is used to implement the cost optimization. The parameters used for the different applications are shown in Table D.4.

Table D.4: qGGMRF parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo	Cameraman Interp
σ_f	0.29	3×10^{-4}	0.29
σ_n	4.58	2×10^{-4}	4.47

- **Total Variation** : The code for TV denoising was downloaded from `http://www.ceremade.dauphine.fr/~peyre/matlab/image/content.html`. It minimizes the cost function :

$$\hat{f} \quad \leftarrow \quad \underset{f}{\operatorname{argmin}} \left\{ \|g - f\|_2^2 + \frac{\lambda_{TV}}{2} \mathrm{TV}(f) \right\} \qquad (D.2)$$

$$\hat{f} \quad \leftarrow \quad \underset{f}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \|g - f\|_2^2 + \frac{c_{TV}}{2} \mathrm{TV}(f) \right\} \qquad (D.3)$$

where $\lambda_{TV} = 2c_{TV}\sigma_n^2$ and $\mathrm{TV}(f)$ represents the total variation operator. The number of iterations was set to 100. This value was found to produce a reasonable denoising result for additive white Gaussian noise of variance σ_n^2. The parameters used for the different applications are shown in Table D.5.

Table D.5: qGGMRF parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo	Cameraman Interp
c_{TV}	0.099	1	0.099
σ_n	3.07	3.46	4.47

- **Discrete Reconstruction (DR) prior** : A denoising operation corresponding to the discrete reconstruction prior (one that restricts the number of classes/values taken by the pixels in the reconstruction to K) is given by

$$(\hat{\mu}, \hat{b}) \leftarrow \underset{\mu, b \in \{1, \cdots, K\}^M}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} \sum_{i=1}^M (g_i - \mu(b_i))^2 + c_{\text{Discrete}} \sum_{\{i,j\} \in \chi} w_{ij} \delta(b_i \neq b_j) \right\} \tag{D.4}$$

where g is a $M \times 1$ vector containing the noisy image, σ_n^2 is the noise variance, b is a $M \times 1$ vector of labels corresponding to each pixel, $\mu : \{1, \cdots, K\} \to \mathbb{R}$ is a function that maps each label to a discrete output level (class mean), χ consists of all pairs of neighboring pixels, w_{ij} weights the interaction between neighboring pixels , δ is an indicator function, and c_{Discrete} is a constant.

The denoising operation above is a non-convex optimization problem and we will describe an algorithm to find a local minimum. To minimize the cost function in (D.4) we use an alternating minimization strategy. The algorithm consists of repeatedly performing the following steps for each pixel i.

  – **Class label update**

$$\hat{b}_i \leftarrow \underset{k \in \{1, \cdots, K\}}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_n^2} (g_i - \hat{\mu}(b_i))^2 + c_{\text{Discrete}} \sum_{j \in \chi_i} w_{ij} \delta(b_j \neq k) \right\} \tag{D.5}$$

  where χ_i is the set of neighbors of pixel i.

  – **Mapping function update**

  Taking derivative of the cost function in (D.4) with respect to each $\mu(k)$ $\forall k \in 1, \cdots, K$ and setting it to zero gives

$$\hat{\mu}(k) \leftarrow \frac{1}{N_k} \sum_{i=1}^M g_i \delta(\hat{b}_i = k) \tag{D.6}$$

where $N_k = \sum\limits_{i=1}^{M} \delta(\hat{b}_i = k)$. Notice that this step simply sets the mapping for a given class to the mean value of the pixels assigned to that class.

Thus the denoising operator corresponding to the discrete reconstruction prior is given by

$$\mathbb{H}(g; \sigma_n^2) = \mu_{\hat{b}}.$$

where $\mu_{\hat{b}} = [\hat{\mu}(\hat{b}_1), \cdots, \hat{\mu}(\hat{b}_M)]^t$.

In practice we set the number of iterations to 10 as we found this to produce good results for the denoising problem. The weights are set to $\frac{1}{12}$ for diagonal neighbors and to $\frac{1}{6}$ for horizontal and vertical neighbors. In order to initialize the class labels b we use Otsu's method [107] on a low pass filtered version of the noisy input g. The parameters used for the different applications are shown in Table D.6.

Table D.6: DR parameters used for different applications

Attribute	Shepp-Logan Tomo	SX Tomo
c_{Discrete}	4	3000
σ_n	20	6×10^{-8}
K	6	5

VITA

VITA

S. V. Venkatakrishnan received his B.Tech. degree in electronics and communication engineering from National Institute of Technology Tiruchirappalli, India, in 2007, M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2009, and is currently pursuing the Ph.D. degree in the School of Electrical and Computer Engineering, Purdue University.

From 2009 to 2010, he worked as a research and development engineer at Baker Hughes Inc., Houston, TX on logging while drilling - imaging magnetometers. His current research interests are in statistical signal processing, inverse problems, machine learning, and computational imaging with applications in electron microscopy and X-ray imaging.