Reconstruction, Classification, and Segmentation for Computational Microscopy

by

Se Un Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2013

Doctoral Committee:

Professor Alfred O. Hero, Chair
Professor Jeffrey Fessler
Assistant Professor Raj Rao Nadakuditi
Assistant Professor Samantha Daly

Dedicated to my loving family

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Alfred Hero, for his guidance throughout the years of my PhD study. With his help and encouragement, I could walk through to the end of the road in this demanding PhD program. Especially, our interactions have broadened my intellectual spectra and inspired me to think in more mathematical and statistical ways. I intellectually revere him for his brilliance and critical thinking, personally admire him for his endless passion in research wherever he is, and respect his professional perfection in his work, leadership, and flawless presentation. Thanks to him, I had great experiences in conducting research and could appreciate how rewarding it is to devise never-invented solutions to challenging problems, even with multiple perspectives. I wish and believe that in future, like him, I will lead a group, conduct research, and teach.

I would also like to show my great appreciation for my committee members, Professors Fessler, Nadakuditi, and Daly, for their valuable comments and feedback on my work, especially their careful reading of the dissertation. I am sincerely grateful to all of my collaborators and the members of the Hero group, particularly Nicolas Dobigeon, whose invaluable feedback to my work shaped me into a better researcher in my first few years of the PhD. I also thank Dr. Dan Rugar for his insightful comments on the MRFM project; Dr. Marc DeGraef for his providing the EBSD dictionary to us and comments on the EBSD work; Professor Chris Miller for collaborating the SDSS project; and Dennis Wei, Ami Wiesel, and Greg Newstadt for having fruitful

discussions with me.

My heartfelt appreciation goes to the Korea Foundation for Advanced Studies, for granting me support to initiate my study in financially difficult years. I was really fortunate to be selected as a scholar in this prestigious foundation.

Special thanks go to all of my friends with whom I worked, had fun, did sports, and traveled all around. I could recharge myself and manage frustrating moments thanks to them.

Finally, I can never express enough my appreciation to my parents, Mr. Jae Young Park and Mrs. Hyunja Ha. Their encouragement, visits, and even casual hellos, specifically during my difficult times, have made me more strong and independent. My brother, Sejin Park, also helped me strengthen ties with family and relax within family. I owe much to my family for always believing in me and encouraging me to achieve my goals.

All these people have made my years in Ann Abor one of the best times of my life. I cherish all the memories.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

ABSTRACT

Reconstruction, Classification, and Segmentation for Computational Microscopy

by

Se Un Park

Chair: Professor Alfred O. Hero

This thesis treats two fundamental problems in computational microscopy: image reconstruction for magnetic resonance force microscopy (MRFM) and image classification for electron backscatter diffraction (EBSD). In MRFM, as in many inverse problems, the true point spread function (PSF) that blurs the image may be only partially known. The image quality may suffer from this possible mismatch when standard image reconstruction techniques are applied. To deal with the mismatch, we develop novel Bayesian sparse reconstruction methods that account for possible errors in the PSF of the microscope and for the inherent sparsity of MRFM images. Two methods are proposed: a stochastic method and a variational method. They both jointly estimate the unknown PSF and unknown image. Our proposed framework for reconstruction has the flexibility to incorporate sparsity inducing priors, thus addressing ill-posedness of this non-convex problem, Markov-Random field priors, and can be extended to other image models. To obtain scalable and tractable solutions, a dimensionality reduction technique is applied to the highly nonlinear PSF space. The experiments clearly demonstrate that the proposed methods have superior

performance compared to previous methods.

In EBSD we develop novel and robust dictionary-based methods for segmentation and classification of grain and sub-grain structures in polycrystalline materials. Our work is the first in EBSD analysis to use a physics-based forward model, called the dictionary, to use full diffraction patterns, and that efficiently classifies patterns into grains, boundaries, and anomalies. In particular, unlike previous methods, our method incorporates anomaly detection directly into the segmentation process. The proposed approach also permits super-resolution of grain mantle and grain boundary locations. Finally, the proposed dictionary-based segmentation method performs uncertainty quantification, i.e. p-values, for the classified grain interiors and grain boundaries. We demonstrate that the dictionary-based approach is robust to instrument drift and material differences that produce small amounts of dictionary mismatch.

CHAPTER I

Introduction

This thesis treats two fundamental problems in computational microscopy: image reconstruction and image segmentation. These problems are treated in the context of the 3D reconstruction of magnetic resonance force microscopy (MRFM) images and the segmentation of electron backscatter diffraction (EBSD) microscopy images, respectively. MRFM is an emerging technology for volumetric imaging of atomic spin systems, while EBSD is a mature technology for imaging polycrystalline structures in materials science. In MRFM we develop new Bayesian sparse reconstruction methods that account for errors in the point spread function (PSF) of the microscope. In EBSD we develop novel and robust dictionary-based methods for segmentation and classification of grain and sub-grain structures in polycrystalline materials.

Microscopic imaging technology has been one of the core driving forces in the physical and engineering sciences. Most modern microscope technology uses computational imaging techniques to reconstruct images of the sample from blurred and noisy physical measurements from the microscope. Computational microscopy consists of using numerical algorithms and mathematical models to perform the reconstructions. Among the many operations used in computational microscopy are: image reconstruction; image segmentation; image classification; object recognition; and quantitative labeling. This thesis concentrates on the first three operations.

Many algorithms for image reconstruction can be viewed mathematically as deconvolution of the measured raw image, which is inherently degraded by a partially unknown point spread function (PSF). The quality degradation of images caused by the PSF is usually modeled as a linear convolution of an image with a blur function or blur kernel. The observed image is usually corrupted by additive noise as well. Image deconvolution must recover the unknown image from the PSF-degraded and noise contaminated raw images.

The objective of image segmentation is to partition the image into contiguous regions defined by a network of region boundaries. The objective of image classification is to categorize all pixels into one of a number of pre-determined labels. Image segmentation is a special case of image classification with two classes: a pixel is at a boundary (class 1) or is not at a boundary (class 2). There are two categories of classification methods: supervised and unsupervised methods. Supervised classification involves training data where the true labels are provided with a set of training data. The classifier is specified using the training data to best match the training labels to the pixels under some criterion, e.g. 0-1 loss or hinge loss. Once trained, the classifier is applied to classify test data with unknown labels. Examples of supervised classifiers include the maximum likelihood classifier, where labels are assigned to the class of highest likelihood. Another type of supervised classifier is the nearest-neighbor/minimum-distance classifier that finds the closest neighbor in the label set for a pixel in the image. Unsupervised classification, unlike supervised classification, does not require ground truth data for labeling; thus the classification is data-driven. An example of unsupervised classification is clustering. In this thesis we mainly consider unsupervised classification and segmentation problems. We next describe the main problems that we consider in image reconstruction and in image segmentation.

1.1 Semi-blind Image Reconstruction

Many deconvolution techniques require prior knowledge of the device response, i.e., the PSF. However, in many practical situations, the true PSF is either unknown or, at best, only partially known. For example, in an optical system uncertainties in the PSF can be due to light diffraction, apparatus/lense aberration, defocusing, or image motion [6, 7]. Such imperfections are common in general imaging systems including MRFM, where there can exist additional errors in the PSF model under the magnetic resonance condition [8]. In such circumstances, the PSF required in the reconstruction process is mismatched with the true PSF. The quality of standard image reconstruction techniques may suffer from this model mismatch. To deal with this mismatch, blind deconvolution methods have been proposed to estimate the unknown image and the PSF jointly. A blind deconvolution algorithm estimates the image and PSF without exact prior knowledge of the PSF. Without further constraints blind deconvolution is intrinsically ill-posed because the solution pairs for the image and PSF are not unique. Thus, blind deconvolution algorithms typically require additional constraints on the image or the PSF, e.g., the image has unit norm or the PSF is smooth. When prior knowledge of the PSF is available, these methods are referred to as semi-blind deconvolution [9, 10] or myopic deconvolution [11, 12, 13]. The ill-posedness of the semi-blind deconvolution problem depends on the extent to which the prior knowledge of PSF is informed.

The primary objective of Chapters II-IV is to address semi-blind deconvolution of MRFM images and explore ways to exploit prior knowledge of image or PSF models, at the same time proposing computationally tractable algorithms. Our approach is based on formulating the semi-blind deconvolution task as an image estimation problem in a Bayesian setting. Bayesian estimation offers a flexible framework to solve complex model-based problems. Prior information can be easily included within the model, leading to an implicit regularization of the ill-posed problem. In addition,

the Bayes framework produces posterior estimates of uncertainty, via posterior variance and posterior confidence intervals. Bayesian inference can often involve difficult analytical and numerical hurdles, such as, non-convex optimization, Monte Carlo simulation, and numerical integration. One widely advocated Bayesian strategy uses approximations to the minimum mean square error (MMSE) or maximum a posteriori (MAP) estimators based on sampling from the posterior distribution. Generation of these samples can often be accomplished using Markov chain Monte Carlo methods (MCMC) [14]. Another approach bypasses Monte Carlo by replacing the posterior by a simpler surrogate function which can be maximized numerically but that yields a close approximation to the MAP estimator. Variational Bayes (VB) approximations fall into this class of approximations. Chapter III takes the MCMC approach while Chapter IV takes the VB approach to the MRFM blind deconvolution problem.

The MRFM semi-blind image restoration problem that we will address in this work was previously studied within a hierarchical Bayesian framework [15] with partially known blur functions in many applications of imaging [16, 17]. Several authors [16, 17] model the deviation of the PSF as uncorrelated zero mean Gaussian noise. The authors of [18] considered an extension of this model to a non-sparse, simultaneous autoregression (SAR) prior model for both the image and point spread function. Papers on single motion deblurring in photography [19, 20] use heavier tailed distributions for the image gradient and an exponential form for the PSF. The algorithm in [19] separately identifies the PSF using a multi-scale approach to perform conventional image restoration. The authors of [20] proposed an image prior to reduce ringing artifacts from blind deconvolution of photo images.

This thesis considers an alternative image and PSF model, which is appropriate for our data but significantly different from those developed for natural images, astronomy, and photography. Our suggested model imposes sparsity on the image and an empirical Bayes prior on the PSF. The contribution of this work is to develop

and apply novel Bayesian approaches to a joint estimation of PSFs and images. We initially assume the desired image is sparse, corresponding to the natural sparsity of the molecular image of interest. We use the image prior as a weighted sum of a part inducing sparsity and a continuous distribution, a positive truncated Laplacian and atom at zero prior[1] [21]. Similar priors have been applied to estimating mixtures of densities [22, 23, 24] and sparse, nonnegative hyperspectral unmixing [25]. In the MCMC approach, described in Chapter III, a Monte Carlo simulation is used to approximate the full posterior distribution, from which MAP estimates of the MRFM image and the PSF can be easily computed. In our variational approach, described in Chapter IV, we additionally introduce a hidden label variable for the contribution of the discrete mass (empty pixel) along with a continuous density function (non-empty pixel). Similar to our 'hybrid' mixture model, inhomogeneous gamma-Gaussian mixture models have been proposed in [26].

In addition to proposing a new hierarchical MRFM image model, in Chapters III and IV we propose solutions that exploit prior information on the MRFM PSF. We represent the PSF on a truncated orthogonal basis, where the basis elements are the singular vectors in the singular value decomposition of the family of perturbed nominal PSFs. A Gaussian prior model specifies a log quadratic Bayes prior on deviations from the nominal PSF. Also, a vague PSF prior of the improper uniform distribution is used. Our approach is related to the recent papers of Tzikas *et al.* [27] and Orieux *et al.* [28]. In [27] a pixel-wise, space-invariant Gaussian kernel basis is assumed with a gradient based image prior. Orieux *et al.* introduced a Metropolis-within-Gibbs algorithm to estimate the parameters that tune the device response. Their strategy [28] focuses on reconstruction with smoothness constraints and requires recomputation of the entire PSF at each step of the algorithm. This approach is computationally expensive, especially for complex PSF models such as that in the

[1] A Laplace distribution as a prior distribution acts as a sparse regularization using ℓ_1 norm. This can be seen by taking negative logarithm on the distribution.

physical instrument for our MRFM application. We propose an alternative strategy that consists of estimating the deviation from a given nominal PSF, and this has a computational advantage over other previous methods, independent of complexity of the PSF generation model. More precisely, the nominal point response of the device is assumed to be known and the true PSF is modeled as a small perturbation around the nominal response. Since we only need to estimate linear perturbations around the nominal PSF relative to a low dimensional precomputed and truncated basis set, this leads to reduction in computational complexity and an improvement in convergence as compared to previous approaches [27] and [28].

In Chapter III the MCMC approach is developed for semi-blind deconvolution of MRFM images. This optimization uses samples generated by a Markov Chain Monte Carlo algorithm to estimate the unknown image, PSF, and parameters. This MCMC approach has been successfully adopted in numerous imaging problems such as image segmentation, denoising, and deblurring [29, 14]. With this framework we approximate the full posterior distribution of the PSF and the image. However, these sampling methods have one disadvantage: convergence is slow and the number of samples that are needed in estimation can be large.

In Chapter IV the variational approximation to the posterior distribution is developed to address the drawbacks of Monte Carlo integration. This approach has been widely applied with success to many different engineering problems [30, 31, 32, 33]. Within the semi-blind imaging literature, for instance, the VB framework has been implemented with a Gaussian prior [34] or a Student's-t prior [27] for the PSF modeling.

This VB method approximates distributions directly by attempting to minimize a distributional distance between the true posterior and a simple approximation, or surrogate, for the true posterior. Specifically, the posterior is estimated by minimizing the Kullback-Leibler (KL) distance between the model and the empirical distribution.

This approximation has been extensively exploited to conduct inference in graphical models [35]. If properly designed, this approach can produce an analytical posterior distribution from which Bayesian estimators can be efficiently computed. Compared to MCMC, VB methods are deterministic and have lower computational complexity, since they avoid Monte Carlo simulations. Additionally, VB approaches do not require as much storage space as does MCMC. However, unlike MCMC approaches, VB approaches do not yield an estimate of the full posterior.

Variational Bayes approaches have other limitations. First, convergence to the true parameters is not guaranteed, even though the VB parameter estimates will be asymptotically normal and equal to the maximum likelihood estimator under suitable conditions [36]. Second, global optimization might be infeasible due to non-convexity in the parameter space. Third, the models that fit the method are limited. In other words, variational Bayes approximations can be easily implemented for only a limited number of statistical models. For example, this method is difficult to apply when latent variables have distributions that do not belong to the exponential family or have mixture distributions (e.g. a discrete distribution [37]), as they do in our proposed sparse image model. For mixture distributions, variational estimators in Gaussian mixtures and in exponential family converge locally to maximum likelihood estimator [38, 39]. The theoretical convergence properties for sparse mixture models, such as our proposed model, are as yet unknown. This has not hindered the application of VB to problems in our sparse image mixture model. Another possible intrinsic limit of the variational Bayes approach, particularly in (semi)-blind deconvolution, is that the posterior covariance structure cannot be effectively estimated nor recovered, unless the true joint distributions have independent individual distributions. This is primarily because VB algorithms are based on minimizing the KL-divergence between the true distribution and the VB approximating distribution, typically a separable factorization approximation with respect to the individual parameters.

Another contribution of this work is the development and implementation of a VB algorithm for mixture distributions in a hierarchical Bayesian model. Our variational Bayes algorithm iterates on a hidden variable domain associated with the mixture coefficients. This results in an algorithm that is faster and more scalable for equivalent image reconstruction qualities than the proposed MCMC approach described in Chapter III.

We experimentally evaluate our MCMC and VB semi-blind algorithms on both simulated and experimental MRFM data obtained from our collaborators in MRFM at IBM. The proposed algorithms are quantitatively compared to state-of-the-art algorithms including the total variation (TV) prior for the PSF [40] and natural sharp edge priors for images with PSF regularization [41]. We also compare our algorithm to basis kernels [27], the mixture model algorithm of Fergus *et al.* [19], and the related method of Shan *et al.* [20] under a motion blur model. Both of our suggested algorithms perform better than these algorithms for estimation of sparse images and smooth PSF.

We illustrate the proposed methods on real data from magnetic resonance force microscopy experiments. The principles of this new 3D imaging technology were first introduced by Sidles [42, 43, 44], who described its potential for achieving 3D atomic scale resolution. In 1992 and 1996, Rugar *et al.* [45, 46] reported experiments that demonstrated the feasibility of MRFM and produced the first MRFM images. More recently, MRFM volumetric spatial resolutions of less than $10nm$ have been demonstrated for imaging a biological sample [4]. The signal provided by MRFM is a so-called force map that is the 3D convolution of the atomic spin distribution and the point spread function (PSF) [47]. This formulation casts the estimation of the spin density from the force map as an inverse problem. Several approaches have been proposed to solve this inverse problem, i.e., to reconstruct the unknown image from the measured force map. Basic algorithms rely on Wiener filters [48, 49, 46] whereas

others are based on iterative least-squares reconstruction approaches [47, 4, 50]. More recently, this problem has been addressed within the Bayesian estimation framework [21, 2]. The drawback of these approaches is that they require prior knowledge of the PSF. However, in many practical situations of MRFM imaging, the exact PSF, i.e., the response of the MRFM tip, is only partially known. Therefore, this MRFM image reconstruction problem can be cast as a semi-blind deconvolution problem with a high-resolution, sparse image model. In such circumstances, the PSF used in the reconstruction algorithm is mismatched to the true PSF of the microscope and the quality of standard MRFM image reconstruction will suffer if this mismatch is not taken into account. To mitigate the effects of the PSF mismatch on MRFM sparse image reconstruction, a non-Bayesian alternating minimization (AM) algorithm [3], which showed robust performance though not scalable, was proposed by Herrity *et al.* In this thesis, we apply our methods to real MRFM tobacco virus data and demonstrate that our proposed semi-blind reconstruction methods better account for this partial knowledge.

Lastly, as a blueprint for future work, we provide extensions to the image and PSF model. At the end of Chapter III we extend our model to incorporate more general image properties, particularly smoothness[2]. Specifically, we present the more generalized image model using Markov random fields (MRFs). MRF type priors have been widely used to exploit local information of inhomogeneous image regions, such as feature, color, or texture, according to the specific neighborhood system for adaptive image segmentation, denoising, interpolation, and deconvolution. The technique using MRF priors promises a generalization of image modeling due to their capacity to incorporate local correlations and desired constraints. For example, an MRF image model using several information features can be used in segmentation

[2]Image smoothness is often assumed as an important feature in natural images, alongside the flatness of the intensity level within boundaries of objects, piece-wise connectivity, and clear edge structures. However, because our applications do not show these properties of boundary or flatness, we do not focus on these.

[51, 52] and as a smoothness prior [53, 54]. Bayesian image reconstruction with MRF priors using an expectation-maximization (EM) algorithm was previously investigated in [55], and its variation using mean field theory was proposed in [56]. A widely-used technique in MRF modeling involves computing images from tomography data [57, 58] with several types of priors using a Bayesian approach. This random field model also fits well into our Bayesian approach with the increased complexity of the solution due to the correlation of sites. Evidently, unlike the spatially homogeneous model in Chapter III, the suggested MRF model does not assume independence of each site. As an extension to the PSF model, at the end of Chapter IV, we provide several approaches to address the space variance of PSFs. This extension is motivated by astronomical data where PSFs should be assumed spatially inhomogeneous but can be approximately homogeneous locally. The PSFs can be found from stellar images and then used for PSF estimation of other sites, usually for galaxies. Another assumption for the data in computational astronomy is the sparsity nature of the image. Thus, our Bayesian formulation casts this problem into again semi-blind sparse image reconstruction.

1.2 Physics-based Dictionary-pursuit for Electron Backscatter Diffraction

In Chapter V we turn to segmentation and classification of grain structures in materials science EBSD microscopy. Electron backscatter diffraction (EBSD) is an efficient and convenient tool, compared to other microscopes, to analyze the crystal orientation, texture of materials, and the spatial locations of grains and grain boundaries. EBSD is particularly helpful in analyzing microstructures of crystalline materials. These crystalline materials have spatial symmetry or lattice patterns with their forming elements. These orderly placed elements characterize backscattered

diffraction patterns. However, there is irregularity in the structure of crystalline materials, which affects performance of materials. Such an analysis of the structures, therefore, merits attention. Specifically, the identification of anomalous parts such as pores, fracture, and additives, is critical in the use of the material.

Unlike the case for MRFM treated in Chapters II-IV, there is at this date no accurate and tractable linear forward model for EBSD that can be used to mathematically relate the crystal orientations to the measurements. However, an accurate Monte Carlo simulation of the forward model has recently been developed by De-Graef's group [5, 59]. Using this simulation approach, a dictionary of diffraction patterns can be generated. The dictionary consists of a set of diffraction patterns and is densely indexed by crystal orientations over a dense but finite grid of Euler angles. This dictionary accounts for electron transport properties of the sample, sample-detector geometry, and noise. With a computational forward model [5, 59] that pre-computes simulated diffraction patterns, we can estimate the neighborhood dictionary elements that are closest to a measured diffraction pattern at a location in the sample. This neighborhood is the feature used to segment and classify the pixels on the sample according to crystal orientations and possibly anomalous structures that are not in the dictionary. This anomaly detection is a novel feature of the proposed method that cannot be performed in the conventional EBSD analysis, thanks to the use of the dictionary. The dictionary design is matched to specific operating parameters of the microscope, e.g., acceleration voltage and beam angle, and the sample, e.g., crystal type (nickel, gold, diamond, etc). Thus there could exist a mismatch between the dictionary and the observation when acquired under different operating conditions. In this thesis we show preliminary results indicating that the dictionary approach to segmentation is robust to mismatch.

In the dictionary approach, a point in the sample producing a diffraction pattern is represented in two ways. One representation is the distribution of correlations

between the observed pattern and the synthetic patterns of dictionary elements. The other is a small number of indices corresponding to large correlations. From these statistics, we identify all grains, grain boundaries, and anomalies.

We establish the robustness of the dictionary approach by applying the dictionary, which was designed for a nickel-base alloy IN100, to a different alloy LSHR[3] acquired using a different microscope. We propose a method of uncertainty quantification to define confidence intervals for the discovered grain/boundary structures. This confidence interval is specifically adapted to the dictionary approach that we adopt in Chapter V. The dictionary approach leads naturally to an algorithm for generating super-resolution maps of grain structures and their boundaries from the low resolution grain maps of the conventional EBSD analysis.

1.3 Outline of Thesis

The rest of this thesis is organized as follows. Chapter II provides the background details for MRFM and the (semi-)blind deconvolution problem. Chapter III formulates and solves the imaging deconvolution problem in a hierarchical Bayesian framework using stochastic samplers. Chapter IV covers the variational methodology and our proposed solutions to the MRFM semi-blind deconvolution problem. Chapter V provides solutions to the anomaly detection and classification problems in EBSD imaging. Chapter VI discusses our findings and concludes.

1.4 Publications

The following publications were produced based on the work presented in this dissertation.

[3]Low Solvus High Refractory: 'low solvus' property indicates that it contributes to material resistance to crack quenching. 'high refractory' means that a material with this property has high tensile strength and creep resistance. This is a nickel base superalloy and is used in the disks of gas turbine engines.

Journal articles

[1] Se Un Park, Nicolas Dobigeon, and Alfred O. Hero, "Variational semi-blind sparse deconvolution with orthogonal kernel bases and its application to MRFM," *Signal Processing,* vol. 94, no. 0, pp. 386 - 400, 2014.

[2] Se Un Park, Nicolas Dobigeon, and Alfred O. Hero, "Semi-blind sparse image reconstruction with application to MRFM," *IEEE Trans. Image Processing,* vol. 21, pp. 3838 -3849, sept. 2012.

Articles in conference proceedings

[1] Se Un Park, Dennis Wei, Marc DeGraef, Megna Shah, Jeff Simmons, and Alfred O. Hero, "EBSD image segmentation using a physics-based forward model," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Sept. 2013, in press.

[2] Se Un Park, Dennis Wei, Marc DeGraef, Megna Shah, Jeff Simmons, and Alfred O. Hero, "EBSD image segmentation using a physics-based pattern dictionary," in *Microscopy and Microanalysis 2013*, Aug. 2013, in press.

[3] Se Un Park, Nicolas Dobigeon, and Alfred O. Hero, "Variational semi-blind sparse image reconstruction with application to MRFM," in *Proc. Computational Imaging Conference in SPIE Symposium on Electronic Imaging Science and Technology,* SPIE, Jan 2012.

[4] Se Un Park, Nicolas Dobigeon, and Alfred O. Hero, "Myopic sparse image reconstruction with application to MRFM," in *Proc. Computational Imaging Conference in SPIE Symposium on Electronic Imaging Science and Technology,* SPIE, Jan. 2011.

CHAPTER II

Semi-blind Deconvolution for Magnetic Resonance Force Microscopy

In this chapter, we present a short background discussion on magnetic resonance force microscopy (MRFM) as well as the need for sparse image reconstruction and blind estimation.

2.1 A Brief Introduction to MRFM

Magnetic resonance imaging (MRI) is a powerful tool for imaging the density of nuclear spins in materials, both living and inert. However, since conventional nuclear magnetic resonance detectors are only capable of detecting and localizing large spin ensembles, the spatial resolution and sensitivity of MRI has limited its use in atomic scale microscopy. Indeed, while atomic force microscopy (AFM) is capable of spatial resolution on the order of 10's of Angstroms or less, the spatial resolution of MRI is only on the order of 10's of micrometers [60]. Thus conventional MRI cannot be used to image at atomic scales. On the other hand, while MRI can be used to image sub-surface spin density, AFM is limited to surface or near surface imaging.

Recently, AFM and MRI have been combined into a single modality, called magnetic resonance force microscopy (MRFM), to combine the high subsurface sensitivity

of MRI with the high spatial resolution of AFM. MRFM has demonstrated sensivity and resolution down to the single-spin level [42, 43]. MRFM was a major breakthrough which enables nondestructive sub-nanometer-scale 3D imaging. MRFM is capable of imaging small spin ensembles below the surface, to a depth of 100 nm, with atomic resolution [1]. This capability could potentially be applied to imaging the 3D chemical composition of nanostructures and biological structures [4]. The applications may include deciphering the structure and interactions of proteins and directly imaging macromolecules that cannot be easily analyzed by using other modalities, such as X-ray and conventional MRI technologies. MRFM could complement the established surface-scanning techniques including scanning electron micrograph (SEM). In this thesis we develop robust statistical models and image reconstruction algorithms for MRFM.

The principle of MRFM is based on the mechanical measurement of attonewton-scale magnetic forces between a magnetic tip and spins in a target sample. The detection of this extremely small force was recently realized with a ultrasensitive cantilever, by using spin relaxation processes and the detection of statistical polarization in small spins [13] [61].

In [1], the basic elements of an MRFM apparatus are shown in Fig. 2.1; a silicon cantilever and an attached magnetic tip are used to detect the interacting force between the electron spin in a sample and the magnetic tip. The sensing was performed at a low temperature to reduce the noise and the relaxation rate of the spins.

A magnetic field and the inhomogeneous field from the magnetic tip determines the "resonant slice" in the 3D space. This resonant slice corresponds to the point spread function of this imaging system and measures the signal/spin energy if a point in the sample matches the conditions for electron spin resonance. The non-invasive penetration depth of MRFM depends on the size of the resonance slice that is generated under the sensitive resonance condition.

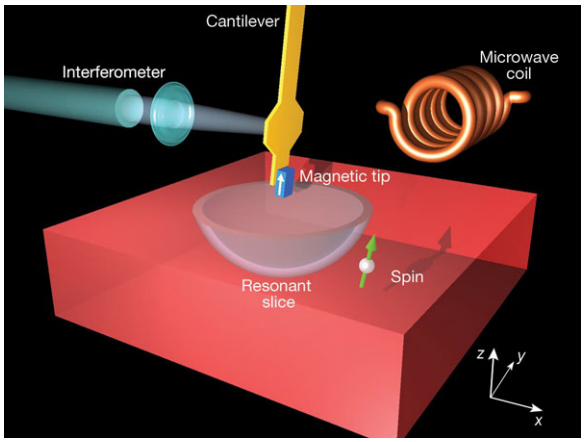


Figure 2.1: The magnetic tip is attached at the end of an ultrasensitive cantilever. The resonant slice represents the area where the field from the magnetic tip satisfies the condition for magnetic resonance. Under the resonance condition, the frequency of the vibrating cantilever shifts correspondingly to the exerted magnetic force caused by the inversion of the spin [1].

To generate a force map, a recently developed protocol, "iOSCAR" (oscillating cantilever-driven adiabatic reversals), has been used [61]. In this protocol, the change in the vibration frequency of the cantilever, corresponding to the tip-sample interactions, is recorded. Then, the average of the square of the retrieved signal (the signal energy) is evaluated, where this energy of the measurement is the sum of energy of spin and noise.

To obtain 3D data, the magnetic tip mechanically scans in three dimensions with respect to the sample. Due to the hemispherical shape of the resonance slice, the retrieved data does not directly exhibit the spin distribution in the sample. Rather, each scanned data point is a weighted contribution of spin signals at different 3-dimensional positions (both laterally and vertically), depending on the relative position of the tip to the sample point.

The capability of MRFM has been made possible by numerous technical advances. Among these advances are the physical modeling of the MRFM PSF and the deconvolution algorithms, which convert magnetic force measurements into a 3D map of particle density. In Chapter III and IV of this thesis, we will concentrate on the PSF

modeling and the related deconvolution for further development of MRFM imaging.

2.2 Sparse Image Reconstruction for MRFM

The MRFM imaging problem is different from others due to the rather special forward model that characterizes the MRFM point spread function and due to the sparsity of atomic scale images of materials. The MRFM image is naturally sparse because molecules sparsely occupy 3D space. Thus, most of the image regions would be empty and only few portions of the image would have significant, observable signals.

Sparse image reconstruction has been studied for over a decade. Recently, this theme has been actively investigated in statistics (LASSO) and signal processing (compressive sensing). The intuitive formulation of the sparse image reconstruction problem is to combine the data fitting term, for deconvolution and denoising, with the sparsity inducing term. A straightforward approach to produce sparse solutions is to constrain the number of the nonzero elements in a solution. However, this problem is non-convex, combinatorial, and NP-hard. Thus this exact approach is usually replaced by a convex continuous optimization.

Among approaches that have been proposed to solve this difficult problem, matching pursuit type algorithms can produce solutions with the desired number of nonzero elements (desired l_0 norm[1] value). However, in image processing applications, these greedy algorithms usually do not efficiently exploit the symmetry of the convolution kernel matrix and the retrieved feasible solutions tend to be overly sparse.

Another approach used to induce sparse solutions is the convexification of the l_0 norm to the l_1 norm. A benefit of this relaxation from l_0 to l_1 is that, by using convex programming, the solution can be efficiently obtained and it converges to the global optimal point. Moreover, in the perspective of high dimensional data analysis [62], the following lemma states that convex solutions, by using a l_1 penalty, bound the

[1]This is not a real norm, but by convention we call it a l_0 norm in this thesis.

optimal solution, under the assumption that the solution is confined in a bounded convex set.

Lemma Let $K_{n,s} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_1 \leq \sqrt{s}\}$ and $S_{n,s} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$. Then, $\text{conv}(S_{n,s}) \subset K_{n,s} \subset 2\text{conv}(S_{n,s})$, where $\text{conv}(K)$ is defined to be the convex hull of a set K.

In this sense, $K_{n,s}$ can be considered as a set of approximately sparse signals, because it is almost the same as the convex hull of $S_{n,s}$.

One popular algorithm using the l_1 norm is the LASSO type estimator [63]. The improved version of this estimator traces the regularization parameter values along the solution path. This tuning of the regularization parameter is not trivial and requires an effort to produce appropriately sparse solutions.

This tuning issue can often be handled by adopting a data-driven approach. The empirical Bayes approach has the capability to automatically estimate the tuning parameter values, provided that the suitable prior distributions are used. This prior knowledge (a prior distribution $\text{P}(X)$) is compensated by the data fidelity term ($\text{P}(Y|X)$, the conditional distribution of Y given X) according to the Bayes rule, as seen in Eq. (2.1). The resulting posterior distribution ($\text{P}(X|Y)$) is then maximized to produce the maximum a posteriori (MAP) estimate.

Bayes Rule

$$\text{P}(X|Y) = \text{P}(Y|X)\text{P}(X)/\text{P}(Y), \tag{2.1}$$

with well-defined random variables X and Y and a probability measure P.

To obtain sparse solutions, one can use connections between a Bayesian prior distribution and the convex penalty term; there is a clear link between a specific Bayesian configuration and convex optimization using the l_1 norm. Assuming a Laplacian distribution as a prior distribution for the true image, the maximization of the posterior

distribution, proportional to the prior distribution and the likelihood function, is equivalent to the minimization of the negative of the log (the argument of the exponential function) of the distribution. For the positive signal values, we can use an exponential distribution, or a positively truncated Laplacian distribution, as the prior distribution for the image. However, the estimation method using random sampling does not produce sparse solutions, because the probability measure at the value 0, $\mathrm{P}(X = 0)$, is zero for Lebesgue continuous distributions such as the exponential distribution. To counter this, several authors have proposed putting a discrete mass at 0 to produce explicit zero values in the unknown signals. In the MRFM literature, this approach was used for Bayesian sparse image reconstruction algorithms [21, 2]. This prior is the starting point for the Bayes sparse reconstruction approach taken in this thesis.

2.3 Semi-blind Image Reconstruction Problems for MRFM

The physical MRFM PSF model, suggested by Mamin *et al.* [8], is represented by a set of functions of several parameters, including the mass of the cantilever probe, the ferromagnetic constant of the probe tip, and external field strength. Using this model and specific parameter values, Mamin *et al.* assumed a nominal and fixed PSF of the microscope for their experiments. The PSF parameters determine the resonance condition to generate a valid PSF. If the condition is not satisfied, then the PSF is degenerate and an observation cannot be made. Moreover, the variation of the PSF with respect to the tuning parameters is large, so the suggested physical PSF model is sensitive to parameter errors. The mismatch between the nominal PSF, generated by the physical model [8], and the true, unknown response of the MRFM tip can exist, due to the imperfection of the model and measurement errors. If this mismatch is not compensated in the MRFM image reconstruction, the reconstruction result will exhibit induced artifacts due to the mismatch. A direct approach of

estimating the parameters in the parametric form of the true PSF is difficult due to the highly non-linear nature of the parametric PSF and the sensitivity of the tuned PSF shape to the parameters. Furthermore, it not practical to take a brute-force approach, which generates possible MRFM PSFs while estimating the 'correct' PSF by trying several parameter values. This is because evaluating the highly non-linear and complex equations of the tuning function is time-consuming and the convergence is not guaranteed. In this thesis, we propose a different semi-blind deconvolution approach to reconstruction of MRFM images with PSF mismatch.

2.4 Blind Deconvolution Problems and Ill-posedness

Before tackling the MRFM imaging problem with PSF mismatch, we review the relevant background on blind deconvolution.

The objective function to minimize, for a simple blind deconvolution problem, can be formulated as follows,

$$L := \|\mathbf{y} - \boldsymbol{\kappa} * \mathbf{x}\|, \tag{2.2}$$

where $\mathbf{y}$ is noisy blurred observation, $\boldsymbol{\kappa}$ is a PSF, $*$ is the convolution operator, and $\mathbf{x}$ is the unknown true signal. There are infinitely many pairs of $(\boldsymbol{\kappa}, \mathbf{x})$ explaining the observed $\mathbf{y}$. Among these pairs, we call the pair of $\delta(\cdot)$ function, or identity convolutional kernel, and the observation itself $\mathbf{y}$, the trivial solution: $(\boldsymbol{\kappa}, \mathbf{x}) = (\delta(\cdot), \mathbf{y})$.

To appreciate the ill-posedness of the solution to $\min_{\mathbf{x},\boldsymbol{\kappa}} L$, consider the one-dimensional blind deconvolution problem that is formulated as follows:

$$L := |y - \kappa x|, \tag{2.3}$$

where the deconvolution in 1D is reduced to multiplication and y and x are scalar variables. Evidently, there exist many solutions that minimize the objective function L, in fact that make it equal to 0. We first note that the objective function L is not

convex; the sublevel set inside the contour, for example $L \leq 0.5$, is not a convex set (Fig. 2.2), and the two by two Hessian matrix of L, by taking derivatives of L with respect to x and κ, is not positive semi-definite.

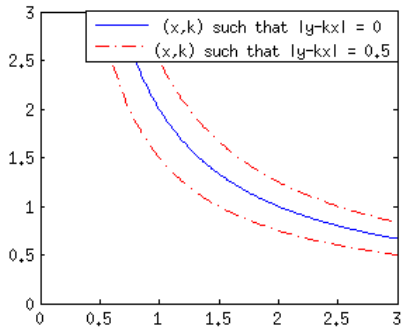


Figure 2.2: A 1-dimensional blind deconvolution problem ($y = 2$). The solution set ($L = 0$) is the blue solid line. The sublevel set inside the contour $L = 0.5$ in red dashes is not convex.

An approach to solving this ill-posed deconvolution problem is to add an ℓ_2-distance image penalty to the objective function of the form $\|\mathbf{x}\|$. This strategy avoids trivial solutions. However, this is not effective since the MAP estimator would favor an image solution $\|\mathbf{x}\| \to 0$ and a PSF solution $\|\boldsymbol{\kappa}\| \to \infty$. This is because the infinitesimal image norm always shrinks the penalty term and the attenuated scale is compensated by the amplified norm of the PSF in the convolution term. This necessitates the regularization of both the image and PSF. Additionally, the joint maximum a posteriori (MAP) estimator $(\hat{\boldsymbol{\kappa}}, \hat{\mathbf{x}})$ suffers from being trapped into the trivial solution. In other words, an MAP estimator, under a broad class of priors, would favor a blurry image ($\mathbf{x} \approx \mathbf{y}$) over a sharp one [64]. To remedy this situation, the marginalized MAP solution for the blur kernel, $\hat{\boldsymbol{\kappa}}$, followed by an MMSE estimate for the image seems promising [64]. Alternatively, an estimator based on edge detection can be used to avoid trivial or degenerate solutions. This is, however, inapplicable to the MRFM model, since the MRFM image characteristics and PSF model show no evidence of image edges or natural motion-blurring type PSFs. In this thesis, the

strategy to constrain the solution space is to reduce the feasible PSF space into much lower dimensions, which is elaborated in the next chapter.

2.5 Minimax Optimization to Blind Deconvolution

As a demonstrative example of a general blind deconvolution solution, we briefly present a minimax approach. The solution using this approach is optimal under the worst case PSF. In the imaging literature, this approach is often criticized as being too conservative and pessimistic. Nonetheless, it is instructive to consider minimax solutions as they yield tight lower bounds on performance over the space of suitably constrained unknown PSFs.

A minimax approach to blind deconvolution is formulated as follows:

$$\min_{\mathbf{x}} \max_{\|\mathbf{D}\|_{\mathbf{W}} \leq \epsilon} \|\mathbf{y} - (\mathbf{H} + \mathbf{D})\mathbf{x}\|_2^2 + c(\mathbf{x}), \tag{2.4}$$

where $\|\mathbf{D}\|_{\mathbf{W}} = \sqrt{tr(\mathbf{D}\mathbf{W}\mathbf{D}^{\mathbf{T}})}$ is defined as a weighted Frobenius norm, $c(\mathbf{x})$ is a penalty term for $\mathbf{x}$, and $\mathbf{W}$ is a diagonal matrix. This problem is also called a robust penalized least squares (RPLS) problem. Here, we do not assume any parameterization of point spread function or any prior knowledge of the system matrix. Rather, we wish to retain the optimal solution under the maximum uncertainty level ϵ in the unknown PSF. The variation $\mathbf{D}$ of the mismatch is constrained in terms of matrix norm, which is stated in the minimax form.

2.5.1 General solution

Using the triangular inequality, we obtain

$$\max_{\|\mathbf{D}\|_{\mathbf{W}} \leq \epsilon} \|\mathbf{y} - (\mathbf{H} + \mathbf{D})\mathbf{x}\|_2 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 + \epsilon\|\mathbf{x}\|_2, \tag{2.5}$$

if the following condition is satisfied

$$\mathbf{D} = -\epsilon \frac{\mathbf{u}\mathbf{x}^{\mathbf{T}}\mathbf{R}}{\|\mathbf{x}\|_{\mathbf{R}}}, \|\mathbf{x}\|_{\mathbf{R}} = \sqrt{\mathbf{x}^{\mathbf{T}}\mathbf{R}\mathbf{x}}\ , \mathbf{R} = \mathbf{W}^{-\mathbf{1}}, \tag{2.6}$$

$$\mathbf{u} = \begin{cases} \frac{\mathbf{y}-\mathbf{H}\mathbf{x}}{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|} & \text{when } \mathbf{y} - \mathbf{H}\mathbf{x} \neq \mathbf{0} \\ \text{Any unit norm vector} & \text{when } \mathbf{y} - \mathbf{H}\mathbf{x} = \mathbf{0}\ . \end{cases} \tag{2.7}$$

Since monotonicity of squared power is preserved, we obtain

$$\max_{\|\mathbf{D}\|_{\mathbf{W}} \leq \epsilon} \|\mathbf{y} - (\mathbf{H} + \mathbf{D})\mathbf{x}\|_2^2 = (\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 + \epsilon\|\mathbf{x}\|_2)^2. \tag{2.8}$$

Then, an equivalent optimization to (2.4) is:

$$\min_{\mathbf{x}} (\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 + \epsilon\|\mathbf{x}\|_2)^2 + c(\mathbf{x}), \tag{2.9}$$

under the condition (2.6) and (2.7).

2.5.2 Solution with an ℓ_1 penalty

When $c(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, a relaxation of the sparsity norm $\|\mathbf{x}\|_0$, we can further simplify the objective function. To deal with the square of sum, we express

$$(a+b)^2 = \min_{0 \leq t \leq 1} \frac{a^2}{t} + \frac{b^2}{1-t}, \tag{2.10}$$

where the minimization is achieved when $t = \frac{a}{a+b}$ and we interpret $x/0$ as 0 if $x = 0$ and ∞ otherwise.

Now the original problem in (2.4) is equivalent to

$$\min_{\mathbf{x}} \min_{0 \leq t \leq 1} \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2}{t} + \frac{\epsilon^2\|\mathbf{x}\|_2^2}{1-t} + \lambda\|\mathbf{x}\|_1, \tag{2.11}$$

where the minimization is achieved when $t = \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2}{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2+\epsilon\|\mathbf{x}\|_{\mathbf{R}}}$.

2.5.2.1 Derivation of nonsingular case

When $\mathbf{y} \neq \mathbf{H}\mathbf{x}$ and $\|\mathbf{x}\|_{\mathbf{R}} \neq \mathbf{0}$, we majorize the objective function as follows:

$L_1 := (\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2 + \epsilon\|\mathbf{x}\|_2)^2$

$= \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2}{t^{opt}} + \frac{\epsilon^2\|\mathbf{x}\|_2^2}{1-t^{opt}}$ (where t^{opt} is the optimal minimizing t)

$\leq \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2}{t} + \frac{\epsilon^2\|\mathbf{x}\|_2^2}{1-t}$

$= \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}'\|_2^2}{t} + \frac{\epsilon^2\|\mathbf{x}'\|_2^2}{1-t} + (\mathbf{x}-\mathbf{x}')^T[-\frac{2}{t}\mathbf{H}^{\mathbf{T}}(\mathbf{y} - \mathbf{H}\mathbf{x}') + \frac{2\epsilon^2}{1-t}\mathbf{x}'] + (\mathbf{x}-\mathbf{x}')^T[\frac{1}{t}\mathbf{H}^{\mathbf{T}}\mathbf{H} + \frac{\epsilon^{\mathbf{2}}}{\mathbf{1-t}}\mathbf{I}](\mathbf{x}-\mathbf{x}')$

$\leq \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}'\|_2^2}{t} + \frac{\epsilon^2\|\mathbf{x}'\|_2^2}{1-t} + (\mathbf{x} - \mathbf{x}')^T[-\frac{2}{t}\mathbf{H}^{\mathbf{T}}(\mathbf{y} - \mathbf{H}\mathbf{x}') + \frac{2\epsilon^2}{1-t}\mathbf{x}'] + [\frac{C}{t} + \frac{\epsilon^2}{1-t}]\|\mathbf{x} - \mathbf{x}'\|^{\mathbf{2}}$

$=: Q(\mathbf{x}; \mathbf{x}')$,

where $C \geq \|\mathbf{H}^{\mathbf{T}}\mathbf{H}\|$, $t := \frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2}{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2+\epsilon\|\mathbf{x}\|_{\mathbf{R}}}$, and the second inequality holds if $\mathbf{x} = \mathbf{x}'$.

Let $a = \frac{C}{t} + \frac{\epsilon^2}{1-t}$ and $\mathbf{e} := \mathbf{y} - \mathbf{H}\mathbf{x}'$. For the evaluation of $\arg\max_{\mathbf{x}} Q(\mathbf{x}; \mathbf{x}')$, the terms of our interest in $Q(\mathbf{x}; \mathbf{x}')$ with respect to $\mathbf{x}$ are: $a\|\mathbf{x}\|^2 + (\frac{2\epsilon^2}{1-t} - 2a)\mathbf{x}^T\mathbf{x}' - \mathbf{x}^T\frac{2}{t}\mathbf{H}^T\mathbf{e}$. The ith component of this equation is $ax_i^2 - 2ax_ix_i' + x_i[-\frac{2}{t}\mathbf{h_i^T}\mathbf{e} + \frac{2\epsilon^2}{1-t}x_i']$, where $\mathbf{h}_i$ is an ith column vector of $\mathbf{H}$. Combining with regularization term, $\lambda\|\mathbf{x}\|_1$, and using quadratic optimization technique with ℓ_1 penalty, we update x_i by using the following soft-thresholding rule:

$$x_i \leftarrow T_{\lambda/2a}[\frac{1}{a}(\frac{1}{t}h_i^T e + \frac{C}{t}x_i')], \tag{2.12}$$

where $T_k(x)$ is a soft-threshold function of x with threshold level k.

2.5.2.2 Derivation of singular case

Two singular cases, $\mathbf{y} = \mathbf{H}\mathbf{x}$ or $\|\mathbf{x}\|_{\mathbf{R}} = \mathbf{0}$, can happen when the estimate is a perfect solution to the unconstrained optimization problem, or the initial guess is zero vector, respectively. The former case is of less concern than the latter, because

we have at least found a good solution to the unconstrained opimization. From the latter case, restarting from 2.9, we can re-initialize $\mathbf{x}$ by using classical optimization methods with a penalty term. For instance, we could use the iterative thresholding with an ℓ_1 norm penalty.

2.5.3 Discussion on minimax approach

We formulate the general minimax problem and provide the solution to the problem with an ℓ_1 penalty for sparse deconvolution. The general tight lower bound is derived in (2.9) with the conditions (2.6) and (2.7). For the minimax problem with an ℓ_1 penalty, we derive an iterative algorithm in (2.12) to find the optimal solution. However, we do not take this approach in this thesis because the minimax solution does not capture the properties of convolution matrix, resulting inferior image quality compared to that from standard deconvolution algorithms using a nominal, slightly mismatched point spread function.

CHAPTER III

A Stochastic Approach to Hierarchical Bayesian Semi-blind Sparse Deconvolution[1]

In this chapter, we present a hierarchical Bayesian approach to semi-blind sparse deconvolution. We construct a stochastic algorithm within a Markov chain Monte Carlo (MCMC) framework. For this algorithm, basic principles of random sampling methods are investigated. We also present a convergent sampling method for the PSF estimation and an extension of the image model using Markov random fields. We start to formulate forward imaging and PSF model.

3.1 Forward Imaging and PSF Model

Let $\mathbf{X}$ denote the $l_1 \times \ldots \times l_n$ unknown n-D positive spin density image to be recovered (e.g., $n = 2$ or $n = 3$) and $\mathbf{x} \in \mathbb{R}^M$ denote the vectorized version of $\mathbf{X}$ with $M = l_1 l_2 \ldots l_n$. This image is to be reconstructed from a collection of $P(= M)$ measurements $\mathbf{y} = [y_1, \ldots, y_P]^T$ via the following noisy transformation:

$$\mathbf{y} = T(\boldsymbol{\kappa}, \mathbf{x}) + \mathbf{n}, \tag{3.1}$$

[1]This chapter is partially based on the papers [37, 13].

where $T(\cdot,\cdot)$ is the n-dimensional convolution operator or the mean response function $\mathbb{E}[\mathbf{y}|\boldsymbol{\kappa},\mathbf{x}]$, $\mathbf{n}$ is a $P \times 1$ observation noise vector and $\boldsymbol{\kappa}$ is the kernel modeling the response of the imaging device.

A typical PSF for MRFM is shown in Mamin *et al.*[8] for horizontal and vertical MRFM tip configurations. In (3.1) $\mathbf{n}$ is an additive Gaussian[2] noise, independent of $\mathbf{x}$, distributed according to $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_P)$. The PSF is assumed to be known up to a perturbation $\Delta\boldsymbol{\kappa}$ about a known nominal $\boldsymbol{\kappa}_0$:

$$\boldsymbol{\kappa} = \boldsymbol{\kappa}_0 + \Delta\boldsymbol{\kappa}. \tag{3.2}$$

In the MRFM application the PSF is described by an approximate parametric function that depends on the experimental setup. Based on the physical parameters (gathered in the vector $\boldsymbol{\zeta}$) tuned during the experiment (external magnetic field, mass of the probe, etc.), an approximation $\boldsymbol{\kappa}_0$ of the PSF can be derived. However, due to model mismatch and experimental errors, the true PSF $\boldsymbol{\kappa}$ may deviate from the nominal PSF $\boldsymbol{\kappa}_0$.

If a vector of the nominal values of parameters $\boldsymbol{\zeta}_0$ for the parametric PSF model $\boldsymbol{\kappa}_{gen}(\boldsymbol{\zeta})$ is known, then direct estimation of a parameter deviation, $\Delta\boldsymbol{\zeta}$, can be performed by evaluation of $\boldsymbol{\kappa}_{gen}(\boldsymbol{\zeta}_0 + \Delta\boldsymbol{\zeta})$. However, in MRFM, as shown by Mamin *et al.* [8], $\boldsymbol{\kappa}_{gen}(\boldsymbol{\zeta})$ is a nonlinear function with many parameters that are required to satisfy 'resonance conditions' to produce a meaningful MRFM PSF. This makes direct estimation of the PSF difficult.

In this thesis, we take a similar approach to the 'basis expansions' in [65, Chap. 5], [27] for approximation of the PSF deviation $\Delta\boldsymbol{\kappa}$ as linear models. Our model for this deviation is that $\Delta\boldsymbol{\kappa}$ can be expressed as a linear combination of elements of an

[2] $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

a priori known basis $\boldsymbol{\kappa}_k$, $k = 1, \ldots, K$,

$$\Delta\boldsymbol{\kappa} = \sum_{k=1}^{K} \lambda_k \boldsymbol{\kappa}_k, \tag{3.3}$$

where $\{\boldsymbol{\kappa}_k\}_{k=1,\ldots,K}$ is a set of basis functions for the PSF perturbations and λ_k, $k = 1, \ldots, K$ are unknown coefficients. To emphasize the influence of these coefficients on the actual PSF, $\boldsymbol{\kappa}$ will be denoted $\boldsymbol{\kappa}(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]^T$. With these notations, (3.1) can be rewritten:

$$\mathbf{y} = T(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x}) + \mathbf{n} = \mathbf{H}(\boldsymbol{\lambda})\,\mathbf{x} + \mathbf{n}, \tag{3.4}$$

where $\mathbf{H}(\boldsymbol{\lambda})$ is an $P \times M$ matrix that describes convolution by the PSF kernel $\boldsymbol{\kappa}(\boldsymbol{\lambda})$.

We next address the problem of estimating the unobserved image $\mathbf{x}$ and the PSF perturbation $\Delta\boldsymbol{\kappa}$ under sparsity constraints given the measurement $\mathbf{y}$ and the bilinear function $T(\cdot, \cdot)$.

3.2 Hierarchical Bayesian Model

3.2.1 Likelihood function

Under the hypothesis that the noise in (3.1) is Gaussian, the observation model likelihood function takes the form

$$f\left(\mathbf{y}|\mathbf{x}, \boldsymbol{\lambda}, \sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{P}{2}} \exp\left(-\frac{\|\mathbf{y} - T(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x})\|^2}{2\sigma^2}\right), \tag{3.5}$$

where $\|\cdot\|$ denotes the standard ℓ_2 norm: $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x}$. This likelihood function will be denoted $f(\mathbf{y}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\mathbf{x}, \boldsymbol{\lambda}, \sigma^2\}$.

3.2.2 Parameter prior distributions

In this section, we introduce prior distributions for the parameters $\boldsymbol{\theta}$.

3.2.2.1 Image prior

As the prior distribution for x_i, we adopt a mixture of a mass at zero and a single-sided exponential distribution:

$$f\left(x_i|w,a\right) = (1-w)\delta\left(x_i\right) + \frac{w}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}\left(x_i\right), \tag{3.6}$$

where $w \in [0,1], a \in [0,\infty), \delta\left(\cdot\right)$ is the Dirac function, $\mathbb{R}_+^*$ is a set of real open interval $(0,\infty)$ and $\mathbf{1}_{\mathbb{E}}\left(x\right)$ is the indicator function of the set $\mathbb{E}$:

$$\mathbf{1}_{\mathbb{E}}\left(x\right) = \begin{cases} 1, & \text{if } x \in \mathbb{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

By assuming the components x_i to be a conditionally independent $(i = 1,\ldots,M)$ given w, a, the following conditional prior distribution is obtained for the image $\mathbf{x}$:

$$f\left(\mathbf{x}|w,a\right) = \prod_{i=1}^{M}\left[(1-w)\delta\left(x_i\right) + \frac{w}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}\left(x_i\right)\right]. \tag{3.8}$$

This image prior is similar to the LAZE distribution (weighted average of a Laplacian probability density function (pdf) and an atom at zero) used, for example, in Ting *et al.* [66, 21]. As motivated by Ting *et al.* and Dobigeon *et al.* [21, 2], the image prior in (3.8) has the interesting property of enforcing some pixel values to be zero[3], reflecting the natural sparsity of the MRFM images. Furthermore, the proposed prior in (3.6) ensures positivity of the pixel values (spin density) to be estimated.

[3]A Laplace distribution as a image prior distribution acting as a sparse regularization using ℓ_1 penalty does not produce zero values by using MCMC sampling methods because the Lebesque measure at a point of this continuous density is zero.

3.2.2.2 PSF parameter prior

We assume that the parameters $\lambda_1, \ldots, \lambda_K$ are *a priori* independent and uniformly distributed over known intervals associated with error tolerances centered at 0. Specifically, define the interval

$$\mathcal{S}_k = [-\Delta\lambda_k, \Delta\lambda_k] \tag{3.9}$$

and define the distribution of $\boldsymbol{\lambda}$ as

$$f(\boldsymbol{\lambda}) = \prod_{k=1}^{K} \frac{1}{2\Delta\lambda_k} \mathbf{1}_{\mathcal{S}_k}(\lambda_k), \tag{3.10}$$

with $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]^T$. In our experiment, $\Delta\lambda_k$'s are set to be large enough to be non-informative, i.e., an improper, flat prior.

3.2.2.3 Noise variance prior

A non-informative Jeffreys' prior is selected as prior distribution for the noise variance:

$$f\left(\sigma^2\right) \propto \frac{1}{\sigma^2} \tag{3.11}$$

This model is equivalent to an inverse gamma prior with a non-informative Jeffreys' hyperprior, which can be seen by integrating out the variable corresponding to the hyperprior [2].

3.2.3 Hyperparameter priors

Define the hyperparameter vector associated with the image and noise variance prior distributions as $\boldsymbol{\Phi} = \{a, w\}$. In our hierarchical Bayesian framework, the estimation of these hyperparameters requires prior distributions in the hyperparameters.

3.2.3.1 Hyperparameter a

A conjugate inverse-Gamma distribution is assumed for hyperparameter a:

$$a|\boldsymbol{\alpha} \sim \mathcal{IG}\left(\alpha_0, \alpha_1\right), \tag{3.12}$$

with $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^T$. The fixed hyperparameters α_0 and α_1 have been chosen to produce a vague prior, i.e., $\alpha_0 = \alpha_1 = 10^{-10}$.

3.2.3.2 Hyperparameter w

A uniform distribution on the simplex $[0, 1]$ is selected as prior distribution for the mean proportion of non-zero pixels:

$$w \sim \mathcal{U}\left([0, 1]\right). \tag{3.13}$$

Assuming that the individual hyperparameters are independent the full hyperparameter prior distribution for $\boldsymbol{\Phi}$ can be expressed as:

$$\begin{aligned} f\left(\boldsymbol{\Phi}|\boldsymbol{\alpha}\right) &= f\left(w\right) f\left(a\right) \\ &\propto \frac{1}{a^{\alpha_0+1}} \exp\left(-\frac{\alpha_1}{a}\right) \mathbf{1}_{[0,1]}\left(w\right) \mathbf{1}_{\mathbb{R}^+}\left(a\right), \end{aligned} \tag{3.14}$$

3.2.4 Posterior distribution

The posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ is:

$$f\left(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{y}\right) \propto f\left(\mathbf{y}|\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) f\left(\boldsymbol{\Phi}\right), \tag{3.15}$$

with

$$f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) = f\left(\mathbf{x}|a, w\right) f\left(\boldsymbol{\lambda}\right) f\left(\sigma^2\right), \tag{3.16}$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{\Phi})$ have been defined in (3.5) and (3.14). The conjugacy of priors in this hierarchical structure allows one to integrate out the parameters σ^2, and the hyperparameter $\boldsymbol{\Phi}$ in the full posterior distribution (3.15), yielding:

$$f(\mathbf{x}, \boldsymbol{\lambda}|\mathbf{y}, \alpha_0, \alpha_1) \propto \int f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{y}) \, \mathrm{d}w \mathrm{d}a \mathrm{d}\sigma^2 \propto \tag{3.17}$$
$$\frac{\mathcal{B}e(1+n_1, 1+n_0)}{\|\mathbf{y} - T(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x})\|^N} \frac{\Gamma(n_1+\alpha_0)}{(\|\mathbf{x}\|_1 + \alpha_1)^{n_1+\alpha_0}} \prod_{k=1}^{K} \frac{1}{2\Delta\lambda_k} \mathbf{1}_{\mathcal{S}_k}(\lambda_k),$$

where $\mathcal{B}e$ is the beta function and Γ is the gamma function.

The next section presents the Metropolis-within-Gibbs algorithm [67] that generates samples distributed according to the posterior distribution $f(\mathbf{x}, \boldsymbol{\lambda}|\mathbf{y})$. These samples are then used to estimate $\mathbf{x}$ and $\boldsymbol{\lambda}$.

3.3 Metropolis-within-Gibbs Algorithm for Semi-blind Sparse Image Reconstruction

We describe in this section a Metropolis-within-Gibbs sampling strategy that allows one to generate samples $\left\{\mathbf{x}^{(t)}, \boldsymbol{\lambda}^{(t)}\right\}_{t=1,\ldots}$ distributed according to the posterior distribution in (3.17). As sampling directly from (3.17) is a difficult task, we will instead generate samples distributed according to the joint posterior $f(\mathbf{x}, \boldsymbol{\lambda}, \sigma^2|\mathbf{y}, \alpha_0, \alpha_1)$. Sampling from this posterior distribution is done by alternatively sampling one of $\mathbf{x}, \boldsymbol{\lambda}, \sigma^2$ conditioned on all other variables [68, 2]. The contribution of this work to [2] is to present an algorithm that simultaneously estimates both the image and PSF. The algorithm results in consistently stable output images and PSFs.

The main steps of our proposed sampling algorithm are given in subsections 3.3.1 through 3.3.3 (see also Algorithm 1).

Algorithm 1 Metropolis-within-Gibbs sampling algorithm for semi-blind sparse image reconstruction

% Initialization:
Sample the unknown image $\mathbf{x}^{(0)}$ from pdf in (3.8),
Sample the noise variance $\widetilde{\sigma}^{2(0)}$ from the pdf in (3.11),
% Iterations:
for $t = 1, 2, \ldots,$ **do**
    Sample hyperparameter $w^{(t)}$ from the pdf in (3.19),
    Sample hyperparameter $a^{(t)}$ from the pdf in (3.20),
    For $i = 1, \ldots, M$, sample the pixel intensity $x_i^{(t)}$ from the pdf in (3.21),
    For $k = 1, \ldots, K$, sample the PSF parameter $\lambda_k^{(t)}$ from the pdf in (3.23), by using Metropolis-Hastings step (see Algo. 2),
    Sample the noise variance $\widetilde{\sigma}^{2(t)}$ from the pdf in (3.26),
end for

3.3.1 Generation of samples according to $f(\mathbf{x}\,|\boldsymbol{\lambda}, \sigma^2, \mathbf{y}, \alpha_0, \alpha_1)$

To generate samples distributed according to $f(\mathbf{x}\,|\boldsymbol{\lambda}, \sigma^2, \mathbf{y}, \alpha_0, \alpha_1)$, it is convenient to sample according to $f(\mathbf{x}, w, a\,|\boldsymbol{\lambda}, \sigma^2, \mathbf{y}, \alpha_0, \alpha_1)$ by the following 3-step procedure.

3.3.1.1 Generation of samples according to $f(w\,|\mathbf{x}, \alpha_0, \alpha_1)$

The conditional posterior distribution of w is

$$f(w\,|\mathbf{x}) \propto (1-w)^{n_0+1-1} w^{n_1+1-1}, \tag{3.18}$$

where $n_1 = \|\mathbf{x}\|_0$ and $n_0 = M - \|\mathbf{x}\|_0$. Therefore, generation of samples according to $f(w\,|\mathbf{x})$ is achieved as follows:

$$w|\mathbf{x} \sim \mathcal{B}e\,(1+n_1, 1+n_0)\,. \tag{3.19}$$

3.3.1.2 Generation of samples according to $f(a\,|\mathbf{x})$

The joint posterior distribution (3.15) yields:

$$a\,|\mathbf{x}, \alpha_0, \alpha_1 \sim \mathcal{IG}\,(\|\mathbf{x}\|_0 + \alpha_0, \|\mathbf{x}\|_1 + \alpha_1)\,. \tag{3.20}$$

3.3.1.3 Generation of samples according to $f(\mathbf{x}\,|w, a, \lambda, \sigma^2, \mathbf{y})$

The posterior distribution of each component x_i $(i = 1, \ldots, M)$ given all other variables is derived as:

$$f\left(x_i|w, a, \lambda, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right) \propto (1 - w_i)\delta\left(x_i\right) + w_i\phi_+\left(x_i|\mu_i, \eta_i^2\right), \tag{3.21}$$

where $\mathbf{x}_{-i}$ stands for the vector $\mathbf{x}$ whose ith component has been removed and μ_i and η_i^2 are defined as the following:

$$\eta_i^2 = \frac{\sigma^2}{\|\mathbf{h}_i\|^2}, \mu_i = \eta_i^2\left(\frac{\mathbf{h}_i^T\mathbf{e}_i}{\sigma^2} - \frac{1}{a}\right), \tag{3.22}$$

with $\mathbf{h}_i$ and $\mathbf{e}_i$ defined in Appendix 3.6.

In (3.21), $\phi_+\left(\cdot, m, s^2\right)$ stands for the pdf of the truncated Gaussian distribution defined on $\mathbb{R}_+^*$ with hidden mean m and hidden variance s^2. Therefore, from (3.21), $x_i|w, a, \lambda, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}$ is a Bernoulli truncated-Gaussian variable with parameter (w_i, μ_i, η_i^2).

To summarize, generation of samples distributed according to $f\left(\mathbf{x}\,|w, \sigma^2, a, , \mathbf{y}\right)$ can be performed by updating the coordinates of $\mathbf{x}$ using M Gibbs moves (requiring generation of Bernoulli truncated-Gaussian variables). A pixel-wise fast and recursive sampling strategy is presented in Appendix 3.6 and an accelerated sparsity enforcing simulation scheme is described in Appendix 3.7. The methods in these two appendices are proposed for efficient drawing of samples.

3.3.2 Generation of samples according to $f(\boldsymbol{\lambda}|\mathbf{x}, \sigma^2, \mathbf{y})$

The posterior distribution of the parameter λ_k conditioned on the unknown image $\mathbf{x}$, the noise variance σ^2 and the other PSF parameters $\{\lambda_j\}_{j\neq k}$ is

$$f\left(\lambda_k|\boldsymbol{\lambda}_{-k}, \mathbf{x}, \sigma^2, \mathbf{y}\right) \propto \exp\left[-\frac{\|\mathbf{y} - T(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x})\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathcal{S}_k}(\lambda_k), \tag{3.23}$$

with $\boldsymbol{\lambda}_{-k} = \{\lambda_j\}_{j\neq k}$. We summarize in Algorithm 2 a procedure for generating samples distributed according to the posterior in (3.23) using a Metropolis-Hastings step with a random walk proposition [67] from a centered Gaussian distribution. In order to sample efficiently, the detailed procedure of how to choose an appropriate value of the variance s_k^2 of the proposal distribution for λ_k is described in Section 3.3.4. At iteration t of the algorithm, the acceptance probability of a proposed state $\lambda_k^\star$ is:

$$\rho_{\lambda_k^{(t)} \to \lambda_k^\star} = \min\left(1, a_k \mathbf{1}_{\mathcal{S}_k}(\lambda_k^\star)\right), \tag{3.24}$$

with

$$2\sigma^2 \log a_k = \left\|\mathbf{y} - T\left(\boldsymbol{\kappa}\left(\lambda_k^{(t)}\right), \mathbf{x}\right)\right\|^2 - \|\mathbf{y} - T(\boldsymbol{\kappa}(\lambda_k^\star), \mathbf{x})\|^2. \tag{3.25}$$

Computing the transformation $T(\cdot, \cdot)$ at each step of the sampler can be computationally costly. Appendix 3.6 provides a recursive strategy to efficiently sample according to $f(\boldsymbol{\lambda}|\mathbf{x}, \sigma^2, \mathbf{y})$.

Algorithm 2 Sampling according to $f(\lambda_k|\boldsymbol{\lambda}_{-k}, \mathbf{x}, \sigma^2, \mathbf{y})$

1: Sample $\varepsilon \sim \mathcal{N}\left(0, s_p^2\right)$,
2: Propose $\lambda_k^\star$ according to $\lambda_k^\star = \lambda_k^{(t)} + \varepsilon$,
3: Draw $u \sim \mathcal{U}([0, 1])$,
4: Set $\lambda_k^{(t+1)} = \begin{cases} \lambda_k^\star, & \text{if } u \leq \rho_{\lambda_k^{(t)} \to \lambda_k^\star}, \\ \lambda_k^{(t)}, & \text{otherwise.} \end{cases}$

where $\mathcal{U}(\mathbb{E})$ stands for the uniform distribution on the set $\mathbb{E}$.

3.3.3 Generation of samples according to $f\left(\sigma^2 \left| \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda} \right.\right)$

Samples $(\sigma^2)^{(t)}$ are generated according to the inverse gamma posterior

$$f(\sigma^2 \left| \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda} \right.) = \mathcal{IG}\left(\frac{P}{2}, \frac{\left\| \mathbf{y} - T\left(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x}\right)\right\|^2}{2}\right). \tag{3.26}$$

3.3.4 Adaptive tuning of an acceptance rate in the random-walk sampling

For an efficient sampling of $\lambda_k, k = 1, \ldots, K$, from the desired distribution $\pi(\lambda_k) = f\left(\lambda_k | \boldsymbol{\lambda}_{-k}, \mathbf{x}, \sigma^2, \mathbf{y}\right)$, we need to properly tune the acceptance rate of the samples from the proposal distribution. A careful selection of a step size is critical for convergence of the method. For example, if the step size is too large, most of the iterations will be rejected and the sampling algorithm will be inefficient. On the other hand, if the step size is too small, most of the random walk moves are accepted but these moves are slow to cover the probable space of the desired distribution, and the method is once again inefficient.

The transition density of Metropolis-Hastings sampling is $q(\lambda^{(t)}, \lambda^{\star(t)}) acc(\lambda^{(t)}, \lambda^{\star(t)})$, where $q(\lambda^{(t)}, \lambda^{\star(t)})$ is the proposal density from $\lambda^{(t)}$ and $acc(\lambda^{(t)}, \lambda^{\star(t)})$ is the acceptance probability for the move from $\lambda^{(t)}$ to $\lambda^{\star(t)}$. Here we denote λ_k by λ without a subscript for simplicity. We set $q(\lambda^{(t)}, \lambda^{\star(t)})$ to be a Gaussian density function of $\lambda^{\star(t)}$, denoted by $q(\lambda^{(t)}, \lambda^{\star(t)}) = q(\lambda^{\star(t)} - \lambda^{(t)}) = \phi(\lambda^{\star(t)}; \lambda^{(t)}, s^2)$ with a mean $\lambda^{(t)}$ and a variance s^2, which produces a random walk sample path. Since $q(\cdot, \cdot)$ is symmetrical, $acc_s(\lambda^{(t)}, \lambda^{\star(t)}) = \min\left(1, \frac{\pi(\lambda^{\star(t)}) q(\lambda^{\star(t)}, \lambda^{(t)})}{\pi(\lambda^{(t)}) q(\lambda^{(t)}, \lambda^{\star(t)})}\right) = \min\left(1, \frac{\pi(\lambda^{\star(t)})}{\pi(\lambda^{(t)})}\right) = \rho_{\lambda^{(t)} \to \lambda^{\star(t)}}$, as derived in (3.24). Then the acceptance probability from a parameter value $\lambda^{(t)}$ is

$$acc_s(\lambda^{(t)}) = \int_{\lambda^{\star(t)}} q(\lambda^{(t)}, \lambda^{\star(t)}) acc_s(\lambda^{(t)}, \lambda^{\star(t)}) d\lambda^{\star(t)}.$$

The acceptance rate with a scale parameter s, acting as a step size, can be expressed as $acc_s = \int_\lambda \pi(\lambda) acc_s(\lambda) d\lambda$.

We evaluate these integrations by using Monte Carlo methods;

$acc_s \approx \frac{1}{n_1}\sum_{t=1}^{n_1} acc_s(\lambda^{(t)})$ with $\lambda^{(t)} \sim \pi(\lambda^{(t)})$, and $acc_s(\lambda^{(t)}) \approx \frac{1}{n_2}\sum_{t=1}^{n_2} acc_s(\lambda^{(t)}, \lambda^{\star(t)})$ with $\lambda^{\star(t)} \sim q(\lambda^{(t)}, \lambda^{\star(t)})$. In practice, this empirical version of the integration value is evaluated as

$$acc_s \approx \frac{1}{W}\sum_{t=1}^{W} acc_s(\lambda^{(t)}, \lambda^{\star(t)}), \tag{3.27}$$

after the burn-in period. Therefore we can evaluate the acceptance rate with s by averaging the Boolean variables of $acc_s(\lambda^{(t)}, \lambda^{\star(t)}), t = 1, \ldots, W$, over a time-frame window of length W with realizations $\{\lambda^{(t)}, \lambda^{\star(t)}\}_t$. In short, we iteratively update s to achieve an appropriate acceptance rate, acc_s, as described in Algorithm 3:

Algorithm 3 Tuning s in the Gaussian proposal density $q(\cdot, \cdot)$

Select upper and lower limits acc_H and acc_L. At each time $t = W, 2W, 3W, \ldots$, tune s via the following steps:

1: Evaluate acc_s using (3.27) for the given time-frame window,

2: Update $s \leftarrow \begin{cases} s \times c, & \text{if } acc_s > acc_H, \\ s \div c, & \text{if } acc_s < acc_L, \\ s, & \text{otherwise.} \end{cases}$

In practice, we fix the variance of the instrumental distribution at the end of a burn-in period. Consequently, the transition kernel will be fixed and this guarantees both ergodicity and stationary distribution. In our experiment, we set a conservative acceptance range, $acc_H = 0.6, acc_L = 0.4$, referring to [67], and $W = 20, c = 4$. This strategy can also be applied to the direct parameter estimation described in Section 3.3.5.

If one wants to update the variance after the burn-in period and ensure the ergodicity and stationary distribution, one can use the proposed update rules. In [69], an adaptive version of the random walk Metropolis algorithm is proposed. This algorithm can automatically tunes s such that the asymptotic acceptance rate converges to the target rate, under various regularity conditions. The adaptive random walk step size $s^{(t)}$, the standard variation of the proposed distribution, is updated to $s^{(t+1)}$

as follows:

$$s' = s^{(t)} + \gamma(acc_{s^{(t)}}(\lambda^{(t)}, \lambda^{\star(t)}) - \tau), \tag{3.28}$$

$$s^{(t+1)} = p(s'), \tag{3.29}$$

where τ is a target acceptance rate, $(\gamma^{(t)})$ is a positive sequence of real numbers such that $\gamma^{(t)} \propto t^{-\lambda}$ for $1/2 < \lambda \leq 1$, and $p(s)$ is a projection function such that $p(s) = s$ for $s \in [a, b]$ $(0 < a < b), p(s) = a$ for $s < a$, and $p(s) = b$ for $s > b$. The values for a and b can be set in a conservative way to ensure the optimal s is in $[a, b]$.

3.3.5 Direct sampling of PSF parameter values

As described in Section 2.3, in the MRFM experiments, the direct estimation of PSF parameters is difficult because of the nonlinearity of $\boldsymbol{\kappa}_{gen}$ and the slow evaluation of $\boldsymbol{\kappa}_{gen}(\boldsymbol{\zeta}')$ given a candidate value $\boldsymbol{\zeta}'$. However, if $\boldsymbol{\kappa}_{gen}$ is simple and evaluated quickly, then a direct sampling of parameter values can be performed. To apply this sampling method, instead of calculating a linearized convolution kernel, $\boldsymbol{\kappa}(\boldsymbol{\lambda})$, we evaluate the exact model PSF, $\boldsymbol{\kappa}_{gen}(\boldsymbol{\zeta})$, in (3.23) and (3.25). Also the proposed parameter vector $\boldsymbol{\zeta}^\star$ correspondingly replaces a coefficient vector $\lambda^\star$ and the updated PSF is used in the estimation of other variables. This strategy turns out to be similar to the approach adopted by Orieux *et al.* in [28].

3.4 Results

In this section we present simulation results that compare the proposed semi-blind Bayesian deconvolution algorithms of Section 3.3 with the non-blind method [2], the AM algorithm [3], and other blind deconvolution methods. Here a nominal PSF $\boldsymbol{\kappa}_0$ was selected that corresponds to the mathematical MRFM point response model proposed by Mamin *et al.* [8].

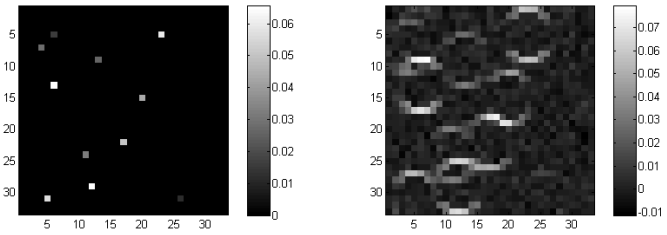
Using our MCMC algorithm described in Sec. 3.3, the MMSE estimators of image and PSF are approximated by ensemble averages over the generated samples after the burn-in period. The joint MAP estimator is selected among the drawn samples, after the stationary distribution is achieved, such that it maximizes the posterior distribution $f(\mathbf{x}, \boldsymbol{\lambda}|\mathbf{y})$ [70].

Table 3.1: Parameters used to compute the MRFM PSF.

Parameter		Value
Description	Name	
Amplitude of external magnetic field	B_{ext}	9.4×10^3 G
Value of B_{mag} in the resonant slice	B_{res}	1.0×10^4 G
Radius of tip	R_0	4.0 nm
Distance from tip to sample	d_0	6.0 nm
Cantilever tip moment	m	4.6×10^5 emu
Peak cantilever oscillation oscillation	x_{pk}	0.8 nm
Maximum magnetic field gradient	G_{max}	125

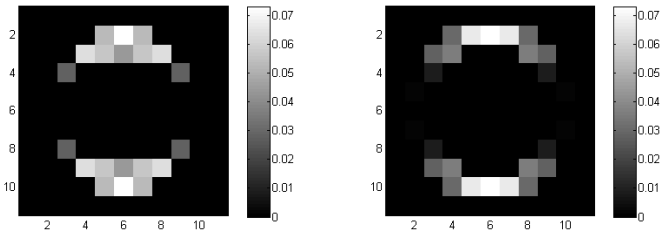
3.4.1 Results on simulated sparse images

We performed simulations of MRFM measurements for PSF and image models similar to those described in Dobigeon *et al.* [2]. The signal-to-noise ratio was set to SNR = 10dB. Several 32×32 synthetic sparse images, one of which is depicted in Fig. 3.1(a), were used to produce the data and were estimated using the proposed Bayesian method. The assumed PSF $\boldsymbol{\kappa}_0$ is generated following the physical model described in Mamin *et al.* [8] when the physical parameters are tuned to the values displayed in Table 3.1. This yields a 11×11 2-dimensional convolution kernel represented in Fig. 3.2(a). We assume that the true PSF $\boldsymbol{\kappa}$ comes from the same physical model where the radius of the tip and the distance from tip to sample have been mis-specified with 2% error as $R = R_0 - 2\% = 3.92$ and $d = d_0 + 2\% = 6.12$. This



(a) Sparse true image ($\|\mathbf{x}\|_0 = 11$) (b) Raw MRFM observation

Figure 3.1: Simulated true image and MRFM raw image exhibiting superposition of point responses (see Fig. 3.2) and noise.



(a) Nominal MRFM PSF (b) True MRFM PSF

Figure 3.2: Assumed PSF $\boldsymbol{\kappa}_0$ and actual PSF $\boldsymbol{\kappa}$.

leads to the convolution kernel depicted in Fig. 3.2(b). The observed measurements $\mathbf{y}$, shown Fig. 3.1(b) are a 32×32 image of size $P = 1024$.

The proposed algorithm requires the definition of K basis vectors $\boldsymbol{\kappa}_k$, $k = 1, \ldots, K$, that span a subspace representing possible perturbations $\Delta\boldsymbol{\kappa}$. We empirically determined this basis using the following PSF variational eigendecomposition approach. A set of 5000 experimental PSFs $\tilde{\boldsymbol{\kappa}}_j$, $j = 1, \ldots, 5000$, were generated following the model described in Mamin *et al.* [8] with parameters d and R randomly drawn according to Gaussian distribution[4] centered at the nominal values d_0, R_0, respectively. Then

[4]We used a PSF generator provided by Dan Rugar's group at IBM [8]. The variances of the Gaussian distributions are carefully tuned so that their standard deviations produce a minimal volume ellipsoid that contains the set of valid PSF's of the form specified in [8].

a standard principal component analysis (PCA) of the residuals $\{\tilde{\boldsymbol{\kappa}}_j - \boldsymbol{\kappa}_0\}_{j=1,\ldots,5000}$, by allowing the maximum variance over the parameters that produce valid MRFM PSFs, was used to identify $K = 4$ principal axes that are associated with the basis vectors $\boldsymbol{\kappa}_k$. The necessary number of basis vectors, $K = 4$ here, was determined empirically by detecting a knee at the scree plot shown in Fig. 3.3. The first four eigenfunctions, corresponding to the first four largest eigenvalues, explain 98.69% of the observed perturbations. The 4 principal patterns of basis vectors are depicted in Fig. 3.4.

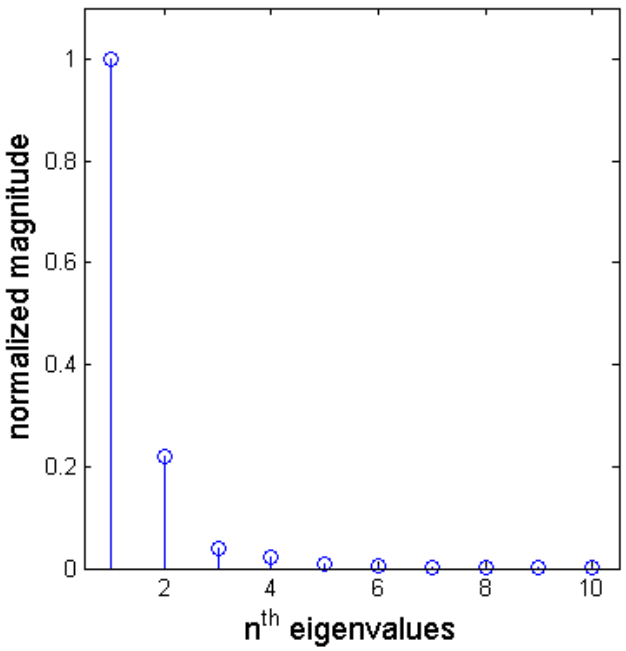


Figure 3.3: Scree plot of residual PCA approximation error in l_2 norm (magnitude is normalized up to the largest value, i.e. $\lambda_{max} := 1$).

The proposed semi-blind Bayesian reconstruction algorithm was applied to estimate both the sparse image and the PSF coefficients of $\boldsymbol{\kappa}_k$'s, using the prior in (3.6). From the observation shown in Fig. 3.1(b) the PSF estimated by the proposed algorithm is shown in Fig. 3.5(a) and is in good agreement with the true one. The corresponding maximum *a posteriori* estimate (MAP) of the unknown image is depicted in Fig. 3.6(a). The obtained coefficients of the PSF-eigenfunctions are close to

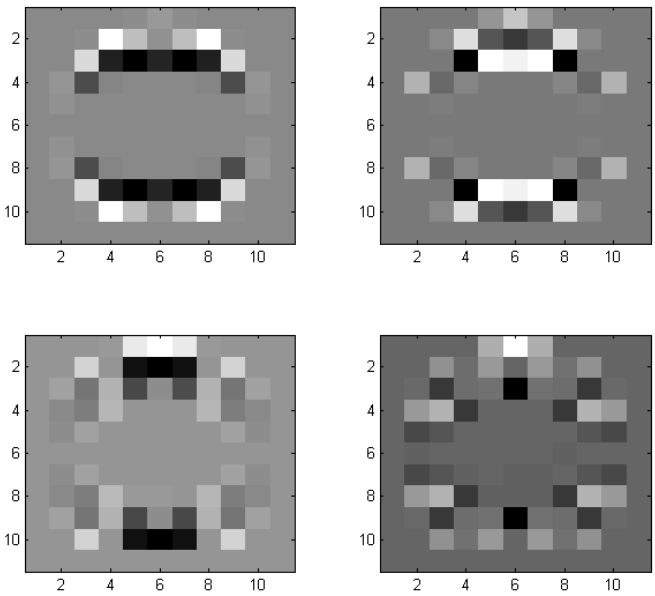


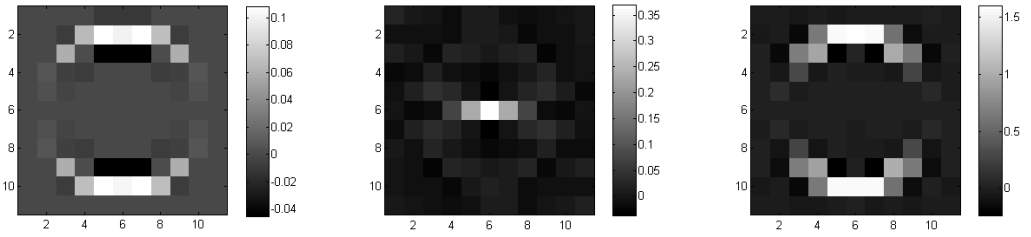
Figure 3.4: The $K = 4$ principal vectors ($\boldsymbol{\kappa}_k$) of the perturbed PSF, identified by PCA.

true coefficients (Fig. 3.7).

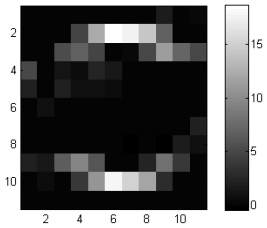
3.4.2 Comparison to other methods

For comparison to a non-blind method, Fig. 3.6(b) shows the estimate using the Bayesian non-blind method [2] with a mismatched PSF. Fig. 3.6(c) shows the estimate generated by the AM algorithm. The nominal PSF described in Section 3.4.1 is used in the AM algorithm and hereafter in other semi-blind algorithms, and the parameter values of AM algorithm were set empirically according to the procedure in [3]. Our proposed algorithm visually appears to outperform the others (Fig. 3.6) while preserving fast convergence (Fig. 3.7).

Quantitative comparisons were obtained by generating different noises in 100 independent trials for a fixed true image. Here, six different true images with six corresponding different sparsity levels ($\|\mathbf{x}\|_0 = 6, 11, 18, 30, 59, 97$) were tested. Fig. 3.8 presents the two histograms of the results with the six sets in the corresponding two



(a) Proposed method (b) Amizic's method (c) Almeida's method



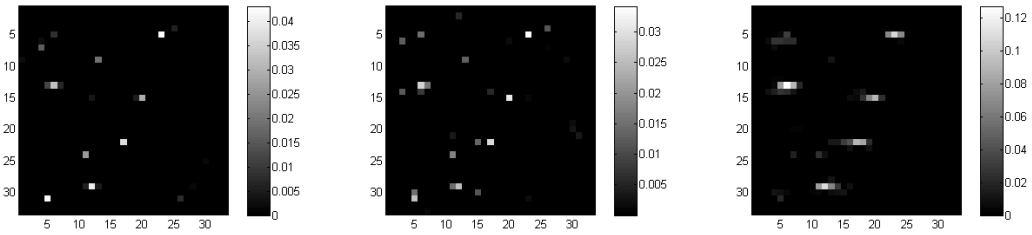
(d) Tzikas' method

Figure 3.5: Estimated PSF of MRFM tip using our semi-blind method, Amizic's method (using TV prior), Almeida's method (using progressive regularization), and Tzikas' method (using the kernel basis PSF model), respectively. For fairness, we used sparse image priors for the methods. (See Section 3.4.2 for details on the methods.)

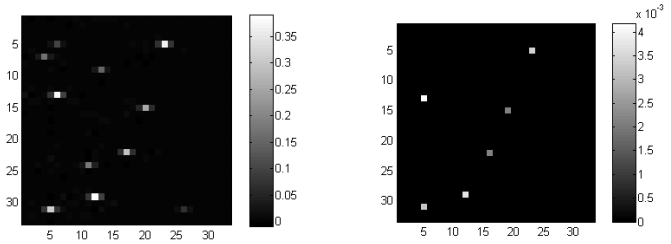
error criteria, $\|\hat{\mathbf{x}} - \mathbf{x}\|^2$, $\|\hat{\mathbf{x}}\|_0$, respectively, both of which indicate that our method performs better and is more stable than the other two methods.

Fig. 3.9 shows reconstruction error performance for several measures of error used in Ting *et al.* [21] and Dobigeon *et al.* [2] to compare different reconstruction algorithms for sparse MRFM images. Notably, compared to the AM algorithm that aims to compensate 'blindness' of the unknown PSF and the previous non-blind method, our method reveals a significant performance gain under most of the investigated performance criteria and sparsity conditions.

In addition to the AM and non-blind comparisons shown in Fig. 3.8, we made direct comparisons between our sparse MRFM reconstruction method and several



(a) MAP, proposed method (b) MAP, Bayesian non-blind method with $\boldsymbol{\kappa}_0$ (c) AM



(d) Almeida's method (e) Tzikas' method

Figure 3.6: The true sparse image and estimated images from Bayesian non-blind, AM, our semi-blind, Almeida's, and Tzikas' methods.

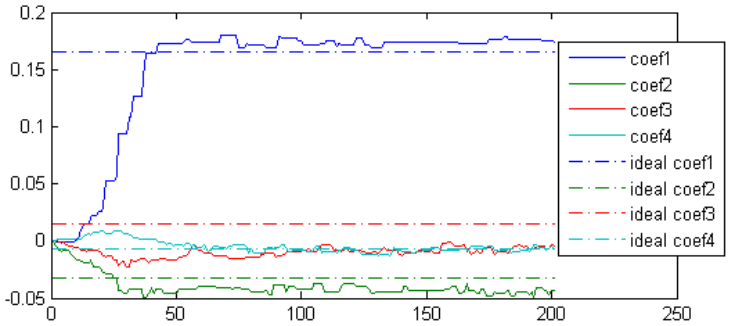
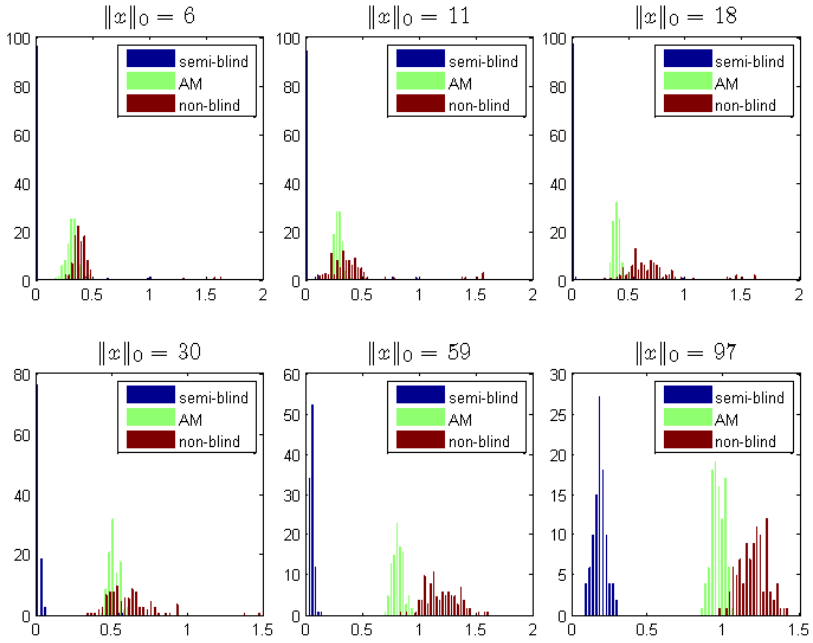
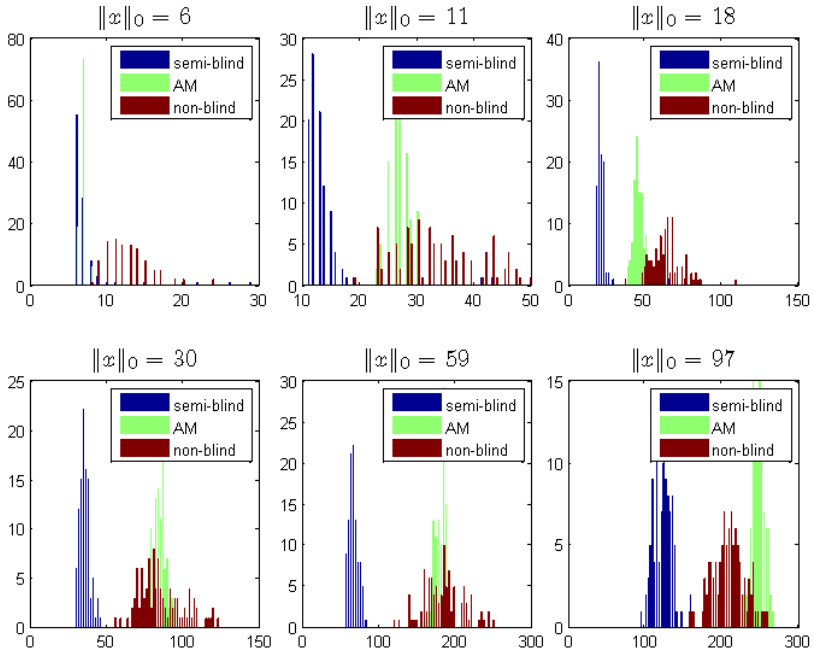


Figure 3.7: Estimated PSF coefficients for 4 PCs over 200 iterations. These curves show fast convergence of our algorithm. 'Ideal coefficients' are the projection values of the true PSF onto the space spanned by four principal PSF basis.

state-of-the-art blind image reconstruction methods [40, 41, 27, 19, 20]. In all cases the algorithms were initialized with the nominal, mismatched PSF and were applied to a sparse MRFM-type image like in Fig. 3.1. For a fair comparison, we made a

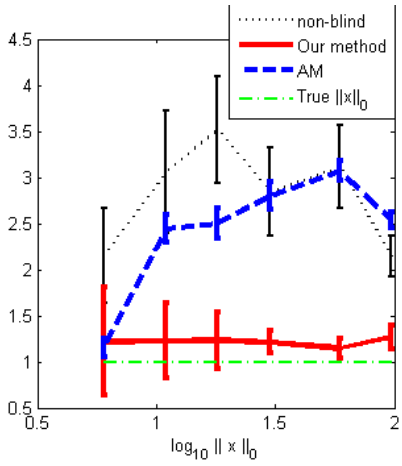


(a) Histograms of the normalized l_2 norm errors. x-axis is $\|\frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|}\|_2^2/\|\mathbf{x}\|_0$.

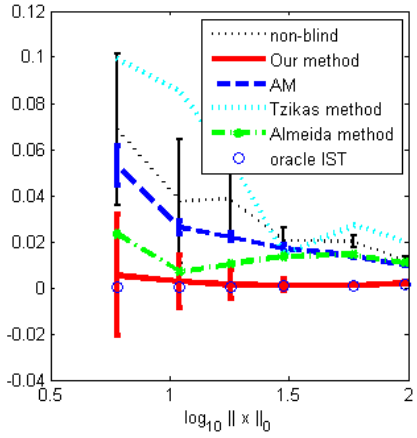


(b) Histograms of the l_0 measures. x-axis is $\|\hat{\mathbf{x}}\|_0$.

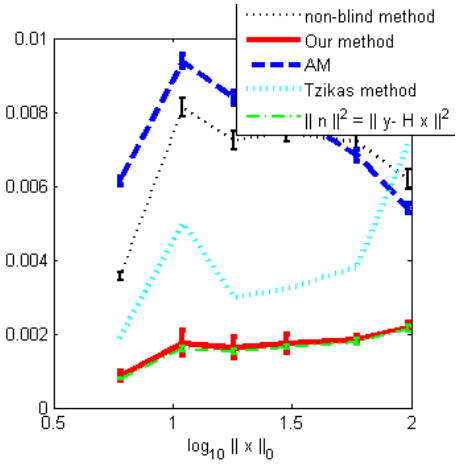
Figure 3.8: Histograms of l_2 and l_0 norm of the reconstruction error. Note that the proposed semi-blind reconstructions exhibit smaller mean error and more concentrated error distribution than the non-blind method of [2] and the alternating minimization method of [3].



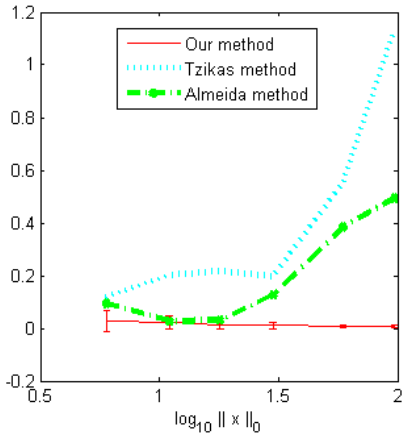
(a) $\|\hat{\mathbf{x}}\|_0/\|\mathbf{x}\|_0$: estimated sparsity. Normalized true level is 1.



(b) $\|\frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|}\|_2^2/\|\mathbf{x}\|_0$: normalized error in reconstructed image



(c) $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2/\|\mathbf{x}\|_0$: residual (projection) error



(d) $\|\frac{\hat{\boldsymbol{\kappa}}}{\|\hat{\boldsymbol{\kappa}}\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|}\|_2^2$, as a performance gauge of our myopic method. At the initial stage of the algorithm, $\|\frac{\boldsymbol{\kappa}_0}{\|\boldsymbol{\kappa}_0\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|}\|_2^2 = 0.5627$

Figure 3.9: Error bar graphs of results from our myopic deconvolution algorithm. For several image $\mathbf{x}$'s of different sparsity levels, errors are illustrated with standard deviations. (Some of the sparsity measure and residual errors are too large to be plotted together with results from other algorithms.)

sparse prior modification in the image model of other algorithms. The total variation (TV) based prior for the PSF suggested by Amizic *et al.* [40] was also implemented. The obtained PSF from this method was considerably worse than the one estimated by our proposed method (see Fig. 3.5(b)) resulting in an extremely poor quality image

deconvolution[5].

The recent blind deconvolution method proposed by Almeida *et al.* [41] utilizes the 'sharp edge' property in natural images, with initial, high regularization in order to effectively evaluate the PSF. This iterative approach has a sequentially decreasing regularization parameter to reconstruct fine details of the image. Adapted to sparse images, this method performs worse than our method, in terms of image and PSF estimation errors. The PSF and image estimates from Almeida's method are presented in Fig. 3.5(c) and 3.6(d), respectively.

Tzikas *et al.* [27] uses a similar PSF model to our method using basis kernels. However, no sparse image prior was assumed in [27] making it unsuitable for sparse reconstruction problems such as the MRFM problem considered in this work. For a fair comparison, we applied the suggested PSF model [27] along with our sparse image prior. The results of PSF and image estimation and the performance graph are shown in Fig. 3.5(d), Fig. 3.6(e), and Fig. 3.9, respectively. In terms of PSF estimation error, our algorithm outperforms the others.

We also compared against the mixture model-based algorithm of Fergus *et al.* [19], and the related method of Shan *et al.* [20], which are proposed for blind deconvolution of shaking/motion blurs and do not incorporate any sparsity penalization. When applied to the sparse MRFM reconstruction problem the algorithms of [19] and [20] performed extremely poorly (produced divergent or trivial solutions). This poor performance is likely due to the fact that the shape of the MRFM PSF and the sparse image model are significantly different from those in blind deconvolution of camera shaking/motion blurs. The generalized PSF model of [19] and [20] with the sparse image prior is Tzikas' model [27], which is described above.

We used the Iterative Shrinkage/Thresholding (IST) algorithm with a true PSF to lower bound our myopic reconstruction algorithm. The IST algorithm effectively

[5]Because this PSF is wrongly estimated and similar to the $2D$ delta function, the image estimate looks similar to the denoised version of observation, so we omit the image estimate.

reconstructs images with a sparsity constraint [71]. From Fig. 3.9(b) the performance of our result is as good as that of the oracle IST. In Table 3.2 we present comparisons of computation time[6] of the proposed sparse semi-blind Bayes reconstruction to that of several other algorithms.

Table 3.2: Computation time of algorithms (in seconds), for the data in Fig. 3.1.

Proposed method	19.06
Bayesian nonblind [2]	3.61
IST [71]	0.09
AM [3]	0.40
Almeida's method [41]	5.63
Amizic's method [40]	5.69
Tzikas' method [27]	20.31

3.4.3 Application to 3D MRFM image reconstruction

In this section, we apply the semi-blind Bayesian reconstruction algorithm to the 3D MRFM tobacco virus data [4] shown in Fig. 3.10. The necessary modification for our algorithm to apply to 3D data is simple because the extension of our 2D pixel-wise sampling method requires only one more added dimension to extend to 3D basis vectors and 3D convolution kernel. As seen in Appendix 3.6, the voxel-wise update of a vectorized image $\mathbf{x}$ can be generalized to nD data. This scalability is another benefit of our algorithm. The implementation of AM algorithm is impractical due to its slow convergence rates [3]. Here we only consider Bayesian methods. The additive noise is assumed Gaussian consistently with [45, 4], so the noise model in paragraph 3.2.1 is applied here.

The PSF basis vectors were obtained from a standard PCA and the number of principal components (PC) in the PSF perturbation was selected as 4 based on detecting the knee in a scree plot. The same interpolation method as used in [2] was

[6]Matlab is used under Windows 7 Enterprise and HP-Z200 (Quad 2.66 GHz) platform.

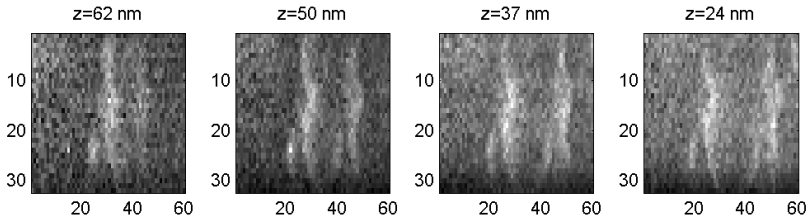
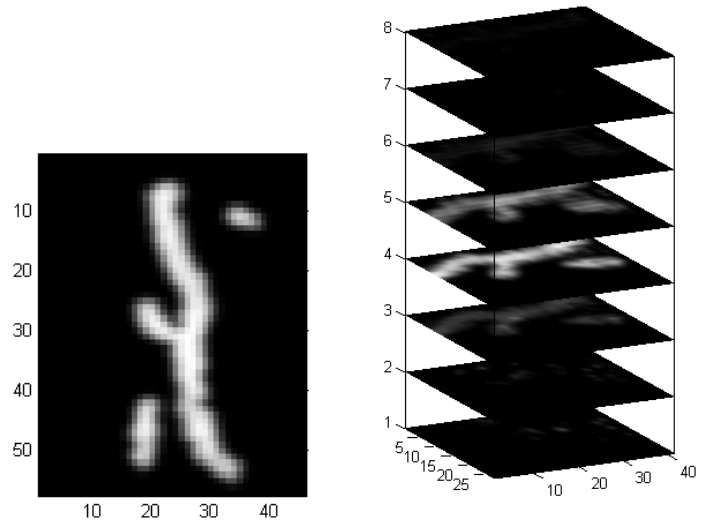


Figure 3.10: Observed data at various tip-sample distances z.

adopted to account for unequal spatial sampling rates in the supplied data for the PSF domain and the image domain.

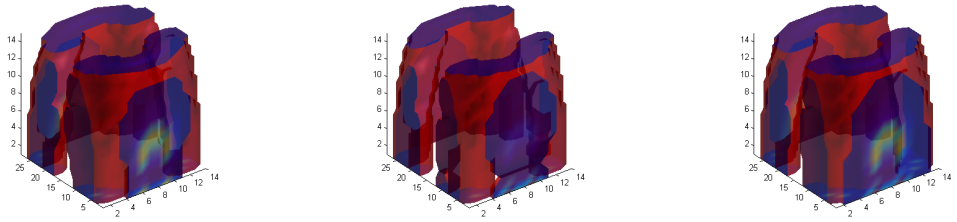
In the PSF and image domains, along z axis, the grid in PSF signal space is 3 times finer than the observation sampling density, because the PSF sampling rate along z-axis is 3 times higher than the data sampling rate is. To interpolate this lower sampled data, we implemented a version of the Bayes MC reconstruction that compensates for unequal projection sampling in z directions using the interpolation procedure [2].



(a) Ground truth synthetic virus image obtained from data in Degen *et al* [4].

(b) Semi-blind reconstruction of the synthetic virus data. Only the z-planes that have non-zero image intensity are shown.)

Figure 3.11: Results of applying the proposed semi-blind sparse image reconstruction algorithm to synthetic 3D MRFM virus image.

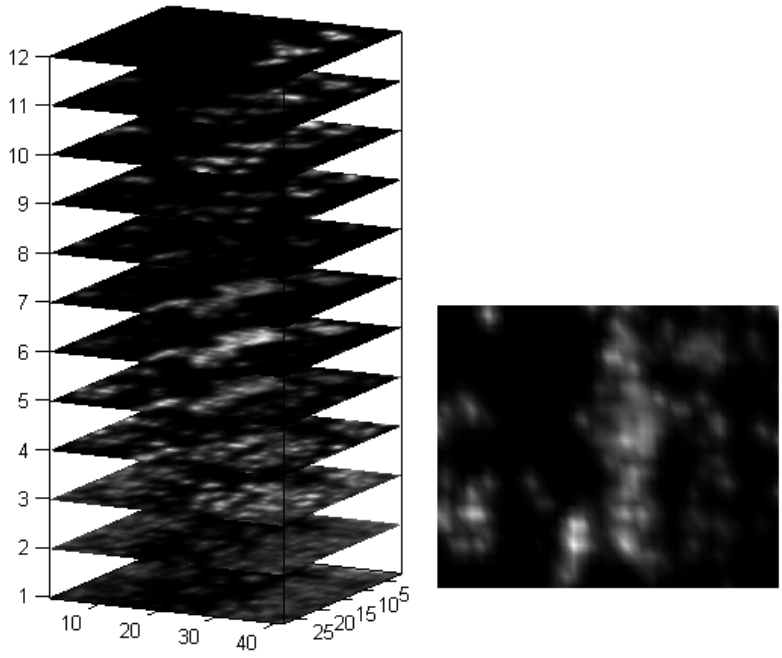


(a) True PSF. (b) Initial, mismatched PSF. (c) Estimated PSF.

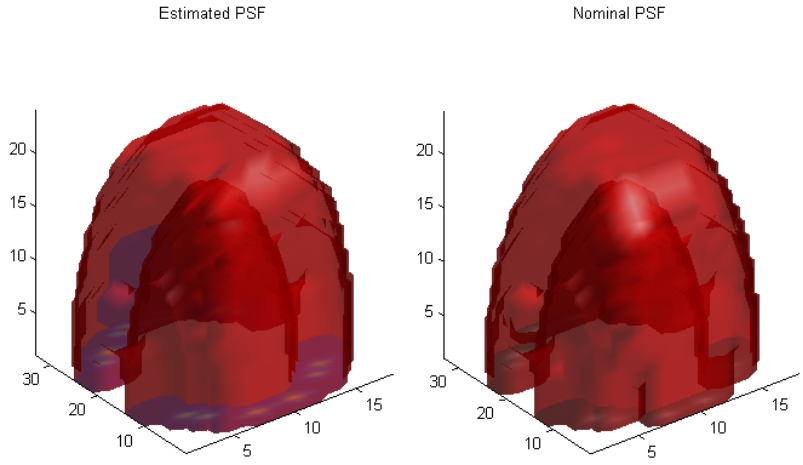
Figure 3.12: PSF estimation result.

To demonstrate that the proposed 3D MCMC semi-blind reconstruction algorithm is capable of reconstruction in the presence of significant MRFM PSF mismatch, we first applied it to a simulated version of the experimental data shown in Fig. 3.10. We used the scanning electron microscope (SEM) virus image reported in Degen *et al.* [4] to create a synthetic 3D MRFM virus image (by following the forward model in (3.1)), one slice of which is shown in Fig. 3.11(a). This 3D image was then passed through the MRFM forward model, shown in Fig. 3.12(a), and 10dB Gaussian noise was added. The mismatched PSF (Fig. 3.12(b)) was used to initialize our proposed semi-blind reconstruction algorithm. After 40 iterations the algorithm reduced the initial normalized PSF error $\| \frac{\boldsymbol{\kappa}_0}{\|\boldsymbol{\kappa}_0\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|} \|^2$ from 0.7611 to $\| \frac{\hat{\boldsymbol{\kappa}}}{\|\hat{\boldsymbol{\kappa}}\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|} \|^2 = 0.0295$. Fig. 3.11(b) and Fig. 3.12(c) show the estimated image and the estimated PSF, respectively.

We next applied the proposed semi-blind reconstruction algorithm to the actual experimental data shown in Fig. 3.10 for which there is neither ground truth on the MRFM image or on the MRFM PSF. The image reconstruction results are shown in Fig. 3.13. The small difference between the nominal PSF and the estimated PSF indicates that the estimated PSF is close to the assumed PSF. We empirically validated this small difference by verifying that multiple runs of the Gibbs sampler gave low variance PSF residual coefficients. We conclude from this finding that the PSF model of Degen *et al.* [4] is in fact nearly Bayes optimal for this experimental data. The blind image reconstruction shown in Fig. 3.13 is similar to the image reconstruction

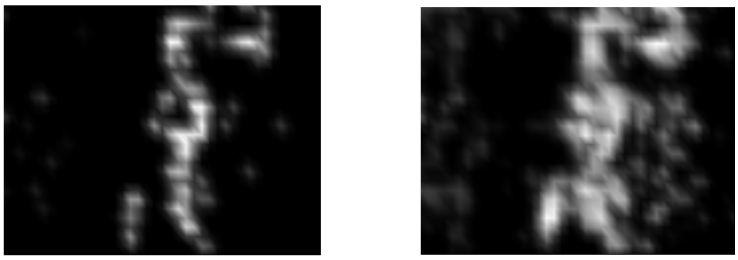


(a) MAP estimate in 3D and the estimated image on 6th plane, showing a virus particle.



(b) Estimated (left) and nominal (right) PSFs. $\| \frac{\hat{\boldsymbol{\kappa}}}{\|\hat{\boldsymbol{\kappa}}\|} - \frac{\boldsymbol{\kappa}_0}{\|\boldsymbol{\kappa}_0\|} \|^2 = 0.0212$. The difference between two is small. (Hard thresholding with level $= \max(PSF) \times 10^{-4}$ is applied for visualization.)

Figure 3.13: Semi-blind MC Bayes method results and PSF coefficient curves. $\Delta z = 4.3nm$, pixel spacing is $8.3nm \times 16.6nm$ in $x \times y$, respectively. The size of (x, y) plane is $498nm \times 531.2nm$. Smoothing is applied for visualization.



(a) MMSE solution. Gray level image intensity values range from 0 (black) to 7.34×10^{-12} (white).

(b) The pixel-wise square root of the image variance. White color indicates a high variance. Gray level image intensity values range from 0 (black) to 3.29×10^{-12} (white).

Figure 3.14: The posterior mean and variance at the 6th plane of the estimated image (Fig. 3.13(a)).

in Degen *et al.* [4] obtained from applying the Landweber reconstruction algorithm with the nominal PSF.

Using the MCMC generated posterior distribution obtained from the experimental MRFM data, we generated confidence intervals, posterior mean and posterior variance of the pixel intensities of the unknown virus image. The posterior mean and variance are presented in Fig. 3.14 for selected pixels. In addition, to demonstrate the match between the estimated region occupied by the virus particle and the actual region we evaluated Bayesian p-values for object regions. The Bayesian p-value for a specific region R_i having non-zero intensity is $pv(R_i) = \mathrm{P}(\{I_k = 1\}_{k \in R_i} | \mathbf{y})$ where P is a probability measure and I_k is an indicator function at the kth voxel. Assuming voxel-wise independence, the p-values are easily computed from the posterior distribution and provide a level of a posteriori confidence in the statistical significance of the reconstruction. We found that over 95% of the Bayesian p-values were greater than 0.5 for the non-zero regions of the reconstruction.

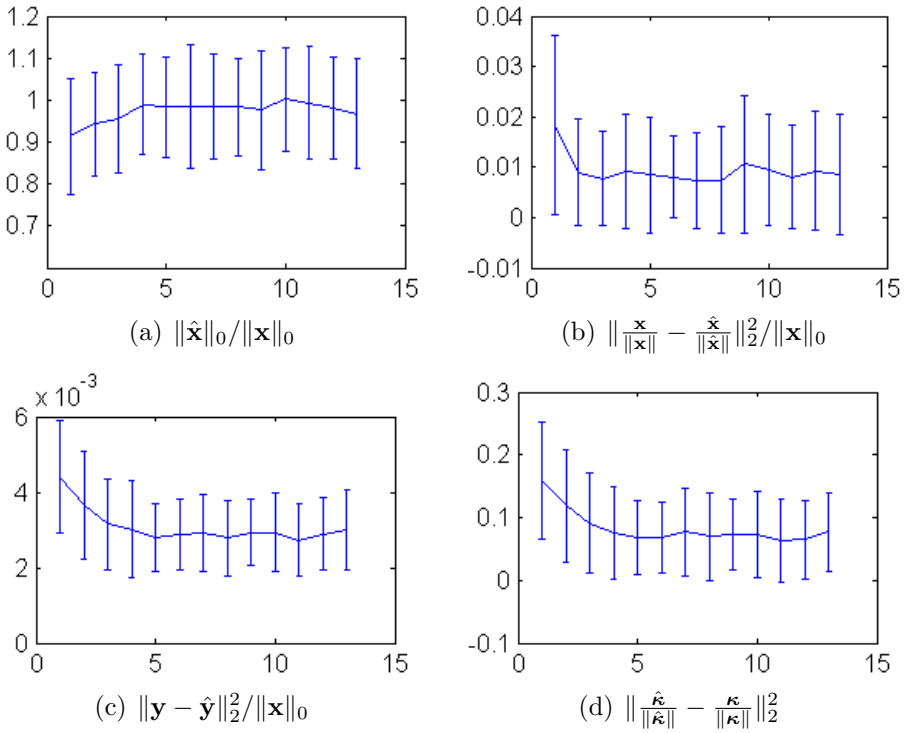
3.4.4 Discussion

Joint identifiability is a common issue underlying all blind deconvolution methods. (e.g., scale ambiguity.) Even though the unicity of our solution is not proven, given the conditions that 1) $\text{span}(\boldsymbol{\kappa}) = \boldsymbol{\kappa}_0 + \text{span}(\sum \boldsymbol{\kappa}_i)$ does not cover a kernel of a delta function, $\boldsymbol{\kappa} = \delta(\cdot)$, and 2) the PSF solution is restricted to this linear space of $\boldsymbol{\kappa}_0, \boldsymbol{\kappa}_i$'s, the equation (3.17) for the MAP criteria promises a reasonable solution that is not trivial such as $\hat{\mathbf{x}} = \mathbf{y}$. Due to this restriction and the sparse nature of the image to be estimated, we can reasonably expect that the solution provided by the algorithm is close to the true PSF. A study of unicity of the solution would be worthwhile but is beyond the scope of the current work as it would require study of the complicated and implicit fixed points of the proposed Bayes objective function.

Note that proposed sparse image reconstruction algorithm can be extended to exploit sparsity in other domains, such as the wavelet domain. In this case, if we define $\mathbf{W}$ to be the transformation matrix on $\mathbf{x}$, the proposed semi-blind approach can be applied to reconstruct the transformed signal $\mathbf{x}' = \mathbf{W}\mathbf{x}$. However, instead of assigning the single sided exponential distribution as prior for $\mathbf{x}'$, a double sided Laplacian distribution might be used to cover the negative values of the pixels. The estimation procedure for PSF coefficients, noise level, and hyperparameters would be the same. For image estimation, the vector $\mathbf{h}_i$ used in (3.22) would be replaced with the ith column of $\mathbf{H}\mathbf{W}^{-1}$.

To investigate the sensitivity to the number of principal PSF components K, we evaluate the corresponding performances in Fig. 3.15. The number of principal PSF components $K = 4$ used in our experiments was determined from the scree plot of relative eigenvalues. Our choice $K = 4$ is validated by the fact that for $K \geq 4$ the performance saturates (see Fig. 3.15). Generally, the selection of the number of the necessary components can be formulated as a model selection problem. Within our Bayesian MCMC framework, we could adopt a reversible jump Markov

chain Monte Carl (RJMCMC) approach [72] to estimate K. In this approach, a reversible Markov chain sampler chooses or jumps between parameter subspaces of different dimensionality. This method produces the probability for each model, which is an empirical distribution of accepted K values along the Markov chain. This is more flexible than our current algorithm but a drawback of this approach is that the computational requirements are high: successive reconstructions need to be performed for the different values of K generated by the Markov chain.



(a) $\|\hat{\mathbf{x}}\|_0/\|\mathbf{x}\|_0$ (b) $\|\frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|}\|_2^2/\|\mathbf{x}\|_0$

(c) $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2/\|\mathbf{x}\|_0$ (d) $\|\frac{\hat{\boldsymbol{\kappa}}}{\|\hat{\boldsymbol{\kappa}}\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|}\|_2^2$

Figure 3.15: Mean-performance graphs for the different number of principal PSF components (K in x axis) from 100 independent trials for each K value and $\|\mathbf{x}\|_0 = 11$. Vertical bars are standard deviations. (d) Initial condition: $\|\frac{\boldsymbol{\kappa}_0}{\|\boldsymbol{\kappa}_0\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|}\|_2^2 = 0.5627$.

3.5 Conclusion and Future Work

3.5.1 Conclusion

We have proposed an extension of the method of the non-blind Bayes reconstruction in Dobigeon *et al.* [2] that simultaneously estimates a partially known PSF and the unknown but sparse image. The method uses a prior model on the PSF that reflects a nominal PSF and uncertainty about the nominal PSF. In our algorithm the values of the parameters of the convolution kernel were estimated by a Metropolis-within-Gibbs algorithm, with an adaptive mechanism for tuning random-walk step size for fast convergence. Our approach can be used to empirically evaluate the accuracy of assumed nominal PSF models in the presence of model uncertainty. In our sparse reconstruction simulations, we demonstrated that the semi-blind Bayesian algorithm has improved performance as compared to the AM reconstruction and other blind deconvolution algorithms and non-blind Bayes method under several criteria.

3.5.2 Future work: image model using Markov random fields

We have developed semi-blind sparse deconvolution methods for MRFM data under the sparsity assumption at a high resolution. However, if the resolution is low, this sparsity assumption needs to be replaced with a smoothness assumption on image objects. In this section, we formulate general image models that produce continuous, smooth image objects using Markov random fields, as opposed to our independent pixel model. We reserve the development and application to real data as our future work.

The assumption of smoothness in image objects implies a high spacial correlation between the adjacent pixels. However, the image priors we have suggested so far assume independent pixel distributions, thus leading to effective sparse estimation. To extend this model to capture the spacial correlation with a similar Bayesian per-

spective, we first propose prior distributions based on Markov random field (MRF) models. These models exploit correlation information in local regions with defined neighborhood relationships for pixels or sites of an image. The MRF model fits well with our site-wise update algorithms, even though neither the posterior nor prior distribution of an entire image is in the closed form. The difficulty in obtaining closed form solutions for the whole image is no obstacle in our estimation process because we use a pixel-wise sampling technique, for which a closed form posterior distribution exists in the form of a Bernoulli truncated exponential distribution family.

We first introduce notations in our Markov Random Field (MRF) image model and state the relevant definitions and theories in MRF.

3.5.2.1 Markov random fields

We hereby consider the image variable $\mathbf{x}$ as a MRF, following the definitions below.

Definition: Random Field

Let S be a finite set, with elements denoted by s and called sites, and let Λ be a finite set called the phase space. A random field on S with phases in Λ is a collection $X = \{X(s)\}_{s \in S}$ of random variables $X(s)$ with values in Λ.

Note that $\mathcal{N}_s$, a subset of S, is the neighborhood of site s, where the neighbors are mutual. Roughly speaking, $\mathcal{C}$ is a clique if and only if any two distinct sites of it are mutual neighbors.

Definition: Markov Random Field

The random field X is called a Markov random field (MRF) with respect to the neighborhood system $\mathbb{N} = \{\mathcal{N}_s\}$, if for all sites $s \in S$ the random variables $X(s)$ and $X(S - (\{s\} \cup \mathcal{N}_s))$ are independent given $X(\mathcal{N}_s)$.

Definition: Gibbs Potential

A Gibbs Potential on Λ^S relative to the neighborhood system $\mathbb{N}$ is a collection $\{V_{\mathcal{C}}\}_{\mathcal{C} \subset S}$

of functions $V_{\mathcal{C}} : \Lambda^S \to \mathbf{R} \cup \{\infty\}$ such that

(i) $V_{\mathcal{C}} \equiv 0$ is not a clique

(ii) for all $x, x' \in \Lambda^S$ and all $\mathcal{C} \subset S$, $(x(\mathcal{C}) = x'(\mathcal{C})) \Rightarrow (V_{\mathcal{C}}(x) = V_{\mathcal{C}}(x'))$.

Definition: Gibbs distribution

The energy function $\mathcal{E} : \Lambda^S \to \mathbf{R} \cup \{\infty\}$ is said to derive from the potential $\{V_{\mathcal{C}}\}_{\mathcal{C} \subset S}$ if $\mathcal{E} = \sum_{\mathcal{C}} V_{\mathcal{C}}(x)$.

With this energy function, we call a probability distribution

$$\pi_T(x) = \frac{1}{Z_T} \exp(-\frac{1}{T}\mathcal{E}(x)) \tag{3.30}$$

a Gibbs distribution.

Hammersley-Clifford Theorem

X is a MRF and $\forall x, \mathrm{P}(X = x) > 0$ if and only if $\mathrm{P}(X = x)$ has the form of a Gibbs distribution.

Finally, we use the fact that the simulated distributions converge to stationary distributions. In other words, in simulation of random fields with an appropriate stochastic sampler, the state gets to the most stable state in terms of the energy $\mathcal{E}$ [68].

3.5.2.2 An adaptive image model

The image prior can be categorized into two classes in terms of spatial intra-dependency of pixels: 1) a spatially homogeneous prior and 2) a spatially inhomogeneous prior. By decomposing image density into the parts of the background and foreground, the following mixture forms of density are obtained.

1) An independent prior: $f(\mathbf{x}) = \prod_i \delta(x_i)(1 - w) + \frac{1}{a} \exp(-\frac{x_i}{a})w$. This image prior distribution is used in Chapter III.

2) A spatially dependent prior:

$$f(x_i|\mathbf{x}_{-i}, \Phi) = (1 - w_i)f(x_i|x_i = 0) + w_i f(x_i|x_i \neq 0, \mathbf{x}_{-i}, \Phi) \tag{3.31}$$

$$= (1 - w_i)\delta(x) + w_i g(x_i|\mathbf{x}_{-i}, \Phi), \tag{3.32}$$

where x_i is ith component of $\mathbf{x}$, $w_i = \mathrm{P}(x_i \neq 0|\mathbf{x}_{-i})$ is pixel occupancy probability, $g(x_i|\mathbf{x}_{-i}, \Phi)$ is conditional density given hyperparameters Φ, $\mathbf{x}_{-i} = \{x_j\}_{j\in\mathcal{N}_i}$, and $\mathcal{N}_i$ is neighborhood indices of x_i. Compared to the spatially independent prior, this prior does not have a closed form for the whole sites or $\mathbf{x}$.

In this section, we investigate the spatially dependent prior distribution. This prior distribution can be configured from two aspects. Firstly, the sparsity rate w_i can be defined to be dependent on neighborhood. This spatially adaptive w_i would help induce sparsity among zero-dominant neighbors and smoothness among signal/nonzero-dominant neighbors. Secondly, the conditional density, the signal prior distribution $g(\cdot)$, can be set up as a function of neighbors. For example, $g(\cdot)$ can be a normal distribution with the mean from the mean of neighbors and the variance from the variance of sampled neighbor values.

A prior with a spatially invariant conditional density

In (3.32), we can set a non-sparsity rate $w_i = h(\frac{|\{x_j : j\in\mathcal{N}_i, x_j\neq 0\}|}{|\mathcal{N}_i|})$ and $g(x_i) = g_a(x_i)$, where $h(\cdot)$ is an increasing function mapping from $[0, 1]$ to $[0, 1]$, and $g_a(x_i)$ is an exponential distribution with parameter $1/a$. We expect the h function to be almost linear for the non-sparsity rate to be approximately proportional to the non-sparsity ratio of neighborhood. (In practice, a small margin (ϵ) from 0 and 1 seems necessary to avoid stability of estimation. Thus, the range of $h(\cdot)$ is $[\epsilon, 1 - \epsilon]$.)

The pixel-wise posterior image density is

$$f(x_i|y, x_{-i}, \sigma^2, a) = (1 - w_i')\delta(x_i) + w_i'\phi_+(x_i|\mu_i, \eta_i^2), \tag{3.33}$$

where $\phi_+(\cdot|\mu_i, \eta_i^2)$ is truncated Gaussian distribution, $w_i' = w(x_i, x_{\mathcal{N}_i}) = \frac{u_i w_i}{u_i w_i + (1-w_i)} w_i$, and u_i is similarly derived as in Section 3.3.1.3. In short, x_i has a posterior distribution which has a spatial-variant rate w_i', different from the rate w_i in the prior distribution, which is, again, the conditionally Bernoulli-truncated Gaussian distribution.

A prior with a spatially variant conditional density $g(x_i|\mathbf{x}_{-i}, \Phi)$

The more general image prior density than the one proposed in (3.33) has a spatially variant conditional density $g(\cdot)$. We define $g(x_i|\mathbf{x}_{-i}, \Phi)$ to be normal distribution with mean and variance (m_i, σ_i^2), where m_i is the mean of nonzero neighbors, $\sigma_i^2 = c \min(|\mathcal{N}_i|/|\mathcal{N}_i \cap I_1|, |\mathcal{N}_i|)$, I_1 is a set of sites of (estimated) foreground pixels (I_0 is similarly defined for background pixels), and c is a variable to be estimated. The result posterior distribution has the same functional form as the one in (3.33) with different coefficient values, which are similarly derived. The scaling factor for the variance in the prior density c is assumed to follow the inverse gamma distribution with parameter values $(1, l_c/2)$, where the hyperparameter $f(l_c) \propto \frac{1}{l_c}$. Then, the posterior for c becomes $\mathcal{IG}(c; \frac{|I_1|}{2}, \sum_{x_s \neq 0} \frac{(x_s - m_s)^2}{2\sigma_i^2})$

3.5.2.3 MAP criteria

We present the MAP criteria for the prior of a spatially variant conditional density. Because the closed form criteria, for the whole sites $\mathbf{x}$, to determine MAP is difficult to obtain due to the intercorrelation of pixels, we suggest an approximation of $f(\mathbf{x}|\mathbf{y})$. We can define $Z(c)$ to be a normalization constant such that $f(\mathbf{x}|c) = \frac{1}{Z(c)} \prod_i f(x_i|c, x_{N_i})$. This partition function $Z(c)$ is evaluated by integrating $\prod_i f(x_i|c, x_{N_i})$ over $\mathbf{x} = [x_1, ..., x_P]^T$, which is almost infeasible. Thus, we approximate $f(\mathbf{x}|c)$ by assuming that the partition function is independent of c. Also, after reaching the steady state of the Markov Chain, we assume $\prod_{i \in I_0}(1 - w_i) \prod_{i \in I_1} w_i$ and $\prod_{i \in I_1}(k_i)^{-0.5}$ can be treated as constants, thus independent of c and $\mathbf{x}$. Under these assumptions, $f(\mathbf{x}|c) \tilde{\propto} \prod_i f(x_i|c, x_{N_i}) \propto \prod_{i \in I_1} k_i c^{-\frac{n_1}{2}} \exp(-\sum_{i \in I_1} \frac{(x_i - m_i)^2}{2k_i c})$, where $\tilde{\propto}$

means 'approximately proportional to'.

The result MAP criteria is then approximated as

$$f(\mathbf{x}|\mathbf{y})\tilde{\propto}||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^N \frac{\Gamma(\alpha_1)}{\beta_1^{\alpha_1}} \prod_{i\in I_1} \frac{1}{k_i^{0.5}}, \tag{3.34}$$

with $\alpha_1 = \frac{n_1}{2}, \beta_1 = -\sum_{i\in I_1} \frac{x_i - m_i}{2k_i}$, and $\Gamma(\cdot)$ a gamma function.

3.6 Appendix 1: Fast Recursive Sampling Strategy

In our proposed iterative Bayesian algorithm, repeated evaluations of the transformation $T(\boldsymbol{\kappa}(\boldsymbol{\lambda}), \mathbf{x})$ can be computationally expensive. For example, at each iteration of the proposed Bayesian myopic deconvolution algorithm, one must generate x_i from its conditional distribution $f(x_i|w, a, \lambda, \sigma^2, \mathbf{x}_{-i}, \mathbf{y})$, which requires the calculation of $T(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i)$ where $\tilde{\mathbf{x}}_i$ is the vector $\mathbf{x}$ whose ith element has been replaced by 0. Moreover, sampling according to the conditional posterior distributions of σ^2 and λ_k (3.26) and (3.23) requires computations of $T(\boldsymbol{\kappa}, \mathbf{x})$.

By exploiting the bilinearity of the transformation $T(\cdot, \cdot)$, we can reduce the complexity of the algorithm. We describe below a strategy, similar to those presented in [2, App. B], which only requires computation of $T(\cdot, \cdot)$ at most $M \times (K+1)$ times. First, let $\mathbf{I}_M$ denote the $M \times M$ identity matrix and $\mathbf{u}_i$ its ith column. In a first step of the analysis, the M vectors $\mathbf{h}_i^{(0)}$ $(i = 1, \ldots, M)$

$$\mathbf{h}_i^{(0)} = T(\boldsymbol{\kappa}_0, \mathbf{u}_i), \tag{3.35}$$

and KM vectors $\mathbf{h}_i^{(k)}$ $(i = 1, \ldots, M,\ k = 1, \ldots, K)$

$$\mathbf{h}_i^{(k)} = T(\boldsymbol{\kappa}_k, \mathbf{u}_i), \tag{3.36}$$

are computed. Then one can compute the quantity $T(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i)$ and $T(\boldsymbol{\kappa}, \mathbf{x})$ at any stage

of the Gibbs sampler without evaluating $T(\cdot,\cdot)$, based on the following decomposition

$$T(\boldsymbol{\kappa},\mathbf{x}) = \sum_{i=1}^{M} x_i \mathbf{h}_i^{(0)} + \sum_{k=1}^{K} \lambda_k \sum_{i=1}^{M} x_i \mathbf{h}_i^{(k)}. \tag{3.37}$$

The resulting procedure to update the ith coordinate of the vector $\mathbf{x}$ is described in Algorithm 4 below.

Algorithm 4 Efficient simulation according to $f(\mathbf{x}\,|w,a,\sigma^2,\mathbf{y})$

At iteration t of the Gibbs sampler, for $i = 1,\ldots,M$, update the ith coordinate of the vector

$$\mathbf{x}^{(t,i-1)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},x_i^{(t-1)},x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T$$

via the following steps:

1: compute $\mathbf{h}_i = \mathbf{h}_i^{(0)} + \sum_{k=1}^{K} \lambda_k \mathbf{h}_i^{(k)}$,
2: set $T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right) = T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i-1)}\right) - x_i^{(t-1)}\mathbf{h}_i$,
3: set $\mathbf{e}_i = \mathbf{x} - T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right)$,
4: compute μ_i, η_i^2 and w_i as defined in [6],
5: draw $x_i^{(t)}$ according to (3.21),
6: set $\mathbf{x}^{(t,i)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},x_i^{(t)},x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T$,
7: set $T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i)}\right) = T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right) + x_i^{(t)}\mathbf{h}_i$.

Note that in step 7. of the algorithm above, $T(\boldsymbol{\kappa},\mathbf{x})$ is recursively computed. Once all the coordinates have been updated, the current $T(\boldsymbol{\kappa},\mathbf{x})$ can be directly used to sample according to the posterior distribution of the noise variance in (3.26). Moreover, this quantity can be used to sample according to the conditional posterior distribution of λ_k in (3.23). More precisely, evaluating $T\left(\boldsymbol{\kappa}\left(\lambda_k^{\star}\right),\mathbf{x}\right)$ in the acceptance probability (3.25) can be recursively evaluated as follows

$$T\left(\boldsymbol{\kappa}\left(\lambda_k^{\star}\right),\mathbf{x}\right) = T\left(\boldsymbol{\kappa}\left(\lambda_k^{(t)}\right),\mathbf{x}\right) - \left(\lambda_k^{(t)} - \lambda_k^{\star}\right)\sum_{i=1}^{M} x_i \mathbf{h}_i^{(k)}. \tag{3.38}$$

3.7 Appendix 2: Sparsity Enforcing Selective Sampling

Since we have estimated the 'overall sparsity', $1 - \hat{w}$, of $\mathbf{x}$ from (3.19), we can expedite the sampling procedure of $\mathbf{x}$ by selectively sampling only significant portions of entire pixels of $\mathbf{x}$. As a result, we expect $(1 - \hat{w}) \times 100\%$ of pixel domain of $\mathbf{x}$ to be zero, which will not need to be re-sampled.

At time t, in order to approximate the quantile $(1 - \hat{w})$ of $\{w_i^{(t)}\}_{i=1,\ldots,M}$ in (3.21) we evaluate the $(1 - \hat{w})$ quantile value of the previously obtained $\{w_i^{(t-1)}\}_{i=1,\ldots,M}$. This approximation accelerates the computation because the exact calculation of $\{w_i^{(t)}\}_{i=1,\ldots,M}$ requires sampling of all x_i's. Let $q = quantile(\{w_i^{(t-1)}\}_{i=1,\ldots,M}, 1 - \hat{w})$ and $w_{thr} = \max(q, 1 - \hat{w})$. When $w_i^{(t)}$ for $x_i^{(t)}$ from (3.21) is less than w_{thr}, then $x_i^{(t)}$ is not updated or is set to zero. Because MCMC sampling is computationally expensive, especially for large size images, this suggestion can be restricted to the burn-in period to save computations.

In our experiment, the selective sampling of $\mathbf{x}$ applied after 3 or 4th iterations produce equally good results compared to the conventional MCMC sampling methods, while reducing computation time by $30-50\%$ for non-blind sparse reconstruction with a fixed PSF and by $10 - 30\%$ for the semi-blind sparse reconstruction.

CHAPTER IV

Variational Bayesian Semi-blind Deconvolution[1]

In this chapter, we present a variational Bayesian method of joint image reconstruction and PSF estimation when the PSF of the imaging device is only partially known. The variational principle allows minimization of the distributional distance between the target posterior distribution and the approximate distribution. This minimization produces feasible solutions when appropriate factorization of approximated distributions is assumed. However, the mixture-model type distributions that are adopted for the image prior distributions require label variables to fit into the variational Bayes framework. Therefore, we introduce this latent variable for image labeling into background and foreground groups. To exploit the sparsity of images as performed in Chapter III, the similar sparsity inducing prior distribution is used for the individual pixel values.

Simulation results clearly demonstrate that the semi-blind deconvolution algorithm compares favorably with previous Markov chain Monte Carlo (MCMC) version of myopic sparse reconstruction. It significantly outperforms mismatched non-blind algorithms that rely on the assumption of the perfect knowledge of the PSF. This algorithm is faster, more scalable for equivalent image reconstruction qualities in Chpater III. The algorithm is illustrated on real data from magnetic resonance force

[1]This chapter is partially based on the papers [73, 74].

microscopy (MRFM). Lastly, as an extension to the space-invariance assumption of the PSF, we consider possible solutions and reserve for a future work.

4.1 Formulation

The formulation in this section is the almost same as the one introduced in Chapter III except the latent variable $\mathbf{z}$. For completeness, the formulation is briefly reproduced.

4.1.1 Image model

As in Chapter III, the image model is defined as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = T(\boldsymbol{\kappa}, \mathbf{x}) + \mathbf{n}, \tag{4.1}$$

where $\mathbf{y}$ is a $P \times 1$ vectorized measurement, $\mathbf{x} = [x_1, \ldots, x_M]^T \succeq 0$ is an $M \times 1$ vectorized sparse image to be recovered, $T(\boldsymbol{\kappa}, \cdot)$ is a convolution operator with the PSF $\boldsymbol{\kappa}$, $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_M]$ is an equivalent system matrix, and $\mathbf{n}$ is the measurement noise vector assumed to be Gaussian, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_P)$. Again, the PSF $\boldsymbol{\kappa}$ is assumed to be unknown but a nominal PSF estimate $\boldsymbol{\kappa}_0$ is available.

4.1.2 PSF basis expansion

The nominal PSF $\boldsymbol{\kappa}_0$ is assumed to be generated with known parameters (gathered in the vector $\boldsymbol{\zeta_0}$) tuned during imaging experiments. However, due to model mismatch and experimental errors, the true PSF $\boldsymbol{\kappa}$ may deviate from the nominal PSF $\boldsymbol{\kappa}_0$. If the generation model for PSFs is complex, direct estimation of a parameter deviation, $\Delta\boldsymbol{\zeta} = \boldsymbol{\zeta_{true}} - \boldsymbol{\zeta_0}$, is difficult.

As proposed in Chapter III, we model the PSF $\boldsymbol{\kappa}$ (resp. $\{\mathbf{H}\}$) as a perturbation about a nominal PSF $\boldsymbol{\kappa}_0$ (resp. $\{\mathbf{H}^0\}$) with K basis vectors $\boldsymbol{\kappa}_k$, $k = 1, \ldots, K$, that

span a subspace representing possible perturbations $\Delta\boldsymbol{\kappa}$. We empirically determined this basis using the variational eigendecomposition approach. If there is no PSF generation model, then we can decompose the support region of the true (suspected) PSF to produce an orthonormal basis. The necessary number of the bases is again chosen to explain most support areas that have major portion/energy of the desired PSF. This approach is presented in our experiment with Gaussian PSFs.

We use a basis expansion to present $\boldsymbol{\kappa}(\boldsymbol{\lambda})$ as the following linear approximation to $\boldsymbol{\kappa}$,

$$\boldsymbol{\kappa}(\boldsymbol{\lambda}) = \boldsymbol{\kappa}_0 + \sum_{i=1}^{K} \lambda_i \boldsymbol{\kappa}_i. \tag{4.2}$$

The objective is to estimate the unknown image, $\mathbf{x}$, and the linear expansion coefficients $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]^T$.

4.1.3 Determination of priors

The priors on the PSF, the image, and the noise are constructed as latent variables in a hierarchical Bayesian model. We use a notation $p(\cdot)$ as a probability measure function, instead of $f(\cdot)$ used in Chapter III, to differentiate our use of $p(\cdot)$ and $q(\cdot)$, the approximation of $p(\cdot)$.

4.1.3.1 Likelihood function

Under the hypothesis that the noise in (4.1) is white Gaussian, the likelihood function takes the form

$$p\left(\mathbf{y}|\mathbf{x}, \boldsymbol{\lambda}, \sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{P}{2}} \exp\left(-\frac{\left\|\mathbf{y} - T\left(\boldsymbol{\kappa}\left(\boldsymbol{\lambda}\right), \mathbf{x}\right)\right\|^2}{2\sigma^2}\right), \tag{4.3}$$

4.1.3.2 Image and label priors

We start from the prior that is a convex combination of an atom at zero and an exponential distribution:

$$p(x_i|a,w) = (1-w)\delta(x_i) + wg(x_i|a). \tag{4.4}$$

In (4.4), $w = \mathrm{P}\left(x_i \neq 0\right)$ is the prior probability of a non-zero pixel and $g(x_i|a) = \frac{1}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}(x_i)$ is a single-sided exponential distribution.

A distinctive property of the image prior (4.4) is that it can be expressed as a latent variable model

$$p(x_i|a,z_i) = (1-z_i)\delta(x_i) + z_i g(x_i|a), \tag{4.5}$$

where the binary variables $\{z_i\}_1^N$ are independent and identically distributed and indicate if the pixel x_i is active

$$z_i = \begin{cases} 1, & \text{if } x_i \neq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

and have the Bernoulli probabilities: $z_i \sim Ber(w)$.

The prior distribution of pixel value x_i in (4.4) can be rewritten conditionally upon latent variable z_i

$$p\left(x_i|z_i=0\right) = \delta\left(x_i\right),$$
$$p\left(x_i|a,z_i=1\right) = g\left(x_i|a\right),$$

which can be summarized in the following factorized form

$$p(x_i|a,z_i) = \delta(x_i)^{1-z_i} g(x_i|a)^{z_i}. \tag{4.7}$$

By assuming each component x_i to be conditionally independent given z_i and a, the following conditional prior distribution is obtained for $\mathbf{x}$:

$$p(\mathbf{x}|a,\mathbf{z}) = \prod_{i=1}^{N} \left[\delta(x_i)^{1-z_i} g(x_i|a)^{z_i}\right] \tag{4.8}$$

where $\mathbf{z} = [z_1, \ldots, z_M]$.

This factorized form will turn out to be crucial for simplifying the variational Bayes reconstruction algorithm in Section 4.2.

4.1.3.3 PSF parameter prior

We assume that the PSF parameters $\lambda_1, \ldots, \lambda_K$ are independent and each is uniformly distributed over intervals

$$\mathcal{S}_k = [-\Delta\lambda_k, \Delta\lambda_k]\,. \tag{4.9}$$

These intervals are specified a priori and are associated with error tolerances of the imaging instrument. The joint prior distribution of $\boldsymbol{\lambda}$ is therefore:

$$p(\boldsymbol{\lambda}) = \prod_{k=1}^{K} \frac{1}{2\Delta\lambda_k} \mathbf{1}_{\mathcal{S}_k}(\lambda_k)\,. \tag{4.10}$$

4.1.3.4 Noise variance prior

A conjugate inverse-Gamma distribution with parameters ς_0 and ς_1 is assumed as the prior distribution for the noise variance (See Appendix 4.6 for the details of this distribution):

$$\sigma^2|\varsigma_0,\varsigma_1 \sim \mathcal{IG}(\varsigma_0,\varsigma_1)\,. \tag{4.11}$$

The parameters ς_0 and ς_1 will be fixed to a number small enough to obtain a vague hyperprior, unless we have good prior knowledge.

4.1.4 Hyperparameter priors

The values of the hyperparameters $\{a, w\}$ greatly impact the quality of the deconvolution, leading to a second level of hierarchy in the Bayesian paradigm. The definitions of distributions for hyperpriors are presented below.

4.1.4.1 Hyperparameter a

A conjugate inverse-Gamma distribution is assumed for the Laplacian scale parameter a:

$$a|\boldsymbol{\alpha} \sim \mathcal{IG}\left(\alpha_0, \alpha_1\right), \tag{4.12}$$

with $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^T$. The parameters α_0 and α_1 will be fixed to a number small enough to obtain a vague hyperprior, unless we have good prior knowledge.

4.1.4.2 Hyperparameter w

We assume a Beta random variable with parameters (β_0, β_1), which are iteratively updated in accordance with data fidelity. The parameter values will reflect the degree of prior knowledge and we set $\beta_0 = \beta_1 = 1$ to obtain a non-informative prior. (See Appendix 4.7 for the details of this distribution)

$$w \sim \mathcal{B}(\beta_0, \beta_1). \tag{4.13}$$

4.1.5 Posterior distribution

The conditional relationships between variables is illustrated in Fig. 4.1. The resulting posterior of hidden variables given the observation is

$$p(\mathbf{x}, a, \mathbf{z}, w, \boldsymbol{\lambda}, \sigma^2|\mathbf{y}) \;\propto\; p(\mathbf{y}|\mathbf{x}, \sigma^2, \boldsymbol{\lambda}) \times p(\mathbf{x}|a, \mathbf{z})p(\mathbf{z}|w)p(w)p(a)p(\boldsymbol{\lambda})p(\sigma^2). \tag{4.14}$$

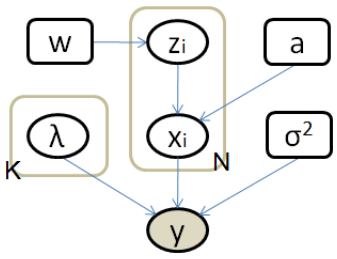


Figure 4.1: Conditional relationships between variables. A node at an arrow tail conditions the node at the arrow head.

Since it is too complex to derive exact Bayesian estimators from this posterior, a variational approximation of this distribution is proposed in the next section.

4.2 Variational Approximation

4.2.1 Basics of variational inference

In this section, we show how to approximate the posterior densities within a variational Bayes framework. Denote by $\mathbf{U}$ the set of all hidden parameter variables including the image variable $\mathbf{x}$ in the model, denoted by $\mathcal{M}$. The hierarchical model implies the Markov representation $p(\mathbf{y}, \mathbf{U}|\mathcal{M}) = p(\mathbf{y}|\mathbf{U}, \mathcal{M})p(\mathbf{U}|\mathcal{M})$. Our objective is to compute the posterior $p(\mathbf{x}|\mathbf{y}, \mathcal{M}) = \int p(\mathbf{y}|\mathbf{U}, \mathcal{M})p(\mathbf{U}|\mathcal{M})d\mathbf{U}_{\setminus\mathbf{x}}/p(\mathbf{y}|\mathcal{M})$, where $\mathbf{U}_{\setminus\mathbf{x}}$ is a set of variables in $\mathbf{U}$ except $\mathbf{x}$. Let q be any arbitrary distribution of $\mathbf{U}$. Then

$$\ln p(\mathbf{y}|\mathcal{M}) = \mathcal{L}(q) + \mathrm{KL}(q\|p) \tag{4.15}$$

with

$$\mathcal{L}(q) = \int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{y}, \mathbf{U}|\mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U} \tag{4.16}$$

$$\mathrm{KL}(q\|p) = -\int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{U}|\mathbf{y}, \mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U}. \tag{4.17}$$

We observe that maximizing the lower bound $\mathcal{L}(q)$ is equivalent to minimizing the

Kullback-Leibler (KL) divergence $\mathrm{KL}(q\|p)$. Consequently, instead of directly evaluating $p(\mathbf{y}|\mathcal{M})$ given $\mathcal{M}$, we will specify a distribution $q(\mathbf{U}|\mathcal{M})$ that approximates the posterior $p(\mathbf{U}|\mathbf{y},\mathcal{M})$. The best approximation maximizes $\mathcal{L}(q)$. We present Algorithm 5 that iteratively increases the value of $\mathcal{L}(q)$ by updating posterior surrogate densities. To obtain a tractable approximating distribution q, we will assume a factorized form as $q(\mathbf{U}) = \prod_j q(\mathbf{U}_j)$ where $\mathbf{U}$ has been partitioned into disjoint groups $\mathbf{U}_j$. Subject to this factorization constraint, the optimal distribution $q^*(\mathbf{U}) = \prod_j q^*(\mathbf{U}_j)$ is given by

$$\ln q_j^*(\mathbf{U}_j) = \mathrm{E}_{\setminus \mathbf{U}_j}\left[\ln p(\mathbf{U},\mathbf{y})\right] + (\text{const}), \quad \forall j \tag{4.18}$$

where $\mathrm{E}_{\setminus \mathbf{U}_j}$ denotes the expectation[2] with respect to all factors $\mathbf{U}_i$ except $i = j$. We will call $q^*(\mathbf{U})$ the posterior surrogate for p.

4.2.2 Suggested factorization

Based on our assumptions on the image and hidden parameters, the random vector is $\mathbf{U} \triangleq \{\boldsymbol{\theta}, \boldsymbol{\phi}\} = \{\mathbf{x}, a, \mathbf{z}, w, \boldsymbol{\lambda}, \sigma^2\}$ with $\boldsymbol{\theta} = \{\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}\}$ and $\boldsymbol{\phi} = \{a, w, \sigma^2\}$. We propose the following factorized approximating distribution

$$q(\mathbf{U}) = q(\mathbf{x}, a, \mathbf{z}, w, \boldsymbol{\lambda}, \sigma^2) = q(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) q(a, w, \sigma^2). \tag{4.19}$$

Ignoring constants[3], (4.18) leads to

$$\begin{aligned} \ln q(a, w, \sigma^2) = &\underbrace{\mathrm{E}_{\setminus a} \ln p(\mathbf{x}|a,\mathbf{z}) p(a)}_{\ln q(a)} + \\ &\underbrace{\mathrm{E}_{\setminus w} \ln p(\mathbf{z}|w) p(w)}_{\ln q(w)} + \underbrace{\mathrm{E}_{\setminus \sigma^2} \ln p(\mathbf{y}|\mathbf{x},\sigma^2) p(\sigma^2)}_{\ln q(\sigma^2)} \end{aligned} \tag{4.20}$$

[2] In the sequel, we use both $\mathrm{E}[\cdot]$ and $\langle\cdot\rangle$ to denote the expectation. To make our expressions more compact, we use subscripts to denote expectation with respect to the random variables in the subscripts. These notations with the subscripts of '$\setminus\mathbf{v}$' denote expectation with respect to all random variables except for the variable $\mathbf{v}$. e.g. $\mathrm{E}_{\setminus \mathbf{U}_j}$

[3] In the sequel, constant terms with respect to the variables of interest can be omitted in equations.

which induces the factorization

$$q(\boldsymbol{\phi}) = q(a)q(w)q(\sigma^2). \tag{4.21}$$

Similarly, the factorized distribution for $\mathbf{x}$, $\mathbf{z}$ and $\boldsymbol{\lambda}$ is

$$q(\boldsymbol{\theta}) = \left[\prod_i q(x_i|z_i)\right] q(\mathbf{z})q(\boldsymbol{\lambda}) \tag{4.22}$$

leading to the fully factorized distribution

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}) = \left[\prod_i q(x_i|z_i)\right] q(a)q(\mathbf{z})q(w)q(\boldsymbol{\lambda})q(\sigma^2) \tag{4.23}$$

4.2.3 Approximating distribution q

In this section, we specify the marginal distributions in the approximated posterior distribution required in (4.23). More details are described in Appendix 4.9. The parameters for the posterior distributions are evaluated iteratively due to the mutual dependence of the parameters in the distributions for the hidden variables, as illustrated in Algorithm 5.

4.2.3.1 Posterior surrogate for a

$$q(a) = \mathcal{IG}(\tilde{\alpha}_0, \tilde{\alpha}_1), \tag{4.24}$$

with $\tilde{\alpha}_0 = \alpha_0 + \sum\langle z_i \rangle, \tilde{\alpha}_1 = \alpha_1 + \sum\langle z_i x_i \rangle$.

4.2.3.2 Posterior surrogate for w

$$q(w) = \mathcal{B}(\tilde{\beta}_0, \tilde{\beta}_1), \tag{4.25}$$

with $\tilde{\beta}_0 = \beta_0 + N - \sum\langle z_i\rangle, \tilde{\beta}_1 = \beta_1 + \sum\langle z_i\rangle$.

4.2.3.3 Posterior surrogate for σ^2

$$q(\sigma^2) = \mathcal{IG}(\tilde{\varsigma}_0, \tilde{\varsigma}_1), \tag{4.26}$$

with $\tilde{\varsigma}_0 = P/2 + \varsigma_0$, $\tilde{\varsigma}_1 = \langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle/2 + \varsigma_1$, and $\langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle = \|\mathbf{y} - \langle\mathbf{H}\rangle\langle\mathbf{x}\rangle\|^2 + \sum \mathrm{var}[x_i]\,[\|\langle\boldsymbol{\kappa}\rangle\|^2 + \sum_l \sigma_{c_l}\|\boldsymbol{\kappa}_l\|^2] + \sum_l \sigma_{c_l}\|\mathbf{H}^l\langle\mathbf{x}\rangle\|^2$, where σ_{c_l} is the variance of the Gaussian distribution $q(c_l)$ given in (4.32) and $\mathrm{var}[x_i]$ is computed under the distribution $q(x_i)$ defined in the next section and described in Appendices 4.9.3.

4.2.3.4 Posterior surrogate for x

We first note that

$$\ln q(\mathbf{x}, \mathbf{z}) = \ln q(\mathbf{x}|\mathbf{z})q(\mathbf{z}) = \mathrm{E}\left[\ln p(\mathbf{y}|\mathbf{x}, \sigma^2)p(\mathbf{x}|a, \mathbf{z})p(\mathbf{z}|w)\right]. \tag{4.27}$$

The conditional density of $\mathbf{x}$ given $\mathbf{z}$ is $p(\mathbf{x}|a, \mathbf{z}) = \prod_i^N g_{z_i}(x_i)$, where $g_0(x_i) \triangleq \delta(x_i), g_1(x_i) \triangleq g(x_i|a)$. Therefore, the conditional posterior surrogate for x_i is

$$q(x_i|z_i = 0) = \delta(x_i), \tag{4.28}$$

$$q(x_i|z_i = 1) = \phi_+(\mu_i, \eta_i), \tag{4.29}$$

where $\phi_+(\mu, \sigma^2)$ is a positively truncated-Gaussian density function with the hidden mean μ and variance σ^2, $\eta_i = 1/[\langle\|\mathbf{h}_i\|^2\rangle\langle 1/\sigma^2\rangle]$, $\mu_i = \eta_i[\langle\mathbf{h}_i^T\mathbf{e}_i\rangle\langle 1/\sigma^2\rangle - \langle 1/a\rangle]$, $\mathbf{e}_i = \mathbf{y} - \mathbf{H}\mathbf{x}_{-i}$, $\mathbf{x}_{-i}$ is $\mathbf{x}$ except for the ith entry replaced with 0, and $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. Therefore,

$$q(x_i) = q(z_i = 0)\delta(x_i) + q(z_i = 1)\phi_+(\mu_i, \eta_i), \tag{4.30}$$

which is a Bernoulli truncated-Gaussian density.

4.2.3.5 Posterior surrogate for z

For $i = 1, \ldots, N$,

$$q(z_i = 1) = 1/[1 + C_i'] \text{ and } q(z_i = 0) = 1 - q(z_i = 1), \tag{4.31}$$

with $C_i' = \exp(C_i/2 \times \tilde{\varsigma}_0/\tilde{\varsigma}_1 + \mu_i\tilde{\alpha}_0/\tilde{\alpha}_1 + \ln\tilde{\alpha}_1 - \psi(\tilde{\alpha}_0) + \psi(\tilde{\beta}_0) - \psi(\tilde{\beta}_1))$. ψ is the digamma function and $C_i = \langle\|\mathbf{h}_i\|^2\rangle(\mu_i^2 + \eta_i) - 2\langle\mathbf{e}_i^T\mathbf{h}_i\rangle\mu_i$.

4.2.3.6 Posterior surrogate for c

For $j = 1, \ldots, K$,

$$q(c_j) = \phi(\mu_{c_j}, \sigma_{c_j}), \tag{4.32}$$

where $\phi(\mu, \sigma)$ is the probability density function for the normal distribution with the mean μ and variance σ, $\mu_{c_j} = \dfrac{\langle\mathbf{x}^T\mathbf{H}^{jT}\mathbf{y} - \mathbf{x}\mathbf{H}^{jT}\mathbf{H}^0\mathbf{x} - \sum_{l\neq j}\mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^l c_l\mathbf{x}\rangle}{\langle\mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^j\mathbf{x}\rangle}$, and $1/\sigma_{c_j} = \langle 1/\sigma^2\rangle\langle\mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^j\mathbf{x}\rangle$.

Algorithm 5 VB semi-blind image reconstruction algorithm

```
% Initialization:
Initialize estimates ⟨x^(0)⟩, ⟨z^(0)⟩, and w^(0), and set λ = 0 to have κ̂^(0) = κ_0,
% Iterations:
for t = 1, 2, ..., do
    Evaluate α̃_0^(t), α̃_1^(t) in (4.24) by using ⟨x^(t−1)⟩, ⟨z^(t−1)⟩,
    Evaluate β̃_0^(t), β̃_1^(t) in (4.25) by using ⟨z^(t−1)⟩,
    Evaluate ς̃_0^(t), ς̃_1^(t) in (4.26) from ⟨‖y − Hx‖^2⟩,
    for i = 1, 2, ..., N do
        Evaluate necessary statistics (μ_i, η_i) for q(x_i|z_i = 1) in (4.28),
        Evaluate q(z_i = 1) in (4.31),
        Evaluate ⟨x_i⟩, var[x_i],
        For l = 1, ..., K, evaluate μ_{c_l}, 1/σ_{c_l} for q(c_l) in (4.32),
    end for
end for
```

The final iterative algorithm is presented in Algorithm 5, where required *shap-*

ing parameters under distributional assumptions and related statistics are iteratively updated.

4.3 Simulation Results

We first present numerical results obtained for Gaussian and typical MRFM PSFs, shown in Fig. 4.2 and Fig. 4.6, respectively. Then the proposed variational algorithm is applied to a tobacco virus MRFM data set. There are many possible approaches to selecting hyperparameters, including the non-informative approach of [37] and the expectation-maximization approach of [23]. In our experiments, hyper-parameters ς_0, ς_1, α_0, and α_1 for the densities are chosen based on the framework advocated in Chapter III. This leads to the vague priors corresponding to selecting small values $\varsigma_0 = \varsigma_1 = \alpha_0 = \alpha_1 = 1$. For w, the noninformative initialization is made by setting $\beta_0 = \beta_1 = 1$, which gives flexibility to the surrogate posterior density for w. The resulting prior Beta distribution for w is a uniform distribution on $[0, 1]$ for the mean proportion of non-zero pixels.

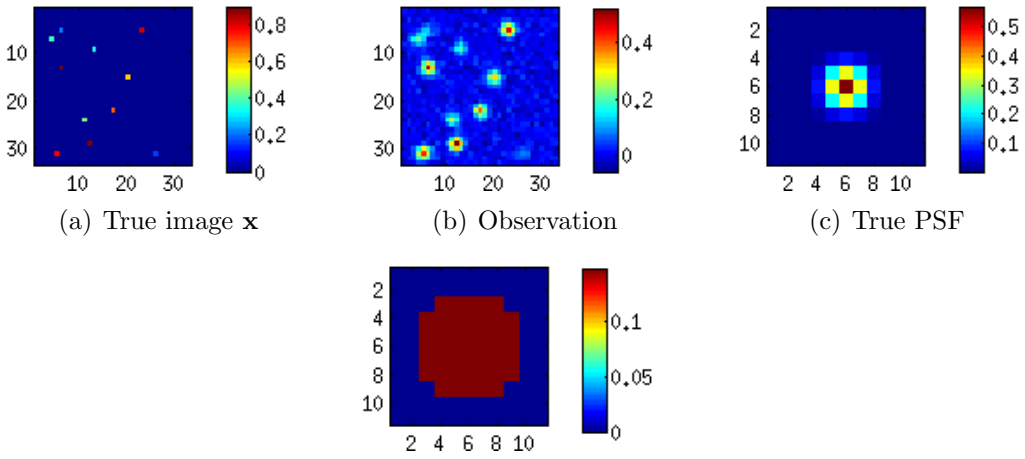
$$w \sim \mathcal{B}(\beta_0, \beta_1) \sim \mathcal{U}([0, 1]) . \tag{4.33}$$

The initial image used to initialize the algorithm is obtained from one Landweber iteration [75].

4.3.1 Simulation with Gaussian PSF

The true image $\mathbf{x}$ used to generate the data, observation $\mathbf{y}$, the true PSF, and the initial, mismatched PSF are shown in Fig. 4.2. Some quantities of interest, computed from the outputs of the variational algorithm are depicted as functions of the iteration number in Fig. 4.3. These plots indicate that convergence to the steady state is achieved after few iterations. In Fig. 4.3, $\mathrm{E}[w]$ and $\mathrm{E}[1/a]$ get close to the true level

but $\mathrm{E}\left[1/\sigma^2\right]$ shows a deviation from the true values. This large deviation implies that our estimation of noise level is conservative; the estimated noise level is larger than the true level. This relates to the large deviation in projection error from noise level (Fig. 4.3(a)). The drastic changes in the initial steps seen in the curves of $\mathrm{E}\left[1/a\right], \mathrm{E}\left[w\right]$ are due to the imperfect prior knowledge (initialization). The final estimated PSF and reconstructed image are depicted in Fig. 4.4, along with the reconstructed variances and posterior probability of $z_i \neq 0$. We decomposed the support region of the true PSF to produce orthonormal bases $\{\boldsymbol{\kappa}_i\}_i$ shown in Fig. 4.5. We extracted 4 bases because these four PSF bases clearly explain the significant part of the true Gaussian PSF. In other words, little energy resides outside of this basis set in PSF space. The reconstructed PSF clearly matches the true one, as seen in Fig. 4.2 and Fig. 4.4. Note that the restored image is slightly attenuated while the restored PSF is amplified because of intrinsic scale ambiguity.



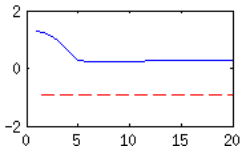
(a) True image $\mathbf{x}$ (b) Observation (c) True PSF

(d) Mismatched PSF

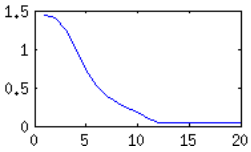
Figure 4.2: Experiment with Gaussian PSF: true image, observation, true PSF, and mismatched PSF ($\boldsymbol{\kappa}_0$).

4.3.2 Simulation with MRFM type PSFs

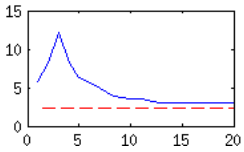
The true image $\mathbf{x}$ used to generate the data, observation $\mathbf{y}$, the true PSF, and the initial, mismatched PSF are shown in Fig. 4.6. The PSF models the PSF of the



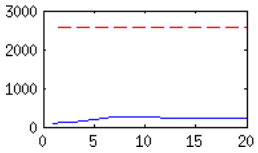
(a) $\log \|\mathbf{y} - \mathrm{E}\mathbf{H}\mathrm{E}\mathbf{x}\|^2$ (solid line) and noise level (dashed line)



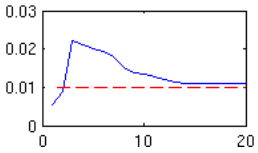
(b) $\log \|\mathbf{x}_{true} - \mathrm{E}\mathbf{x}\|^2$



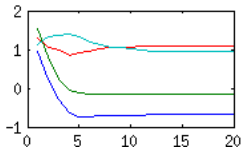
(c) $\mathrm{E}[1/a]$ (solid line) and true value (dashed line)



(d) $\mathrm{E}\left[1/\sigma^2\right]$ (solid line) and true value (dashed line)



(e) $\mathrm{E}[w]$ (solid line) and true value (dashed line)



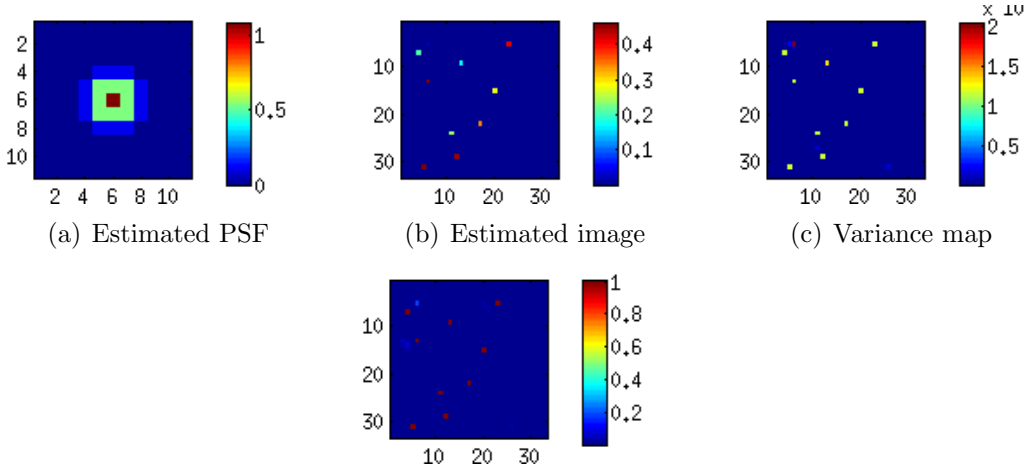
(f) $\mathrm{E}[\mathbf{c}]$. Four PSF coefficients.

Figure 4.3: Result of Algorithm 5: curves of residual, error, $\mathrm{E}\left[1/a\right], \mathrm{E}\left[1/\sigma^2\right], \mathrm{E}\left[w\right], \mathrm{E}\left[\mathbf{c}\right]$, as functions of number of iterations. These curves show how fast the convergence is achieved.

MRFM instrument, derived by Mamin *et al.* [8]. The convergence of the algorithm is achieved after the 10th iteration. The reconstructed image can be compared to the true image in Fig. 4.7, where the pixel-wise variances and posterior probability of $z_i \neq 0$ are rendered. The PSF bases are obtained by the procedure proposed in Section 4.1.2 with the simplified MRFM PSF model and the nominal parameter values [21]. Specifically, by detecting a knee $K = 4$ at the scree plot, explaining more than 98.69% of the observed perturbations as described in Chapter III, we use the first four eigenfunctions, corresponding to the first four largest eigenvalues. The resulting $K = 4$ principal basis vectors are depicted as in Chapter III. The reconstructed PSF with the bases clearly matches the true one, as seen in Fig. 4.6 and Fig. 4.7.

4.3.3 Comparison with PSF-mismatched reconstruction

The results from the variational deconvolution algorithm with a mismatched Gaussian PSF and a MRFM type PSF are presented in Fig. 4.9 and Fig. 4.10, respectively; the relevant PSFs and observations are presented in Fig. 4.2 in Section 4.3.1 and in Fig. 4.6 in Section 4.3.2, respectively. Compared with the results of our VB semi-blind



(a) Estimated PSF (b) Estimated image (c) Variance map

(d) Weight map

Figure 4.4: (a) Restored PSF, (b) image, (c) map of pixel-wise (posterior) variance, and (d) weight map. $\hat{\boldsymbol{\kappa}} = \mathrm{E}\boldsymbol{\kappa}$ is close to the true one. A pixel-wise weight shown in (d) is the posterior probability of the pixel being a nonzero signal.

algorithm (Algorithm 5), shown in Fig. 4.4 and Fig. 4.7, the reconstructed images from the mismatched non-blind VB algorithm in Fig. 4.9 and Fig. 4.10, respectively, inaccurately estimate signal locations and blur most of the non-zero values.

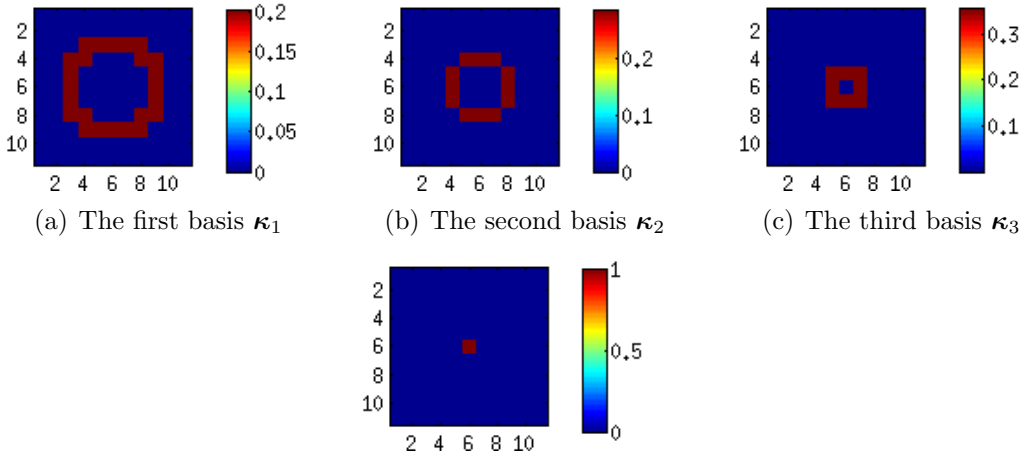
Additional experiments (not shown here) establish that the PSF estimator is very accurate when the algorithm is initialized with the true image.

4.3.4 Comparison with other algorithms

To quantify the comparison, we performed experiments with the same set of four sparse images and the MRFM type PSFs as used in Chapter III. By generating 100 different noise realizations for 100 independent trials with each true image, we measured errors according to various criteria. We tested four sparse images with sparsity levels $\|\mathbf{x}\|_0 = 6, 11, 18, 30$.

Under these criteria[4], Fig. 4.11 visualizes the reconstruction error performance for

[4]Note that the ℓ_0 norm has been normalized. The true image has value 1; $\|\hat{\mathbf{x}}\|_0/\|\mathbf{x}\|_0$ is used for MCMC method; $\mathrm{E}[w] \times N/\|\mathbf{x}\|_0$ for variational method since this method does not produce zero pixels but $\mathrm{E}[w]$.
Note also that, for our simulated data, the (normalized) true noise levels are $\|\mathbf{n}\|^2/\|\mathbf{x}\|_0 =$



(a) The first basis $\boldsymbol{\kappa}_1$ (b) The second basis $\boldsymbol{\kappa}_2$ (c) The third basis $\boldsymbol{\kappa}_3$

(d) The fourth basis $\boldsymbol{\kappa}_4$

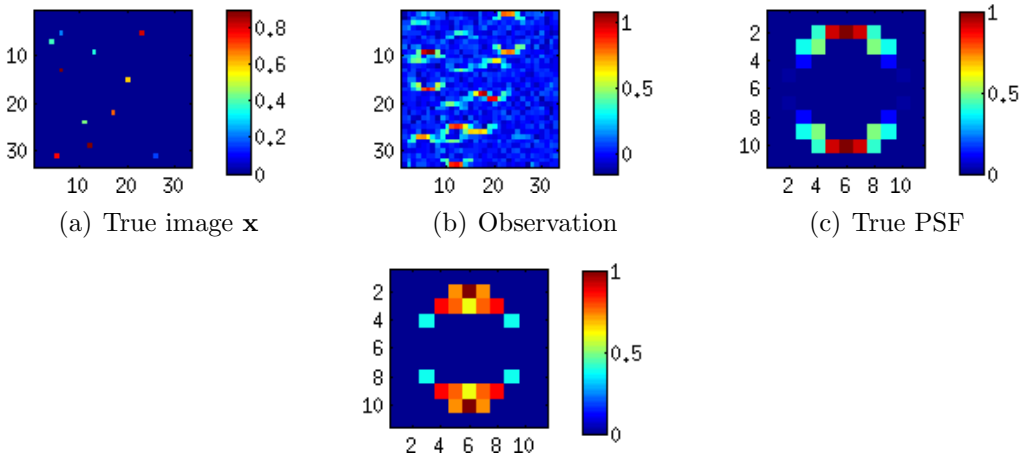
Figure 4.5: PSF bases, $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_4$, for Gaussian PSF.

several measures of error. From these figures we conclude that the VB semi-blind algorithm performs at least as well as the previous MCMC semi-blind algorithm. In addition, the VB method outperforms AM [3] and the mismatched non-blind MCMC [2] methods. In terms of PSF estimation, for very sparse images the VB semi-blind method seems to outperform the MCMC method. Also, the proposed VB semi-blind method converges more quickly and requires fewer iterations. For example, the VB semi-blind algorithm converges in approximately 9.6 seconds after 12 iterations, but the previous MCMC algorithm takes more than 19.2 seconds after 40 iterations to achieve convergence[5].

In addition, we made comparisons between our sparse image reconstruction method and other state-of-the-art blind deconvolution methods [34, 40, 41, 27, 19, 20], as shown in Chapter III. These algorithms were initialized with the nominal, mismatched PSF and were applied to the same sparse image as our experiment above. For a fair comparison, we made a sparse prior modification in the image model of other algorithms, as needed. Most of these methods do not assume or fit into the sparse model

$0.1475, 0.2975, 0.2831, 0.3062$ for $\|\mathbf{x}\|_0 = 6, 11, 18, 30$, respectively.

[5]The convergence here is defined as the state where the change in estimation curves over time is negligible.



(a) True image $\mathbf{x}$ (b) Observation (c) True PSF

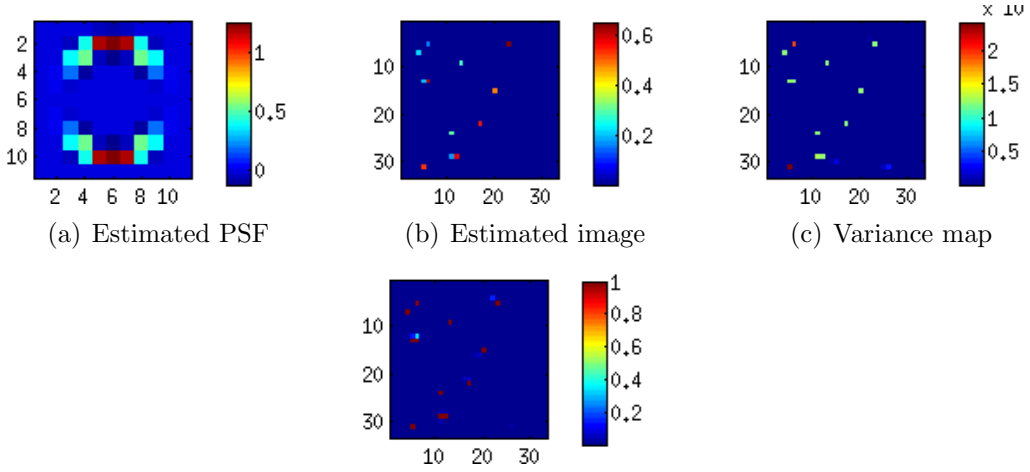
(d) Mismatched PSF

Figure 4.6: Experiment with simplified MRFM PSF: true image, observation, true PSF, and mismatched PSF ($\boldsymbol{\kappa}_0$).

in our experiments, thus leading to poor performance in terms of image and PSF estimation errors. Among these tested algorithms, two of them, proposed by Tzikas *et al.* [27] and Almeida *et al.* [41], produced non-trivial and convergent solutions and the corresponding results are compared to ours in Fig. 4.11. By using basis kernels the method proposed by Tzikas *et al.* [27] uses a similar PSF model to ours. Because a sparse image prior is not assumed in their algorithm [27], we applied their suggested PSF model along with our sparse image prior for a fair comparison. The method proposed by Almeida *et al.* [41] exploits the sharp edge property in natural images and uses initial, high regularization for effective PSF estimation. Both of these perform worse than our VB method as seen in Fig. 4.11. The remaining algorithms [34, 40, 19, 20], which focus on photo image reconstruction or motion blur, either produce a trivial solution ($\hat{\mathbf{x}} \approx \mathbf{y}$) or are a special case of Tzikas's model [27].

To show lower bound our myopic reconstruction algorithm, we used the IST algorithm with a true PSF to effectively restore sparse images with a sparsity constraint. We demonstrate comparisons of the computation time[6] of our proposed reconstruction algorithm to that of others in Table 4.1.

[6]Matlab is used under Windows 7 Enterprise and HP-Z200 (Quad 2.66 GHz) platform.



(a) Estimated PSF (b) Estimated image (c) Variance map

(d) Weight map

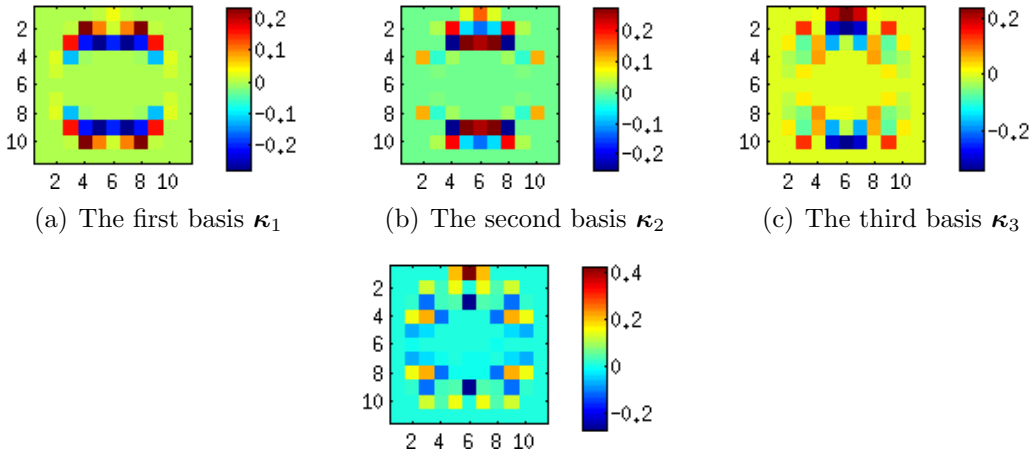
Figure 4.7: Restored PSF and image with pixel-wise variance and weight map. $\hat{\boldsymbol{\kappa}} = \mathrm{E}\boldsymbol{\kappa}$ is close to the true one.

Table 4.1: Computation time of algorithms (in seconds), for the data in Fig. 4.6.

Our method	9.58
semi-blind MC in Chapter III	19.20
Bayesian nonblind [2]	3.61
AM [3]	0.40
Almeida's method [41]	5.63
Amizic's method [40]	5.69
Tzikas's method [27]	20.31
(oracle) IST [71]	0.09

4.3.5 Application to tobacco mosaic virus (TMV) data

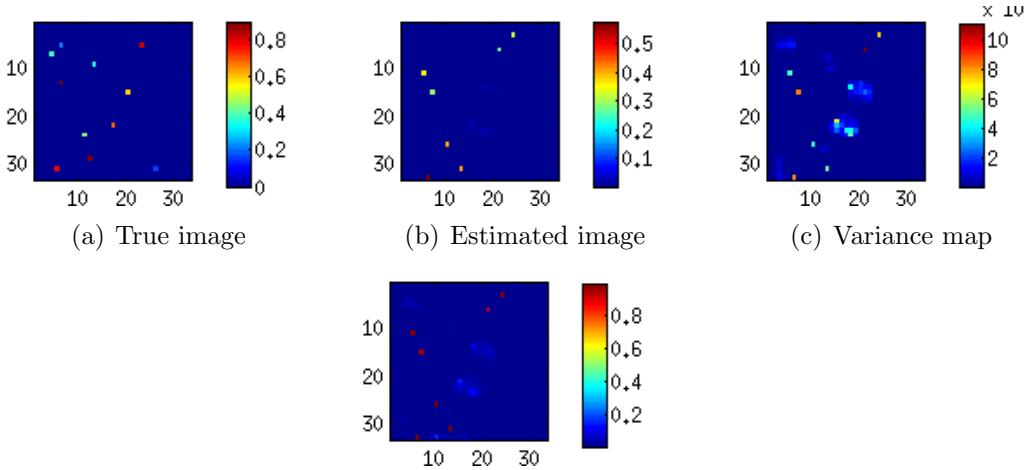
We applied the proposed variational semi-blind sparse deconvolution algorithm to the tobacco mosaic virus data, made available by our IBM collaborators [4], shown in Fig. 3.10 of Chapter III. Our algorithm is easily modifiable to these 3D raw image data and 3D PSF with an additional dimension in dealing with basis functions to evaluate each voxel value x_i. The noise is assumed Gaussian [45, 4] and the four PSF bases are obtained by the procedure proposed in 4.1.2 with the physical MRFM PSF model and the nominal parameter values [8]. The reconstruction of the 6th layer is



(a) The first basis $\boldsymbol{\kappa}_1$ (b) The second basis $\boldsymbol{\kappa}_2$ (c) The third basis $\boldsymbol{\kappa}_3$

(d) The fourth basis $\boldsymbol{\kappa}_4$

Figure 4.8: PSF bases, $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_4$, for MRFM PSF.



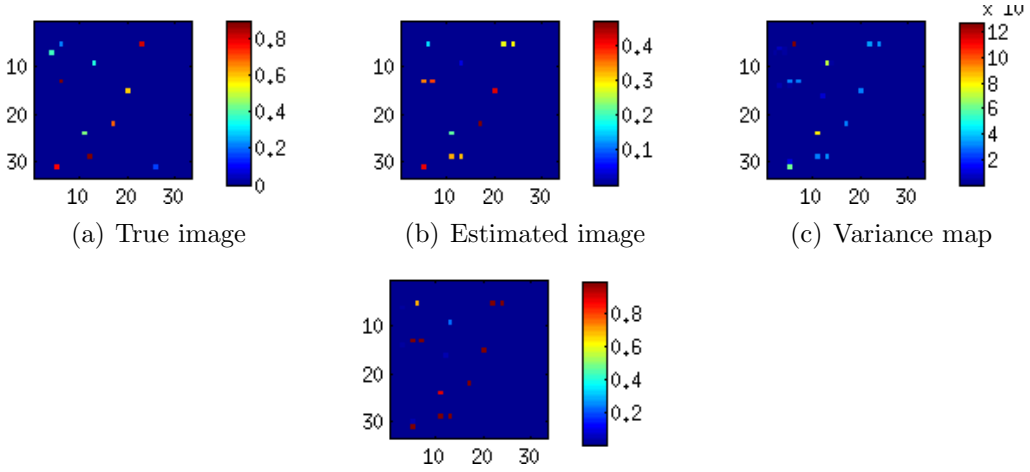
(a) True image (b) Estimated image (c) Variance map

(d) Weight map

Figure 4.9: (mismatched) Non-blind result with a mismatched Gaussian PSF.

shown in Fig. 4.12(a), and is consistent with the results obtained by other methods. (see [37, 2].) The estimated deviation in PSF is small, as predicted in Chapter III.

While they now exhibit similar smoothness, the VB and MCMC images are still somewhat different since each algorithm follows different iterative trajectory in the high dimensional space of 3D images, thus converging possibly to slightly different stopping points near the maximum of the surrogate distribution. We conclude that the two images from VB and MCMC are comparable in that both represent the 2D SEM image well, but VB is significantly faster.



(a) True image (b) Estimated image (c) Variance map

(d) Weight map

Figure 4.10: (mismatched) Non-blind result with a mismatched MRFM type PSF.

4.4 Discussion

4.4.1 Solving scale ambiguity

In blind deconvolution, joint identifiability is a common issue. For example, because of scale ambiguity, the unicity cannot be guaranteed in a general setting. It is not proven in our solution either. However, the shift/time ambiguity issue noticed in [76] is implicitly addressed in our method using a nominal and basis PSFs. Moreover, our constraint on the PSF space using a basis approach effectively excludes a delta function as a PSF solution, thus avoiding the trivial solution. Secondly, the PSF solution is restricted to this linear spanning space, starting form the initial, mismatched PSF. We can, therefore, reasonably expect that the solution provided by the algorithm is close to the true PSF, away from the trivial solution or the initial PSF.

To resolve scale ambiguity in a MCMC Bayesian framework, stochastic samplers are proposed in [76] by imposing a fixed variance on a certain distribution[7]. Another approach to resolve the scale ambiguity is to assume a hidden scale variable that is multiplied to the PSF and dividing the image (or vice versa.), where the scale is

[7]We note that this MCMC method designed for 1D signal deconvolution is not efficient for analyzing 2D and 3D images, since the grouped and marginalized samplers are usually slow to converge requiring hundreds of iterations [76].

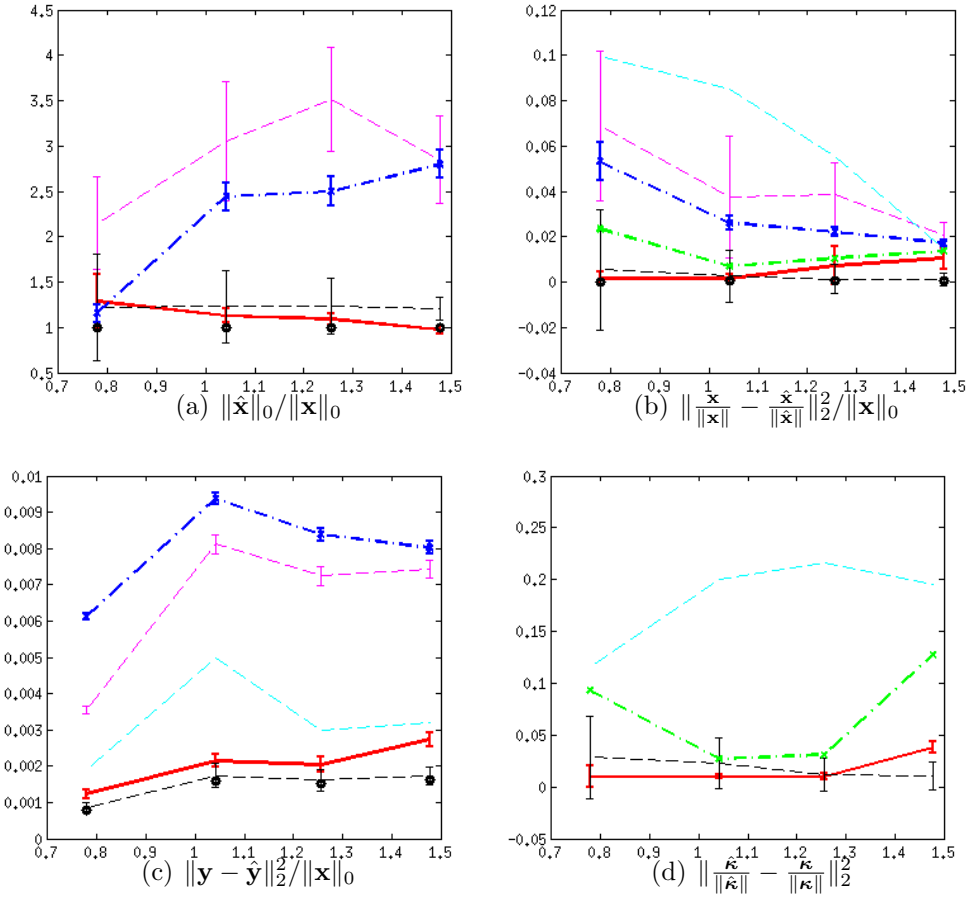
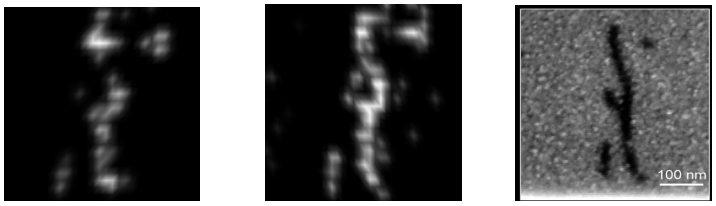


Figure 4.11: For various image sparsity levels (x-axis: $\log_{10}\|\mathbf{x}\|_0$), performance of several blind, semi-blind, and nonblind deconvolution algorithms: the proposed method (red), AM (blue), Almeida's method (green), Tzikas's method (cyan), semi-blind MC (black), mismatched nonblind MC (magenta). Errors are illustrated with standard deviations. (a): Estimated sparsity. Normalized true level is 1 (black circles). (b): Normalized error in reconstructed image. For the lower bound, information about the true PSF is only available to the oracle IST (black circles). (c): Residual (projection) error. The noise level appears in black circles. (d): PSF recovery error, as a performance gauge of our semi-blind method. At the initial stage of the algorithm, $\|\frac{\boldsymbol{\kappa}_0}{\|\boldsymbol{\kappa}_0\|} - \frac{\boldsymbol{\kappa}}{\|\boldsymbol{\kappa}\|}\|_2^2 = 0.5627$. (Some of the sparsity measure and residual errors are too large to be plotted together with results from other algorithms.)



(a) VB estimate (b) MC estimate (c) SEM [4]

Figure 4.12: (a) estimated virus image by VB, (b) estimated virus image by our stochastic method in Chapter III, and (c) virus image from electron microscope [4].

drawn along each iteration of the Gibbs sampler [77].

4.4.2 Exploiting spatial correlations

Our Bayesian hierarchical model (Fig. 4.1) does not account for possible spatial dependencies that might exist in the image. Spatial dependency can be easily incorporated in the model by adding a spatial latent variable with an associated prior distribution. This can be accomplished, for example, by adding a hidden Markov random field model to the vector $\mathbf{x}$ in Fig. 4.1 and we investigate this MRF image model using MCMC at the end of Chapter III. Examples of Markov random field models that have been applied to imaging problems similar to ours are Ising or Potts models [78], Gauss-Markov random fields [79], and Hierarchical Dirichlet processes [80]. Bayesian inference of the hidden parameters of such model is feasible using Monte Carlo and Gibbs sampling, as in [81] and [80], and using variational Bayes EM [82]. Spatial dependency extensions of our model using VB is a worthwhile and interesting topic for future study but will not be pursued further in this work.

4.5 Conclusion and Future Work

4.5.1 Conclusion

We suggested a novel variational solution to a semi-blind sparse deconvolution problem. Our method uses Bayesian inference for image and PSF restoration with a sparsity-inducing image prior via the variational Bayes approximation. The proposed variational Bayes algorithm iterates on a hidden variable domain associated with the mixture coefficients, as an extension of Chapter III. Its power in automatically producing all required parameter values from the data merits further attention for the extraction of image properties and retrieval of necessary features.

From the simulation results, we conclude that the performance of the VB method competes with MCMC methods in sparse image estimation, while requiring fewer computations. Compared to a non-blind algorithm whose mismatched PSF leads to imprecise and blurred signal locations in the restored image, the VB semi-blind algorithm correctly produces sparse image estimates. The benefits of this solution compared to our stochastic solution in Chapter III are: 1) faster convergence, 2) stability of the method, and 3) memory efficiency.

4.5.2 Future work: extension to PSF model for application to computational astronomy

In this section, the global invariance assumption of the point spread function is relaxed. This work is inspired by computational astronomy, where the PSF is space-variant, even in an observation acquired from one charge-coupled device (CCD). This imperfection of the telescope modality challenges the joint PSF and image estimation process, because the violation of the PSF invariance further exacerbates the ill-conditioned problem of semi-blind deconvolution. We address this problem by assuming local invariance near the astronomical object of interest, rather than working

on the direct simultaneous estimation of the PSF and image.

4.5.2.1 PSF model extension for SDSS project

The extension of the PSF model is motivated by the recent success of the Sloan Digital Sky Survey (SDSS), the global project that retrieves a huge amount of astronomical image data to map the universe [83]. Within this project, we focus on estimating galactic objects and their light profiles or shapes. In observations from ground telescopes, the galaxies recorded are adversely affected by the turbulence of the media, the imperfection of optical instruments, the variable observational conditions, the limit of the imaging resolution, and additive noise. In order to offset some of these adverse effects, the raw data from optical telescopes are pre-processed to remove artificial receiver effects and eliminate satellite scratches. Especially, because there are multiple CCDs that may have different light sensitivity (receiver efficiency) and that may overlap in the imaging scope, this CCD effect is re-calibrated [84]. In these pre-processed images, the stars and galaxies are represented as small dots and ellipses set within a void space, which appears as a black background.

Although the pre-processing has removed some of the adverse effects in the images, additional procedures are required to eliminate residual noise and blur effects caused by some inherent characteristics of optics. The first necessary process is denoising, which typically involves applying classical least-square type approaches with appropriate regularization. The second is deblurring, which involves deconvolution in usual circumstances where a known point spread function (PSF) or sensor response is used to sharpen the image. The main problems associated with deconvolution of the galactic objects are two-fold: first, the conventional assumption of a known PSF does not apply to our astronomical data and second, the PSF is not generally space-invariant over the image ranges in our data. The second issue especially exacerbates the deblurring process. Therefore, unlike suppressing noise, deblurring requires more

caution and different approaches from standard deconvolution methods.

By addressing these two problems, we propose a solution to the estimation of both the galaxy image and the PSF. One key assumption of our solution is to consider a stellar light response as a candidate PSF of the region. If the region of interest is relatively small or the variation of the PSFs is negligible, we can assume local space invariance of the PSF to be estimated. Under this premise, our proposed semi-blind image reconstruction method can apply to the data.

4.5.2.2 Estimation of a spatially variant PSF

Assuming local space invariance of the PSF, we can initialize PSFs near the image region of interest. Since the data points (measured PSFs) are not regularly spaced nor dense, but sparsely located, we should resort to interpolation/extrapolation approaches that do not require such a restriction. There are several strategies for guessing an initial PSF satisfying this condition.

The simplistic approach is to assume the light profile of the nearest star[8] as the initial guess of the PSF. This is equivalent to the interpolation using nearest neighbors (Voronoi tessellation). The second approach is to take the weighted average of the stars nearby. Since we have a catalog of the coordinates of galaxies and stars, it seems reasonable to use this distance information. This averaging can be applied to the generalized nearest-neighbor approach (natural neighbor interpolation) with distance-dependent weights for neighbors. For the weights, different powers of distance can be considered and the inverse distance weighting (IDW) provides the model, as follows.

If we have measured data points on $\{x_i\}_{i=1,...,N}$, e.g. the PSF at the coordinate x_i or $\boldsymbol{\kappa}_i := \boldsymbol{\kappa}(x_i)$ for $i = 1, ..., N$ is known, we use the following interpolation schemes to predict $\boldsymbol{\kappa}(x)$ at $x \in \mathbb{R}^2$.

[8]Stars are categorized among astronomical objects by using their light intensity profiles in the SDSS project into the format of Flexible Image Transport System (FITS).

$$\boldsymbol{\kappa}(x) = \sum_{i=0}^{N} \frac{w_i(x)\boldsymbol{\kappa}_i}{\sum_{j=0}^{N} w_j(x)}, \tag{4.34}$$

where $w_i(x) = \frac{1}{l_i^p}$ is a weighting function and l_i is a distance metric between x and x_i or the distance between a galaxy of interest and a ith star nearby in the same field or CCD.

This approach can be interpreted as an optimization over the PSF field by minimizing the ℓ_2 norm difference, as follows:

$$\boldsymbol{\kappa}_0 = \underset{\boldsymbol{\kappa}}{\operatorname{argmin}} \sum_i \frac{\|\boldsymbol{\kappa} - \boldsymbol{\kappa}_i\|^2}{l_i^p} = \frac{\sum_i \boldsymbol{\kappa}_i / l_i^p}{\sum_i 1/l_i^p}, \tag{4.35}$$

where $\boldsymbol{\kappa}_i$ is the light profile of the ith star. For example, for $N = 2$ and $p = 1$, $\boldsymbol{\kappa}_0 = \frac{l_2}{l_1+l_2}\boldsymbol{\kappa}_1 + \frac{l_1}{l_1+l_2}\boldsymbol{\kappa}_2$. This initialization is reasonable if the galaxy is between these stars. The limitation of this approach is that the estimation does not explain the gradient or the change of the light profiles over the extended space.

The third approach that addresses this gradual change is the combination of linear extrapolation and interpolation. In other words, this is to interpolate among projected $\boldsymbol{\kappa}_i$s. As shown in Fig. 4.13, these linearly projected light profiles ($\boldsymbol{\kappa}(x)$) are evaluated with the coordinates of the projected points. Each orthogonal projection point is made for the galaxy considering each pair of stars. The initial guess is performed as follows,

$$\boldsymbol{\kappa}_0 = \underset{\boldsymbol{\kappa}}{\operatorname{argmin}} \sum_{i<j} \frac{\|\boldsymbol{\kappa} - \boldsymbol{\kappa}_{i,j}\|^2}{l_{i,j}^p} = \frac{\sum_{i<j} \boldsymbol{\kappa}_{i,j} / l_{i,j}^p}{\sum_{i<j} 1/l_{i,j}^p}, \tag{4.36}$$

where $\boldsymbol{\kappa}_{i,j}$ is a spatially linear (extrapolated) function whose value at the coordinate of $\boldsymbol{\kappa}_i$ or $\boldsymbol{\kappa}_j$ is exactly $\boldsymbol{\kappa}_i$ or $\boldsymbol{\kappa}_j$, respectively. Similar to the approach above, we use $p = 1$, and set $l_{i,j} = 1/(1/l_i + 1/l_j)$, instead of the distance between $\boldsymbol{\kappa}_{i,j}$ and the galaxy, to take into account the actual distance between the stars and galaxy.

We present several initialization maps based on assigning nearest neighbors, av-

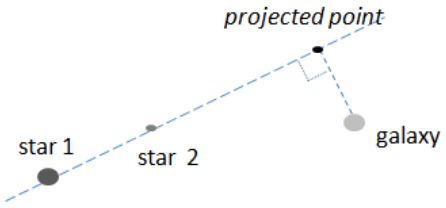


Figure 4.13: Initial guess of PSF based on the line projection (extrapolation of PSF in a 2D space).

eraging neighbors, and averaging extrapolated PSFs in Fig. 4.14. We also note that the difference between two initializations of (4.34) and (4.35) reduces as the number or density of PSFs increases.

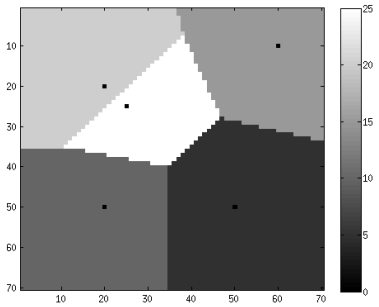
We present a simple result from astronomical image data in Fig. 4.15, by using our proposed variational Bayes method in Chapter IV and a mean of local stars as an initial PSF.

4.6 Appendix 1: Inverse Gamma Distribution

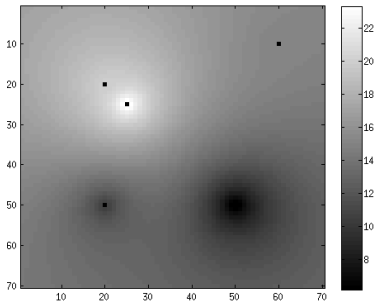
The density of an inverse Gamma random variable $X \sim \mathcal{IG}(a, b)$ is $\frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-\frac{b}{x})$, for $x \in (0, \infty)$. $\mathrm{E}X^{-1} = a/b$ and $\mathrm{E}\ln(X) = \ln(b) - \psi(a)$.

4.7 Appendix 2: Beta Distribution

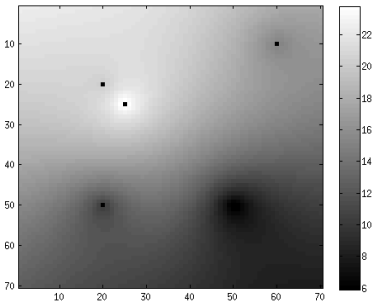
The density of a Beta random variable $X \sim \mathcal{B}(a, b)$ is $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^{b-1}(1-x)^{a-1}$, for $x \in (0, 1)$, with $\Gamma(c) = \int_0^\infty t^{c-1} e^{-t} dt$. The mean of $\mathcal{B}(a, b)$ is $\frac{b}{a+b}$ and $\mathrm{E}\ln(\mathcal{B}(a, b)) = \psi(b) - \psi(a + b)$, where ψ is a digamma function.



(a) Nearest neighbor map

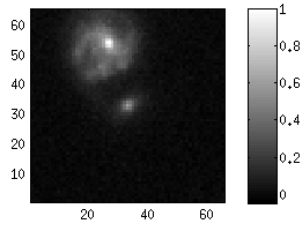


(b) Averaged neighbor map from (4.34)

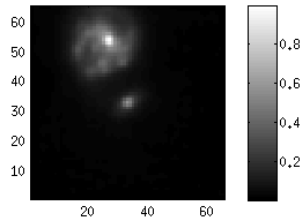


(c) Averaging extrapolated-PSF map from (4.35)

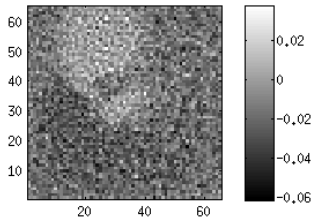
Figure 4.14: PSF initialization maps with $p = 1$. The measured levels can be seen from (a) and the location of PSFs are rendered as black dots.



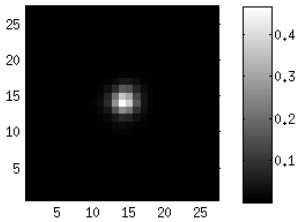
(a) Noisy and blurry observation $\mathbf{y}$



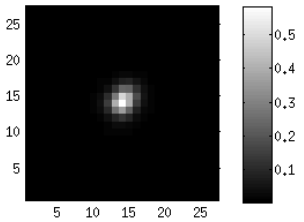
(b) Projected observation $\hat{\mathbf{y}}$



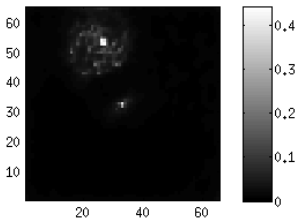
(c) Difference between observation and the projected observation



(d) Initial PSF $\boldsymbol{\kappa}_0$



(e) Estimated PSF $\hat{\boldsymbol{\kappa}}$



(f) Deconvolved galaxy image

Figure 4.15: Semi-blind deconvolution of a galaxy image from SDSS.

4.8 Appendix 3: Positively Truncated Gaussian Distribution

The density of a truncated Gaussian random variable x_i is denoted by $x_i \sim \mathcal{N}_+(x_i; \mu, \eta)$, and its statistics used in this work are

$$\begin{aligned}
\mathrm{E}\left[x_i | x_i > 0\right] &= \mathrm{E}\left[\mathcal{N}_+(x_i; \mu, \eta)\right] \\
&= \mu + \sqrt{\eta}\frac{\phi(-\mu/\sqrt{\eta})}{1 - \Phi_0(-\mu/\sqrt{\eta})}, \\
\mathrm{E}\left[x_i^2 | x_i > 0\right] &= \mathrm{var}[x_i | x_i > 0] + (\mathrm{E}\left[x_i | x_i > 0\right])^2 \\
&= \eta + \mu(\mathrm{E}\left[x_i | x_i > 0\right]),
\end{aligned}$$

where Φ_0 is a cumulative distribution function for the standard normal distribution.

4.9 Appendix 4: Derivations of $q(\cdot)$

In this section, we derive the posterior densities defined by variational Bayes framework in Section 4.2.

4.9.1 Derivation of $q(\boldsymbol{\lambda})$

We denote the expected value of the squared residual term by $R = \mathrm{E}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$. For $\lambda_l, l = 1, \ldots, K$,

$$\begin{aligned}
R =& \mathrm{E}\|\mathbf{y} - \mathbf{H}^0\mathbf{x} - \sum_{l \neq j} \mathbf{H}^l\mathbf{x}\lambda_l - \mathbf{H}^j\mathbf{x}\lambda_j\|^2 \\
=& \lambda_j^2\langle \mathbf{x}^T{\mathbf{H}^j}^T\mathbf{H}^j\mathbf{x}\rangle - 2\lambda_j\langle \mathbf{x}^T{\mathbf{H}^j}^T\mathbf{y} - \mathbf{x}{\mathbf{H}^j}^T\mathbf{H}^0\mathbf{x} \\
& - \sum_{l \neq j} \mathbf{x}^T{\mathbf{H}^j}^T\mathbf{H}^l\lambda_l\mathbf{x}\rangle + const,
\end{aligned}$$

where $\mathbf{H}^j$ is the convolution matrix corresponding to the convolution with $\boldsymbol{\kappa}_j$. For $i \neq j$ and $i, j > 0$, $\mathrm{E}(\mathbf{H}^i\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = tr({\mathbf{H}^i}^T\mathbf{H}^j(\mathrm{cov}(\mathbf{x}) + \langle\mathbf{x}\rangle\langle\mathbf{x}^T\rangle)) = (\mathbf{H}^i\langle\mathbf{x}\rangle)^T(\mathbf{H}^j\langle\mathbf{x}\rangle)$,

since $tr(\mathbf{H}^{i^T}\mathbf{H}^j\mathrm{cov}(\mathbf{x})) = tr(\mathbf{H}^i D^T \mathbf{H}^j D) = \sum_k d_k^2 \mathbf{h}_k^i \mathbf{h}_k^j = 0$. Here, $\mathrm{cov}(\mathbf{x})$ is approximated as a diagonal matrix $D^2 = \mathrm{diag}(d_1^2, \ldots, d_n^2)$. This is reasonable, especially when the expected recovered signal $\hat{\mathbf{x}}$ exhibits high sparsity. Likewise, $\mathrm{E}(\mathbf{H}^0\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = \boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_j \sum_i \mathrm{var}[x_i] + (\mathbf{H}^0\langle\mathbf{x}\rangle)^T(\mathbf{H}^j\langle\mathbf{x}\rangle)$ and $\mathrm{E}(\mathbf{H}^j\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = \|\boldsymbol{\kappa}_j\|^2 \sum_i \mathrm{var}[x_i] + \|\mathbf{H}^j\langle\mathbf{x}\rangle\|^2$.

Then, we factorize $\mathrm{E}\left[-\frac{R}{2\sigma^2}\right] = -\frac{(\lambda_j - \mu_{\lambda_j})^2}{2\sigma_{\lambda_j}}$, with $\mu_{\lambda_j} = \frac{\langle \mathbf{x}^T\mathbf{H}^{j^T}\mathbf{y} - \mathbf{x}\mathbf{H}^{j^T}\mathbf{H}^0\mathbf{x} - \sum_{l\neq j}\mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^l\lambda_l\mathbf{x}\rangle}{\langle \mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^j\mathbf{x}\rangle}$, $1/\sigma_{\lambda_j} = \langle 1/\sigma^2\rangle\langle \mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^j\mathbf{x}\rangle$.

If we set the prior, $p(\lambda_j)$, to be a uniform distribution over a wide range of the real line that covers error tolerances, we obtain a normally distributed variational density $q(\lambda_j) = \phi(\mu_{\lambda_j}, \sigma_{\lambda_j})$ with its mean μ_{λ_j} and variance σ_{λ_j} defined above, because $\ln q(\lambda_j) = \mathrm{E}\left[-\frac{R}{2\sigma^2}\right]$. By the independence assumption, $q(\boldsymbol{\lambda}) = \prod q(\lambda_j)$, so $q(\boldsymbol{\lambda})$ can be easily evaluated.

4.9.2 Derivation of $q(\sigma^2)$

We evaluate R ignoring edge effects; $R = \|\mathbf{y} - \langle\mathbf{H}\rangle\langle\mathbf{x}\rangle\|^2 + \sum \mathrm{var}[x_i][\|\langle\boldsymbol{\kappa}\rangle\|^2 + \sum_l \sigma_{\lambda_l}\|\boldsymbol{\kappa}_l\|^2] + \sum_l \sigma_{\lambda_l}\|\mathbf{H}^l\langle\mathbf{x}\rangle\|^2$. $\|\boldsymbol{\kappa}\|^2$ is a kernel energy in ℓ_2 sense and the variance terms add uncertainty, due to the uncertainty in $\boldsymbol{\kappa}$, to the estimation of density. Applying (4.18), (ignoring constants)

$$\begin{aligned}
\ln q(\sigma^2) &= \mathrm{E}_{\setminus\sigma^2}\left[\ln p(\mathbf{y}|\mathbf{x}, \boldsymbol{\lambda}, \sigma^2)p(\sigma^2)p(\mathbf{x}|a, w)p(w)p(a)\right] \\
&= \mathrm{E}_{\mathbf{x},\boldsymbol{\lambda}}\left[\ln p(\mathbf{y}|\mathbf{x}, \sigma^2)\right] + \ln p(\sigma^2) \\
&= -\frac{\mathrm{E}_{\mathbf{x},\boldsymbol{\lambda}}\left[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\right]}{2\sigma^2} - \frac{P}{2}\ln\sigma^2 + \ln p(\sigma^2). \\
\mathcal{IG}(\tilde{\varsigma}_0, \tilde{\varsigma}_1) &\triangleq q(\sigma^2) = \mathcal{IG}(P/2 + \varsigma_0, \langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle/2 + \varsigma_1).
\end{aligned}$$

($\mathbb{E}_{\setminus\sigma^2}$ denotes expectation with respect to all variables except σ^2.)

4.9.3 Derivation of $q(\mathbf{x})$

For $x_i, i = 1, \ldots, N$, $R = \mathrm{E}\|\mathbf{e}_i - \mathbf{h}_i x_i\|^2$ with $\mathbf{e}_i = \mathbf{y} - \mathbf{H}\mathbf{x}_{-i} = \mathbf{y} - \mathbf{H}^0\mathbf{x}_{-i} - \sum_l \mathbf{H}^l\lambda_l\mathbf{x}_{-i}$, $\mathbf{h}_i = [\mathbf{H}^0 + \sum \mathbf{H}^l\lambda_l]_i = \mathbf{h}_i^0 + \sum \mathbf{h}_i^l\lambda_l = (i\text{th column of } \mathbf{H})$. Ignoring constants, $R = \langle\|\mathbf{h}_i\|^2\rangle x_i^2 - 2\langle\mathbf{h}_i^T\mathbf{e}_i\rangle x_i$.

Using the orthogonality of the kernel bases and uncorrelatedness of λ_l's, we derive the following terms (necessary to evaluate R): $\langle\|\mathbf{h}_i\|^2\rangle = \|\mathbf{h}_i^0\|^2 + \sum_l \sigma_{\lambda_l}\|\mathbf{h}^l{}_i\|^2$ and, $\langle\mathbf{h}_i^T\mathbf{e}_i\rangle = \langle\mathbf{h}_i^T\rangle(\mathbf{y} - \langle\mathbf{H}\rangle\langle\mathbf{x}_{-i}\rangle) - \sum_l \mathrm{var}[\lambda_l]\mathbf{h}_i^{lT}\mathbf{H}^l\langle\mathbf{x}_{-i}\rangle$.

Then, $\mathrm{var}[x_i] = w_i'\mathrm{E}\left[x_i^2|x_i > 0\right] - w_i'^2(\mathrm{E}\left[x_i|x_i > 0\right])^2$, $\mathrm{E}\left[x_i\right] = w_i'\mathrm{E}\left[x_i|x_i > 0\right]$, where $w_i' = q(z_i = 1)$ is the posterior weight for the normal distribution and $1 - w_i'$ is the weight for the delta function. The required statistics of x_i that are used to derive the distribution above are obtained by applying Appendix 4.8.

4.9.4 Derivation of $q(\mathbf{z})$

To derive $q(z_i = 1) = \langle z_i\rangle$, we evaluate the unnormalized version $\hat{q}(z_i)$ of $q(z_i)$ and normalize it. $\ln\hat{q}(z_i = 1) = \mathrm{E}_{\setminus z_i}\left[-\frac{\|\mathbf{e}_i - \mathbf{h}_i x_i\|^2}{2\sigma^2} - \ln a - \frac{x_i}{a} + \ln w\right]$ with $x_i \sim N_+(\mu_i, \eta_i)$ and $\ln\hat{q}(z_i = 0) = \mathrm{E}_{\setminus z_i}\left[-\frac{\|\mathbf{e}_i\|^2}{2\sigma^2} + \ln(1 - w)\right]$ with $x_i = 0$. The normalized version of the weight is $q(z_i = 1) = 1/[1 + C_i']$. $C_i' = \exp(\ln\hat{q}(z_i = 0) - \ln\hat{q}(z_i = 1)) = \exp(C_i/2 \times \langle 1/\sigma^2\rangle + \mu\langle 1/a\rangle + \langle\ln a\rangle + \langle\ln(1-w) - \ln w\rangle = \exp(C_i/2 \times \tilde{\varsigma}_0/\tilde{\varsigma}_1 + \mu\tilde{\alpha}_0/\tilde{\alpha}_1 + \ln\tilde{\alpha}_1 - \psi(\tilde{\alpha}_0) + \psi(\tilde{\beta}_0) - \psi(\tilde{\beta}_1))$. ψ is a digamma function and $C_i = \langle\|\mathbf{h}_i\|^2\rangle(\mu_i^2 + \eta_i) - 2\langle\mathbf{e}_i^T\mathbf{h}_i\rangle\mu_i$.

CHAPTER V

EBSD Image Segmentation Using a Physics-Based Forward Model[1]

In this chapter, we present a novel classification approach for the EBSD analysis that utilizes a physics-based forward model (dictionary). Our proposed methods address the deficiencies of conventional EBSD analysis; without a suitable forward model, conventional methods do not exploit full diffraction patterns and discard much information except for only a few features in computing crystal orientations. Furthermore, the conventional approaches do not have the capability of detecting anomalous structures, which critically affect the material use. Another problem is the low resolution map of grain and boundary structures. By using the dictionary, we use full diffraction images to classify the EBSD patterns into the categories of grain interiors, grain boundaries, and two anomaly types. This framework can also be used to build a super-resolution grain/boundary map.

5.1 Introduction

Electron backscatter diffraction, EBSD, is used to perform quantitative microstructure analysis of crystalline materials on a millimeter to nanometer scale [87, 88].

[1]This chapter is partially based on the papers [85, 86].

In crystalline materials (including typical metals, minerals, and ceramics), the constituent atoms, molecules, or ions that form the material are arranged in an ordered or repetitive pattern in three spatial dimensions. Some crystalline materials are formed of an aggregate of single crystal grains, and this type of material is called polycrystalline. Steel, aluminum, and many materials frequently used in engineering are polycrystalline, and the structure of these polycrystalline materials can be analyzed by using EBSD [87, 89]. This analysis helps identify specific material imperfections, because the structure of the materials is not perfect; the lattice within the single crystal grains can contain defects or anomalies. These imperfections affect the properties of the material, depending on the type of anomaly such as pores and impurities.

To examine the microstructure of such materials, aside from EBSD techniques, one can use chemical etching, optical microscopy and scanning electron microscopy [87]. Compared to these modalities, the advantage of EBSD analysis is the potential to measure crystal orientation, analyze the texture of the material, and identify the position of all grains and grain boundaries. In this sense, an EBSD analysis complements the conventional analysis techniques by combining microstructure and crystallography sensitivities.

Using EBSD analysis, two important characteristics of polycrystalline materials are investigated [87]. One characteristic is that the crystals in the different grains have different orientations; the edges/faces of the crystal lattice are oriented in different directions in different grains. The other is that these materials have boundary regions, called grain boundaries or edges, where the different grains contact each other spatially [90].

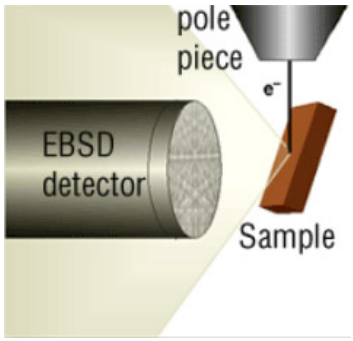
Study of the grain boundary is important, because grain boundaries or interfaces between grains are critical in the quality of material use. In other words, the orientation in the grain boundary or edge can significantly affect material properties, because defects such as fracture and corrosion, for instance, can be initiated at grain

edges. Since different grains have different orientations, EBSD can help identify the boundary using the difference in the orientations of grains on adjacent sides of the boundary.

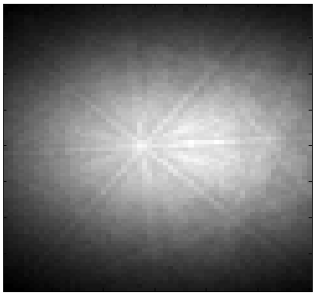
In EBSD, as seen in Fig. 5.1, a beam of electrons is directed onto a tilted crystalline material sample. The electrons then interact with the atoms in the crystal lattice, and a part of the electrons are diffracted by atomic layers in the sample and generate visible lines in the fluorescent phosphorus screen. The image appearing on the screen is called a diffraction pattern. A typical pattern consisting of straight lines (Kikuchi lines) is shown in Fig. 5.1(b). This pattern is essentially a projection of the geometry of the crystal lattice planes, conveying information on the crystalline structure and orientation of the grain. The symmetry of the pattern is closely related to the crystal structure at the beam injection point in the sample. Diffraction patterns are determined by the orientation of a crystal. Indeed, the physical (magnetic or electrical) characteristics of polycrystalline materials are intimately related to the distribution of crystal orientations. Based on the diffraction pattern at each point in the sample, an EBSD analysis produces orientation maps.

In conventional EBSD analysis, the orientations and widths of selected diffraction lines are measured and matched against a pre-computed index for the material under study. The indexing yields the crystal orientation, commonly described by three Euler angles with respect to a reference frame, for the volume illuminated by the beam. By repeating the process on a grid of points in the sample, an orientation map or image is produced. The image is then segmented into grains by thresholding norm differences between the Euler angles. The EBSD measurements can also be fused with other modalities in multimodal analyses [92, 93].

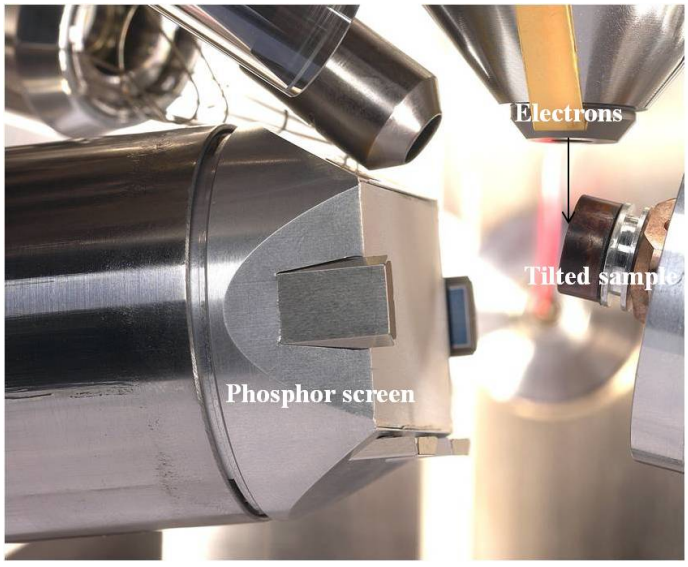
In contrast to the standard indexing approach described above, in this work we use the full diffraction patterns to perform segmentation. This allows us to exploit information that is normally discarded in the indexing process, particularly the char-



(a) An EBSD system. An electron beam is scattered by atoms in different crystal planes in the sample, producing a diffraction pattern on the detector screen. Image taken from [91].



(b) Typical EBSD pattern from IN100 material. Data courtesy of Dr. Megna Shah at AFRL.



(c) The experimental set up for EBSD. Image taken from [87].

Figure 5.1: Electron Backscatter Diffraction

acteristics of the background intensity profile. With this additional information, grain boundaries and anomalous points can be detected as explicit classes at the same time as grains are segmented. Anomaly detection is an important capability, since anomalies may correspond to defects or contaminants that affect the material properties. In addition, processing the diffraction patterns directly avoids problems with indexing when the observed pattern is a poor match to the database of pre-computed expected patterns, as occurs for example at grain boundaries.

Image segmentation techniques have been proposed from various perspectives, including clustering methods [94], region-growing methods [95], level set methods [96, 97], and a recently proposed inverse diffusion and expectation-maximization algorithm for materials images [98]. This work is distinguished by our use of a detailed physics-based model, proposed by Callahan *et al.* [5], for the material, the EBSD interaction, and the experimental setup. We use a collection of the simulated diffraction patterns, also called a dictionary, generated by this model for a large densely sampled set of crystal orientations. Segmentation is done using features that measure similarity between observed diffraction patterns and elements in the dictionary. To the best of our knowledge, we are the first to propose this type of dictionary-based segmentation for EBSD using a physical forward model. The advantage of our approach is its greater robustness to instrument blur and noise and its ability to detect anomalies not represented in the dictionary. At the same time, the large size of both the dictionary and the experimental dataset, together with the high dimension of the diffraction patterns, necessitate the use of highly efficient computational methods. Finally, along with the segmentation results, under our model we refine boundaries to super-resolve boundary locations and to provide statistical evidence, or p-values, for the discovered boundaries by performing hypothesis testing.

5.2 Dictionary-based Forward Model

One difficulty in EBSD analysis is that there is no simple model that accurately describes the physical mechanism that causes the measured diffraction pattern. As an alternative we use a computational forward model proposed by DeGraef [5, 59] that generates EBSD diffraction patterns[2] using electron transport theory and a stochastic scattering model. We use a pre-computed set of patterns, that we call a dictionary, generated numerically from the forward model. The dictionary was provided to us by Marc DeGraef of Carnegie Mellon University.

5.2.1 Diffraction pattern dictionary

As described by Callahan *et al.* [5] the process of generating the dictionary consists of three steps repeated for all Euler angles. The first step models the interaction of the electron beam with a location on the sample using the Schrödinger equation with a Bloch wave ansatz [99]. The backscattered electron yield is calculated for a set of directions covering the hemisphere of all possible exit directions. The second step is to interpolate the intensities over the hemisphere onto the pixel locations on the collecting detector. This interpolation step uses an equal-area projection of the hemisphere onto a square or hexagonal grid [100] to permit standard bilinear interpolation of the intensities. The third step is to model additional instrument effects such as detector quantum efficiency, Poisson noise, coupling optics, and the point spread function and binning mode of the CCD. The EBSD master patterns [5, 59], obtained before the second interpolation step, consider only line patterns and not background or gradation (Fig. 5.2). The dictionary provided to us by DeGraef used a simpler model for the background, consisting of averaging over all diffraction patterns (Fig. 5.3) generated by the master template [5].

The dictionary consists of $M = 281700$ diffraction patterns finely-sampled in

[2]Each pattern is generated by using a sampled Euler angle triplet (ϕ_1, Φ, ϕ_2) [59].

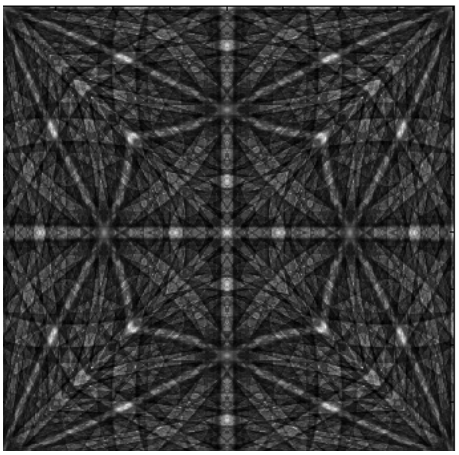
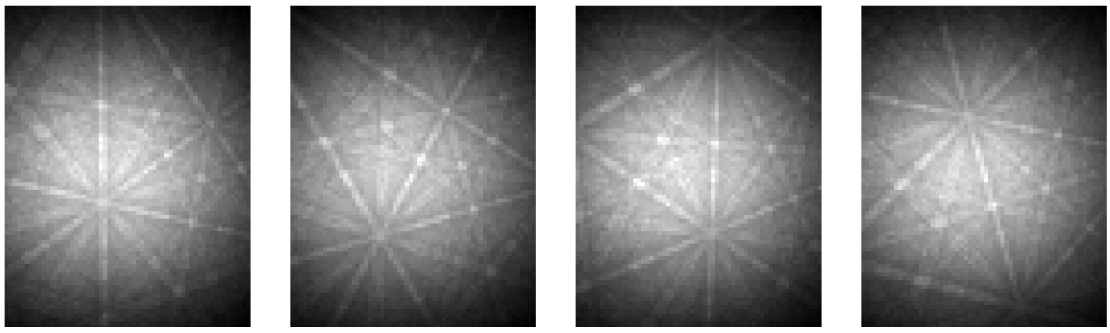


Figure 5.2: The IN100 master pattern defined at the beam energy level 15 keV (From [5], courtesy of Dr. DeGraef.)



(a) index = 3122 (b) index = 25182 (c) index = 93544 (d) index = 202563

Figure 5.3: Several diffraction patterns in the dictionary. Unlike the master pattern in Fig. 5.2, they have a background. The dictionary indices for (a),(b),(c), and (d) are indicated.

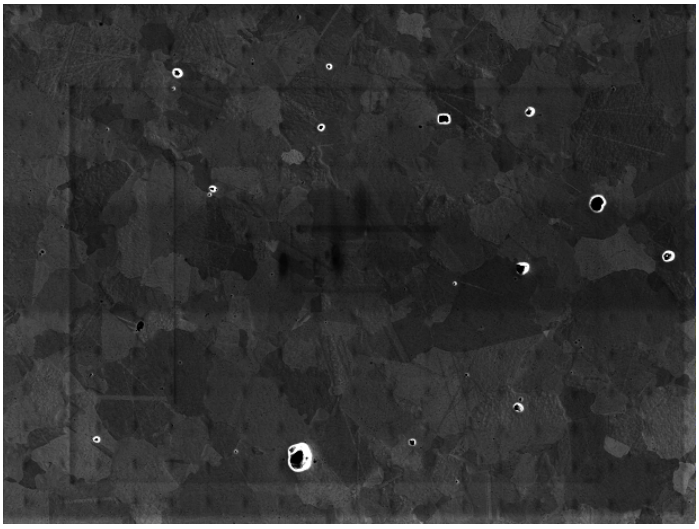


Figure 5.4: The secondary electron image for IN100 sample acquired by Mike Graeber and Megna Shah at AFRL.

orientation (Euler angle) space. The dictionary was generated by Marc DeGraef, using the procedure of [5] explained above, for the cubic (m3m) point symmetry group corresponding to crystals in a Nickel alloy sample (IN100). Due to the 24-fold symmetry of the Euler angles over this group it was sufficient to sample the patterns in a fundamental zone occupying 1/24th of the sphere. Each pattern was downsampled to a 80 times 60 image. The goodness of fit of the pattern dictionary is demonstrated in [5], where it produces simulated patterns close to experimental patterns acquired by Mike Graeber and Megna Shah at AFRL. The secondary electron image of the sample is shown in Fig. 5.4.

5.2.2 Observation model

Here we develop a mathematical model for experimentally observed diffraction patterns using the dictionary described in Section 5.2.1. We represent the dictionary of 281700 diffraction patterns as a 4800×281700 matrix $\mathbf{M}$ with each column $\mathbf{m}_i$ corresponding to a vectorized 80×60 diffraction pattern. In the case where a linear combination of patterns in the dictionary 'perfectly' explain the observed pattern

$\mathbf{y}_j$ measured at the jth sample point, the weight vector $\mathbf{a}_j = [a_{j1}, ..., a_{jM}]^T$ can be expressed by the following observation equation, where the coefficients a_{ji} satisfy the positivity and sum-up-to 1 constraint ($\sum_i a_{ji} = 1$ and $a_{ji} \geq 0$):

$$\mathbf{y}_j = \sum_i \mathbf{m}_i a_{ji} + \mathbf{n}_j, \tag{5.1}$$

where $\mathbf{n}_j$ is the residual error that accounts for noise and model mismatch.

If we model the background as approximately the same at each location on the sample, we may express each observed pattern $\mathbf{y}_j$ as

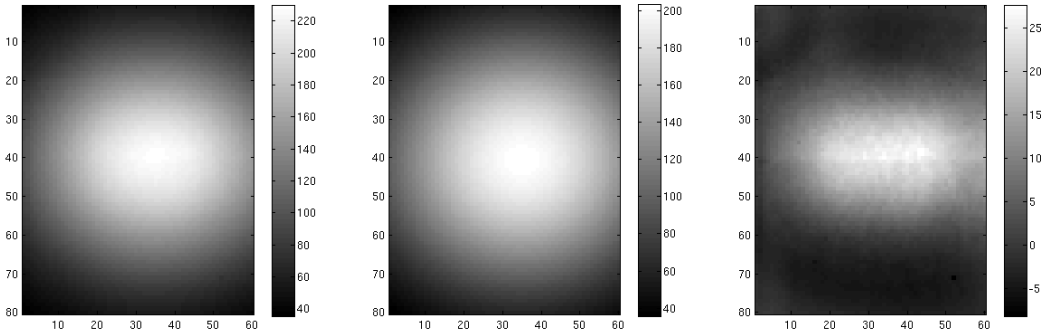
$$\mathbf{y}_j = \alpha_j \mathbf{b}_\mathbf{Y} + \tilde{\mathbf{y}}_j, \tag{5.2}$$

where $\mathbf{b}_\mathbf{Y}$ is the common background component, α_j is a scale factor, and $\tilde{\mathbf{y}}_j$ is orthogonal to $\mathbf{b}_\mathbf{Y}$. We assume that similar decomposition $\mathbf{m}_i = \beta_i \mathbf{b}_\mathbf{M} + \tilde{\mathbf{m}}_i$ holds for the elements of the dictionary.

This decomposition[3] separates the background from the master pattern components of the observed pattern. Least squares estimation of $\tilde{\mathbf{m}}_j$ and $\tilde{\mathbf{y}}_j$ under the model (5.2) results in compensating for background by subtracting the dictionary mean intensity and observation mean intensity from the dictionary and observation, respectively. After this subtraction/projection-away procedure, the characteristic lines in the diffraction patterns in $\{\tilde{\mathbf{y}}_j\}$ and $\{\tilde{\mathbf{m}}_j\}$ are more distinguishable, as seen in the uncompensated patterns in Fig. 5.7 and in the compensated patterns in Fig. 5.9. Both the compensated and uncompensated patterns will be used for segmentation, as presented below.

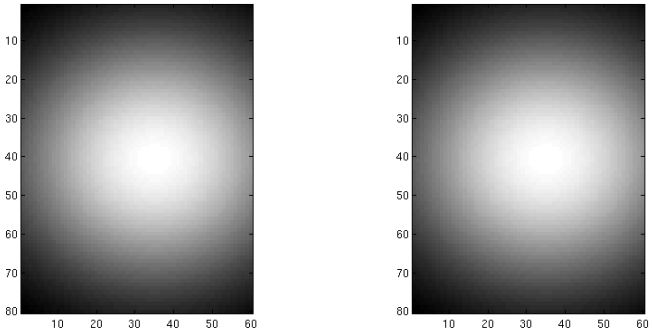
This least squares compensation corresponds closely to projection of the measured pattern onto the subspace that is orthogonal to the first principal component of the

[3]This process has an additional benefit; it compensates the dictionary-observation mismatch as seen in Fig. 5.5.



(a) Mean of observation patterns (b) Mean of dictionary patterns (c) Difference

Figure 5.5: Bias between the forward model (dictionary) and observation



(a) The first singular vector from the SVD analysis of the dictionary (b) Mean of dictionary patterns

Figure 5.6: The closeness of the first singular vector to the mean of the dictionary.

dictionary. Indeed we found that the singular value decomposition (SVD) produces a principal component that is very close to the mean of the patterns, for both the dictionary and observations over the sample; the normalized difference is on the order of 10^{-6} (Fig. 5.6).

Given the orthogonal decomposition in (5.2) and its counterpart for the dictionary, we propose the following linear model that accounts for noise, mismodeling error and possible anomalous pixels.

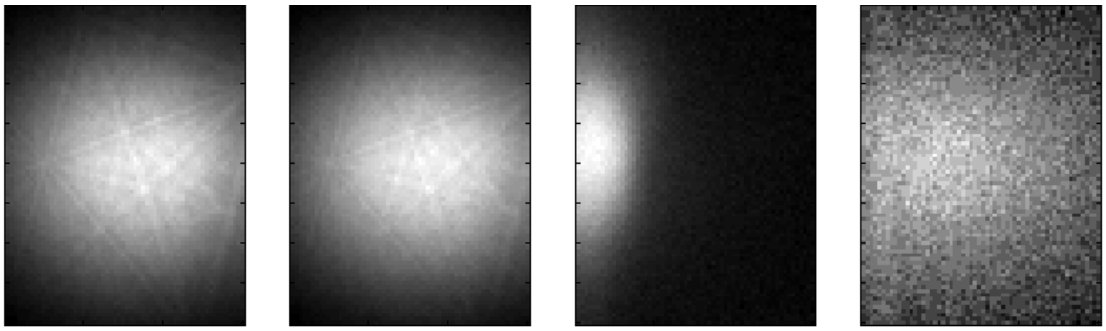
$$\tilde{\mathbf{y}}_j = \sum_i \tilde{\mathbf{m}}_i a_{ij} + \mathbf{n}_j + \mathbf{z}_j. \tag{5.3}$$

Specifically, $\mathbf{n}_j$ represents measurement noise and $\mathbf{z}_j$ is a gross error that represents anomalous patterns and is present only for a small number of indices j. For a pattern $\tilde{\mathbf{y}}_j$ corresponding to a single crystal orientation in the interior of a grain, the number of significant coefficients a_{ij} is expected to be small, corresponding to nearby orientations. For grain boundaries, the coefficients would be close to zero except in a few neighborhoods of the dictionary corresponding to patterns of the grains adjacent to the boundary. Thus we can expect that for non-anomalous patterns $\tilde{\mathbf{y}}_j$, the coefficients $\{a_{ij}\}$ are sparse, i.e., all but a few coefficients are zero. The reader may better appreciate this discussion by referring to Fig. 5.13 and accompanying text below.

Under the model in (5.3), an observed pattern $\mathbf{y}_j$ can be represented by a least squares estimator of the coefficients $\{a_{ij}\}$, or by its sparse version using sparse linear regression [63]. However, due to the high dimension of the dictionary $\mathbf{M}$ (4800×281700), such methods are difficult to implement. In this work, we focus on a simpler dictionary matching pursuit approach that can be interpreted as a special case of a very stringent sparse linear regression representation. This approach uses inner products of each sample pattern with all patterns in the dictionary, as described in Section 5.3. We note that if $\mathbf{n}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{z}_j = \mathbf{0}$, and $\tilde{\mathbf{y}}_j$ is known to be well-approximated by a single dictionary element, then inner products are sufficient to identify the most likely dictionary element and estimate the value of the single nonzero a_{ij}.

5.3 Dictionary Matching Pursuit for Segmentation

In this section, we discuss our dictionary matching pursuit method for segmenting experimental diffraction patterns. The method segments the sample into four classes corresponding to (1) grain interiors, (2) grain boundaries, and two types of anomalies, (3) one with a shifted background pattern and the other (4) where the background is unshifted but the pattern is very noisy. A sample pattern is classified by a classifier



(a) (b) (c) (d)

Figure 5.7: Raw uncompensated diffraction patterns for four representative pixels: (a) grain interior, (b) grain boundary, (c) background-shifted, and (d) high noise.

trained on two sets of dictionary features: (1) the inner products between the sample and the dictionary and (2) locally shared neighborhoods in the dictionary. We describe below these dictionary-derived features used in the classifier before presenting the DT in Section 5.4.

5.3.1 Uncompensated inner-product features for anomaly detection

The first feature used to perform segmentation measures the similarity of an observed diffraction pattern to patterns in the dictionary with the background components included. Since the background typically dominates the total energy in the patterns (more than 99.7% in both our dictionary and experimental dataset), this feature essentially measures the similarity between background profiles and can be a criteria for detecting anomalous patterns. In other words, the anomaly detection can be performed by analyzing the distribution of these measures; outliers in the distribution would be dissimilar to the background pattern.

Define the normalized inner product between uncompensated patterns $\mathbf{m}_i$ in the dictionary and uncompensated observed sample pattern $\mathbf{y}_j$:

$$\rho_{ij} = \frac{\mathbf{m}_i^T \mathbf{y}_j}{\|\mathbf{m}_i\| \, \|\mathbf{y}_j\|}. \tag{5.4}$$

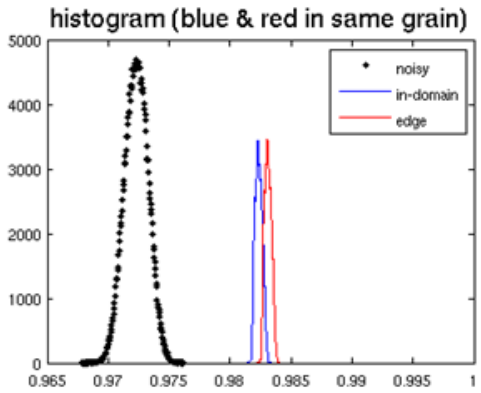


Figure 5.8: Histograms of uncompensated inner products for the three different pixel locations.

Define the mean inner product

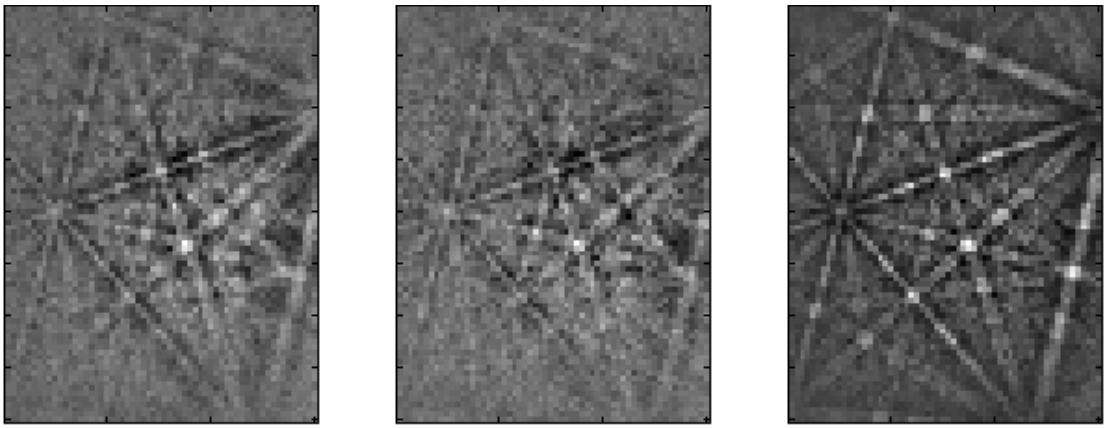
$$\bar{\rho}_j = \frac{1}{M} \sum_{i=1}^{M} \rho_{ij} \tag{5.5}$$

as a measure of background similarity, where $M = 281700$.

We present several histograms of $\rho_{\cdot j}$ for three different patterns (j: pixel locations for interior grain, edge, and noisy patterns) in Fig. 5.8. This measure ρ_{ij} is quite useful to exclude anomalous patterns but is not capable of effectively separating between the grain interior and boundary patterns. For further separation of the in-grain and boundary patterns, we introduce compensated inner-product features in the next section. These features are used in deriving neighborhood similarity features in Section 5.3.3.

5.3.2 Compensated inner-product features

We perform compensation of the diffraction patterns to achieve better matching of the dictionary to non-anomalous pixels on the sample. Toward this end, as explained above, we use the mean $\bar{\mathbf{y}}$ of the observed patterns as an estimate for the observation background $\mathbf{b}_{\mathbf{Y}}$, and similarly for the dictionary. We then compute the



(a) (b) (c)

Figure 5.9: (a)(b) Background-compensated patterns corresponding to Figs. 5.7(a) and 5.7(b). (c) Best-match dictionary element for both (a) and (b).

background-subtracted patterns $\tilde{\mathbf{y}}_j$ from (5.2), where the scale factor α_j is found to be $\mathbf{y}_j^T \mathbf{b}_\mathbf{Y} / \mathbf{b}_\mathbf{Y}^T \mathbf{b}_\mathbf{Y}$. The same is done for the dictionary to determine $\tilde{\mathbf{m}}_i$.

Figures 5.9(a) and 5.9(b) show the result of removing the background from the experimental patterns in Figs. 5.7(a) and 5.7(b). The diffraction lines that characterize crystal orientation are preserved. Fig. 5.9(c) shows the background-subtracted dictionary element with the largest normalized inner product with the patterns in Fig. 5.9(a) and 5.9(b), which happens to be the same in both cases. The correspondence between diffraction lines suggests that close matches in orientation space can be found by considering background-compensated normalized inner products.

Using this compensation for a pixel j on the sample, we define the normalized inner product values between elements of the dictionary and the normalized EBSD pattern at j:

$$\tilde{\rho}_{ij} = \frac{\tilde{\mathbf{m}}_i^T \tilde{\mathbf{y}}_j}{\|\tilde{\mathbf{m}}_i\| \, \|\tilde{\mathbf{y}}_j\|}, i = 1, \ldots, M. \tag{5.6}$$

In Fig. 5.11, we show the histograms of the inner products $\{\rho_{ij}\}_{i=1}^M$ for two types of patterns found at different locations js on the IN100 sample. Note that the in-grain pattern has values of $\tilde{\rho}_{ij}$ $(i = 1, ..., M)$ that are less concentrated than they are

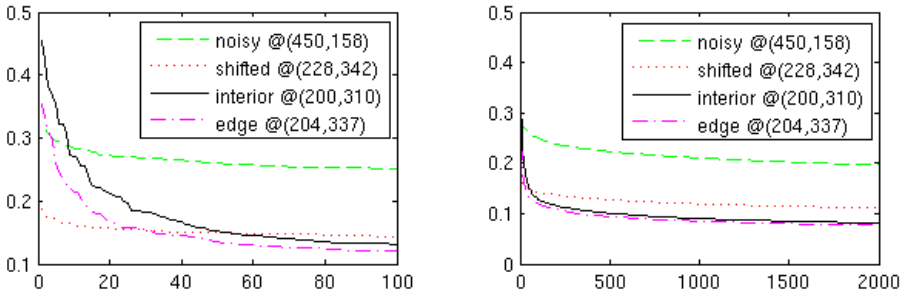
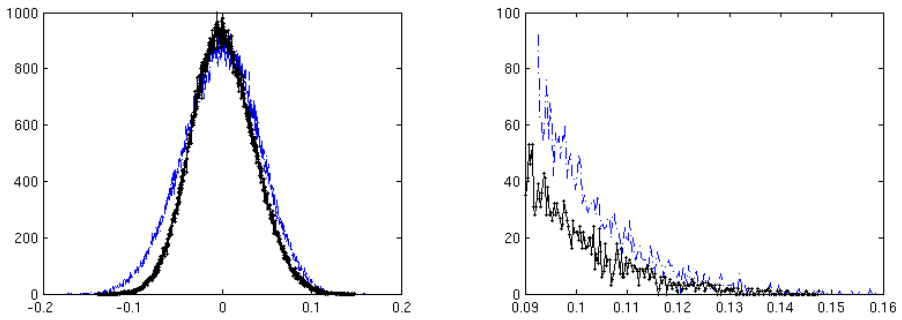


Figure 5.10: Normalized inner products sorted by decreasing absolute value for four representative pixels - interior, boundary, background-shifted, and high noise. (left: the 100 largest values, right: the 2000 largest values) The initial part of the curves indicates that interior and edge pixels are highly correlated with several dictionary entries, after which the curves steeply decrease. After the 100 largest values, the curves corresponding to the interior and edge pixels appear to converge to a low asymptote, whereas the curves for noisy and background-shifted pixels do not decay much.

for a noisy pattern. This is due to the fact that the in-grain sample patterns have better matches to the dictionary than noisy patterns, thus having a thicker right tail (Fig. 5.11(b)). A few large inner product values are in the right tail of the distribution and the effect of this tail is seen in high order statistics, such as skewness and kurtosis of the inner product histogram. We note that the skewness (Fig. 5.12(c)) and kurtosis (Fig. 5.12(d)) images differentiate in-grain and edge pixels, while preserving uniformity within grains. The entropy seems to best delineate the boundaries (Fig. 5.12(f)). The different properties captured by the higher order moments and entropy reflect the fact the distribution of orientations can be quite different over different grains.

5.3.3 Neighborhood similarity for grain segmentation

Grain segmentation is performed based on a neighborhood similarity measure derived from the compensated inner products. The reader will recall that these are the inner products computed after the background components have been removed from both the observations and the dictionary.



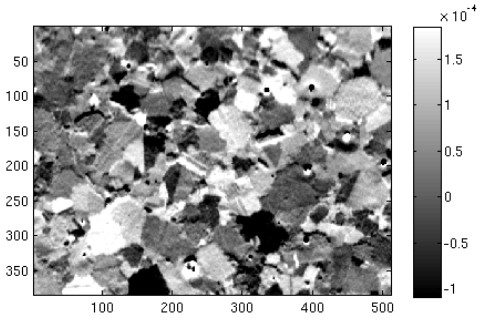
(a) (b)

Figure 5.11: (a) Histograms of normalized inner products $\tilde{\rho}_{\cdot j}$s between compensated patterns for an in-grain IN100 sample location (thin blue) and a sample location at an anomaly in the sample (thick black). The histogram for a boundary pattern overlaps the histogram for an in-grain pattern and is not drawn. The evaluated range of $\tilde{\rho}_{\cdot j}$s is inside [-0.2 0.2]. (b) The right tails of histograms in (a).

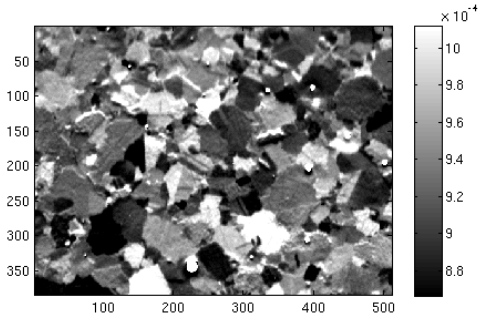
Based on Fig. 5.9 and the model in (5.3), we represent each observed pattern with the dictionary elements having the highest normalized inner products with the observed pattern. Denote by I_j the indices of the dictionary elements chosen for the observation at pixel j; I_j can be regarded as a proxy for the set of significant coefficients a_{ij} in (5.3). It is clear from Fig. 5.9 that more than one representative is required to differentiate between interior patterns (Fig. 5.9(a)) and boundary patterns (Fig. 5.9(b)) since the single best matches can be identical. In addition, using multiple representatives improves robustness against noise and discretization of the orientation space.

To determine the number of dictionary elements required, we examine in Fig. 5.10 the decay of the sorted normalized inner products for the four different types of patterns. Fig. 5.10 suggests that a few tens of dictionary elements are sufficient to capture the nearest neighbors for interior and boundary patterns. For concreteness, we choose the number of representatives $|I_j|$ to be 40 after multiple experiments.

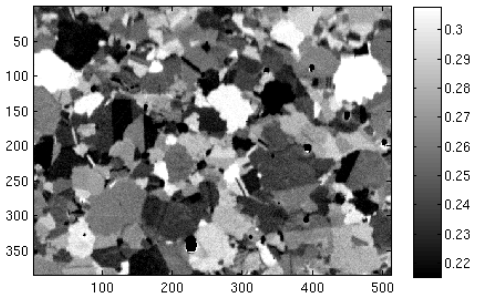
Given the sets I_j, we define a dictionary-based measure of similarity between a pixel j and its neighbors in a 3×3 neighborhood $\mathcal{N}_j$ centered at but not including



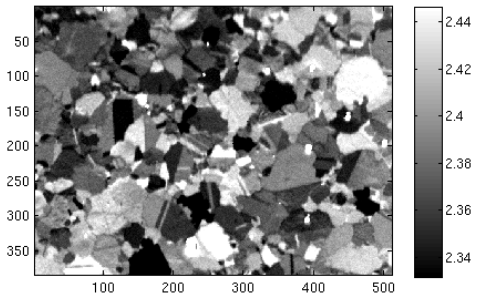
(a) Mean



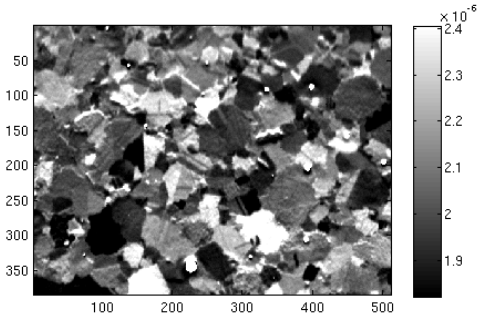
(b) Variance



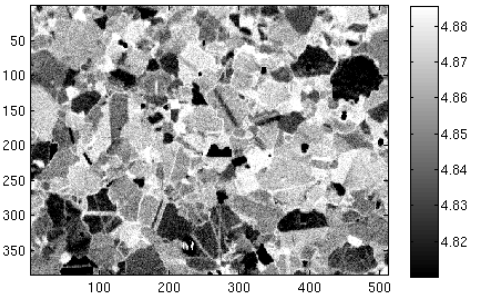
(c) Skewness



(d) Kurtosis



(e) The 4th moment



(f) Entropy

Figure 5.12: Properties of distributions of normalized inner products ($\tilde{\rho}_{\cdot j}$s). (thresholded to show values between 5% and 95% quantiles.)

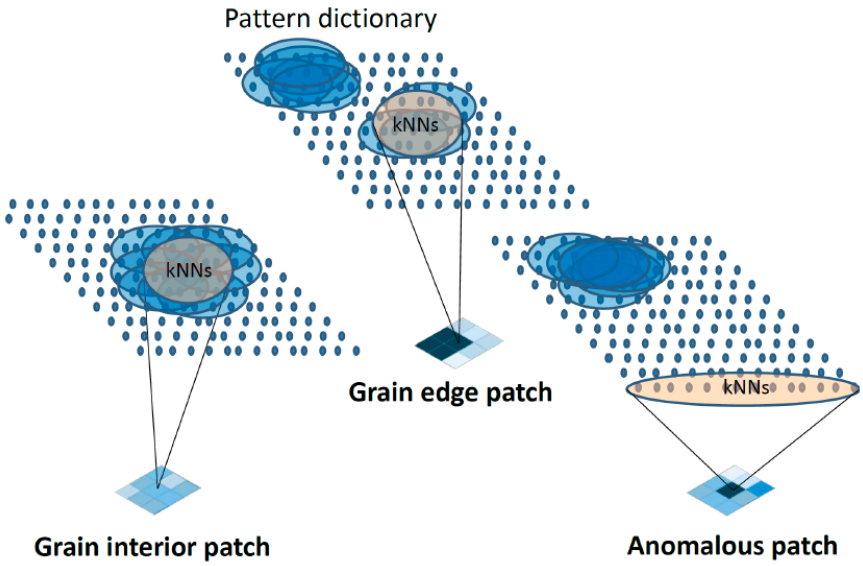


Figure 5.13: Diagram of kNNs (k-nearest neighbors) in the dictionary (elements as dots) using ℓ_2 inner product for grain interior, grain edge, and anomalous patches. Different types of patches would have different degrees of concentration of neighbors.

j. Specifically, the similarity between patterns at pixels j and $k \in \mathcal{N}_j$ is given by $s_{jk} = |I_j \cap I_k|$, and the average similarity of pixel j with its neighbors is

$$s_j = \frac{\sum_{k \in \mathcal{N}_j} s_{jk}}{|\mathcal{N}_j|}. \tag{5.7}$$

The reasoning behind this membership approach is represented in Fig. 5.13. A grain interior patch would have the most concentrated neighbors compared to grain edges and anomalies. A grain edge patch can have overlaps from spatially adjacent grain interior patches, while an anomalous patch can have dispersed neighbors.

5.3.3.1 Gaussian mixture model for discriminating in-grain vs boundary patterns

Using the neighborhood similarity measure, we can differentiate between grain interiors and boundaries. For the IN100 nickel sample studied in Section 5.6, we find that the empirical distribution of the neighborhood similarities resembles a mixture

of two Gaussian densities of different mean and different variance (See Figs. 5.16). Thus, we propose a maximum likelihood approach to separate these components by using an expectation-maximization (EM) algorithm [22]. This estimate will be used to determine an optimal decision threshold on the neighborhood similarity measures for assigning labels of in-grain versus boundary to each pixel.

Let $\mathbf{x} = (x_1, ..., x_n)$ be a sequence of the neighborhood similarity measures where n is the number of observed pixels. We model that this sequence is an i.i.d. sample from a Gaussian mixture density of the form $\tau_1 f(x; \mu_1, \sigma_1^2) + \tau_2 f(x; \mu_2, \sigma_2^2)$ where $\tau_1 = \pi$ ($\tau_2 = 1 - \pi$, $\pi \in [0, 1]$) is the mixture coefficient and $f(x; \mu, \sigma^2)$ is a Gaussian density with mean μ and variance σ^2. An equivalent stochastic representation of the mixture model is obtained by hypothesizing an i.i.d. sequence of latent variables $\mathbf{z} = (z_1, \ldots, z_n)$ which labels each x_i with its (unknown) provenance from either distribution $f(x; \mu_1, \sigma_1^2)$ or from $f(x; \mu_2, \sigma_2^2)$. Specifically, conditioned on $z_i = 1$ and $z_i = 2$, respectively, the distribution of x_i is distributed with one of the above densities. The unknown parameters $\theta = \left(\tau, \mu_1, \mu_2, \sigma_1, \sigma_2\right)$ of the mixture model can be estimated by maximum likelihood estimation. For mixture models, such as the one we are considering here, the expectation-maximization (EM) algorithm can be applied. This algorithm alternates between computing the expectation of a quantity called the "complete data log-likelihood" and maximizing this computed quantity over the unknown parameters.

The complete data likelihood function for the mixture model is the joint distribution of the observed and unobserved variables $\mathbf{x}$ and τ and can be written as:

$$P(\mathbf{x}, \mathbf{z}|\theta) = \prod_{i=1}^{n} \sum_{j=1}^{2} \mathbf{1}(z_i = j)\ \tau_j\ f(x_i; \mu_j, \sigma_j), \tag{5.8}$$

where $\mathbf{1}(\cdot)$ is an indicator function and f is the (assumed) Gaussian probability density function with mean parameter μ_j and variance parameter σ_j, where $j = 1$ for in-grain and $j = 2$ for boundary patterns, respectively.

The E step and M step are given explicitly below by following the derivation in [101]:

E step: From the current estimate $\theta^{(t)}$, Bayes theorem leads to the following update form for the conditional weights:

$$T_{j,i}^{(t)} := \mathrm{P}(Z_i = j | X_i = x_i; \theta^{(t)}) = \frac{\tau_j^{(t)}\ f(x_i; \mu_j^{(t)}, \sigma_j^{(t)})}{\tau_1^{(t)}\ f(x_i; \mu_1^{(t)}, \sigma_1^{(t)}) + \tau_2^{(t)}\ f(x_i; \mu_2^{(t)}, \sigma_2^{(t)})}. \tag{5.9}$$

M step: This step maximizes expectation of log likelihood obtained from the previous E step. The update equations are:

$$\tau_j^{(t+1)} = \frac{\sum_{i=1}^n T_{j,i}^{(t)}}{\sum_{i=1}^n (T_{1,i}^{(t)} + T_{2,i}^{(t)})} = \frac{1}{n} \sum_{i=1}^n T_{j,i}^{(t)}, \tag{5.10}$$

and

$$\mu_1^{(t+1)} = \frac{\sum_{i=1}^n T_{1,i}^{(t)} x_i}{\sum_{i=1}^n T_{1,i}^{(t)}} \text{ and } \sigma_1^{(t+1)} = \frac{\sum_{i=1}^n T_{1,i}^{(t)} (x_i - \mu_1^{(t+1)})(x_i - \mu_1^{(t+1)})^\top}{\sum_{i=1}^n T_{1,i}^{(t)}} \tag{5.11}$$

$$\text{and } \mu_2^{(t+1)} = \frac{\sum_{i=1}^n T_{2,i}^{(t)} x_i}{\sum_{i=1}^n T_{2,i}^{(t)}} \text{ and } \sigma_2^{(t+1)} = \frac{\sum_{i=1}^n T_{2,i}^{(t)} (x_i - \mu_2^{(t+1)})(x_i - \mu_2^{(t+1)})^\top}{\sum_{i=1}^n T_{2,i}^{(t)}}. \tag{5.12}$$

Given ML estimates $\theta = \hat{\theta}$, Bayes optimal decision on whether an observed pattern at pixel j is in-grain ($z_j = 2$) or on the boundary ($z_j = 1$) is the MAP estimator $\hat{z}_j = \arg\max_{z_j=1,2} \mathrm{P}(z_j | x_j; \theta)$. This is equivalent to a Bayes optimal threshold on x_j, which is the solution to the following equation for x:

$$\tau_1 f(x; \mu_1, \sigma_1) = \tau_2 f(x; \mu_2, \sigma_2). \tag{5.13}$$

This equation for the optimal decision threshold is quadratic.

5.3.3.2 Uncertainty quantification of discovered boundaries using p-value

By performing hypothesis testing, we evaluate uncertainty measures for the discovered boundaries. As above, define the 3×3 (9 pixel) spatial neighborhood $\mathcal{N}_j$ of pixel j on the sample. Similarly, for each of the pixels in $\mathcal{N}_j$ let I_i denote the indices of the 40 nearest neighbors of $i \in \mathcal{N}_j$ in the pattern dictionary. Consider testing the hypothesis "H_0: the neighborhood of pixel j is in-grain" versus the hypothesis "H_1: the neighborhood of pixel j is not in-grain" based on the neighborhood similarity sequences for $\mathcal{N}_j$. Given I_i, under H_0 we model the neighborhood overlap ratio $\hat{p} = |I_j \cap I_i|/40$, $i \neq i$, as i.i.d. distributed according to a Binomial distribution with parameters $n = 40$ and $p = p_0$. Here the value p_0 measures the nominal sharing ratio of neighbors between adjacent locations, and can be determined empirically. A test of H_0 vs H_1 of some level α, e.g. $\alpha = 0.01$, compares the empirical observation $n\hat{p}$ to the $1 - \alpha$ quantile of the Binomial cumulative distribution with parameter n, p_0. This one sided test can be used to define a p-value on the observed value of $\hat{p} = |I_i \cap I_j|/40$. Specifically, the p-value for H_0 is defined as the probability under H_0 that the $Binom(40, p_0)$ distributed random variable V is greater than the observed value $\hat{p}$. Small p-values indicate that the pixel neighborhood $\mathcal{N}_j$ is not inside a grain and provides strong evidence to reject H_0.

5.4 Decision Tree

We adopt a binary decision tree (DT) framework [102] to classify pixels into one of 4 classes, thereby accomplishing segmentation and anomaly detection simultaneously. The decision tree first decides on one of 2 classes, anomalous and non-anomalous pixels, and subsequently refines the decision to one of 4 classes: anomalous background shifted, anomalous non-background shifted, non-anomalous in-grain, and non-anomalous boundary. The proposed DT classifier is illustrated in Fig. 5.14. We first

test at node 1 the binary hypothesis that a pixel is an anomaly or not by thresholding the mean uncompensated inner product $\bar{\rho}_j$ in (5.5). If the node 1 test labels the pixel as anomalous then the DT tests at node 2 whether the pixel is a background-shifted anomaly or not using a second threshold on $\bar{\rho}_j$. The shift in background makes the value of $\bar{\rho}_j$ significantly lower than for the other anomaly. Finally if the node 1 test assigns a label of non-anomalous pixel, the DT tests whether the pixel is an in-grain vs a boundary pixel at node 3 using the neighborhood intersection similarity measure (5.7), which is computed from background-compensated diffraction patterns. Interior pixels have high similarity values because they share many common neighbors in the dictionary with adjacent pixels. On the other hand, boundary pixels can be viewed as mixtures of adjacent interior pixels and therefore have lower similarities because the set of nearest neighbors is only partially shared with those on either side of the boundary.

The only tuning parameters in our DT classifier are the decision thresholds on the similarity measures used at each node of the tree. When there is ground truth data available these thresholds can be selected by cross-validated minimization of nodal deviance measures such as mis-classification rate, Gini index, or entropy [103]. Parameter selection is generally more difficult however when the learning is unsupervised as in our case. Our application of DT classifiers differs from previous imaging applications, e.g., land cover classification in remote sensing [104, 105], in several important ways. First, our classifier is a hybrid DT in that it uses different features (background-compensated and uncompensated patterns) and similarity measures (inner products and neighborhood intersections). Second, unlike standard non-parametric DT classifiers, our DT is informed by a physics model through the generated dictionary of diffraction patterns. As a consequence, the tree structure of our DT classifier is fixed and need not be learned from the data.

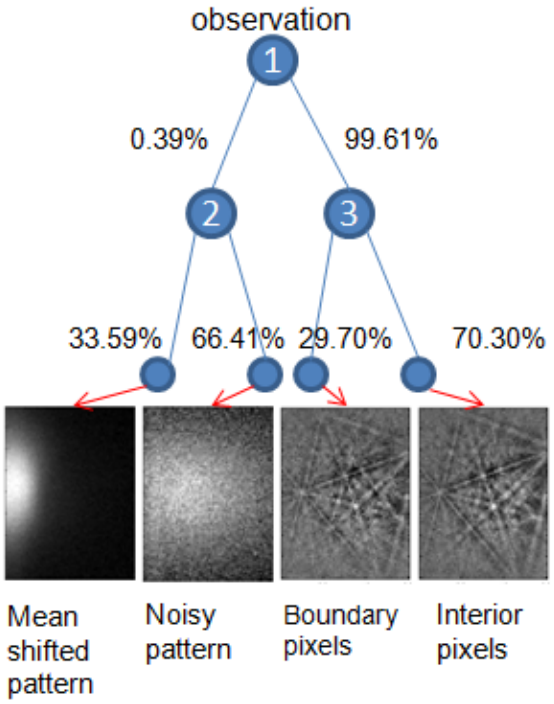


Figure 5.14: Decision tree. The background similarity criterion (5.5) is used at nodes 1 and 2 while the neighborhood similarity criterion (5.7) is used at node 3. The division of the population at each parent node is shown above the branches. Representative diffraction patterns are also shown.

5.5 Super-resolution of Grain Boundaries

Once the DT has labeled the boundary pixels the neighborhood model developed in Section 5.3.3 can be used to obtain a refined estimate of the boundary locations. These boundary pixels can be expected to be close to patterns that are an amalgam of the diffraction patterns in the grains that share the boundary. We propose to evaluate the variation of the dictionary neighborhood of a pixel as it translates along a line normal to the boundary (see Fig. 5.15). We constrain the following discussion to 2-grain boundaries - boundaries that separate two grains, called respectively G_1 and G_2. The approach described below can be extended to tri-grain or higher order boundaries. We define the pairwise consistency measure for the pixel j as it translates along the line normal to a 2-grain boundary, as follows:

$$c_1(j) = |\mathcal{N}_{G_1} \cap I_j|, \tag{5.14}$$

$$c_2(j) = |\mathcal{N}_{G_2} \cap I_j|, \tag{5.15}$$

where $\mathcal{N}_{G_1}$ is the indices of the 40 most common dictionary patterns occurring in the different kNN neighborhoods ($k = 40$) found by scanning pixels within grain G_1. $\mathcal{N}_{G_2}$ is defined similarly by scanning grain G_2.

A refinement of the location of the edge is found by solving for the 'crossing' point of the two curves c_1 and c_2 (See the bottom figure in Fig. 5.15). Thus, along the normal vector for each edge, we evaluate the following super-resolved edge location by linearly interpolating the curves up to the super-resolved pixel resolution.

$$\text{argmin}_j |c_1(j) - c_2(j)|, \text{ or} \tag{5.16}$$

$$\text{argmax}_j |c_1(j) \times c_2(j)|, \tag{5.17}$$

where j is in the super-resolved grid. We present results in the next section using two

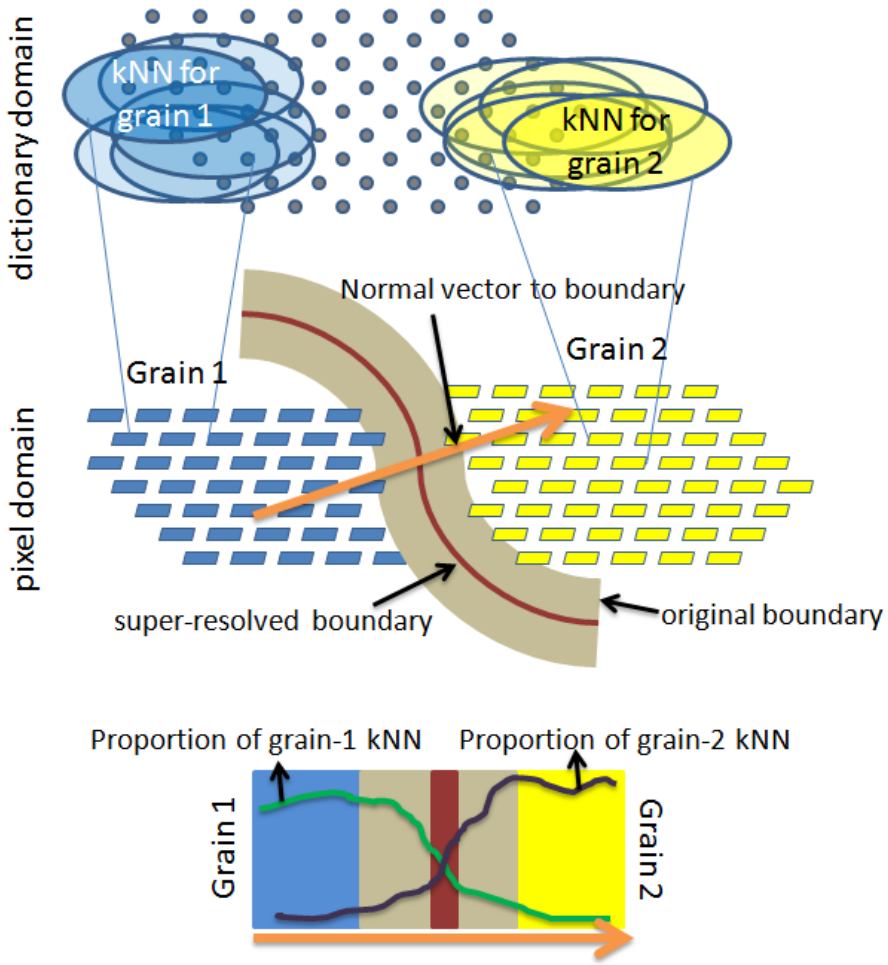


Figure 5.15: Diagram for the super-resolution method using common dictionary kNN neighborhoods of two different grains. A refined estimate of the boundary is obtained by evaluating the crossing point of the pairwise consistency measures (5.14) and (5.15) in bottom figure.

versions of the super-resolution algorithm. One uses simple vertical and horizontal scans, ignoring true normal directions to edges. The other version approximates the normal directions and the detailed algorithm is presented in Appendix 2. The first simple algorithm is a special case of the second algorithm.

5.6 Results

5.6.1 Application to IN100 data

Here we apply the method described in Section 5.3 to EBSD data obtained from a section of the Ni-base alloy IN100. In the next section we apply it to a different (LSHR) sample. The dimensions of the sample slice are 512 x 384 pixels. The diffraction pattern corresponding to each pixel in the sample is 80 x 60 pixels in size.

For the IN100 sample, ground truth class labels are not available. Thus, to set the thresholds in the DT classifier in Fig. 5.14, we consider the empirical distributions (obtained from the experimental sample) of the background similarity $\bar{\rho}_j$ (5.5) and the neighborhood similarity s_j (5.7). The distribution of $\bar{\rho}_j$ can be divided into three parts by inspection: a uniform distribution with a small number of samples over $[\rho_{\min}, t_1]$, an increasing tail over $[t_1, t_2]$, and a large peak over $[t_2, 1]$. We choose the first threshold $t_1 = 0.815$ to separate background-shifted patterns at node 1. From Fig. 5.16(a), which shows the histogram of $\bar{\rho}_j$ between 0.99 and 1, we select $t_2 = 0.996$ as the threshold to classify high-noise patterns at node 2. Lastly, at node 3 we use $|I_j| = 40$ dictionary representatives and consider the empirical histogram of the neighborhood similarity s_j shown in Fig. 5.16(b). We set the threshold at the 0.3 quantile at 31.25, which corresponds to the knee in the histogram. The percentages in Fig. 5.14 indicate the division of the parent population at each node resulting from the above thresholds. Note that the percentages below node 3 are not 30% and 70% exactly because the neighborhood similarities are quantized. Choosing a threshold by fitting a 2-component Gaussian mixture model (using the EM algorithm in Section 5.3.3.1) to Fig. 5.16(b) yields a very similar value (see Fig. 5.16(c)).

In Fig. 5.17(a), we show a neighborhood similarity map computed according to (5.7) for $|I_j| = 40$. As expected, interior pixels have larger similarity values while boundary pixels and anomalies have smaller similarities since they are less likely to

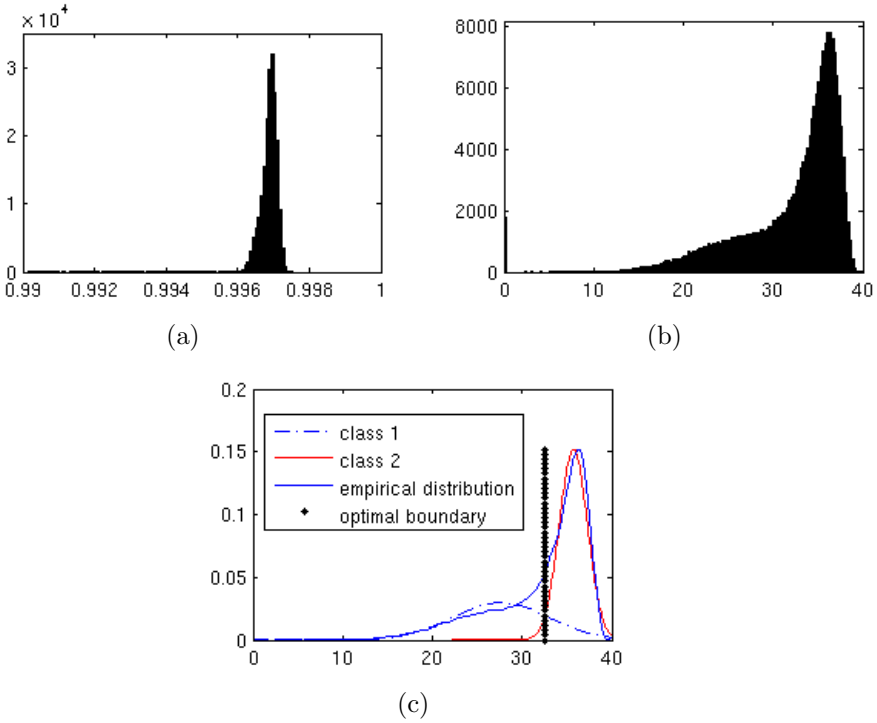


Figure 5.16: IN100 data. (a) Empirical histogram of the background similarity measure $\bar{\rho}_j$ between 0.99 and 1. The threshold t_2 at node 2 in Fig. 5.14 is set to 0.996. (b) Empirical histogram of the neighborhood similarity measure s_j. The threshold at node 3 in Fig. 5.14 is set to the 0.3 quantile at 31.25. (c) The Bayes optimal decision boundary as an optimal threshold level is 32.54 under the Gaussian mixture model.

share common neighbors in the dictionary with spatially adjacent pixels. Similarity maps evaluated with different numbers of nearest neighbors, $|I_j| = 20$, 40, 200, and 2000, yield equally good performance visually (not shown). For comparison, we show a similarity map in Fig. 5.17(b) computed[4] using only the observed patterns, specifically by averaging the uncompensated inner products between a pattern at a given pixel and patterns at adjacent pixels. Comparing Figs. 5.17(b) and 5.17(a), it can be seen that the map in Fig. 5.17(b) has blurrier edges near the upper boundary of the sample[5]. This suggests that the dictionary-based approach has a denoising effect that results in sharper segmentation.

Fig. 5.17(c) shows the segmented image produced by our DT classifier using the thresholds specified above. In terms of segmenting grains, our result is consistent with the image in Fig. 5.17(d) from a conventional segmentation algorithm that uses Euler angles, specifically as implemented in the software DREAM.3D [106]. However, the proposed method also identifies anomalies (colored red and blue in Fig. 5.17(c)) that are either not detected (black clusters) or misclassified as one or more grains in Fig. 5.17(d). This is due to the fact that the conventional segmentation algorithm is not aware of anomalies and has difficulty extracting meaningful Euler angles from anomalous diffraction patterns, as indicated by the black clusters.

As additional evidence for the discovered boundaries, we evaluated p-values under our neighborhood model following the procedure in Section 5.3.3.2. The null hypothesis H_0 is as described in that section with the parameter value p_0 in the null binomial distribution set to 0.8. For IN100 data, visual inspection and the EM algorithm both estimate the ratio of approximately 0.8 ($40 \times 0.8 = 32$, which is close to the threshold we use.) as the optimal decision threshold level (Fig. 5.18). The derived p-value map for the null hypothesis H_0: grain is provided in Fig. 5.18, which is based on a binomial

[4]This quantity is defined as $\frac{\sum_{k \in \mathcal{N}_j} \rho^a_{jk}}{|\mathcal{N}_j|}$, where $\rho^a_{jk} = \frac{\mathbf{y}_j^T \mathbf{y}_k}{\|\mathbf{y}_j\| \|\mathbf{y}_k\|}$.

[5]These blurry edges are also found in a similarity map computed using only 'compensated' observed patterns.

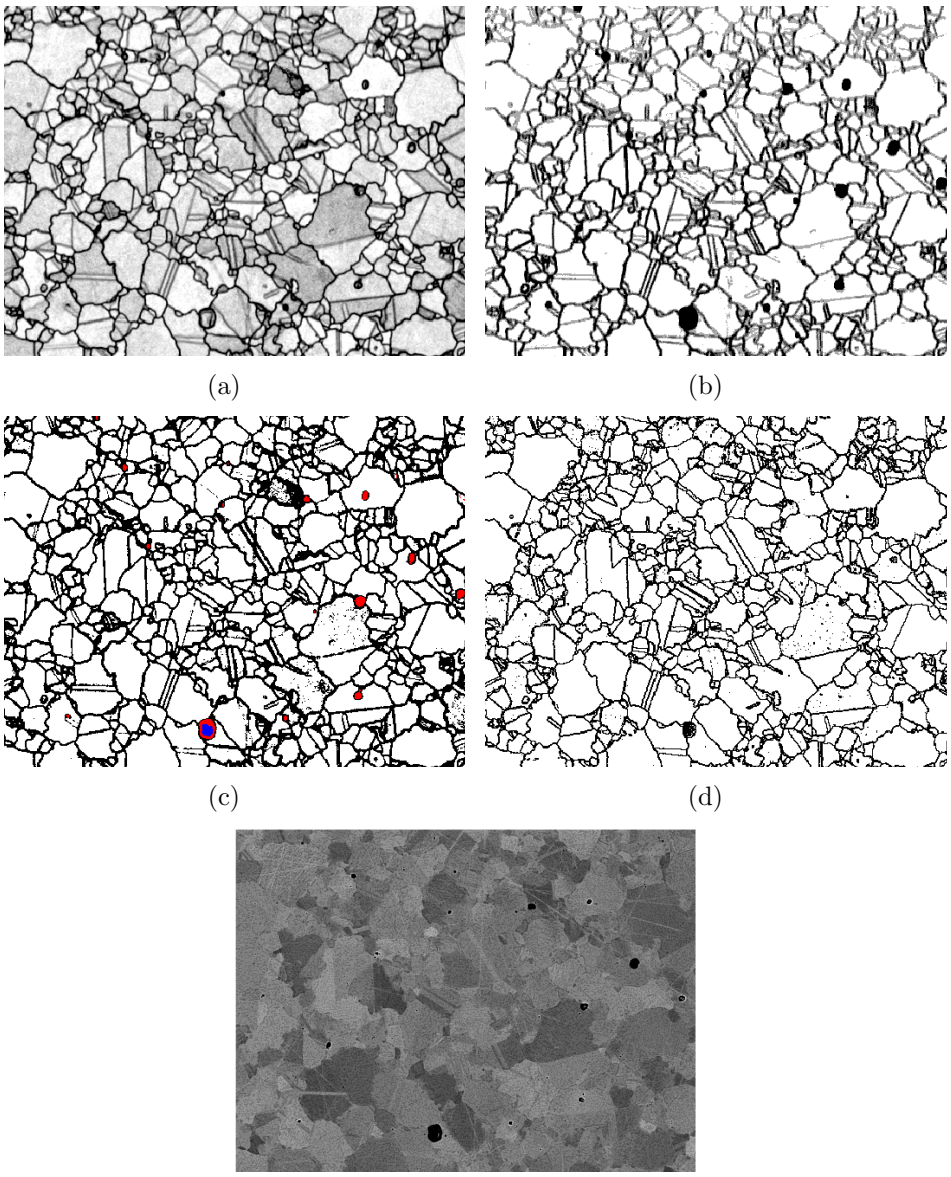


Figure 5.17: IN100 data. (a) Neighborhood similarity map with $|I_j| = 40$. The grayscale runs between the 0.05 and 0.95 quantiles of the similarity values to suppress the visual effect of outliers. (b) Similarity map derived from observed patterns only. The upper parts of the sample are blurry and the boundary structure is not as clear as in the dictionary-based map (a). (c) Segmentation result from our decision tree in Fig. 5.14. Grain interiors in white, boundaries in black, noisy pixels in red, and background-shifted pixels in blue. (d) Segmentation result from DREAM.3D. Black clusters represent pixels that cannot be classified. (a horizontal white line located in the lower left side, visible in other figures, is missing here.) (e) back-scattered electron (BSE) image as a near ground truth from a different high resolution microscope.

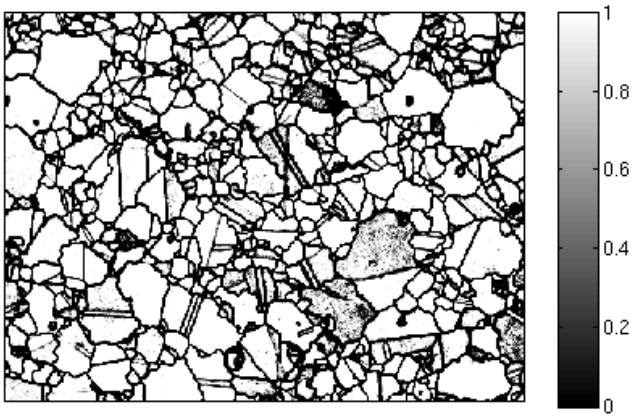


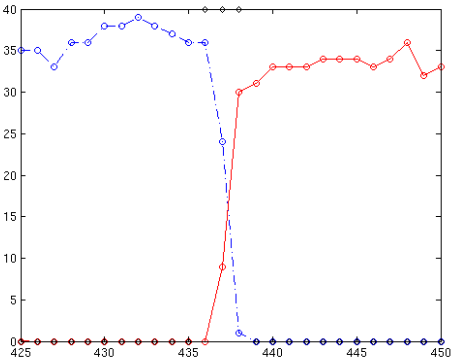
Figure 5.18: For IN100 data, p-values as empirical evidence for edge. H_0: grain. the average sharing ratio of dictionary elements is 0.8 among spatial neighbors.

CDF. We conclude that we have discovered strong evidence for the segmented edges.

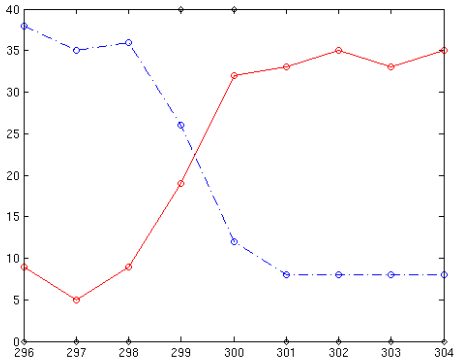
Furthermore, we implemented the super-resolution method in Section 5.5 for refining the boundary location estimates. The refined location of an edge lies near the contact point of the two curves $c_1(j), c_2(j)$ swept out over a line normal to the edge. The resulting super-resolved edges (Fig. 5.20(b)), from vertical and horizontal scans not a normal direction scan, are more highly resolved than the original (Fig. 5.20(a)) after super-resolution processing. Several other examples are illustrated (Fig. 5.21). This simple scanning of x and y directions (horizontally and vertically, called 'xy scanning') is a special case of Algo. 6) in Appendix 2. Along a scanned direction, we evaluate pairwise-consistency curves such as in Fig. 5.19 and find the 'crossing' coordinate where the two curves meet. Rather than marking one pixel[6] that has the minimum difference between two curves, we mark multiple pixels to prevent disconnectivity in some areas (e.g. blobs or blurry parts).

The super-resolution results by scanning more extensive directions following Algo. 6 in Appendix 2 are provided in Fig. 5.22. This algorithm uses several more directions

[6]For rendering, the pixel location is quantized.



(a) Between two different grains, scanned from (425,90) to (490,90).



(b) Between two similar grains, scanned from (100,296) to (100,304).

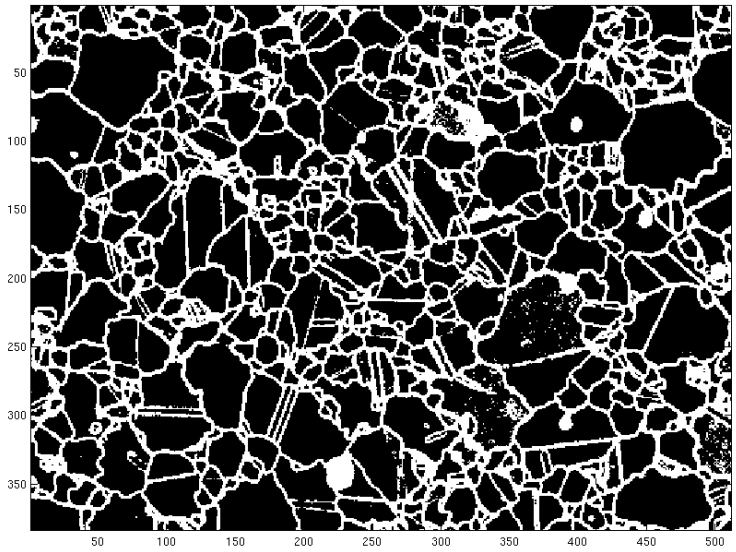
Figure 5.19: Pairwise comparison of the common membership. A blue dash curve indicates the number of common members along the y-axis between a point at the coordinate in the graph and the grain at the left end point. A red solid curve indicates the number of common members along the y-axis between a point at the coordinate in the graph and the grain at the right end point. The black diamond dots are boundary labels; if the value of a boundary label is non-zero, it is a boundary.

than the two directions (x and y directions) to find an approximate normal direction to an edge. We used 12 directions with an angular resolution of 15 degree. The detailed partial views from Fig. 5.22 are illustrated in Fig. 5.23. The results in Fig. 5.22 seem outperform the results from simple xy scanning in terms of the less noisy edge locations (Fig. 5.23).

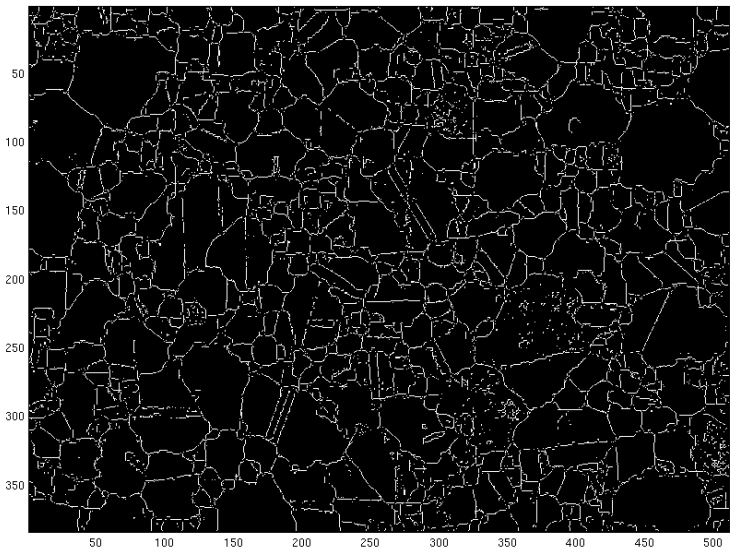
In Fig. 5.24, using the proposed super-resolution method we present several consistency curves for thin, thick, and asymmetric edges. These show different grain mantle changes. The investigation of grain mantle and boundary properties are reserved for future work, which requires estimation of grain orientations.

5.6.2 Application to LSHR data

To evaluate the robustness of the dictionary to model mismatch, we applied our methods to the LSHR data set. This dataset was collected from a different microscope at AFRL than the one used to acquire the IN100 dataset. Recall that the dictionary

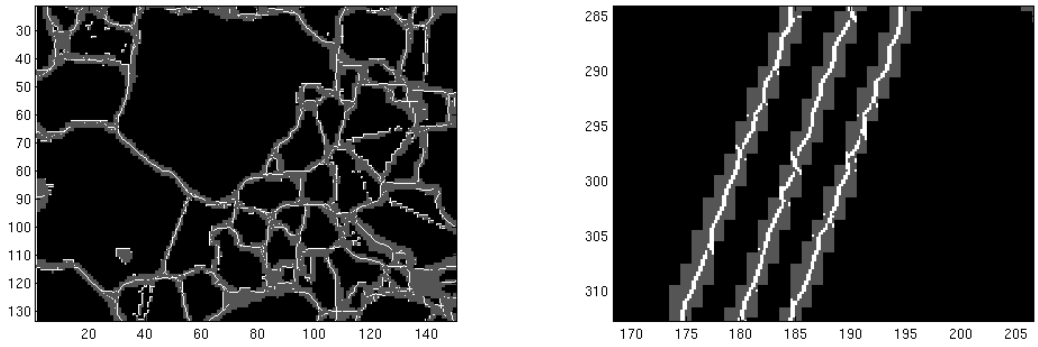


(a) Original edges



(b) Super-resolved edges

Figure 5.20: The original (a) and (5x) super-resolved (b) edges for IN100 by using xy scanning. Thin connected edges look disconnected due to the rendering effect (the finite resolution in the downsized image to fit the size of the paper). For connectivity map, see Fig. 5.21.



(a) (b)

Figure 5.21: Partial views of Fig. 5.20 with super-resolved edges in white and original edge in gray. The boundary pixels are painted at the 5x higher resolution.

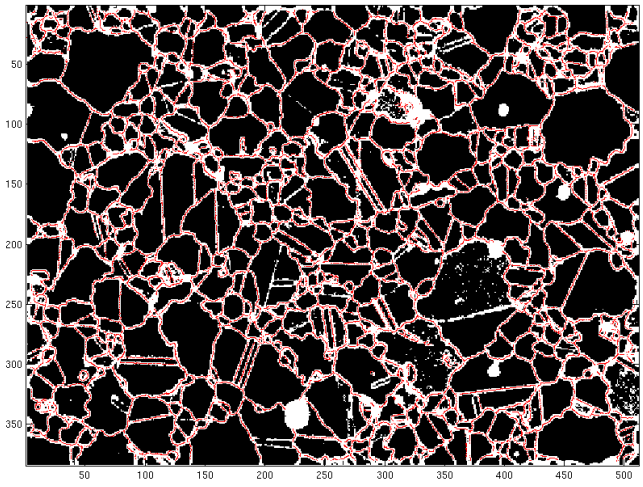
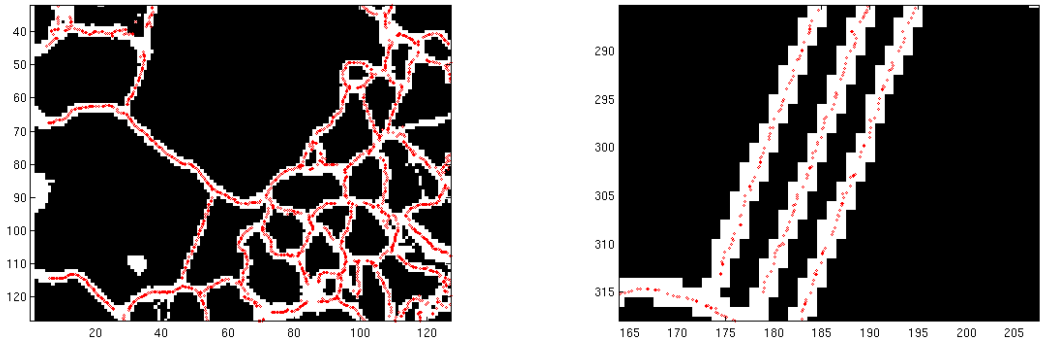


Figure 5.22: For IN100 data, super-resolution result by using Algo. 6 in Appendix 2.



(a) (b)

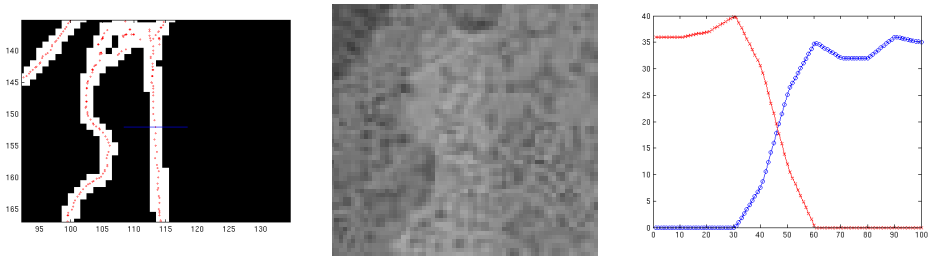
Figure 5.23: Partial views of Fig. 5.22 with super-resolved edges in red dots and original edge in gray. The refined boundary 'locations' are marked by red circles at higher resolution than the resolution of the 'painted' boundary pixels in Fig. 5.20 and 5.21.

was generated by Marc DeGraef specifically tuned to the IN100 sample and the sample acquisition parameters of the IN100 microscope. We conclude that the results shown here (see Fig. 5.25) indicate a strong degree of robustness of the dictionary to model mismatch.

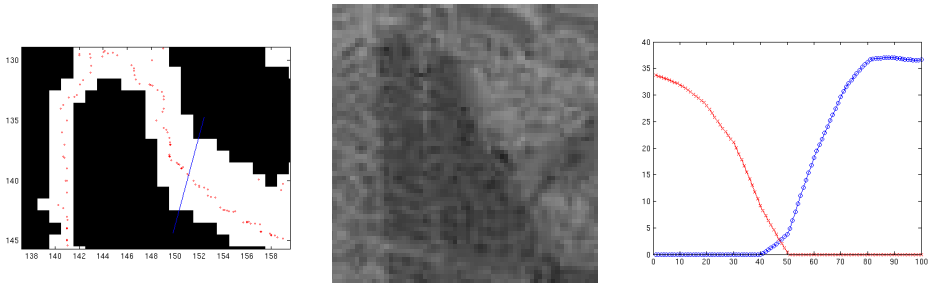
Fig. 5.26(a) presents the empirical histogram of the background similarity measures. Most values and the peak locations are less than those from IN100 in Fig. 5.16(a). This is because LSHR diffraction patterns do not perfectly match the dictionary developed for IN100. However, there is no significant outlier, unlike the IN100 data, and there are few outliers in the distribution of the background similarity measures of the LSHR data. This is again demonstrated in the decision map in Fig. 5.27(c). This is consistent with the higher quality and more exacting standard with which the LSHR sample was produced.

The threshold value for boundary classification is 28.63, based on visual segmentation (35% percentile value in Fig. 5.26(b)). This value is close to the value 29.52, which is derived from a Gaussian mixture model using the EM algorithm (Fig. 5.26(c)).

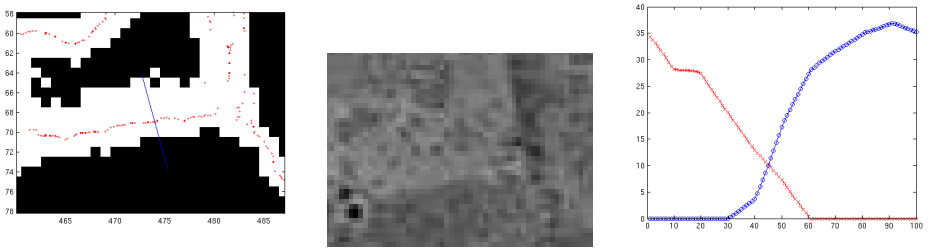
In Fig. 5.27(a), we show a neighborhood similarity map computed according to



(a) Thin edge (b) BSE image (c) Consistency curve for (a)

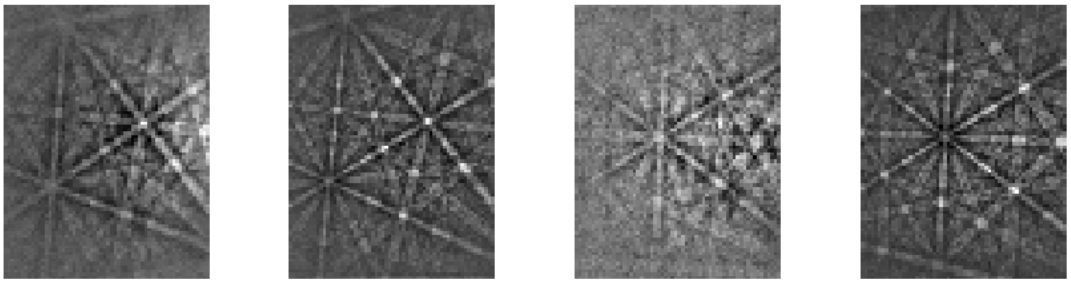


(d) Thick edge (e) BSE image (f) Consistency curve for (d)



(g) Asymmetric edge (h) BSE image (i) Consistency curve for (g)

Figure 5.24: Several examples of different grain mantle transitions.

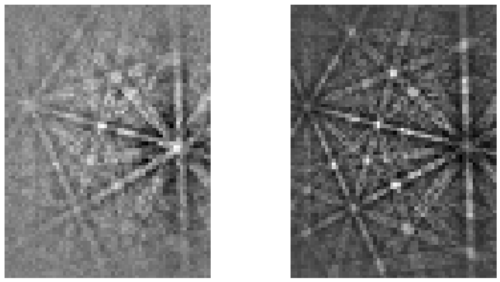


(a) observation at (460,380)

(b) best match for (a)

(c) observation at (336,66)

(d) best match for (c)



(e) observation at (364,316)

(f) best match for (e)

Figure 5.25: Three representative patterns from the LSHR sample collected with a different microscope at AFRL than that for which the dictionary was designed and the best matches in the dictionary, designed for the microscope that acquired the IN100 sample. The observation patterns are of good match to the dictionary patterns.

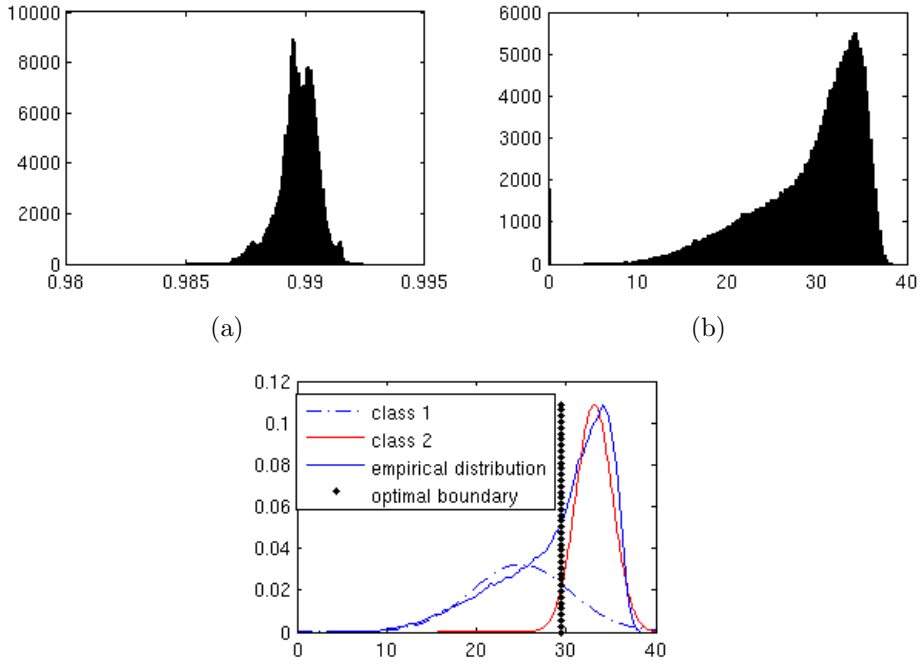


Figure 5.26: LSHR data. (a) Empirical histogram of the background similarity measure $\bar{\rho}_j$. The threshold t_2 is set to 0.9865. (b) Empirical histogram of the neighborhood similarity measure s_j. The threshold for boundary segmentation is set to the 0.35 quantile at 28.63. (c) The Bayes optimal decision boundary as an optimal threshold level is 29.52, under the Gaussian mixture model.

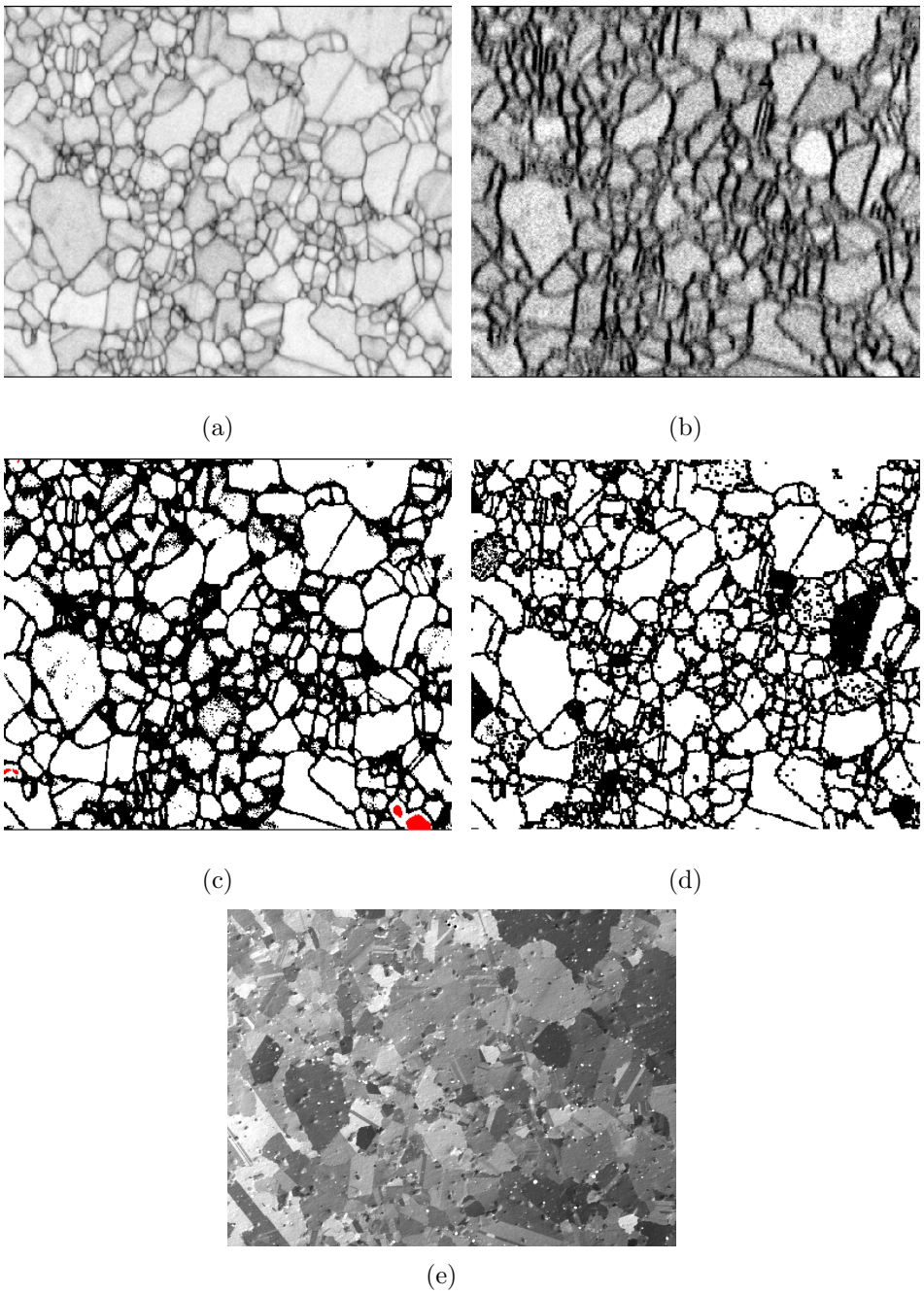


Figure 5.27: LSHR data. (a) Neighborhood similarity map with $|I_j| = 40$. (b) Similarity map derived from observed patterns only. Most parts of the sample are blurry and the boundary structure is not as clear as in the dictionary-based map (a). For visualization, 5% - 95% quantile values are used for thresholding. (c) Segmentation result from our decision rules. Grain interiors in white, boundaries in black, anomalous pixels in red. (d) Segmentation result using OEM estimated Euler angles. Black colors denote where Euler angles are different from those of neighboring pixels by at least 5 degree for each Euler angle component. Some black chunks inside grains are mis-classified as edges. (e) back-scattered electron (BSE) image

(5.7) for $|I_j| = 40$. As previously demonstrated with the IN100 results, interior pixels for LSHR have larger similarity values while boundary pixels and anomalies have smaller similarities. This map is compared to the autocorrelation map (Fig. 5.27(b)) using only the observed patterns. Comparing these two figures, we again clearly demonstrate that our proposed similarity measures produce a denoised and deblurred boundary structure. Fig. 5.27(c) illustrates the segmented image produced by our DT classifier using the thresholds specified above. This result is consistent with the images in Fig. 5.27(d), obtained by thresholding norm differences between Euler angles (EBSD machine outputs), and in Fig. 5.27(e), from a back-scattered electron microscope. Again, the benefit of the dictionary DT approach is that it identifies anomalies that are undetected, misclassified, or unclassified using conventional EBSD methods.

We evaluated p-values as evidence for edges by following the procedure in Section 5.3.3.2. The null hypothesis of grain or not-grain assumes a binomial parameter value to be 0.7, because in the LSHR data, visual inspection and EM both estimate the ratio of approximately 0.7 ($40 \times 0.7 = 28$. This is close to the threshold we use.) as the optimal decision threshold level (Fig. 5.28). The p-value map for the null hypothesis H_0: grain is provided in Fig. 5.28, and its computation is based on a binomial CDF. We again conclude that we have discovered strong evidence for the found edges.

The super-resolution results by using a simple xy scanning in Section 5.5 to refine the boundary locations are presented in Fig. 5.29. Several examples as partial magnified views from Fig. 5.29 are shown in Fig. 5.30. Our algorithm in Algo. 6 produces the super-resolution results in Fig. 5.31 by using more extensive scanning of line angles. The detailed partial views from Fig. 5.31 are illustrated in Fig. 5.32.

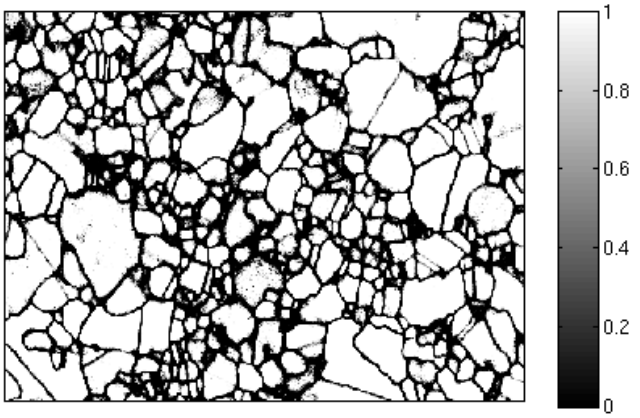


Figure 5.28: For LSHR data, p-values as empirical evidence for edge based on a binomial CDF. H_0: grain. the average sharing ratio of dictionary elements is 0.7 among spatial neighbors.

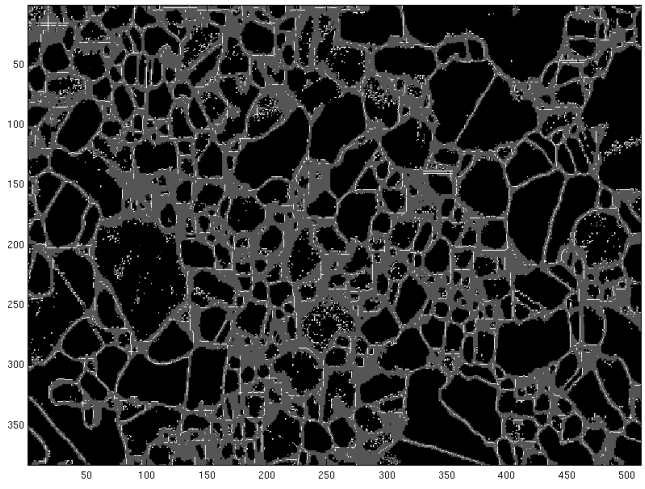
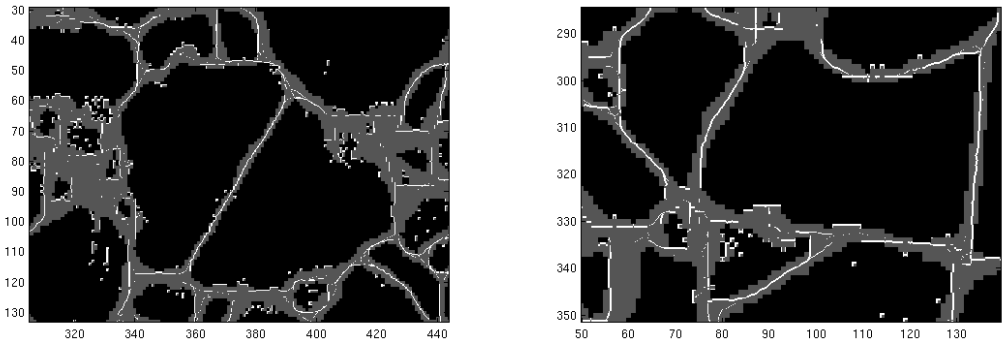


Figure 5.29: For LSHR data, (5x) super-resolved edges by using xy scanning with super-resolved edges in white and original edge in gray.



(a) (b)

Figure 5.30: Partial views of Fig. 5.29 with super-resolved edges in white and original edge in gray.

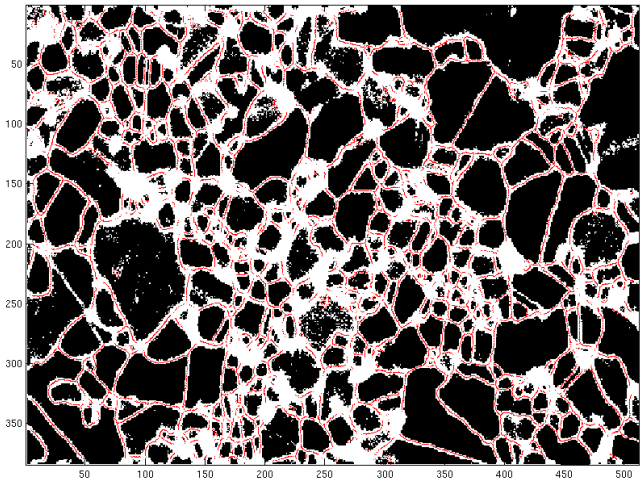
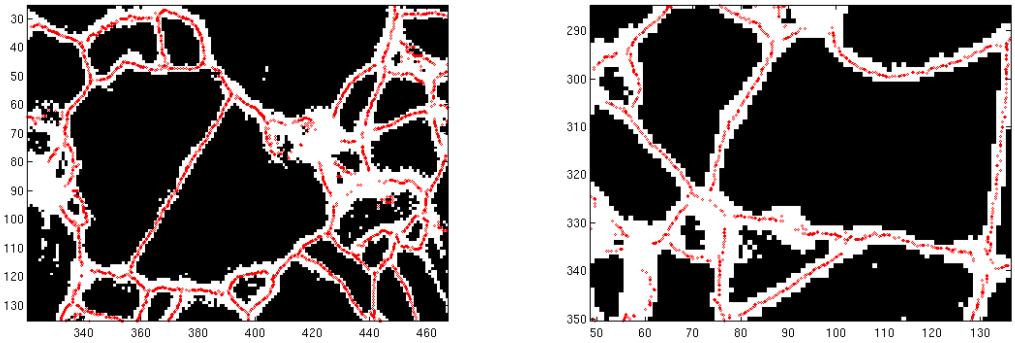


Figure 5.31: For LSHR data, super-resolution result by using Algo. 6.



(a) (b)

Figure 5.32: Partial views of Fig. 5.31 with super-resolved edges in red dots and original edge in gray.

5.7 Conclusion and Future Work

5.7.1 Conclusion

We have proposed a simultaneous classification and anomaly detection method for EBSD images. This is novel in its use of full diffraction patterns instead of Euler angles or equivalent representations of orientation. Our work is the first to exploit this dictionary of reference diffraction patterns through a physics-based forward model. Our decision tree classifier uses features that measure similarity between observed patterns and dictionary elements and is efficient enough to handle large sets of high-dimensional patterns. The method succeeds in not only segmenting grain interiors from boundaries but also discovering and identifying two types of anomalies. The edge enhancement is another benefit gained from using both a neighborhood model and the physics-based forward model. The statistical significance of edges can be gauged by p-values using hypothesis testing and can be more precisely located through super-resolution.

5.7.2 Future work

An important issue for future work is to reduce computation time required for implementation of the dictionary approach. The search of the dictionary for kNN matches requires computing and evaluating inner products between the dictionary patterns and patterns observed at every sample pixel. This takes several hours to complete. A good initialization, such as OEM estimated Euler angles[7], is a worthwhile direction for investigating speed-up. To accomplish this, however, the OEM angles would need to be mapped to the dictionary angles, requiring interpolation and a consistently defined fundamental zone. Another direction is to develop methods explicitly designed for estimating Euler angles at the grain level and to estimate misorientation. The former requires a consistent definition of fundamental zone of Euler angles. A potential improvement of the dictionary is to make the sampling more uniform over Euler angles (see Appendix 5.8).

Estimation of grain-wise orientations requires a pixel-level estimation and collection of the statistics of constituent pixels of a grain. The pixel-level estimates can be obtained after evaluating the weight factors for individual dictionary elements, which are sparse in the grain interior patterns. Thus, classical sparse estimation methods, combined with the suitable regularization on weights, would produce desired solutions. The grain-wise orientations together with grain shapes can be used to characterize different grain and boundary properties such as misorientations and twin[8] grain/boundary areas.

[7]These angles are outputs from an EBSD scanner combined with its proprietary estimation algorithm.

[8]Twin grains interfacing each other have a mirrored symmetry in terms of their crystal orientations.

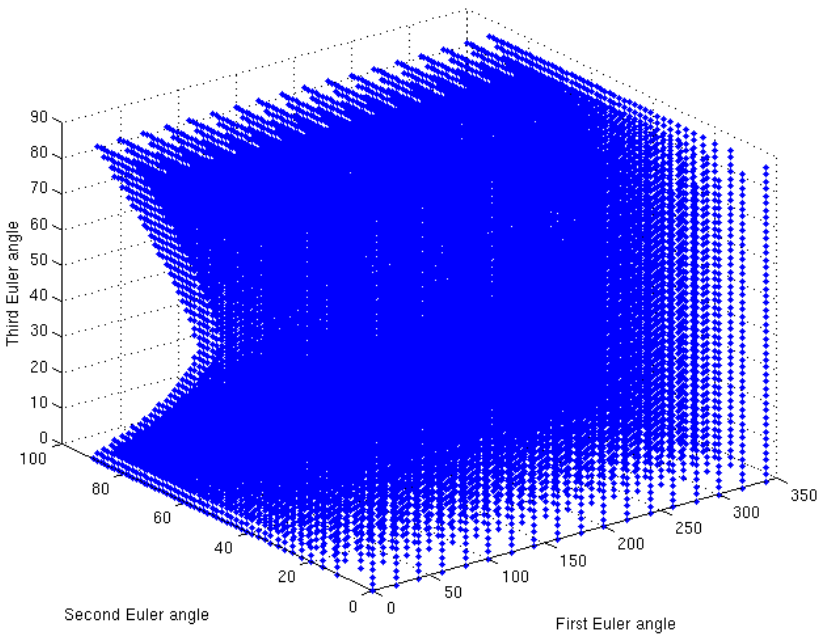


Figure 5.33: Euler angles sampled in the dictionary. The first Euler angle values are downsampled by a factor of 10.

5.8 Appendix 1: Discussion on Uniformity in the Dictionary

Each element in the dictionary we are using is a diffraction pattern that corresponds to triplet Euler angles (ϕ_1, Φ, ϕ_2). We test the uniformity of these angles and image patterns. The sampling schedule along triplet Euler angles is represented in Fig. 5.33. Note that along the second Euler angle, samples in the mid range (in the upper left side of the Fig. 5.33) are not drawn, due to possible degeneracy of the Euler angle[9] [5, 59, 90]. To better understand this non-uniform sampling scheduling within the Euler angle space, we further investigated uniformity in the dictionary, using other representations of the dictionary elements.

We first investigated the image space using an inner product metric. We randomly selected 4000 out of 281,700 samples and evaluated their individual ℓ_2 normalized inner product values with all other image samples, as seen in Fig. 5.34(a), where

[9]Degenerate Euler angles do not have a unique representation

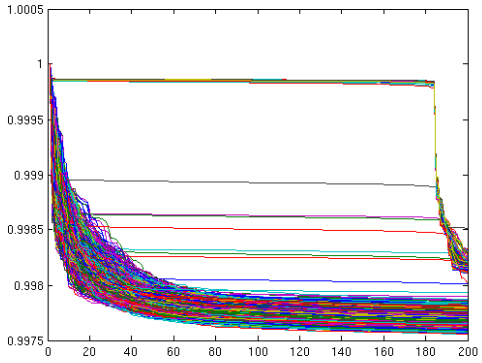
the inner product values for only the top 200 values out of 281,700 are shown. The inner product curves show that some of the diffraction patterns are not uniformly sampled but rather densely sampled. The upper curves have larger values (close to 1) than other curves for the index of approximately 180 along the x-axis in Fig. 5.34(a). Here, each curve corresponds to one of the 4000 selected patterns. In other words, the selected dictionary elements for these upper curves have closer neighbors than the others and correspond to the samples whose second Euler angles are equal or close to zero. These angles are close to the degenerate Euler angle values. We validated this finding by running the experiment multiple times.

Another uniformity test is performed in the quaternion space. Since Euler angles can be converted into unit quaternion vectors up to their signs, we checked whether these quaternion vectors in a four dimensional space are uniformly distributed on the three dimensional spherical manifold. Specifically, under cubic symmetry, one triplet Euler angle has 24 equivalent Euler angles and 48 unit quaternion vectors [90]. This factor of 2 is due to the additional vectors that are the negative version of the 24 transformed vectors. We performed a uniformity test on this axial data of quaternion vectors by randomly selecting 4000 triplet Euler angles sampled in the dictionary and transforming them to (4000×48) unit quaternion vectors. We performed the Bingham test, which is a simple test of uniformity in the axial data $\{\mathbf{x}_i\}$. The reasoning behind this test is that uniformity is rejected if the sample scatter matrix

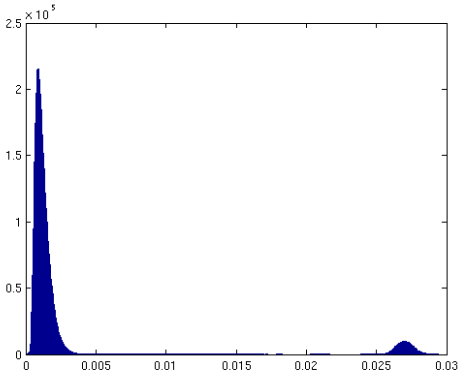
$$\mathbf{T} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

is far from its expected value $p^{-1}\mathbf{I}_p$ with $p = 4$ for the unit quaternion vectors [107]. The derived test statistic is

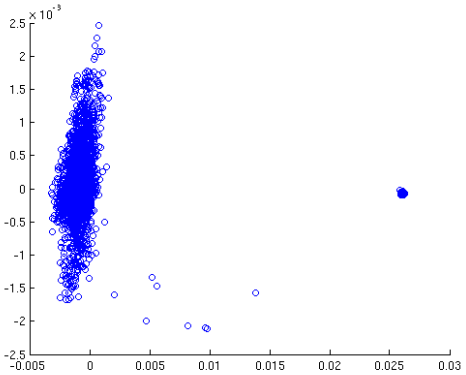
$$S = \frac{p(p+2)}{2} n(\mathrm{tr}(\mathbf{T}^2) - p^{-1}). \tag{5.18}$$



(a) The sorted inner product values for 4000 randomly selected patterns from the dictionary. Each curve, corresponding to one of the 4000 patterns, consists of the top 200 values from all the inner product values between a selected pattern and all other patterns in the dictionary.

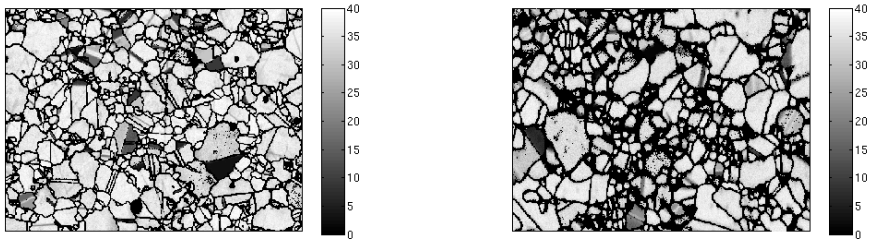


(b) The Euclidean distances among the selected 4000 curves shown in (a).



(c) MDS (multidimesional scaling) result of curves in (a)

Figure 5.34: Uniformity check of the dictionary patterns.



(a) In-grain consistency map for IN100 (b) In-grain consistency map for LSHR

Figure 5.35: Maps of consistency measures for in-grain pixels: $c_i(j)$ is similarly defined as (5.14) but with j is in the $i$the grain.

The uniformity is rejected for large values of S. The distribution of S, under uniformity, is approximately $\chi^2_{(p-1)(p+2)/2}$ with an error on the order of $O(n^{-1})$, where χ^2_m is a chi-squared distribution with the degree of freedom of m. The value of S in our data is greater than 10^5, thus the p-value is significant, and the test rejects the hypothesis that data is uniformly distributed on the sphere.

Therefore, the dictionary is not uniformly sampled in the unit quaternion space nor in the image space with the normalized ℓ_2 inner product metric. The non-uniformity, especially in the image space, may affect the performance and accuracy of dictionary-observation diffraction pattern matching. Indeed, the calculation of grain misorientations depends on the concentration of Euler angles or their surrogates (dictionary elements). To perform an accurate evaluation of this concentration using our dictionary matching pursuit methods, we need uniformly distributed dictionary elements over the ℓ_2 metric space. The drawback of the current non-uniformity is demonstrated in the evaluated consistency measuers within grains in Fig. 5.35 for IN100 and LSHR. We believe that more uniform dictionary can reduce the within-grain variability, as currently observed in Fig. 5.35.

However, our membership-based approach to classification, segmentation, and anomaly detecetion is robust against this non-uniformity and insensitive to the densely sampled image patterns. This is possible because our nearest neighbor model in the dictionary does not assume a uniform or fixed distance between different sample

patterns.

5.9 Appendix 2: Super-resolution algorithm

We present a super-resolution algorithm following the guideline in Section 5.5 that is more general than a simple but efficient xy scanning method used in the result section.

For each boundary pixel, Algorithm 6 tests many lines that cross the pixel and have a width W but different angles (e.g., from 0 to 165 degree with the increment of 15 in our experiment). For one of the lines, the algorithm evaluates the multiplied consistency degree with respect to the two adjacent grains, finds the super-resolved location maximizing the multiplied consistency, and evaluates the FWHM. Among the different directions of lines, the algorithm chooses the one having the least FWHM. This direction is the near-normal direction to the tested boundary and the new location maximizing the multiplied consistency is the super-resolved boundary location for the original boundary pixel. In our experiment, we used $W = 1$ and $\Delta\theta = 15$ degree to have $N_a = 12$.

Algorithm 6 Super-resolution algorithm

% Initialization:
Define the width W and the angular resolution $\Delta\theta$ of the scanning line to have N_a lines having angles of $\{\theta_k\}_{k=1}^{N_a}$ per pixel,
% Iterations:
for each boundary pixel j **do**
  for angle θ_k, $k = 1, 2, \ldots, N_a$ **do**
    Evaluate two most popular grains (defined to be G_1 and G_2) that are swept by the curve having angle θ_k,
    Evaluate consistency measures $c_1(t)$ and $c_2(t)$ in (5.14) and (5.15) where t is along the scanning line (super-resolved),
    Evaluate $c(k) := \text{argmax}_t c_1(t)c_2(t)$ and $FWHM(k) :=$ the full width at half maximum of $c(k)$,
  end for
  Evaluate the super-resolved boundary location $c(k_{opt})$ with $k_{opt} = \text{argmin}_k FWHM(k)$,
end for

CHAPTER VI

Conclusions

In this thesis, different aspects of microscopic imaging are investigated. For the MRFM data, we propose several solutions to the image reconstruction problem when we have imperfect knowledge of the point spread function that blurs the image. We address this mismatch by using our semi-blind deconvolution algorithms within a Bayesian framework. This semi-blind sparse image reconstruction is motivated by the MRFM imaging challenges and its sparse nature and PSF model. Our endeavor within a Bayesian perspective can be classified into two categories. First, a stochastic sampler is used that guarantees the convergence of the simulated distributions to the target distributions. The second category is our deterministic method that converges to solutions, by approximating the complex posterior distributions to the distributions of factored forms in their prior configuration. We also present our detailed blueprint for a future work. As an extension to the image model, we propose a random field model to incorporate smoothness, in contrast to sparsity, of images. Another extension involves the PSF field. We investigated methods dealing with the practical issues in the spatial invariance assumption of PSF, when it can be relaxed.

For the EBSD microscopy, we propose a novel approach to perform classification and anomaly detection by using a physics-based forward model or the dictionary. Under the model, we efficiently compute similarity measures between the dictionary

and data. From these statistics, raw EBSD patterns are classified into grains, boundaries, and two types of anomalies. This classification is robust to noise, instrumental differences, and material samples. Especially, this work is the first to delineate grain boundaries at a higher resolution than in the observation dimension and quantify the uncertainty for the discovered boundaries in a polycrystalline material.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, "Single spin detection by magnetic resonance force microscopy," *Nature*, vol. 430, no. 6997, pp. 329–332, 2004.

[2] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Processing*, vol. 18, pp. 2059–2070, Sept. 2009.

[3] K. Herrity, R. Raich, and A. O. Hero, "Blind reconstruction of sparse images with unknown point spread function," in *Proc. Computational Imaging Conference in IS&TSPIE Symposium on Electronic Imaging Science and Technology* (C. A. Bouman, E. L. Miller, and I. Pollak, eds.), vol. 6814, (San Jose, CA, USA), SPIE, Jan. 2008.

[4] C. L. Degen, M. Poggio, H. J. Mamin, C. T. Rettner, and D. Rugar, "Nanoscale magnetic resonance imaging," *Proc. Nat. Academy of Science*, vol. 106, pp. 1313–1317, Feb. 2009.

[5] P. G. Callahan and M. DeGraef, "Dynamical EBSD patterns: I. pattern simulations," *Microsc. Microanal.*, 2013, in press.

[6] R. Ward and B. Saleh, "Deblurring random blur," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 35, pp. 1494–1498, oct 1987.

[7] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, pp. 43–64, may 1996.

[8] J. Mamin, R. Budakian, and D. Rugar, "Point response function of an MRFM tip," tech. rep., IBM Research Division, Oct. 2003.

[9] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline, "Semi-blind deconvolution of neural impulse response in fMRI using a Gibbs sampling method," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal (ICASSP)*, vol. 5, pp. 601–604, May 2004.

[10] G. Pillonetto and C. Cobelli, "Identifiability of the stochastic semi-blind deconvolution problem for a class of time-invariant linear systems," *Automatica*, vol. 43, pp. 647–654, April 2007.

[11] P. Sarri, G. Thomas, E. Sekko, and P. Neveux, "Myopic deconvolution combining Kalman filter and tracking control," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal (ICASSP)*, vol. 3, pp. 1833–1836, 1998.

[12] G. Chenegros, L. M. Mugnier, F. Lacombe, and M. Glanc, "3D phase diversity: a myopic deconvolution method for short-exposure images: application to retinal imaging," *J. Opt. Soc. Am. A*, vol. 24, pp. 1349–1357, May 2007.

[13] S. U. Park, N. Dobigeon, and A. O. Hero, "Myopic sparse image reconstruction with application to MRFM," in *Proc. Computational Imaging Conference in IS&TSPIE Symposium on Electronic Imaging Science and Technology* (C. A. Bouman, I. Pollak, and P. J. Wolfe, eds.), vol. 7873, SPIE, Jan. 2011.

[14] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer, 2 ed., 2004.

[15] R. Molina, "On the hierarchical Bayesian approach to image restoration: applications to astronomical images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 1122–1128, nov 1994.

[16] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration from partially known blurs," *IEEE Trans. Image Processing*, vol. 9, pp. 1784–1797, oct 2000.

[17] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially known blurs," *Optical Eng*, vol. 41, pp. 1845–1854, 2002.

[18] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Processing*, vol. 15, pp. 3715–3727, 2006.

[19] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, (New York, NY, USA), pp. 787–794, ACM, 2006.

[20] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graphics*, vol. 27, no. 3, p. 1, 2008.

[21] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Processing*, vol. 18, pp. 1215–1227, June 2009.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[23] N. Nasios and A. Bors, "Variational learning for Gaussian mixture models," *IEEE Trans. Systems, Man, Cybernet. Part B*, vol. 36, pp. 849–862, Aug. 2006.

[24] A. Corduneanu and C. M. Bishop, "Variational Bayesian model selection for mixture distributions," in *Proc. Conf. Artificial Intelligence and Statistics*, pp. 27–34, 2001.

[25] K. Themelis, A. Rontogiannis, and K. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 60, pp. 585–599, feb. 2012.

[26] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu, "A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI.," *Neuroimage*, vol. 41, pp. 941–69, July 2008.

[27] D. Tzikas, A. Likas, and N. Galatsanos, "Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors," *IEEE Trans. Image Processing*, vol. 18, pp. 753–764, April 2009.

[28] F. Orieux, J.-F. Giovannelli, and T. Rodet, "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution," *J. Opt. Soc. Am. A*, vol. 27, pp. 1593–1607, July 2010.

[29] W. R. Gilks, *Markov Chain Monte Carlo In Practice*. Chapman and Hall/CRC, 1999.

[30] C. M. Bishop, J. M. Winn, and C. C. Nh, "Non-linear Bayesian image modelling," in *Proc. Eur. Conf. Comput. Vis. (EECV)*, pp. 3–17, Springer-Verlag, 2000.

[31] Z. Ghahramani and M. J. Beal, "Variational inference for Bayesian mixtures of factor analysers," in *Proc. Adv. in Neural Inf. Process. Syst. (NIPS)*, pp. 449–455, MIT Press, 2000.

[32] M. J. Beal and Z. Ghahramani, "The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures," in *Bayesian Stat.*, vol. 7, pp. 453–464, 2003.

[33] J. Winn, C. M. Bishop, and T. Jaakkola, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, pp. 661–694, 2005.

[34] S. Babacan, R. Molina, and A. Katsaggelos, "Variational Bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Processing*, vol. 18, pp. 12–26, jan. 2009.

[35] H. Attias, "A variational Bayesian framework for graphical models," in *Proc. Adv. in Neural Inf. Process. Syst. (NIPS)*, pp. 209–215, MIT Press, 2000.

[36] A. M. Walker, "On the asymptotic behaviour of posterior distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 31, no. 1, pp. 80–88, 1969.

[37] S. U. Park, N. Dobigeon, and A. O. Hero, "Semi-blind sparse image reconstruction with application to MRFM," *IEEE Trans. Image Processing*, vol. 21, pp. 3838 –3849, sept. 2012.

[38] B. Wang and D. Titterington, "Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values," in *Proc. Conf. Uncertainty in Artificial Intelligence (UAI)*, pp. 577–584, AUAI Press, 2004.

[39] B. Wang and M. Titterington, "Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model," *Bayesian Anal.*, vol. 1, no. 3, pp. 625–650, 2006.

[40] B. Amizic, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian blind image deconvolution with parameter estimation," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, pp. 626–630, Aalborg (Denmark), August 2010.

[41] M. Almeida and L. Almeida, "Blind and semi-blind deblurring of natural images," *Trans. Image Processing*, vol. 19, pp. 36–52, jan. 2010.

[42] J. A. Sidles, "Noninductive detection of single-proton magnetic resonance," *Appl. Phys. Lett.*, vol. 58, pp. 2854–2856, June 1991.

[43] J. A. Sidles, "Folded stern-gerlach experiment as a means for detecting nuclear magnetic resonance in individual nuclei," *Phys. Rev. Lett.*, vol. 68, pp. 1124–1127, Feb 1992.

[44] J. A. Sidles, J. L. Garbini, K. J. Bruland, D. Rugar, O. Züger, S. Hoen, and C. S. Yannoni, "Magnetic resonance force microscopy," *Rev. Mod. Phys.*, vol. 67, pp. 249–265, Jan 1995.

[45] D. Rugar, C. S. Yannoni, and J. A. Sidles, "Mechanical detection of magnetic resonance," *Nature*, vol. 360, pp. 563–566, Dec. 1992.

[46] O. Züger, S. T. Hoen, C. S. Yannoni, and D. Rugar, "Three-dimensional imaging with a nuclear magnetic resonance force microscope," *J. Appl. Phys.*, vol. 79, pp. 1881–1884, Feb. 1996.

[47] S. Chao, W. M. Dougherty, J. L. Garbini, and J. A. Sidles, "Nanometer-scale magnetic resonance imaging," *Review Sci. Instrum.*, vol. 75, pp. 1175–1181, April 2004.

[48] O. Züger and D. Rugar, "First images from a magnetic resonance force microscope," *Applied Physics Letters*, vol. 63, no. 18, pp. 2496–2498, 1993.

[49] O. Züger and D. Rugar, "Magnetic resonance detection and imaging using force microscope techniques," *J. Appl. Phys.*, vol. 75, pp. 6211–6216, May 1994.

[50] C. L. Degen, M. Poggio, H. J. Mamin, C. T. Rettner, and D. Rugar, "Nanoscale magnetic resonance imaging. Supporting information," *Proc. Nat. Academy of Science*, vol. 106, Feb. 2009.

[51] Z. Kato, T. chuen Pong, and S. G. Qiang, "Multicue mrf image segmentation: Combining texture and color features," in *International Conference on Pattern Recognition*, pp. 660–663, 2002.

[52] V. E. Johnson, "A model for segmentation and analysis of noisy images," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. pp. 230–241, 1994.

[53] S. Li, "On discontinuity-adaptive smoothness priors in computer vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, pp. 576 –586, jun 1995.

[54] V. E. Johnson, W. H. Wong, X. Hu, and C. tu Chen, "Image restoration using gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 413–425, 1991.

[55] T. Hebert and R. Leahy, "A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors," *Medical Imaging, IEEE Transactions on*, vol. 8, pp. 194 –202, jun 1989.

[56] J. Zhang, "The mean field theory in em procedures for blind markov random field image restoration," *Image Processing, IEEE Transactions on*, vol. 2, pp. 27 –40, jan 1993.

[57] R. Leahy and X. Yan, "Incorporation of anatomical mr data for improved functional imaging with pet," in *Information Processing in Medical Imaging* (A. Colchester and D. Hawkes, eds.), vol. 511 of *Lecture Notes in Computer Science*, pp. 105–120, Springer Berlin / Heidelberg, 1991.

[58] D. Higdon, J. Bowsher, V. Johnson, T. Turkington, D. Gilland, and R. Jaszczak, "Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data," *Medical Imaging, IEEE Transactions on*, vol. 16, pp. 516 –526, oct. 1997.

[59] M. DeGraef, "Multi-modal data in materials," in *Data Fusion: For the Detection of Rare and Anomalous Events*, (Dayton, OH, USA), Dec. 2012.

[60] J. M. Tyszka, S. E. Fraser, and R. E. Jacobs, "Magnetic resonance microscopy: recent advances and applications," *Current Opinion in Biotechnology*, vol. 16, no. 1, pp. 93 – 99, 2005.

[61] H. J. Mamin, R. Budakian, B. W. Chui, and D. Rugar, "Detection and Manipulation of Statistical Polarization in Small Spin Ensembles," *Physical Review Letters*, vol. 91, no. 20, 2003.

[62] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *ArXiv e-prints*, Sept. 2011.

[63] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[64] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, pp. 2354–2367, 2011.

[65] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Aug. 2003.

[66] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 1261–1264, Oct. 2006.

[67] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer-Verlag, 1999.

[68] P. Bremaud, *Markov Chains: Gibbs fields, Monte Carlo Simulation, and Queues.* Springer Verlag, 1999.

[69] Y. F. Atchadé and J. S. Rosenthal, "On adaptive markov chain monte carlo algorithms," *Bernoulli*, vol. 11, no. 5, pp. pp. 815–828, 2005.

[70] J.-M. Marin and C. P. Robert, *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York, NY, USA: Springer, 2007.

[71] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[72] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[73] S. U. Park, N. Dobigeon, and A. O. Hero, "Variational semi-blind sparse image reconstruction with application to MRFM," in *Proc. Computational Imaging Conference in IS&T SPIE Symposium on Electronic Imaging Science and Technology* (C. A. Bouman, I. Pollak, and P. J. Wolfe, eds.), vol. 8296, pp. 82960G–82960G–11, SPIE, Jan. 2012.

[74] S. U. Park, N. Dobigeon, and A. O. Hero, "Variational semi-blind sparse deconvolution with orthogonal kernel bases and its application to MRFM," *Signal Processing*, vol. 94, no. 0, pp. 386 – 400, 2014.

[75] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," *Amer. J. Math.*, vol. 73, pp. 615–624, July 1951.

[76] D. Ge, J. Idier, and E. L. Carpentier, "Enhanced sampling schemes for MCMC based blind BernoulliGaussian deconvolution," *Signal Processing*, vol. 91, no. 4, pp. 759 –772, 2011.

[77] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, pp. 1059–1074, april 2010.

[78] J. Besag and P. J. Green, "Spatial statistics and Bayesian computation," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 55, no. 1, pp. 25–37, 1993.

[79] M. A. T. Figueiredo and J. M. N. Leitao, "Unsupervised image restoration and edge location using compound Gauss-Markov random fields and the MDL principle," *IEEE Trans. Image Processing*, vol. 6, pp. 1089–1102, 1997.

[80] R. Mittelman, N. Dobigeon, and A. Hero, "Hyperspectral image unmixing using a multiresolution sticky HDP," *IEEE Trans. Signal Processing*, vol. 60, no. 4, pp. 1656–1671, 2012.

[81] F. Forbes and G. Fort, "Combining Monte Carlo and mean-field-like methods for inference in hidden Markov random fields," *IEEE Trans. Image Processing*, vol. 16, no. 3, pp. 824–837, 2007.

[82] F. Forbes and N. Peyrard, "Hidden Markov random field model selection criteria based on mean field-like approximations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1089–1101, 2003.

[83] The SDSS collaboration, "The Sloan Digital Sky Survey - Mapping the Universe." http://sdss.org, 2013. accessed Jan 7, 2013.

[84] R. Lupton, J. E. Gunn, Z. Ivezic, G. R. Knapp, S. Kent, and N. Yasuda, "The SDSS imaging pipelines," in *ASP Conf. Ser. Astronomical Data Analysis Software and Systems X*, vol. 238, 2001.

[85] S. U. Park, D. Wei, M. DeGraef, M. Shah, J. Simmons, and A. O. Hero, "EBSD image segmentation using a physics-based forward model," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Sept. 2013, in press.

[86] S. U. Park, D. Wei, M. DeGraef, M. Shah, J. Simmons, and A. O. Hero, "EBSD image segmentation using a physics-based pattern dictionary," in *Microscopy and Microanalysis 2013*, Aug. 2013, in press.

[87] Oxford Instruments, "BASICS OF EBSD - PRINCIPLE SYSTEM COMPONENTS." http://ebsd.com, 2013. accessed February 5, 2013.

[88] J. R. Michael and J. I. Goldstein, "EBSD characterization of materials: From meteorites to welds to MEMS," *Microsc. Microanal.*, vol. 11, pp. 1324–1325, 8 2005.

[89] E. J. Mittemeijer and P. Scardi, *Diffraction analysis of the microstructure of materials*, vol. 68. Springer, 2004.

[90] A. Schwartz, M. Kumar, B. Adams, and D. Field, *Electron Backscatter Diffraction in Materials Science*. Springer, 2009.

[91] Center for Nanoscience and Nanotechnology, Hebrew University of Jerusalem, "High Resolution Scanning Electron Microscope Sirion: Basics and Tutorials." http://nanoscience.huji.ac.il/unc/sem_basics.htm, 2013. accessed February 5, 2013.

[92] H. Nordmark, M. Di Sabatino, M. Acciarri, J. Libal, S. Binetti, E. J. Ovrelid, J. C. Walmsley, and R. Holmestad, "EBIC, EBSD and TEM study of grain boundaries in multicrystalline silicon cast from metallurgical feedstock," in *Photovoltaic Specialists Conference, 2008. PVSC '08. 33rd IEEE*, pp. 1 –6, may 2008.

[93] Y. Wang, H. Kimura, Y. Akiniwa, and K. Tanaka, "EBSD-AFM hybrid analysis of fatigue slip system and crack initiation in polycrystalline metal under cyclic torsional loading," in *Micro-NanoMechatronics and Human Science, 2005 IEEE International Symposium on*, pp. 223 – 228, nov. 2005.

[94] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, pp. 1101 –1113, nov 1993.

[95] L. Shapiro and G. Stockman, *Computer visión*. Prentice Hall, 2001.

[96] S. Osher and N. Paragios, *Geometric Level Set Methods in Imaging,Vision,and Graphics*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.

[97] T. Chan and L. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, pp. 266 –277, feb 2001.

[98] H.-C. Chuang, L. Huffman, M. Comer, J. Simmons, and I. Pollak, "An automated segmentation for nickel-based superalloy," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pp. 2280 –2283, oct. 2008.

[99] A. Winkelmann, C. Trager-Cowan, F. Sweeney, A. P. Day, and P. Parbrook, "Many-beam dynamical simulation of electron backscatter diffraction patterns," *Ultramicroscopy*, vol. 107, pp. 414–421, 2007.

[100] D. Roşca, "New uniform grids on the sphere," *Astronomy and Astrophysics*, vol. 520, pp. 1–4, Oct. 2010.

[101] G. McLachlan and D. Peel, *Finite Mixture Models*. Wiley series in probability and statistics: Applied probability and statistics, Wiley, 2004.

[102] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. Chapman & Hall/CRC, 1984.

[103] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer Series in Statistics, 2009.

[104] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 399–409, 1997.

[105] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sensing of Environment*, vol. 86, no. 4, pp. 554–565, 2003.

[106] BlueQuartz Software, "DREAM.3D: Digital Representation Environment for Analyzing Microstructure in 3D." http://dream3d.bluequartz.net, 2013. accessed February 5, 2013.

[107] K. Mardia and P. Jupp, *Directional statistics*. Wiley series in probability and statistics, Wiley, 2000.