# COMPUTATIONAL IMAGING AND INVERSE TECHNIQUES FOR HIGH-RESOLUTION AND INSTANTANEOUS SPECTRAL IMAGING

BY

SEVINC FIGEN OKTEM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Farzad Kamalabadi, Chair
Professor Richard E. Blahut, Co-Chair
Professor Yoram Bresler
Doctor Joseph M. Davila, NASA Goddard Space Flight Center

# ABSTRACT

In this thesis, we develop a class of novel spectral imaging techniques that enable capabilities beyond the reach of conventional methods. Each development is based on computational imaging, which involves distributing the spectral imaging task between a physical and a computational system and then digitally forming images of interest from multiplexed measurements by means of solving an inverse problem. In particular, in the first approach, a nonscanning spectral imaging technique is developed to enable performing spectroscopy over a two-dimensional instantaneous field-of-view. This technique combines a parametric estimation approach with a slitless spectrometer configuration. In the second approach, a spectral imaging technique with an optical device known as a photon sieve is developed to achieve superior spatial and spectral resolutions relative to conventional filter-based spectral imagers. This technique relies on the wavelength-dependent focusing property of the photon sieve, and multiplexed measurements recorded by a photon sieve imaging system with a moving detector. In each of these two techniques, multiplexed measurements are combined with an image formation model and then the resultant inverse problem is solved computationally for image reconstruction. The associated inverse problems, which can be viewed as multiframe image deblurring problems, are formulated in a Bayesian estimation framework to incorporate the additional prior statistical knowledge of the targeted objects. Computationally efficient algorithms are then designed to solve the resulting nonlinear optimization problems. In addition to the development of each technique, Bayesian Cramer-Rao bounds are also obtained to characterize the estimation uncertainties and performance limits, as well as to explore the optimized system design. The effectiveness of the spectral imaging techniques are illustrated for an application in remote sensing of the solar atmosphere. Lastly, the phase retrieval problem, another inverse problem that arises in the photon-sieve imaging setting with coherent illumi-

nation, is studied to devise computationally efficient algorithms. As a whole, the developed spectral imaging techniques enable finer spectral information in the form of higher temporal, spatial, and spectral resolutions. This will enhance the unique diagnostic capabilities of conventional spectral imaging systems in applications as diverse as physics, chemistry, biology, medicine, astronomy and remote sensing.

*To my sweet family,*
*for their support, encouragement, and unconditional love.*

# ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Prof. Farzad Kamalabadi and Prof. Richard E. Blahut for being mentors to me in many more ways than one. Above all, I am deeply indebted to Prof. Kamalabadi for his continuous support and encouragement in supervising my thesis, as well as for giving the opportunity to work with him on exciting problems in spectral imaging. Likewise, I will always be indebted to Prof. Blahut for enabling me to enter the field of computational imaging and inverse problems, as well as for all his caring efforts to bring out the best in me, and all the freedom that he gave me.

Besides my advisors, I am very fortunate to know and work with Dr. Joe Davila, whose innovative ideas on solar spectroscopy became the basis for this study. I would like to thank him for all the helpful discussions and support, as well as for graciously sharing his office with me during my internship. I am also grateful to Prof. Bresler for always being accessible and providing insightful perspectives on improving this thesis. Moreover, I would like to thank Prof. O'Sullivan and Prof. Snyder for many useful discussions on phase retrieval and reconstruction algorithms, and Prof. Ozaktas for always being there for me.

I am thankful to NASA Headquarters for two years of financial support as a Doctoral Research Fellow. More specifically, this work has been supported under the Earth and Space Science Fellowship Program - Grant "NNX12AL74H". I also thank Peggy Wells and Nancy Morris for their kind help with various administrative matters.

This thesis would have not been possible without all the friendship. I will remain forever indebted to three great people, Ghazale, Gizem, and Mik, who have been my family here and have stood with me in all the tough times as well as the good. I am also grateful to the past and present members of my lab, Sara, Navid, Qiaomin, Vineet, and Jay, for their kind

help and friendship, as well as for stimulating discussions. I would like to thank everybody else who accompanied me in this PhD journey, including Sena, Seda, Giray, Cagdas, Ozlem, Levent, Mujde, Gulcin, Ozan, Gamze, Ergin, Deniz, Esra, Aly, Taylor, Sai, and Noyan, with whom I shared joy and sorrow. Special thanks go to my far-away friends from college and high school, Hakan, Ezgi, Hilal, Pinar, Celil, Esra, Ata, for sticking with me and keeping me smiling all the way through.

And last of all, I want to thank my family from the bottom of my heart for giving me all the love, happiness, and freedom that I could ask for. They have been there for me every step of the way and have always loved me unconditionally. Without their love and support, I would not have become what I am now.

# TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION . . . . . 1
1.1 Dissertation goals and organization . . . . . 2

CHAPTER 2 NONSCANNING (INSTANTANEOUS) SPECTRAL IMAGING . . . . . 6
2.1 Introduction . . . . . 6
2.2 Parametric forward problem . . . . . 9
2.3 Inverse problem . . . . . 15
2.4 Dynamic programming algorithm . . . . . 17
2.5 Sample application in solar spectral imaging . . . . . 26
2.6 Conclusion . . . . . 33

CHAPTER 3 HIGH-RESOLUTION SPECTRAL IMAGING WITH PHOTON-SIEVES . . . . . 35
3.1 Introduction . . . . . 35
3.2 Part 1: Image formation with photon sieves . . . . . 37
3.3 Fresnel image formation models . . . . . 38
3.4 Approximate image formation models . . . . . 42
3.5 Numerical comparisons of the models . . . . . 44
3.6 Part 2: Computational spectral imaging with photon sieves . . 44
3.7 Forward problem . . . . . 46
3.8 Inverse problem . . . . . 49
3.9 Sample application in solar spectral imaging . . . . . 51
3.10 Conclusion . . . . . 55

CHAPTER 4 PERFORMANCE LIMITS IN COMPUTATIONAL SPECTRAL IMAGING . . . . . 56
4.1 Introduction . . . . . 56
4.2 Image formation model . . . . . 57
4.3 Bayesian Cramer-Rao error bounds . . . . . 58
4.4 Performance limits for instantaneous spectral imaging . . . . . 62
4.5 Conclusion . . . . . 66

CHAPTER 5 COMPUTATIONAL METHODS FOR PHASE RETRIEVAL . . . . . 71
5.1 Introduction . . . . . 71
5.2 Applications . . . . . 71
5.3 Problem definition and characteristics . . . . . 72
5.4 Existing algorithms . . . . . 74
5.5 Connection of Schulz-Snyder algorithm to well-known methods 87
5.6 Global optimization methods . . . . . 96
5.7 Conclusion . . . . . 109

CHAPTER 6 CONCLUSIONS . . . . . 110

APPENDIX A PROOF OF THEOREM 1 . . . . . 113

APPENDIX B DERIVATION OF THE FISHER INFORMATION MATRIX . . . . . 117

APPENDIX C A NECESSARY CONDITION FOR CONVEXITY IN ALTERNATING-MINIMIZATION . . . . . 120

REFERENCES . . . . . 121

# CHAPTER 1

# INTRODUCTION

*Spectral imaging* or *imaging spectroscopy* is a fundamental diagnostic technique in the physical sciences with application in diverse fields such as physics, chemistry, biology, medicine, astronomy, and remote sensing [1, 2]. Spectral imagers enable sensing properties of a scene based on measurement of radiated energy interacting with matter. The measured spectrum (i.e. radiation intensity as a function of wavelength) provides a means for uniquely identifying the physical, chemical, and biological properties of targeted objects [3, 4]. This makes spectral imaging a useful diagnostic tool in various applications including remote sensing of astrophysical plasmas, environmental monitoring, resource management, biomedical diagnostics, industrial inspection, and surveillance, among many others.

For example, in astrophysical imaging of space plasmas, which was our initial motivation for this study, energy transitions of the constituent matter in the plasma produce spectral emission lines. Measurements of the emitted spectrum provide estimates of the parameters of these spectral lines, which are essential for inferring the plasma parameters (such as density, temperature, and flow speed of the radiating ions) [5, 6]. Such measurements enable the investigation of the complex plasma behavior by revealing how particles and energy flow through the radiating plasma [7, 8].

The objective of spectral imaging is to form images of a scene as a function of wavelength. For a two-dimensional scene, this requires simultaneously obtaining a three-dimensional data: one for spectral and two for spatial dimensions. As illustrated in Fig. 1.1, the slices of this data cube represent images of the scene at different wavelengths, whereas the data at a single spatial position give the full spectrum emitted from that position.

However, obtaining this three-dimensional spectral data cube with inherently two-dimensional detectors poses intrinsic limitations on the spatio-spectral extent of the technique. To address this limitation, conventional

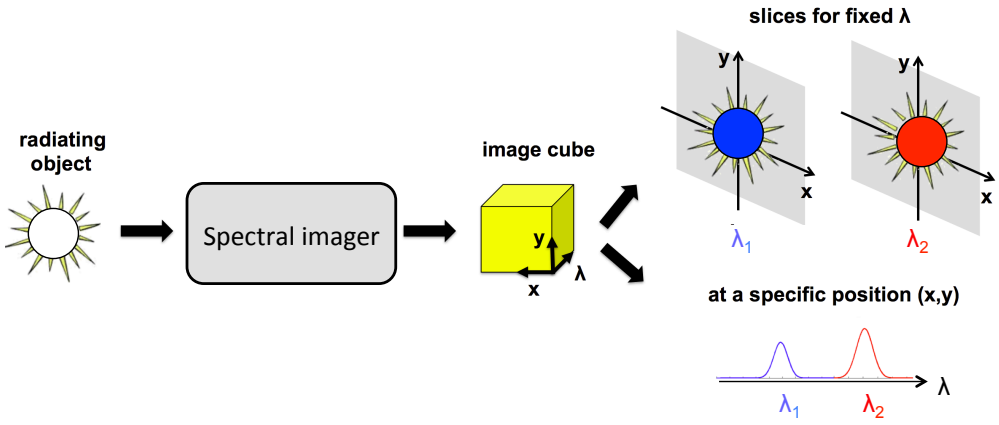



Figure 1.1: A cartoon of what spectral imaging is about. The goal of the spectral imager is to measure the radiation intensity of a two-dimensional source as a function of space and wavelength. The data that it generates has two spatial dimensions and one spectral dimension, so it is a data cube. The slices of this data cube give images of the scene at different wavelengths, whereas the data at a single spatial position give the full spectrum emitted from that position.

spectral imaging techniques rely on a scanning process to build up the three-dimensional (3D) *data cube* from a series of two-dimensional (2D) measurements that are acquired sequentially. Typically this is done by using a spectrometer with a long slit and scanning the scene spatially, or by using an imager with a series of spectral filters and scanning the scene spectrally. In the former case (referred to as rastering or push broom) only a thin slice of the scene is observed at a time, whereas in the latter case only one spectral band is observed at a time. Similarly, Fourier and Hadamard transform based spectrometers perform scanning in a transform domain (through their movable parts) to build up the three-dimensional data cube [1].

## 1.1 Dissertation goals and organization

The overall goal of this dissertation is to develop a class of novel spectral imaging techniques motivated by the limitations of conventional spectral imaging systems. Each development is based on computational imaging, which involves distributing the spectral imaging task between a physical and a computational system and then digitally forming images of interest from multiplexed measurements by means of solving an inverse problem.

Conventional spectral imaging techniques purely rely on physical systems, which impose inherent limitations on the performance such as temporal, spatial, and spectral resolutions. On the other hand, computational spectral imaging techniques enable to overcome these physical limitations by passing on some of the burden to a computational system. The added computational part provides flexibility to combine information from different multiplexed measurements, as well as to incorporate the additional prior knowledge about the objects of interest into the image formation process. In particular, the prior knowledge used in this study consists of the statistical knowledge of the spatial and spectral distributions, and the information that the spectra are composed of discrete lines (for which the techniques are designed).

Each computational imaging technique is developed in three steps. First, a novel optical system is used to overcome the inherent physical limitations in spatial, spectral, and temporal resolutions of conventional systems. Second, the inverse problem for image reconstruction is formulated by combining multiplexed measurements with an image formation model and using a Bayesian estimation framework. And third, computationally efficient algorithms are designed to solve the resulting nonlinear optimization problems. In addition to the development of each technique, Bayesian Cramer-Rao bounds are also obtained to characterize the estimation uncertainties and performance limits, as well as to explore the optimized system design. The effectiveness of the spectral imaging techniques are illustrated in an application for remote sensing of the solar atmosphere, which was the initial motivation for this study. The thesis describes all these aspects, with a particular focus on the inverse problems involved.

The results of our study are presented in the rest of this thesis as follows:

### 1.1.1 Computational imaging for instantaneous spectral imaging

Due to the intrinsic limitation of two-dimensional detectors in capturing inherently three-dimensional spectral data cube, spectral imaging techniques conventionally rely on a spatial or spectral scanning process, which renders them unsuitable for dynamic scenes. In Chapter 2, a nonscanning (instantaneous) spectral imaging technique is developed to enable performing spec-

troscopy over a two-dimensional instantaneous field-of-view. Hence this technique offers the additional capability of an instantaneous two-dimensional field-of-view over the conventional slit spectrometers. Here the physical parameters of interest are estimated by combining the multiplexed measurements of a slitless spectrometer with a parametric model and then solving the resultant inverse problem computationally. The associated inverse problem, which can be viewed as a multiframe semiblind deblurring problem (with shift-variant blur), is formulated as a maximum posterior (MAP) estimation problem since in many such experiments prior statistical knowledge of the physical parameters can be well estimated. Subsequently, an efficient dynamic programming algorithm is developed to find the global optimum of the nonconvex MAP problem. Lastly, the algorithm and the effectiveness of the spectral imaging technique are illustrated for an application in solar spectral imaging.

### 1.1.2 Computational imaging for high-resolution spectral imaging with photon sieves

The photon sieve, a modification of the Fresnel zone plate, is a new class of diffractive image forming devices that opens up new possibilities for high resolution imaging and spectroscopy, especially at ultra-violet (UV) and x-ray regime. In Chapter 3, we develop a novel computational photon sieve imaging modality that enables superior spatial and spectral resolutions relative to conventional filter-based spectral imagers. This technique relies on the wavelength-dependent focusing property of the photon sieve, and multiplexed measurements recorded by a photon sieve imaging system with a moving detector. First, we derive exact and approximate Fresnel imaging formulas that relate the output of a photon sieve imaging system to its input, originating from either a coherent or incoherent extended source. For the spatially incoherent illumination, we then study the problem of recovering the individual spectral images from the superimposed and blurred measurements of the proposed photon sieve system. This inverse problem, which can be viewed as a multiframe deconvolution problem involving multiple objects, is formulated as a maximum posterior estimation problem, and solved using a fixed-point algorithm. The performance and effectiveness of the proposed

technique are illustrated for an application in solar spectral imaging through computer simulations.

### 1.1.3 Fundamental performance limits for computational spectral imaging

In the first two chapters, we develop a class of novel computational spectral imaging techniques that enable capabilities beyond the reach of the conventional methods. Since an inversion is required for the reconstruction of the spectral imaging information from the noisy measurements, a rigorous theory is essential for quantitative characterization of the performance of the techniques. In Chapter 4 we develop such a theory using the Bayesian Cramer-Rao lower bounds. The lower bounds for estimation uncertainties are presented for a fairly general image formation model, and then used to explore the performance limits of the instantaneous spectral imaging technique. Via Monte Carlo simulations, the tightness of the bounds and performance of the developed MAP algorithm are evaluated under different observing scenarios and instrument design considerations for an application in solar spectral imaging. The developed framework allows us not only to characterize the fundamental precision limits, but also to explore the design requirements that render these imaging modalities effective.

### 1.1.4 Computational methods for phase retrieval

Phase retrieval problems arise in the photon-sieve imaging setting with coherent illumination. These problems are generalizations of the classical phase retrieval problem, which is the recovery of a signal from the magnitude of its Fourier transform. In Chapter 5, we analyze and compare important algorithms for the classical phase retrieval problem, which is notoriously difficult to solve due to the nonlinearity involved. In particular, we derive the Schulz-Snyder phase retrieval algorithm as an alternating minimization method, and discuss its advantages and drawbacks. An annealing-type Schulz-Snyder algorithm, a hybrid method that incorporates annealing-type global optimization methods, is also proposed to avoid convergence to nonglobal solutions.

# CHAPTER 2

# NONSCANNING (INSTANTANEOUS) SPECTRAL IMAGING

## 2.1 Introduction

The objective of spectral imaging is to form images of a scene as a function of wavelength. For a two-dimensional scene, this requires simultaneously obtaining a three-dimensional data: one for spectral and two for spatial dimensions. However, obtaining this three-dimensional data with inherently two-dimensional detectors poses intrinsic limitations on the spatio-spectral extent of the technique. To address this limitation, conventional spectral imaging techniques rely on a scanning process to build up the three-dimensional (3D) *data cube* from a series of two-dimensional (2D) measurements that are acquired sequentially. Typically this is done by using a spectrometer with a long slit and scanning the scene spatially, or by using an imager with a series of spectral filters and scanning the scene spectrally. In the former case (referred to as rastering or push broom) only a thin slice of the scene is observed at a time, whereas in the later case only one spectral band is observed at a time. Similarly, Fourier and Hadamard transform based spectrometers perform scanning in a transform domain (through their movable parts) to build up the three-dimensional data cube [1]. As a result, these conventional methods are effective for scenes that remain stationary during the scanning process involved.

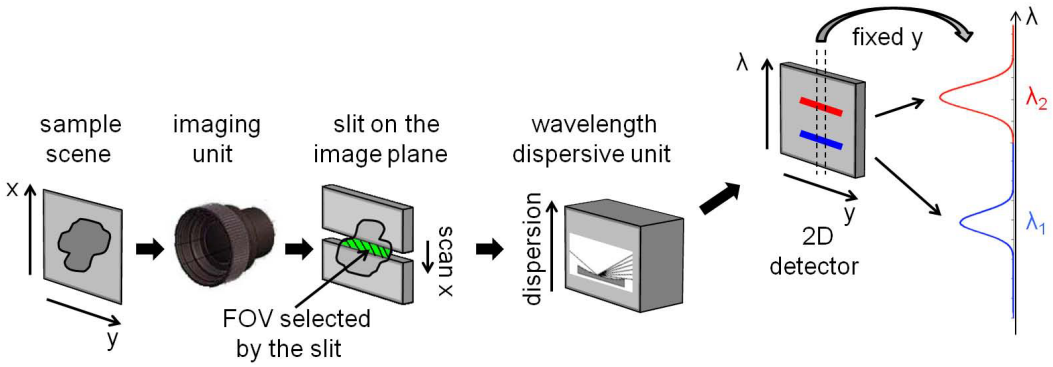
More recently, methods that reconstruct the three-dimensional data cube from a single-shot measurement have been proposed by using *tomographic* approaches [10–14], and *coded apertures* [15, 16]. The main idea in the *tomographic* approach is to build up the three-dimensional data cube from

A preliminary version of the results of this chapter has been recently presented in [9].

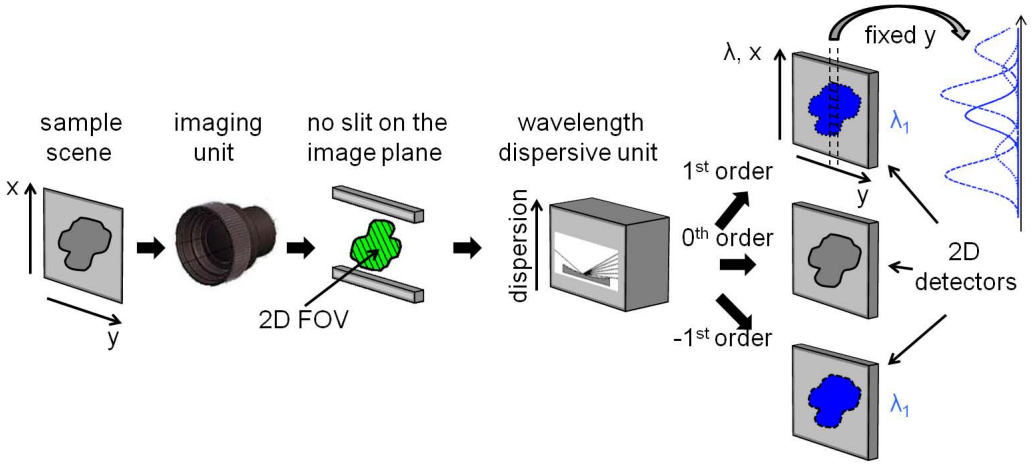
its two-dimensional projections, where the projections are obtained through spectrally dispersed images of the scene [10], each dispersed in a different direction and by a different amount (through different optical diffraction orders of a grating). These tomographic approaches require a large set of projections (i.e. dispersed images) to be captured at once, hence demanding a large detector area to be used. The resulting cost generally limits either the FOV or the resolution. On the other hand, instantaneous *coded aperture* techniques [15, 16] approach this problem by acquiring only one coded projection, while simultaneously requiring that the scene be sparse in some transform domain. Although the imposed sparsity assumption may not hold in general, for the cases where such restriction holds, the three-dimensional data cube can be reconstructed from the single two-dimensional measurement using compressive sensing methods.

In this chapter, we develop a *parametric* approach to achieve instantaneous (nonscanning) spectral imaging. Figure 2.1b shows the schematic depiction of the system involved. This system is quite similar to a conventional slit (push broom) spectrometer which is illustrated in Fig. 2.1a; but unlike a slit spectrometer, this system has an instantaneous *two-dimensional* FOV (rather than a one-dimensional FOV limited by a slit). Also, because dispersing the two-dimensional input image causes an overlap of spatial and spectral information on the 2D detector, *multiple* spectrally-dispersed images of the scene are acquired simultaneously in order to gather the needed information for reconstruction of the spectral imaging information. This idea is similar to the previous snapshot tomographic approaches [10–14], but different in that our technique relies on a parametric approach to reconstruct the 3D data cube from significantly smaller number of dispersed images (hence demanding a smaller detector area).

Our approach treats the problem of estimating the physical parameters of interest from the measurements of this instantaneous spectral imager by using a parametric model for the measurements. Based on this parametric model, the estimation problem can be viewed as a one-dimensional multiframe semiblind deblurring problem with shift-variant blur, where multiple blurred images of the same scene are obtained through multiple dispersed images, each with a different diffraction order. We formulate the inverse problem as a maximum posterior (MAP) estimation problem since in many such experiments prior statistical knowledge of the physical parameters can




(a) Conventional slit spectrometer




(b) Developed instantaneous spectral imager

Figure 2.1: Comparison of (a) conventional slit spectrometer versus (b) developed instantaneous spectral imager. In both cases, an imaging unit (e.g. lens, mirror) focuses a 2D scene on an image plane. In a slit spectrometer (a), a narrow slit lies on the image plane to limit the FOV to a 1D portion of the scene. The light that passes through the slit is input to a wavelength dispersive unit, which generally consists of a collimator optics (e.g., lens), a dispersive element (e.g., diffraction grating), and a focusing optics (e.g., lens). Each spectral line in the incoming light is dispersed according to the wavelength and imaged onto a 2D detector. To obtain the spectrum of the entire 2D scene, the slit is moved within the image plane to scan the scene spatially. For the instantaneous spectral imager (b), the slit is widened to achieve an instantaneous 2D FOV. This causes an overlap of spatial and spectral information on the detector: dispersed spectral lines from all spatial positions within the FOV are superimposed. To decompose this superimposed data computationally, multiple spectrally dispersed images with different diffraction orders are recorded using multiple detectors (in this case three detectors for the orders +1, 0, and -1).

be well estimated. The resulting nonconvex MAP problem is solved by developing an efficient dynamic programming algorithm, which is an extension of a previously proposed algorithm for maximum likelihood parameter estimation of superimposed signals [17, 18]. The developed algorithm, whose preliminary version appeared in [9], yields parameter estimates that are close to the global optimum of the MAP problem. A local optimization algorithm initialized with these estimates can then be used to obtain the desired global optimum. Through numerical investigations, we verify the results of a Cramer- Rao bound analysis in [19] demonstrating that the physical parameters can be estimated with the same order of accuracy as the conventional slit spectroscopy, while enabling a two-dimensional FOV at the same time.

This chapter is organized as follows. In Section 2.2, we introduce the parametric forward model (for the dispersed images). The inverse problem is formulated in Section 2.3 as a MAP problem. Section 2.4 presents the dynamic programming algorithm for efficiently solving the MAP problem. Numerical simulation results for an application in solar spectral imaging are presented in Section 2.5. Section 2.6 concludes the chapter.

## 2.2 Parametric forward problem

In a slit spectrometer (see Fig. 2.1a), a narrow slit lies on the image plane of a lens or mirror, hence limiting the field-of-view to a one-dimensional portion of the scene. As a result, only a thin slice of the scene is observed at a time. The light that passes through the slit then enters into a wavelength dispersive unit where each spectral line in the incoming light is dispersed according to wavelength and imaged onto a two-dimensional detector [1]. Because the dispersion plane is aligned to be perpendicular to the long side of the slit, one dimension in the detector corresponds only to the wavelength ($\lambda$) whereas the other dimension corresponds to the spatial dimension admitted through the slit ($y$). (Hence the spatial and spectral information do not overlap on the detector, and each resulting dispersed spectral line is associated with a single position on the scene.) To obtain spectral information of an entire two-dimensional scene, the scene is scanned spatially using the slit, i.e., the one-dimensional instrument FOV is pointed to a series of adjacent spatial ($x$) positions on the scene, with a narrow slit exposure taken at each pointing

location.

This approach is not suitable for dynamic scenes that evolve on time scales faster than the scanning process involved. For example, in solar spectroscopy, the scanning takes on the order of minutes (to cover an active/dynamic region of interest) whereas the physical processes occurring in the solar plasma change on the order of seconds [12].

To overcome this limitation with an instantaneous spectral imager, the width of the entrance slit is increased to obtain an instantaneous two-dimensional FOV (see Fig. 2.1b). (Hence a two-dimensional image of the scene is allowed at the image plane of the imaging unit.) Then the light from the two-dimensional scene enters into the dispersive unit where each spectral line in the incoming beam is dispersed and imaged onto a two-dimensional detector. Because now the input to the dispersive system is two-dimensional, dispersion causes an overlap of spatial and spectral information on the detector. More specifically dispersed spectral lines from all positions along the spatial dimension that is parallel to the dispersion plane are now superimposed at the output. To overcome the difficulty of decomposing this superimposed data, *multiple* spectrally dispersed images (of the two-dimensional scene) are recorded simultaneously using multiple detectors. These dispersed images differ by the amount and direction of dispersion as determined by different diffraction (spectral) orders. In particular, a negative diffraction order indicates the reversal of the dispersion direction, and higher diffraction orders indicate larger amounts of dispersion.

### 2.2.1 Parametric image formation model

In the parametric model, dispersed images are expressed as superposition of dispersed spectral lines from different spatial positions on the scene. If the dispersion plane of the dispersive unit is aligned to be parallel to the columns of pixels on the detector (as illustrated in Fig. 2.1b), then spectra from neighboring columns are not mixed; that is, in the dispersed image, only spectral lines from positions along a single column are superimposed. This allows us to treat the two-dimensional problem as a one-dimensional problem where each column of the dispersed image is modeled independently.

Then considering a column of pixels of length $M$, the dispersed spectral

lines from all of these pixels are superimposed; hence the observed intensity at any detector pixel is the sum of contributions from all of these spectral lines. Fig. 2.2 illustrates this superposition on a single column of pixels.

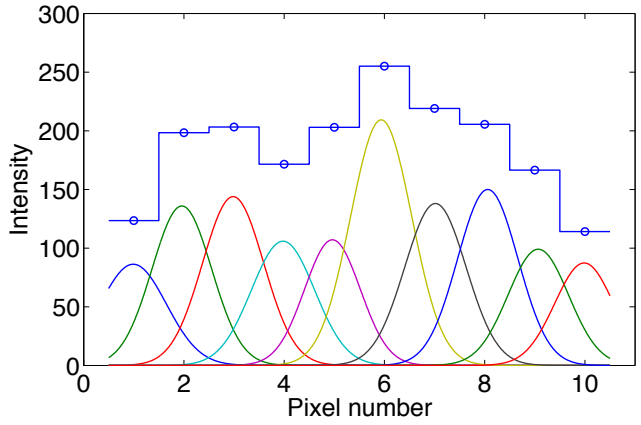



Figure 2.2: One column of the simulated dispersed image (in a ten-pixel detector column), where colored spectral lines are associated with different pixels, and bars correspond to the total observed intensity at each pixel.

In the parametric model, dispersed spectral line from pixel $m'$ is modeled to have a Gaussian shape, and is characterized by three parameters of interest: an integrated line intensity (strength) $f_{m'}$, a finite line width (broadening) $\Delta_{m'}$, and a line center shift (Doppler shift) $\epsilon_{m'}$. Here the Gaussian assumption is due to thermal (Doppler) broadening [8, 20, 21]. The finite line width is the result of the thermal motions of the emitting particles along the line of sight (in the radiating scene), and Doppler shifts in wavelength, i.e., variations in the line center position, are associated with coherent flows along the line of sight.

Subsequently, the observed intensity at any detector pixel is given by the sum of contributions from all of these parametric spectral lines. In particular, the contribution of each spectral line (to the total intensity observed at a pixel) is given by the integrated intensity of the spectral line over that pixel. Let $x'$ be a continuous variable in pixel units that denotes the vertical location on the detector, and let the $m$th detector pixel correspond to the range $m - 1/2 \leq x' < m + 1/2$, where $m = 1, \ldots, M$. Then the contribution of the spectral line at pixel $m'$ to the total intensity at pixel $m$ can be found by

integrating the Gaussian spectral line over the $m$th pixel:

$$\begin{aligned} c_{m'} &= \int_{m-1/2}^{m+1/2} \frac{f_{m'}}{\sqrt{2\pi}|a|\Delta_{m'}} \exp\left[-\frac{(x'-m'-a\,\epsilon_{m'})^2}{2(a\Delta_{m'})^2}\right] dx' \\ &= \frac{f_{m'}}{\sqrt{2\pi}|a|\Delta_{m'}} \int_{-1/2}^{1/2} \exp\left[-\frac{(x'+m-m'-a\,\epsilon_{m'})^2}{2(a\Delta_{m'})^2}\right] dx' \\ &= f_{m'} \frac{\mathrm{erf}(t_2)-\mathrm{erf}(t_1)}{2} \end{aligned} \tag{2.1}$$

where the error function $\mathrm{erf}(t)$ is

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x'^2) dx', \tag{2.2}$$

$$t_{1,2} = \frac{m-m' \mp 1/2 - a\epsilon_{m'}}{\sqrt{2}\mid a \mid \Delta_{m'}} \tag{2.3}$$

and both line width $\Delta_{m'}$ and Doppler shift $\epsilon_{m'}$ are measured in pixel units and in the first diffraction order ($a = +1$). At higher orders, these are scaled by the spectral order $a$ because the physical spread of any spectral interval on the detector is enlarged by the order $a$ (since the dispersion amount varies linearly with order) [4]. Also a negative sign in the order indicates reversal of the direction of dispersion, which affects the relative direction of line center shift (Doppler shift).

Then the total intensity at the detector pixel $m$, denoted by $y_m^a$, is given by the sum of contributions from all spectral lines at pixels $m' = 1, \ldots, M$:

$$y_m^a = \sum_{m'=1}^{M} f_{m'} \frac{\mathrm{erf}(t_2)-\mathrm{erf}(t_1)}{2} \quad \text{if } a \neq 0 \tag{2.4}$$

for all $m = 1, \ldots, M$. This is true for any dispersed image with order $a \neq 0$. On the other hand, for the zero order image, which is the result of direct imaging without any dispersion, the total intensity at each pixel is simply the integrated line intensity of the spectral line at that pixel:

$$y_m^0 = f_m \tag{2.5}$$

Therefore, the complete parametric model for the intensities of the $a$th

order dispersed image can be expressed as

$$y_m^a = \sum_{m'=1}^{M} f_{m'} \phi_{m-m'}^a(\mathbf{\Theta}_{m'}) \tag{2.6}$$

where $\mathbf{\Theta}_{m'} = [\epsilon_{m'},\ \Delta_{m'}]$, and the contribution amount $\phi_{m-m'}^a(\mathbf{\Theta}_{m'})$ is

$$\phi_{m-m'}^a(\mathbf{\Theta}_{m'}) = \begin{cases} \frac{\mathrm{erf}(t_2)-\mathrm{erf}(t_1)}{2} & \text{if } a \neq 0; \\ \delta_{m-m'} & \text{if } a = 0 \end{cases} \tag{2.7}$$

with $\delta_m$ denoting the Kronecker delta function. The amount of contributions from superimposed signals is known up to the parameters $\mathbf{\Theta}_{m'}$ and contains the nonlinearity in the model.

The unknowns in this model are the spectral line parameters $\mathbf{\Psi}_{m'} = [\mathbf{\Theta}_{m'}, f_{m'}]$ satisfying the constraints $\mathbf{\Psi}_{m'} \in \Omega$. The constraint set $\Omega$ equals to $\Lambda \times \Pi$, where $\Lambda$ and $\Pi$ denote the constraint sets for $\mathbf{\Theta}_{m'}$ and $f_{m'}$, respectively: $\Lambda = \{(\epsilon_{m'}, \Delta_{m'}) \in \mathbb{R} \times \mathbb{R}^+ : \mid \epsilon_{m'} \mid \leq \epsilon_{\max}$ and $\Delta_{\min} \leq \Delta_{m'} \leq \Delta_{\max}\}$, and $\Pi = \{f_{m'} \in \mathbb{R}^+ : f_{\min} \leq f_{m'} \leq f_{\max}\}$. (The number of superimposed signals is known, and equal to the number of pixels $M$.)

If we define the vectors $\mathbf{y}^a = [y_1^a \dots y_M^a]^\top$, $\mathbf{f} = [f_1 \dots f_M]^\top$, $\boldsymbol{\epsilon} = [\epsilon_1 \dots \epsilon_M]^\top$, and $\boldsymbol{\Delta} = [\Delta_1 \dots \Delta_M]^\top$ (with the superscript $^\top$ denoting the transpose of a vector), each of these vectors has the size $M$. Then, based on the parametric model, each dispersed image $\mathbf{y}^a$ can be viewed as a blurred version of the same input object $\mathbf{f}$ with a different spatially-varying filter of unknown parameters $\boldsymbol{\epsilon}$ and $\boldsymbol{\Delta}$. On the other hand, when the order is zero, there is no dispersion and hence no blur on the input image; that is $\mathbf{y}^0 = \mathbf{f}$.

We note that continuum background is neglected in this model. This requires that either the background is relatively small compared to the sum of the contributions of the dispersed spectral lines, or it can be subtracted from the measurements through pre-processing. Moreover, each dispersed image is assumed to be monochromatic, hence due to a single spectral line with known central wavelength. This will be approximately true either when one spectral line is dominant (i.e. strongest) in the passband of the instrument (i.e. the spectrum filtered by a spectral filter is dominated by a single spectral line), or when the FOV is limited such that adjacent spectral lines with different wavelengths do not cause spatial extent of each monochromatic image to overlap on the detector. Our goal in the inversion will be to estimate the

spectral line parameters associated with this single central wavelength, but over a two-dimensional FOV.

### 2.2.2 Observation model with noise

In vector-matrix form, the observation model with noise is given by

$$\tilde{\mathbf{y}}^a = \sum_{m'=1}^{M} f_{m'} \mathbf{h}^a_{m'}(\mathbf{\Theta}_{m'}) + \mathbf{n}^a \tag{2.8}$$

where

$$\mathbf{h}^a_{m'}(\mathbf{\Theta}_{m'}) = [\phi^a_{1-m'}(\mathbf{\Theta}_{m'}), \ldots, \phi^a_{M-m'}(\mathbf{\Theta}_{m'})]^\top \tag{2.9}$$

and $\mathbf{n}^a = [n^a_1 \ldots n^a_M]^\top$ is the noise vector with $n^a_m \sim N(0, \sigma^2_a)$ representing white Gaussian noise that is uncorrelated across both different pixels $m$ and orders $a$. In practice this noise model is valid when the following conditions are satisfied: first, photon noise is the dominant source of noise in the measurements rather than the thermal and readout noise (which can be ensured by sufficiently long integration time); second, a strong spectral line is measured through all pixels (so that the values of $y^a_m$ are large enough to well approximate the Poisson noise as Gaussian noise); third, signal-to-noise ratio (SNR) is sufficiently high (such that noise standard deviation can be approximated as constant over each dispersed image).

Here we are interested in the case that multiple dispersed images at different diffraction orders are simultaneously available. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ be the set of all orders that are measured with $N$ being the number of different orders, and $\{\sigma_{a_1}, \sigma_{a_2}, \ldots, \sigma_{a_N}\}$ be the corresponding noise standard deviations for these measurements. Defining the $M \times M$ matrix

$$\mathbf{H}^a(\mathbf{\Theta}) = [\mathbf{h}^a_1(\mathbf{\Theta}_1), \ldots, \mathbf{h}^a_M(\mathbf{\Theta}_M))] \tag{2.10}$$

with $\mathbf{\Theta} = [\mathbf{\Theta}_1^\top, \ldots, \mathbf{\Theta}_M^\top]^\top$, the model for each order $a$ can be rewritten more compactly as

$$\tilde{\mathbf{y}}^a = \mathbf{H}^a(\mathbf{\Theta})\, \mathbf{f} + \mathbf{n}^a \tag{2.11}$$

Then by stacking all measured dispersed images into a single vector, $\tilde{\mathbf{y}}$, the

complete model becomes

$$\tilde{\mathbf{y}} = \mathbf{H}(\boldsymbol{\Theta})\mathbf{f} + \mathbf{n} \tag{2.12}$$

where

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{\mathbf{y}}^{a_1} \\ \tilde{\mathbf{y}}^{a_2} \\ \vdots \\ \tilde{\mathbf{y}}^{a_N} \end{bmatrix}, \quad \mathbf{H}(\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{H}^{a_1}(\boldsymbol{\Theta}) \\ \mathbf{H}^{a_2}(\boldsymbol{\Theta}) \\ \vdots \\ \mathbf{H}^{a_N}(\boldsymbol{\Theta}) \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}^{a_1} \\ \mathbf{n}^{a_2} \\ \vdots \\ \mathbf{n}^{a_N} \end{bmatrix}$$

## 2.3 Inverse problem

In the inverse problem, the goal is to estimate the unknown spectral line parameters $\mathbf{f}$ and $\boldsymbol{\Theta}$ from the measurements $\tilde{\mathbf{y}}$ based on the model (2.12). This problem can be viewed as a *multiframe*, *semiblind* deblurring problem with *shift-variant* blur. Here *semiblind* refers to the fact that although the parametric form of each blur is known, the blur parameters $\boldsymbol{\Delta}$ and $\boldsymbol{\epsilon}$ are unknown and must therefore be estimated jointly with the original image $\mathbf{f}$. The term *multiframe* refers to the availability of multiple blurred images of the same object through different diffraction orders.

We formulate the inverse problem as a maximum posterior (MAP) estimation problem that incorporates prior knowledge of the statistics of the spectral line parameters. Such priors can be obtained from the measurements of existing conventional push broom spectrometers, as will be illustrated in Section 2.5.1. Incorporation of prior information helps to regularize the inverse problem, hence preventing (excessive) noise magnification that would result from over fitting to the noisy data.

Treating the parameter vectors $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ as *independent* random vectors, the MAP estimates of $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ from the measurements $\tilde{\mathbf{y}}$ are given by

$$\arg \max_{\substack{\mathbf{f}\in\Pi^M \\ [\boldsymbol{\epsilon},\boldsymbol{\Delta}]\in\Lambda^M}} p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})\, p(\mathbf{f}) p(\boldsymbol{\Delta}) p(\boldsymbol{\epsilon}) \tag{2.13}$$

where $p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ represents the conditional probability density function (pdf) of $\tilde{\mathbf{y}}$ given $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ (equivalently, the likelihood function of the unknown parameters), and $p(\mathbf{f})$, $p(\boldsymbol{\Delta})$, and $p(\boldsymbol{\epsilon})$ denote the prior distributions. These prior distributions specify the probability of each parameter indepen-

dently from the observed data; hence they describe the information we have on each parameter prior to observing the data.

On the other hand, the conditional pdf $p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ comes from the noisy observation model in (2.8) and (2.12), and has the following form:

$$p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon}) = \prod_{i=1}^{N} \frac{1}{(\sqrt{2\pi}\sigma_{a_i})^M} e^{-\frac{1}{2\sigma_{a_i}^2}||\tilde{\mathbf{y}}^{a_i} - \mathbf{H}^{a_i}(\boldsymbol{\Theta})\mathbf{f}||_2^2} \tag{2.14}$$

The role of this pdf in the estimation problem is to force the estimates of $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ to match the observation model closely.

Note that a noisy observation of $\mathbf{f}$ is available through the zeroth order image, $\tilde{\mathbf{y}}^0$. This gives an immediate statistical model for $\mathbf{f}$. Hence if $\tilde{\mathbf{y}}^0$ is observed at a sufficiently high SNR such that the conditional pdf $p(\tilde{\mathbf{y}}^0 \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ is more concentrated around the true value of $\mathbf{f}$ as compared to the prior of $\mathbf{f}$ (yielding the prior $p(\mathbf{f})$ to be effectively constant where the conditional pdf $p(\tilde{\mathbf{y}}^0 \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ is nonzero), then the prior of $\mathbf{f}$ is not necessary to yield a useful estimate. For this reason, we ignore $p(\mathbf{f})$ in the MAP formulation, which yields a simpler form of a separable nonlinear least-squares problem [22], and to a more efficient estimation algorithm (as will be discussed in the next section). However, we note that if it were desired to keep the prior $p(\mathbf{f})$ in the MAP formulation, the dynamic programming algorithm could still be derived to solve the MAP problem, but with less efficiency.

For the priors $p(\boldsymbol{\Delta})$ and $p(\boldsymbol{\epsilon})$, we assume that parameters at different pixels are independently distributed. After combining all of these together, and taking the logarithm of (2.13), the MAP estimation problem becomes

$$\min_{\substack{\mathbf{f}\in\Pi^M \\ [\boldsymbol{\epsilon},\boldsymbol{\Delta}]\in\Lambda^M}} \sum_{i=1}^{N} \frac{1}{2\sigma_{a_i}^2}||\tilde{\mathbf{y}}^{a_i} - \mathbf{H}^{a_i}(\boldsymbol{\Theta})\mathbf{f}||_2^2 - \sum_{m=1}^{M}(\log p(\Delta_m) + \log p(\epsilon_m)) \tag{2.15}$$

We can express this as

$$\min_{\substack{\mathbf{f}\in\Pi^M \\ \boldsymbol{\Theta}\in\Lambda^M}} ||\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta})\mathbf{f}||_W^2 + \sum_{m=1}^{M} \Gamma(\boldsymbol{\Theta}_m) \tag{2.16}$$

where $\Gamma(\boldsymbol{\Theta}_m) = -2\log p(\Delta_m) - 2\log p(\epsilon_m)$ is the regularization functional arising from priors, and $W$ denotes the weighted norm in (2.15) where for each diffraction order the sum of squared residuals is weighted with the re-

ciprocal of the noise variance at that order. That is, in vector-matrix form, $||.||_W^2 = (.)^*\mathbf{W}(.)$ with $\mathbf{W}$ being the inverse covariance matrix for the data, and superscript $^*$ denoting the conjugate transpose of a vector.

This inverse problem belongs to the class of *separable nonlinear least-squares problems* [22] with regularization. In these problems the observation model is a linear combination/superposition of parametrically prespecified nonlinear functions/signals. The parameters of interest can be grouped into two categories: parameters that affect the observations in a linear fashion ($\mathbf{f}$ in our case), and parameters that affect the observations in a nonlinear fashion ($\mathbf{\Theta}$ in our case). We note that implicit in the formulation of the estimation problem is that (2.16) has a unique global minimum.

## 2.4 Dynamic programming algorithm

We now focus on developing an efficient and globally converging algorithm for solving our MAP problem. Note that this MAP estimation problem requires solving a nonlinear and nonconvex optimization problem with possibly local minima. (Nonconvexity of the problem is apparent since the Hessian matrix of the objective function is not positive semidefinite for all feasible points [23].) This can create difficulty in efficiently finding the global minimum of the problem using local optimization methods such as gradient-descent type methods [24] and expectation-maximization type algorithms [25]. This is because such methods converge to one of the local minima depending on the initialization. Many of the methods proposed for the general separable nonlinear least-squares problems [22] and for problems involving superimposed signals (see, for example, [26–28]) are such local optimization methods.

While efficient local optimization methods may suffer from convergence to local minima, global optimization methods can guarantee convergence to the global solution. There are two types of *general-purpose* global optimization methods: deterministic versus stochastic. The deterministic global optimization methods (such as exhaustive search, branch and bound method [29]) can guarantee convergence to a global solution within a certain tolerance value; but it is not practical to employ these methods in most applications because of their high computational complexity [30]. On the other hand, stochastic

methods (such as simulated annealing) lower the computational cost in return for weaker guarantees for global convergence (in probabilistic sense) [31].

Here we develop an efficient global optimization method, *specialized* for our problem, which combines the strengths of deterministic and stochastic (global optimization) methods: global convergence guarantee within a certain tolerance value, as the deterministic approaches, and lower computational complexity, as the stochastic approaches. The key idea in this method is to perform a computationally efficient search that is equivalent to exhaustive search by exploiting the special form of the objective function to optimize, which arises from the limited interaction of superimposed signals (i.e. Gaussian line profiles) with few of their closest neighbors. This special form, so-called as *local interaction* in [17], yields to a Markovian-like property of the globally optimal solutions (in the sense of deterministic dependence, rather than statistical dependence). This allows us to break the optimization problem into smaller subproblems, which are then recursively solved by performing an exhaustive search in a reduced space. The resulting dynamic programming (DP) algorithm has a computational cost that is linear in the number of superimposed signals as opposed to the exponential cost in exhaustive search.

### 2.4.1 Local interaction signal model

The development of the dynamic programming algorithm relies on one major assumption: local interaction, implying that each superimposed signal interacts (overlaps) with only a few of its closest neighbors. Let $r \geq 1$ be the number of closest neighbors with which each superimposed signal overlaps on both sides. Then, mathematically the *local interaction* model [17] is expressed as

$$\mathbf{h}_i(\boldsymbol{\Theta}_i)^* \mathbf{h}_j(\boldsymbol{\Theta}_j) \approx 0 \text{ for } |i-j| > r \tag{2.17}$$

This requires that the $i$th and $j$th columns of $\mathbf{H}(\boldsymbol{\Theta})$, denoted by $\mathbf{h}_i(\boldsymbol{\Theta}_i)$ and $\mathbf{h}_j(\boldsymbol{\Theta}_j)$ (associated with the $i$th and $j$th superimposed signals), are approximately orthogonal if they are separated by more than $r$ columns. (Note that the $i$th column of $\mathbf{H}(\boldsymbol{\Theta})$, $\mathbf{h}_i(\boldsymbol{\Theta}_i)$, contains the contributions of the $i$th superimposed signal to measurements at all detector pixels.)

Suitability of this local interaction model to our problem follows from the

Gaussian shape of superimposed signals and Cramer-Rao bound analysis. Note that in our observation model, superimposed signals are Gaussian line profiles, each centered around a different pixel on the detector (hence most of its energy is concentrated around that pixel). Therefore, for each Gaussian, the interaction (overlap) is limited to the closest neighboring Gaussians and is determined by the width of the Gaussian, which itself is determined by the amount of dispersion in the instrument (i.e. higher dispersion results in wider width, hence larger interaction). Therefore, the extent of interaction, $r$, depends on the amount of dispersion, which is a design choice for the instrument.

On the other hand, a study of the Cramer-Rao bound, a lower bound on the error standard deviation of unbiased estimators [32], reveals that large dispersion (such that more than a few Gaussians overlap with each other) is not an optimal design choice because it results in significantly large errors in the parameter estimates [19]. That is, useful instrument models can be restricted without loss of generality to Gaussian line profiles which interact only with few of their closest neighbors (typically, $r \leq 5$).

A more general discussion of suitability of the local interaction model to a wide class of separable least-squares problems has been given in [17]. It was shown, based on Cramer-Rao bound analysis, that in many instances superimposed signals interacting with more than a few neighbors cannot be separated to any meaningful accuracy; hence useful models can be restricted to those with local interaction.

Other than the amount of dispersion in the instrument, there are other factors that affect the choice of $r$ from the algorithmic point of view. As will be discussed in the next two sections, the choice of $r$, which impacts the approximation in (2.17), provides a mechanism for making a tradeoff between the accuracy of DP estimates (in terms of closeness to the desired MAP estimates) and computational complexity.

### 2.4.2 Dynamic programming algorithm

The dynamic programming algorithm presented in this section is an extension of a previously proposed method for maximum likelihood parameter estimation of superimposed signals [17,18]. This algorithm was presented for

the maximum likelihood problem and when each superimposed signal interacts with only one neighbor on both sides. Here we extend the algorithm to the MAP framework (that involves priors) and to superimposed signals interacting with arbitrary number of neighbors. A preliminary version of the extended MAP algorithm was presented in our paper [9].

For simplicity and without loss of generality, we ignore the weights in the least-squares term of the MAP functional in (2.16), and treat the problem with identical weights of unity. The more general case can be simply handled within this framework after scaling each measurement vector ($\tilde{\mathbf{y}}^{a_i}$) and measurement matrix ($\mathbf{H}^{a_i}(\mathbf{\Theta})$) by the standard deviation of the corresponding measurement noise ($\sigma_{a_i}$).

The dynamic programming algorithm breaks the MAP optimization problem into smaller subproblems that are related to each other recursively. This recursive multistage optimization process is enabled by the local interaction model as follows: for any parameter set $(\mathbf{f}, \mathbf{\Theta})$, any extent of interaction $r$ with $1 \leq r \leq M-1$, and any pixel $k$ in the range $1 \leq k < M-r$, the objective function in the MAP formulation can be decomposed as follows:

$$
\begin{aligned}
&\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta})\mathbf{f}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m) \\
=&\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 \\
&+\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]})\mathbf{f}_{[k+r+1:M]}\|^2 \\
&-\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m) \\
&+2\mathrm{Re}\{\mathbf{f}^*_{[1:k]}\mathbf{H}^*(\mathbf{\Theta}_{[1:k]})\mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]})\mathbf{f}_{[k+r+1:M]}\} \\
\approx&\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 \\
&+\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]})\mathbf{f}_{[k+r+1:M]}\|^2 \\
&-\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m) \\
\triangleq&\tilde{J}(\mathbf{\Psi})
\end{aligned}
$$

where $\mathbf{\Theta}_{[i:j]}$ denotes $[\mathbf{\Theta}_i\, \mathbf{\Theta}_{i+1}\, \ldots\, \mathbf{\Theta}_j]$, $\mathbf{f}_{[i:j]}$ denotes the corresponding similar representation, and $\mathbf{H}(\mathbf{\Theta}_{[i:j]})$ is a submatrix composed of $i$th to $j$th columns of $\mathbf{H}(\mathbf{\Theta})$. The approximate equality holds from the local interaction assumption in (2.17); hence, when the local interaction model holds, solving the MAP

problem is equivalent to minimizing $\tilde{J}(\mathbf{\Psi})$.

This decomposed objective function has the generic functional form of

$$\tilde{J}(\mathbf{\Psi}) = \tilde{J}_1(\mathbf{\Psi}_{[1:k]}, \mathbf{\Psi}_{[k+1:k+r]}) + \tilde{J}_2(\mathbf{\Psi}_{[k+1:k+r]}, \mathbf{\Psi}_{[k+r+1:M]}) \tag{2.18}$$

with $\mathbf{\Psi}_i = (\mathbf{\Theta}_i, f_i)$, where the function $\tilde{J}_1(.)$ contains the first term of $\tilde{J}(\mathbf{\Psi})$, and $\tilde{J}_2(.)$ contains the next two terms, in addition to the prior terms. This form enables us to efficiently find the global optimum of $\tilde{J}(\mathbf{\Psi})$ via dynamic programming [33–35]. This is because given $\mathbf{\Psi}_{[k+1:k+r]}$ for any $k$, the variables $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_k$ and $\mathbf{\Psi}_{k+r+1}, \ldots, \mathbf{\Psi}_M$ are decoupled. As a result, if $\tilde{J}(\mathbf{\Psi})$ is optimized for a given $\mathbf{\Psi}_{[k+1:k+r]}$, then the optimal values of $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_k$ are a function of only $\mathbf{\Psi}_{[k+1:k+r]}$, and hence can be denoted as $\mathbf{\Psi}^*_{[1:k]}(\mathbf{\Psi}_{[k+1:k+r]})$, and obtained by optimizing only $\tilde{J}_1(.)$. This property of globally optimal solutions is similar to the Markov property of random processes (where in our case deterministic dependence replaces the role of statistical dependence).

This shows that our problem satisfies the *principle of optimality* of the theory of dynamic programming [33]: subsets of an optimal solution of the original problem are themselves optimal solutions to its subproblems. This allows us to efficiently solve the high-dimensional problem by solving smaller subproblems that are related to each other recursively. More specifically, if we define the $k$th subproblem as finding $\mathbf{\Psi}^*_{[1:k]}(\mathbf{\Psi}_{[k+1:k+r]})$ for any given $\mathbf{\Psi}_{[k+1:k+r]}$, then it can be solved by using the solution of the $(k-1)$th subproblem:

$$\mathbf{\Psi}^*_{[1:k]}(\mathbf{\Psi}_{[k+1:k+r]}) = \arg \min_{\substack{\mathbf{\Psi}_k \in \Omega \\ \mathbf{\Psi}_{[1:k-1]} \in \mathbf{\Psi}^*_{[1:k-1]}(\mathbf{\Psi}_{[k:k+r-1]})}} \tilde{J}_1(\mathbf{\Psi}_{[1:k]}, \mathbf{\Psi}_{[k+1:k+r]}) \tag{2.19}$$

This limits the search for $\mathbf{\Psi}_{[1:k-1]}$ to a reduced set given by the solution of the $(k-1)$th subproblem, and hence yields a significant computational gain over the exhaustive search of the original problem. (Indeed, if the $(k-1)$th subproblem has a unique solution, this reduced set contains only one solution.) The global minimum of $\tilde{J}(\mathbf{\Psi})$ can then be computed recursively through $M-r$ stages, where at the $k$th stage the $k$th subproblem is solved through recursion, while $k$ increases from 1 to $M-r$. Here we note that an alternative extension of [17] to $r > 1$ case can perform the recursion differently by solving the $k$th subproblem using the solution of the $(k-r)$th subproblem, but this would result in higher computational cost.

As a final observation, we note that each subproblem can also be simplified. Explicitly, the $k$th subproblem is

$$\min_{\substack{\boldsymbol{\Theta}_{[1:k]}\in\Lambda^k \\ \mathbf{f}_{[1:k]}\in\Pi^k}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 + \sum_{m=1}^{k}\Gamma(\boldsymbol{\Theta}_m) \quad (2.20)$$

Here, the minimization over $\boldsymbol{\Theta}_{[1:k]}$ can be solved separately by eliminating $\mathbf{f}_{[1:k]}$ from (2.20) based on the *variable projection* technique of separable nonlinear least-squares problems [22]. This results in the following equivalent problem:

$$\min_{\boldsymbol{\Theta}_{[1:k]}\in\Lambda^k} \|\mathbf{P}^{\perp}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:k]})}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}]\|^2 + \sum_{m=1}^{k}\Gamma(\boldsymbol{\Theta}_m) \quad (2.21)$$

where $P^{\perp}_{\mathbf{A}} = I - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ is the projection matrix onto the orthogonal complement of the column space of $\mathbf{A}$.

With all these observations, the steps in the dynamic programming algorithm are summarized below [9].

1. *Initialization stage* ($k = 1$)*:*

   (a) For each $(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]}) \in \Omega^r$, solve the following problem

   $$\hat{\boldsymbol{\Theta}}_{[1:1]}(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]}) = \arg\min_{\boldsymbol{\Theta}_1\in\Lambda} \|\mathbf{P}^{\perp}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:1]})}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[2:1+r]})\mathbf{f}_{[2:1+r]}]\|^2 + \Gamma(\boldsymbol{\Theta}_1)$$

   through exhaustive search over $\boldsymbol{\Theta}_1 \in \Lambda$.

   (b) Record the optimal values as a function of $\boldsymbol{\Theta}_{[2:1+r]}$:

   $$\boldsymbol{\Theta}^*_{[1:1]}(\boldsymbol{\Theta}_{[2:1+r]}) = \{\boldsymbol{\Theta}_{[1:1]} \in \Lambda : \boldsymbol{\Theta}_{[1:1]} = \hat{\boldsymbol{\Theta}}_{[1:1]}(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]}) \text{ for some } \mathbf{f}_{[2:1+r]} \in \Pi^r\}$$

2. *Update stages* ($k = 2, \ldots, M - r$)*:*

(a) For each $(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) \in \Omega^r$, solve the following problem

$$\hat{\boldsymbol{\Theta}}_{[1:k]}(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) = \arg \min_{\substack{\boldsymbol{\Theta}_k \in \Lambda \\ \boldsymbol{\Theta}_{[1:k-1]} \in \boldsymbol{\Theta}^*_{[1:k-1]}(\boldsymbol{\Theta}_{[k:k+r-1]})}} \|\mathbf{P}^{\perp}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:k]})}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 + \sum_{m=1}^{k} \Gamma(\boldsymbol{\Theta}_m)$$

through exhaustive search over $\boldsymbol{\Theta}_{[1:k-1]} \in \boldsymbol{\Theta}^*_{[1:k-1]}(\boldsymbol{\Theta}_{[k:k+r-1]})$ and $\boldsymbol{\Theta}_k \in \Lambda$.

(b) Record the optimal values as a function of $\boldsymbol{\Theta}_{[k+1:k+r]}$:

$$\boldsymbol{\Theta}^*_{[1:k]}(\boldsymbol{\Theta}_{[k+1:k+r]}) = \{\boldsymbol{\Theta}_{[1:k]} \in \Lambda^k : \boldsymbol{\Theta}_{[1:k]} = \hat{\boldsymbol{\Theta}}_{[1:k]}(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) \text{ for some } \mathbf{f}_{[k+1:k+r]} \in \Pi^r\}$$

3. *Final stage:*

(a) To obtain the final estimate of $\boldsymbol{\Theta}$, solve the following problem

$$\hat{\boldsymbol{\Theta}} = \arg \min_{\substack{\boldsymbol{\Theta}_{[M-r+1:M]} \in \Lambda^r \\ \boldsymbol{\Theta}_{[1:M-r]} \in \boldsymbol{\Theta}^*_{[1:M-r]}(\boldsymbol{\Theta}_{[M-r+1:M]})}} \|\mathbf{P}^{\perp}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:M]})}\tilde{\mathbf{y}}\|^2 + \sum_{m=1}^{M} \Gamma(\boldsymbol{\Theta}_m) \tag{2.22}$$

through exhaustive search over $\boldsymbol{\Theta}_{[M-r+1:M]} \in \Lambda^r$ and $\boldsymbol{\Theta}_{[1:M-r]} \in \boldsymbol{\Theta}^*_{[1:M-r]}(\boldsymbol{\Theta}_{[M-r+1:M]})$.

(b) Estimate of $\mathbf{f}$ is then given by

$$\hat{\mathbf{f}} = [\mathbf{H}^*(\hat{\boldsymbol{\Theta}})\mathbf{H}(\hat{\boldsymbol{\Theta}})]^{-1}\mathbf{H}^*(\hat{\boldsymbol{\Theta}})\tilde{\mathbf{y}} \tag{2.23}$$

The relation of the dynamic programming algorithm to the MAP problem is stated in the following theorem. This theorem is a generalization of Theorem 5 of [17].

**Theorem 1.** *If the local interaction model holds exactly for some $r \geq 1$, i.e.*

$$\mathbf{h}_i(\boldsymbol{\Theta}_i)^*\mathbf{h}_j(\boldsymbol{\Theta}_j) = 0 \text{ for } |i-j| > r \tag{2.24}$$

*for all $i, j = 1, \ldots, M$, then the estimates obtained with the dynamic programming algorithm (with this value of $r$) are same as the MAP estimates*

*obtained by solving* (2.16).

*Proof.* See Appendix. □

This theorem shows that the exact MAP estimates can be obtained with the DP algorithm when the local interaction model holds exactly (i.e. with exact orthogonality). However, for our spectral imaging problem, exact orthogonality as in (2.24) is not possible because of the Gaussian nature of the overlapping signals, hence there is some unavoidable deviation from exact orthogonality. Fortunately, it has been shown, under some regularity conditions, that the dynamic programming algorithm is robust in the sense that deviations from exact orthogonality continuously perturb the DP estimates from the exact MAP estimate, and moreover, the resulting deviation from the exact MAP estimate is upper-bounded by a constant that is proportional to the deviation from exact orthogonality [18]. Therefore, for any well-conditioned problem, if the deviation from orthogonality is small enough, then the DP estimates are close to the desired MAP estimate. As a result, the accuracy of the DP estimates is controlled by the amount of deviation from exact orthogonality, which is indeed controlled by the choice of $r$. Therefore, as the value of $r$ is increased, the accuracy of DP estimates will be improved. In practice, the DP estimates can be used as initialization for a local optimization method to obtain the exact global MAP estimate.

As a final remark, we note that this parameter estimation algorithm is quite general, and it can be applied to other problems involving different superimposed signals and priors. Superimposed signal models and the resulting separable nonlinear least-squares problems are of wide interest in various applications such as sensor array processing, communications, imaging, robotics, and vision [22]. Two commonly encountered problems are estimation of frequency and amplitude of superimposed sinusoids, and estimation of position, width, and amplitude of overlapping pulses of given shape (as our problem) [17]. The dynamic programming algorithm is applicable to any such separable nonlinear least-squares problem for which the local interaction signal model is suited.

### 2.4.3 Computational aspects

We now consider the computational requirements of the generalized dynamic programming algorithm. Note that the nonconvex minimization problem at each stage is solved through exhaustive search over the parameter space restricted by the constraint sets. This requires discretization of the search space to a finite number of parameter values. Let $q$ be the number of quantization levels used in exhaustive search for each scalar parameter, and $n$ and $p$ be the number of scalar parameters in each $\mathbf{\Theta}_m$ and $f_m$, respectively. (In our problem, we have $n = 2$ and $p = 1$ with the parameters in $\mathbf{\Theta}_m$ being $\Delta_m$ and $\epsilon_m$.)

For the minimization problem at each stage, we need to evaluate objective functions of the form

$$\|\mathbf{P}^{\perp}_{\mathbf{H}(\tilde{\Theta}_1)}(\tilde{\mathbf{y}} - \mathbf{H}(\tilde{\Theta}_2)\mathbf{f})\|^2 + \Gamma((\tilde{\Theta}_1)) \tag{2.25}$$

We will simply state the computational requirement in terms of the number of function evaluations of this form (although these function evaluations have different costs at different stages).

At the $k$th stage, the objective function needs to be evaluated for all possible values of $\mathbf{\Theta}_k$ and $\mathbf{\Theta}_{[1:k-1]}$. Assuming that each subproblem has a unique solution, the recorded set $\mathbf{\Theta}^*_{[1:k-1]}(\mathbf{\Theta}_{[k:k+r-1]})$ has at most the size of the vector $\mathbf{f}_{[k:k+r-1]}$; therefore, there are at most $(q^p)^r$ different values for $\mathbf{\Theta}_{[1:k-1]}$. With $q^n$ possible values for $\mathbf{\Theta}_k$, the objective function at the $k$th stage needs to be evaluated $q^{rp+n}$ times. Moreover, this exhaustive search is repeated for every possible value of $(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})$, hence $q^{(p+n)r}$ times. Because there are a total of $M-r-1$ update stages, the total number of objective function evaluations at these stages is $(M-r-1) \times q^{(p+n)r} \times q^{rp+n} = (M-r-1)q^{r(2p+n)+n}$. Using a similar argument, the initialization and final stages require $q^{r(p+n)+n}$ and $q^{r(p+n)}$ function evaluations, respectively. Therefore, it follows that the total computational effort of the dynamic programming algorithm is of $\mathcal{O}(q^{r(2p+n)+n})$, while the exhaustive search of the original problem over the entire parameter space is of $\mathcal{O}(q^{M(p+n)})$. Hence the computational cost is exponential only in the number of interacting signals, $r$, while linear in the total number of superimposed signals, $M$, as opposed to the exponential cost in $M$ in exhaustive search of the original problem. With typically $M \gg r$,

this shows the computational efficiency of the dynamic programming algorithm compared to the exhaustive search of the original problem over the entire parameter space.

As mentioned before, there exists a bounded discrepancy between DP estimates and the global MAP estimates because of the approximation in the local interaction model. A second source of discrepancy will arise from performing the optimization (exhaustive search) at each stage over a discretized parameter space (rather than over continuous values of the parameters). Clearly, the discretization needs to be fine enough to remain close to the desired global MAP estimates. In practice, a local optimization method initialized with DP estimates will be used subsequently in order to refine these estimates and obtain the exact global MAP estimate.

As a final remark, we note that the computational complexity of the DP algorithm can be further reduced through parallel implementations of the dynamic programming algorithm [36], or through the approximate version of the algorithm [37] which has significantly lower computational cost.

## 2.5 Sample application in solar spectral imaging

In this section, we illustrate the performance of the instantaneous spectral imaging technique and the MAP estimation framework for an application in solar spectral imaging [7]. For this, we consider a prominent solar emission line in the extreme ultraviolet (EUV) regime, with a central wavelength of $\lambda_0 = 195.12$ Å. Our goal is to estimate the parameters of this emission line (consisting of integrated intensity, line width and Doppler shift parameters) within a two-dimensional FOV from the observations of the instantaneous spectral imager. These emission line parameters yield estimates of the physical parameters of the solar plasma (such as the temperature, density, and flow speed of the ion emitting this spectral line), and hence enable the investigation of the dynamic plasma behavior.

### 2.5.1 Estimation of the prior distributions

To apply the MAP approach, we need to specify the prior distributions of line widths, $\Delta_m$, and Doppler shifts, $\epsilon_m$. The choice of these priors is application-

dependent. Formerly, parameters at different pixels are assumed to be independently distributed. Here we further treat them as having the same distribution, hence as independent and identically distributed random variables from pixel-to-pixel. Therefore two density distributions, one for line widths $\Delta_m$ and one for Doppler shifts $\epsilon_m$, need to be estimated.

For this, we use observations obtained with a conventional push broom (slit) spectrometer [38]. Each observation with a slit is associated with a 1D portion of the scene admitted through the slit; hence line widths and Doppler shifts obtained from a slit data can be viewed as 1D realizations of $\Delta_m$ and $\epsilon_m$ over a column of pixels. Figure 2.3 shows the histograms of line widths and Doppler shifts obtained from a large set of slit data, where each data is obtained at a different time and different slit position (corresponding to different 1D realizations). For density estimation, the histograms are normalized by the number of total observations (so that bin counts sum to one). Here we note that the parameters in the histograms are shown in pixel units, rather than in physical units, in order to match the units in the parametric model. The implicit step in this conversion is discussed in [19].

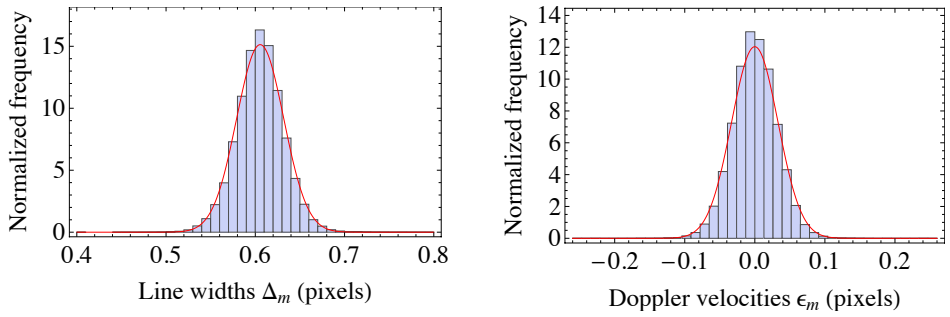



Figure 2.3: Normalized histograms of line widths and Doppler velocities for the solar spectral imaging application. Fitted distributions are shown in red on top of the histograms.

Because parameter values are clustered around one value in the histograms, Gaussian distribution is used to model their density distributions:

$$\Delta_m \sim N(\mu_\Delta, \sigma_\Delta^2), \tag{2.26}$$

$$\epsilon_m \sim N(\mu_\epsilon, \sigma_\epsilon^2) \tag{2.27}$$

with each parameter i.i.d. over pixels, $(\mu_\Delta, \sigma_\Delta^2)$ denoting the mean and variance of the distribution of line widths, and $(\mu_\epsilon, \sigma_\epsilon^2)$ denoting the mean and variance of the distribution of Doppler shifts. These mean and variance parameters are respectively estimated from the data using the maximum likelihood approach, which are then given by the sample mean and sample variance [39]. Gaussian distributions with these estimated parameters are shown in red on top of the histograms. The resulting prior terms to be used in the MAP estimation are given by

$$\Gamma(\mathbf{\Theta}_m) = \frac{(\Delta_m - \mu_\Delta)^2}{\sigma_\Delta^2} + \frac{(\epsilon_m - \mu_\epsilon)^2}{\sigma_\epsilon^2} \tag{2.28}$$

for all $m = 1, \ldots, M$.

### 2.5.2 Numerical results

Computer simulation results are presented to demonstrate the effectiveness of the parametric MAP approach for estimating the spectral line parameters from the measurements of instantaneous spectral imager. For this, we work with the simulated measurements of the instantaneous spectral imager for the solar application. We consider a column of pixels of length 50 on a detector (i.e. $M = 50$). Spectral line parameters $f_{m'}$, $\Delta_{m'}$, and $\epsilon_{m'}$ associated with each pixel $m'$ are randomly and independently generated according to their prior probability distributions (given in Section 2.5.1) for $m' = 1, 2, \ldots, 50$. Then the measurement of the spectral imager, $\tilde{\mathbf{y}}^a$, along this detector column is simulated based on the parametric model in equation (2.11) (as the superposition of Gaussian line profiles with these spectral line parameters). Such simulated measurements are obtained for three orders $a \in \{0, +1, -1\}$. Also for each order, the additive noise term, $\mathbf{n}^a$, is randomly and independently generated according to Gaussian distribution, where each component has zero mean and variance of $\sigma^2$. Fig. 2.4 shows an example of the resulting noisy measurements with a noise standard deviation of $\sigma = 2$.

In order to estimate the spectral line parameters from these noisy measurements, the dynamic programming algorithm is used with the extent of interaction $r = 2$, hence with the model that each Gaussian signal interacts with its two closest neighbors on both sides. To define the constraint

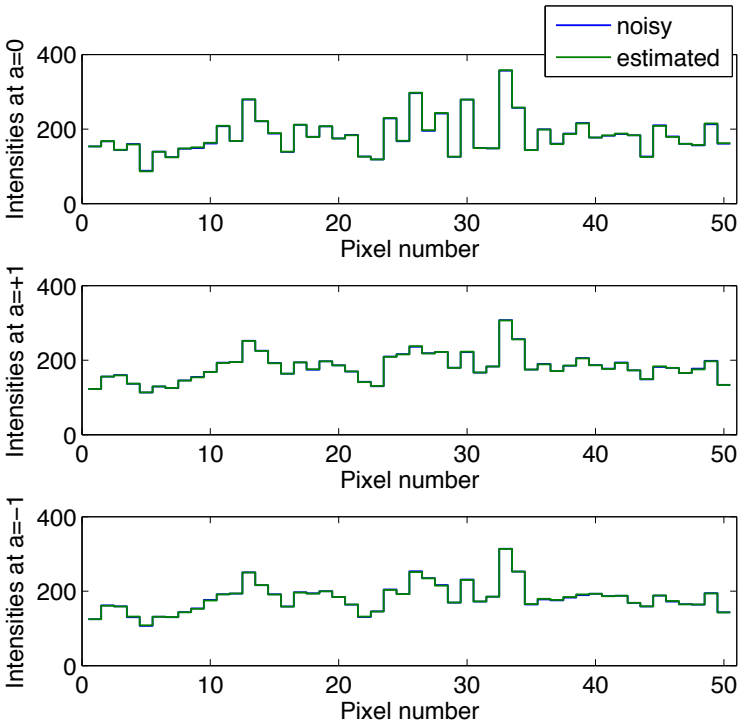



Figure 2.4: Noisy observations and the corresponding fit along one column of the fifty-pixel detector.

sets involved, line width and Doppler shift parameters are restricted to the ranges observed in the histograms (see Fig. 2.3). Constraining all integrated intensity parameters to a single range likewise would require a large amount of quantization levels and hence computational load, because the dynamic range of intensities over all pixels is large (see Fig. 2.4). Instead, a different range is assigned to each integrated intensity parameter at a different pixel. In particular, each intensity parameter is constrained to lie around the value of the zeroth order measurement at that pixel, since the zeroth order measurements are Gaussian distributed around the true integrated intensities.

The parameter space restricted by these ranges must also be discretized (to a finite number of values) for the exhaustive search in the DP algorithm. A straightforward option is uniform discretization [17] of each range where the number of quantization levels is chosen based on the Cramer-Rao error bounds of the parameters [19]. Instead, here we choose a nonuniform quantization grid to take into account the normal distribution of the parameters.

More specifically, the grid is designed by dividing each parameter range to regions of equal probability, rather than of equal length. The resulting grid is more dense around the mean of the parameter, where most of its realizations will lie.

The estimates obtained with the dynamic programming algorithm are refined by a gradient-based interior-point algorithm (a local optimization method) applied to the MAP problem. For the evaluation of the estimates, estimated parameters are compared with the true parameter values by using the root-mean-square (RMS) error: $\sqrt{\sum_{m=1}^{M}(f_m - \hat{f}_m)^2/M}$ (similarly for $\boldsymbol{\Delta}$ and $\boldsymbol{\epsilon}$).

Figure 2.5 shows the estimates of integrated intensities, line widths, and Doppler shifts obtained from the observations in Fig. 2.4. As illustrated in Fig. 2.4, the estimated parameters yield estimated observations that are almost same as the given observations; RMS errors between given and estimated observations are typically less than 1 for all orders 0, +1, and −1. Moreover, RMS errors for the parameter estimates are typically less than 2 for intensities, and 0.02 (pixels) for line widths and Doppler shifts. When converted to the physical units, this estimation with as low as three measured orders has the same order of accuracy as the state-of-the-art slit spectroscopy used for this application [38], which suffers from the limitation of a 1D FOV. Note that measuring more than three orders can help further to reduce the errors in the parameter estimates. The quantification of the amount of improvement with additional orders is a topic of future study.

To evaluate the performance of the parametric MAP approach further, we investigate the effect of the noise standard deviation (hence SNR) on the estimation accuracy of the spectral line parameters. For this, Monte Carlo simulations are performed for a total of 40 random parameter sets, and the numerical averages of RMS errors from these runs are computed for cases with varying noise standard deviation.

Fig. 2.6 shows the average RMS errors of the parameter estimates as a function of noise standard deviation. To understand the improvement in the accuracy of estimates as compared to a trivial estimate where all parameters are set to their known mean values without any estimation (more specifically, line width and Doppler shift estimates are set to mean values in their prior distributions and integrated intensity estimates are set to the zeroth order measurements), the RMS error of this trivial estimate is also shown in the

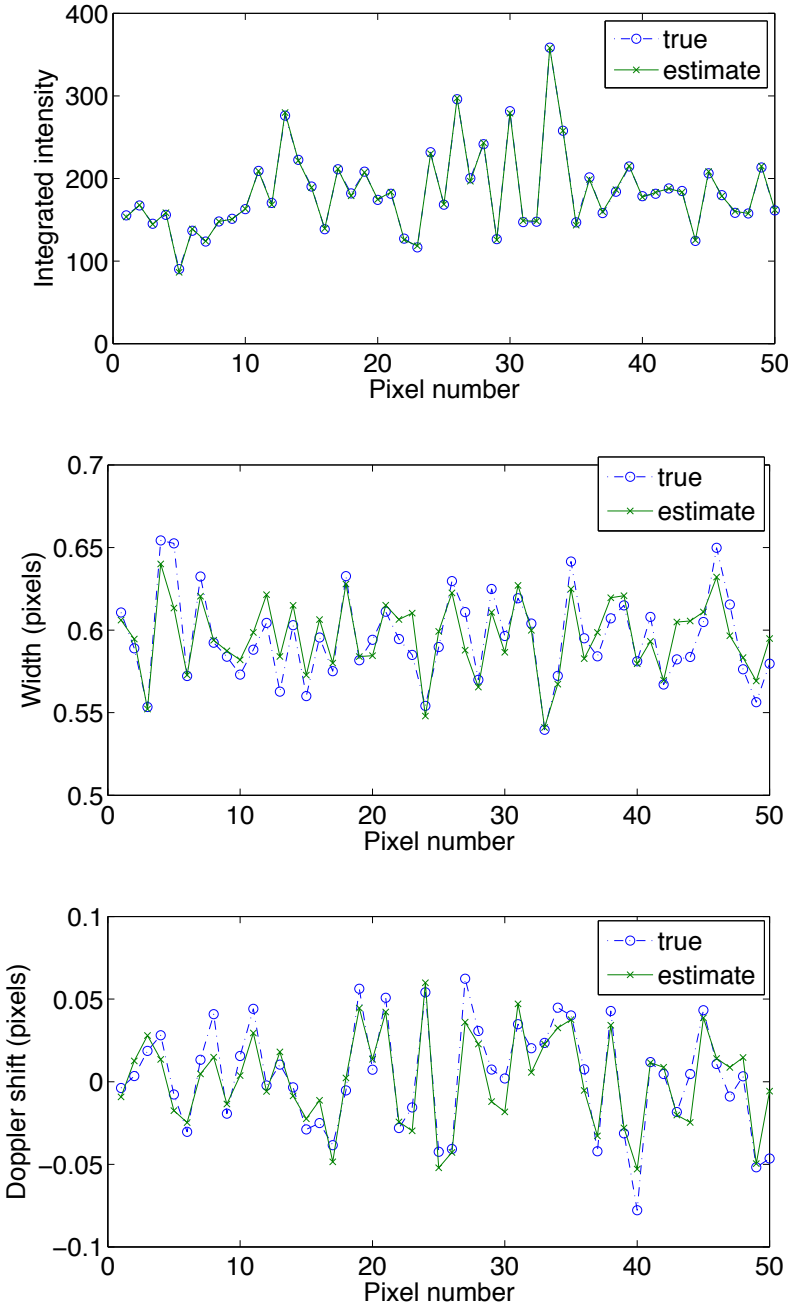



Figure 2.5: Estimates of integrated intensities, line widths, and Doppler shifts obtained with the dynamic programming algorithm, for observing the EUV solar emission line with $\{0, +1, -1\}$ orders, and noise standard deviation of 2. Blue and green lines correspond to true and estimated parameters, respectively.

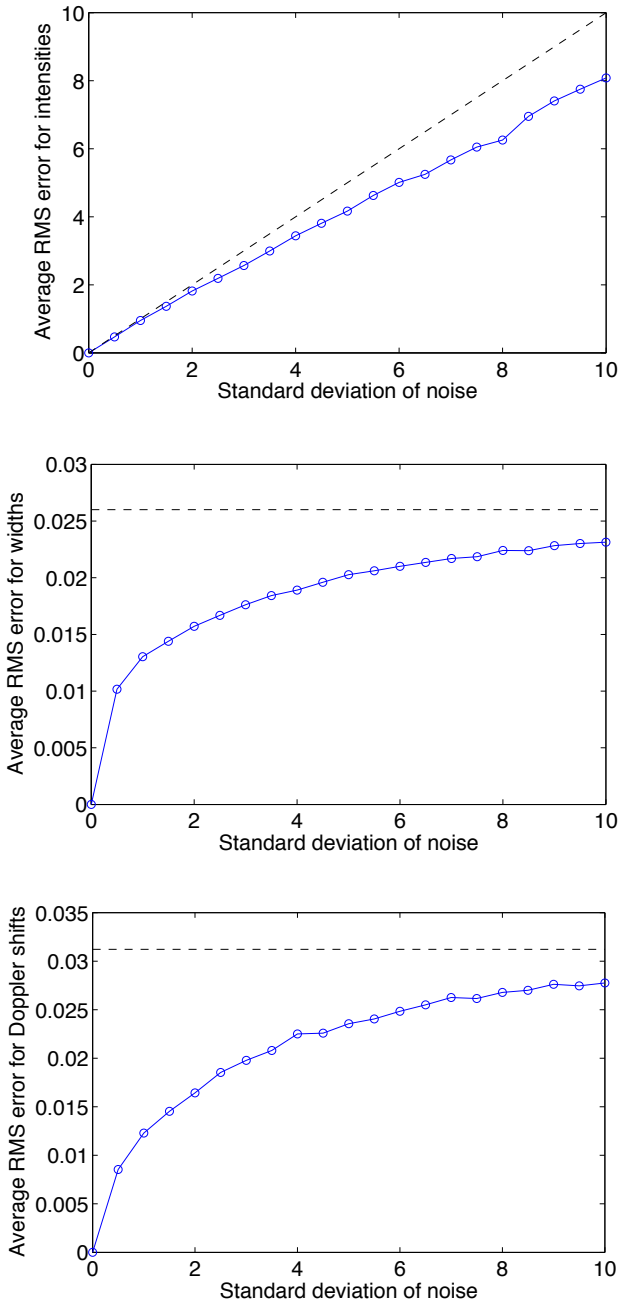



Figure 2.6: RMS errors for the estimates of intensities, widths, and Doppler shifts as a function of the noise standard deviation when $\{0, +1, -1\}$ orders are measured along a 50-pixel detector column (i.e. $M = 50$). For comparison, dashed lines show the RMS errors of the trivial estimates where all parameters are set to their known mean values, without any estimation.

figures with a dashed line. As seen, for the noise-free case with $\sigma = 0$, the true parameter values are always obtained. Moreover, the errors for integrated intensities strongly depend on the noise std since the zeroth order measurement directly provides a noisy observation of integrated intensities. In fact, for the low noise regime (up to a noise std of 5) the intensity estimates obtained with the DP algorithm do not show significant improvement over the zeroth order measurement.

For the line widths and Doppler shifts, the dependence on the noise std is weaker at the high noise regime. This is because, in this regime, the estimation is highly dominated by the priors (rather than the measurements). Also we note that the estimation accuracy is comparable to the slit spectroscopy when the noise std is smaller than 4 (corresponding to an SNR of $\sim 50$ when SNR is defined as the ratio of the signal mean to the standard deviation of the noise). To achieve similar accuracy at higher noise levels, more spectral orders (than three) will be needed.

## 2.6 Conclusion

We have presented a new spectral imaging modality with a slitless configuration that admits two-dimensional instantaneous FOV. In this instantaneous spectral imaging technique, spectrally dispersed images of a two-dimensional scene are simultaneously measured in several diffraction orders. The parameters of the spectral lines (within the scene) are then estimated by using these measurements with a parametric model and by solving the resultant inverse problem computationally. The associated inverse problem can be viewed as a multiframe semiblind deblurring problem with shift-variant blur, and is tackled here by using a MAP estimation framework where the prior distributions of the spectral line parameters are estimated from the measurements of existing slit spectrometers. An efficient dynamic programming algorithm is developed to find the global optimum of the resulting nonconvex MAP problem. This algorithm yields parameter estimates that are close to the global optimum of the MAP problem, which can then be refined by using a local optimization method.

We have investigated the application of the technique in solar spectral imaging. Computer simulation results suggest that spectral line parame-

ters can be estimated with the same order of accuracy as the conventional slit spectroscopy, but with the added benefit of providing an instantaneous two-dimensional field-of-view. Moreover, this estimation accuracy is achievable with as low as three dispersed images. This illustrates the advantage of the parametric approach over the tomographic approaches [10–14] which generally require significantly larger number of dispersed images.

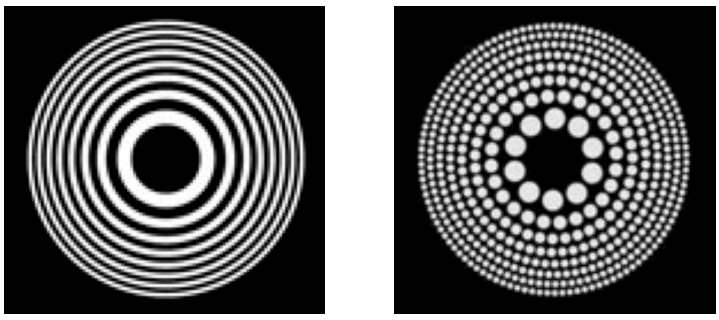
To conclude, this parametric approach to spectral imaging offers the means for effective estimation of spectral line parameters over an instantaneous two-dimensional FOV. The estimated spectral line parameters can be used to infer physical parameters of a radiating medium (such as the temperature, density, and flow speed of the particles involved in the radiation). These inferred parameters enable the investigation of the dynamic behavior by revealing how particles and heat flow through the radiating medium. Such a capability resulting from the presented technique is particularly useful for studying the spectra of dynamic scenes in a wide variety of space remote sensing applications.

# CHAPTER 3

# HIGH-RESOLUTION SPECTRAL IMAGING WITH PHOTON-SIEVES

## 3.1 Introduction

A photon sieve is a modification of a Fresnel zone plate in which open zones are replaced by a large number of circular holes (see Fig. 3.1). It has been proposed as a superior image forming device than the Fresnel zone plate [42], to be especially used at UV and x-ray wavelengths where refractive lenses are not available due to strong absorption of materials, and reflective mirrors are difficult to manufacture to achieve near diffraction-limited resolution. In fact, at these shorter wavelengths, surface roughness and figure errors often limit resolution of reflective optics to a level that is significantly lower than the diffraction limit [2-4].



(a) (b)

Figure 3.1: (a) Fresnel zone plate. (b) Photon sieve.

The parts of this chapter have been presented in [40].  Some other parts of this chapter will be published in a forthcoming paper [41].

Photon sieves, just like Fresnel zone plates, offer diffraction-limited imaging performance with relaxed manufacturing tolerances. Advantages over Fresnel zone plates are improved spatial resolution for a given smallest fabricable structure and suppression of higher diffraction orders, through quasi-random variations in the distribution and diameter of the holes [42]. They also consist of one connected piece and have less sensitivity to manufacturing errors, enabling simple and low-cost fabrication [4-6]. As a result, this new class of lightweight diffractive image forming devices opens up new possibilities for high resolution imaging and spectroscopy.

Many such photon sieve imaging systems have been suggested at visible, UV, and x-ray wavelengths, some of which are also fabricated and tested to demonstrate diffraction-limited high spatial resolution [1, 3, 4, 7-10]. However, because the focal length of the photon sieve is wavelength-dependent (causing *chromatic aberration*), its use has been generally restricted to monochromatic sources [42–44]. To operate with broad or multispectral illumination, methods for reducing the chromatic aberration have been developed to focus different wavelengths onto the same focal plane [45–47].

In this chapter we present a new photon sieve imaging modality that, conversely, takes advantage of chromatic aberration. The fact that different wavelengths are focused at different distances from the photon sieve is exploited to develop a novel multispectral imaging technique. In contrast to traditional spectral imagers employing a series of wavelength filters, the proposed technique relies on a simple optical system, but requires powerful reconstruction methods to form spectral images computationally. In addition to diffraction-limited high spatial resolution enabled by photon sieves, this technique can also achieve higher spectral resolution than the conventional spectral imagers.

In the first part of this chapter, we present exact and approximate Fresnel imaging formulas that relate the output of a photon sieve imaging system to its input, either when the source is coherent or incoherent. These imaging relations for photon sieve are crucial for effectively analyzing and solving the inverse problems that arise from the new imaging modalities enabled by photon sieves. The results presented in this part have been appeared in our paper [40].

In the second part of this chapter, we will use these imaging formulas in the

development of a novel computational spectral imaging technique with photon sieves. The idea is to use a photon sieve imaging system with a moving detector which records images at different planes. Because the focal length of the photon sieve is wavelength-dependent, each measurement consists of superimposed images of different wavelength sources, with each individual image being either in focus or out of focus. The image of each wavelength source is then recovered by combining these multiplexed measurements with a mathematical model of the imaging system and solving the resultant inverse problem computationally. The promising aspects of the technique in terms of spectral and spatial resolutions are illustrated for EUV solar spectral imaging through numerical simulations. The results presented in this part have been recently appeared in our paper [41].

## 3.2 Part 1: Image formation with photon sieves

We first derive exact and approximate Fresnel imaging formulas that relate the output of a photon sieve imaging system to its input. In the literature the focusing properties of photon sieves and the design procedure have been analyzed through the calculation of Fresnel-Kirchoff diffraction integrals [42] and approximate Fresnel integrals [48]. While the analysis in [42] was only for point sources, the approximate treatment of extended objects [48] has been limited to half of the imaging system (from the photon sieve plane to the measurement plane). In this part, we study the exact and complete image formation process for extended objects.

For this, we consider a photon sieve imaging system in which the photon sieve lies between an extended source and an image (measurement) plane (see Fig. 3.2). We provide closed-form Fresnel imaging formulas that relate the image formed at the measurement plane to the complex amplitude of the source, being either coherent or incoherent. The relations are given in terms of convolutions and Fourier transforms (rather than complicated integrals) to provide a more transparent understanding of the imaging process. This form also offers a fast way of simulating the imaging system.

We also present similar coherent and incoherent imaging equations for an approximate model (consisting of infinite series of lenses), which is known to be equivalent to the Fresnel zone plate [49, 50]. The two-dimensional

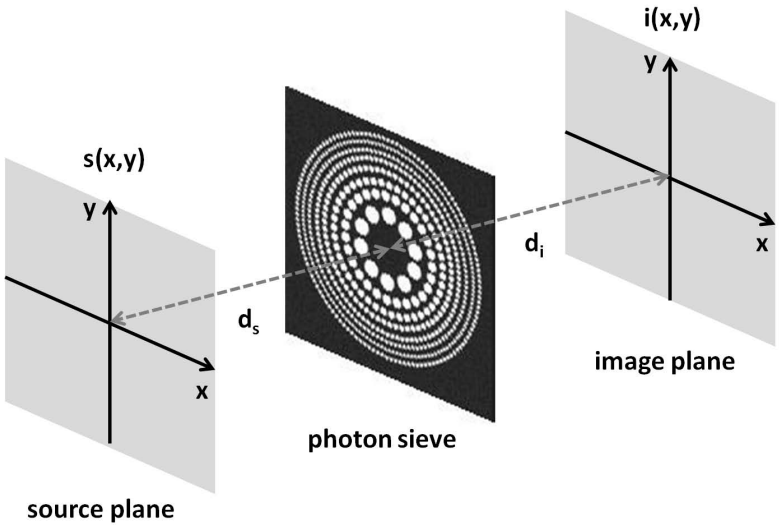



Figure 3.2: Schematic view of the photon sieve imaging system [40]. © [2013] IEEE. Reprinted with permission.

pointspread function of this approximate model is compared with that of the photon sieve, to verify their agreement.

Both the exact and the approximate imaging relations are crucial for effectively analyzing and solving subsequent inverse problems that arise from the new imaging modalities enabled by photon sieves.

## 3.3 Fresnel image formation models

### 3.3.1 Notation

- $s(x, y)$: incident field in the source plane
- $i_n(x, y)$: diffracted field from the $n$th pinhole at the image plane
- $i(x, y)$: total diffracted field from photon sieve at the image plane, hence $i(x, y) = \sum_n i_n(x, y)$
- $(x_n, y_n)$: central location of the $n$th pinhole
- $d_n$: diameter of the $n$th pinhole
- $a_n(x, y)$: aperture function of the $n$th pinhole (taking value 1 inside the pinhole)

### 3.3.2 Imaging system and assumptions

The photon sieve can be used to form images either with coherent or incoherent light, both of which will successively be studied. We consider the imaging system in Fig. 3.2. Here $d_s$ and $d_i$ denote the distances from the source and image planes to the plane where the photon sieve resides.

The input illumination function, located at the $xy$ source plane, is first assumed to be a space-varying, coherent, monochromatic wave of wavelength $\lambda$. Its complex baseband representation is then composed of a linear superposition of plane waves traveling in a continuum of directions [50]:

$$s(x,y) = \int\int a(\alpha,\beta) e^{i2\pi \frac{\alpha x+\beta y}{\lambda}} d\alpha\, d\beta \tag{3.1}$$

where $a(\alpha,\beta)$ is the angular spectrum of the wave, and can be expressed in terms of the Fourier transform $S(f_x,f_y)$ of the input $s(x,y)$ as $a(\alpha,\beta) = 1/\lambda^2 S(\alpha/\lambda,\beta/\lambda)$.

### 3.3.3 Fresnel imaging formula for coherent sources

The input wavefront $s(x,y)$ propagates toward a photon sieve located at a distance $d_s$, as shown in Fig. 3.2. Our goal here is to mathematically relate the input $s(x,y)$ to the image $i(x,y)$ formed at a distance $d_i$ from the photon sieve. The relation given in terms of convolutions and Fourier transforms will provide an easier understanding and analysis of the imaging process, and also suggest an efficient way for simulating the photon sieve imaging system.

We use the Fresnel approximation for diffraction. Let $u(x,y)$ be the field distribution just before the photon sieve. Then $u(x,y)$ is related to $s(x,y)$ through propagation in free space:

$$u(x,y) = s(x,y) * \frac{1}{i\lambda d_s} e^{i\pi \frac{x^2+y^2}{\lambda d_s}}$$

where the convolution is with the pointspread function of the free space in the Fresnel approximation [50,51] (with the constant phase shift term dropped). This convolution can be re-written as a chirp multiplication followed by a

Fourier transform followed by a second chirp multiplication:

$$u(x,y) = \frac{1}{i\lambda d_s} e^{i\pi \frac{x^2+y^2}{\lambda d_s}} \tilde{F}\{s(x',y') e^{i\pi \frac{x'^2+y'^2}{\lambda d_s}}\} \left( \frac{x}{\lambda d_s}, \frac{y}{\lambda d_s} \right)$$

Let $v_n(x,y)$ be the field distribution just after the $n$th pinhole of photon sieve. Then, the relation between the field distributions just before and after the photon sieve is

$$v_n(x,y) = a_n(x,y) u(x,y)$$

where $a_n(x,y)$ is the aperture function of the $n$th pinhole given by the following circle function [50]:

$$a_n(x,y) = \text{circ}\left( \frac{x-x_n}{d_n}, \frac{y-y_n}{d_n} \right)$$

The diffracted field, $i_n(x,y)$, from $n$th pinhole in the image plane is related to $v_n(x,y)$ through propagation in free space, so as before

$$i_n(x,y) = \frac{1}{i\lambda d_i} e^{i\pi \frac{x^2+y^2}{\lambda d_i}} \tilde{F}\{v_n(x',y') e^{i\pi \frac{x'^2+y'^2}{\lambda d_i}}\} \left( \frac{x}{\lambda d_i}, \frac{y}{\lambda d_i} \right)$$

Let us first rewrite the term involving the Fourier transform:

$$\begin{aligned}
&\tilde{F}\{v_n(x,y) e^{i\pi \frac{x^2+y^2}{\lambda d_i}}\} \\
&= \tilde{F}\left\{ \frac{1}{i\lambda d_s} e^{i\pi \frac{x^2+y^2}{\lambda/\Delta}} a_n(x,y) \tilde{F}\{s(x',y') e^{i\pi \frac{x'^2+y'^2}{\lambda d_s}}\} \left( \frac{x}{\lambda d_s}, \frac{y}{\lambda d_s} \right) \right\} \\
&= \frac{1}{i\lambda d_s} \tilde{F}\{e^{i\pi \frac{x^2+y^2}{\lambda/\Delta}} a_n(x,y)\} * \tilde{F}\left\{ \tilde{F}\{s(x',y') e^{i\pi \frac{x'^2+y'^2}{\lambda d_s}}\} \left( \frac{x}{\lambda d_s}, \frac{y}{\lambda d_s} \right) \right\}
\end{aligned}$$

where $\Delta = 1/d_i + 1/d_s$. The second term above equals to $(\lambda d_s)^2 s(-\lambda d_s x, -\lambda d_s y)$ $\exp(i\pi\lambda d_s(x^2+y^2))$. Then

$$i_n(x,y) = -\frac{d_s}{d_i} e^{i\pi \frac{x^2+y^2}{\lambda d_i}} \left[ \frac{i\lambda}{\Delta} e^{-i\pi \frac{x^2+y^2}{\Delta \lambda d_i^2}} * A_n(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}) * \tilde{s}(x,y) \right]$$

where $A_n(f_x, f_y)$ is the Fourier transform of $a_n(x,y)$, and $\tilde{s}(x,y)$ is a *scaled and attenuated version of* $s(x,y)$ given by $s(-\frac{d_s}{d_i}x, -\frac{d_s}{d_i}y) e^{i\pi d_s(x^2+y^2)/(\lambda d_i^2)}$. The total diffracted field $i(x,y)$ in the image plane can be found by sum-

ming $i_n(x, y)$ over all pinholes $n$:

$$i(x, y) = -\frac{d_s}{d_i} e^{i\pi \frac{x^2+y^2}{\lambda d_i}} \left[ g_{\lambda,d_i}(x, y) * \tilde{s}(x, y) \right] \tag{3.2}$$

where $g_{\lambda,d_i}(x, y)$ is the *coherent pointspread function of the photon sieve* at wavelength $\lambda$ and distance $d_i$ given by [40]

$$g_{\lambda,d_i}(x, y) = i\frac{\lambda}{\Delta} e^{-i\pi \frac{x^2+y^2}{\Delta \lambda d_i^2}} * A(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}) \tag{3.3}$$

Here $A(x, y) = \sum_n A_n(x, y)$, which is in our case a sum of jinc functions [50]. Hence the formed image is a scaled, inverted, and chirp multiplied version of the input signal $s(x, y)$ which is filtered in the frequency domain by the chirp interpolated aperture functions.

### 3.3.4 Fresnel imaging formula for incoherent sources

We have seen that the photon sieve forms an image of the object $s(x, y)$ when a spatially coherent wave originates from the object. In this section, we show for spatially incoherent illumination that the photon sieve still produces images, but in intensity only (rather than in complex amplitude).

We consider a spatially incoherent wave originating from the source $s(x, y)$, which can be expressed [50] as

$$\rho(x, y, t) = s(x, y)w(x, y, t) \tag{3.4}$$

where $w(x, y, t)$ is a spatially white random process with the property of $E[w(x, y, t)w^*(x', y', t)] = \delta(x - x', y - y')$. As a result, the autocorrelation of $\rho(x, y, t)$ is given by

$$E[\rho(x, y, t)\rho^*(x', y', t)] = |s(x, y)|^2 \delta(x - x', y - y')$$

We derive an expression for the image intensity, which is described by the

expected value of the squared amplitudes observed at the image plane [40]:

$$
\begin{aligned}
&E[|i(x,y,t)|^2] \\
&= \left(\frac{d_s}{d_i}\right)^2 \iiiint E[\rho(-\frac{d_s}{d_i}\xi, -\frac{d_s}{d_i}\eta, t)\rho^*(-\frac{d_s}{d_i}\xi', -\frac{d_s}{d_i}\eta', t)] \\
&g_{\lambda,d_i}(x-\xi, y-\eta) {g_{\lambda,d_i}}^*(x-\xi' y-\eta') e^{-\frac{i\pi(\xi^2+\eta^2-\xi'^2-\eta'^2)}{\lambda d_i^2/d_s}} d\xi d\xi' d\eta d\eta' \\
&= \left(\frac{d_s}{d_i}\right)^2 \iint \left| s\left(-\frac{d_s}{d_i}\xi, -\frac{d_s}{d_i}\eta\right)\right|^2 |g_{\lambda,d_i}(x-\xi, y-\eta)|^2 d\xi d\eta \\
&= \left(\frac{d_s}{d_i}\right)^2 \left| s\left(-\frac{d_s}{d_i}x, -\frac{d_s}{d_i}y\right)\right|^2 * |g_{\lambda,d_i}(x,y)|^2
\end{aligned} \tag{3.5}
$$

where $g_{\lambda,d_i}(x,y)$ is the coherent optical pointspread function of the photon sieve obtained in the earlier section. Hence the ensemble average of an incoherent image, which is independent of time, is the convolution of intensities, rather than complex amplitudes.

## 3.4 Approximate image formation models

A photon sieve is a modification of a Fresnel zone plate, with only the zones replaced by holes. Because of this relation, many studies of the photon sieve and design criteria rely on viewing it as a Fresnel zone plate [52]. Similarly here we review a lens model given for the Fresnel zone plate [49], and use it as an approximate model for the photon sieve. We give the coherent and incoherent imaging formulas for this approximate model, which (as we will see later) will provide a very efficient way of approximately simulating photon sieve systems.

The Fresnel zone plate generates large number of diffracted orders, and each order, $m$, comes to focus at a focal distance

$$f_m = \frac{r_n^2}{mn\lambda} = \frac{r_1^2}{m\lambda} = \frac{f_1}{m} \tag{3.6}$$

and with the diffraction efficiencies [49, Chap. 9]

$$\eta_m = \begin{cases} 1/4 & \text{if } m = 0 \\ 1/m^2\pi^2 & \text{if } m \text{ odd} \\ 0 & \text{if } m \text{ even} \end{cases} \tag{3.7}$$

where $r_n$ is the distance between the end of the $n$th zone and the origin. Hence 25% of the incident radiation is transmitted in the forward direction ($m = 0$) without being focused, 10% comes to focus on the first order focus, $f_1$, 1% comes to focus on the third order focus, $f_3$, and so on.

Attwood follows Goodman's approach [51] and uses the above observation to model the zone plate as an infinite series of thin lenses, one for each odd order [49]. The reasoning behind this model is that the wavefront curvature of each diffracted order $m$ has a step-wise shape (where each step is due to one zone), and for a zone plate of many zones, say $N > 100$, the wavefront can be approximated by a continuous radial phase advance $\phi(r) = \pi r^2/(\lambda f_m)$, corresponding to a lens of focal length $f_m$. The transmittance function of this model is then

$$t(x,y) = \sum_{\text{odd } m} \frac{1}{|m\pi|}\, e^{-i\pi \frac{r^2}{\lambda f_m}}\, a_l(x,y) \tag{3.8}$$

where $a_l(x,y)$ is the aperture function of the lens determined by the outer diameter of the photon sieve.

Below we give the coherent and incoherent imaging equations for this lens model, which relate the image $i(x,y)$ to the source $s(x,y)$ when the photon sieve in Fig. 3.2 is replaced by the lens model consisting of infinite series of lenses.

**Coherent case:**

At the image plane the $m$th order diffracted field from the $m$th lens is given by

$$i_m(x,y) = -\frac{d_s}{d_i} e^{i\pi \frac{x^2+y^2}{\lambda d_i}}\, [\tilde{s}(x,y) * \tilde{g}_m(x,y)] \tag{3.9}$$

where $\tilde{g}_m(x,y)$ is the *coherent point spread function of the $m$th lens*, and its form depends whether the $m$th lens is in focus or out of focus [50]:

$$\tilde{g}_m(x,y) = \begin{cases} A_l(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}) & \text{if } \epsilon_m = 0 \\ A_l(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i}) * i\frac{\lambda}{\epsilon_m} e^{-i\pi \frac{x^2+y^2}{\epsilon_m \lambda d_i^2}} & \text{if } \epsilon_m \neq 0 \end{cases}$$

where $\epsilon_m = 1/d_i + 1/d_s - 1/f_m$ and $A_l(f_x, f_y)$ is the Fourier transform of the aperture function of the lens, $a_l(x,y)$.

Then the total diffracted field resulting from all orders is [40]

$$i(x,y) = -\frac{d_s}{d_i} e^{i\pi \frac{x^2+y^2}{\lambda d_i}} \left[ \tilde{s}(x,y) * \sum_{\text{odd } m} \frac{1}{|m\pi|} \tilde{g}_m(x,y) \right] \tag{3.10}$$

**Incoherent case:**

Similar to the incoherent imaging with photon sieves, the complex amplitudes in the above convolution are replaced with their intensities.

## 3.5 Numerical comparisons of the models

Here we numerically compare the optical pointspread functions of the photon sieve and the approximate lens model. For this, we first design a sample photon sieve for EUV imaging. By following the design steps explained in [52], for the wavelength of 50nm, we choose the outer diameter of the photon sieve as 5cm, and the diameter of the smallest hole as 0.38mm. This resulted in a photon sieve with focal length of 25m and number of virtual zones of 50, where in each white zone the fraction of open area due to holes is chosen as 0.6.

As shown in Fig. 3.3, the resulting photon sieve and the corresponding lens model have very similar psfs, as expected. Hence the model consisting of infinite series of lenses provides a very good approximation to the photon sieve, with the advantage of requiring less computations for the simulation of the imaging system. As a result, this model can be used for an approximate, but simpler, analysis of the inverse problems arising from the new imaging modalities enabled by photon sieves.

## 3.6 Part 2: Computational spectral imaging with photon sieves

So far we have derived exact and approximate imaging formulas that relate the output of a photon sieve imaging system to its input. These imaging formulas will now be used in the development of a novel spectral imaging technique with photon sieves.

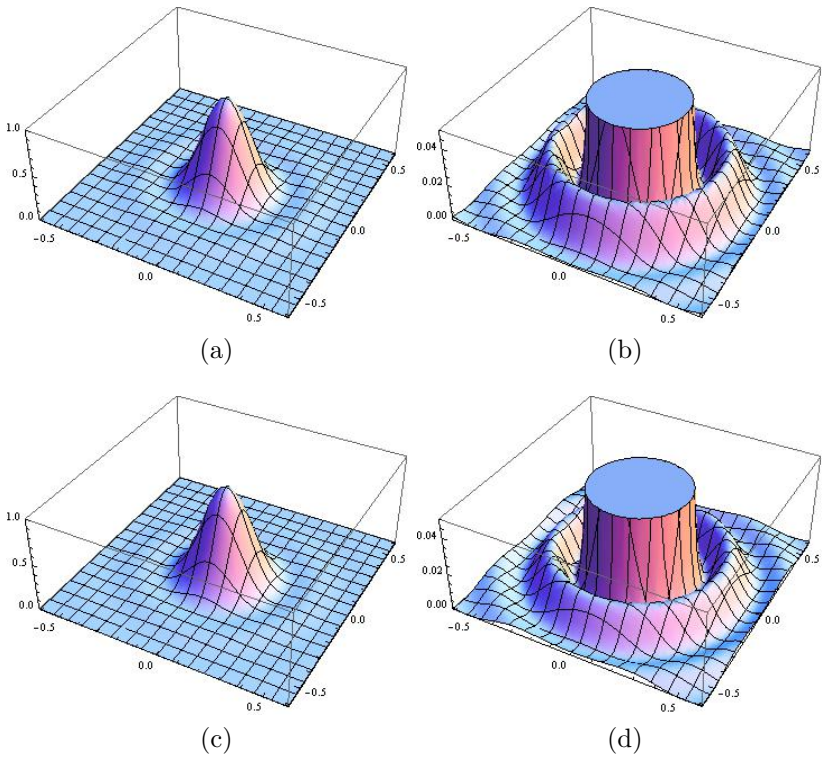



Figure 3.3: Intensities of the pointspread functions of a sample photon sieve (a) and the corresponding lens model (c) at the first order focus, and their zoomed versions (b)-(d) respectively (spatial axes are measured in millimeters) [40]. © [2013] IEEE. Reprinted with permission.

As mentioned earlier, the idea in this technique is to use a photon sieve imaging system with a moving detector which records images at different planes (see Fig. 3.4). Because the focal length of the photon sieve depends on the wavelength, each measurement consists of superimposed images of different wavelength sources, with each individual image being either in focus or out of focus. For spatially incoherent illumination, we study the problem of recovering the individual deblurred images from these superimposed measurements. We first formulate the discrete forward problem using the closed-form Fresnel imaging formulas derived. The inverse problem is then a multiframe deconvolution problem involving multiple objects, and is formulated as a maximum posterior (MAP) estimation problem by incorporating prior statistical knowledge about the targeted scenes. The resulting nonlinear optimization problem is then solved using a fixed-point iterative algorithm [53,54]. At the end, the performance of the technique is illustrated

for EUV spectral imaging.

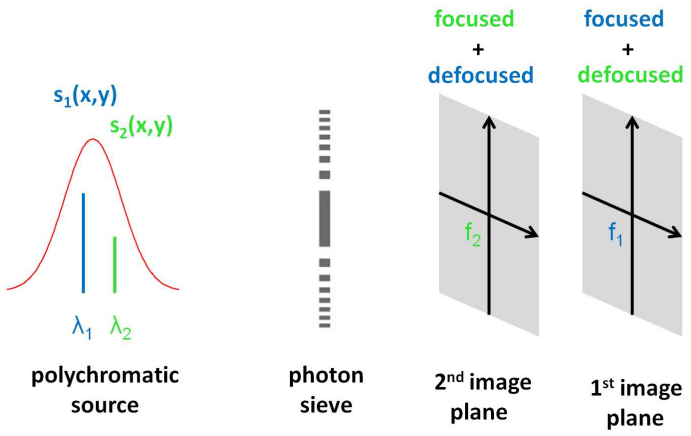



Figure 3.4: Illustration of how the photon sieve spectral imaging system works for a polychromatic source of two uncorrelated monochromatic components, with wavelengths $\lambda_1$ and $\lambda_2$. A moving detector takes measurements at two planes $f_1$ and $f_2$. At the focal plane of the first source, $f_1$, a focused image of the first source overlaps with a defocused image of the second source, and a similar observation at the focal plane of the second source, $f_2$. The individual images will be recovered from these superimposed data by means of solving an inverse problem.

## 3.7 Forward problem

### 3.7.1 Photon sieve spectral imaging system

We consider the photon sieve imaging system in Fig. 3.5, where the detector is moved to record intensity measurements at $K$ different planes. Here $d_s$ and $d_k$ denote the distances from the source and $k$th measurement plane to the plane where the photon sieve resides, with $k = 1, \ldots, K$. As input illumination, we consider a polychromatic source consisting of $P$ spatially incoherent monochromatic sources, each with a different wavelength $\lambda_p$ where $p = 1, \ldots, P$. These monochromatic sources are also assumed to be mutually incoherent [50]. Although, in this study, we focus on the spatially incoherent case where the photon sieve produces images in intensity only, the concepts readily generalize to the coherent or partially coherent case as well.

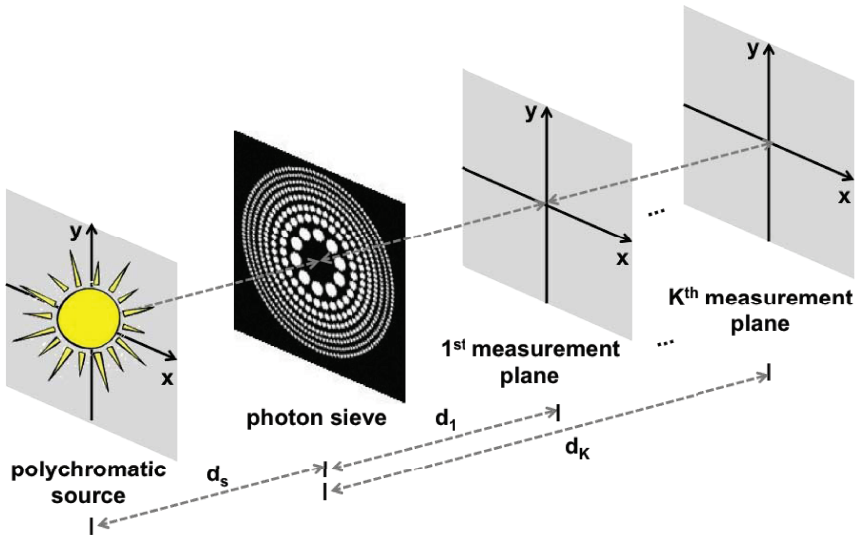



Figure 3.5: The photon sieve spectral imaging system [41].

### 3.7.2 Continuous image formation model

Our first goal is to mathematically relate the intensity of the input sources, $f_{\lambda_p}(x, y)$, to the images formed at the measurement planes. For continuous sources, we have derived the image formation models in Section 3.2. By using these models, the intensity $t_k(x, y)$ observed at the $k$th measurement plane is [41]

$$t_k(x, y) = \sum_{p=1}^{P} s_p(x, y) * g_{\lambda_p, d_k}(x, y) \tag{3.11}$$

where $s_p(x, y) = \frac{d_s^2}{d_k^2} f_{\lambda_p}(-\frac{d_s}{d_k}x, -\frac{d_s}{d_k}y)$ is a scaled version of the source intensity $f_{\lambda_p}(x, y)$, and $g_{\lambda_p, d_k}(x, y)$ is the incoherent point-spread function (PSF) of the photon sieve at wavelength $\lambda_p$ and distance $d_k$, given by [40]

$$g_{\lambda_p, d_k}(x, y) = \left| i\frac{\lambda_p}{\Delta_k} e^{-i\pi \frac{x^2+y^2}{\Delta_k \lambda_p d_k^2}} * A\left(\frac{x}{\lambda_p d_k}, \frac{y}{\lambda_p d_k}\right) \right|^2 \tag{3.12}$$

Here $\Delta_k = 1/d_s + 1/d_k$, and $A(x, y)$ is the Fourier transform of the total aperture function of the photon sieve (sum of jinc functions [50] corresponding to the Fourier transform of the circular aperture functions of each hole on the photon sieve).

An approximate, but simpler, model to (3.12) exists when the number of virtual zones $N > 100$ [40]. In this case, the photon sieve can be replaced by a series of lenses with appropriate parameters. The resulting approximate PSF requires less computation; in particular, when the measurement plane is at the first-order focus, the PSF is just a squared jinc function.

### 3.7.3 Discrete forward model

Because the measurements will be acquired digitally and also input images will be reconstructed on a digital computer, a discrete forward model is required. We obtain such a discrete model based on the band limitedness of the continuous functions involved. Note that the PSF $g_{\lambda_p,d_k}$ is band limited to a circle of diameter $2D/(\lambda_p d_k)$. This is because the argument of the absolute value in (3.12) has a circular frequency support, whose diameter $D/(\lambda_p d_k)$ is determined by the outer diameter $D$ of the photon sieve. (With the incoherent PSF given by the absolute square of this function, the frequency support of the PSF is given by the convolution of this circular support with itself, resulting in a circular support of twice diameter.)

The band limitedness of $t_k$ then directly follows from the band limitedness of the PSF. Also note that all high frequencies of $s_p$ that lie outside the frequency support of the PSF are lost at the output, as a result of inherent *diffraction-limit* [49, 51]. Hence the nullspace of this PSF operator is nonempty. For this reason, we restrict our attention to input functions of same bandwidth only, and aim for recovering the band limited version of $s_p$, which is defined as $\tilde{s}_p(x,y) \equiv s_p(x,y) * \mathrm{jinc}((2D/\lambda_p d_1)\, x, (2D/\lambda_p d_1)\, y)$. The input $s_p$ in (3.11) can be replaced with its band limited version $\tilde{s}_p$, without affecting the measurement $t_k$. Therefore, all functions in the continuous observation model can be assumed band limited.

By replacing each continuous band limited function with its discrete representation with sinc basis, the continuous convolution operation in (3.11) reduces to a discrete convolution:

$$t_k[m,n] = \sum_{p=1}^{P} \tilde{s}_p[m,n] * g_{\lambda_p,d_k}[m,n] \tag{3.13}$$

where $t_k[m,n]$, $\tilde{s}_p[m,n]$, and $g_{\lambda_p,d_k}[m,n]$ are uniformly sampled versions of their continuous forms, e.g. $t_k[m,n] = t_k(m\Delta, n\Delta)$ for some $\Delta$ smaller than the Nyquist sampling interval. We will assume that $t_k[m,n]$, i.e. uniformly sampled version of the continuous observations, is approximately the same as the detector measurements with a pixel size of $\Delta$, i.e., the averaged intensity over a pixel. We also assume that the size of the input objects are limited to the detector range determined by $N \times N$ pixels, i.e., $m, n = 0, \ldots, N-1$.

Using lexicographic ordering and linearity of the convolution operator, the model can be cast in matrix-vector form:

$$\boldsymbol{t}_k = \sum_{p=1}^{P} \boldsymbol{H}_{k,p} \tilde{\boldsymbol{s}}_p \tag{3.14}$$

where $\boldsymbol{H}_{k,p}$ is an $N^2 \times N^2$ block Toeplitz matrix corresponding to the convolution operation with $g_{\lambda_p,d_k}$. By combining the measurement data from all measurement planes, we get [41]

$$\boldsymbol{t} = \boldsymbol{H}\tilde{\boldsymbol{s}}, \tag{3.15}$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{1,1} & \dots & \boldsymbol{H}_{1,P} \\ \vdots & & \vdots \\ \boldsymbol{H}_{K,1} & \dots & \boldsymbol{H}_{K,P} \end{bmatrix} \tag{3.16}$$

where $\boldsymbol{t} = [\boldsymbol{t}_1^T | \dots | \boldsymbol{t}_K^T]^T$, $\tilde{\boldsymbol{s}} = [\tilde{\boldsymbol{s}}_1^T | \dots | \tilde{\boldsymbol{s}}_p^T]^T$, and $\boldsymbol{H}$ is a $KN^2 \times PN^2$ matrix. In practice, $t$ is observed in the presence of noise, hence the complete observation model is

$$\boldsymbol{y} = \boldsymbol{t} + \boldsymbol{w} = \boldsymbol{H}\tilde{\boldsymbol{s}} + \boldsymbol{w} \tag{3.17}$$

where $\boldsymbol{w} = [\boldsymbol{w}_1^T | \dots | \boldsymbol{w}_K^T]^T$ is the additive noise vector with $(\boldsymbol{w}_k)_i \sim N(0, \sigma_k^2)$ representing white Gaussian noise that is uncorrelated across pixels $i$ and measurement planes $k$.

## 3.8 Inverse problem

In the inverse problem, the goal is to recover the unknown intensities, $\tilde{\boldsymbol{s}}$, of sources of different wavelengths from the intensity measurements, $\boldsymbol{y}$, of the photon sieve system. This inverse problem is a multiframe deconvolution problem from measurements of superimposed and blurred data. (Each measurement is composed of focused/defocused images of different sources.) This deconvolution problem is inherently ill-posed, and as the distance between different measurement planes or different wavelengths decreases, the problem becomes more ill-conditioned due to the increase in the linear dependency of the rows and columns of $\boldsymbol{H}$, respectively.

There are a variety of approaches to solving ill-posed linear inverse problems [55]. We consider a stochastic inversion approach based on MAP esti-

mation to incorporate the prior statistical knowledge of the targeted scenes. The MAP estimate of $\tilde{\boldsymbol{s}}$ from the measurements $\boldsymbol{y}$ is given by

$$\arg\max_{\tilde{\boldsymbol{s}}} p(\boldsymbol{y} \mid \tilde{\boldsymbol{s}})\, p(\tilde{\boldsymbol{s}}) \tag{3.18}$$

where $p(\boldsymbol{y} \mid \tilde{\boldsymbol{s}})$ denotes the conditional probability density function of $\boldsymbol{y}$ given $\tilde{\boldsymbol{s}}$ (arising from the observation model in (3.17)), and $p(\tilde{\boldsymbol{s}})$ denotes the prior distribution of $\tilde{\boldsymbol{s}}$ (specifying the probability of $\tilde{\boldsymbol{s}}$ independently from the observed data).

By taking the logarithm of (3.18), the MAP estimation problem turns into a regularized linear least-squares problem, which involves the minimization of the following functional:

$$\Phi(\hat{\boldsymbol{s}}) = ||\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{s}}||_W^2 + \alpha^2 R(\hat{\boldsymbol{s}}) \tag{3.19}$$

where $R$ and $\alpha$ are respectively the regularization functional and parameter arising from $p(\hat{\boldsymbol{s}})$ (i.e., $R(\hat{\boldsymbol{s}}) \propto -\log p(\hat{\boldsymbol{s}})$), and $W$ is a weight chosen according to the noise standard deviation at different measurement planes.

One common regularization functional is the quadratic regularization for which $R(\hat{\boldsymbol{s}}) = ||L\hat{\boldsymbol{s}}||_2^2$ with an appropriately chosen operator $L$ (often a derivative operator). This is associated with a Gaussian prior. This choice of prior leads to a quadratic optimization problem with a stable solution, and results in globally smooth reconstructions. Global smoothness is due to the fact that the regularization functional penalizes large variations in the reconstructed function due to the underlying Gaussian prior. In situations where the underlying object is not globally smooth, the prior can be replaced, for example, with a Laplacian prior, which will imply a regularization function with $l_1$ norm. This will penalize only the total variation in the reconstructed function, and allow the preservation of the edges when they fit the data, which is due to the larger tails of Laplacian prior compared to the Gaussian prior.

Here we use a general $l_p$-based regularization [41]:

$$\Phi(\hat{\boldsymbol{s}}) = ||\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{s}}||_W^2 + \alpha^2 ||\boldsymbol{D}\hat{\boldsymbol{s}}||_p^p \tag{3.20}$$

where $\boldsymbol{D}$ is a discrete approximation to the gradient operator. The $l_p$ norm implies a generalized Gaussian prior of the form $p(\hat{\boldsymbol{s}}) \propto \prod_i \exp(-|(\boldsymbol{D}\hat{\boldsymbol{s}})_i|^p)$, and hence allows reconstructions of globally smooth objects to sharper ob-

jects, depending on the prior knowledge in a specific application. When $p \neq 2$, the resulting nonlinear optimization problem does not have a closed-form solution, and numerical techniques are used to find the solution. One such approach is a fixed-point iterative algorithm [54], a special case of the "half-quadratic regularization" method [53], which obtains the reconstruction as the solution of a series of approximating quadratic problems.

## 3.9 Sample application in solar spectral imaging

Here we present numerical simulations to illustrate the high spatial and spectral resolution enabled by the proposed spectral imaging technique. For this, we consider a polychromatic input source generating two quasi-monochromatic waves at close (but different) EUV wavelengths: $\lambda_1 = 33.4$ nm and $\lambda_2 = 33.5$ nm (i.e., $P = 2$). Moreover, the photon sieve system records the intensities at the two focal planes, $f_1$ and $f_2$, corresponding to wavelengths $\lambda_1$ and $\lambda_2$ (i.e., $K = 2$). Then at the first focal plane, the measurement consists of a focused image of the first source overlapped with a defocused image of the second source, and vice versa at the other focal plane.

For the photon sieve, a sample design in [52] for EUV solar imaging is considered, with the outer diameter of the photon sieve as 25 mm, and the diameter of the smallest hole as 5 µm. This results in a photon sieve with first-order focal lengths of $f_1 = 3.742$ m and $f_2 = 3.731$ m, and Abbe's diffraction resolution limit of 5 µm [49, 51]. The pixel size on the detector is then chosen as 2.5 µm to match the diffraction-limited resolution of the system (i.e., corresponding to Nyquist rate).

In our first experiment, solar EUV scenes of size $128 \times 128$ are used as the inputs to the photon sieve system. However, since the resolution of the existing solar spectral imagers are below the diffraction-limited resolution considered here, it is not possible to obtain a realistic (high-resolution) input to the simulated system. Instead, we use these solar images as if they were images of some other sun-like object, and illustrate the diffraction-limited resolution for this case. Our goal is to illustrate that diffraction-limited high-resolution can be achieved for imaging objects with similar characteristics.

Using the forward model in (3.13), we generated a data set corresponding to $\boldsymbol{y}$ at the signal-to-noise ratio of $\sim 31$dB. Fig. 3.6 shows the resulting

measurements at the two focal planes together with the contributions of each source and the corresponding PSFs of the system. The reconstructed images with $l_2$-norm regularization are shown in Fig. 3.7 for the two sources, together with the only-diffraction-limited versions of the original scenes, for comparison. This suggests that the proposed system achieves near diffraction-limited resolution, with the absolute percentage difference between reconstructions and diffraction-limited images less than 15% in this case. For this experiment, $p = 2$ is chosen for the regularization (prior) because the diffraction-limited objects of interest, as shown in Fig. 3.7b-3.7d, are nearly globally smooth.

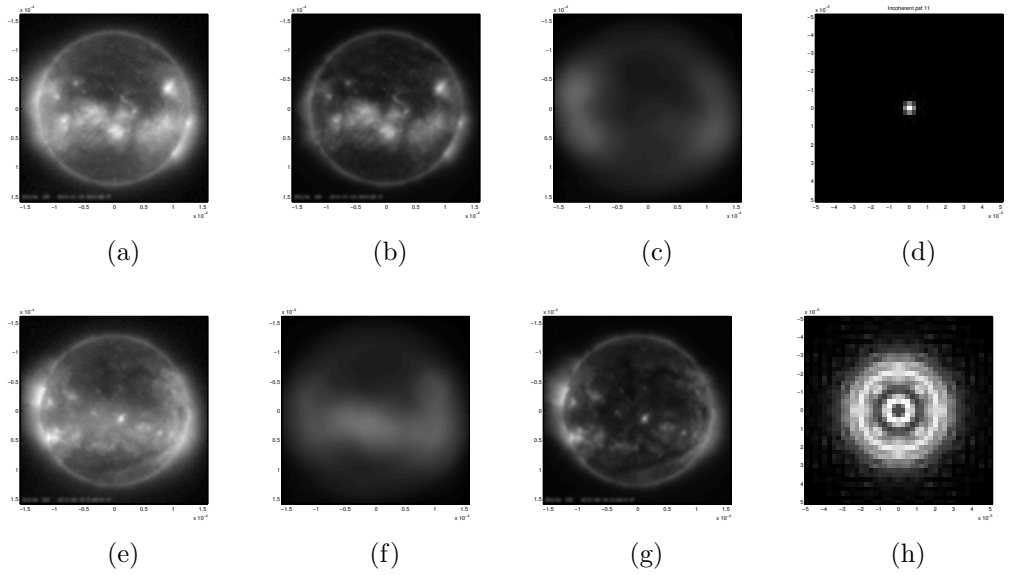
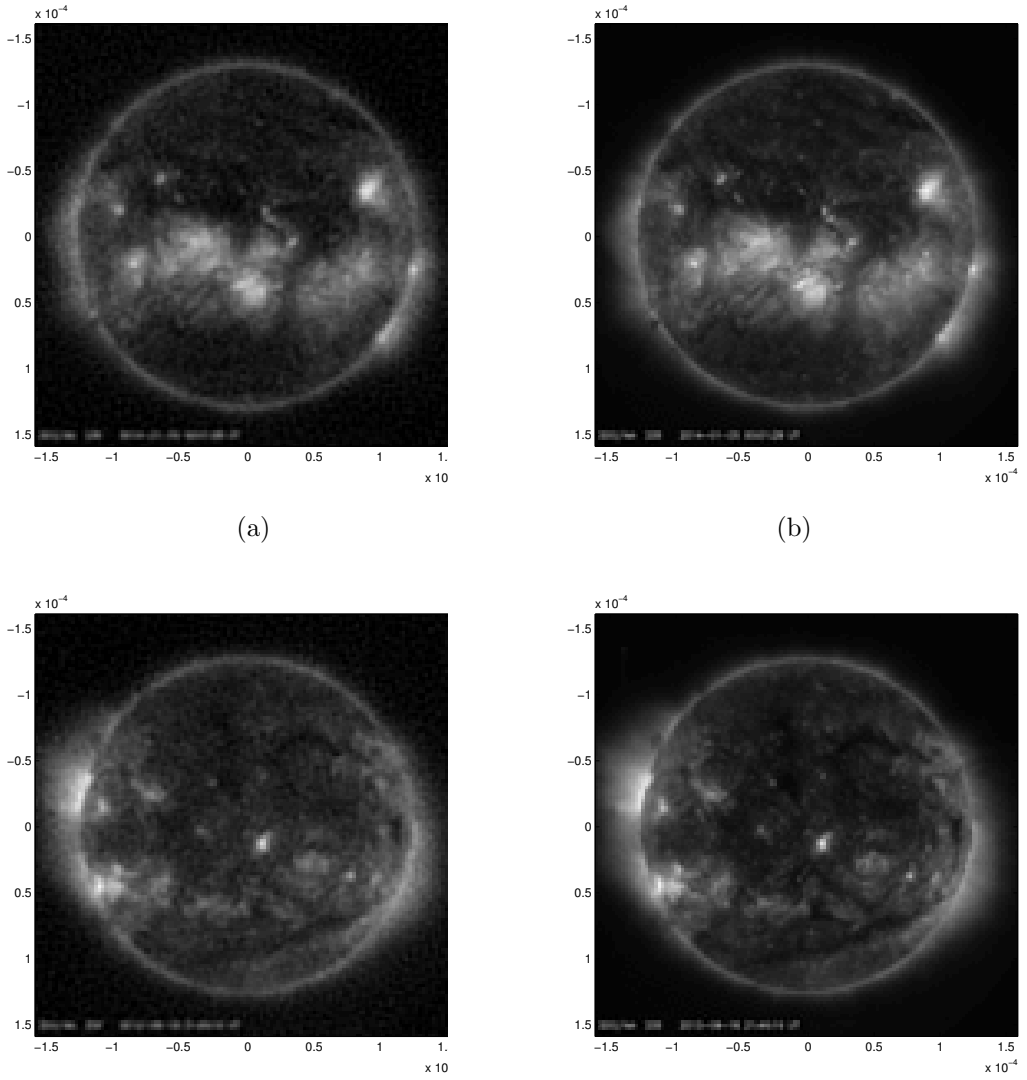


Figure 3.6: Measured intensities at the first and second focal planes (a)-(e), the underlying images of the first source at the first and second focal planes (b)-(f), the underlying images of the second source at the first and second focal planes (c)-(g), and sampled and zoomed point-spread functions of the system for the focused and defocused cases (d)-(h), respectively [41].

In our second experiment, we use sharper images of letters $U$ and $A$ (of size $31 \times 31$) as our source intensities. Our goal is now to illustrate the significance of proper prior choice for achieving diffraction limit or even beyond. As before, the measurements at two focal planes are simulated at the signal-to-noise ratio of $\sim$ 38dB. Because the diffraction-limited objects have sharp edges, as shown in Fig. 3.8b-3.8e, $l_1$-norm regularization (Laplacian prior) is used for the reconstructions. As shown in Fig. 3.8, diffraction-limited resolution is achieved with this regularization, where the absolute percent-



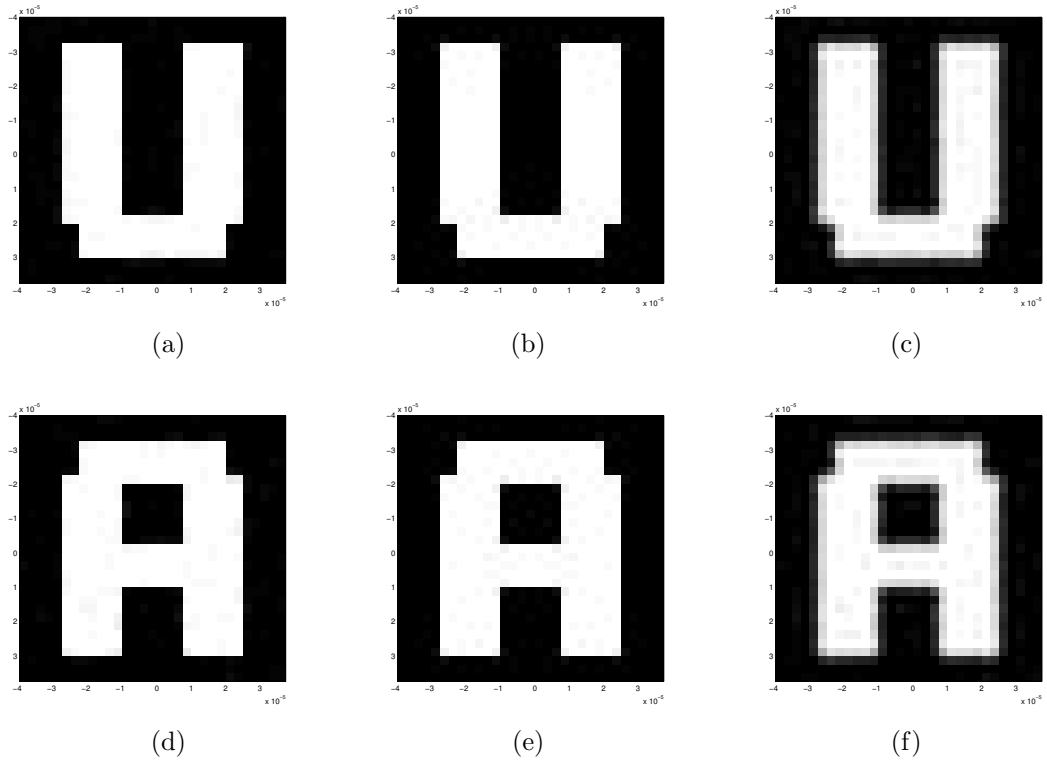
(a) (b)

(c) (d)

Figure 3.7: Estimated intensities of the first and second sources (a)-(c), and diffraction-limited images of the original sources (b)-(d), respectively [41].

age difference between reconstructions and diffraction-limited images is less than 4%. (Also shown in Fig. 3.8 are reconstructions with $l_2$ regularization, which show the effect of smoothing imposed by an incorrect prior.) Moreover, the reconstructions with $l_1$ regularization are slightly sharper than the diffraction-limited versions, and in fact more similar to the original scenes. This illustrates the possibility of achieving high-resolution even beyond the diffraction-limit with the use of prior knowledge of the targeted scenes. This

advantage of *computational* spectral imaging is easier to observe when diffraction causes larger smoothing on the original scenes of letter U and A. In this case $l_1$ regularization can still recover the sharp edges that are associated with the high frequencies lost due to the diffractive imaging process.



(a) (b) (c)

(d) (e) (f)

Figure 3.8: Estimated intensities of the first and second sources using $l_1$-based regularization (a)-(d), only diffraction-limited images of the original sources (b)-(e), estimated intensities using $l_2$-based regularization (c)-(f), respectively [41].

The above experiments illustrate the possibility of achieving diffraction-limited spatial resolution and even beyond with the proposed spectral imaging technique. Another important advantage, which is slightly hidden in these experiments, is the higher spectral resolution achieved compared to the conventional spectral imagers with wavelength filters. Note that the sources have wavelengths 33.4 nm and 33.5 nm, resulting in a spectral resolution of 0.1nm, which is less than 0.3% of the central wavelength of each source. Such a high spectral resolution is not possible to achieve with the state-of-the-art EUV wavelength filters, which can at best have a spectral resolution of 10% of the central wavelength [56]. This becomes an issue when this 10%

spectral band contains more than one spectral line, in which case resolving the individual lines is not possible. Lastly, these promising aspects of the technique can be improved further by taking more measurements than the unknown intensities (such as obtaining measurements at the intermediate planes), which will help to remedy the ill-posed nature of the deconvolution problem.

## 3.10 Conclusion

In this chapter, we have derived exact and approximate Fresnel imaging formulas that relate the output of a photon sieve imaging system to its input, originating from either a coherent or incoherent extended source. These imaging formulas are used in the development of a novel computational spectral imaging technique with photon sieves. In contrast to traditional filter-based spectral imagers, this technique relies on a simple optical system, but requires powerful image processing methods to form spectral images computationally. This new generation of spectral imagers with photon sieves not only offers near diffraction-limited spatial resolution, but also provides several orders of magnitude higher spectral resolution compared to filter-based spectral imagers. Indeed, the technique offers the possibility of separating nearby spectral components that would not otherwise be possible using wavelength filters. These aspects are particularly useful in applications that require high-resolution spectral analysis in the presence of nearby spectral components.

# CHAPTER 4

# PERFORMANCE LIMITS IN COMPUTATIONAL SPECTRAL IMAGING

## 4.1 Introduction

Computational imaging and sensing is an important field that enables new imaging capabilities by distributing the imaging task between a physical and a computational system. In this framework, an image of interest is computationally formed from the physical observations by means of solving an inverse problem. In the earlier two chapters, we show how to utilize the computational imaging framework to develop a class of novel spectral imaging techniques. Since an inversion is required to infer the spectral imaging information from the noisy measurements, a rigorous theory is essential for quantitative characterization of the performance of these techniques.

In this chapter, we develop such a theory for the characterization of the fundamental performance limits, and in particular, seek answers to the following questions:

- What minimum data is required for the estimation of the spectral imaging information at a desired precision?
- How much improvement in the precision of estimates is expected with additional measurements?
- What is the maximum expected precision (minimum uncertainty) in the estimation under different observing scenarios, and which instrument design considerations can achieve it?

A preliminary version of the results of this chapter has been presented in [19]. F. S. Oktem, F. Kamalabadi, and J. M. Davila, "Cramer-Rao bounds and instrument optimization for slitless spectroscopy," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 2169–2173. 

Our tool is the Bayesian Cramer-Rao lower bound theory [57], which provides a lower bound on the mean-square estimation error of any Bayesian estimator. We first obtain the Bayesian Cramer-Rao lower bounds for the estimation of the parameters in a fairly general image formation model. The performance limits of the instantaneous spectral imaging technique are then explored by using the error bounds, which are derived in terms of important instrument design considerations including the diffraction orders to measure, dispersion scale, signal-to-noise ratio, and number of pixels. Our treatment also provides a framework for exploring the optimal choices of these design considerations. For an application in solar spectral imaging, the tightness of the bounds and the performance of the MAP estimator are evaluated via Monte Carlo simulations and under different observing scenarios. As a whole, the Bayesian Cramer-Rao framework not only allows us to explore the requirements that render this new imaging modality effective, but also yields to optimal design choices to minimize the unavoidable estimation errors due to noise.

In the spectroscopy literature, error bounds have been obtained for a simpler problem of fitting a single Gaussian line to measurements in a maximum likelihood sense [20, 58]. On the other hand, the problem in our instantaneous spectral imaging setting is equivalent to fitting a superposition of multiple Gaussians to the measurements. For this setting, the Cramer-Rao error bounds, which provides a lower bound on the variance of *unbiased* estimators, have been presented in our paper [19]. Other related works to this chapter include the error bounds derived for problems that have similar forms as the image formation model studied here, and appear in different contexts such as parameter estimation of superimposed signals [37], localization of EEG and MEG sources [59], array signal processing [60], and multiframe blind deconvolution [61].

## 4.2 Image formation model

We return to the general form of image formation model introduced in Chapter 2, which also includes as its special case the image formation model in Chapter 3. As before, this image formation model relates the measurements to the spectral imaging information that we want to reconstruct, and is given

by

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{y} + \mathbf{n}, \\ \mathbf{y} &= \mathbf{H}(\boldsymbol{\Theta})\mathbf{f} \end{aligned} \tag{4.1}$$

Here $\tilde{\mathbf{y}}$ denotes a noisy measurement vector of length $N$, and $\mathbf{y}$ denotes its noise-free version. The noise vector $\mathbf{n}$ represents a zero-mean Gaussian vector, whose components are uncorrelated and have common variance $\sigma^2$. Moreover, the vectors $\boldsymbol{\Theta}$ and $\mathbf{f}$ represent unknown or known parameters which together capture the spectral imaging information. The vector $\mathbf{f}$ is of size $M$, and $\boldsymbol{\Theta}$ is of size $MP \times 1$ and is given by $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1\, \boldsymbol{\Theta}_2 \ldots \boldsymbol{\Theta}_M]^\top$. Here $\boldsymbol{\Theta}_i = [\theta_{i1}\, \theta_{i2} \ldots \theta_{iP}]$ contains all the parameters associated with the $i$th column of the matrix $\mathbf{H}(\boldsymbol{\Theta})$, which is denoted by $\mathbf{h}_i(\boldsymbol{\Theta}_i)$. The system (blur) matrix $\mathbf{H}(\boldsymbol{\Theta})$, which has size $N \times M$, can be (i) shift-variant or shift-invariant, and (ii) fully or partially known depending on whether $\boldsymbol{\Theta}$ is known. This general framework includes as its special case many semiblind and nonblind deblurring problems involving shift-variant/invariant blur and single/multiple channels.

## 4.3 Bayesian Cramer-Rao error bounds

In the inverse problem, the goal is to estimate the unknown parameters $\mathbf{f}$, $\boldsymbol{\Theta}$, or both from the measurements $\tilde{\mathbf{y}}$. As in earlier chapters, we are interested in solving this inverse problem in a Bayesian estimation framework. This allows us to incorporate the prior knowledge of the statistics of the unknown parameters, and helps to improve the stability of the inverse problem.

Bayesian Cramer-Rao bound (CRB) theory [57,62–64] gives a lower bound on the mean-square estimation error (MSE) of any Bayesian estimator. It is expressed in terms of the Bayesian information matrix $\boldsymbol{J}_B$ whose elements are

$$[\boldsymbol{J}_B]_{ij} = E_{\tilde{\mathbf{y}},\boldsymbol{\Psi}} \left[ \frac{\partial \ln p(\tilde{\mathbf{y}}, \boldsymbol{\Psi})}{\partial \psi_i} \frac{\partial \ln p(\tilde{\mathbf{y}}, \boldsymbol{\Psi})}{\partial \psi_j} \right] \tag{4.2}$$

with $\boldsymbol{\Psi} = [\mathbf{f}^\top\, \boldsymbol{\Theta}^\top]^\top$ in our case. The mean-square error of any Bayesian estimator $\hat{\boldsymbol{\Psi}}$ is then lower bounded by

$$E_{\tilde{\mathbf{y}},\boldsymbol{\Psi}}[(\hat{\psi}_i(\tilde{\mathbf{y}}) - \psi_i)^2] \geq [\boldsymbol{J}_B^{-1}]_{ii} \tag{4.3}$$

with $\hat{\psi}_i(\tilde{\mathbf{y}})$ denoting any Bayesian estimator of the scalar parameter $\psi_i$ (corresponding to the $i$th element of $\boldsymbol{\Psi}$). This bound is subject to the following regularity conditions [63]:

1. The joint distribution $p(\tilde{\mathbf{y}}, \boldsymbol{\Psi})$ is absolutely continuous with respect to $\psi_i$ almost everywhere for all $i = 1, \ldots, M(P+1)$.

2. $\lim_{\psi_i \to \pm\infty} \psi_i \, p(\tilde{\mathbf{y}}, \boldsymbol{\Psi}) = 0$ almost everywhere for all $i = 1, \ldots, M(P+1)$.

3. $\boldsymbol{J}_B$ is nonsingular.

The conditions are given for the case that the joint distribution $p(\tilde{\mathbf{y}}, \boldsymbol{\Psi})$ is strictly positive for all $\boldsymbol{\Psi} \in \mathcal{R}^{M(P+1)}$. The bound also holds for the case that the parameter space is a subset of $\mathcal{R}^{M(P+1)}$ under obvious modifications of the regularity conditions [64]. We further note that the bound is attainable if and only if the posterior distribution of $\boldsymbol{\Psi}$ given $\tilde{\mathbf{y}}$ is Gaussian [57, 63].

The Bayesian information matrix $\boldsymbol{J}_B$ can be expressed as

$$\boldsymbol{J}_B = \boldsymbol{J}_D + \boldsymbol{J}_P \tag{4.4}$$

where the first term depends on the data and the noise distribution, and the second term depends only on the prior distributions $p(\boldsymbol{\Psi})$. In particular, the $(i, j)$th elements of these terms are given by

$$[\boldsymbol{J}_D]_{ij} = -E_{\tilde{\mathbf{y}}, \boldsymbol{\Psi}} \left[ \frac{\partial^2 \ln p(\tilde{\mathbf{y}}|\boldsymbol{\Psi})}{\partial \psi_i \, \partial \psi_j} \right] = E_{\boldsymbol{\Psi}} \left[ [\boldsymbol{J}_F(\boldsymbol{\Psi})]_{ij} \right], \tag{4.5}$$

$$[\boldsymbol{J}_P]_{ij} = -E_{\boldsymbol{\Psi}} \left[ \frac{\partial^2 \ln p(\boldsymbol{\Psi})}{\partial \psi_i \, \partial \psi_j} \right] \tag{4.6}$$

Here $\boldsymbol{J}_F(\boldsymbol{\Psi})$ denotes the Fisher information matrix whose $(i, j)$th element is given by

$$[\boldsymbol{J}_F(\boldsymbol{\Psi})]_{ij} = -E_{\tilde{\mathbf{y}}|\boldsymbol{\Psi}} \left[ \frac{\partial^2 \ln p(\tilde{\mathbf{y}}|\boldsymbol{\Psi})}{\partial \psi_i \, \partial \psi_j} \right] \tag{4.7}$$

Fisher information matrix alone gives a lower bound on the variance of *unbiased* estimators [32, 57]. On the other hand, Bayesian information matrix enables to bound the MSE averaged over the prior distribution and has the advantage that it does not require any unbiasedness condition.

Hence the computation of the Bayesian Cramer-Rao bounds involves computing the data and prior terms in the Bayesian information matrix. The

prior term, $\boldsymbol{J}_P$, has a closed-form for many frequently encountered distributions, and the data term, $\boldsymbol{J}_D$, can be obtained by taking the expectation of the Fisher information matrix over $\boldsymbol{\Psi}$. (We note that if non-informative priors that satisfy the regularity conditions are used, then only the computation of the data term is required since the prior term reduces to zero.)

The Fisher information matrix has the following closed-form under our image formation model in (4.1):

$$\boldsymbol{J}(\boldsymbol{\Psi}) = \begin{bmatrix} \boldsymbol{J}_{\mathbf{f},\mathbf{f}} & \boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}} \\ \boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}}^{\top} & \boldsymbol{J}_{\boldsymbol{\Theta},\boldsymbol{\Theta}} \end{bmatrix} \tag{4.8}$$

where

$$\boldsymbol{J}_{\mathbf{f},\mathbf{f}} = \frac{1}{\sigma^2}\,\mathbf{H}^{\top}(\boldsymbol{\Theta})\mathbf{H}(\boldsymbol{\Theta}), \tag{4.9}$$

$$\boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}} = \frac{1}{\sigma^2}\,\mathbf{H}^{\top}(\boldsymbol{\Theta})\boldsymbol{D}(\boldsymbol{\Theta})\boldsymbol{G}(\mathbf{f}), \tag{4.10}$$

$$\boldsymbol{J}_{\boldsymbol{\Theta},\boldsymbol{\Theta}} = \frac{1}{\sigma^2}\,\boldsymbol{G}^{\top}(\mathbf{f})\boldsymbol{D}^{\top}(\boldsymbol{\Theta})\boldsymbol{D}(\boldsymbol{\Theta})\boldsymbol{G}(\mathbf{f}) \tag{4.11}$$

with

$$\boldsymbol{D}_i(\boldsymbol{\Theta}_i) = \left[\frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i1}} \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i2}} \cdots \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{iP}}\right], \tag{4.12}$$

$$\boldsymbol{D}(\boldsymbol{\Theta}) = [\boldsymbol{D}_1(\boldsymbol{\Theta}_1)\; \boldsymbol{D}_2(\boldsymbol{\Theta}_2) \ldots \boldsymbol{D}_M(\boldsymbol{\Theta}_M)], \tag{4.13}$$

$$\boldsymbol{G}(\mathbf{f}) = \mathrm{diag}(\mathbf{f}) \otimes \boldsymbol{I}_{\mathrm{P}\times\mathrm{P}} \tag{4.14}$$

Here $\mathrm{diag}(\mathbf{f})$ is a diagonal matrix with the elements of $\mathbf{f}$ on the diagonal, $\otimes$ is the Kronecker product, and $\boldsymbol{I}_{P\times P}$ is $P \times P$ identity matrix. This form of Fisher information matrix can be regarded as a specialized version of the Fisher information matrices obtained earlier in the context of parameter estimation of superimposed signals [37] and localization of EEG and MEG sources [59]. (The derivation for obtaining this closed-form expression is provided in Appendix for completeness.) The data term of Bayesian information matrix, $\boldsymbol{J}_D$, is then given by the expectation of this Fisher information matrix over $\boldsymbol{\Psi}$; but this expectation is generally not analytically tractable, and must be computed through Monte Carlo simulation or numerical integration.

### 4.3.1 Error bounds for transformed quantities of interest

The error bounds presented in the previous section are for the estimation of parameters $\mathbf{f}$ and $\boldsymbol{\Theta}$ which directly appear in the image formation model. However the goal in spectral imaging is often to estimate some physical quantities of interest that are related to these model parameters through a transformation. In such settings, one would actually be interested in the error bounds for the estimation of these transformed parameters, rather than the model parameters. Moreover the goal in instrument optimization is generally to minimize the estimation errors of some physical quantities with respect to design parameters. This requires us to relate the model parameters to these physical quantities and design parameters, and then to obtain the error bounds for the physical quantities as a function of the design parameters.

In this section, we discuss how the error bounds for a set of transformed parameters can be obtained from the error bounds for the model parameters presented in the previous section. Let the transformed parameter vector of interest be $\boldsymbol{\Psi}'$, which is related to the model parameter vector $\boldsymbol{\Psi}$ by

$$\boldsymbol{\Psi}' = \boldsymbol{\gamma}(\boldsymbol{\Psi}) \tag{4.15}$$

Then, under some regularity conditions, the Bayesian Cramer-Rao bounds for $\boldsymbol{\Psi}'$ can be obtained from the error bounds for $\boldsymbol{\Psi}$ by using the following error propagation formula [62]:

$$E_{\tilde{\mathbf{y}},\boldsymbol{\Psi}}[(\hat{\boldsymbol{\gamma}}_i(\boldsymbol{\Psi}) - \boldsymbol{\gamma}_i(\boldsymbol{\Psi}))^2] \geq \left[\boldsymbol{T}_{\boldsymbol{\gamma}}\, \boldsymbol{J}_B^{-1}\, \boldsymbol{T}_{\boldsymbol{\gamma}}^{\top}\right]_{ii} \tag{4.16}$$

where $\boldsymbol{J}_B$ is the Bayesian information matrix for the model parameters $\boldsymbol{\Psi}$, and the matrix

$$\boldsymbol{T}_{\boldsymbol{\gamma}} = -E_{\boldsymbol{\Psi}}\left[\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}}\right] \tag{4.17}$$

takes into account the effect of transformation $\boldsymbol{\gamma}$, with $\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}}$ representing the Jacobian matrix of the transformation whose $ij$th element is given by $\frac{\partial \boldsymbol{\gamma}_i(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}_j}$.

## 4.4 Performance limits for instantaneous spectral imaging

We will now use this general framework to characterize the performance limits of the instantaneous spectral imaging technique developed in Chapter 2, as well as to explore the optimized instrument design.

### 4.4.1 Bayesian information matrix

In order to compute the error bounds, we first need to specify the prior and data terms of the Bayesian information matrix given in (4.6) and (4.5). By treating the unknown parameters $\boldsymbol{\Delta}$ and $\boldsymbol{\epsilon}$ as i.i.d. Gaussian random variables as in Section 2.5, the prior term $\boldsymbol{J}_P$ is given by

$$[\boldsymbol{J}_P]_{ij} = \begin{cases} 1/\sigma_\Delta^2 & \text{if } i = j = 3k+2; \\ 1/\sigma_\epsilon^2 & \text{if } i = j = 3k+3; \\ 0 & \text{otherwise} \end{cases} \tag{4.18}$$

where $k = 1, \ldots, M$.

The data term $\boldsymbol{J}_D$ can be obtained by evaluating the expectation of the Fisher information matrix in (4.8) over $\boldsymbol{\Delta}$ and $\boldsymbol{\epsilon}$, either through Monte Carlo simulation or numerical integration. For the computation of the Fisher matrix, the specific form of $\boldsymbol{D}_i(\boldsymbol{\Theta}_i)$ in (4.12) is required, which can be obtained from the forward model in Section 4.2 as follows:

$$\boldsymbol{D}_i(\boldsymbol{\Theta}_i) = \begin{bmatrix} \frac{\partial\, h_i^{a_1}(\boldsymbol{\Theta}_i)}{\partial\, \Delta_i} & \frac{\partial\, h_i^{a_1}(\boldsymbol{\Theta}_i)}{\partial\, \epsilon_i} \\ \frac{\partial\, h_i^{a_2}(\boldsymbol{\Theta}_i)}{\partial\, \Delta_i} & \frac{\partial\, h_i^{a_2}(\boldsymbol{\Theta}_i)}{\partial\, \epsilon_i} \\ \vdots & \vdots \\ \frac{\partial\, h_i^{a_N}(\boldsymbol{\Theta}_i)}{\partial\, \Delta_i} & \frac{\partial\, h_i^{a_N}(\boldsymbol{\Theta}_i)}{\partial\, \epsilon_i} \end{bmatrix} \tag{4.19}$$

If $a_k \neq 0$, the elements of each of these column vectors can be computed as

$$\left[\frac{\partial\, h_i^{a_k}(\boldsymbol{\Theta}_i)}{\partial\, \Delta_i}\right]_j = \frac{f_i}{\sqrt{\pi}\Delta_i}\left(e^{-t_1^2}t_1 - e^{-t_2^2}t_2\right), \tag{4.20}$$

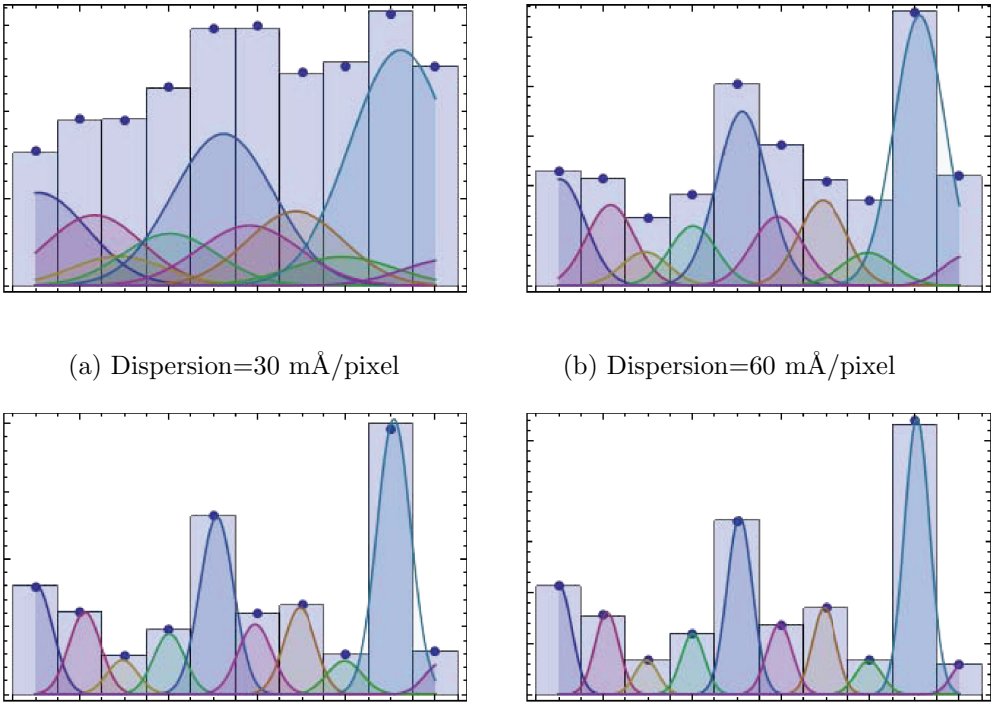
$$\left[\frac{\partial\, h_i^{a_k}(\boldsymbol{\Theta}_i)}{\partial\, \epsilon_i}\right]_j = \frac{f_i\, a_k}{\sqrt{2\pi} \mid a_k \mid \Delta_i}\left(e^{-t_1^2} - e^{-t_2^2}\right) \tag{4.21}$$

for $i, j = 1, \ldots, M$, and zero if $a_k = 0$, with the parameter $t_{1,2}$ given by

$$t_{1,2} = \frac{j - i \mp 1/2 - a_k \epsilon_i}{\sqrt{2} \mid a_k \mid \Delta_i} \tag{4.22}$$

### 4.4.2 Mapping for physical quantities of interest

An important design parameter in the instantaneous spectral imaging setting is the dispersion scale (reciprocal dispersion) $D$, which represents the wavelength range corresponding to a single pixel, and is measured here in mÅ/pixel. Low dispersion scale means large dispersion in the instrument as is illustrated in Fig. 4.1.



(a) Dispersion=30 mÅ/pixel

(b) Dispersion=60 mÅ/pixel

(c) Dispersion=90 mÅ/pixel

(d) Dispersion=120 mÅ/pixel

Figure 4.1: Simulated measurements at different dispersion scales, where colored Gaussians correspond to individual pixel contributions, bars correspond to the total contribution at each pixel, and circles correspond to the observation at that pixel with noise added.

The model parameters $\Delta_i$ and $\epsilon_i$ (measured in pixel units) can be related to the actual physical quantities they represent (measured in physical units) by

using the dispersion scale. More explicitly, the associated physical quantities $w_i$ and $v_i$, respectively denoting the line width in wavelength units and the line-of-sight velocity, can be expressed in terms of the model parameters as follows:

$$w_i = D\,\Delta_i, \tag{4.23}$$

$$v_i = \frac{c(D\epsilon_i)}{1000\lambda_0} \tag{4.24}$$

Here $w_i$ is measured in mÅ, $v_i$ is measured in km/s, $c$ denotes the speed of light in km/s, and $\lambda_0$ represents the central wavelength in Å. The second relation is obtained using the Doppler shift formula: $\frac{\Delta\lambda}{\lambda_0} = \frac{v}{c}$. This gives the relation between the line-of-sight velocity $v$ and the resulting wavelength shift $\Delta\lambda$ of the central wavelength $\lambda_0$.

Then the error bounds for $f_i$, $w_i$, $v_i$ can be obtained from the error bounds for $f_i$, $\Delta_i$, $\epsilon_i$ by using the relation in (4.16). By defining the new parameter set as $\mathbf{\Psi}'_i = [f_i \quad w_i \quad v_i]^\top$, and the old parameter set being $\mathbf{\Psi}_i = [f_i \quad \Delta_i \quad \epsilon_i]^\top$, the transformation involved is given by [19]

$$\mathbf{\Psi}'_i = \gamma(\mathbf{\Psi}_i) = \begin{bmatrix} f_i & D\Delta_i & \dfrac{Dc}{1000\lambda_0}\epsilon_i \end{bmatrix}^\top \tag{4.25}$$

The Jacobian matrix $\frac{\partial\boldsymbol{\gamma}(\mathbf{\Psi})}{\partial\mathbf{\Psi}}$ and also its expectation $\boldsymbol{T}_{\boldsymbol{\gamma}}$ in (4.17) is a block diagonal matrix with each block given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & \frac{Dc}{1000\lambda_0} \end{bmatrix}$$

The error bounds for the physical parameters $\mathbf{\Psi}'$ can then be obtained by using (4.16).

We note that the error bounds for the physical parameters are functions of the instrument design choices including the diffraction orders to be measured, standard deviation of the noise, dispersion scale, and the number of pixels. All of these design choices are significant in determining the amount of information available from the slitless data, and their effect on the reconstruction accuracy can be explored using the derived error bounds. The given formulation is also general enough to allow consideration of other physical

quantities of interest and design parameters if needed.

### 4.4.3 Numerical results for instrument optimization

Thus far, the analytic framework developed is quite general and can be applied to any particular application of instantaneous spectral imaging. In this section, we illustrate the usefulness of these concepts for an application in solar spectral imaging.

We consider the measurements of an EUV emission line emitted from the Sun, with a central wavelength of $\lambda_0 = 195$ Å. Our goal is to explore the effectiveness of the instantaneous spectral imaging technique in estimating the parameters of this emission line over an two-dimensional field-of-view, as well as to find the optimal design choices for the diffraction orders, dispersion scale, noise standard deviation, and number of pixels. The optimality criteria should be chosen based on the science objectives of the experiments performed with the instrument. As an example, we consider here a commonly used criterion: the one that minimizes the average root MSE (RMSE) for the estimates of $f_i$, $w_i$, and $v_i$, where the average is over all pixels $i$. This optimality criterion is evaluated under various observing scenarios with different diffraction orders, and as a function of one design parameter at a time while fixing the others.

To evaluate the tightness of the bounds and performance of the MAP estimation method, we also perform Monte Carlo simulations under different observing scenarios and instrument design considerations. For this, the spectral line parameters are generated randomly according to their modeled distributions in solar spectroscopy, which are described in Section 2.5.1. The averaged RMSE of the MAP estimates are computed by using a total of 50 random parameter sets. Figures 4.2, 4.3 and 4.4 show the Bayesian Cramer-Rao error bounds (straight line) and the errors obtained with the MAP estimator (dotted points) on the same plot for a variety of different observing scenarios.

The plots provide important insights about how to optimally operate this spectral imaging technique in order to minimize the estimation errors. First of all, as shown in Fig. 4.2 there is an optimal dispersion regime to operate, appearing in the range of $40 - 50$ mA/pixel for all scenarios, and this

optimal range is similar for both the line width and line-of-sight velocity estimates. On the other hand, as seen in Fig. 4.3, the number of pixels in the measurements do not significantly affect the average RMSE.

As seen in Fig. 4.4, for the noise-free case with $\sigma = 0$, the true parameter values are always obtained. Moreover, the errors for integrated intensities strongly depend on the noise std since the zeroth order measurement directly provides a noisy observation of integrated intensities. In fact, for the low noise regime (up to a noise std of 5) the intensity estimates obtained with the MAP algorithm do not show significant improvement over the noisy zeroth order measurement. For the line widths and Doppler shifts, the dependence on the noise std is weaker at the high noise regime. This is because, in this regime, the estimation is highly dominated by the priors (rather than the measurements). Also we note that with orders $\{0, +1, -1\}$, the estimation accuracy is comparable to the slit spectroscopy when the noise std is smaller than 4 (corresponding to an SNR of $\sim 50$ when SNR is defined as the ratio of the signal mean to the standard deviation of the noise). To achieve similar accuracy at higher noise levels, more spectral orders (than three) will be needed.

When we compare the various observing scenarios with different orders, the case with the diffraction orders $\{0, +1, -1\}$ operated at a nearly-optimal dispersion appears to be the most cost-effective one. Yet additional fourth and fifth orders provide an improvement of up to 25%. Nevertheless, if operated in an optimal regime with the orders $\{0, +1, -1\}$, the performance of the instantaneous spectral imager is comparable to the conventional method of slit spectrograph (in terms of precision of the estimates), but with the additional advantage of a large instantaneous FOV.

## 4.5 Conclusion

By using Bayesian Cramer-Rao lower bound theory, we have developed a general framework for quantitatively characterizing the performance of the computational spectral imaging techniques (described in the earlier two chapters) in terms of their reconstruction accuracy. The derived error bounds are used to explore the performance limits of the instantaneous spectral imaging technique under various different observing scenarios. The analysis also

provides a framework for exploring the optimal choices of the design considerations. Numerical results indicate that if operated in an optimal dispersion regime with three diffraction orders $\{0, +1, -1\}$, the physical parameters can be estimated with the same order of accuracy as the state-of-the-art slit spectroscopy, but with the added benefit of an instantaneous two-dimensional field-of-view.

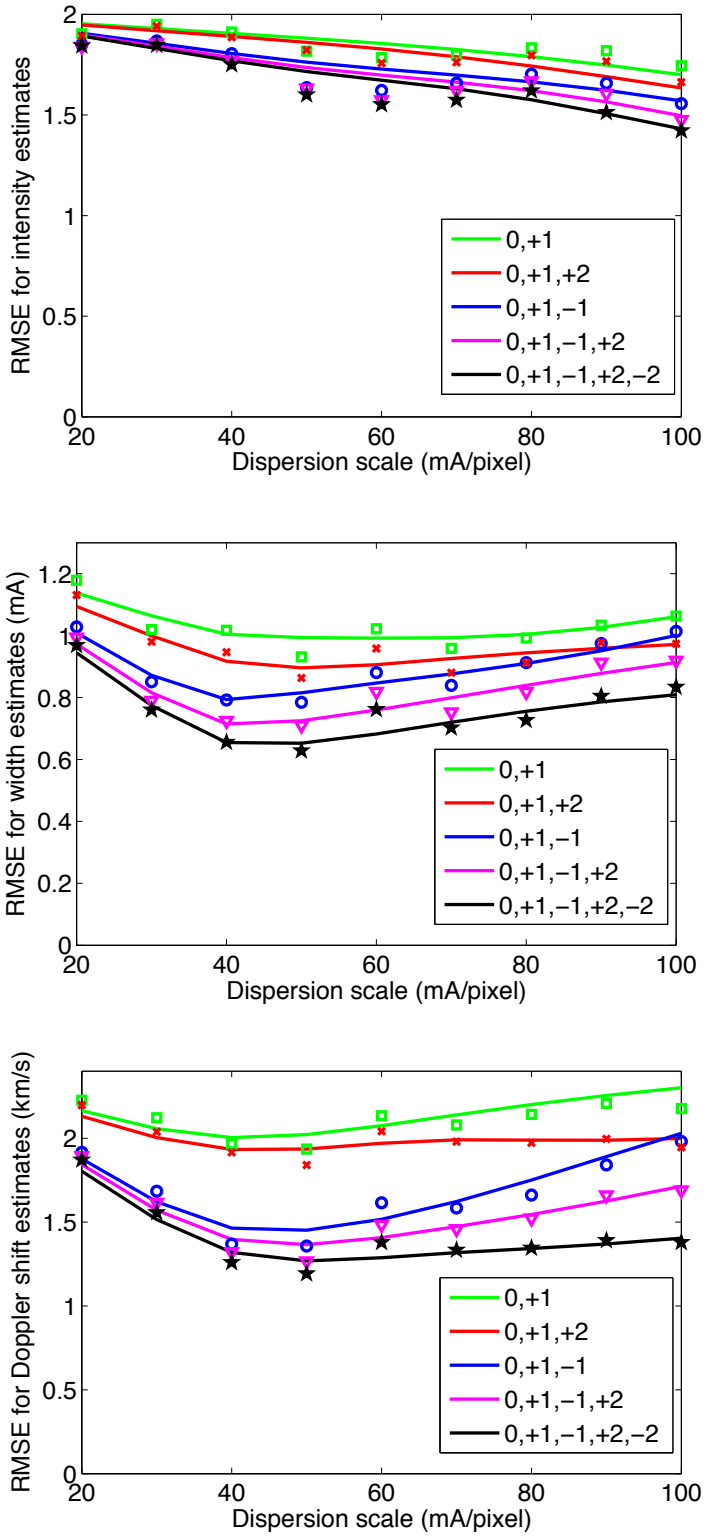



Figure 4.2: Average RMSE for the intensity, width, and Doppler shift estimates as a function of the dispersion scale with $\sigma = 2$ and $M = 10$. The Bayesian Cramer-Rao error bounds are shown with straight lines, and the simulated errors are shown with dotted points, for four different observing scenarios.

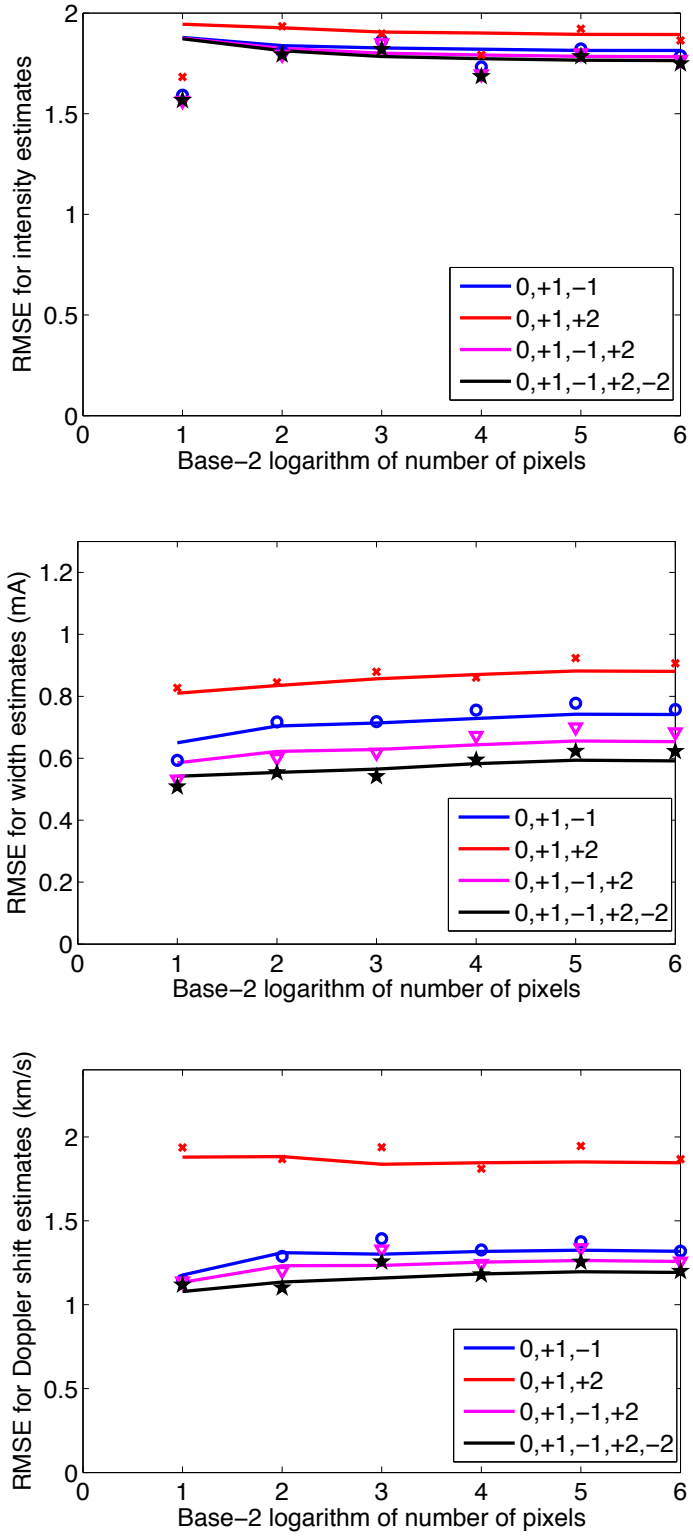



Figure 4.3: Average RMSE for the intensity, width, and Doppler shift estimates as a function of the number of pixels with $D = 40$ m$\mathcal{A}$/pixel and $\sigma = 2$. The Bayesian Cramer-Rao error bounds are shown with straight lines, and the simulated errors are shown with dotted points, for four different observing scenarios.

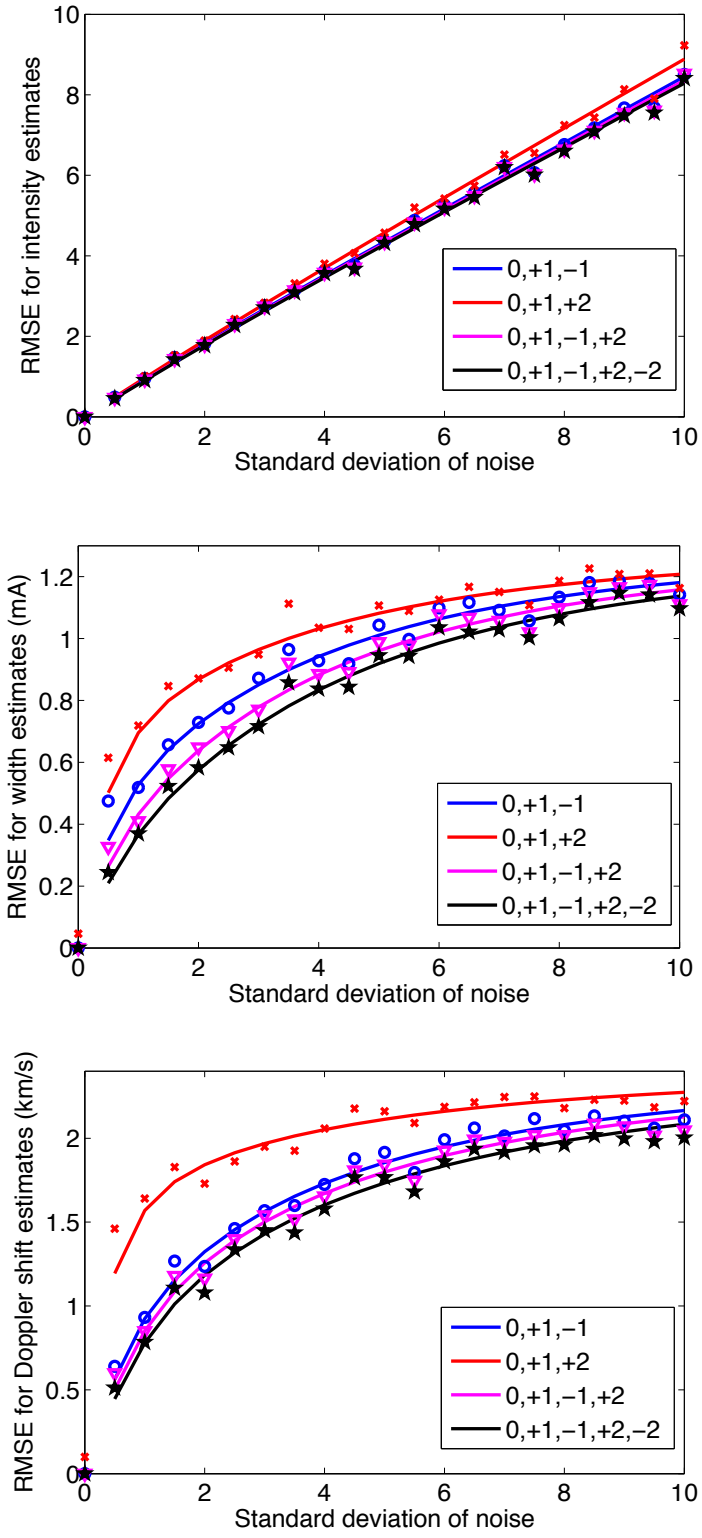



Figure 4.4: Average RMSE for the intensity, width, and Doppler shift estimates as a function of the noise standard deviation with $D = 40$ m$\mathcal{A}$/pixel and $M = 10$. The Bayesian Cramer-Rao error bounds are shown with straight lines, and the simulated errors are shown with dotted points, for four different observing scenarios.

# CHAPTER 5

# COMPUTATIONAL METHODS FOR PHASE RETRIEVAL

## 5.1 Introduction

Phase retrieval problems arise in the photon-sieve imaging setting with coherent illumination. These problems are generalizations of the classical phase retrieval problem, which is the recovery of a signal from the magnitude of its Fourier transform.

In this chapter, we review and compare important algorithms for the classical phase retrieval problem. In particular, we establish the relation of the Schulz-Snyder (SS) phase retrieval algorithm, which has not received much attention to date, to several well-known algorithms, including alternating-minimization, expectation-maximization, and blind Richardson-Lucy algorithms, and gradient-descent methods. These connections allow the algorithm to be seen in a new light, making many of its convergence properties, advantages and drawbacks apparent. The gained understanding yields new insights to improve the algorithm in terms of reliability and speed. In particular, we propose a hybrid method by incorporating annealing-type global optimization methods to avoid convergence to nonglobal solutions. This hybrid method and the connection of the SS algorithm to alternating-minimization have been presented in [66].

## 5.2 Applications

The classical phase retrieval problem appears in a variety of different applications such as diffraction (lensless) imaging, astronomical imaging, x-ray crystallography, and microscopy. One of the most important applications of the problem is in diffraction imaging, in which magnitude of x-ray or electron

diffraction patterns are used to form images of a targeted object [67, chap. 6]. A far-field diffraction pattern is the Fourier transform of the diffracting object; however, only the magnitude, but not the phase, of the diffraction pattern can be measured by some practical light detectors such as charge-coupled devices (CCDs). Then the recovery of the object requires solving the classical phase retrieval problem. This is illustrated in Fig. 5.1.

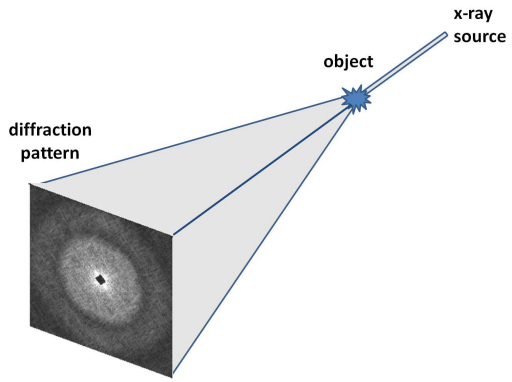



Figure 5.1: Illustration of diffraction imaging.

Another application arises in astronomy where the resolution of a telescope image is limited by the atmospheric turbulence [67, chap. 7]. The solution to the limited resolution lies within a variety of interferometric techniques. With these techniques, it is possible to measure the spatial coherence function in the far field, which is proportional to the Fourier transform of the object intensity. However, it is usually not possible to get an accurate estimate of the phase of the coherence function. Therefore, the recovery of the object intensity is only possible with the tools of classical phase retrieval.

## 5.3 Problem definition and characteristics

To pose the classical problem, let $|F(\mu)|$ denote the Fourier magnitude of an *unknown* function $f(x)$. Then given the noisy observation $|\tilde{F}(\mu)|$ of the Fourier magnitude and some prior information about the function itself, the (inverse) problem can be stated as follows: find a feasible function $f(x)$ whose Fourier magnitude $|F(\mu)|$ is equal to or close to the given Fourier magnitude $|\tilde{F}(\mu)|$ in some sense. Since the autocorrelation of a function and the squared magnitude of its Fourier transform constitute a Fourier transform pair, i.e.

$|F(\mu)|^2 \Leftrightarrow R_f(x) = f(x) * f^*(-x)$, the problem can be equivalently stated as follows: find a feasible function $f(x)$ whose autocorrelation $R_f(x)$, is equal to or close to the given autocorrelation $\tilde{R}(x)$ in some sense. Here solution feasibility refers to satisfying the known constraints (determined from prior information).

Important characteristics of the problem are listed below.

1. The functions $f(x)$, $-f(x)$, $f(x-x_0)$, and $f^*(-x)$ have the same Fourier transform magnitude, where $x_0$ is a real constant. As a result, there exist intrinsic *ambiguities* in sign, translation, and rotation by 180 degrees. These ambiguities will result in a set of functions all consistent with the measured data. However, any of these functions is acceptable as a solution because the goal is generally to obtain the form of the function, but not its orientation. After the recovery process, it might be possible to remove these ambiguities with additional information.

2. Fourier magnitude $|F(\mu)|$ *uniquely* determines $f(x)$ (to within trivial ambiguities) for *almost all* two or higher dimensional real-valued discrete signals with finite support. A sufficient condition for uniqueness is the irreducibility of the z-transform; that is, the function cannot be expressed as the convolution of two or more signals each of which is not a delta function [68, 69]. This condition is almost always satisfied in two (or higher) dimensions because almost all polynomials in two (or more) variables are irreducible [70].

3. Uniqueness still holds for two or higher dimensional real-valued discrete signals with finite support if the Fourier magnitude is known at discrete frequencies, but not at continuum of frequencies. A discrete signal which is zero outside $0 \leq x \leq N-1$ is uniquely defined by the magnitude of its $M$-point DFT provided that $M \geq 2N-1$ [69] (hence twofold oversampling in the Fourier domain for each dimension).

4. The recovery problem is highly *nonlinear* (e.g. nonlinear relation between $f(x)$ and $R_f(x)$).

5. The problem is often *ill-conditioned* [50]; that is, the solution varies rapidly with changes in the given data.

6. Because the problem is equivalent to recovering a signal from its autocorrelation, it is a special case of the blind deconvolution problem, in which a signal is to be recovered from measurements of its convolution with an unknown point spread function. This is so because the unknown point spread function is simply $f^*(-x)$ for the phase retrieval problem.

## 5.4 Existing algorithms

Even when a unique solution exists for the classical phase retrieval problem, solving the problem numerically is widely accepted as difficult. This is due to the quadratic measurements involved and the resulting nonlinear and nonconvex formulations of the inverse problem. To date, there is no practical algorithm with guaranteed exact recovery (equivalently with global convergence). Well-known algorithms for the classical problem suffer from convergence to nonglobal solutions, instability, involving trial and error, and unreliability with missing convergence proofs. Examples are Fienup's iterative algorithms [71], which are widely used, and Schulz-Snyder algorithm [72], both of which have nice properties in terms of image generality and computational efficiency. Here we review and compare these general-purpose phase retrieval algorithms, together with a more detailed analysis of the Schulz-Snyder algorithm since it has not received much attention to date. A review of various other phase retrieval algorithms can be found in [67].

### 5.4.1 Fienup's iterative transform algorithms

To date, Fienup's iterative transform algorithms are regarded as the most successful practical algorithms for solving the phase retrieval problem both in terms of computational complexity, sensitivity to noise, and image generality. These algorithms were initially proposed as heuristic strategies; however, their remarkable success in a wide variety of applications has later motivated their analysis from a theoretical standpoint. The iterative transform algorithms apply projection operators at each iteration in order to find a solution in the intersection of constraint sets. They have two important versions: the error-reduction algorithm (ER) [65] and the hybrid input-output algorithm

(HIO) [71].

The error-reduction algorithm, being the simplest version of the iterative transform algorithms, is mainly a descendant of the Gerchberg-Saxton algorithm [73]. It switches back and forth between the space and frequency domains in each of which the available constraints are imposed. The constraints are often nonnegativity and support in the space domain and known Fourier magnitude in the frequency domain.

The error-reduction algorithm can be viewed in different ways: as a form of gradient-descent method for minimizing the squared error in Fourier magnitudes [71] or as an alternating projection algorithm onto nonconvex sets [67, chap. 8]. It has been shown that the iterations monotonically reduce the squared error criteria [71], and the algorithm is said to converge when the reduction in the error is less than a threshold. The convergence speed is generally very slow, but it guarantees linear convergence [74]. From its equivalence to a gradient-descent method, it is clear that this algorithm cannot guarantee convergence to a global solution. Indeed, convergence to local minima has been commonly observed in many different applications [75, 76].

To avoid convergence of ER to local minima, many variations of the ER algorithm have been proposed [71,74,77,78]. Among these, the hybrid input-output (HIO) algorithm has been recognized empirically to be the most successful. Although its convergence behavior cannot be completely analyzed because of the nonlinearity in projections [79], in practice it often converges to a reasonably good solution for a wide variety of applications. Its ability to escape from local minima has been demonstrated empirically for proper choice of its parameter [74, 77]. Unlike ER which can be interpreted as a *local* minimizer as with all gradient methods, HIO and its variants have been regarded as heuristic *global* minimizers which use feedback to reach a global solution [74].

Unlike ER, HIO does not force the iterates to satisfy the constraints, but it uses the iterates eventually to drive the algorithm to a solution that satisfy the constraints. Hence the algorithm can be interpreted as an alternating projection algorithm with relaxed (space-domain) constraints. It can also be related to a lifted saddle point optimization problem whose fixed point is in the intersection of constraint sets. We quote Marchesini [78] here: "They seek the saddle point of the difference of two antagonistic error metrics with respect to feasible and unfeasible spaces defined by the constraints. They

move along the steepest-descent direction in the feasible space and steepest-ascent direction in the unfeasible space." Note that ER only moves in the steepest-descent direction in the feasible space. An improved version of HIO in terms of reliability and speed has also been proposed by optimizing the step size in the saddle point problem [74, 78].

HIO is empirically observed to be faster than the ER algorithm. Moreover, it has some capabilities of escaping from local minima, and often converges to reasonably good solutions empirically [74]. However, some initializations can cause the algorithm to converge to unsatisfactory solutions associated with local minima [74]. In addition, there is no proof of convergence, and the algorithm is unstable for some parameter values (hence sensitive to parameter choice). Furthermore, it cannot converge in the noisy case because the intersection of the constraint sets is empty and the algorithm seeks for a point in the intersection [79]. For each problem, the best possible combination of HIO and ER should be figured out for stabilization and refinement purposes. Therefore, an element of trial and error is involved and care is required during its application.

Fienup [71] has emphasized that "An approach that would be superior to the ones considered here would be one that minimizes the Fourier-domain error while inherently satisfying the object-domain constraints". The Schulz-Snyder algorithm, which will be considered next, addresses this issue and provides a method that minimizes a certain error criterion while automatically satisfying the nonnegativity, support and sum constraints. An important drawback of the algorithm, however, is that it can converge to nonglobal solutions, as will be discussed in detail later on. But this is a characteristic of all known methods too for the classical phase retrieval problem.

### 5.4.2 Schulz-Snyder phase retrieval algorithm:

An iterative algorithm for recovering *nonnegative* real signals from autocorrelation measurements has been developed by Schulz and Snyder [72]. (Note that the autocorrelation of a function and the squared magnitude of its Fourier transform constitute a Fourier transform pair.) Although this method was initially proposed for recovering signals from any $m$th-order correlation, here we consider only the $m = 2$ case for the phase retrieval. We re-

fer to this information-theoretic image formation algorithm as *Schulz-Snyder algorithm*.

To pose the problem, we borrow the notation mostly from [72]. The signal of interest $\{f(x) : x \in \mathcal{X}\}$ is assumed to be discrete, nonnegative and real-valued with some finite support $\mathcal{X} \subset \mathcal{R}^n$. The autocorrelation of $f$ is then defined as

$$R_f(y) = \sum_{x \in \mathcal{X}} f(x) f(x+y) = \sum_{x \in \mathcal{X}} f(x) f(x-y) \tag{5.1}$$

and its support is given by $\mathcal{Y} \subset \mathcal{R}^n$ where

$$\mathcal{Y} = \{y : y = x_1 - x_2, (x_1, x_2) \in \mathcal{X}^2\} \tag{5.2}$$

If we restrict our attention to two-dimensional images as in many applications, $n$ is 2, and $x$ is a shorthand for the two-dimensional spatial image coordinate vector. Moreover, the supports are represented by finite sets $\mathcal{X} = \{0, 1, \ldots, N-1\} \times \{0, 1, \ldots, M-1\}$ and $\mathcal{Y} = \{-(N-1), \ldots, N-1\} \times \{-(M-1), \ldots, M-1\}$.

Given the noisy autocorrelation measurements $\{\tilde{R}(y) : y \in \mathcal{Y}\}$, the goal is to find a nonnegative signal $f$ whose autocorrelation $R_f$ is equal to or close to the given autocorrelation $\tilde{R}$ in some discrepancy measure $D$. This is equivalent to solving the following optimization problem [72]:

$$\min_{f \geq 0} D(\tilde{R}, R_f) \tag{5.3}$$

The discrepancy measure used for the Schulz-Snyder algorithm is the *Csiszar's distance* (also known as I-divergence) and is defined [80] as

$$D(\tilde{R}, R_f) = \sum_y \tilde{R}(y) \ln \frac{\tilde{R}(y)}{R_f(y)} - \sum_y \tilde{R}(y) + \sum_y R_f(y) \tag{5.4}$$

This information-theoretic measure reduces to the Kullback-Leibler distance [81] (also known as relative entropy) when the nonnegative functions are restricted to probability distributions, or more generally to functions whose sums are equal. To sum up, the Schulz-Snyder algorithm aims to minimize the Csiszar's distance between the measured autocorrelation and the autocorrelation of the estimated signal.

The problem stated here is deterministic and it does not involve any

stochastic functions. Note that minimizing the Csiszar's distance for the deterministic problem is equivalent to maximizing the log-likelihood of the Poisson distributed data. Hence the Schulz-Snyder algorithm can also be related to a maximum-likelihood problem.

The recursion in the Schulz-Snyder algorithm is given by [72]

$$f^{k+1}(x) = f^k(x) \frac{1}{\sqrt{\tilde{R}_0}} \left[ \frac{1/2(\tilde{R}(x) + \tilde{R}(-x))}{R_{f^k}(x)} * f^k(x) \right] \tag{5.5}$$

where $f^k(x)$ is the signal estimate at the $k$th iteration. This recursion, which has been suggested without a formal derivation [72], decreases the Csiszar's distance monotonically and always converges to a limit. In Section 5.5, we present two different formal derivations of the algorithm, lacking in the paper of Schulz and Snyder [72], based on alternating-minimization [66] and expectation-maximization [25, 82] methods. Hence the SS algorithm can be interpreted as an alternating minimization algorithm applied to a lifted non-convex optimization problem (or equivalently an expectation maximization algorithm for maximizing the log-likelihood of Poisson-distributed measurements.) We also discuss its relation to blind Richardson-Lucy algorithm, and gradient-descent type methods.

### 5.4.3 Analysis of the Schulz-Snyder algorithm

In the next section, we establish the relation of the Schulz-Snyder phase retrieval algorithm to several well-known algorithms. These connections allow the Schulz-Snyder algorithm to be seen in a new light, making many of its convergence properties, advantages and drawbacks apparent. Here we list these advantages and drawbacks, some of which were shown by Schulz and Snyder [72], and some of which are new here.

*Some nice features:*

- The Csiszar's distance between observed and estimated autocorrelations, $D(\tilde{R}, R_f)$, (or equivalently the log-likelihood function) is monotonically reduced at each iteration, and converges to a limit. More precisely,

$$D(\tilde{R}, R_{f^k}) - D(\tilde{R}, R_{f^{k+1}}) \geq 2\sqrt{\tilde{R}_0}\, D(f^{k+1}, f^k) \geq 0 \tag{5.6}$$

where $\tilde{R}_0 = \sum_y \tilde{R}(y)$, and with equality if and only if $f^{k+1} = f^k$.

- Nonnegativity and support constraints are automatically satisfied at each iteration provided that the initial estimate satisfies these constraints.

- The iterates also satisfy a sum constraint: $\sum_x f^k(x) = \sqrt{\tilde{R}_0}$. (Any function that satisfies the Kuhn-Tucker conditions also sums to the same constant.)

- The algorithm is easy to implement, applicable to all nonnegative real objects, and computationally tractable. Each iteration requires 4 FFT computations (together with proper zero-padding to avoid circular convolution).

*Some weaknesses:*

- The algorithm does not guarantee convergence to global solutions, like any gradient-descent method applied to the nonconvex optimization problem in (5.3) (nonconvexity will be illustrated later). The algorithm may converge to one of the many stationary points of the objective function, depending on the initialization, and therefore it should be run with multiple initializations. (It is also useful to refine the estimates obtained with a global optimization method.)

- It has slow convergence near a minimum which is inherent to all first-order methods.

We now illustrate both of these weaknesses. We first analyze the algorithm when it is converging to a global solution to illustrate that even when the algorithm converges to a global solution rather than a nonglobal solution, the convergence can be quite slow, especially when it is close to the minimum. For this consider the original image and its autocorrelation shown in Figures 5.2a and 5.2c. The estimate of the image and its autocorrelation obtained after $5 \times 10^4$ iterations are given in Figures 5.2b and 5.2d. Fig. 5.2e shows the Csiszar's distance between true and estimated autocorrelations, which is the objective function to be minimized, as a function of the iteration number. A similar plot for the Csiszar's distance between true and estimated images is given in Fig. 5.2f. As shown in plots, the image estimate $f^k$ and the

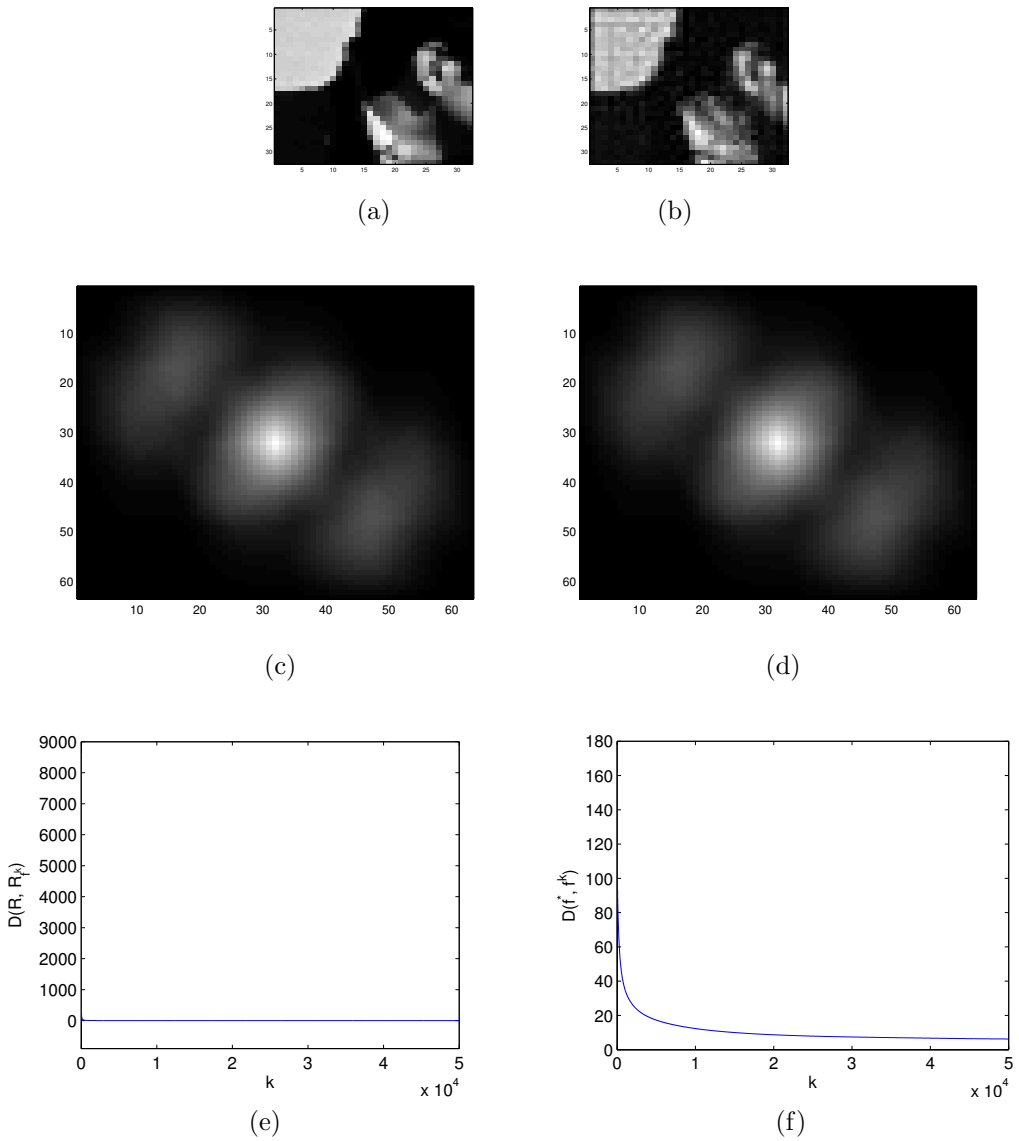



Figure 5.2: (a) Original Image. (b) Autocorrelation of the original image. (c) Image estimate. (d) Autocorrelation estimate. (e) Csiszar's distance between true and estimated autocorrelations. (f) Csiszar's distance between true and estimated images.

autocorrelation estimate $R_{f^k}$ converge with different rates. In particular, the distance between autocorrelations rapidly decreases in the first few iterations and then the autocorrelation estimate is not significantly improved in the subsequent iterations. On the other hand, the image estimate converges slower and the speed of convergence becomes worse as the estimate gets closer to the solution. This indicates that the image estimate fits to the large

scale structure quickly, but to details slowly. This behavior has also been observed for the nonblind Richardson-Lucy algorithm [83], which is also an alternating minimization algorithm.

Secondly, we illustrate the nonconvexity of the optimization problem and the resulting issue of convergence to nonglobal solutions. To illustrate the nonconvexity of the objective function, let us consider a simple two-dimensional image with L-shaped support: $\tilde{f} = [0.1\ 0;\ 0.1\ 0.8]$. Given the autocorrelation $\tilde{R}$ of this image, Fig. 5.3 shows the objective function $D(\tilde{R}, R_f)$ over the set of all feasible images $f$. The feasible set is a triangle containing all L-shaped nonnegative images that sum to one. (We note that the feasible set is always a simplex.) Here we assume that the support is known. As a result, only three variables are unknown, one of which is redundant because of the sum constraint. As shown in the figure, the objective function has two local minimums in addition to the global minimum at $\tilde{f}$. Note that even with a simple image, two local minimums exist. In fact, for practical large images, the number of local minimums will be also large.

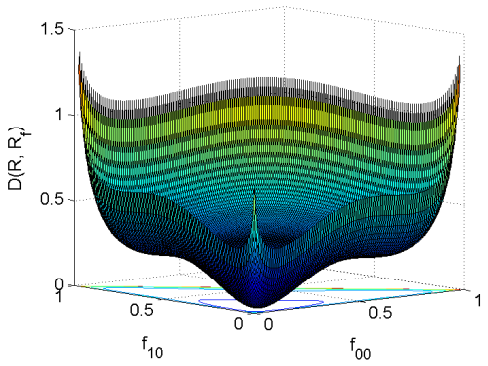



Figure 5.3: Illustration of the nonconvexity of the objective function for a simple 2D phase retrieval problem.

As we will discuss in Section 5.5.4, Schulz-Snyder algorithm is a gradient-descent type method. Since the objective function is also multimodal (i.e. contains more than one minima), the output highly depends on the choice of the initial point. The progress of the algorithm with different initializations is shown in Fig. 5.4. This illustrates that the algorithm can converge to a global or a local minimum depending on the starting point.

We further illustrate convergence to nonglobal solutions for a more realistic image of size $64 \times 64$ shown in Fig. 5.5a. The estimate when initialized with

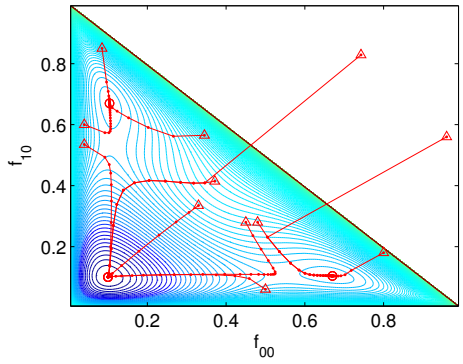



Figure 5.4: Behavior of the algorithm for different starting points. Lower triangular region corresponds to the feasible set. The contour lines represent the objective function $D(\tilde{R}, R_f)$. The global minimum is on the bottom left corner. The local minimums are on bottom right and top left corners.

a uniform image is shown in Fig. 5.5b. This estimate indeed corresponds to a nonglobal solution. This is because plotting the Csiszar's distance between original and estimated images, shortly $D(\tilde{f}, f_k)$, as in Fig. 5.5c shows that after some iterations the image estimate starts moving away from the original image, which indicates its convergence to a point different than the global minimum. Such an increase in $D(\tilde{f}, f_k)$ does not occur when the estimate converges to the global solution. On the other hand, the Csiszar's distance between autocorrelations, which is the objective function here, monotonically decreases with the number of iterations (see Fig. 5.5f). We also note that uniform image is an obvious choice for initialization since it is the center of the feasible set defined by a simplex. With this initialization, the algorithm often converges to a nonglobal solution whose closeness to the global minimum is image-dependent. This example corresponds to a case when the resulting estimate is substantially similar to the original image.

As a final remark, we note that convergence of the Schulz-Snyder algorithm to local minima has been reported before through some numerical experiments [84]. However, it is more appropriate to view these radically different reconstructions as ambiguous nonunique solutions rather than local minimums. The reasoning is that all of these experiments are based on underestimated supports, which intrinsically yield ambiguous nonunique solutions. (It is well known that when the support is underestimated, solving the phase retrieval problem often yields ambiguous nonunique solutions regardless of

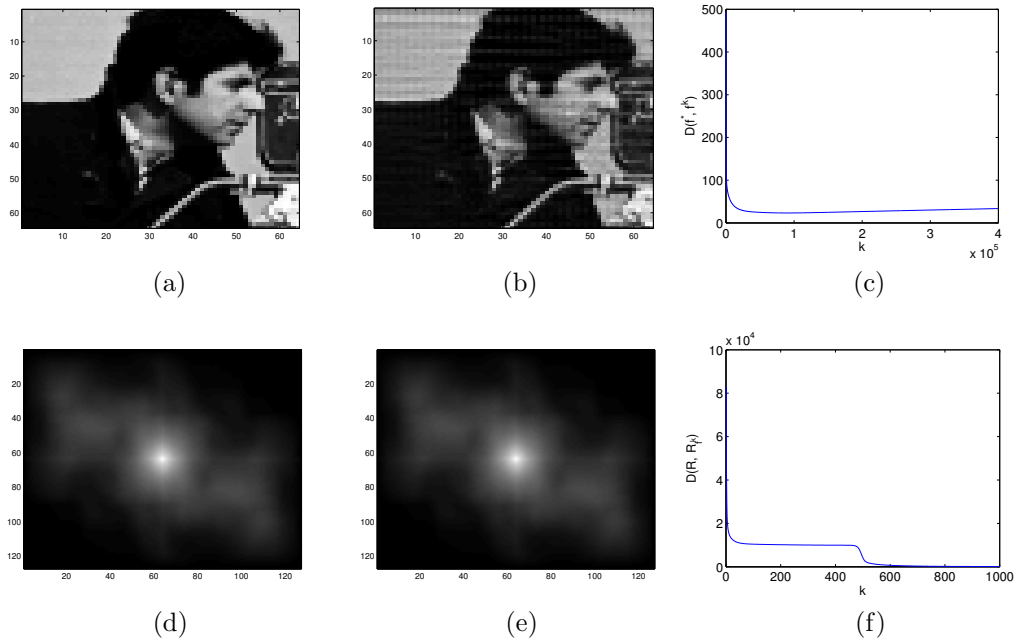


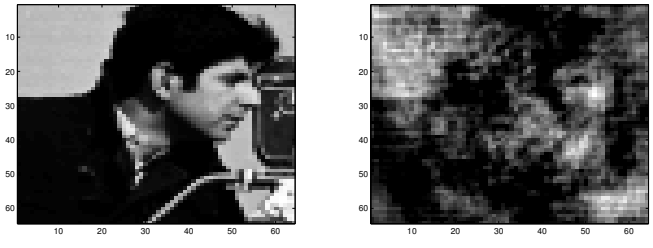

Figure 5.5: (a) Original Image. (b) Nonglobal estimate from uniform initialization (c) Csiszar's distance between the original and estimated images as a function of iterations. (d) Noiseless autocorrelation measurement. (e) Autocorrelation of the estimated image. (f) Csiszar's distance between autocorrelations as a function of iterations.

the algorithm used [85–87].)

### 5.4.4 Comparison of SS algorithm with Fienup's algorithms

Here we compare the Schulz-Snyder algorithm with the widely used Fienup's algorithms (ER and HIO). To the best of our knowledge, such a comparison has not been presented in the literature. Figures 5.6, 5.7, 5.8, and 5.9 show some sample reconstructions from noiseless autocorrelation measurements when SS, ER, and HIO algorithms are used with uniform image initialization. The parameter of the HIO algorithm is set to $\beta = 0.5$.

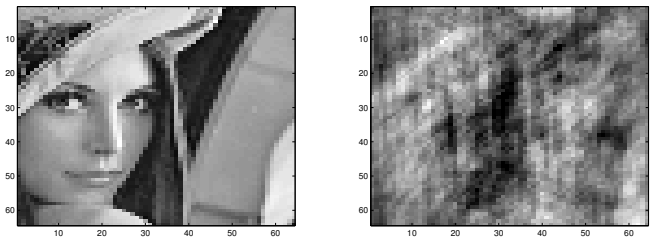
For comparison, we first group these three algorithms into two categories: local optimization methods and heuristic global optimization methods. As noted before, both ER and SS belong to the first category of local optimization methods since both are gradient-descent type methods applied to different nonconvex formulations of the problem (the former with the squared



(a) Original image (b) ER estimate

(c) HIO estimate (d) SS estimate

Figure 5.6: SS algorithm versus Fienup's algorithm.

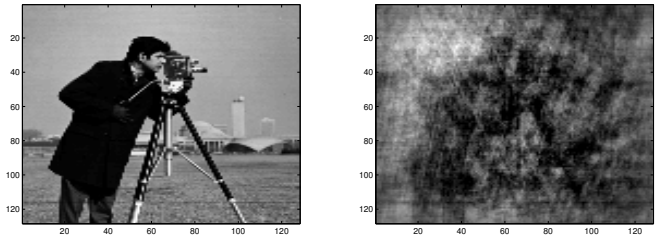


(a) Original image (b) ER estimate

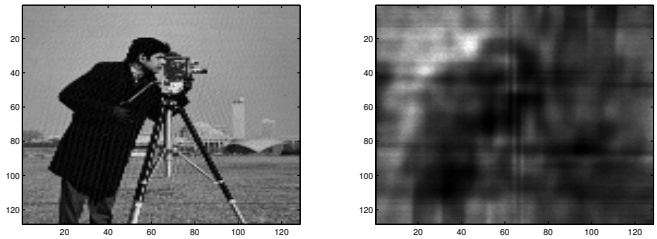
(c) HIO estimate (d) SS estimate

Figure 5.7: SS algorithm versus Fienup's algorithm.

error and latter with Csiszar's distance). As a result, both algorithms suffer from converging to nonglobal solutions depending on the initialization.

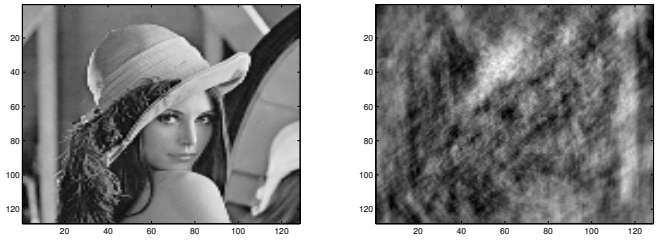


(a) Original image (b) ER estimate

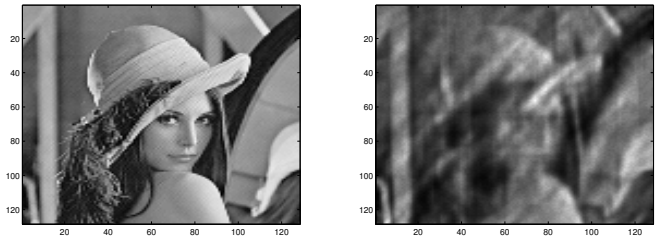


(c) HIO estimate (d) SS estimate

Figure 5.8: SS algorithm versus Fienup's algorithm.



(a) Original image (b) ER estimate



(c) HIO estimate (d) SS estimate

Figure 5.9: SS algorithm versus Fienup's algorithm.

Another common property is that they both satisfy the space-domain constraints (support and nonnegativity) at every iteration. These two algorithms

differ in terms of computational cost; the ER and SS algorithms respectively require 2 and 4 FFTs in each iteration. In terms of reconstruction quality, the SS algorithm often yields better estimates than the ER algorithm (as also seen in sample reconstructions). The ER algorithm usually quickly converges to a nonglobal solution (in less than 1500 iterations for images of size no larger than $128 \times 128$). On the other hand, the SS algorithm is much slower and usually requires more than 10000 iterations to converge to a solution, although its estimates are often better than the estimates of the ER (though it is still a nonglobal solution).

The HIO algorithm, on the other hand, belongs to the second group of heuristic global optimization methods; it is a heuristic variation of the ER algorithm to achieve global convergence. In particular, it relaxes the space domain constraints in ER so that both the reconstructions and the convergence speed are improved over ER (it often converges less than 4000 iterations). This algorithm, although sensitive to the choice of the parameter, offers the possibility of obtaining estimates close to the global solution in an efficient way. As illustrated in the presented numerical results, reconstructions from noiseless measurements are often very similar to the original images, showing a superior reconstruction quality than both ER and SS algorithm. This result is not very surprising noting the differences in their purposes: local optimization versus global optimization. However, this at the same time suggests that exploring the variations of the SS algorithm to achieve global convergence may compete with the HIO algorithm, which is a heuristic variation of the ER algorithm that performs worse than SS.

In Section 5.6, we will explore this possibility by incorporating annealing-type global optimization methods into the gradient-descent type SS algorithm intending to achieve global convergence. Before we exploit the gradient-descent nature of the algorithm, we first discuss the connections of the Schulz-Snyder algorithm to several well-known methods including gradient-descent type methods, alternating-minimization, expectation-maximization, and blind Richardson-Lucy algorithms.

## 5.5 Connection of Schulz-Snyder algorithm to well-known methods

There is an inherent relation between blind Richardson-Lucy algorithm, expectation-maximization, alternating-minimization and gradient-descent type methods. First of all, the blind Richardson-Lucy algorithm has been derived based on the EM algorithm [88]. In addition, many authors noted that the EM can be viewed as an alternating-minimization procedure [89,90] or as a gradient-descent type algorithm [90]. In this section, we will point out the relationship of Schulz-Snyder algorithm to all these algorithms. In the meantime, two different formal derivations of the algorithm, lacking in the paper of Schulz and Snyder [72], based on alternating-minimization [66] and expectation-maximization [25, 82] methods will be presented.

### 5.5.1 Alternating-minimization method

Several well-known algorithms including the Blahut-Arimoto algorithm for computing the channel capacity, Blahut's algorithm for computing the rate-distortion function, expectation-maximization algorithms including the Richardson Lucy nonblind deconvolution algorithm can all be described as alternating-minimization procedures (see [90] for a unified viewpoint). Here we arrive at the Schulz-Snyder algorithm using an alternating-minimization approach and show it to be an alternating minimization algorithm applied to nonconvex optimization [66]. This will also constitute a formal derivation of the algorithm since it has been given by Schulz and Snyder without a formal derivation.

To apply the alternating minimization approach, we first lift the phase retrieval problem to a higher space. Consider the optimization problem given in (5.3), where a minimum must also satisfy $\sum_x f(x) = \sqrt{\tilde{R}_0}$ [72]. Hence we aim to solve the following optimization problem:

$$\min_{\substack{\sum_x f(x)=\sqrt{\tilde{R}_0} \\ f\geq 0}} \sum_y \tilde{R}(y) \ln \frac{\tilde{R}(y)}{R_f(y)} - \sum_y \tilde{R}(y) + \sum_y R_f(y) \tag{5.7}$$

Using the convex decomposition lemma [50, p. 382], we lift $R_f(y)$ by express-

ing it as a minimum in a larger space:

$$-\ln R_f(y) = \min_{\substack{\sum_x h(x|y)=1 \\ h(x|y)\geq 0}} \sum_x h(x|y) \ln \left( \frac{h(x|y)}{f(x)f(x+y)} \right) \tag{5.8}$$

This equality follows directly from the application of Jensen's inequality to convex functions, in particular to negative logarithm. Substituting (5.8) in $D(\tilde{R}, R_f)$, and using the constraints $\sum_x h(x|y) = 1$ and $\sum_x f(x) = \sqrt{\tilde{R}_0}$, we obtain

$$D(\tilde{R}, R_f) = \min_{\substack{\sum_x h(x|y)=1 \\ h(x|y)\geq 0}} \sum_y \sum_x \tilde{R}(y) h(x|y) \ln \frac{\tilde{R}(y)h(x|y)}{f(x)f(x+y)} \tag{5.9}$$

This relation is one of the variational representations in [90] obtained by minimizing the joint I-divergence.

Using this variational representation in (5.7) leads to a double minimization problem and motivates the alternating-minimization algorithm for the phase retrieval problem (Joseph A. O'Sullivan, personal communication, May 2010). The idea is to analytically solve the problem over each single variable, and then iterate by alternately minimizing. Here we alternate by minimizing the lifted term in (5.9) over $f(x)$ and $h(x|y)$. Note that this term is convex in $f$ for fixed $h$ and convex in $h$ for fixed $f$, but not convex in the pair $(f, h)$ as will be showed later.

To minimize over $h$ for fixed $f$, the necessary and sufficient condition is explicitly given by

$$h(x|y) = \frac{f(x)f(x+y)}{\sum_x f(x)f(x+y)} \tag{5.10}$$

Similarly to minimize over $f$ for fixed $h$, the condition is

$$f(x) = \frac{\sum_y \tilde{R}(y)(h(x|y) + h(x-y|y))}{2\sqrt{\tilde{R}_0}} \tag{5.11}$$

These equations are obtained from Kuhn-Tucker conditions. The first condition also follows directly from the equality condition in Jensen's inequality.

The iterations in the alternating-minimization are then given by [66]

$$h^k(x|y) = \frac{f^k(x)f^k(x+y)}{\sum_x f^k(x)f^k(x+y)} \tag{5.12}$$

$$f^{k+1}(x) = \frac{\sum_y \tilde{R}(y)(h^k(x|y) + h^k(x-y|y))}{2\sqrt{\tilde{R}_0}} \tag{5.13}$$

where $f^0$ is any positive function inside the known support. Merging these two iterations into one gives exactly the recursion (5.5) in the Schulz-Snyder algorithm. Thus, we arrived at the Schulz-Snyder algorithm by following a rigorous derivation.

Because the objective function lifted to higher dimensions is not convex in pair $(f, h)$, clearly the alternating-minimization procedure cannot guarantee convergence to a global solution. Nonconvexity in pair $(f, h)$ follows directly from the nonconvexity of the original objective function $D(\tilde{R}, R_f)$ (as illustrated in Fig. 5.3), because convexity of the original objective function is a necessary condition for the convexity of the lifted objective function as showed in the Appendix.

Based on alternating minimization derivation, the Schulz-Snyder algorithm can be viewed as upper bound minimization, as similar to the discussion in [91]. In addition, we can interpret the algorithm from the system viewpoint, considering the autocorrelation function $R_f(y)$ as the output of the filter $f(-x)$ with input $f(x)$. Because $h(x|y) = f(x)f(x+y)/R_f(y)$, we have

$$\sum_y h(x|y)R_f(y) = \sqrt{\tilde{R}_0}f(x) \tag{5.14}$$

That is, the auxiliary function $h(x|y)$ is an inverse filter that gives $f(x)$ from the autocorrelation $R_f(y)$. Similarly, because we have

$$\sum_y h(x-y|y)R_f(y) = \sqrt{\tilde{R}_0}f(x) \tag{5.15}$$

we can also consider $h(x-y|y)$ as an inverse filter. Hence, if we knew $h(x|y)$, we could simply compute $f(x)$ by passing the given autocorrelation measurements $\tilde{R}(y)$ through the filters $h(x|y)$ and/or $h(x-y|y)$. However, these filters also depend on the unknown function $f(x)$.

Nevertheless, we can interpret the alternating minimization algorithm based

on this observation. At every iteration $k$, an inverse filter $h^k(x|y)$ is computed based on the current estimate $f^k(x)$. Then $\tilde{R}(y)$ is passed through the filters $h^k(x|y)$ and $h^k(x-y|y)$ as given in (5.13). The output is the new estimate $f^{k+1}(x)$. If $h^k(x|y)$ is the true inverse filter, then the estimate $f^{k+1}(x)$ will be the same as the previous estimate $f^k(x)$; that is, the inverse filters will recover the true image. Otherwise, the steps will be repeated until a fixed point is reached. Note that all local minimizers are also fixed points of the recursion.

### 5.5.2 Expectation-maximization method

The applicability of the EM to all deterministic *linear* inverse problems with positivity constraints has been described in [82, 92]. This is motivated by the observation that such deterministic problems (although lacking stochastic components) can be interpreted as statistical estimation problems from incomplete data. As a result, these problems can be related to the maximum-likelihood estimation (MLE) and thus solved with the EM. Moreover, the resulting EM algorithm converges to the minimizer of the objective function of the deterministic problem when the error is formulated by using the Csiszar's distance.

In this study we apply the same observation to our *nonlinear* deterministic problem. This similarly allows us to interpret the problem as a statistical estimation problem from incomplete data. Then EM method is used to solve the resulting MLE problem. As we will see, the EM algorithm applied to this maximum-likelihood problem is the same as the Schulz-Snyder algorithm that results from the minimization of the Csiszar's distance. All properties of the EM algorithm when applied to linear inverse problems are shared with the Schulz-Snyder algorithm, excluding the global convergence property due to the nonlinearity (and nonconvexity) involved.

We note that the approach here is different than assuming a statistical model for the noise and deriving an EM algorithm for the resulting stochastic problem. We have earlier seen that the AM algorithm derived for the deterministic problem formulated with Csiszar's distance is equivalent to the EM algorithm derived for the stochastic problem formulated with Poisson noise. Since this relation is obvious now, we do not repeat the same dis-

cussion here. Instead we focus on the deterministic problem and show how to relate it to an independent stochastic problem (without assuming a noise model for the original problem). We then derive an EM algorithm using this relation, which again turns out to be the Schulz-Snyder algorithm.

In the absence of statistical noise, the phase retrieval problem requires a solution $f$ to the nonlinear system of equations $\tilde{R}_j = \sum_{i=0}^{N-1} f_i f_{i+j}$ for $j \in \{-(N-1), \ldots, N-1\}$. Here $\mathcal{X} = \{0, 1, \ldots, N-1\}$, $\tilde{R}$ is known and $f$ is positive. Note that we change the notation for an easier interpretation of images as probability mass functions. By assuming that $\tilde{R}$ is positive and summable, i.e. $0 < \sum_j \tilde{R}_j < \infty$, the problem can be rescaled by $\sum_j \tilde{R}_j$. Therefore, without loss of generality, we can assume that $\sum_j \tilde{R}_j = 1$ and consequently $\sum_i f_i = 1$. This observation allows us to regard $f$ and $\tilde{R}$ as probability mass functions (pmfs).

The statistical interpretation of the phase retrieval problem can then be posed as follows. Let $X$ and $Y$ be independent random variables with pdf $f$. Then, $Z = Y - X$ has the distribution $f_i * f_{-i}$, which is also equal to known $\tilde{R}$. The goal is now to estimate $f$ from infinitely many iid observations of $Z$. Here $\{Z_k\}$ for $k = 1, \ldots, K$ is the incomplete-observed data whereas $\{X_k, Z_k\}$ can be chosen as the complete-unobserved data with $K$ being the total number of measurements.

Let us first formulate the MLE of $f$ based on incomplete data $\{Z_k\}$. The log-likelihood function in this case is

$$\begin{aligned} \Lambda(f) &= \log \prod_{k=1}^{K} P(Z_k = z_k), \\ &= \sum_j K_j \log \sum_i f_i f_{i+j}, \\ &\propto \sum_j \hat{\tilde{R}}_j \log \sum_i f_i f_{i+j} \end{aligned} \tag{5.16}$$

where $K_j$ is the number of times the outcome $Z = j$ is observed and $\hat{\tilde{R}}_j$ is its relative frequency. As $K \to \infty$, from SLLN the empirical distribution $\hat{\tilde{R}}$ converges to $\tilde{R}$. With this connection, it is now clear that maximizing the likelihood in (5.16) is equivalent to minimizing the Csiszar's distance in (5.3). As we have discussed in Section 5.4.2, an analytical solution to this problem does not exist (without expressing it as a double minimization). In order to

solve the MLE analytically, we will first pretend that the complete data are available. Then the estimate of the complete data will be obtained from the previous iteration of the EM algorithm. Note that compared to alternating-minimization approach, this involves a different way of lifting the problem in order to solve it explicitly.

To formulate the MLE of $f$ from complete data $\{X_k, Z_k\}$, the log-likelihood can be expressed as follows:

$$\Lambda(f) = \sum_j \sum_i K_{ij} \log(f_i f_{i+j}) \tag{5.17}$$

where $K_{ij}$ is the number of times the outcome $\{Z = j, X = i\}$ is observed. The goal is to maximize (5.17) over $f$ subject to $\sum_i f_i = 1$ and $f > 0$. Taking the derivative of the Lagrangian function with respect to $f_i$ gives the MLE of $f_i$ as follows:

$$\hat{f}_i = \frac{\sum_k \mathbf{1}_{\{X_k=i\}}}{2K} + \frac{\sum_k \mathbf{1}_{\{X_k+Z_k=i\}}}{2K} \tag{5.18}$$

Because $X_k$'s are unobserved, we need to find the expectation of $\mathbf{1}_{\{X_k=i\}}$ and $\mathbf{1}_{\{X_k+Z_k=i\}}$ given $Z_k = z_k$:

$$E[\mathbf{1}_{\{X_k=i\}}|Z_k = z_k] = \frac{P(Z_k = z_k|X_k = i)f_i}{\sum_i P(Z_k = z_k|X_k = i)f_i} \tag{5.19}$$

Then, the expectation of the first term in (5.18) is given by

$$\frac{1}{2K}\sum_k \frac{P(Z_k = z_k|X_k = i)f_i}{\sum_i P(Z_k = z_k|X_k = i)f_i} = \frac{1}{2}\sum_j \frac{\hat{\hat{R}}_j f_i f_{i+j}}{\sum_i f_i f_{i+j}}$$

Similarly, the expectation of the second term in (5.18) is

$$\frac{1}{2K}\sum_k \frac{P(Z_k = z_k|X_k = i - z_k)f_{i-z_k}}{\sum_i P(Z_k = z_k|X_k = i - z_k)f_{i-z_k}} = \frac{1}{2}\sum_j \frac{\hat{\hat{R}}_j f_i f_{i-j}}{\sum_i f_i f_{i-j}}$$

Substituting these in (5.18) results in the following EM iteration for the phase retrieval problem:

$$f_i^{k+1} = \frac{1}{2} f_i^k \sum_j \frac{\hat{\hat{R}}_j (f_{i+j}^k + f_{i-j}^k)}{\sum_i f_i^k f_{i+j}^k} \tag{5.20}$$

As the number of samples increases to infinity, from SLLN the empirical

distribution $\hat{\tilde{R}}$ approaches the true distribution $\tilde{R}$, which is known. This is equivalent to saying that the known true distribution $\tilde{R}$ can be used instead of $\hat{\tilde{R}}$. If we further scale back the functions $\tilde{R}$ and $f$, the EM iteration in (5.20) reduces to the iteration in the Schulz-Snyder algorithm (see (5.5)).

### 5.5.3 Blind Richardson-Lucy algorithm

The blind deconvolution problem is the recovery of a signal from noisy measurements of its convolution with an unknown point spread function. The blind Richardson-Lucy algorithm has been developed to solve this problem, which can be derived based on EM algorithm or minimization of Csiszar's distance [50, 88, 93, 94]. The phase retrieval problem is a special case of the blind deconvolution problem. Moreover, the Schulz-Snyder algorithm closely resembles the blind RL algorithm both in terms of its recursion and derivation. Thus one would expect connections between these two algorithms, which will be the subject of this section.

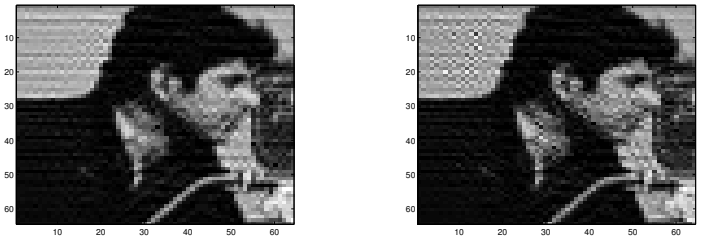
If $\{v(x) : x \in \mathcal{Y}\}$ denote the given measurements of the convolution of unknown functions $f(x)$ and $h(x)$ with $\sum_y v(x) = v_0$, without loss of generality we can impose that $\sum_x h(x) = \sqrt{v_0}$ and consequently $\sum_x f(x) = \sqrt{v_0}$. Then the iterations in the blind Richardson-Lucy algorithm are given by

$$h^{k+1}(x) = h^k(x) \frac{1}{\sqrt{v_0}} \left[ \frac{v(x)}{h^k(x) * f^k(x)} * f^k(-x) \right] \tag{5.21}$$

$$f^{k+1}(x) = f^k(x) \frac{1}{\sqrt{v_0}} \left[ \frac{v(x)}{h^k(x) * f^k(x)} * h^k(-x) \right] \tag{5.22}$$

The phase retrieval problem is a blind deconvolution problem with $h(x) = f(-x)$. Even without using this knowledge, we can apply the blind Richardson-Lucy algorithm to the given autocorrelation measurements $\tilde{R}(y)$. Then in the case of global convergence (which is not guaranteed by the algorithm), it will produce estimates of the image and its mirror and we will have $h^k(x) \to f(-x)$ and $f^k(x) \to f(x)$, or vice versa. However, when the algorithm does not converge to a global solution, the reconstructed functions $h$ and $f$ are not even required to be mirror images of each other (see Figures 5.10a and 5.10b for the recovery of the image in Fig. 5.5a using the blind RL algorithm). For example it has been observed with a nonideal ini-

tial guess that the reconstruction of $h$ can result in a delta-function while the reconstruction of $f$ is just the observed convolution $v$. Thus, the blind Richardson-Lucy algorithm applied directly to the phase retrieval problem can converge to a different set of stationary points than the Schulz-Snyder algorithm. We would expect this set to be larger since the additional information that $h(x) = f(-x)$ is not exploited.



(a) (b)

Figure 5.10: Phase retrieval using blind Richardson-Lucy algorithm: (a) $f(x)$, (b) $h(-x)$.

On the other hand, if we use the additional information that $h(x) = f(-x)$ in the blind Richardson-Lucy algorithm, we can obtain the Schulz-Snyder algorithm exactly. To see this, substitute $h^k(x) = f^k(-x)$ in (5.21) and (5.22) and make the change of variable $x' = -x$ in the first equality to obtain

$$f^{k+1}(x) = f^k(x)\frac{1}{\sqrt{\tilde{R}_0}}\left[\frac{\tilde{R}(-x)}{f^k(x) * f^k(-x)} * f^k(x)\right] \tag{5.23}$$

$$f^{k+1}(x) = f^k(x)\frac{1}{\sqrt{\tilde{R}_0}}\left[\frac{\tilde{R}(x)}{f^k(x) * f^k(-x)} * f^k(x)\right] \tag{5.24}$$

In the absence of noise, i.e. when $\tilde{R}(x) = \tilde{R}(-x)$, these iterations are essentially the same iteration, which is equivalent to the Schulz-Snyder iteration (consider (5.5) with the noise-free condition). In the case of noisy measurements, i.e. when $\tilde{R}(x) \neq \tilde{R}(-x)$, these iterations give rise to different updates. In this circumstance one approach would be to repeat one iteration after the other. Another approach is to merge the iterations into one by summing them together. This second approach will exactly result in the Schulz-Snyder algorithm.

### 5.5.4 Gradient-descent methods

For completeness, we also review the relationship between the Schulz-Snyder algorithm and gradient-descent methods [72]. The derivative of the objective function $D(\tilde{R}, R_f)$ with respect to $f(x)$ is given by

$$\frac{\partial D(\tilde{R}, R_f)}{\partial f(x)} = -\sum_y \frac{\tilde{R}(y)}{R_f(y)}(f(x+y) + f(x-y)) + 2\sqrt{\tilde{R}_0} \tag{5.25}$$

Then, the recursion in Schulz-Snyder algorithm as given in (5.5) can be rewritten as follows [72]:

$$\begin{aligned} f^{(k+1)}(x) &= \frac{1}{2\sqrt{\tilde{R}_0}} f^{(k)}(x) \left( 2\sqrt{\tilde{R}_0} - \frac{\partial D(\tilde{R}, R_{f^k})}{\partial f^k(x)} \right), \\ &= f^{(k)}(x) - \frac{1}{2\sqrt{\tilde{R}_0}} f^{(k)}(x) \frac{\partial D(\tilde{R}, R_{f^k})}{\partial f^k(x)} \end{aligned} \tag{5.26}$$

This recursion is equivalent to a gradient-descent type approach because the term multiplying the gradient is nonnegative. Note that the step size is not a constant, but is proportional to the current value of the function. In this sense, the algorithm is equivalent to a *weighted* gradient-descent method. We further note that equation (5.26) shows the recursion in terms of a main term and a correction term. The correction term is zero if and only if the gradient is zero provided the function is nonzero.

Two weaknesses of the Schulz-Snyder algorithm become more clear by its interpretation as a first-order gradient-descent type method. These are its

1. inability to guarantee global convergence like all gradient search methods applied to nonconvex problems,

2. slow convergence near the minimum that is inherent to all first-order methods.

Although second-order methods like Newton's method converge faster, they require computation, inversion, and storage of the Hessian, which will be prohibitive for large phase retrieval problems. Furthermore, although other first-order gradient methods such as conjugate gradient might also provide faster convergence, they additionally require line search. Therefore, the advantage of the Schulz-Snyder algorithm over standard gradient-descent methods is

that it decreases the objective function monotonically without requiring a step size computation (and while satisfying the image constraints automatically). This advantage has been also mentioned for a more general class of algorithms, known as concave-convex procedure [95].

## 5.6 Global optimization methods

The phase retrieval problem generally requires solving a nonconvex optimization problem with multiple minimums. However, many of the proposed phase retrieval algorithms are either directly based on gradient methods or they can be interpreted as gradient methods. As a result, such algorithms suffer from convergence to nonglobal solutions. In this section, we apply global optimization methods to the classical phase retrieval problem in order to obtain approximate solutions for the global minimum. We propose two methods:

- a simulated annealing algorithm applied to the nonconvex optimization problem involving minimization of the Csiszar's distance,
- a hybrid method to combine the computational efficiency of Schulz-Snyder algorithm with the global convergence ability of annealing-type algorithms.

The first approach requires high computational effort for large images since the optimization involves a large number of unknowns. To reduce the computational complexity, a hybrid of efficient Schulz-Snyder algorithm and global optimization methods is also proposed. For example, a straightforward hybrid is to use the output of SS algorithm as an input to a global optimization method, or vice versa, which are known as multistage methods that consist of a global and a local phase [96]. Here we propose a better hybrid method [66] which is promising in terms of improving the reliability and reducing the computational cost.

### 5.6.1 Related work

The simulated annealing (SA) method was previously applied to the classical phase retrieval problem [97]. The objective function to be minimized was

chosen as the Euclidean distance between measured and estimated autocorrelations. This method has been reported to be reliable and successful in producing approximate global solutions even in the presence of noise. However, the major drawbacks are its high computational cost and slow convergence rate. In particular, the method is not applicable to images of size larger than 64. Moreover, for each different image, care must be exercised when choosing the parameters such as the cooling schedule and scale of perturbation (for example, the number of local minimums increases dramatically with the size of the image [98]). A general rule needs to be established for the cooling strategy. This method was also generalized to the blind deconvolution problem in [99], which is reported to suffer from the same problem of computational inefficiency [87].

To improve the convergence rate of the SA algorithm, it was also proposed to first run the Fienup's algorithm and then use its output as the starting point for the SA [98]. This combination, in general, allows the SA to start with a lower temperature value, which in turn speeds up the convergence. However, the performance of the algorithm highly depends on the performance of the Fienup's algorithm, which might vary with different images and different initializations. Each time the cooling schedule needs to be adjusted accordingly; otherwise, the SA algorithm can be trapped in a local minimum.

### 5.6.2 Simulated annealing algorithm applied to minimization of Csiszar's distance

There exist many deterministic and stochastic global optimization methods. Deterministic methods (such as branch and bound) are more reliable than the stochastic methods since they can guarantee convergence to a global solution within a specified tolerance value. However, these methods are often very slow, and thus not applicable to high dimensional problems [30]. In fact, the phase retrieval problem generally requires an optimization over a high number of unknowns. When an image of $N$ by $M$ pixels is to be reconstructed, the number of unknowns is $N \times M$, which will be a huge number for large images.

This fact suggests the use of stochastic global optimization methods in-

stead. These methods can give fairly good solutions in a quicker way, but with a weaker convergence guarantee; namely, they ensure convergence in a probabilistic sense [96]. The possible candidates are cross entropy method, simulated annealing, and particle swarm optimization, among many others [96].

Our choice here is the simulated annealing algorithm with Metropolis criterion. This choice is due to its ease of implementation and its considerable success in many image recovery problems. The simulated annealing algorithm is to be applied to the same nonconvex optimization problem as in Schulz-Snyder algorithm. We repeat it here for convenience:

$$\begin{aligned} \min_{f} \quad & D(\tilde{R}(x), f(x) * f(-x)) \\ \text{subject to} \quad & \sum_{x} f(x) = \sqrt{\sum_{x} \tilde{R}(x)} \\ & f \geq 0 \end{aligned}$$

To apply the method, we need to specify the following:

1. generation of new candidate points,
2. a way to handle the constraints,
3. a cooling schedule and its parameters.

All these choices can dramatically affect the performance of the algorithm in terms of computational efficiency or reliability. Therefore, care should be taken for their determination.

**1. Generation of new candidate points satisfying the constraints:** In every iteration of the simulated annealing algorithm, a new candidate point satisfying the constraints is generated through random perturbation of the previous estimate. Common approaches for random perturbation rely on generating random variables that are (i) Gaussian distributed, (ii) Cauchy distributed, (iii) uniformly distributed in a hypersphere, (iv) uniformly distributed in a hypercube [31, 100, 101]. Another variation is to change all the components of a point at once or only one component at a time. Note that new candidate points generated in one of the above mentioned ways do not necessarily satisfy the constraints. This additionally requires proper

handling of the constraints. Common approaches for this are the following: (i) acceptance-rejection method, (ii) projection, (iii) penalty methods.

In our problem the constraints are nonnegativity and summation to a known constant. Our goal is to impose these constraints to the candidate points in the most efficient way possible. For example, the acceptance rejection method does not constitute an efficient way to satisfy the nonnegativity constraint. The reason is that rejections due to generated negative pixel values can introduce significant computational cost for images of some zero or very small pixel values. Instead, we choose to satisfy the nonnegativity constraint automatically during the random perturbation process. Among all mentioned approaches for the perturbation, this is possible if the perturbations are uniformly distributed in a hypercube in the positive quadrant. That is, the range of the hypercube never takes negative values (by taking into account the values of each component of the previous estimate). Gaussian and Cauchy distributed perturbations cannot guarantee to satisfy the nonnegativity constraint because they cannot be bounded. Besides, for spherically uniform distributed perturbations, it is also hard to specify the radius to satisfy nonnegativeness. Moreover, we choose to vary all components at once instead of one at a time, for the same purpose of efficiency. Furthermore, the projection method is used to satisfy the sum constraint. That is, after generating a nonnegative candidate point, the point is rescaled to sum to the given constant.

A rule of thumb to set the maximum amount of perturbation is to enable escaping from local minima in few iterations. Because we estimate the radius of local regions for local minima as 0.01 at most (based on some test images), we set the perturbation radius to the fixed value of 0.002. (Alternatively, one can also decrease the amount of maximum possible perturbation as the iterations proceed, which may yield faster convergence.)

Let us now summarize the steps in our simulated annealing algorithm with Metropolis criterion. At iteration $k$,

- $\hat{f}^k \sim$ Uniformly distributed in a nonnegative neighborhood of $f^{k-1}$
- $\hat{f}^k \leftarrow \hat{f}^k \times f_0/(\sum_x \hat{f}^k(x))$ where $f_0 = \sqrt{\sum_x \tilde{R}(x)}$
- $f^k = \hat{f}^k$ with probability $p^k = \min\left(\exp\left(-\frac{D(\tilde{R},\hat{f}^k * \hat{f}^k) - D(\tilde{R},f^{k-1}*f^{k-1})}{T^{k-1}}\right), 1\right)$
  $f^k = f^{k-1}$ with probability $1 - p^k$

- Decrease $T^k$ from $T^{k-1}$

**2. Cooling schedule:**

A cooling schedule relates the new temperature $T^k$ to the previous temperature $T^{k-1}$, and effectively determines the acceptance probability of the new estimate. In many nonconvex optimization problems, it has been observed that the performance of the annealing-type algorithm is not very sensitive to the form of cooling schedule [31, chap. 8]. However, there also exist problems for which the choice of the schedule makes a significant difference. As we will discuss in this section, the phase retrieval problem is also of this type. Still, once a good schedule is found, it generally works for different instances of the same problem with minor modifications in the parameters.

There are mainly two different approaches for cooling: static cooling and dynamic cooling [101]. In static cooling, the cooling parameters are fixed during the iterations of the algorithm. On the other hand, in dynamic cooling, the parameters are adaptively changed. Although dynamic schedules expand the use of the algorithm to many different instances of the problem, they generally require more development time. In our first analysis, we focus on static schedules.

Common choices for the temperature decay rate $T^k$ in static schedules can be listed as follows [31, 100, 101]:

- Exponential decay: $\alpha^k$ $(\alpha \approx 1)$
- Logarithmic decay: $1/\log(k+1)$
- Faster logarithmic decay: $1/((k+1)^{\alpha/2}\log(k+1))$ $(\alpha \in (0,1))$

For discrete-valued problems, the commonly used schedule is the logarithmic decay. On the other hand, for continuous-valued problems (like phase retrieval problem), the most common choice is the exponential cooling schedule. This cooling is sometimes performed in a stepwise manner where the temperature is decreased only after certain number of iterations (such as every 10 or 100 iterations).

Given the initial and final temperatures, and maximum allowable number of iterations, one can solve for the unknown parameters in the cooling schedule. The initial temperature is often chosen such that the initial acceptance ratio is not less than a desired value. For this, one should first estimate the

maximum change $\Delta D_{\max}$ in the objective function between any two neighboring solutions. Then, a lower bound for the initial acceptance probability is given by $p^1 = \exp(-\Delta D_{\max}/T^1)$. Then, one can set the initial temperature by solving for $T^1$ when $p^1$ is specified:

$$T^1 = -\Delta D_{\max}/\log p^1 \tag{5.27}$$

We consider five different cooling schedules as candidates for the annealing process (see Fig. 5.11). These include exponential schedules with different rates, logarithmic schedule, and some linear combinations of exponential and logarithmic schedules. Because the temperature only appears inside the acceptance probability of the new estimate, viewing the lower bound for the acceptance probability, defined by $\exp(-\Delta D_{\max}/T^k)$, provides a better way of understanding the effect of different cooling schedules. This is illustrated in Fig. 5.12. As seen in this figure, the candidates considered provide a good range of different cooling schedules.

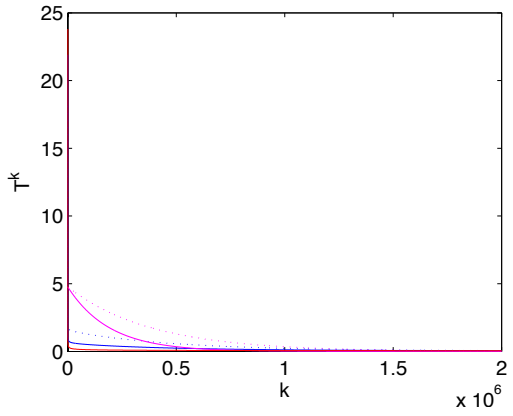



Figure 5.11: The change in temperature as a function of iterations for different cooling schedules: exponential schedules (pink, pink-dotted), logarithmic schedule (red), linear combination of exponential and logarithmic schedules (blue, blue-dotted).

In the numerical experiments, only the rapid exponential cooling schedule, denoted by pink line, yield good reconstructions. Therefore, this schedule is chosen for the problem, and its parameters are slightly optimized through trial and error. The final values of the parameters used in the experiments

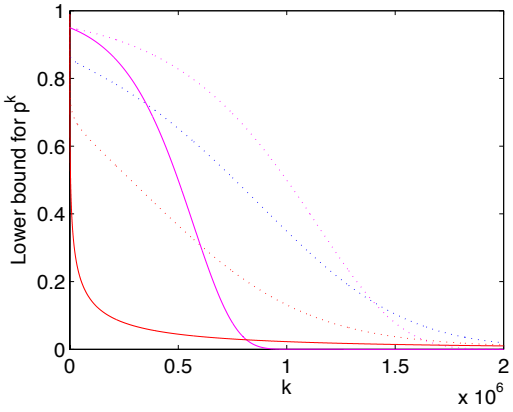



Figure 5.12: The change in the lower bound of acceptance probability as a function of iterations for different cooling schedules: exponential schedules (pink, pink-dotted), logarithmic schedule (red), linear combination of exponential and logarithmic schedules (red-dotted, blue-dotted).

are listed below.

- Lower bound for initial acceptance probability: 0.95
- Initial and final temperatures: 130 and $4 \times 10^{-3}$
- The minimum number of iterations: $2 \times 10^6$
- Stopping rule: 100 successively rejected estimates

**3. Numerical results:**

The developed simulated annealing algorithm is applied to the noise-free autocorrelation of the image shown in Fig. 5.13a. The output of the algorithm after $2.1 \times 10^6$ iterations is given in Fig. 5.13b. Fig. 5.13c shows the Csiszar's distance between true and estimated autocorrelations, which is the objective function to be minimized, as a function of iterations. A similar plot for the Csiszar's distance between true and estimated images is given in Fig. 5.13d.

This image estimate corresponds to an approximate global solution. One way to argue this is through the observation that similar results are obtained with many other random initializations. Moreover, in none of these experiments is a constant increase in the distance between estimated and true images observed (which was before interpreted as an indicator for convergence to nonglobal solutions). However, although the SA algorithm yields

approximate global solutions in this example, the computation time is almost 15 times that of the Schulz-Snyder algorithm when it is converging to the global solution. We also note that the total acceptance ratio in this example is 72%.

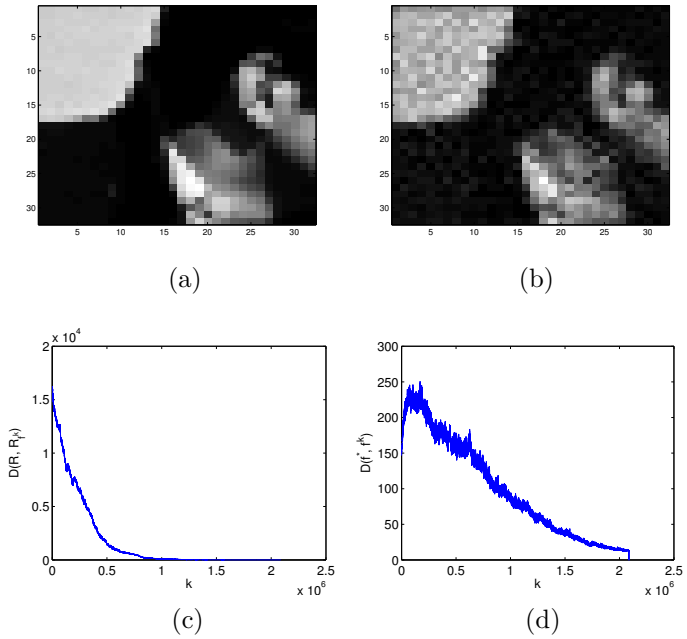



Figure 5.13: (a) Original image. (b) Image estimate obtained with SA. (c) Csiszar's distance between true and estimated autocorrelations. (d) Csiszar's distance between true and estimated images.

Because the output of SA is often a rough estimate of the original image, there are a number of different ways to improve its output: running SA with a decaying perturbation amount, running SS with the output of the SA (to refine the SA estimates), or averaging SA estimates from different initializations. Fig. 5.14 shows the resulting estimates from SA algorithm with cooled perturbation (where exponential schedule is chosen for updating the maximum perturbation amount), and from SS algorithm applied to the output of the SA algorithm. With these additions, the computation time approaches to almost 20 times of the Schulz-Snyder algorithm. Based on many repeated experiments, we conclude that the second method that exploits the SS algorithm provides a better improvement on the SA output compared to the SA algorithm with cooled perturbation. This also requires less computation. Fig. 5.15a illustrates the application of this method to an image of larger

size in order to improve the estimate of SA. Figures 5.15b and 5.15c show the output of the SA algorithm after $2.3 \times 10^6$ iterations and the output after improving this estimate using the SS algorithm, respectively. Here we again note that the total computation time is at least 20 times more that of pure SS (provided that it globally converges). As a final remark, the total acceptance ratio in this case is 73%.

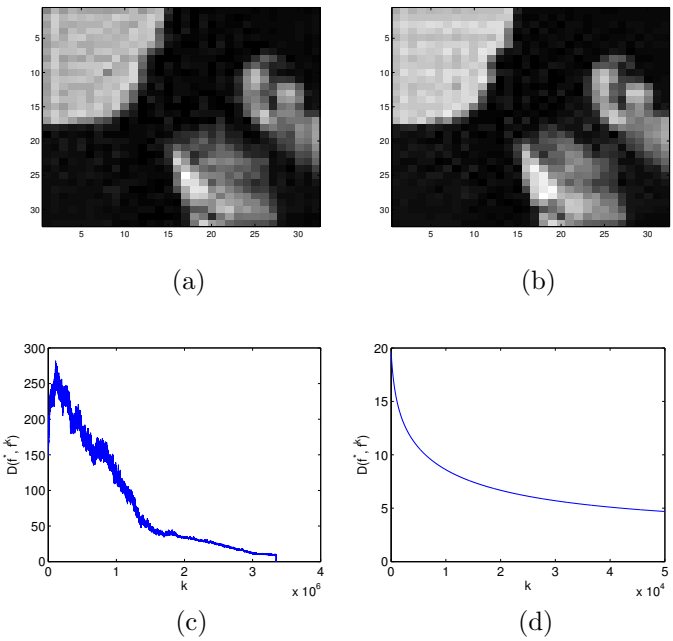
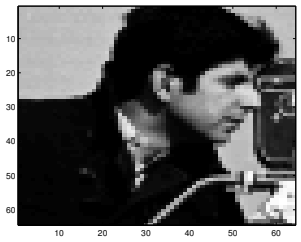


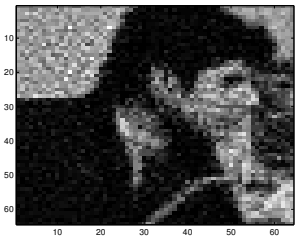

Figure 5.14: Image estimates and corresponding error curves for SA algorithm with cooled perturbation (a and c), and for SS algorithm applied to the output of regular SA algorithm (b and d).

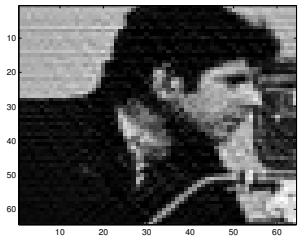
### 5.6.3 Hybrid method based on SS algorithm

Although the SA algorithm developed for the nonconvex problem can yield approximate global solutions, it has high computational cost for images of practical size. In fact, it requires at least 20 times more computation time than a successful reconstruction instance of the SS algorithm. To obtain a global optimization method with less computational cost, we propose a hybrid method to combine the better computational efficiency of Schulz-Snyder algorithm with the global convergence ability of annealing-type algorithms [66].



(a)





(b) (c)

Figure 5.15: (a) Original image. (b) Image estimate obtained with SA. (c) Improved image estimate with SS.

The basic idea of slow cooling for global optimization can also be implemented in other ways than the simulated annealing algorithm. For example, an approach for global optimization from noisy measurements is given in stochastic approximation [31, chap. 8] as

$$f_{k+1} = f_k - a_k G_k(f_k) + b_k w_k \tag{5.28}$$

where $G_k$'s are noisy gradient measurements. (Stochastic approximation methods are a family of stochastic optimization algorithms that attempt to find extrema of functions which cannot be computed directly, but only estimated via noisy observations.) Many authors have proved that these types of iterations achieve global convergence in probability when $a_k$ and $b_k$ have certain forms [31, chap. 8]. The essential idea here is to inject randomness to the gradient-descent type iteration. Note that without the term $b_k w_k$, this is just a gradient-descent type method. To achieve global convergence, a Monte Carlo random term $w_k$ scaled by a decaying coefficient $b_k$ is added to the iteration. The injected random term $b_k w_k$ helps escaping from a local minimum by adding variation to the recursion. This approach is similar to the SA in the sense that the algorithm sometimes accepts a poorer value of

the objective function in the hopes of leaving a local minimum. Besides, the possibility of leaving a minimum is decreasing as the iterations proceed (in order to guarantee convergence in a finite amount of iterations).

As we have discussed in Section 5.5, the Schulz-Snyder algorithm is equivalent to a gradient-descent type approach. That is, the first part of the iteration in (5.28) is achieved by the Schulz-Snyder iteration. To achieve global convergence, we can inject a similar random term $b_k w_k$ to the iteration and let $b_k \to 0$. This will have similar benefits in terms of global convergence as the iteration for stochastic approximation.

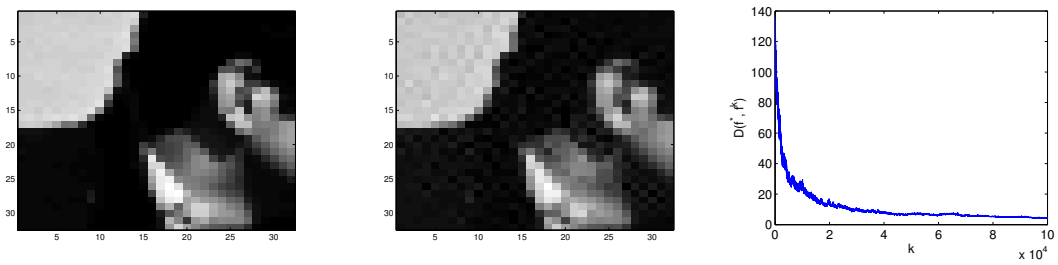
To implement this method, we first need to determine the cooling schedule for $b_k$ and the way to generate the perturbation term $w_k$. As before, the perturbation term is generated from a uniform distribution in the nonnegative neighborhood of the current estimate (in order to preserve nonnegativity). The maximum perturbation amount is again chosen as the maximum possible change at a pixel in the pure SS algorithm. The perturbed point is also rescaled to sum to the desired known constant. For the cooling schedule on $b_k$, curves similar to the ones in Fig. 5.12 are considered for the lower bounds of the acceptance probability. Good reconstructions are observed for a logarithmic-type decay given by $b_k = (1/(k+1))^{(Mk^a)}$ with parameters $M$ and $a$. These parameters are chosen similarly using the initial and final temperatures, and maximum allowable number of iterations.

To summarize the annealing-type SS algorithm, the steps at iteration $k$ are [66]

- $\tilde{f^k}(x) \leftarrow f^{k-1}(x) \frac{1}{\sqrt{\tilde{R}_0}} \left[ \frac{1/2(\tilde{R}(x)+\tilde{R}(-x))}{R_{f^{k-1}}(x)} * f^{k-1}(x) \right]$
- $\hat{f^k} \sim$ Uniformly distributed in the nonnegative neighborhood of $\tilde{f^k}$ defined by a hypercube of sides $b_k$
- $f^k \leftarrow \hat{f^k} \times f_0 / (\sum_x \hat{f^k}(x))$ where $f_0 = \sqrt{\tilde{R}_0}$
- Decrease $b_k$ using $b_k = (\frac{1}{k+1})^{Mk^a}$

The annealing-type SS algorithm is applied to the original image shown in Fig. 5.16a. The output of the algorithm after $10^5$ iterations is given in Fig. 5.16b. Fig. 5.16c shows the Csiszar's distance between the true and estimated images. We observe that image estimates of the regular SS algorithm with 200 different initializations do not have smaller distance to the original

image when compared to the estimate of the annealing-type SS algorithm. Moreover, the required computation time is almost same as the regular SS algorithm.




(a) (b) (c)

Figure 5.16: (a) Original image. (b) Image estimate obtained with annealing-type SS algorithm. (c) Csiszar's distance between true and estimated images.

We also apply the annealing-type SS algorithm with the same parameters to a bigger image shown in Fig. 5.17a. The output of the algorithm after $2 \times 10^5$ iterations is given in Fig. 5.17b. Similarly, Fig. 5.17c shows the Csiszar's distance between true and estimated images. This also illustrates superior reconstruction than the standard SS algorithm, but with almost same amount of computations. We also note that the algorithm gave similar reconstructions with all generated random initializations. Fig. 5.17d shows the average of such ten reconstructions; hence, as shown, averaging the estimates obtained with different initializations also improves the reconstructions. The convergence rates of the annealing-type algorithm and the standard SS algorithm (with uniform initialization) are also shown in Fig. 5.17e. This illustrates that both algorithms have sublinear convergence where the convergence is to the global solution for the annealing-type algorithm whereas the Schulz-Snyder algorithm converges to a nonglobal solution.

To summarize, the annealing-type SS algorithm provides superior reconstructions than the standard SS algorithm, and offers the possibility of global convergence without introducing significant additional computational cost to the SS algorithm. In fact, it only requires slightly more iterations for convergence due to the injected randomness. Note that both the proposed method and HIO are heuristic global optimization methods for the classical phase retrieval problem; hence a comparison between them is legitimate. First of all, both algorithms offer the possibility of obtaining the global solution with

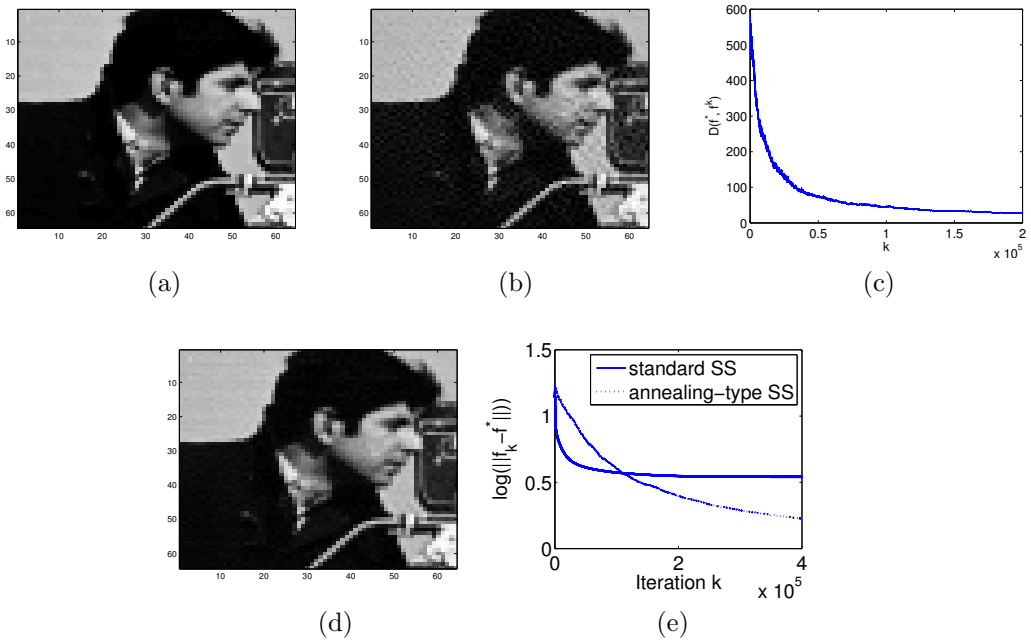



Figure 5.17: (a) Original image. (b) Image estimate obtained with annealing-type SS algorithm. (c) Csiszar's distance between true and estimated images. (d) The average of ten estimates of annealing-type algorithm. (e) Convergence rate.

proper parameter values. But, as a result, they are both sensitive to parameter choices. Adaptive versions of the annealing-type SS algorithm can be explored to overcome this drawback. On the other hand, although these algorithms yield similar reconstructions, HIO is significantly faster (for example, few seconds vs half an hour). This is predicted to arise because HIO modifies the ER algorithm such that it improves not only the reconstructions but also the convergence speed over the ER (through relaxation of space domain constraints). On the other hand, the annealing-type SS algorithm only improves the reconstructions while slightly slowing down the SS. In this approach, the space domain constraints are still imposed at every iteration. Hence we can conclude that, without accelerating them, both the SS algorithm and its variant are not competitive with Fienup's heuristic HIO algorithm, and the classical phase retrieval problem is still lacking a formal method of solution.

## 5.7 Conclusion

In this chapter, we have analyzed and compared important algorithms for the classical phase retrieval problem. In particular, we have derived the Schulz-Snyder phase retrieval algorithm as an alternating minimization method, and discussed its advantages and drawbacks. An annealing-type Schulz-Snyder algorithm, a hybrid method that incorporates annealing-type global optimization methods, has also been developed to avoid convergence to nonglobal solutions.

# CHAPTER 6

# CONCLUSIONS

In this thesis, we developed a class of novel computational spectral imaging techniques that enable capabilities beyond the reach of conventional methods. For each development, we combined novel multiplexed measurements with an image formation model and then computationally solved the resultant inverse problem for image reconstruction. The resulting class of instantaneous spectral imagers can estimate the spectral line parameters with the same order of accuracy as the state-of-the-art slit spectroscopy, but with the added benefit of an instantaneous two-dimensional field-of-view. Similarly, the new generation of spectral imagers with photon sieves offer not only near diffraction-limited spatial resolution, but also several orders of magnitude higher spectral resolution compared to the state-of-the-art filter-based spectral imagers. As a result, these techniques enable finer spectral information in the form of higher temporal, spatial, and spectral resolutions, which will provide improved diagnostic capabilities in applications as diverse as physics, chemistry, biology, medicine, astronomy and remote sensing.

We now provide a brief summary of each chapter. In Chapter 2, we presented a parametric approach to spectral imaging which offers the means for performing spectroscopy over an instantaneous two-dimensional FOV. This technique employs a slitless spectrometer configuration that measures spectrally dispersed images of the two-dimensional scene in different diffraction orders. We estimated the parameters of the spectral lines within the scene by combining these multiplexed measurements with a parametric model and formulating the resultant inverse problem in a MAP estimation framework. We then developed an efficient dynamic programming algorithm to find the global optimum of the resulting nonconvex MAP problem. We investigated the application of the technique in solar spectral imaging. Numerical results suggest that spectral line parameters can be estimated with the same order of accuracy as the conventional slit spectroscopy, but with the added benefit

of an instantaneous two-dimensional FOV. The estimated parameters can be used to infer physical properties of a radiating medium, and the instantaneous capability is particularly useful for studying dynamic phenomena in applications that require spectral analysis around a spectral line, such as in space remote sensing applications.

In Chapter 3, we derived exact and approximate Fresnel imaging formulas that relate the output of a photon sieve imaging system to its input, originating from either a coherent or incoherent extended source. These imaging formulas were then used in the development of a novel computational spectral imaging technique. The technique employs a photon sieve imaging system with a moving detector in order to exploit the wavelength-dependent focusing property of the photon sieve. For the spatially incoherent illumination, we studied the problem of recovering the individual spectral images from the superimposed and blurred measurements of the photon sieve system. We formulated this inverse problem as a MAP problem, which was then solved using a fixed-point algorithm. The effectiveness of the developed technique was illustrated for an application in solar spectral imaging. The results suggest that the technique offers not only near diffraction-limited spatial resolution, but also several orders of magnitude higher spectral resolution compared to filter-based spectral imagers. This provides the possibility of separating nearby spectral components that would not otherwise be possible using wavelength filters, and will be particularly useful in applications that require high-resolution spectral analysis.

By using Bayesian Cramer-Rao lower bound theory in Chapter 4, we developed a general framework for quantitatively characterizing the performance of the computational spectral imaging techniques in terms of their reconstruction accuracy. The derived error bounds were then used to explore the performance limits of the instantaneous spectral imaging technique under various different observing scenarios. Via Monte Carlo simulations, we evaluated the tightness of the bounds and the performance of the developed MAP algorithm for an application in solar spectral imaging. The developed framework allows us not only to characterize the fundamental precision limits, but also to explore the optimal choices of the design considerations.

Phase retrieval problems arise in the photon-sieve spectral imaging setting with coherent illumination; however, this type of illumination has not been the focus of this study yet, and will be a future research direction. These

problems are generalizations of the classical phase retrieval problem, which was the focus of Chapter 5. Here we analyzed and compared important algorithms for the classical phase retrieval problem. In particular, we derived the Schulz-Snyder phase retrieval algorithm as an alternating minimization method, and discussed its advantages and drawbacks. An annealing-type Schulz-Snyder algorithm, a hybrid method that incorporates annealing-type global optimization methods, was also developed to avoid convergence to nonglobal solutions.

# APPENDIX A

# PROOF OF THEOREM 1

This proof is a generalization of the proof of Theorem 5 in [17] to $r \geq 1$ case, and is presented here for completeness.

Let $\hat{\mathbf{\Theta}}$ denote the estimate obtained with the DP algorithm, and $\{\mathbf{\Theta}^o, \mathbf{f}^o\}$ denote the exact MAP estimate given by solving (2.16), hence

$$\{\mathbf{\Theta}^o, \mathbf{f}^o\} = \arg \min_{\substack{\mathbf{f} \in \Pi^M \\ \mathbf{\Theta} \in \Lambda^M}} ||\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta})\mathbf{f}||_W^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m) \tag{A.1}$$

Our goal is show that $\hat{\mathbf{\Theta}} = \mathbf{\Theta}^o$. Note that this will imply that $\hat{\mathbf{f}} = \mathbf{f}^o$ since $\hat{\mathbf{f}}$ is obtained from $\hat{\mathbf{\Theta}}$ using (2.23).

First, we use induction to prove that

$$\hat{\mathbf{\Theta}}_{[1:k]}(\mathbf{\Theta}^o_{[k+1:k+r]}, \mathbf{f}^o_{[k+1:k+r]}) = \mathbf{\Theta}^o_{[1:k]} \quad \text{for } k = 1, 2, \ldots, M - r \tag{A.2}$$

For this, similar to (2.23), we set

$$\begin{aligned} &\hat{\mathbf{f}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) \\ &\quad = \mathbf{H}^*(\hat{\mathbf{\Theta}}_{[1:k]})\mathbf{H}(\hat{\mathbf{\Theta}}_{[1:k]})]^{-1}\mathbf{H}^*(\hat{\mathbf{\Theta}}_{[1:k]})(\tilde{\mathbf{y}} - \mathbf{\Theta}_{[k+1:k+r]}\mathbf{f}_{[k+1:k+r]}) \end{aligned} \tag{A.3}$$

to the optimal values of the problem (2.20). Then, from the update equation of $\mathbf{\Theta}$ in the algorithm, we have

$$\begin{aligned} &\{\hat{\mathbf{\Theta}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}), \hat{\mathbf{f}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})\} \\ &= \arg \min_{\substack{\mathbf{\Theta}_k \in \Lambda \\ \mathbf{\Theta}_{[1:k-1]} \in \mathbf{\Theta}^*_{[1:k-1]}(\mathbf{\Theta}_{[k:k+r-1]}) \\ \mathbf{f}_{[1:k]} \in \Pi^k}} \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 \\ &\qquad + \sum_{m=1}^{k} \Gamma(\mathbf{\Theta}_m) \end{aligned} \tag{A.4}$$

We first prove the base case of induction for $k = 1$:

$$
\begin{aligned}
&\{\hat{\boldsymbol{\Theta}}_{[1:1]}(\boldsymbol{\Theta}^o_{[2:1+r]}, \mathbf{f}^o_{[2:1+r]}), \hat{\mathbf{f}}_{[1:1]}(\boldsymbol{\Theta}^o_{[2:1+r]}, \mathbf{f}^o_{[2:1+r]})\} \\
&\quad = \arg\min_{\substack{\boldsymbol{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:1]})f_1\|^2 + \Gamma(\boldsymbol{\Theta}_1) \\
&\quad = \arg\min_{\substack{\boldsymbol{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:1]})f_1\|^2 + \Gamma(\boldsymbol{\Theta}_1) + \sum_{m=2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \\
&\qquad\qquad + \|\mathbf{H}(\boldsymbol{\Theta}^o_{[2+r:M]})\mathbf{f}^o_{[2+r:M]}\|^2 \\
&\qquad\qquad - 2\mathrm{Re}\{(\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:1]})f_1)^* \mathbf{H}(\boldsymbol{\Theta}^o_{[2+r:M]})\mathbf{f}^o_{[2+r:M]}\} \\
&\quad = \arg\min_{\substack{\boldsymbol{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[2:M]})\mathbf{f}^o_{[2:M]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:1]})f_1\|^2 + \Gamma(\boldsymbol{\Theta}_1) + \sum_{m=2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \\
&\quad = \{\boldsymbol{\Theta}^o_1, f^o_1\} \qquad\qquad (A.5)
\end{aligned}
$$

The second equality follows from (2.24) which implies that

$$\mathbf{H}(\boldsymbol{\Theta}_{[1:1]})^* \mathbf{H}(\boldsymbol{\Theta}^o_{[2+r:M]}) = 0$$

and from the fact that adding terms that are independent of $\boldsymbol{\Theta}_1$ and $f_1$ do not affect the minimization. The fourth equality follows from the optimality of $\{\boldsymbol{\Theta}^o, \mathbf{f}^o\}$.

Now, for the inductive step, suppose that

$$\hat{\boldsymbol{\Theta}}_{[1:i]}(\boldsymbol{\Theta}^o_{[i+1:i+r]}, \mathbf{f}^o_{[i+1:i+r]}) = \boldsymbol{\Theta}^o_{[1:i]} \quad \text{for some } i \in \{1, 2, \ldots, M - r\} \qquad (A.6)$$

We want to prove that the statement holds for $i + 1$. Similar to the base

step, we have

$$
\begin{aligned}
&\{\hat{\boldsymbol{\Theta}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]}), \hat{\mathbf{f}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]})\} \\
&\quad = \arg \min_{\substack{\boldsymbol{\Theta}_{i+1}\in\Lambda \\ \boldsymbol{\Theta}_{[1:i]}\in\boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]}) \\ \mathbf{f}_{[1:i+1]}\in\Pi^{i+1}}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]})\mathbf{f}^o_{[i+2:i+r+1]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})\mathbf{f}_{[1:i+1]}\|^2 \\
&\qquad\qquad + \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m) + \|\mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]})\mathbf{f}^o_{[i+r+2:M]}\|^2 \\
&\qquad\qquad -2\mathrm{Re}\{(\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]})\mathbf{f}^o_{[i+2:i+r+1]} \\
&\qquad\qquad\qquad -\mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})\mathbf{f}_{[1:i+1]})^*\mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]})\mathbf{f}^o_{[i+r+2:M]}\} \\
&\qquad\qquad + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \\
&\quad = \arg \min_{\substack{\boldsymbol{\Theta}_{i+1}\in\Lambda \\ \boldsymbol{\Theta}_{[1:i]}\in\boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]}) \\ \mathbf{f}_{[1:i+1]}\in\Pi^{i+1}}} \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:M]})\mathbf{f}^o_{[i+2:M]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})\mathbf{f}_{[1:i+1]}\|^2 \\
&\qquad\qquad + \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m) + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m)
\end{aligned} \tag{A.7}
$$

where the first equality follows from (A.4), and from (2.24) which implies that $\mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})^*\mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]}) = 0$. Also, by the optimality of $\{\boldsymbol{\Theta}^o, \mathbf{f}^o\}$,

$$
\begin{aligned}
\{\boldsymbol{\Theta}^o_{[1:i+1]}, \mathbf{f}^o_{[1:i+1]}\} = \arg \min_{\substack{\boldsymbol{\Theta}_{[1:i+1]}\in\Lambda^{i+1} \\ \mathbf{f}_{[1:i+1]}\in\Pi^{i+1}}} &\|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:M]})\mathbf{f}^o_{[i+2:M]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})\mathbf{f}_{[1:i+1]}\|^2 \\
&+ \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m) + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m)
\end{aligned} \tag{A.8}
$$

If we now compare (A.7) and (A.8), the objective functions are the same, while the constraint sets are different with the latter containing the former. But because the minimizer $\boldsymbol{\Theta}^o_{[1:i+1]}$ of (A.8) is yet an element of the smaller constraint set in (A.7), it must also be the minimizer of (A.7). (This follows because $\boldsymbol{\Theta}^o_{[1:i]} = \hat{\boldsymbol{\Theta}}_{[1:i]}(\boldsymbol{\Theta}^o_{[i+1:i+r]}, \mathbf{f}^o_{[i+1:i+r]})$ by the assumption (A.6), and hence $\boldsymbol{\Theta}^o_{[1:i]} \in \boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]})$.) Therefore,

$$
\{\hat{\boldsymbol{\Theta}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]}), \hat{\mathbf{f}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]})\} = \{\boldsymbol{\Theta}^o_{[1:i+1]}, \mathbf{f}^o_{[1:i+1]}\} \tag{A.9}
$$

This completes the induction, hence the statement in (A.2) is proved. With $k = M - r$, this gives that

$$\hat{\mathbf{\Theta}}_{[1:M-r]}(\mathbf{\Theta}^o_{[M-r+1:M]}, \mathbf{f}^o_{[M-r+1:M]}) = \mathbf{\Theta}^o_{[1:M-r]} \tag{A.10}$$

as a result, $\mathbf{\Theta}^o_{[1:M-r]} \in \mathbf{\Theta}^*_{[1:M-r]}(\mathbf{\Theta}_{[M-r+1:M]})$. Using this in (2.22) together with the similar arguments used above, it is easily seen that $\hat{\mathbf{\Theta}} = \mathbf{\Theta}^o$.

# APPENDIX B

# DERIVATION OF THE FISHER INFORMATION MATRIX

The Fisher information matrix under our image formation model can be regarded as a specialized version of the Fisher information matrices obtained earlier in the context of parameter estimation of superimposed signals [37] and localization of EEG and MEG sources [59]. Its derivation is presented here for completeness.

The Fisher information matrix can be obtained by first writing the log-likelihood function, which is given by

$$\ln p(\tilde{\mathbf{y}}|\mathbf{\Psi}) \propto -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(\tilde{y}_n - y_n(\mathbf{\Psi}))^2 \tag{B.1}$$

where the constant term that is independent of the unknown parameters is omitted. By substituting this in Equation (4.7), the elements of the Fisher information matrix can be obtained as

$$[\boldsymbol{J}_F(\mathbf{\Psi})]_{ij} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\frac{\partial y_n(\mathbf{\Psi})}{\partial \psi_i}\frac{\partial y_n(\mathbf{\Psi})}{\partial \psi_j} = \frac{1}{\sigma^2}\left[\left(\frac{\partial \mathbf{y}}{\partial \mathbf{\Psi}}\right)^{\top}\frac{\partial \mathbf{y}}{\partial \mathbf{\Psi}}\right]_{ij} \tag{B.2}$$

where $\frac{\partial \mathbf{y}}{\partial \mathbf{\Psi}}$ is $N \times M(P+1)$ Jacobian matrix of $\mathbf{y}$, that is $[\frac{\partial \mathbf{y}}{\partial \mathbf{\Psi}}]_{ni} = \frac{\partial y_n(\mathbf{\Psi})}{\partial \psi_i}$.

Because the parameter $\mathbf{\Psi}$ is composed of two parts as $\mathbf{\Psi} = [\mathbf{f}^{\top}\ \mathbf{\Theta}^{\top}]^{\top}$, the Jacobian matrix is correspondingly decomposed into two parts:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{\Psi}} = \left[\frac{\partial \mathbf{y}}{\partial \mathbf{f}}\ \frac{\partial \mathbf{y}}{\partial \mathbf{\Theta}}\right] \tag{B.3}$$

The columns of the Jacobian matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{f}}$ can be computed as

$$\frac{\partial \mathbf{y}}{\partial f_i} = \mathbf{h}_i(\mathbf{\Theta}_i) \tag{B.4}$$

hence

$$\frac{\partial \mathbf{y}}{\partial \mathbf{f}} = \mathbf{H}(\boldsymbol{\Theta}) \tag{B.5}$$

On the other hand, the Jacobian matrix $\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}}$ is given by

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}} = \left[ \frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}_1} \; \frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}_2} \cdots \frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}_M} \right] \tag{B.6}$$

where

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}_i} = \left[ \frac{\partial \mathbf{y}}{\partial \theta_{i1}} \; \frac{\partial \mathbf{y}}{\partial \theta_{i2}} \cdots \frac{\partial \mathbf{y}}{\partial \theta_{iP}} \right] \tag{B.7}$$

The columns of this matrix can be computed as

$$\frac{\partial \mathbf{y}}{\partial \theta_{ip}} = f_i \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{ip}} \tag{B.8}$$

hence

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}_i} = f_i \left[ \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i1}} \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i2}} \cdots \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{iP}} \right] \tag{B.9}$$

To obtain a closed-form expression, we define

$$\boldsymbol{D}_i(\boldsymbol{\Theta}_i) = \left[ \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i1}} \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{i2}} \cdots \frac{\partial \mathbf{h}_i(\boldsymbol{\Theta}_i)}{\partial \theta_{iP}} \right] \tag{B.10}$$

$$\boldsymbol{D}(\boldsymbol{\Theta}) = [\boldsymbol{D}_1(\boldsymbol{\Theta}_1) \; \boldsymbol{D}_2(\boldsymbol{\Theta}_2) \ldots \boldsymbol{D}_M(\boldsymbol{\Theta}_M)] \tag{B.11}$$

$$\boldsymbol{G}(\mathbf{f}) = \mathrm{diag}(\mathbf{f}) \otimes \boldsymbol{I}_{\mathrm{P}\times\mathrm{P}} \tag{B.12}$$

where $\mathrm{diag}(\mathbf{f})$ is a diagonal matrix with the elements of $\mathbf{f}$ on the diagonal, $\otimes$ is the Kronecker product, and $\boldsymbol{I}_{P\times P}$ is $P \times P$ identity matrix. Then by substituting (B.9) in Equation (B.6), we obtain

$$\begin{aligned}
\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Theta}} &= [f_1 \boldsymbol{D}_1(\boldsymbol{\Theta}_1) \;\; f_2 \boldsymbol{D}_2(\boldsymbol{\Theta}_2) \;\; \ldots \;\; f_M \boldsymbol{D}_M(\boldsymbol{\Theta}_M)] \\
&= [\boldsymbol{D}_1(\boldsymbol{\Theta}_1) \; \boldsymbol{D}_2(\boldsymbol{\Theta}_2) \ldots \boldsymbol{D}_M(\boldsymbol{\Theta}_M)] \begin{bmatrix} f_1 \boldsymbol{I}_{P\times P} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & f_2 \boldsymbol{I}_{P\times P} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & f_M \boldsymbol{I}_{P\times P} \end{bmatrix} \\
&= \boldsymbol{D}(\boldsymbol{\Theta}) \boldsymbol{G}(\mathbf{f})
\end{aligned} \tag{B.13}$$

Finally by substituting this and (B.5) in Equation (B.3), the Jacobian matrix

of $\mathbf{y}$ is obtained as

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\Psi}} = [\mathbf{H}(\boldsymbol{\Theta}) \quad \boldsymbol{D}(\boldsymbol{\Theta})\boldsymbol{G}(\mathbf{f})] \tag{B.14}$$

Then from Equation (B.2), the Fisher information matrix has the following closed-form:

$$\boldsymbol{J}_F(\boldsymbol{\Psi}) = \begin{bmatrix} \boldsymbol{J}_{\mathbf{f},\mathbf{f}} & \boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}} \\ \boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}}^{\top} & \boldsymbol{J}_{\boldsymbol{\Theta},\boldsymbol{\Theta}} \end{bmatrix} \tag{B.15}$$

where

$$\boldsymbol{J}_{\mathbf{f},\mathbf{f}} = \frac{1}{\sigma^2}\, \mathbf{H}^{\top}(\boldsymbol{\Theta})\mathbf{H}(\boldsymbol{\Theta}) \tag{B.16}$$

$$\boldsymbol{J}_{\mathbf{f},\boldsymbol{\Theta}} = \frac{1}{\sigma^2}\, \mathbf{H}^{\top}(\boldsymbol{\Theta})\boldsymbol{D}(\boldsymbol{\Theta})\boldsymbol{G}(\mathbf{f}) \tag{B.17}$$

$$\boldsymbol{J}_{\boldsymbol{\Theta},\boldsymbol{\Theta}} = \frac{1}{\sigma^2}\, \boldsymbol{G}^{\top}(\mathbf{f})\boldsymbol{D}^{\top}(\boldsymbol{\Theta})\boldsymbol{D}(\boldsymbol{\Theta})\boldsymbol{G}(\mathbf{f}) \tag{B.18}$$

# APPENDIX C

# A NECESSARY CONDITION FOR CONVEXITY IN ALTERNATING-MINIMIZATION

Let $J(f)$ be an objective function that is to be minimized by using an alternating-minimization technique. Suppose a lifted objective function, $L(f,h)$, of the following form is available:

$$J(f) = \min_h L(f,h) \tag{C.1}$$

Then convexity of the original objective function, $J(f)$, is a necessary condition for the convexity of the lifted objective function, $L(f,h)$ (Joseph A. O'Sullivan, personal communication, October 2010). The proof of this statement, due to J. A. O'Sullivan, is presented here for completeness.

To prove this, suppose $J$ is not convex in $f$, then

$$J(\lambda f_1 + (1-\lambda)f_2) > \lambda J(f_1) + (1-\lambda)J(f_2), \tag{C.2}$$

for some $f_1$ and $f_2$. Suppose $h_i = \arg\min_h L(f_i,h)$, that is $J(f_i) = L(f_i,h_i)$ for $i = \{1,2\}$. Then we have

$$\begin{aligned} & L(\lambda f_1 + (1-\lambda)f_2, \lambda h_1 + (1-\lambda)h_2) \\ & \quad \geq \min_h L(\lambda f_1 + (1-\lambda)f_2, h), \\ & \quad = J(\lambda f_1 + (1-\lambda)f_2), \\ & \quad > \lambda J(f_1) + (1-\lambda)J(f_2), \\ & \quad = \lambda L(f_1,h_1) + (1-\lambda)L(f_2,h_2), \end{aligned} \tag{C.3}$$

which shows that $L(f,h)$ is not convex in pair $(f,h)$.

# REFERENCES


[1] G. G. Shepherd, *Spectral Imaging of the Atmosphere*. Academic Press, 2002, vol. 82.

[2] P. Massey, and M. M. Hanson, "Astronomical Spectroscopy," in *Planets, Stars and Stellar Systems*, T. D. Oswalt and H. E. Bond, Eds. Springer Netherlands, 2013, pp. 35–98.

[3] G. A. Shaw and H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, pp. 3–28, 2003.

[4] J. M. Lerner, "Imaging spectrometer fundamentals for researchers in the biosciences–a tutorial," *Cytometry Part A*, vol. 69, no. 8, pp. 712–734, 2006.

[5] J. W. Brosius, R. J. Thomas, J. M. Davila, and E. Landi, "Analysis of a solar active region extreme-ultraviolet spectrum from SERTS-97," *The Astrophysical Journal*, vol. 543, pp. 1016–1026, 2000.

[6] A. J. Coyner and J. M. Davila, "Determination of non-thermal velocity distributions from SERTS linewidth observations," *The Astrophysical Journal*, vol. 742, pp. 115–122, 2011.

[7] K. J. H. Phillips, U. Feldman, and E. Landi, *Ultraviolet and X-ray Spectroscopy of the Solar Atmosphere*. Cambridge University Press, Cambridge, 2008.

[8] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*. John Wiley & Sons, 2008.

[9] F. S. Oktem, F. Kamalabadi, and J. M. Davila, "Parameter estimation for instantaneous spectral imaging," to appear in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[10] T. Okamoto and I. Yamaguchi, "Simultaneous acquisition of spectral image information," *Optics Letters*, vol. 16, no. 16, pp. 1277–1279, 1991.

[11] M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Applied Optics*, vol. 34, no. 22, pp. 4817–4826, 1995.

[12] C. C. Kankelborg and R. J. Thomas, "Simultaneous imaging and spectroscopy of the solar atmosphere: advantages and challenges of a 3-order slitless spectrograph," in *Proc. SPIE*, vol. 4498, 2001, pp. 16–26.

[13] W. R. Johnson, D. W. Wilson, W. Fink, M. Humayun, and G. Bearman, "Snapshot hyperspectral imaging in ophthalmology," *Journal of Biomedical Optics*, vol. 12, no. 1, pp. 014 036–014 036, 2007.

[14] N. Hagen and E. L. Dereniak, "Analysis of computed tomographic imaging spectrometers. I. Spatial and spectral resolution," *Applied Optics*, vol. 47, no. 28, pp. F85–F95, 2008.

[15] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.

[16] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[17] S. F. Yau and Y. Bresler, "Maximum likelihood parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 804–820, 1993.

[18] S. F. Yau and Y. Bresler, "On the robustness of parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Processing*, vol. 44, no. 11, pp. 2825–2836, 1996.

[19] F. S. Oktem, F. Kamalabadi, and J. M. Davila, "Cramer-Rao bounds and instrument optimization for slitless spectroscopy," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 2169–2173.

[20] J. Ireland, "Precision limits to emission-line profile measuring experiments," *The Astrophysical Journal*, vol. 620, p. 1132, 2005.

[21] J. M. Hollas, *Modern Spectroscopy.* John Wiley & Sons, 2004.

[22] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, 2003.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[24] I. Griva, S. G. Nash, and A. Sofer, *Linear and Nonlinear Optimization.* SIAM, 2009.

[25] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[26] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.

[27] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 3, pp. 378–392, 1989.

[28] H. Kwakernaak, "Estimation of pulse heights and arrival times," *Automatica*, vol. 16, no. 4, pp. 367–377, 1980.

[29] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Operations Research*, vol. 14, no. 4, pp. 699–719, 1966.

[30] L. Liberti and N. Maculan, *Global Optimization: From Theory to Implementation*. Springer, 2006, vol. 84.

[31] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, 2005, vol. 65.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[33] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[34] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 1995, vol. 1, no. 2.

[35] M. Sniedovich, *Dynamic Programming: Foundations and Principles*. CRC Press, 2010.

[36] W. Rytter, "On efficient parallel computations for some dynamic programming problems," *Theoretical Computer Science*, vol. 59, no. 3, pp. 297–307, 1988.

[37] S. Yau and Y. Bresler, "A compact Cramér-Rao bound expression for parametric estimation of superimposed signals," *IEEE Trans. Signal Processing*, vol. 40, no. 5, pp. 1226–1230, 1992.

[38] P. R. Young, G. Del Zanna, H. E. Mason et al., "EUV emission lines and diagnostics observed with Hinode/EIS," *Publ. Astron. Soc. Japan*, vol. 59, pp. S857–S864, 2007.

[39] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*. John Wiley & Sons, 2011.

[40] F. S. Oktem, J. M. Davila, and F. Kamalabadi, "Image formation model for photon sieves," in *IEEE Int. Conf. on Image Processing (ICIP)*, 2013, pp. 2373–2377.

[41] F. S. Oktem, F. Kamalabadi, and J. M. Davila, "High-resolution computational spectral imaging with photon sieves," to appear in *IEEE Int. Conf. on Image Processing (ICIP)*, 2014.

[42] L. Kipp, M. Skibowski, R. Johnson, R. Berndt, R. Adelung, S. Harm, and R. Seemann, "Sharper images by focusing soft x-rays with photon sieves," *Nature*, vol. 414, no. 6860, pp. 184–188, 2001.

[43] R. Menon, D. Gil, G. Barbastathis, and H. I. Smith, "Photon-sieve lithography," *J. Opt. Soc. Am. A*, vol. 22, no. 2, pp. 342–345, 2005.

[44] G. Andersen, "Large optical photon sieve," *Optics letters*, vol. 30, no. 22, pp. 2976–2978, 2005.

[45] G. Andersen and D. Tullson, "Broadband antihole photon sieve telescope," *Applied Optics*, vol. 46, no. 18, pp. 3706–3708, 2007.

[46] C. Zhou, X. Dong, L. Shi, C. Wang, and C. Du, "Experimental study of a multiwavelength photon sieve designed by random-area-divided approach," *Applied Optics*, vol. 48, no. 8, pp. 1619–1623, 2009.

[47] P. Gorenstein, J. D. Phillips, and R. D. Reasenberg, "Refractive/diffractive telescope with very high angular resolution for X-ray astronomy," in *Proc. SPIE*, vol. 5900, 2005, pp. 590 018–590 018.

[48] Q. Cao and J. Jahns, "Focusing analysis of the pinhole photon sieve: individual far-field model," *J. Opt. Soc. Am. A*, vol. 19, no. 12, pp. 2387–2393, 2002.

[49] D. Attwood, *Soft X-rays and Extreme Ultraviolet Radiation: Principles and Applications*. Cambridge University Press, Cambridge, 2000.

[50] R. E. Blahut, *Theory of Remote Image Formation*. Cambridge University Press, Cambridge, 2004.

[51] J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. Englewood, Colorado: Roberts, 2005.

[52] J. M. Davila, "High-resolution solar imaging with a photon sieve," in *Proc. SPIE*, vol. 8148, 2011, pp. 81 480O–81 480O.

[53] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.

[54] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 813–824, 1998.

[55] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*. SIAM, 2010, vol. 7.

[56] J. R. Lemen, A. M. Title, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman et al., "The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO)," *Solar Physics*, vol. 275, pp. 17–40, 2012.

[57] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I.* John Wiley & Sons, New York, 1971.

[58] S. Minin and F. Kamalabadi, "Uncertainties in extracted parameters of a gaussian emission line profile with continuum background," *Applied Optics*, vol. 48, no. 36, pp. 6913–6922, 2009.

[59] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for EEG and MEG dipole source localization," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 5, pp. 303–321, 1993.

[60] M. Viberg and A. L. Swindlehurst, "A Bayesian approach to autocalibration for parametric array signal processing," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3495–3507, 1994.

[61] G. Harikumar and Y. Bresler, "Analysis and comparative evaluation of techniques for multichannel blind deconvolution," in *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, 1996, pp. 332–335.

[62] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.

[63] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 338–342, 1988.

[64] R. D. Gill and B. Y. Levit, "Applications of the van Trees inequality: a Bayesian Cramer-Rao bound," *Bernoulli*, pp. 59–79, 1995.

[65] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, pp. 27–29, 1978.

[66] F. S. Oktem and R. E. Blahut, "Schulz-Snyder Phase Retrieval Algorithm as an Alternating Minimization Algorithm," in *Imaging and Applied Optics*, OSA Technical Digest (CD) (Optical Society of America, 2011), paper CMC3.

[67] H. Stark, *Image Recovery: Theory and Application.* Academic, New York, 1987.

[68] Y. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," *Opt. Commun.*, vol. 30, no. 3, pp. 304–308, 1979.

[69] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 2, pp. 140 – 154, 1982.

[70] M. Hayes and J. McClellan, "Reducible polynomials in more than one variable," *Proceedings of the IEEE*, vol. 70, no. 2, pp. 197 – 198, 1982.

[71] J. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[72] T. J. Schulz and D. L. Snyder, "Image recovery from correlations," *J. Opt. Soc. Am. A*, vol. 9, no. 8, pp. 1266–1272, 1992.

[73] R. W. Gerchberg and W. O. Saxton, "A particular algorithm for the determination of phase from image plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[74] S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," *Review of Scientific Instruments*, vol. 78, p. 011301, 2007.

[75] J. Fienup and C. Wackerman, "Phase-retrieval stagnation problems and solutions," *J. Opt. Soc. Am. A*, vol. 3, no. 11, pp. 1897–1907, 1986.

[76] J. Seldin and J. Fienup, "Numerical investigation of the uniqueness of phase retrieval," *J. Opt. Soc. Am. A*, vol. 7, no. 3, pp. 412–427, 1990.

[77] V. Elser, "Phase retrieval by iterated projections," *J. Opt. Soc. Am. A*, vol. 20, no. 1, pp. 40–55, 2003.

[78] S. Marchesini, "Phase retrieval and saddle-point optimization," *J. Opt. Soc. Am. A*, vol. 24, no. 10, pp. 3289–3296, 2007.

[79] H. Takajo, T. Takahashi, and T. Shizuma, "Further study on the convergence property of the hybrid input–output algorithm used for phase retrieval," *J. Opt. Soc. Am. A*, vol. 16, no. 9, pp. 2163–2168, 1999.

[80] I. Csiszar, "Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems," *The Annals of Statistics*, vol. 19, no. 4, pp. 2032–2066, 1991.

[81] S. Kullback, *Information Theory and Statistics*. Wiley, New York, 1959.

[82] Y. Vardi and D. Lee, "From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems," *J. Royal Stat. Society, Series B*, vol. 55, no. 3, pp. 569–612, 1993.

[83] L. Lucy, "Optimum strategies for inverse problems in statistical astronomy," *Astronomy and Astrophysics*, vol. 289, pp. 983–994, 1994.

[84] K. Choi, A. D. Lanterman, and R. Raich, "Convergence of the Schulz-Snyder phase retrieval algorithm to local minima," *J. Opt. Soc. Am. A*, vol. 23, no. 8, pp. 1835–1845, 2006.

[85] T. Isernia, F. Soldovieri, G. Leone, and R. Pierri, "On the local minima in phase reconstruction algorithms," *Radio Science*, vol. 31, no. 6, pp. 1887–1899, 1996.

[86] T. Crimmins, J. Fienup, and B. Thelen, "Improved bounds on object support from autocorrelation support and application to phase retrieval," *J. Opt. Soc. Am. A*, vol. 7, no. 1, pp. 3–13, 1990.

[87] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, pp. 43–64, 1996.

[88] T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach," *J. Opt. Soc. Am. A*, vol. 9, no. 7, pp. 1052–1061, 1992.

[89] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, vol. 1, 1984.

[90] J. A. O'Sullivan, "Alternating minimization algorithms: From Blahut-Arimoto to expectation-maximization," in *Codes, curves, and signals: common threads in communications*, A. Vardy, Ed. Kluwer Academic Publishers, 1998, pp. 173–192.

[91] T. Minka, "Expectation-maximization as lower bound maximization," tutorial available at http://research.microsoft.com/en-us/um/people/minka/papers/em.html, 1998.

[92] Y. Vardi, "Applications of the EM algorithm to linear inverse problems with positivity constraints," in *Image Models (and Their Speech Model Cousins)*, S. E. Levinson and L. Shepp, Eds. Springer Verlag, 1996, pp. 183–198.

[93] F. Tsumuraya, N. Miura, and N. Baba, "Iterative blind deconvolution method using Lucy's algorithm," *Astronomy and Astrophysics*, vol. 282, pp. 699–708, 1994.

[94] D. A. Fish, A. M. Brinicombe, E. R. Pike, and J. G. Walker, "Blind deconvolution by means of the Richardson–Lucy algorithm," *J. Opt. Soc. Am. A*, vol. 12, no. 1, pp. 58–65, 1995.

[95] A. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.

[96] Z. B. Zabinsky, "Random search algorithms," Tech. Rep., Department of Industrial and Systems Engineering, University of Washington, Seattle, WA, 2009.

[97] M. Nieto-Vesperinas and J. Mendez, "Phase retrieval by Monte Carlo methods," *Opt. Commun.*, vol. 59, no. 4, pp. 249–254, 1986.

[98] M. Nieto-Vesperinas, R. Navarro, and F. Fuentes, "Performance of a simulated-annealing algorithm for phase retrieval," *J. Opt. Soc. Am. A*, vol. 5, no. 1, pp. 30–38, 1988.

[99] B. McCallum, "Blind deconvolution by simulated annealing," *Optics Communications*, vol. 75, no. 2, pp. 101–105, 1990.

[100] P. Salamon, P. Sibani, and R. Frost, *Facts, Conjectures, and Improvements for Simulated Annealing*. Society for Industrial Mathematics, 2002.

[101] E. Aarts, J. Korst, and W. Michiels, "Simulated annealing," *Search Methodologies*, pp. 187–210, 2005.