Polarimetric Image Reconstruction Algorithms

by

John R. Valenzuela

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied Physics)
in The University of Michigan
2010

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Professor Stephen C. Rand
Associate Professor Selim Esedoglu
Senior Scientist Brian J. Thelen, Michigan Tech Research Institute
Chief Scientist Richard G. Paxman, General Dynamics – AIS

This dissertation is dedicated to my wife, daughter, and son.

ACKNOWLEDGEMENTS

There are many many people (too many to name) who have helped me along the way to this degree and I have a deep heart-felt thanks to all of them. There are a few individuals however, without whom, I would have stopped my graduate training at the Master's level. Jeff Fessler, Brian Thelen, and Rick Paxman; my gratitude to these three outstanding scientists could not be greater.

My advisor, Jeff, is the model graduate student mentor. Jeff is always genial and encouraging; qualities that are very much appreciated by bewildered graduate students. In addition to his pleasant demeanor Jeff is a scientist of the highest caliber. Working with Jeff has been an unforgettable experience.

My colleagues, Brian Thelen and Rick Paxman, have been more supportive and helpful than can be expressed in words. They have both always been willing to spend time discussing technical aspects of my research as well as providing insights as to which direction to move. I stand indebted to them in more ways than one, I never once provided them with a charge number, check's in the mail guys :-) . I will never forget the help and encouragement I received from them, I only hope that someday I have the opportunity to help someone else like they helped me.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Polarimetric Image Reconstruction Algorithms

by

John R. Valenzuela

Chair: Jeffrey A. Fessler

In the field of imaging polarimetry Stokes parameters are sought and must be inferred from noisy and blurred intensity measurements. Using a penalized-likelihood estimation framework we investigate reconstruction quality when estimating intensity images and then transforming to Stokes parameters (traditional estimator), and when estimating Stokes parameters directly (Stokes estimator). We define our cost function for reconstruction by a weighted least squares data fit term and a regularization penalty. It is shown that under quadratic regularization, the traditional and Stokes estimators can be made equal by appropriate choice of regularization parameters. It is empirically shown that, when using edge preserving regularization, estimating the Stokes parameters directly leads to lower RMS error in reconstruction. Also, the addition of a cross channel regularization term further lowers the RMS error for both methods especially in the case of low SNR.

The technique of phase diversity has been used in traditional incoherent imaging systems to jointly estimate an object and optical system aberrations. We extend the technique of phase diversity to polarimetric imaging systems. Specifically, we describe penalized-likelihood methods for jointly estimating Stokes images and optical system aberrations from measurements that contain phase diversity. Jointly estimating Stokes images and optical system aberrations involves a large parameter space. A closed-form expression for the estimate of the Stokes images in terms of the aberration parameters is derived and used in a formulation that reduces the dimensionality of the search space to the number of aberration parameters only. We compare the performance of the joint estimator under both quadratic and edge-preserving regularization. The joint estimator with edge-preserving regularization yields higher fidelity polarization estimates than with quadratic regularization. Under quadratic regularization, using the reduced-parameter search strategy, accurate aberration estimates can be obtained without recourse to regularization "tuning".

Phase-diverse wavefront sensing is emerging as a viable candidate wavefront sensor for adaptive-optics systems. In a quadratically penalized weighted least squares estimation framework a closed form expression for the object being imaged in terms of the aberrations in the system is available. This expression offers a dramatic reduction of the dimensionality of the estimation problem and thus is of great interest for practical applications. We have derived an expression for an approximate joint covariance matrix for object and aberrations in the phase diversity context. Our expression for the approximate joint covariance is compared with the "known-object" Cramèr-Rao lower bound that is typically used for system parameter optimization. Estimates of the optimal amount of defocus in a phase-diverse wavefront sensor derived from the joint-covariance matrix, the known-object Cramèr-Rao bound, and Monte Carlo simulations are compared for an extended scene and a point object. It is found that our variance approximation, that incorporates the uncertainty of the object, leads to an improvement in predicting the optimal amount of defocus to use in a phase-diverse wavefront sensor.

CHAPTER I

Introduction

Passive optical polarimetric imaging, or simply polarimetry, is an emerging remote sensing technology which complements panchromatic, mulitspectral, and hyperspectral imaging. While spectral signatures carry information about material properties, the polarization state of an optical field across a scene contains information related to surface features of objects within the scene, such as, shape and roughness. Polarimetric imaging systems acquire data that can be used to infer the polarization state of an optical field.

The earth is illuminated by sunlight which is essentially unpolarized, however, light reflected from the earth can have a surprisingly large linear polarization component. The polarization state of reflected light depends strongly on the granularity of the reflecting surface; rough surfaces tend to reflect light in a diffuse fashion leaving the optical field unpolarized. Reflection from smooth surfaces, which is typical of man-made objects, tends to be dominated by specular reflection which can result in a polarizing of the optical field especially if the angle of incidence is near Brewster's angle. Since naturally occurring objects typically have a larger surface granularity than man-made objects, polarimetry offers the potential for improved target detection and identification.

Polarization of an optical field is not directly observable and so must be inferred from intensity measurements. There are several ways of representing the polarization state of an optical field, the most common in the remote sensing context is the Stokes-vector representation. The Stokes vector, $\mathbf{S}$, is defined in terms of optical intensity in the following way: S_0 is the total intensity, S_1 is the difference between the intensity transmitted by a linear polarizer oriented parallel to the x (0° reference) axis and one oriented parallel to the y axis, S_2 is the difference between the intensity transmitted by a linear polarizer oriented at 45° to the x axis and one oriented at 135°, and S_3 is the difference between the intensity transmitted by a right circular polarizer and a left circular polarizer. In the vast majority of remote sensing applications the component S_3 is negligible; for this reason it is typical to work with only the first three components of the Stokes vector.

Polarimeters, like traditional incoherent imaging sensors, have resolution limits that depend on noise and system point-spread function. In remote-sensing applications, degradations in the point-spread function are often due to atmospheric turbulence, residual aberrations in the optical system or misalignment among components in the optical system. As such, polarimetric imaging applications stand to benefit from post-detection processing. Moreover, polarimeters utilizing adaptive optics (AO) would benefit from an optical wavefront sensor that exploits the polarimetric nature of the collected data.

If the polarimeter's point-spread-function is known, image reconstruction techniques can be used to remove noise and blur thereby increasing resolution. However, it is not immediately obvious which space the estimation should be done in. The final representation of the polarization state of the optical field is done in Stokes space; but since $\mathbf{S}_1$ and $\mathbf{S}_2$ are differences of intensity images, their signal-to-ratios are significantly smaller than that of the individual intensity images. The choice is then: estimate in the higher SNR intensity space and then transform to the Stokes space, or estimate directly in the lower SNR Stokes space. If the system point-spread-function is not known, it must be estimated jointly with the Stokes parameters, or in the case of an AO system, there must be a way to estimate the degrading optical wavefront and apply corrections in real time.

When choosing an optical wavefront sensor for an AO system there are numerous candidates all having individual strengths and weaknesses. Phase diversity is an image-based wavefront-sensing technique that allows for the *joint* estimation of object and phase aberrations. The technique of phase diversity has an advantage over other wavefront sensing modalities in that it is explicitly designed to accommodate extended objects. The flexibility to accommodate extended scenes allows the possibility of developing a phase-diverse wavefront sensor for imaging modalities other than traditional panchromatic imaging.

In this dissertation we first explore image reconstruction algorithms for imaging polarimetry. Assuming a monochromatic object and a known system point-spread-function we use a penalized-likelihood estimation framework we investigate reconstruction quality estimating in the space that the data was collected (intensity) and when estimating in the "data-reduced" Stokes space. We consider quadratic, edge-preserving, and inter-channel regularization penalty functions.

In chapter III we extend the technique of phase diversity to polarimetric imaging systems. Specifically, we describe penalized-likelihood methods for jointly estimating (monochromatic) Stokes images and optical system aberrations from measurements that contain phase diversity. Both edge-preserving and quadratic regularization penalty functions are considered. The choice of regularization penalty is considered in light of whether a parameter is nuisance or not.

In chapter IV we consider optimization of system parameters in a phase-diverse wavefront sensor. Most analyses in this area are done using Cramèr-Rao bounds. We consider performance of a specific form of estimator for which wavefront sensing could be implemented in real time. Specifically, we analyze the joint covariance of object (monochromatic) and aberrations for a quadratically regularized weighted least squares estimator.

The main contributions discussed in this work are:

- An analysis of penalized-likelihood image reconstruction techniques for estimating Stokes parameters from polarimetric measurements [95, 97].
- Development of a unified framework for jointly estimating Stokes parameters and aberration parameters from polarimetric measurements that contain phase diversity [96, 98].
- Derivation of an approximate joint covariance of aberration estimates and unknown object in the phase diversity context .

CHAPTER II

Polarimetric Image Reconstruction

2.1 Introduction

In remote sensing and astronomical applications the properties of light that are commonly measured and analyzed include intensity, wavelength, and coherence (interferometry) [93, 7]. In the context of imaging, intensity measurements provide information on scene content while spectral measurements provide additional information that can be used for material classification and target identification. A fourth property of light that is related to imaging is polarization. Polarization varies slowly with wavelength and so tends to be uncorrelated with spectral measurements [93] thereby offering the potential for image enhancements not available with spectral measurements alone. This paper describes methods for estimating polarization information, i.e., Stokes vectors, from polarimetric intensity measurements.

In the context of remote sensing, polarization signatures are used to infer surface features of an object under incoherent illumination [93, 26]. Man made objects tend to have smoother surfaces than natural objects, so the mechanism of reflection is dominated by specular reflection which tends to retain or even enhance any polarization of the source.

The state of polarization of a transverse optical field can be represented in several ways [15]; in this paper we focus on the Stokes vector representation. The Stokes vector is a four component vector, $\mathbf{S} = (S_0, S_1, S_2, S_3)$, whose elements are functions of the optical field. The components of the Stokes vector are defined as follows: S_0 is the total intensity, S_1 is the difference between the intensity transmitted by a linear polarizer oriented parallel to the x (0° reference) axis and one oriented parallel to the y axis, S_2 is the difference between the intensity transmitted by a linear polarizer oriented at 45° to the x axis and one oriented at 135°, and S_3 is the difference between the intensity transmitted by a right circular polarizer and a left circular polarizer. In the vast majority of remote sensing applications the component S_3 is negligible; for this reason it is typical to work with only the first three components of the Stokes vector.

The intensity passed by a linear polarizer, whose transmission axis is oriented at angle θ, may be written in terms of the components of the Stokes vector. The effect of an ideal linear polarizer is to pass that part of the electric field which is along the transmission axis of the polarizer. Let the transmission axis of the polarizer be $\mathbf{p}_\theta = \cos(\theta)\hat{\mathbf{i}} + \sin(\theta)\hat{\mathbf{j}}$ and the electric field be

$$\mathbf{E}\left(t\right) = E_x\left(t\right)\hat{\mathbf{i}} + E_y\left(t\right)\hat{\mathbf{j}}. \tag{2.1}$$

The intensity, $\Gamma(\theta)$, passed by the polarizer is then

$$\begin{aligned} \Gamma\left(\theta\right) &= \langle\left(\mathbf{E}\left(t\right)\cdot\mathbf{p}_\theta\right)^2\rangle \\ &= \langle E_x^2\left(t\right)\rangle\cos^2\left(\theta\right) + \langle E_y^2\left(t\right)\rangle\sin^2\left(\theta\right) \\ &+ 2\langle E_x\left(\theta\right)E_y\left(\theta\right)\rangle\sin\left(\theta\right)\cos\left(\theta\right) \end{aligned} \tag{2.2}$$

where $\langle\cdot\rangle$ indicates time averaging. Using the double angle formulae we rewrite the expression (2.2) for the intensity as

$$\begin{aligned}\Gamma\left(\theta\right) = &\frac{1}{2}\left(\langle E_x^2\left(t\right)\rangle + \langle E_y^2\left(t\right)\rangle\right)\\ &+\frac{1}{2}\cos(2\theta)\left(\langle E_x^2\left(t\right)\rangle - \langle E_y^2\left(t\right)\rangle\right)\\ &+\langle E_x\left(t\right)E_y\left(t\right)\rangle\sin(2\theta).\end{aligned} \tag{2.3}$$

The components of the Stokes vector can be written in terms of the electric field as follows [45]

$$S_0 = \Gamma\left(0^\circ\right) + \Gamma\left(90^\circ\right) \qquad = \langle E_x^2(t)\rangle + \langle E_y^2(t)\rangle \tag{2.4}$$

$$S_1 = \Gamma\left(0^\circ\right) - \Gamma\left(90^\circ\right) \qquad = \langle E_x^2(t)\rangle - \langle E_y^2(t)\rangle \tag{2.5}$$

$$S_2 = \Gamma\left(45^\circ\right) - \Gamma\left(135^\circ\right) \quad = 2\langle E_x(t)E_y(t)\rangle. \tag{2.6}$$

By direct substitution, the intensity is related to the Stokes parameters as follows

$$\Gamma\left(\theta\right) = \frac{1}{2}\left(S_0 + S_1\cos(2\theta) + S_2\sin(2\theta)\right). \tag{2.7}$$

For J measurement angles $\{\theta_1, \ldots, \theta_J\}$ equation (2.7) becomes a system of J equations. In matrix form the system is

$$\mathbf{\Gamma} = \mathbf{T}_{J\times 3}\mathbf{S}. \tag{2.8}$$

where $\mathbf{\Gamma} = ((\Gamma(\theta_1), \ldots, \Gamma(\theta_J))$ and $\mathbf{S} = (S_0, S_1, S_2)$. The conventional estimate of the Stokes vector uses the pseudoinverse, indicated by †, of $\mathbf{T}_{J\times 3}$:

$$\hat{\mathbf{S}} = \mathbf{T}_{J\times 3}^\dagger\mathbf{\Gamma}. \tag{2.9}$$

However, the model (2.7) and the estimate (2.9) have ignored noise, blur, and other degradations. The proposed methods overcome these limitations.

2.2 Image Reconstruction Applied to Stokes Vector Imaging

Statistical image reconstruction techniques are applied to data that has been corrupted by non-ideal system effects, i.e., noise and blur. When applying a reconstruction algorithm to polarimetric imagery we are confronted with the question of which image set to reconstruct: the polarimetric intensity images or the underlying Stokes images. While access to the Stokes images is the ultimate goal we must be concerned with the low signal levels in the S_1 and S_2 images. On the other hand, the intensity images do not have this low signal difficulty and so are good candidates for improvement through image reconstruction. We investigate both approaches theoretically and numerically. Estimation of the intensity images is referred to as the traditional approach and estimation of the Stokes vector is referred to as the proposed approach. We explore weighted least squares estimators both with quadratic roughness penalty and with edge-preserving regularization.

Our proposed method for Stokes vector estimation can be generalized to account for optical imperfections such as retardances [94], but for notational simplicity we assume the polarization properties of the optical components are ideal here.

2.2.1 Traditional Image Restoration Approach

In the traditional approach to image restoration we try to recover the uncorrupted images from the noisy images individually. For the polarimetric imaging problem this translates into first restoring the true intensity images, and then converting those images into Stokes space via the linear transformation (2.9). We treat the general case of J polarimetric channels (images) each having a unique associated polarization angle.

Denote the lexicographically ordered data collected in the jth channel by y_j. The system matrix, that represents physical effects such as optical and detector blur, is denoted A_j, and the noise vector by ε_j. The data vectors $\{y_j\}$ and noise vectors ε_j are each of length n_d. The size of the system matrix of the jth channel is $n_d \times n_p$ where n_p is the number of pixels in an individual true intensity image Γ_j.(In general, $n_d \neq n_p$) The model for the jth channel of the collected data is

$$y_j = A_j \Gamma_j + \varepsilon_j. \tag{2.10}$$

For simplicity, we adopt a zero-mean independent-identically-distributed Gaussian noise model:

$$\varepsilon_j \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right). \tag{2.11}$$

Our goal here is to estimate the set of intensity images $\mathbf{\Gamma} = (\Gamma_1, \ldots, \Gamma_J)$ from the set of measurements $\mathbf{y} = (y_1, \ldots, y_J)$. Penalized-likelihood, or MAP, estimators have been used extensively for such image reconstruction problems [103, 14, 6, 95]. Here, a penalized-likelihood estimator for $\mathbf{\Gamma}$ is given by

$$\hat{\mathbf{\Gamma}} = \underset{\mathbf{\Gamma}}{\operatorname{argmin}} \{-\log p(\mathbf{y}|\mathbf{\Gamma}) + R_\Gamma(\mathbf{\Gamma})\} \tag{2.12}$$

where R_Γ is a regularization function for the intensity images. Since the intensity images have no *a priori* coupling, typically one chooses an $R_\Gamma(\mathbf{\Gamma})$ that separates, i.e.,

$$R_\Gamma(\mathbf{\Gamma}) = \sum_{j=1}^{J} R_j(\Gamma_j), \tag{2.13}$$

the likelihood function also separates, so the minimization problem for $\hat{\mathbf{\Gamma}}$ separates into J individual regularized image restoration operations, i.e.,

$$\hat{\Gamma}_j = \underset{\Gamma_j}{\operatorname{argmin}} \{-\log p(y_j|\Gamma_j) + R_j(\Gamma_j)\}, \tag{2.14}$$

where

$$-\log p(y_j|\Gamma_j) = \frac{1}{2\sigma^2}\|y_j - A_j\Gamma_j\|^2. \tag{2.15}$$

After restoring each intensity image, Γ_j, we then estimate the Stokes vector, $\hat{\mathbf{S}}_\Gamma$, using a generalized version of equation (2.9) that is appropriate for images. The generalization to images is accomplished by writing the transformation matrix so that we have a pixel-by-pixel version of equation (2.9): so

$$\hat{\mathbf{S}}_\Gamma = \check{\mathbf{T}}^\dagger \hat{\mathbf{\Gamma}} = \left(\mathbf{T}^\dagger_{J\times 3} \otimes \mathbf{I}_{n_p}\right)\hat{\mathbf{\Gamma}}. \tag{2.16}$$

where $\mathbf{I}_{n_p}$ is the $n_p \times n_p$ identity matrix and $\otimes$ denotes Kronecker product.

2.2.2 Stokes Estimation Model

We now propose a method for estimating $\mathbf{S}$ directly from $\mathbf{y}$. The method for estimating the Stokes vector (images) directly differs from the traditional estimation model in both the likelihood function and the regularization function.

A penalized-likelihood estimator for the Stokes vector is

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmin}}\{-\log p(\mathbf{y}|\mathbf{S}) + R_S(\mathbf{S})\} \tag{2.17}$$

where R_S is a regularization function for the Stokes images. As in the traditional estimator the regularization function R_S typically separates, i.e.,

$$R_S(\mathbf{S}) = \sum_{j=1}^{3} R_j(S_j). \tag{2.18}$$

However, unlike the traditional case the likelihood function $p(\mathbf{y}|\mathbf{S})$ does not separate and so the minimization problem (2.17) is coupled. On the other hand, (2.17) involves fewer unknown parameters than (2.12) because typically $J > 3$.

2.3 Analytical Estimator Analysis: Quadratically Penalized Weighted Least Squares Estimator

This section compares analytically the traditional and Stokes estimators. The degrading effect of the imaging system is taken to be optical blur and represented by the matrix B, i.e., $A_j = B$ in equation (2.10). This matrix is Toeplitz if the system is shift invariant, but the analysis that follows applies to general blur matrices. We assume here that each channel has the same optical and detector blur; this is a reasonable assumption because aberrations and detector effects are not affected by linear polarizers.

We focus on the case of four polarimetric measurements (taken with linear polarizers) at angles $[0°, 45°, 90°, 135°]$; this sets the size of $\mathbf{T}_{J\times 3}$ in equation (2.8) to be 4×3, denoted by $\mathbf{T}_4$ hereafter. The Stokes to intensity transformation, $\mathbf{T}_4$, and the intensity to Stokes transformation, $\mathbf{T}_4^\dagger$, are given by

$$\mathbf{T}_4 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} \quad \text{and} \quad \mathbf{T}_4^\dagger = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

We note for use below that $\mathbf{T}_4'\mathbf{T}_4 = \mathrm{diag}(1, \frac{1}{2}, \frac{1}{2})$, where $'$ indicates conjugate transpose. In fact the analysis that follows applies to any set of polarization angles for which $\mathbf{T}_{J\times 3}'\mathbf{T}_{J\times 3}$ is a diagonal matrix.

2.3.1 Traditional Estimator

The data model (2.10) becomes

$$\mathbf{y} = (\mathbf{I}_4 \otimes B)\,\boldsymbol{\Gamma} + \varepsilon, \tag{2.19}$$

where $\otimes$ is the Kronecker product. For simplicity we consider a quadratic regularizing penalty function that uses vertical and horizontal neighboring pixels. In one dimension the regularizer, $R_\Gamma(\Gamma)$, is written

$$\begin{aligned} R_\Gamma(\Gamma) &= \frac{1}{2}\beta \sum_{j=1}^{J} \sum_{k=2}^{n_p} (\Gamma_{j_k} - \Gamma_{j_{k-1}})^2 \\ &= \frac{1}{2}\beta \sum_{j=1}^{J} \|C\Gamma_j\|^2 \end{aligned} \tag{2.20}$$

where C is a finite differencing matrix, β is the regularization "tuning" parameter, and $\|\cdot\|$ is the Euclidean-norm. Note that it is reasonable to use the same regularization parameter for the different polarimetric channels because there is a high correlation between them. In two dimensions the summation notation becomes cumbersome and so we use linear algebra notation exclusively. Let C be a finite differencing matrix that takes both vertical and horizontal differences, then the regularization function can be succinctly written, noting that $\mathbf{C} \triangleq \mathbf{I}_4 \otimes C$, as

$$R_\Gamma(\boldsymbol{\Gamma}) = \frac{1}{2}\beta \|\mathbf{C}\boldsymbol{\Gamma}\|^2. \tag{2.21}$$

The estimator (2.12) can then be written

$$\hat{\boldsymbol{\Gamma}} = \underset{\boldsymbol{\Gamma}}{\operatorname{argmin}} \{ \frac{1}{2\sigma^2} \|\mathbf{y} - (\mathbf{I}_4 \otimes B)\,\boldsymbol{\Gamma}\|^2 + \frac{1}{2}\beta \|\mathbf{C}\boldsymbol{\Gamma}\|^2 \}. \tag{2.22}$$

We consider the case of unconstrained optimization and so the minimizer must satisfy

$$\nabla_{\mathbf{\Gamma}} \left(\frac{1}{2\sigma^2} \|\mathbf{y} - (\mathbf{I}_4 \otimes B)\, \mathbf{\Gamma}\|^2 + \frac{1}{2}\beta \|\mathbf{C}\mathbf{\Gamma}\|^2 \right) = \mathbf{0}.$$

Solving the above equation for $\hat{\mathbf{\Gamma}}$ and combining with (2.16) yields

$$\hat{\mathbf{S}}_{\Gamma} = \check{\mathbf{T}}^{\dagger} \left(\mathbf{I}_4 \otimes \left(\left[B'B + \sigma^2 \beta R \right]^{-1} B' \right) \right) \mathbf{y}, \tag{2.23}$$

where $R = C'C$. This expression corresponds to separate deblurring of each polarimetric channel followed by converting the restored images into Stokes images.

2.3.2 Stokes Estimator

Noting that the system effects are identical to the case of the traditional estimator we write the data model for the Stokes estimator as follows:

$$\mathbf{y} = (\mathbf{T}_4 \otimes B)\, \mathbf{S} + \varepsilon. \tag{2.24}$$

Following the same procedure as in the case of the traditional estimator we have the following implicitly defined estimator

$$\begin{aligned} \hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmin}} \{ & \frac{1}{2\sigma^2} \|\mathbf{y} - (\mathbf{T}_4 \otimes B)\, \mathbf{S}\|^2 \\ & + \frac{1}{2} \left(\beta_0 \|CS_0\|^2 + \beta_1 \|CS_1\|^2 + \beta_2 \|CS_2\|^2 \right) \}. \end{aligned} \tag{2.25}$$

In this case the three images being estimated are very different and so justify three independent regularization "tuning" parameters. Minimizing with respect to $\mathbf{S}$ leads to the closed form estimator

$$\hat{\mathbf{S}} = \left[\mathbf{T}_4' \mathbf{T}_4 \otimes B'B + \sigma^2 \beta_3 \otimes R \right]^{-1} (\mathbf{T}_4' \otimes B')\, \mathbf{y} \tag{2.26}$$

where $\beta_3 \overset{\triangle}{=} \mathrm{diag}(\beta_0, \beta_1, \beta_2)$. We now analyze the two approaches (2.23) and (2.26).

2.3.3 Spatial Resolution Analysis of the Stokes Estimator

We begin by calculating the mean value of the proposed estimator (2.26). To aid in the calculation define $P = \mathrm{diag}(1, 2, 2)$, and note that $\mathbf{T}_4'\mathbf{T}_4 = P^{-1} = \mathrm{diag}(1, \frac{1}{2}, \frac{1}{2})$. Then using (2.19):

$$\begin{aligned} E[\hat{\mathbf{S}}|\mathbf{S}] &= \left[\mathbf{T}_4'\mathbf{T}_4 \otimes B'B + \sigma^2\beta_3 \otimes R\right]^{-1} \\ &\times (\mathbf{I}_3 \otimes B')\left(P^{-1} \otimes \mathbf{I}_{n_p}\right)\left(\mathbf{T}_4^{\dagger} \otimes \mathbf{I}_{n_p}\right) E[\mathbf{y}|\mathbf{S}] \\ &= \left[\mathbf{T}_4'\mathbf{T}_4 \otimes B'B + \sigma^2\beta_3 \otimes R\right]^{-1} \\ &\times (\mathbf{I}_3 \otimes B')\left(P^{-1} \otimes \mathbf{I}_{n_p}\right)(\mathbf{I}_3 \otimes B)\,\mathbf{S} \\ &= \underbrace{\left[P^{-1} \otimes B'B + \sigma^2\beta_3 \otimes R\right]^{-1}\left(P^{-1} \otimes B'B\right)}_{\mathbf{L}}\mathbf{S} \\ &= \mathbf{LS}, \end{aligned}$$

where $\mathbf{L}$ is a $[3n_p \times 3n_p]$ matrix that acts somewhat like a Wiener filter [49]. Each term in the above expression is uncoupled since the matrices P^{-1} and β_3 are diagonal. To explore the spatial resolution properties of the estimators, we approximate B by a circulant matrix [104]. Let Q be the orthonormal DFT matrix, then the eigen-decompostions B and C are approximated by $Q\Theta Q'$ and $Q\Omega Q'$ respectively. Then

$$\begin{aligned} E[\hat{\mathbf{S}}|\mathbf{S}] &= \left[P^{-1} \otimes Q\Theta'\Theta Q' + \sigma^2\beta_3 \otimes Q\Omega'\Omega Q'\right]^{-1} \\ &\cdot \left(P^{-1} \otimes Q\Theta'\Theta Q'\right)\mathbf{S} \\ &= Q\{\left[P^{-1} \otimes \Theta'\Theta + \sigma^2\beta_3 \otimes \Omega'\Omega\right]^{-1} \\ &\cdot \left(P^{-1} \otimes \Theta'\Theta\right)\}Q'\mathbf{S}. \end{aligned} \tag{2.27}$$

We see that the expectation of $\hat{\mathbf{S}}$ is approximately a filtered version of the true objects with each image having an independent filter.

$$\text{Filter for } S_0 \Longrightarrow \quad L_k = \frac{|B_k|^2}{|B_k|^2 + \beta_0 \sigma^2 |F_k|^2} \tag{2.28a}$$

$$\text{Filter for } S_1 \Longrightarrow \quad L_k = \frac{\frac{1}{2}|B_k|^2}{\frac{1}{2}|B_k|^2 + \beta_1 \sigma^2 |F_k|^2} \tag{2.28b}$$

$$\text{Filter for } S_2 \Longrightarrow \quad L_k = \frac{\frac{1}{2}|B_k|^2}{\frac{1}{2}|B_k|^2 + \beta_2 \sigma^2 |F_k|^2}, \tag{2.28c}$$

where $\{B_k\}$ and $\{F_k\}$ are the DFT coefficients of the first column of B and C respectively and $k = 1, \ldots, n_p$. We see that the S_1 and S_2 channels have a different spatial resolution than the S_0 channel, unless we choose $\beta_1 = \beta_2 = \beta_0/2$. However, in the intensity model all three channels always have identical resolutions.

Matching the spatial resolution of the S_1 and S_2 channels to the S_0 channel also decouples the estimator. That is, by choosing

$$\beta_3 = \beta \operatorname{diag}\left(1, \frac{1}{2}, \frac{1}{2}\right) \tag{2.29}$$

we have

$$\begin{aligned}
\hat{\mathbf{S}} &= \left[\mathbf{T}_4' \mathbf{T}_4 \otimes B'B + \sigma^2 \beta_3 \otimes R\right]^{-1} \left(\mathbf{T}_4' \otimes B'\right) \mathbf{y} \\
&= \left[\operatorname{diag}(1, \frac{1}{2}, \frac{1}{2}) \otimes \left(B'B + \sigma^2 \beta R\right)\right]^{-1} \left(\mathbf{T}_4' \otimes B'\right) \mathbf{y} \\
&= \operatorname{diag}(1, 2, 2) \otimes \left[B'B + \sigma^2 \beta R\right]^{-1} \left(\mathbf{T}_4' \otimes B'\right) \mathbf{y} \\
&= \mathbf{T}_4^{\dagger} \otimes \left[B'B + \sigma^2 \beta R\right]^{-1} B' \mathbf{y} \\
&= \hat{\mathbf{S}}_{\Gamma}.
\end{aligned}$$

In other words, for the choice (2.29), with quadratic regularization the Stokes estimator becomes uncoupled and reduces to the traditional estimator. Next we consider the covariance of the estimator.

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{S}}|\mathbf{S}) &= \mathrm{Cov}(\mathbf{Ly}|\mathbf{S}) \\ &= \sigma^2 \left[P^{-1} \otimes B'B + \sigma^2 \beta_3 \otimes R\right]^{-1} \\ &\cdot \left(P^{-1} \otimes B'B\right) \left[P^{-1} \otimes B'B + \sigma^2 \beta_3 \otimes R\right]^{-1}\end{aligned}$$

In the circulant approximation we have,

$$\begin{aligned}\mathrm{Cov}(\hat{\mathbf{S}}|\mathbf{S}) &= \sigma^2 \left[P^{-1} \otimes Q\Theta'\Theta Q' + \sigma^2 \beta_3 \otimes Q\Omega'\Omega Q'\right]^{-1} \\ &\cdot \left(P^{-1} \otimes Q\Theta'\Theta Q'\right) \\ &\cdot \left[P^{-1} \otimes Q\Theta'\Theta Q' + \sigma^2 \beta_3 \otimes Q\Omega'\Omega Q'\right]^{-1} \\ &= \sigma^2 Q \left[P^{-1} \otimes \Theta'\Theta + \sigma^2 \beta_3 \Omega'\Omega\right]^{-1} \\ &\cdot \left(P^{-1} \otimes \Theta'\Theta\right) \\ &\cdot \left[P^{-1} \otimes \Theta'\Theta + \sigma^2 \beta_3 \Omega'\Omega\right]^{-1} Q'\end{aligned} \tag{2.30}$$

The variance of the ith pixel in each image is

$$\mathrm{Var}\{S_{0_i}|\mathbf{S}\} = \frac{\sigma^2}{n_p} \sum_k \frac{|B_k|^2}{\left(|B_k|^2 + \beta_0 \sigma^2 |F_k|^2\right)^2} \tag{2.31a}$$

$$\mathrm{Var}\{S_{1_i}|\mathbf{S}\} = \frac{\sigma^2}{n_p} \sum_k \frac{2|B_k|^2}{\left(|B_k|^2 + \beta_1 \sigma^2 |F_k|^2\right)^2} \tag{2.31b}$$

$$\mathrm{Var}\{S_{2_i}|\mathbf{S}\} = \frac{\sigma^2}{n_p} \sum_k \frac{2|B_k|^2}{\left(|B_k|^2 + \beta_2 \sigma^2 |F_k|^2\right)^2} \tag{2.31c}$$

For the choice (2.29) the noise in the reconstructed Stokes images $\{S_1, S_2\}$ is twice that of S_0, indicating that the regularization for the $\{S_1, S_2\}$ images may need to be stronger than S_0. In summary, if one used quadratic regularization with regularization parameters as in (2.29), then $\hat{\mathbf{S}}$ and $\hat{\mathbf{S}}_\Gamma$ would be identical. Next we turn to the case of nonquadratic regularization.

2.4 Empirical Studies

2.4.1 Edge-preserving Regularization

In the Stokes images $\{S_1, S_2\}$ the polarization information typically has sharp edges. To recover as much polarization information as possible the regularization function should preserve edges. Since quadratic regularization tends to wash out edges and smooth noise we explore edge preserving regularization using the hyperbolic function $\psi(t;\delta) = \delta^2 \left(\sqrt{1 + \left(\frac{t}{\delta}\right)^2} - 1 \right)$, where δ is a "shape parameter" that controls edge-preserving properties. For fixed δ this function is approximately quadratic for values of $t < \delta$ and approximately linear for $t > \delta$. This behavior will tend to smooth noise and preserve edges. The Stokes and intensity regularizers are written

$$R_\Gamma(\mathbf{\Gamma}) = \sum_{j=1}^{J} \sum_{k=1}^{2n_p} \psi\left([C\Gamma_j]_k; \delta\right)$$

$$R_S(\mathbf{\Gamma}) = \sum_{\ell=0}^{2} \sum_{k=1}^{2n_p} \psi\left([CS_\ell]_k; \delta_\ell\right)$$

The Stokes estimator now has two regularization parameters per Stokes image and the intensity estimator has two regularization parameters in total. Since closed form expressions for the minimizers of these cost functions are intractable we minimize them numerically. The numerical optimization was done with the LBFGS algorithm[57]. To obtain optimal values of all regularization parameter combinations would be computationally burdensome so we chose the β parameters by analyzing the local point spread function [31] of the quadratic estimators. The β parameters are chosen so that the FWHM of the estimator is a prescribed amount in regions where the regularization function is approximately quadratic. The local impulse response is defined by

$$\begin{aligned}\boldsymbol{l}_j\left(\mathbf{S}\right) &= \lim_{\varepsilon\rightarrow 0}\frac{\hat{\mathbf{S}}\left(\overline{y}(\mathbf{S}+\varepsilon\boldsymbol{e}_j)\right)-\hat{\mathbf{S}}\left(\overline{y}\left(\mathbf{S}\right)\right)}{\varepsilon} \\ &= \nabla\hat{\mathbf{S}}\left(\overline{y}\left(\mathbf{S}\right)\right)\nabla\overline{y}\left(\mathbf{S}\right)\boldsymbol{e}_j.\end{aligned} \tag{2.32}$$

For the quadratically penalized weighted least squares estimator with white Gaussian noise the local impulse response of each Stokes image is independent and written[30]

$$\boldsymbol{l}_{k_j} = [B'B+\sigma^2\beta_k R]^{-1}B'B\boldsymbol{e}_{k_j} \tag{2.33}$$

where $k \in \{0,1,2\}$ indicates the Stokes image. Using a local Fourier approximation we can compute this impulse response with FFTs. Since the FFTs are computationally inexpensive we can sweep over the β parameters and choose the one that corresponds to an *a priori* FWHM of the impulse response. By choosing β so that the estimator has controlled noise smoothing properties we can vary the δ parameters to find optimal values.

2.4.2 Cross-Channel Regularization

Polarimetric signatures usually are correlated in the Stokes parameters S_1 and S_2. To exploit this correlation we can introduce a cross-channel regularization term into the cost function. Cross-channel regularization has proved beneficial in multispectral image restoration algorithms [11, 14]. The cross-channel regularization functions we adopt for the traditional and proposed estimators are

$$R_{\text{cross}}\left(\boldsymbol{\Gamma}\right) = \beta_{\text{cross}}\sqrt{\sum_j [C\Gamma_j]^2} \tag{2.34}$$

$$R_{\text{cross}}\left(\mathbf{S}\right) = \beta_{\text{cross}}\sqrt{[CS_0]^2+[CS_1]^2+[CS_2]^2}. \tag{2.35}$$

The traditional and Stokes estimators, using channel-by-channel and cross-channel regularization, are then

$$\begin{aligned}\hat{\mathbf{\Gamma}} &= \underset{\mathbf{\Gamma}}{\operatorname{argmin}}\{\frac{1}{2\sigma^2}\|\mathbf{y} - (\mathbf{I}_4 \otimes B)\,\mathbf{\Gamma}\|^2 \\ &\quad + \beta \sum_j \psi\,(C\Gamma_j;\delta)\} + \beta_{\text{cross}}\sqrt{\sum_j [C\Gamma_j]^2}\end{aligned} \tag{2.36}$$

$$\begin{aligned}\hat{\mathbf{S}} &= \underset{\mathbf{S}}{\operatorname{argmin}}\{\frac{1}{2\sigma^2}\|\mathbf{y} - (\mathbf{T}_{4\times 3} \otimes B)\,\mathbf{S}\|^2 \\ &\quad + \beta_0\psi\,(CS_0;\delta_0) + \beta_1\psi\,(CS_1;\delta_1) + \beta_2\psi\,(CS_2;\delta_2) \\ &\quad + \beta_{\text{cross}}\sqrt{[CS_0]^2 + [CS_1]^2 + [CS_2]^2}\}.\end{aligned} \tag{2.37}$$

The additional parameters β_{cross} were varied over a range of values to find optimal settings.

2.5 Simulation Experiments

Simulation experiments were performed to evaluate both the traditional and proposed estimators as well as the cross-channel regularization. The simulations were conducted using the circulant approximation, this approximation was facilitated by tapering the object to its mean at the boundaries. For the true imagery we used polarimetric images that were collected by General Dynamics Advanced Information Systems in Ypsilanti, Michigan; the polarization angles of the sensor were $\{0°, 45°, 90°, 135°\}$. We added known optical blur and Gaussian noise to the imagery. The system PSF had a FWHM of 1.9 pixels; the PSF was constructed from a phase screen, parameterized by a uniform distribution of the first 20 Zernike polynomials, placed in an annular pupil. We define the SNR of an image by: SNR $= 20\ \log_{10}(\|\bar{y}\|/\|\bar{y} - y\|)$dB where $\bar{y}$ and y are the uncorrupted and corrupted images respectively. The simulation experiments were done at two SNR levels using the same PSF: 25dB and 15dB.

The regularization parameters $[\beta_0, \beta]$ were chosen so that the FWHM of the estimator's PSF was 1.5 pixels under quadratic regularization. We chose to set the parameters $[\beta_1, \beta_2] = 2\beta_0$ for increased noise suppression in the $\{S_1, S_2\}$ channels. The second set of regularization parameters, $[\delta_0, \delta_1, \delta_2, \delta]$, was determined by sweeping each parameter over a range of values and choosing the parameters which yielded a minimum RMS error in the estimate. The cross-channel regularization was evaluated for the Stokes estimator with channel-by-channel edge-preserving regularization. The optimal δ values were used and the parameter β_{cross} was swept over a range to determine an optimal setting. Once the optimal values of the regularization parameters were determined for both data SNR levels the estimator was evaluated over a 100 realization noise ensemble.

2.6 Results

Tables 2.1 and 2.2 compare the RMS estimation errors of four quantities over the noise ensemble: (1) the S_0 estimate, (2) the S_1 estimate, (3) the S_2 estimate, and (4) the estimate of the degree of linear polarization (DOLP). The DOLP is a useful quantity in polarimetric image analysis and is defined by DOLP $= \sqrt{S_1^2 + S_2^2}/S_0$. The proposed and traditional estimators both perform well in the estimation of S_0 at both SNR levels. However, the proposed estimator outperforms the traditional estimator, especially in the 15dB SNR case, on the S_1 and S_2 images as well as the DOLP. The superior performance of the proposed estimator can be attributed to the joint nature of the estimator in which regularization is applied directly to the Stokes images. The addition of the cross-channel regularization term improves both estimators. The cross-channel term improved the traditional estimator more than the proposed estimator in RMS error because of the higher correlation between the intensity channels than between the Stokes channels. The addition of the cross-channel regularization term brought the estimators into near equivalent performance in the high SNR case. In the low SNR case the cross-channel regularization helped both estimators but the proposed estimator maintained superior performance.

Table 2.1: Simulation Results (SNR = 25dB): RMS Error Percentages

	Stokes	**traditional**	**Stokes with cross-channel**	**traditional with cross-channel**
$\hat{S}_0$	$0.71 \pm 0.0035\%$	$0.91 \pm 0.0040\%$	$1.07 \pm 0.0035\%$	$1.30 \pm 0.0040\%$
$\hat{S}_1$	$42.19 \pm 0.17\%$	$62.45 \pm 0.24\%$	$36.97 \pm 0.16\%$	$38.57 \pm 0.17\%$
$\hat{S}_2$	$45.93 \pm 0.19\%$	$67.08 \pm 0.25\%$	$40.58 \pm 0.17\%$	$42.00 \pm 0.19\%$
DOLP	$27.83 \pm 0.11\%$	$39.90 \pm 0.15\%$	$24.78 \pm 0.10\%$	$24.54 \pm 0.11\%$

Table 2.2: Simulation Results (SNR = 15dB): RMS Error Percentages

	Stokes	**traditional**	**Stokes with cross-channel**	**traditional with cross-channel**
$\hat{S}_0$	$2.49 \pm 0.011\%$	$3.10 \pm 0.015\%$	$2.74 \pm 0.012\%$	$3.44 \pm 0.015\%$
$\hat{S}_1$	$61.90 \pm 0.27\%$	$126.88 \pm 0.57\%$	$58.13 \pm 0.25\%$	$69.18 \pm 0.30\%$
$\hat{S}_2$	$65.35 \pm 0.29\%$	$137.32 \pm 0.60\%$	$61.14 \pm 0.26\%$	$75.00 \pm 0.33\%$
DOLP	$43.95 \pm 0.20\%$	$96.59 \pm 0.31\%$	$42.37 \pm 0.18\%$	$46.68 \pm 0.20\%$

Fig. 2.1 show the noisy and blurred data for both SNR levels, Figs. 2.2 and 2.3 show estimates of the Stokes images for SNR levels of 25dB and 15dB respectively, and Figs. 2.4 and 2.5 show estimates of the DOLP for SNR levels of 25dB and 15dB respectively. Figs. 2.6 and 2.7 show estimates of the DOLP with the addition of cross-channel regularization for both the proposed and traditional estimators for SNR levels of 25dB and 15dB respectively.

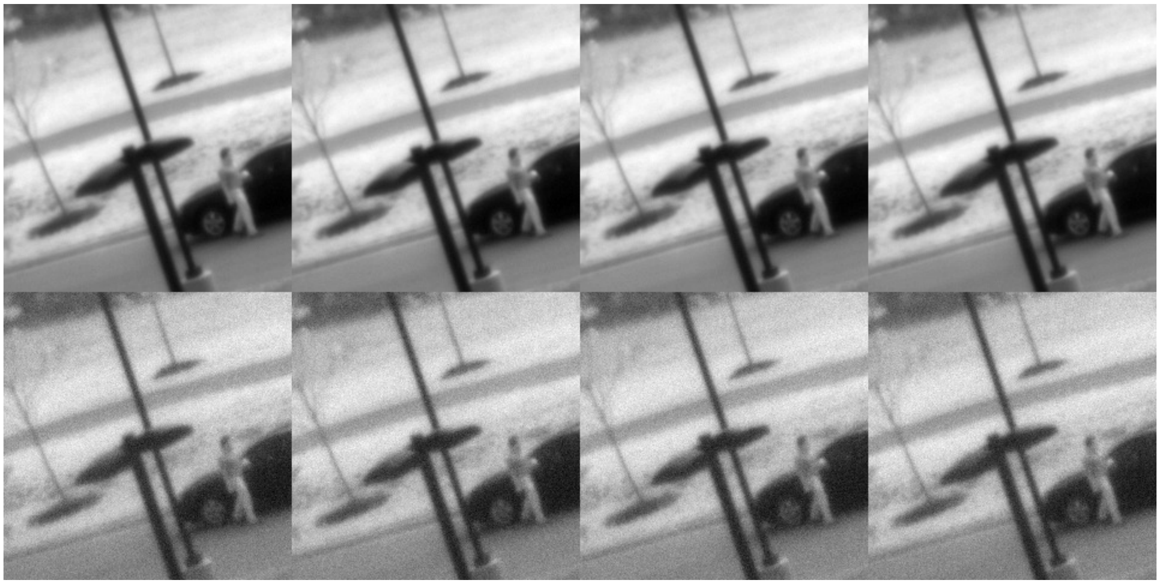


Figure 2.1: Noisy and blurred polarimetric imagery. The First row has an SNR of 25dB and the second row has an SNR of 15dB. From left to right the angle of the polarizer is $\{0°, 45°, 90°, 135°\}$

2.7 Conclusions and Future Work

Estimation of Stokes vectors directly provides estimates with lower overall RMS error as compared with restoring the intensity images and then transforming to Stokes space for interpretation. The addition of a cross-channel regularization term improves interpretability for both the proposed estimator and the tradtitional estimator markedly. In the low (15dB) SNR regime the proposed estimator outperforms the traditional estimator both with and without cross-channel regularization. Future work includes addressing non idealities such aliasing and broadband optical PSF effects. Also, investigating estimator efficiency, convergence properties, and automatic selection of regularization parameters will be done.

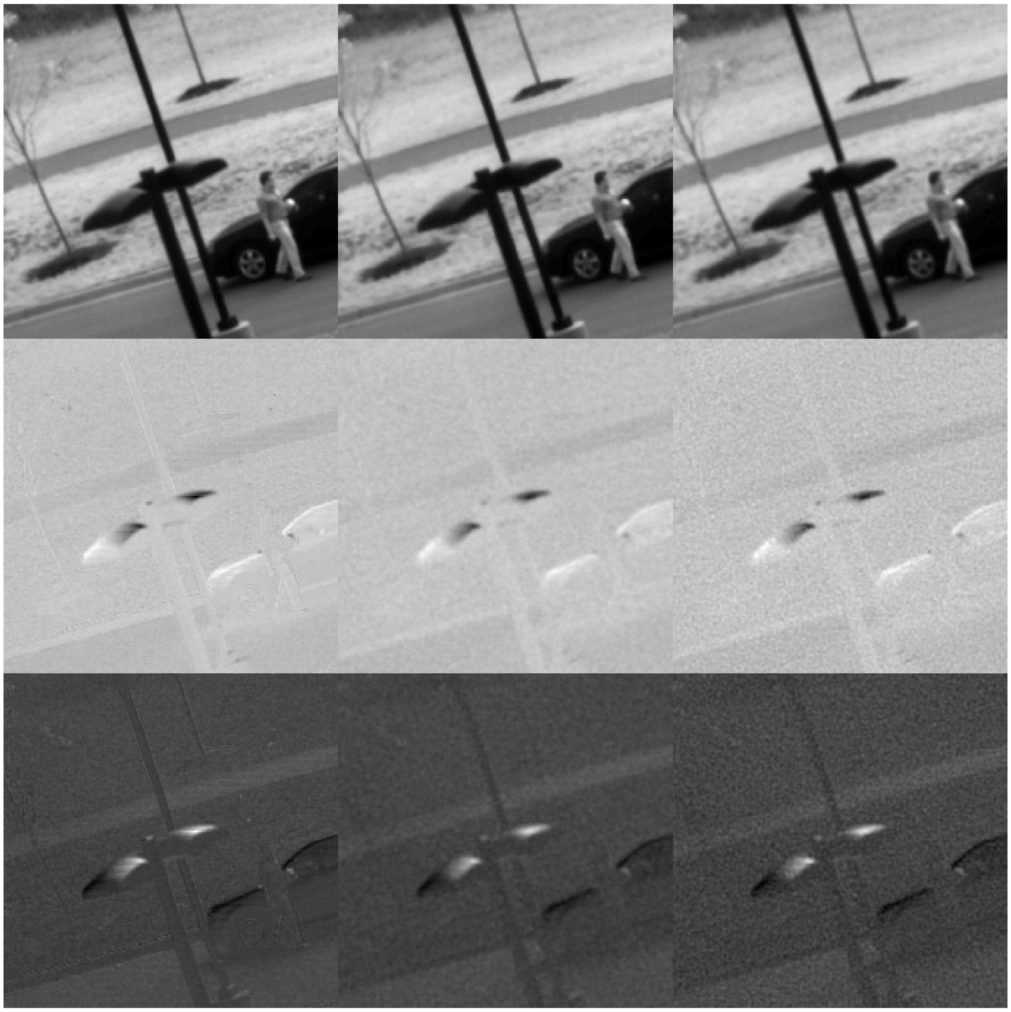


Figure 2.2: Estimates of Stokes images for SNR = 25dB. All rows read from left to right: Pristine, Proposed Method, Traditional Method. First row: S_0, second row: S_1, third row: S_2.

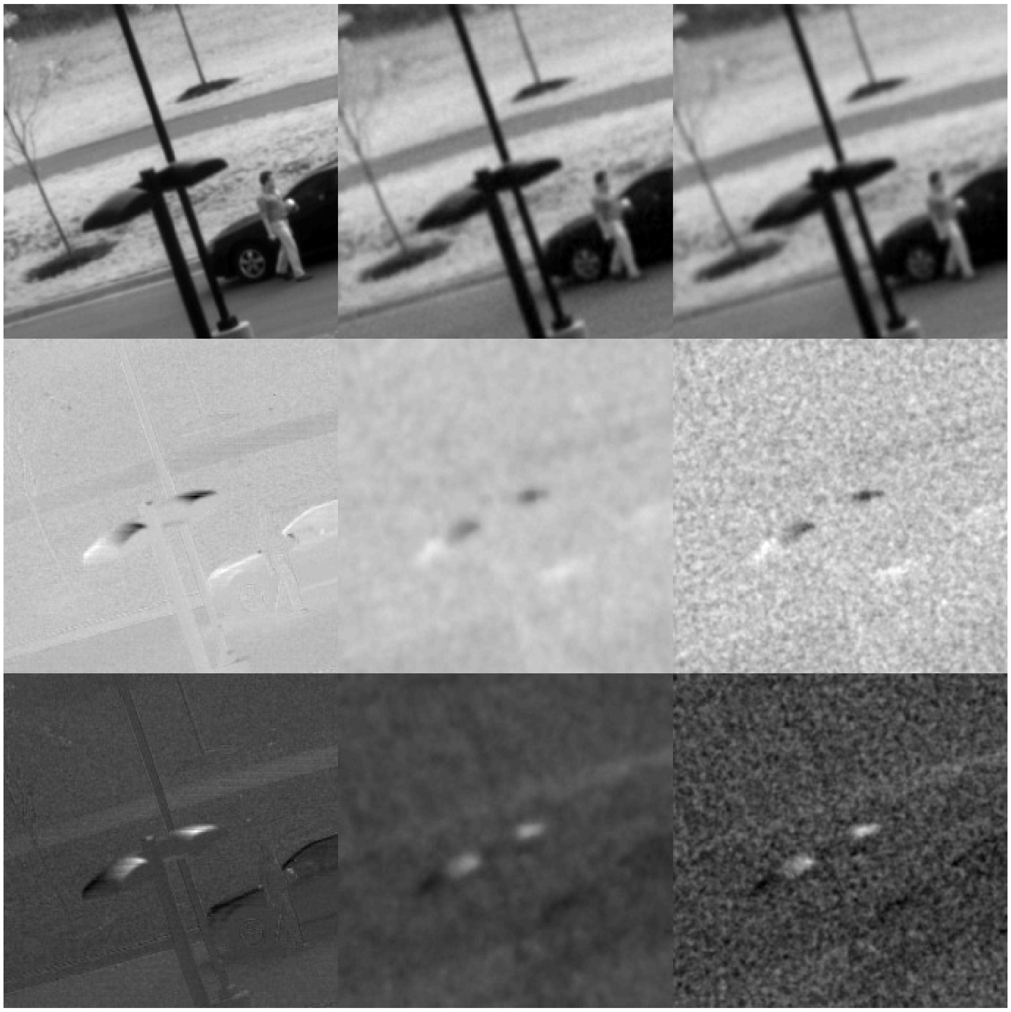


Figure 2.3: Estimates of Stokes images for SNR = 15dB. All rows read from left to right: Pristine, Proposed Method, Traditional Method. First row: S_0, second row: S_1, third row: S_2.

Figure 2.4: Estimates of the DOLP for SNR = 25dB, from left to right: Pristine, Proposed Method, Traditional Method.

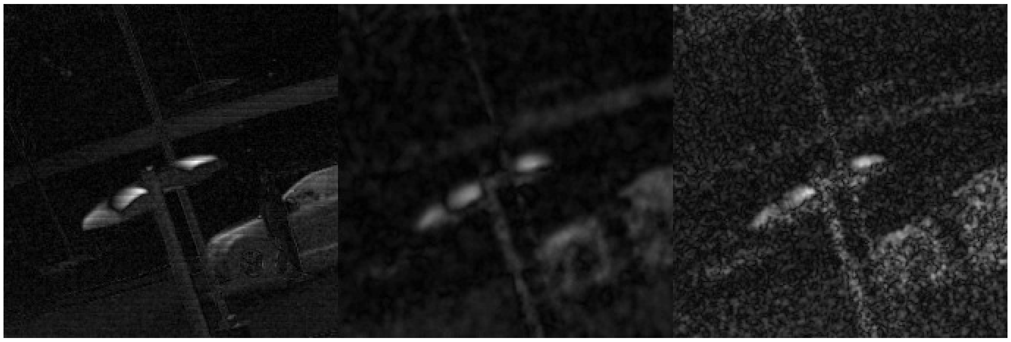


Figure 2.5: Estimates of the DOLP for SNR = 15dB, from left to right: Pristine, Proposed Method, Traditional Method.

Figure 2.6: Estimates of the DOLP for SNR = 25dB, from left to right: Pristine, Proposed method with cross channel regularization, Traditional method with cross channel regularization.

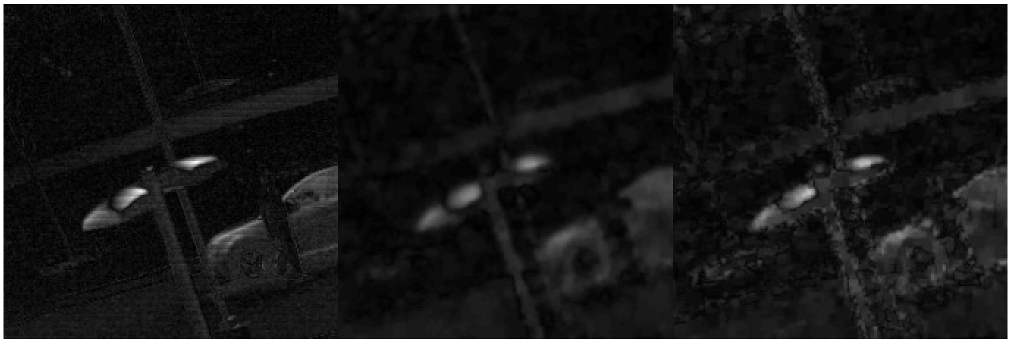


Figure 2.7: Estimates of the DOLP for SNR = 15dB, from left to right: Pristine, Proposed method with cross channel regularization, Traditional method with cross channel regularization.

CHAPTER III

Phase-Diverse Polarimetric Image Reconstruction

3.1 Introduction

Polarimetric imaging systems acquire data that can be used to infer the polarization state of an optical field [93, 7]. The polarization state of an optical field across a scene contains information related to surface features such as shape and roughness [26]. Naturally occurring objects typically have a larger surface granularity than man-made objects, so polarimetry offers the potential for improved target detection and identification over other imaging modalities [37].

The polarization state of a transverse optical field can be specified by the Stokes vector $\mathbf{S} = (S_0, S_1, S_2, S_3)$ [15, 45]. The elements of $\mathbf{S}$ are functions of the optical intensity and defined in the following way: S_0 is the total optical intensity, S_1 is the difference between the optical intensity transmitted by a linear polarizer with pass axis oriented at 0° (reference) and one having pass axis oriented at 90°, S_2 is the difference between the optical intensity transmitted by a linear polarizer with pass axis oriented at 45° and one having pass axis oriented at 135°, and S_3 is the optical intensity transmitted by a right circular polarizer and a left circular polarizer. In the majority of remote-sensing applications the linear polarization state of the optical field is of interest and so the S_3 component is ignored. We adopt this usual simplification of considering only the first three components of the Stokes vector, though the method generalizes easily.

Polarimeters, like traditional incoherent imaging sensors, have resolution limits that depend on noise and system point-spread function. In remote-sensing applications, degradations in the point-spread function are often due to atmospheric turbulence, residual aberrations in the optical system or misalignment among components in the optical system. We previously developed a method for estimating Stokes images directly from polarimetric measurements [97]. That work assumed complete knowledge of the system point-spread function and was thus limited in its range of application. In this paper, we propose methods that overcome this limitation by introducing phase diversity into the polarimetric measurements. In traditional incoherent imaging the technique of phase diversity has been used to jointly estimate the object and optical aberrations in the presence of atmospheric turbulence [69]. Phase diversity requires the simultaneous collection of two or more images that are related via a deterministic phase perturbation. Typically, two images are collected: one is the conventional in-focus image and the second image is acquired on a separate focal plane that is translated along the optical axis thereby inducing a known defocus to the second image. Fig. 3.1 shows a typical phase diversity configuration. A direct extension of the traditional phase diversity strategy to polarimetry would be to acquire

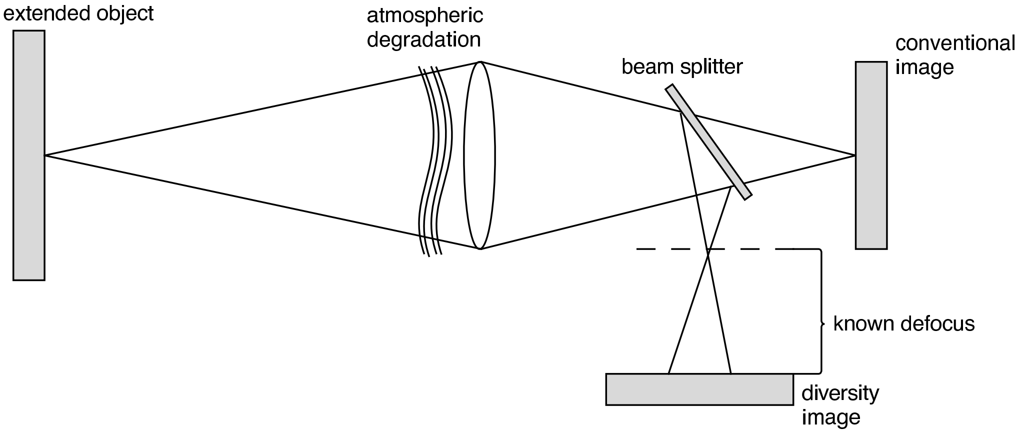


Figure 3.1: Traditional phase-diversity imaging strategy.

two measurements per polarimetric channel; a four-channel polarimeter would be extended to an eight-channel polarimeter. Here we present two algorithms to jointly estimate the Stokes vectors and optical aberrations using a simpler four-channel phase diverse polarimeter. The method could be adapted easily to eight-channel polarimeters and other variations, but a four-channel polarimeter configuration is particularly attractive in terms of cost and complexity of hardware.

One acquisition parameter that must be chosen is the amount of defocus in the diversity channel(s). Choosing the optimal amount of phase diversity for phase-diverse phase-retrieval in a traditional incoherent imaging system was investigated in [53] using the Cramér-Rao lower bound. In this work we also use the Cramér-Rao lower bound for phase-diverse phase-retrieval as a guide in choosing the amount of defocus to introduce into the system.

For simplicity of presentation, all optical-system elements are assumed to be ideal and all polarimetric measurements are assumed to be perfectly registered.

The organization of this paper is as follows. Section II presents the mathematical framework of joint estimation of object and aberrations from polarimetric measurements. Section III formulates a reduced-parameter search strategy. Section IV explores joint estimation numerically with both quadratic and edge-preserving regularization. Sections V and VI give results and concluding remarks.

3.2 Mathematical Framework

3.2.1 Stokes-Vector Imaging

The optical intensity, Γ, at a single point in an imaging system with a linear polarizer in the optical path having pass axis oriented at angle θ to the reference axis can be expressed in terms of the Stokes vector,

$$\Gamma(\theta) = \frac{1}{2}\left[S_0 + S_1\cos(2\theta) + S_2\sin(2\theta)\right]. \tag{3.1}$$

An imaging polarimeter has multiple channels, each with a different polarization angle. For J measurements (channels) at polarization angles $\theta_1, \ldots, \theta_J$, equation (3.1) becomes a system of J equations. In matrix form the system is

$$\begin{bmatrix} \Gamma(\theta_1) \\ \vdots \\ \Gamma(\theta_J) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & \cos(2\theta_1) & \sin(2\theta_1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(2\theta_J) & \sin(2\theta_J) \end{bmatrix}\begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix}. \tag{3.2}$$

When $\Gamma(\theta_j), S_0, S_1$, and S_2 are images each of size $N \times M$, (3.2) can be configured lexicographically to become

$$\mathbf{\Gamma} = \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p}\right)\mathbf{S}, \qquad n_p = NM, \tag{3.3}$$

where $\mathbf{S} = (S_0, S_1, S_2)$ is a $3n_p \times 1$ column vector, $\mathbf{I}_{n_p}$ is the $n_p \times n_p$ identity matrix, $\otimes$ is the Kronecker product, $\mathbf{T}_{J\times 3}$ is the matrix in (3.2), and $\mathbf{\Gamma}$ is a $Jn_p \times 1$ column vector. The conventional estimate of the Stokes images, $\hat{\mathbf{S}}_{\text{conv}}$, is formed by using the pseudo-inverse of $\mathbf{T}_{J\times 3}$ [15]

$$\hat{\mathbf{S}}_{\text{conv}} = \left\{\left[(\mathbf{T}'_{J\times 3}\mathbf{T}_{J\times 3})^{-1}\mathbf{T}'_{J\times 3}\right] \otimes \mathbf{I}_{n_p}\right\}\mathbf{\Gamma}, \tag{3.4}$$

where " $'$ " denotes conjugate transpose. The matrix inverse in (3.4) is guaranteed to exist if $J \geq 3$ and the θ_j are chosen so that $\mathbf{T}_{J\times 3}$ has linearly independent columns. In words, in (3.4) the $J \times 3$ system of equations in (3.2) is solved by least-squares at each voxel independently.

3.2.2 Forward-Imaging Model

The model (3.3) ignores measurement blur and noise. A more complete discrete-discrete forward model for an incoherent imaging system that accounts for space-invariant optical blur and additive noise can be represented by 2D discrete convolution:

$$y_j(n,m) = b_j(n,m) * \ \Gamma_j(n,m) + \varepsilon_j(n,m) \quad n = 1, \ldots, N, \quad m = 1, \ldots, M \tag{3.5}$$

where $y_j(n,m)$ is the data for the jth channel, $b_j(n,m)$ denotes the incoherent point-spread function associated with the jth channel, $\Gamma_j(n,m)$ is the jth channel ideal intensity image, $*$ denotes 2D convolution, and $\varepsilon_j(n,m)$ is additive noise. A matrix-vector representation of (3.5) is

$$\mathbf{y}_j = \mathbf{B}_j \left[(\mathbf{T}_{J\times 3})_j \otimes \mathbf{I}_{n_p} \right] \mathbf{S} + \boldsymbol{\varepsilon}_j, \qquad j = 1, \ldots, J, \tag{3.6}$$

where $\mathbf{B}_j$ denotes a $n_p \times n_p$ Toeplitz matrix whose entries depend on $b_j(n,m)$, $(\mathbf{T}_{J\times 3})_j$ denotes the jth row of $\mathbf{T}_{J\times 3}$, and $\boldsymbol{\varepsilon}_j$ is an additive noise vector. Stacking J channels (each given by (3.6)) yields

$$\mathbf{y} = \mathbf{B} \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p} \right) \mathbf{S} + \boldsymbol{\varepsilon} \tag{3.7}$$

where $\mathbf{y} \overset{\triangle}{=} (\mathbf{y}_1, \ldots, \mathbf{y}_J)$, $\mathbf{B} \overset{\triangle}{=} \mathrm{diag}\{\mathbf{B}_j\}$ is a block diagonal matrix with the single-channel blur matrices on the diagonal, and $\boldsymbol{\varepsilon} \overset{\triangle}{=} (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_J)$.

3.2.3 Point-Spread-Function Parameterization

Ideally the matrices $\mathbf{B}_j$ (or equivalently the PSFs $b_j(n,m)$) would correspond to diffraction-limited PSFs. In practice the PSF is often degraded by known or unknown aberrations. In the presence of aberrations the generalized pupil function for the system can be written

$$\mathcal{P}(x, y) = A(x, y) \exp\left[\imath W(x, y)\right], \tag{3.8}$$

where $A(x, y)$ is a binary aperture function and $W(x, y)$ is an effective optical path-length error. Figure 3.2 shows the geometry that defines the aberration function W. If the system had no aberrations the exit pupil would have a perfect spherical wave emanating from it towards the focal plane. However, when aberrations are present the wavefront leaving the exit pupil departs from the spherical ideal. The aberration function $W(x, y)$ is the path-length error, with respect to a Gaussian reference sphere, accumulated as a ray passes from the reference sphere to the actual wavefront [43]. It is a well known property of space-invariant optical imaging systems that the coherent transfer function is a scaled version of the generalized pupil function and can be

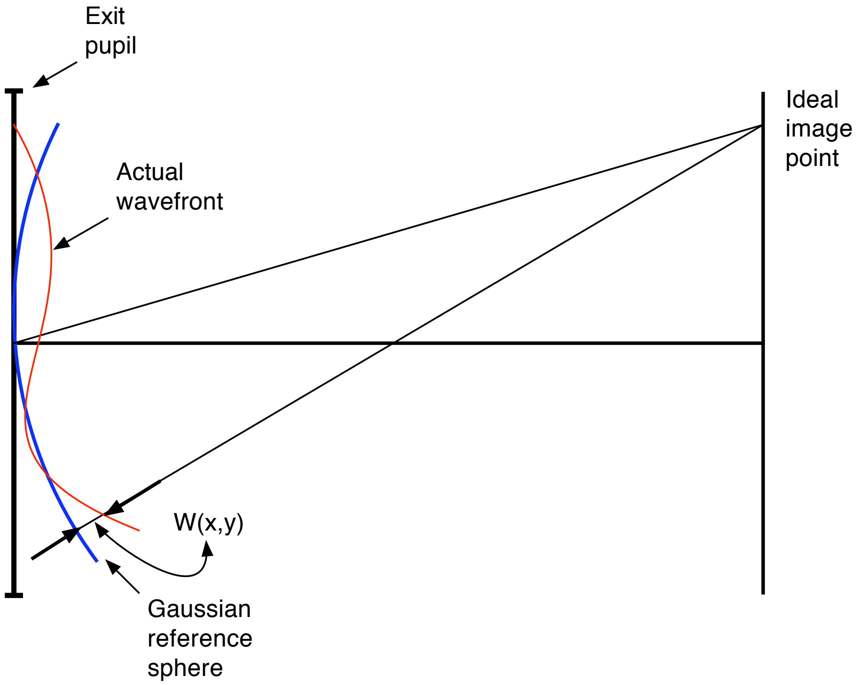


Figure 3.2: Geometry for defining the aberration function $W(x, y)$.

written

$$H(u,v) = A(u,v)\exp\left[\imath W(u,v)\right], \tag{3.9}$$

where (u,v) are frequency domain coordinates [43]. Aberrations in an optical system can be represented using a suitable basis set $\{\varphi_k(u,v)\}$, such as Zernike polynomials [60]. Representing $W(u,v)$ in the basis $\{\varphi_k(u,v)\}$ parameterizes the generalized pupil function:

$$H(u,v;\boldsymbol{\alpha}) = A(u,v)\exp\left[\imath\sum_{k=1}^{K}\alpha_k\varphi_k(u,v)\right] \qquad \text{where } \boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_K). \tag{3.10}$$

Visible regime polarimeter configurations, such as division-of-focal-plane and division-of-amplitude, simultaneously acquire all of the polarimetric channels and so are exposed to identical optical aberrations, i.e., $W(u,v)$ is the same for each channel.

3.2.4 Phase Diversity

To aid in the estimation of aberrations, we propose to introduce phase diversity, typically by different amounts in the different polarimetric channels. Figs. 3.3 and 3.4 show two possible polarimetric-phase-diverse imaging strategies. If the phase diversity function in channel j is denoted $\phi_j(u,v)$, then the generalized pupil function for the jth channel can be written

$$H_j(u,v;\boldsymbol{\alpha}) = A(u,v)\exp\left\{\imath\left[\sum_{k=1}^{K}\alpha_k\varphi_k(u,v) + \phi_j(u,v)\right]\right\}. \tag{3.11}$$

The corresponding incoherent point-spread function, $h_j(x,y)$, and the optical transfer function, $\mathcal{H}_j(u,v)$, can be written in terms of the generalized pupil func-

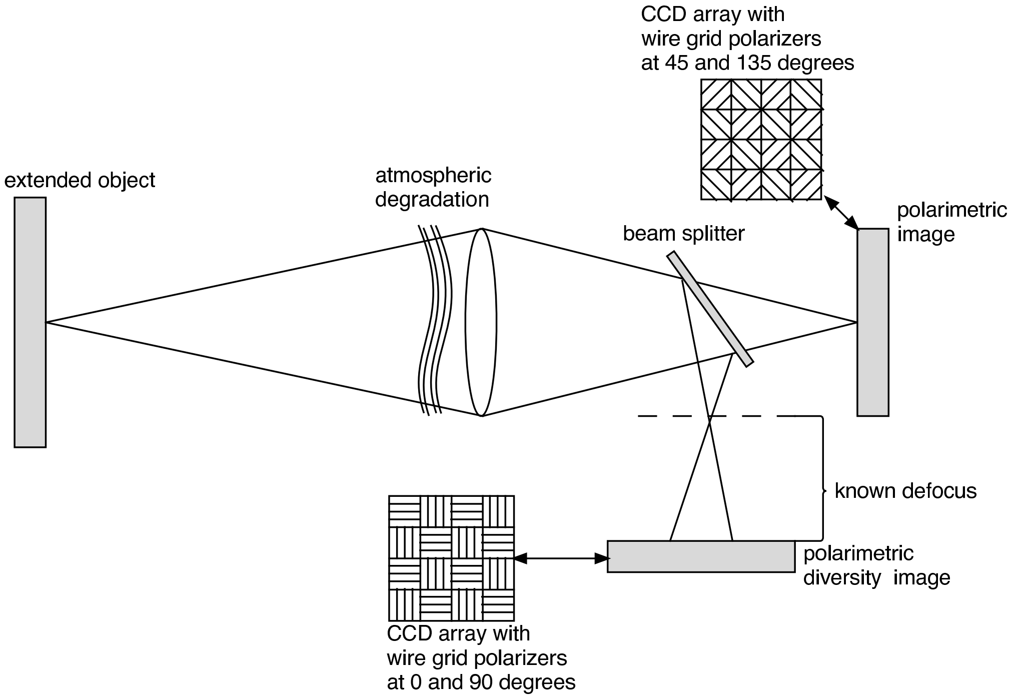


Figure 3.3: Polarimetric phase diversity strategy utilizing the division-of-focal-plane technique.

tion:

$$h_j(x,y;\boldsymbol{\alpha}) = c\left|\mathcal{F}^{-1}\left[H_j(u,v;\boldsymbol{\alpha},\phi_j)\right]\right|^2 \tag{3.12}$$

$$\mathcal{H}_j(u,v;\boldsymbol{\alpha}) = c\ \mathcal{F}\left[\left|\mathcal{F}^{-1}\left[H_j(u,v;\boldsymbol{\alpha},\phi_j)\right]\right|^2\right] \tag{3.13}$$

where $\mathcal{F}[\cdot]$ is the Fourier transform and c is a constant that normalizes the point-spread function to unit volume [53]. The modeled system point-spread function, $b_j(n,m)$, and optical transfer function consist of samples of $h_j(x,y;\boldsymbol{\alpha})$ and $\mathcal{H}_j(u,v;\boldsymbol{\alpha})$ at the Nyquist sampling rate [43], respectively. Consequently, each blur matrix, $\mathbf{B}_j$, is parameterized by the vector $\boldsymbol{\alpha}$. For analysis and implementation we assume periodic boundary conditions on the object so that the blur matrices, $\mathbf{B}_j(\boldsymbol{\alpha})$, are circulant

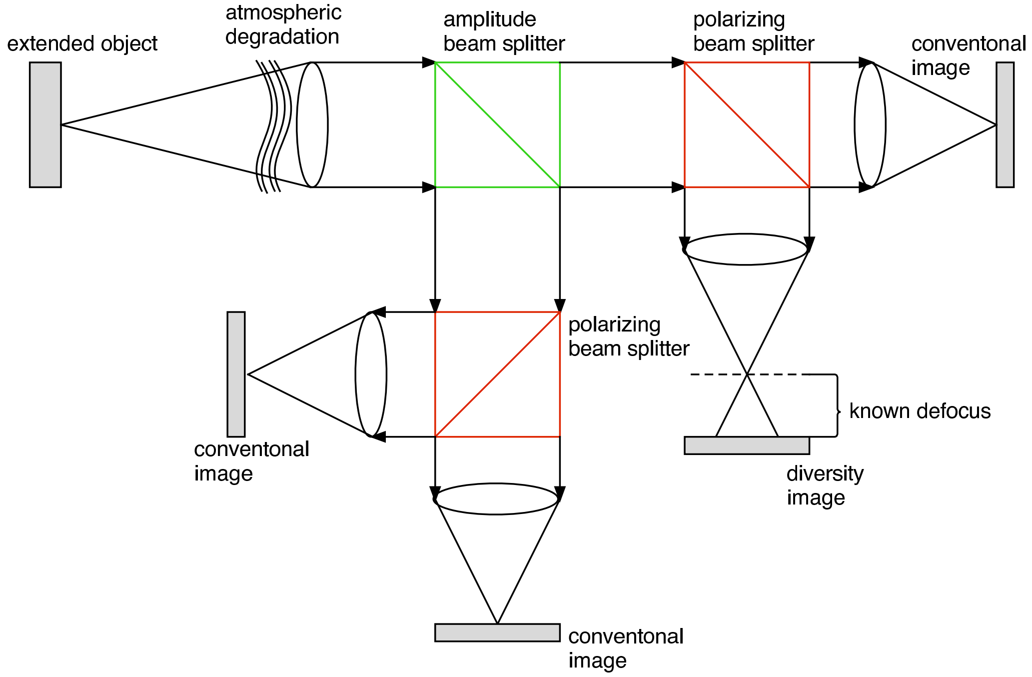


Figure 3.4: Polarimetric phase diversity strategy utilizing the division-of-amplitude technique.

and thus diagonalized by a 2D DFT matrix

$$\mathbf{B}_j(\boldsymbol{\alpha}) = \mathbf{Q}\ \boldsymbol{\Omega}_j(\boldsymbol{\alpha})\ \mathbf{Q}', \tag{3.14}$$

where $\mathbf{Q}$ is a 2D unitary DFT matrix and $\boldsymbol{\Omega}_j(\boldsymbol{\alpha})$ is a diagonal matrix whose entries are the DFT coefficients of the first column of $\mathbf{B}_j(\boldsymbol{\alpha})$.

3.3 Algorithms for Joint Estimation of Stokes Images and Aberrations

This section describes novel algorithms for estimating $\mathbf{S}$ and $\boldsymbol{\alpha}$ jointly under the model (3.7). Under an additive Gaussian noise model $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_p})$ for $j = 1, \ldots, J$, the log-likelihood function for both the object $\mathbf{S}$ and aberration parameters

$\boldsymbol{\alpha}$ is

$$L(\mathbf{S}, \boldsymbol{\alpha}) = -\frac{1}{2\sigma^2} \left\| \mathbf{y} - \mathbf{B}(\boldsymbol{\alpha}) \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p} \right) \mathbf{S} \right\|^2. \tag{3.15}$$

Conventional maximum-likelihood estimation is ineffective for this problem because $\mathbf{B}(\boldsymbol{\alpha})$ is ill-conditioned. Therefore we focus on penalized-likelihood estimators of the form

$$\left(\hat{\mathbf{S}}, \hat{\boldsymbol{\alpha}}\right) = \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}} \left\{ -L(\mathbf{S}, \boldsymbol{\alpha}) + R(\mathbf{S}) \right\} \triangleq \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}} \Psi\left(\mathbf{S}, \boldsymbol{\alpha}\right) \tag{3.16}$$

where $R(\mathbf{S})$ is a regularization term that penalizes an object, $\mathbf{S}$, according to how much it departs from our assumptions about the image properties [28]. In remote sensing $\boldsymbol{\alpha}$ is often a nuisance parameter. However, in an adaptive-optics system with aberration correction capability, $\boldsymbol{\alpha}$ is a parameter of interest. Depending on the task at hand, either $\boldsymbol{\alpha}$ or $\mathbf{S}$ or both can be parameters of interest. The choice of regularization penalty will in general depend on the task, i.e., which parameters are of interest and which are nuisance.

3.3.1 $\boldsymbol{\alpha}$ as a Nuisance Parameter

When the Stokes images, $\mathbf{S}$, are primary interest, then $\boldsymbol{\alpha}$ is a nuisance parameter and a regularization function that reflects *a priori* knowledge about the object should be chosen. Stokes images (S_1, S_2) typically have sharp edges due to man-made objects having stronger polarimetric signatures than naturally occurring objects. To recover as much polarization information as possible the regularization function, $R(\mathbf{S})$, should preserve edges. Since quadratic regularization tends to wash out edges and smooth noise we explore edge-preserving regularization using a hyperbolic potential function $\psi(t; \delta) = \delta^2 \left(\sqrt{1 + \left(\frac{t}{\delta}\right)^2} - 1 \right)$. For fixed δ this function is approximately quadratic for values of $t < \delta$ and approximately linear for $t > \delta$. This behavior will tend to smooth noise and preserve edges. Specifically, we chose $R(\mathbf{S})$ be

$$R(\mathbf{S}) = \sum_{l=0}^{2} \sum_{k=1}^{2n_p} \beta_l \ \psi\left([\mathbf{C}S_l]_k; \delta_l\right), \tag{3.17}$$

where $\mathbf{C}$ is a 2D finite-differencing matrix (horizontal and vertical differences). The estimator (3.16) is then

$$\left(\hat{\mathbf{S}}, \hat{\boldsymbol{\alpha}}\right) = \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \left\| \mathbf{y} - \mathbf{B}(\boldsymbol{\alpha}) \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p}\right) \mathbf{S} \right\|^2 + \sum_{l=0}^{2} \sum_{k=1}^{2n_p} \beta_l \ \psi\left([\mathbf{C}S_l]_k; \delta_l\right) \right\}. \tag{3.18}$$

3.3.2 $\boldsymbol{\alpha}$ As a Parameter of Interest — Reduced Parameter Search Strategy

In [42] it was shown that, for a two channel phase-diversity system under an additive Gaussian noise model, the estimate of the object being imaged could be expressed in terms of the system aberration parameters. This result was generalized in [69] for phase-diverse imaging with an arbitrary number of channels. A similar procedure can be used to derive a closed-form expression for the Stokes images in terms of system aberrations for polarimetric phase-diverse imaging. Deriving this closed-form expression requires the use of a quadratic regularizer that can be diagonalized by the DFT as in (3.14). We focus on quadratic regularizers of the form

$$R(\mathbf{S}) = \frac{1}{2} \left\| (\sqrt{\boldsymbol{\beta}_3} \otimes \mathbf{C}) \mathbf{S} \right\|^2, \tag{3.19}$$

where $\sqrt{\boldsymbol{\beta}_3} \overset{\triangle}{=} \operatorname{diag}\{\sqrt{\beta_0}, \sqrt{\beta_1}, \sqrt{\beta_2}\}$ and $\beta_i > 0 \ \ i = 0, 1, 2$. Using this regularization function (3.16) becomes

$$\left(\hat{\mathbf{S}}, \hat{\boldsymbol{\alpha}}\right) = \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}} \ \frac{1}{2\sigma^2} \left\| \mathbf{y} - \mathbf{B}\left(\boldsymbol{\alpha}\right) \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p}\right) \mathbf{S} \right\|^2 + \frac{1}{2} \left\| (\sqrt{\boldsymbol{\beta}_3} \otimes \mathbf{C}) \mathbf{S} \right\|^2. \tag{3.20}$$

For a fixed aberration vector, $\boldsymbol{\alpha}$, (3.20) is convex in $\mathbf{S}$ and the column gradient of $\hat{\mathbf{S}}$ satisfies the stationary point condition $\nabla_{\mathbf{S}}\Psi(\mathbf{S};\boldsymbol{\alpha}) = \mathbf{0}$, where $\boldsymbol{\Psi}$ is defined in (3.16), which leads to

$$\begin{aligned}\mathbf{S}(\boldsymbol{\alpha}) = {}& \left[\left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{B}\left(\boldsymbol{\alpha}\right)'\mathbf{B}\left(\boldsymbol{\alpha}\right)\left(\mathbf{T}_{J\times 3}\otimes\mathbf{I}_{n_p}\right) + \sigma^2\boldsymbol{\beta}_3\otimes\mathbf{C}'\mathbf{C}\right]^{-1} \\ & \times\left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{B}\left(\boldsymbol{\alpha}\right)'\mathbf{y}.\end{aligned} \tag{3.21}$$

The matrix inverse in (3.21) is guaranteed to exist provided the intersection of the null spaces of the component matrices is the zero vector. Because $\mathbf{C}$ is a first-order finite differencing matrix, the only nonzero vectors in its null space are of the form $\gamma\mathbf{1}$ where $\gamma \in \mathbb{R}$ and $\mathbf{1}$ is the $n_p \times 1$ vector of ones. Therefore, nonzero vectors in the null space of $(\sqrt{\boldsymbol{\beta}_3}\otimes\mathbf{C})$ are of the form $\boldsymbol{v} = (\gamma_1\mathbf{1},\ \gamma_2\mathbf{1},\ \gamma_3\mathbf{1})$ where $\gamma_1, \gamma_2, \gamma_3$ are not all simultaneously zero. It remains to show that $\boldsymbol{v}$ is not in the null space of $\left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{B}\left(\boldsymbol{\alpha}\right)'\mathbf{B}\left(\boldsymbol{\alpha}\right)\left(\mathbf{T}_{J\times 3}\otimes\mathbf{I}_{n_p}\right)$. Now,

$$\begin{aligned}&\left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{B}\left(\boldsymbol{\alpha}\right)'\mathbf{B}\left(\boldsymbol{\alpha}\right)\left(\mathbf{T}_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\boldsymbol{v} \\ &= \left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{Q}\boldsymbol{\Omega}\left(\boldsymbol{\alpha}\right)'\boldsymbol{\Omega}\left(\boldsymbol{\alpha}\right)\underbrace{\mathbf{Q}'\left(\mathbf{T}_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\boldsymbol{v}}_{\boldsymbol{u}},\end{aligned}$$

where the circulant approximation has been used. Observe that $\boldsymbol{u}$ is nonzero only in the DC components. Recall that the optical transfer function, $\boldsymbol{\Omega}_j\left(\boldsymbol{\alpha}\right)$, for space-invariant blur conserves energy and thus does not alter DC components. Thus,

$$\left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{Q}\boldsymbol{\Omega}\left(\boldsymbol{\alpha}\right)'\boldsymbol{\Omega}\left(\boldsymbol{\alpha}\right)\boldsymbol{u} = \left(\mathbf{T}'_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\check{\boldsymbol{v}} \neq \mathbf{0}$$

since $\check{\boldsymbol{v}}$ is a nonzero constant vector.

Substitution of (3.21) into (3.20) yields an “aberration only” objective function,

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\frac{1}{2\sigma^2}\left\|\mathbf{y} - \mathbf{B}\left(\boldsymbol{\alpha}\right)\left(\mathbf{T}_{J\times 3}\otimes\mathbf{I}_{n_p}\right)\mathbf{S}(\boldsymbol{\alpha})\right\|^2 + \frac{1}{2}\left\|(\sqrt{\boldsymbol{\beta}_3}\otimes\mathbf{C})\mathbf{S}(\boldsymbol{\alpha})\right\|^2\right\}. \tag{3.22}$$

The estimate in (3.22) is a *joint* estimate of object and aberrations, that is, minimization over $\boldsymbol{\alpha}$ *implicitly* minimizes over $\mathbf{S}$. Once $\boldsymbol{\alpha}$ has been estimated the object estimate is given by (3.21). We note that this algorithm is a special case of the variable projection method [39].

3.4 Simulation Experiments

We performed simulation experiments to evaluate joint estimation with both edge-preserving regularization (3.18) and quadratic regularization (3.22). The simulations were conducted using the circulant approximation, this approximation was facilitated by tapering the object to its mean at the boundaries. Two situations were considered: 1. the object parameters are of interest and the aberration parameters are nuisance parameters, and 2. the aberration parameters are of interest and the object parameters are nuisance parameters. Because of the significant computational savings afforded by (3.22) we evaluated it with distinct regularization tuned for each object and aberrations. For comparison we also evaluated the conventional estimate (3.4) using the same data without phase diversity. For ground truth, we used polarimetric images collected using a division-focal-plane polarimeter by General Dynamics Advanced Information Systems, Ypsilanti, MI. The linear polarizer pass axes were oriented at $\{0°, 45°, 90°, 135°\}$, and the subsampled polarimetric image size was $[256 \times 256]$ (subsampled from a $[512 \times 512]$ micropolarizer array). The imagery was then corrupted by space-invariant optical blur and additive zero-mean Gaussian noise. The optical blur was constructed using an annular pupil with a phase distortion constructed from Zernike polynomials 4-19 as defined in [60]; the phase distortion had an RMS strength of 0.2 waves. The phase of the generalized pupil function is shown in Fig. 3.5. We define the SNR of an image to be $20\log_{10}(\|\bar{y}\| / \|\bar{y} - y\|)$dB where $\bar{y}$ and y are the noise free and noisy images respectively; the experiments were done at two SNR levels:

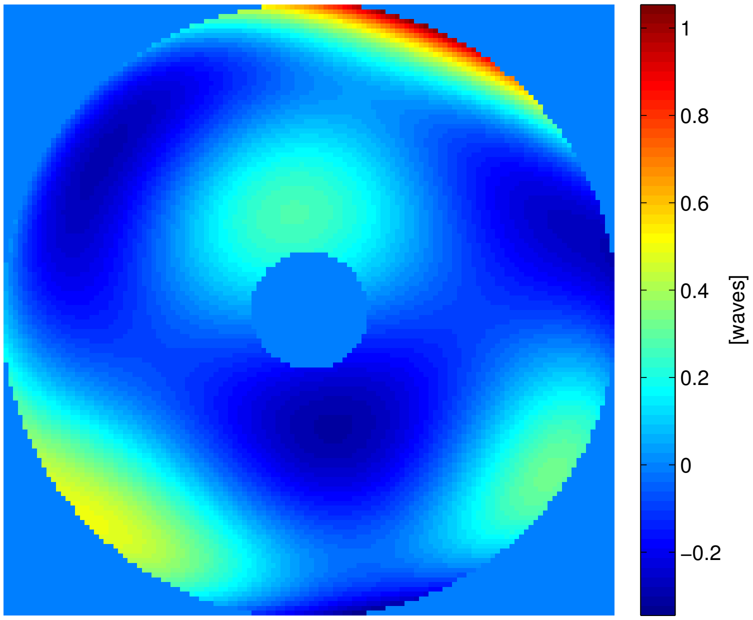


Figure 3.5: Phase of the generalized pupil function in units of waves.

45dB and 25dB. To emulate a traditional phase-diversity configuration the defocused channels were at angles $\{0°, 90°\}$. In this configuration the $\{45°, 135°\}$ channels sum to form the conventional in-focus channel and the $\{0°, 90°\}$ channels sum to form the de-focus channel, see Fig. 3.7. To aid in selecting the amount of defocus to use in the diversity channels we assumed complete knowledge of the object, as in the problem of phase retrieval, in Figs. 3.9 and 3.10 (phase-diverse phase-retrieval) and computed the Cramér-Rao bound for the aberration parameters over a range of defocus values optimized for the estimation of $\boldsymbol{\alpha}$. The Fisher-information matrix is computed from the log-likelihood function in (3.20)

$$\mathbf{F} = \frac{1}{\sigma^2} \left[\nabla_{\boldsymbol{\alpha}} \boldsymbol{\mu}(\boldsymbol{\alpha})\right] \left[\nabla_{\boldsymbol{\alpha}} \boldsymbol{\mu}(\boldsymbol{\alpha})\right]' \tag{3.23}$$

where ∇ denotes the column gradient and $\boldsymbol{\mu}(\boldsymbol{\alpha}) \overset{\triangle}{=} \mathbf{B}(\boldsymbol{\alpha}) \left(\mathbf{T}_{4\times 3} \otimes \mathbf{I}_{n_p}\right) \mathbf{S}$. The Fisher-information matrix was computed and inverted for various values of defocus. Since the Zernike polynomials are orthonormal, the mean of the diagonal elements corresponds to the minimum achievable mean-squared error, $\hat{W}_{\text{MIN}}$, of any unbiased estimator of the degrading wavefront $W(\boldsymbol{\alpha})$. In Fig. 3.6 $\hat{W}_{\text{MIN}}$ is plotted against peak-to-valley defocus. The minimum occurs when the amount of defocus is 1.8 waves peak-to-valley; we used this amount of defocus in the simulations but we note that it is not necessarily the optimal choice for joint estimation of object and aberrations. The blurry and noisy data with and without phase diversity are shown in Figs. 3.7 and 3.8.

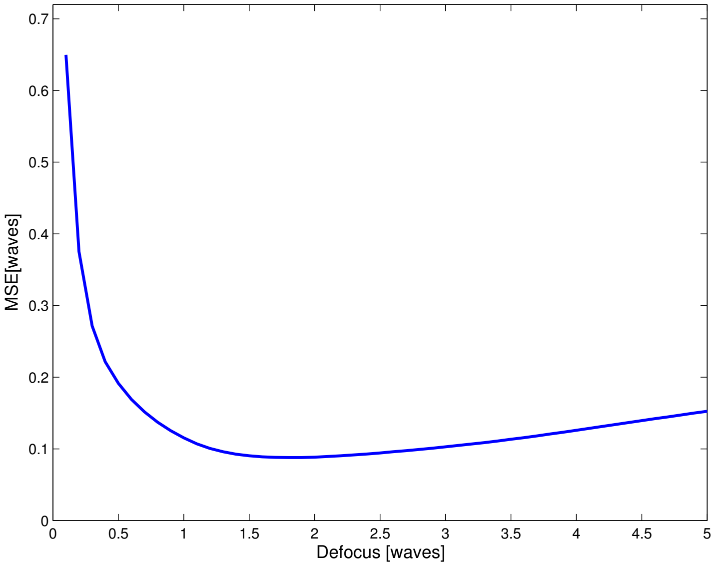


Figure 3.6: Minimum mean squared error as a function of defocus measured from peak to valley, both axes are in units of waves.

Figure 3.7: Data for SNR = 45dB: from left to right: $\{0°, 45°, 90°, 135°\}$. The defocused channels are at $\{0°, 90°\}$.

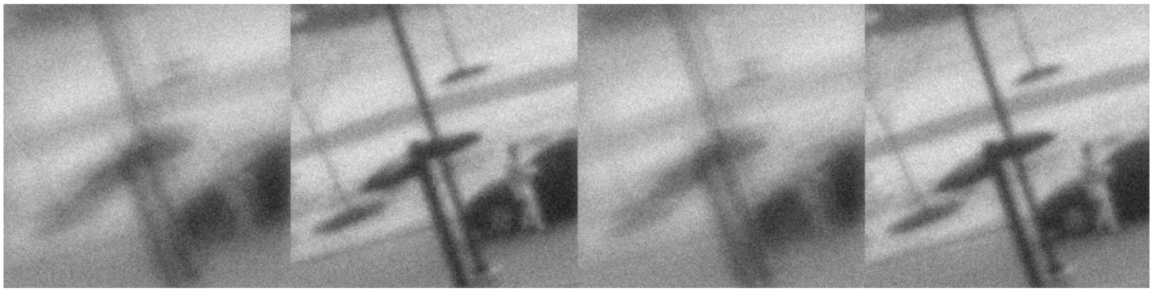


Figure 3.8: Data for SNR = 25dB: from left to right: $\{0°, 45°, 90°, 135°\}$. The defocused channels are at $\{0°, 90°\}$.

Numerical evaluation of (3.18) and (3.22) requires the selection of regularization "tuning" parameters; for (3.18) six parameters must be chosen, $(\beta_0, \beta_1, \beta_2, \delta_0, \delta_1, \delta_2)$, and for (3.22) three parameters must be chosen $(\beta_0, \beta_1, \beta_2)$.

When there is no phase diversity the estimator PSF, $\boldsymbol{l}(\boldsymbol{\alpha})$, of (3.20) for a fixed aberration parameter, $\boldsymbol{\alpha}$, is given by [30]

$$\boldsymbol{l}_k(\boldsymbol{\alpha}) = [\boldsymbol{B}(\boldsymbol{\alpha})'\boldsymbol{B}(\boldsymbol{\alpha}) + \sigma^2\beta_k\mathbf{C}'\mathbf{C}]^{-1}\boldsymbol{B}(\boldsymbol{\alpha})'\boldsymbol{B}(\boldsymbol{\alpha})\boldsymbol{e}_k, \tag{3.24}$$

where $\boldsymbol{B}(\boldsymbol{\alpha})$ is the common blur across channels, $\boldsymbol{e}_k$ is a Kronecker impulse, and $k = 0, 1, 2$ indicates the Stokes image; because the blur is space invariant (3.24) is independent of pixel location. The parameters $(\beta_0, \beta_1, \beta_2)$ in (3.18) were chosen so that in the limit that the hyperbolic potential is approximately quadratic, i.e., (3.18) $\approx$ (3.20), the channel point-spread functions had full width at half maxima (FWHM) of $(r,\ 2r,\ 2r)$, where r is the FWHM of the diffraction-limited point-spread function in the absence of phase diversity, β_0 was calculated using

$$\hat{\beta}_0 = \underset{\beta_0}{\operatorname{argmin}} \left\| \mathrm{FWHM}\left[\boldsymbol{l}_0(\boldsymbol{\alpha})\right] - r \right\|^2, \tag{3.25}$$

and (β_1, β_2) were calculated similarly. Setting the PSF FWHM of $\hat{S}_1$ and $\hat{S}_2$ to twice that of $\hat{S}_0$ is reasonable because of the significantly lower SNR in the S_1 and S_2 images. We generated 20 realizations of $\boldsymbol{\alpha}$ each having RMS phase strengths of 0.2 waves over the pupil. For each aberration realization, (3.25) was solved numerically, then the final values for $(\beta_0, \beta_1, \beta_2)$ were determined by averaging over the ensemble. Once the β parameters were fixed the nonquadratic regularization parameters, $\{\delta_0, \delta_1, \delta_2\}$, were determined by a brute-force multidimensional search for the parameter combination which minimized the normalized RMS error between the Stokes object and the Stokes estimate.

For (3.22) there were three regularization parameters to set for each case. These parameters were determined by a brute-force multidimensional search for the parameter combination which minimized the normalized RMS error between 1. the Stokes object and the Stokes estimate, and 2. the true aberrations and the aberration estimate. The regularization parameters that were "tuned" for object estimation were 10 orders of magnitude larger than those for aberration estimation.

After the regularization parameters were set, the estimators were evaluated over a 20 realization noise ensemble for each of two SNR levels. The initial estimate in each case was formed using (3.4) with the phase-diverse data. Since closed form expressions for the minimizers of (3.18) and (3.22) are not tractable they were minimized numerically. The optimization was done using the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [57]. The minimization of (3.18) required preconditioning due to the different scales of the Stokes images and the aberration parameters. Samples of the Hessian matrix of (3.18) were calculated via finite differences and used in a diagonal preconditioner. The iterative search was stopped when the iteration, k, satisfied $(\Psi_{k+1} - \Psi_k)/\Psi_k < 10^{-10}$, this corresponded to ≈ 200 iterations for (3.18) and ≈ 30 iterations for (3.22).

3.5 Results

Tables 3.1 and 3.2 show normalized RMS estimation errors for each of (3.18), (3.22), and (3.4). The reported errors are of the quantities S_0, the total linear polarization (TPOL) $\sqrt{S_1^2 + S_2^2}$, and wavefront. There is no wavefront error to be reported for the conventional estimate, (3.4), so a value of N/A is listed. Also, the estimates of S_0 and TPOL are listed as N/A for (3.22) when the regularization was tuned for aberration estimation because the estimated images are unrecognizable. The poor object estimates in this case are due to the small values of the regularization parameter. Recall that the object estimate is given by (3.21) which approaches the inverse filter as $\beta \to 0$ and thus greatly amplifies noise. The aberration estimation errors for (3.22) when tuned for object estimation were reasonably good and are included for completeness.

Table 3.1: RMS Error Percentages for SNR = 45dB

cost	**parameter of interest**	S_0	$\sqrt{S_1^2 + S_2^2}$	**wavefront**
edge-preserving	$\mathbf{S}$	$1.8\% \pm 0.01\%$	$36\% \pm 0.3\%$	$3.3\% \pm 0.2\%$
quadratic	$\mathbf{S}$	$1.6\% \pm 0.01\%$	$40\% \pm 0.2\%$	$3.0\% \pm 0.2\%$
quadratic	$\boldsymbol{\alpha}$	N/A	N/A	$1.4\% \pm 0.2\%$
conventional estimate	$\mathbf{S}$	$10\% \pm 0.0013\%$	$60\% \pm 0.11\%$	N/A

Table 3.2: RMS Error Percentages for SNR = 25dB

cost	**parameter of interest**	S_0	$\sqrt{S_1^2 + S_2^2}$	**wavefront**
edge-preserving	$\mathbf{S}$	$6.2\% \pm 0.2\%$	$59\% \pm 1.4\%$	$80\% \pm 4.6\%$
quadratic	$\mathbf{S}$	$6.5\% \pm 0.02\%$	$61\% \pm 1.0\%$	$79\% \pm 0.36\%$
quadratic	$\boldsymbol{\alpha}$	N/A	N/A	$16\% \pm 7\%$
conventional estimate	$\mathbf{S}$	$11\% \pm 0.011\%$	$490\% \pm 1.7\%$	N/A

Figs. 3.9 and 3.10 show object estimates for SNR = 45dB and SNR = 25dB respectively. Each estimate is displayed in RGB format with the RGB channels set

as $[S_0+10\sqrt{S_1^2+S_2^2}, S_0, S_0]$; in this display scheme the polarized elements of the scene are red while the unpolarized elements are in gray scale; the factor of 10 in the red channel was chosen for visual appeal. As expected, the estimates with data at a higher SNR have lower RMS errors and are more visually appealing. Figs. 3.11 and 3.12 show cuts through TPOL reconstructions, at a column having an edge with large polarization content, for SNR = 45dB and SNR = 25dB respectively. The benefit of edge-preserving regularization is apparent in both cases but more pronounced at the 25dB SNR level as the quadratically regularized estimate shows significantly larger blurring across the edge. Figs. 3.13 and 3.14 show the residual wavefronts, that is,

Figure 3.9: Image estimation results for SNR = 45dB. From left to right: object, estimate using edge-preserving regularizer, estimate using quadratic regularizer, and the conventional estimate.

Figure 3.10: Image estimation results for SNR = 25dB. From left to right: object, estimate using edge-preserving regularizer, estimate using quadratic regularizer, and the conventional estimate.

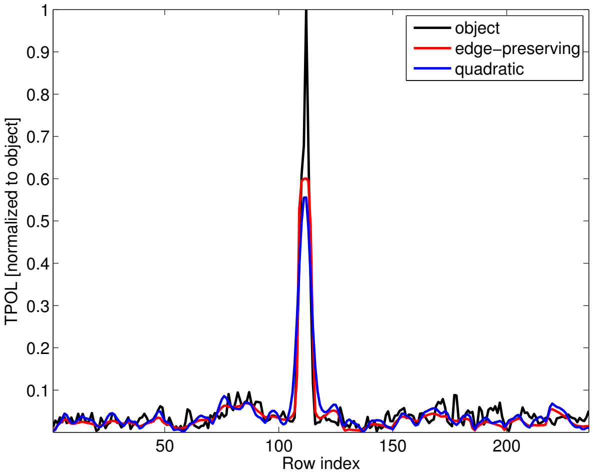


Figure 3.11: Cuts through a column of TPOL for the object and reconstructions with edge-preserving and quadratic regularization at SNR = 45dB.

the estimated wavefront less the true wavefront. The estimates in all cases have lower RMS errors with higher SNR data. At 45dB SNR the wavefront estimation errors are all comparable. At 25dB SNR the wavefront error in using (3.22) (when tuned for aberration estimation) is markedly lower than (3.18) and (3.22) (when tuned for object estimation). This significant reduction in estimation error can be attributed to the regularization being tuned for aberration estimation.

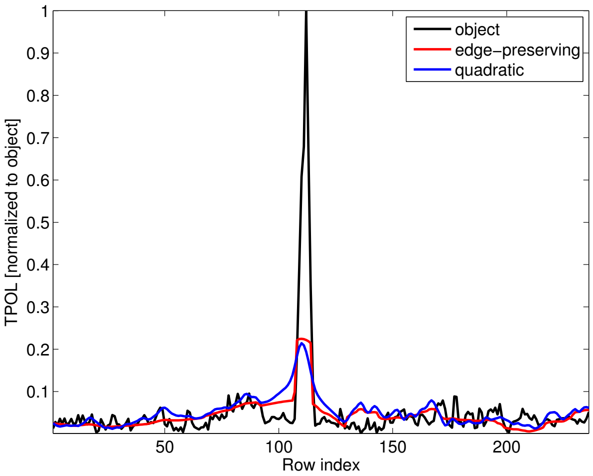


Figure 3.12: Cuts through a column of TPOL for the object and reconstructions with edge-preserving and quadratic regularization at SNR = 25dB.

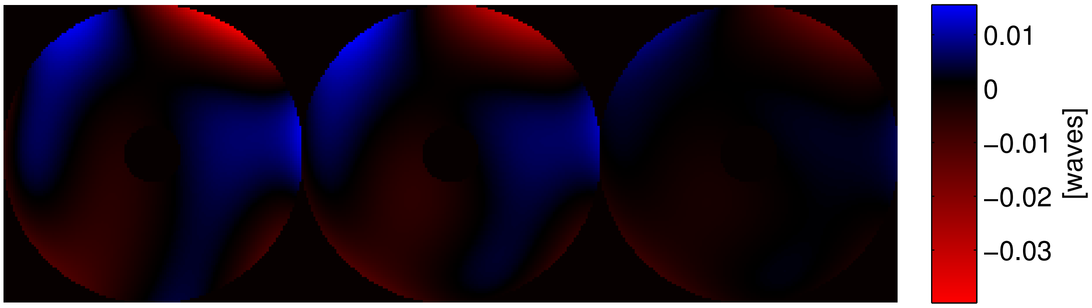


Figure 3.13: Residual wavefront errors for SNR = 45dB. From left to right: edge-preserving regularization, quadratic regularization tuned for object estimation, and quadratic regularization tuned for aberration estimation.

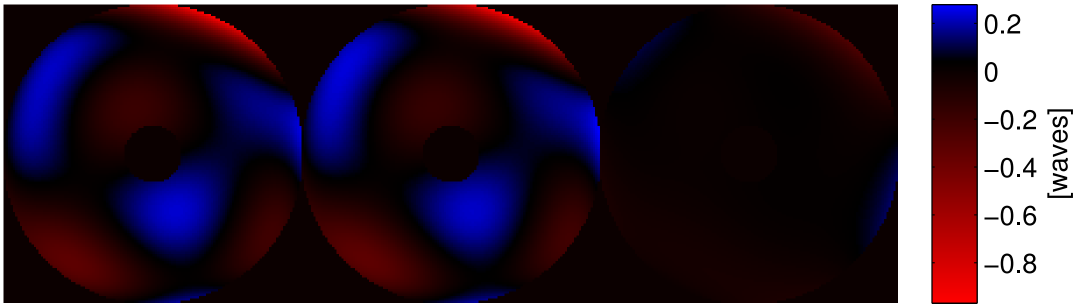


Figure 3.14: Residual wavefront errors for SNR = 25dB. From left to right: edge-preserving regularization, quadratic regularization tuned for object estimation, and quadratic regularization tuned for aberration estimation.

3.6 Conclusions and Future Work

This paper has described two methods, (3.18) and (3.22), for joint estimation of Stokes images and aberrations from polarimetric images with phase diversity. Estimation accuracy follows a task-based hierarchy, i.e., in a joint-estimation framework the choice of algorithm is task dependent. When the task is image restoration (aberrations are nuisance parameters) an algorithm that jointly estimates object and aberrations while incorporating *a priori* knowledge of the *object* is appropriate. However, if the aberration parameters are of interest and the object is a nuisance parameter then a reduced-parameter algorithm should be chosen. We mention that there contrived situations for which aberration estimation in this context will fail. One such circumstance is when the scene is polarized in such a way that the diversity channel receives little or no signal. For example, for a fixed set of polarizer angles $\{0°, 45°, 90°, 135°\}$, if the object is completely polarized along the reference axis (0°), 90° channel will have zero signal; if that channel is the diversity channel the ambiguity in the problem will not be broken and the aberrations will not be able to be estimated.

Future work includes analyzing the bias and covariance of (3.20) and using those expressions to investigate how the choice of diversity channels impacts estimation of Stokes images and aberrations.

CHAPTER IV

Approximation to the object-aberration joint covariance matrix in the phase-diversity context

4.1 Introduction

Incoherent imaging systems have resolution limits that depend, among other things, on known or unknown phase aberrations. Phase aberrations arise from a variety of sources including atmospheric turbulence, misaligned optics within the system, improper mirror figure, and sub-aperture misalignments in multi-aperture systems [40]. Knowledge of system phase aberrations affords either correction through the use of adaptive optics (AO), or post-detection deblurring of the collected imagery via image restoration algorithms.

Phase diversity is an image-based wavefront-sensing technique that allows for the joint estimation of object and phase aberrations. The technique of phase diversity requires the simultaneous collection of two or more images that are related via a deterministic phase perturbation. In the canonical phase diversity configuration, two images are collected: one is the conventional in-focus image and the second image is acquired on a separate focal plane that is translated along the optical axis thereby inducing a known defocus to the second image, see figure 4.1.

For telescopes that require wavefront sensors, such as, telescopes that employ

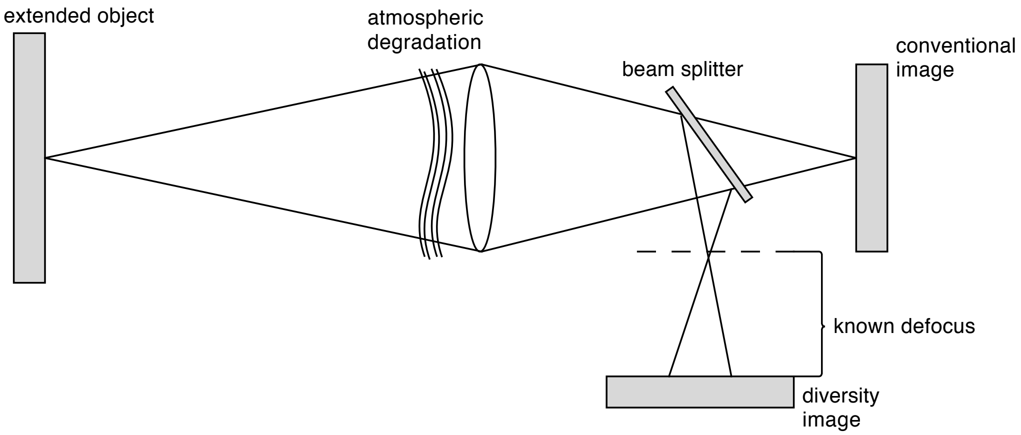


Figure 4.1: Traditional phase-diversity imaging strategy.

AO correction or multi-aperture telescopes that require sub-aperture phasing, phase diversity is a candidate wavefront sensor [71]. When implementing a phase-diverse wavefront sensor there are several acquisition parameters that must be chosen, such as, the type and strength of phase diversity. Several analyses have been done to address the optimality of acquisition parameters in the Cramèr-Rao sense [53, 21, 20, 22, 19], however, these analyses assume complete knowledge of the object being imaged and therefore may give misleading results. Moreover, many phase-diverse wavefront sensors are biased estimators which are not bounded in variance by the classical Cramèr-Rao lower bound. An attractive class of phase-diverse wavefront sensors are quadratically regularized weighted least squares estimators, this is due to the availability of a reduced parameter formulation of the estimator that is explicitly dependent on only the aberrations being estimated. In this paper we present an approximation to the *joint* covariance matrix for a quadratically regularized weighted least squares phase-diverse wavefront sensor; we use this approximation as an analysis tool for aberration estimation performance in the presence of an *unknown* object.

The organization of this chapter is as follows: In section 2 we review the phase diversity concept and algorithms. In section 3 we review approximations to the bias and covariance of implicitly defined estimators. In section 4 we develop an expression for the covariance of aberration estimates for a quadratically regularized weighted least squares phase-diverse wavefront sensor. In section 5 we present simulation experiments, and in sections 6 and 7 we present results and conclusions.

4.2 Review of the phase diversity concept

The technique of phase diversity was first proposed by Gonsalves [42, 41] and was later generalized and put into the framework of maximum-likelihood estimation by Paxman [69]. As an *explicit* maximum-likelihood estimator (MLE) of object and aberrations phase diversity is not a viable wavefront sensing technique owing to the large number of parameters that must be jointly estimated (object pixels + aberration parameters). The viability of phase diversity as an ML wavefront sensor comes from the availability of a closed-form expression for the object in terms of the aberration parameters. In an ML framework that assumes additive Gaussian noise, the implicit function theorems can be invoked and a closed-form expression for the object, in terms of the aberration parameters, can be written . Back substitution of this expression into the MLE reduces the dimensionality of the estimation problem down to only the number of aberration parameters. For completeness we briefly restate these results in the framework of linear algebra.

4.2.1 Forward imaging model

A discrete-discrete forward model for the jth channel of a J channel incoherent imaging system that includes phase diversity, accounts for space-invariant optical blur, and additive noise can be represented by

$$\mathbf{y}_j = t_j \mathbf{B}_j(\boldsymbol{\alpha})\mathbf{S} + \boldsymbol{\varepsilon}_j, \qquad j = 1, \ldots, J, \tag{4.1}$$

where $\mathbf{y}_j$ is an $(n_d \times 1)$ lexicographically ordered data vector, $\mathbf{B}_j(\boldsymbol{\alpha})$ denotes an $(n_p \times n_p)$ Toeplitz matrix that is parameterized by the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$, $\mathbf{S}$ is an $(n_p \times 1)$ lexicographically ordered object vector , $\boldsymbol{\varepsilon}_j$ is an $(n_p \times 1)$ lexicographically ordered noise vector, n_d is the number of collected pixels, and n_p is the number of object pixels. Stacking J channels (each given by (4.1)) yields

$$\mathbf{y} = \mathbf{B}(\boldsymbol{\alpha}) \left(\mathbf{T}_J \otimes \mathbf{I}_{n_p} \right) \mathbf{S} + \boldsymbol{\varepsilon} \tag{4.2}$$

where $\mathbf{y} \overset{\triangle}{=} (\mathbf{y}_1, \ldots, \mathbf{y}_J)$, $\mathbf{T}_J \overset{\triangle}{=} (t_1, \ldots, t_J)$, $\mathbf{B}(\boldsymbol{\alpha}) \overset{\triangle}{=} \mathrm{diag}\{\mathbf{B}_j(\boldsymbol{\alpha})\}$ is a block diagonal matrix with the single-channel blur matrices on the diagonal, $\mathbf{I}_{n_p}$ is the $n_p \times n_p$ identity matrix, $\boldsymbol{\varepsilon} \overset{\triangle}{=} (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_J)$, and $\otimes$ is the Kronecker product.

To aid in the analysis we make the common additional approximation that the $\mathbf{B}_j(\boldsymbol{\alpha})$ are circulant and are thus diagonalized by the unitary 2D-DFT matrix. We discuss the parameterization of the $\mathbf{B}_j(\boldsymbol{\alpha})$ in detail in the next section.

4.2.2 Phase aberration parameterization in the phase diversity context

Phase aberrations in a space-invariant incoherent imaging system are conveniently represented through the generalized pupil function [43]. Let the continuous-space generalized pupil function (also the coherent transfer function in this case) be denoted by $H(u, v)$, then

$$H(u, v) = P(u, v) \exp\left[\imath W(u, v)\right], \tag{4.3}$$

where $P(u, v)$ is a binary aperture function, $W(u, v)$ is a phase aberration function proportional to an effective optical path-length error, and (u, v) are frequency domain coordinates [43]. The phase aberration function, W, can be parameterized by representing it in an appropriate basis. Using a suitable basis, $\{\varphi_k(u, v)\}$, $W(u, v)$ can be expanded leading to

$$H(u,v;\boldsymbol{\alpha}) = P(u,v)\exp\left[\imath\sum_{k=1}^{K}\alpha_k\varphi_k(u,v)\right] \qquad \text{where } \boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_K). \tag{4.4}$$

If the phase diversity function in channel j is denoted $\phi_j(u,v)$, then the generalized pupil function for the jth channel can be written

$$H_j(u,v;\boldsymbol{\alpha}) = P(u,v)\exp\left\{\imath\left[\sum_{k=1}^{K}\alpha_k\varphi_k(u,v) + \phi_j(u,v)\right]\right\}. \tag{4.5}$$

The continuous-space optical transfer function, $\mathcal{H}_j(u,v)$, is then

$$\mathcal{H}_j(u,v;\boldsymbol{\alpha}) = c\ \mathcal{F}\left[\left|\mathcal{F}^{-1}\left[H_j(u,v;\boldsymbol{\alpha})\right]\right|^2\right], \tag{4.6}$$

where $\mathcal{F}[\cdot]$ is the Fourier transform and c is a constant that ensures that the corresponding point-spread function is normalized to unit volume [53]. Let $\boldsymbol{\Omega}_j(\boldsymbol{\alpha})$ denote an $(n_p \times n_p)$ diagonal matrix consisting of samples of (4.6) taken at the Nyquist rate. Then, by invoking the circulant approximation we may write

$$\mathbf{B}_j(\boldsymbol{\alpha}) = \mathbf{Q}\ \boldsymbol{\Omega}_j(\boldsymbol{\alpha})\ \mathbf{Q}' \tag{4.7}$$

where $\mathbf{Q}$ is the unitary 2D-DFT matrix and $'$ indicates conjugate transpose. Consequently, each blur matrix, $\mathbf{B}_j$, is parameterized by the vector $\boldsymbol{\alpha}$.

4.2.3 Maximum-likelihood framework

Under the noise model $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{K}_{\boldsymbol{\varepsilon}})$ a maximum-likelihood estimate of the *joint* parameter vector, $[\mathbf{S}, \boldsymbol{\alpha}]$, for the imaging model (4.2) is given by

$$[\hat{\mathbf{S}}, \hat{\boldsymbol{\alpha}}] = \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}}\, \Phi(\mathbf{S}, \boldsymbol{\alpha}) \tag{4.8}$$

$$= \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}}\, \frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\boldsymbol{\alpha})(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S}\right\|^2_{\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}}, \tag{4.9}$$

where $\mathbf{K}_{\boldsymbol{\varepsilon}}$ is assumed to be nonsingular. For a fixed aberration vector, $\boldsymbol{\alpha}$, (4.9) is convex in $\mathbf{S}$ and so satisfies the stationary-point condition $\nabla_{\mathbf{S}}\Phi(\mathbf{S}, \cdot) = \mathbf{0}$. Using this condition an expression for the object in terms of the aberration parameters can be derived:

$$\mathbf{S}(\boldsymbol{\alpha}) = \left[\mathbf{T}'_J\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{B}(\boldsymbol{\alpha})\mathbf{T}_J\right]^{-1}\mathbf{T}'_J\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{y}. \tag{4.10}$$

We note here that (4.10) contains a matrix inverse and so is not guaranteed to exist for all blur matrices. We can gain insight into the validity of (4.10) by invoking (4.7):

$$\left[(\mathbf{T}'_J \otimes \mathbf{I}_{n_p})\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{B}(\boldsymbol{\alpha})(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\right]^{-1} \tag{4.11}$$

$$= \left[(\mathbf{T}'_J \otimes \mathbf{I}_{n_p})\mathbf{Q}\ \boldsymbol{\Omega}'(\boldsymbol{\alpha})\ \mathbf{Q}'\ \mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\ \mathbf{Q}\ \boldsymbol{\Omega}(\boldsymbol{\alpha})\ \mathbf{Q}'(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\right]^{-1}. \tag{4.12}$$

Observe that each matrix in (4.12) must be invertible. Thus, for (4.10) to exist, the OTFs, $\boldsymbol{\Omega}_j(\boldsymbol{\alpha})$, must not have zeros on the main diagonal. Since imaging system OTFs are a normalized autocorrelation of the corresponding generalized pupil function, we see immediately that systems with circular pupils, sampled at the Nyquist rate, will necessarily have zeros in the OTFs. Traditionally, authors have had success evaluating expressions of the form (4.10) only at points where the matrix inverse exists [12, 13]. As convenient as that is for numerical evaluation, it is an unsatisfying means of regularizing an ill-posed problem. A sufficient condition for regularizing (4.9) while retaining a closed form expression for the object in terms of the aberration parameters is to use a quadratic regularizer that has a positive-definite Hessian. Adding quadratic regularization of the form

$$R(\mathbf{S}) = \beta\frac{1}{2}\left\|\mathbf{C}\mathbf{S}\right\|^2, \tag{4.13}$$

where β is a regularization "tuning" parameter and $\mathbf{C}$ is a matrix such that $\mathbf{C}'\mathbf{C}$ is positive definite, to the objective function in (4.9) provides a sufficient condition for the existence of an expression of the object in terms of the aberration parameters. With regularization of this form (4.9) becomes

$$[\hat{\mathbf{S}}, \hat{\boldsymbol{\alpha}}] = \underset{(\mathbf{S},\boldsymbol{\alpha})}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\boldsymbol{\alpha})(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S}\right\|^2_{\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}} + \beta\frac{1}{2}\left\|\mathbf{C}\mathbf{S}\right\|^2\right\} \tag{4.14}$$

$$\Longleftrightarrow \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\frac{1}{2}\left\|\mathbf{y} - \mathbf{B}(\boldsymbol{\alpha})(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S}(\boldsymbol{\alpha})\right\|^2_{\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}} + \beta\frac{1}{2}\left\|\mathbf{C}\mathbf{S}(\boldsymbol{\alpha})\right\|^2\right\}, \tag{4.15}$$

where

$$\mathbf{S}(\boldsymbol{\alpha}) = \left[\mathbf{T}_J'\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{B}(\boldsymbol{\alpha})\mathbf{T}_J + \beta\mathbf{C}'\mathbf{C}\right]^{-1}\mathbf{T}_J'\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{y}. \tag{4.16}$$

Observe that the matrix inverse in (4.16) exists because $\mathbf{T}_J'\mathbf{B}'(\boldsymbol{\alpha})\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{B}(\boldsymbol{\alpha})\mathbf{T}_J$ is nonnegative definite and $\beta\mathbf{C}'\mathbf{C}$ is positive definite so their sum is positive definite. Estimators of the form (4.14) are biased and not bounded in variance by the classical Cramèr-Rao lower bound. To obtain performance bounds, in terms of mean-squared error, for estimators of the type (4.14) the bias and variance must either be calculable directly or approximated. In the next section we review results on the approximation of the covariance of an implicitly defined estimator.

4.3 Covariance approximation for implicitly defined estimators

In this section we review the results of Fessler [29]. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n_p}) \in \mathbb{R}^{n_p}$ be an unknown real parameter vector that is to be estimated from a measurement vector $\mathbf{Y} = (Y_1, \ldots, Y_{n_d}) \in \mathbb{R}^{n_d}$. Let the estimator be of the form

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}} \Phi(\boldsymbol{\theta}, \mathbf{Y}), \tag{4.17}$$

estimators of this form can be viewed as mappings from the data space to the parameter space. That is, (4.17) can be written $\hat{\boldsymbol{\theta}} = h(\mathbf{Y})$ where $h : \mathbb{R}^{n_d} \to \mathbb{R}^{n_p}$. Under the assumption that $\Phi(\cdot, \mathbf{Y})$ has a global minimum, $\hat{\boldsymbol{\theta}}$, the stationary-point condition is satisfied for each component of the parameter vector at $\hat{\boldsymbol{\theta}}$:

$$\mathbf{0} = \frac{\partial}{\partial \theta_j} \Phi(\boldsymbol{\theta}, \mathbf{Y}) \bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \qquad j = 1, \ldots, n_p. \tag{4.18}$$

At this point we note that the implicit function theorems guarantee the existence of $h(\mathbf{Y}) = (h_1(\mathbf{Y}), \ldots, h_{n_p}(\mathbf{Y}))$ for suitable regular Φ. Thus, we can rewrite (4.18) as

$$\mathbf{0} = \frac{\partial}{\partial \theta_j} \Phi(h(\mathbf{Y}), \mathbf{Y}) \bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \qquad j = 1, \ldots, n_p. \tag{4.19}$$

Now, consider the first order Taylor expansion of $h(\mathbf{Y})$

$$h(\mathbf{Y}) \approx h(\bar{\mathbf{Y}}) + \nabla_{\mathbf{Y}} h(\bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}}) \quad \text{where } \nabla_{\mathbf{Y}} = \left(\frac{\partial}{\partial Y_1}, \ldots, \frac{\partial}{\partial Y_{n_d}} \right). \tag{4.20}$$

Then,

$$\operatorname{Cov}(\hat{\boldsymbol{\theta}}) = \operatorname{Cov}(h(\mathbf{Y})) \approx \nabla_{\mathbf{Y}} h(\bar{\mathbf{Y}}) \operatorname{Cov}(\mathbf{Y}) \nabla_{\mathbf{Y}} h'(\bar{\mathbf{Y}}), \tag{4.21}$$

where we have used the identity $\mathrm{Cov}(\mathbf{A}\mathbf{x}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{x})\mathbf{A}'$. Note that the dependence on h is only through its partial derivatives at $\bar{\mathbf{Y}}$. Applying the chain rule to (4.19) we have

$$\nabla_{\mathbf{Y}}\left[\nabla_{\boldsymbol{\theta}}\Phi(h(\mathbf{Y}), \mathbf{Y})\right] = \mathbf{0}_{n_p \times n_d} \tag{4.22}$$

$$\Rightarrow \nabla^{[2,0]}\Phi(h(\mathbf{Y}), \mathbf{Y})\nabla_{\mathbf{Y}}h + \nabla^{[1,1]}\Phi = \mathbf{0}_{n_p \times n_d} \tag{4.23}$$

$$\Rightarrow \nabla_{\mathbf{Y}}h = -[\nabla^{[2,0]}\Phi(h(\mathbf{Y}), \mathbf{Y})]^{-1}\nabla^{[1,1]}\Phi(h(\mathbf{Y}), \mathbf{Y}), \tag{4.24}$$

where (j,k)th element of the $(n_p \times n_p)$ operator $\nabla^{[2,0]}$ is $\frac{\partial^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_k}$, and the (j,n)th element of the $(n_p \times n_d)$ operator $\nabla^{[1,1]}$ is $\frac{\partial^2}{\partial\boldsymbol{\theta}_j\partial\mathbf{Y}_n}$. Substitution into (4.21), and evaluation at $\bar{\mathbf{Y}}$, yields an approximation for the covariance of $\hat{\boldsymbol{\theta}}$

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \approx \left[-[\nabla^{[2,0]}\Phi(\check{\boldsymbol{\theta}}, \bar{\mathbf{Y}})]^{-1}\nabla^{[1,1]}\Phi(\check{\boldsymbol{\theta}}, \bar{\mathbf{Y}})\right]\mathrm{Cov}(\mathbf{Y})\left[-[\nabla^{[2,0]}\Phi(\check{\boldsymbol{\theta}}, \bar{\mathbf{Y}})]^{-1}\nabla^{[1,1]}\Phi(\check{\boldsymbol{\theta}}, \bar{\mathbf{Y}})\right]', \tag{4.25}$$

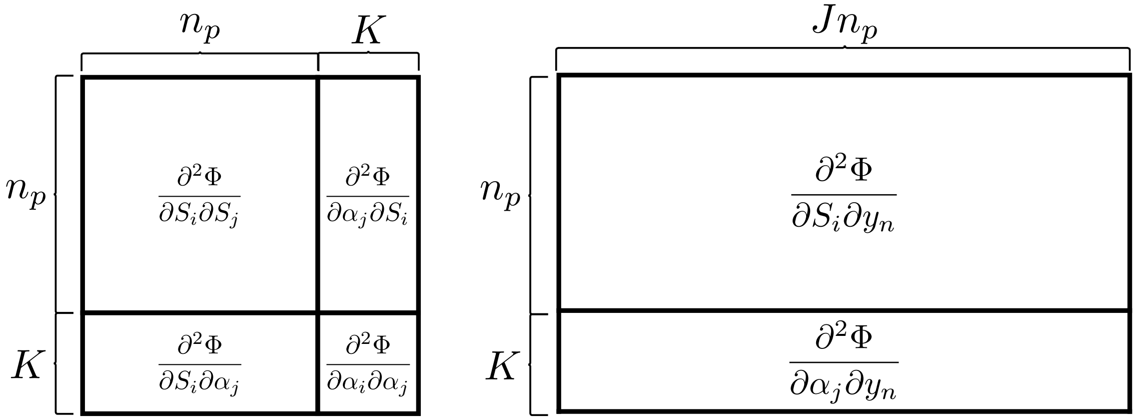
where $\check{\boldsymbol{\theta}} \overset{\triangle}{=} h(\bar{\mathbf{Y}})$.

4.4 Approximation to the covariance of aberration estimates in the phase diversity context

In this section we give an expression for the approximate *joint* covariance matrix for the aberration estimates obtained using (4.14). Let $\boldsymbol{\theta} = [\mathbf{S}, \boldsymbol{\alpha}]$, the operators $\nabla^{[2,0]}$ and $\nabla^{[1,1]}$ are then of sizes $(n_p + K) \times (n_p + K)$ and $(n_p + K) \times Jn_p$ respectively; the operators acting on Φ are shown graphically in Figure 4.2.

Let $\nabla^{[2,0]}\Phi(\mathbf{S}, \boldsymbol{\alpha}) = \mathbf{F}$, and its block-matrix representation be

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix}. \tag{4.26}$$



(a) $\nabla^{[2,0]}\Phi$ (b) $\nabla^{[1,1]}\Phi$

Figure 4.2: partitions of $\nabla^{[2,0]}\Phi$ and $\nabla^{[1,1]}\Phi$

The block elements are given by

$$\mathbf{F}_{11} = (\mathbf{T}'_J \otimes \mathbf{I}_{n_p})\mathbf{B}'\mathbf{K}_\varepsilon^{-1}\mathbf{B}(\mathbf{T}_J \otimes \mathbf{I}_{n_p}) + \beta\mathbf{C}'\mathbf{C} \tag{4.27}$$

$$\mathbf{F}_{12} = (\mathbf{T}'_J \otimes \mathbf{I}_{n_p})\left(\mathbf{B}'\mathbf{K}_\varepsilon^{-1}\frac{\partial\mathbf{B}}{\partial\boldsymbol{\alpha}}\right)(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S} \tag{4.28}$$

$$\mathbf{F}_{21} = \mathbf{F}'_{12} \tag{4.29}$$

$$[\mathbf{F}_{22}]_{(k,l)} = \left\langle \frac{\partial\mathbf{B}}{\partial\alpha_l}(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S}\ ,\ \frac{\partial\mathbf{B}}{\partial\alpha_k}(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S} \right\rangle_{\mathbf{K}_\varepsilon^{-1}}, \tag{4.30}$$

where we have suppressed the dependence on $\boldsymbol{\alpha}$ for brevity. The matrix $\nabla^{[1,1]}\Phi$ can be partitioned as a block-vector:

$$\nabla^{[1,1]}\Phi = \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix}, \tag{4.31}$$

where

$$\mathbf{G} = \left(\nabla_{\mathbf{S}}\left[(\nabla_{\mathbf{y}}\Phi)\right]\right) = -\left(\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{B}(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\right)' \tag{4.32}$$

$$\mathbf{H} = \frac{\partial}{\partial \boldsymbol{\alpha}}(\nabla_{\mathbf{y}}\Phi) = -\mathbf{K}_{\boldsymbol{\varepsilon}}^{-1}\frac{\partial \mathbf{B}}{\partial \boldsymbol{\alpha}}(\mathbf{T}_J \otimes \mathbf{I}_{n_p})\mathbf{S}. \tag{4.33}$$

The covariance matrix of the aberration estimates can be extracted in the following way. Let $\tilde{\boldsymbol{e}}_k = (\mathbf{0}, \boldsymbol{e}_k)$, where $\mathbf{0}$ is a vector of zeros with length n_p, and $\boldsymbol{e}_k$ is a vector of length K having a 1 in the kth location. The covariance of the kth and lth aberration parameter estimates is then

$$\begin{aligned}
\operatorname{Cov}(\hat{\alpha}_k, \hat{\alpha}_l) &= \tilde{\boldsymbol{e}}_k{}' \operatorname{Cov}(\hat{\boldsymbol{\theta}})\tilde{\boldsymbol{e}}_l && (4.34)\\
&\approx \tilde{\boldsymbol{e}}_k{}'[-\nabla^{[2,0]}\Phi]^{-1}[\nabla^{[1,1]}\Phi]\mathbf{K}_{\boldsymbol{\varepsilon}}[\nabla^{[1,1]}\Phi]'[-\nabla^{[2,0]}\Phi]^{-1}\tilde{\boldsymbol{e}}_l && (4.35)\\
&= (\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}[\nabla^{[1,1]}\Phi]'[\nabla^{[2,0]}\Phi]^{-1}\tilde{\boldsymbol{e}}_k)'(\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}[\nabla^{[1,1]}\Phi]'[\nabla^{[2,0]}\Phi]^{-1}\tilde{\boldsymbol{e}}_l), && (4.36)
\end{aligned}$$

where we have used the symmetry of $\nabla^{[2,0]}\Phi$ and assumed $\mathbf{K}_{\boldsymbol{\varepsilon}}$ is symmetric positive definite. When the variance of the individual aberration estimates is desired (4.36) simplifies to

$$\operatorname{Var}(\hat{\alpha}_k) = \left\|\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}[\nabla^{[1,1]}\Phi]'[\nabla^{[2,0]}\Phi]^{-1}\tilde{\boldsymbol{e}}_k\right\|^2. \tag{4.37}$$

The expressions (4.36) and (4.37) can be reduced further by writing out the elements of $[\nabla^{[2,0]}\Phi]^{-1}$:

$$[\nabla^{[2,0]}\Phi]^{-1}(\mathbf{S}, \boldsymbol{\alpha}) = \mathbf{F}^{-1} = \begin{bmatrix} (\mathbf{F}_{11} - \mathbf{F}_{12}\mathbf{F}_{22}^{-1}\mathbf{F}_{21})^{-1} & -\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\Delta^{-1} \\ -\Delta^{-1}\mathbf{F}_{21}\mathbf{F}_{11}^{-1} & \Delta^{-1} \end{bmatrix}, \tag{4.38}$$

where $\Delta = \mathbf{F}_{22} - \mathbf{F}_{21}\mathbf{F}_{11}^{-1}\mathbf{F}_{12}$ is the *Schur* complement. Observe that $\nabla^{[2,0]}\Phi$ is a *joint* Fisher-Information Matrix (FIM) that can be used when $\mathbf{B}(\boldsymbol{\alpha})$ is nonsingular.

Moreover, $\mathbf{F}_{22}$ is the FIM that is typically used when computing the known-object Cramèr-Rao bound on aberration estimates for phase-diverse wavefront sensing. It was first pointed out in [71] that the *joint* Cramèr-Rao bound for aberration estimates consists of a "known object" term less a correction term due to the object being unknown. Continuing we have

$$[\nabla^{[2,0]}\Phi]^{-1}\tilde{\boldsymbol{e}_k} = \begin{bmatrix} (\mathbf{F}_{11} - \mathbf{F}_{12}\mathbf{F}_{22}^{-1}\mathbf{F}_{21})^{-1} & -\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\Delta^{-1} \\ -\Delta^{-1}\mathbf{F}_{21}\mathbf{F}_{11}^{-1} & \Delta^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{e}_k \end{bmatrix} \tag{4.39}$$

$$= \begin{bmatrix} -\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\Delta^{-1}\boldsymbol{e}_k \\ \Delta^{-1}\boldsymbol{e}_k \end{bmatrix} \tag{4.40}$$

$$= \begin{bmatrix} -\mathbf{F}_{11}^{-1}\mathbf{F}_{12} \\ \mathbf{I} \end{bmatrix} \Delta^{-1}\boldsymbol{e}_k. \tag{4.41}$$

The covariance of the kth and lth aberration estimates can now be written

$$\mathrm{Cov}(\hat{\alpha}_k, \hat{\alpha}_l) \approx \left(\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}\left(\mathbf{H}' - \mathbf{G}'\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\Delta^{-1}\boldsymbol{e}_k\right)' \left(\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}\left(\mathbf{H}' - \mathbf{G}'\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\Delta^{-1}\boldsymbol{e}_l\right). \tag{4.42}$$

When the variances of the aberration estimates are desired (4.42) reduces to

$$\mathrm{Var}(\hat{\alpha}_k) \approx \left\|\mathbf{K}_{\boldsymbol{\varepsilon}}^{1/2}\left(\mathbf{H}' - \mathbf{G}'\mathbf{F}_{11}^{-1}\mathbf{F}_{12}\right)\Delta^{-1}\boldsymbol{e}_k\right\|^2. \tag{4.43}$$

Recall that $\Delta^{-1}\boldsymbol{e}_k = [\Delta^{-1}]_k$, the kth column of Δ^{-1}. Also recall that matrix-vector multiplication is a weighted sum of the columns of the matrix with the elements of the vector. So our matrix multiply is a weighted sum of images with the columns of Δ^{-1} as the weights.

4.4.1 Discussion

The expression (4.42) can now, in principle, be evaluated. However, care must be taken when using (4.42) because it depends on the actual aberrations being estimated and the object being imaged. In a practical application one would specify statistical classes of aberrations and objects *a priori* and then run Monte Carlo experiments over those classes. From those data conclusions might then be drawn for the variance of the aberration estimates from a particular statistical class. Also, (4.42) will be of higher value when the estimation error is dominated by variance; as such, our focus will be on imaging scenarios of moderate to low SNR. In the following section we adopt a Monte Carlo approach for evaluating (4.43) in imaging scenarios of moderate to low SNR. We note that Monte Carlo techniques of this kind are commonly used in the AO and wavefront sensing literature [60, 33, 35, 99, 53]. Considering, analytically, the effects of a random aberration vector, $\boldsymbol{\alpha}$, on the covariance approximation will be done in our future work.

4.5 Simulation Experiments

Our simulation experiments follow a straightforward Monte-Carlo paradigm. The simulations were conducted using the circulant approximation, this approximation was facilitated by tapering the object to its mean at the boundaries. For a given phase-diverse wavefront sensing configuration we evaluate (4.43) over an ensemble of phase screens from a particular statistical class. We then sum the variances for each phase screen, and then average those results over the phase screen ensemble collapsing the estimate variances to a scalar quantity. This scalar quantity is a figure of merit for the particular phase-diverse wavefront sensing configuration under consideration.

For comparison, we evaluate the known object Cramèr-Rao bound for the model (4.1). The expression for the elements of the known object FIM are given by (4.30). The

FIM is populated and inverted for each phase screen realization, the diagonal elements are averaged, then we average over the phase screen ensemble to obtain a scalar figure of merit. We also evaluate (4.15) for verification. Equation (4.15) is evaluated over both noise and phase screen ensembles. The reason for this is that we are building up an empirical variance for each phase screen realization and then averaging that over the phase screen ensemble to obtain a scalar quantity for the phase-diverse wavefront sensing configuration under consideration.

The phase diversity configuration that has the least amount of hardware complexity is the canonical configuration in Figure 4.1; for this reason we focus on the canonical configuration. An important parameter to choose in this configuration is the strength of defocus imparted to the out-of-focus image; analyses regarding the selection of this parameter using the known-object Cramèr-rao bound have been reported in the literature [53, 21].

In our experiments we use two objects: (i) an extended scene, and (ii) a point object. The extended scene was of size $[64 \times 64]$ and taken from a larger $[256 \times 256]$ image, the subregion that is processed is indicated by a red box in figure 4.3; the point object is also of size $[64 \times 64]$ pixels. The imagery is corrupted by space-invariant blur and zero mean independent-identically-distributed additive Gaussian noise of variance σ^2. The objects, blurred objects, and blurry and noisy objects are shown in figures 4.3 and 4.4.

In all of the calculations presented here the pupil phase aberrations were parameterized using Zernike polynomials 4 through 19 as defined in [60]; that is, the expansion basis, $\{\varphi_i\}$, is composed of Zernike polynomials. Each phase screen realization, in a 50 realization ensemble, was constructed by first drawing aberration coefficients, $\boldsymbol{\alpha}$, from a uniform distribution over the interval $[-1, 1]$ and using them as weights in a basis expansion over an annular pupil. The resulting phase screen is

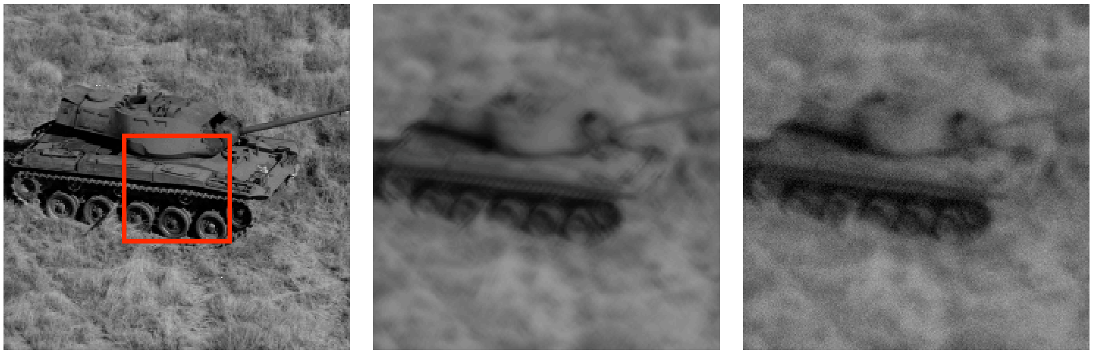


Figure 4.3: From left to right: pristine tank with subregion for processing indicated by the red box, tank with 0.2 RMS waves of optical blur, tank with 0.2 RMS waves of optical blur and additive noise to 25dB

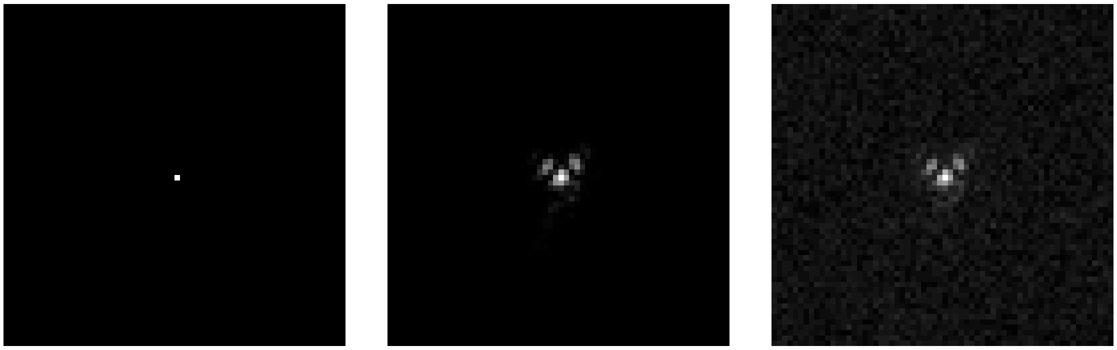


Figure 4.4: From left to right: point object, point with 0.2 RMS waves of optical blur, point with 0.2 RMS waves of optical blur and additive noise to 5dB

then normalized to 0.2 waves RMS over the pupil. The phase of the generalized pupil function for one particular realization in the ensemble is shown in figure 4.5. A word of caution is in order at this point. The use of only the first 15 non-planar Zernike polynomials in a independent identically uniformly distributed manner is not meant to correspond to a particular optical circumstance; however, it may resemble random low order optical fabrication errors. Also, an ensemble of 50 realizations may not be large enough to adequately represent a statistical class. The question of adequate ensemble size requires further research.

The SNR of the extended scene is 25dB and that of the point object is 5dB, where we define the SNR of an image to be $20\log_{10}(\|\bar{y}\| \,/\, \|\bar{y}-y\|)$dB where $\bar{y}$ and y are the

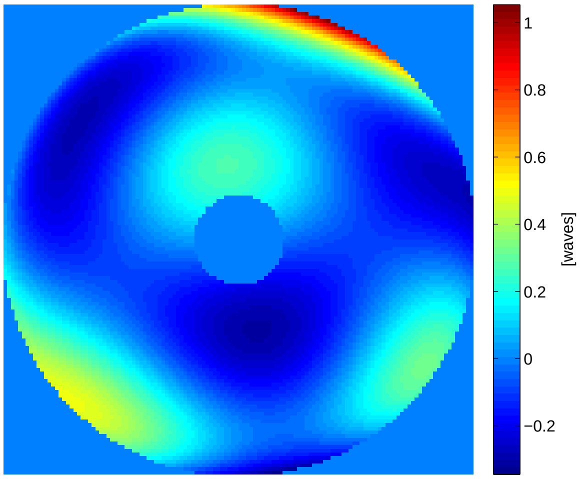


Figure 4.5: A particular wavefront realization in the Monte Carlo ensemble.

noise free and noisy images respectively. We note also that the SNR is only computed over pixels that contain signal. This method of computing the SNR ensures that the SNR of a point object does not depend on the size of the imaging array.

Since a closed form expression for the minimizer of (4.15) is not tractable it was minimized numerically. The optimization was done using the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [57]. The iterative search was stopped when the iteration, k, satisfied $(\Psi_{k+1} - \Psi_k)/\Psi_k < 10^{-10}$, this corresponded to ≈ 30 iterations.

As an example showing that the bulk of the estimation error, in phase-diverse wavefront sensing, comes from variance in moderate to low SNR imaging scenarios we evaluated (4.15) over a 20 realization noise ensemble and a single phase screen realization. Figure 4.6 shows scatter plots of the individual estimates (left) and the average estimates over the ensemble (right); perfect estimation corresponds to estimates lying on the diagonal line. The scatter plot on the right shows that the average estimates are quite close to the true values indicating a small bias; whereas the variance of the individual estimates is comparatively large.

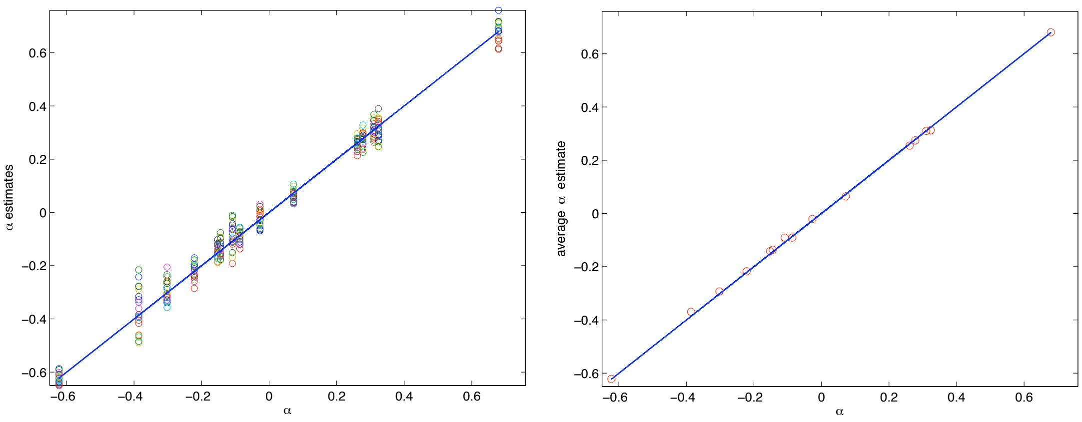


Figure 4.6: Individual aberration estimates and averaged estimates.

The evaluation of (4.43) requires the selection of a regularization "tuning" parameter. Previous experience with phase diversity suggests that the aberration estimates are weakly dependent on the regularization parameter provided the regularization is small. We verify this claim by evaluating (4.43) over our phase screen ensemble for several different regularization parameters. The resulting curve is shown in Fig. 4.7. The curve shows that for regularization parameters smaller than approximately 10^{-6} the standard deviation of the estimates is nearly constant.

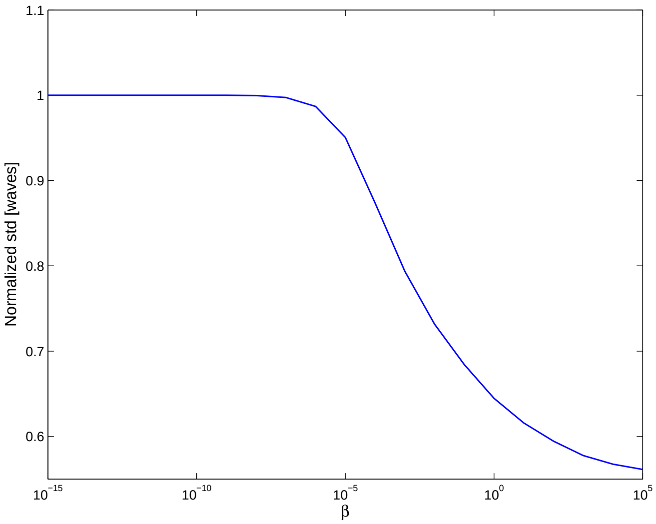


Figure 4.7: Estimated standard deviation as a function of β.

4.6 Results and discussion

The results of our simulation experiments for selection of the defocus parameter are summarized in table 4.1. Figure 4.8 shows comparisons of the average of the estimate standard deviations using (4.43) and the known-object Cramèr-Rao bound using (4.30), as well as the RMSE of the Monte Carlo simulation, for the extended scene over a range of defocus values. An immediate observation is that the Cramèr-Rao bound method is lower than the estimated standard deviations using (4.43). This is consistent with intuition as the Cramèr-Rao bound assumes complete knowledge of the object and one would expect improved estimator performance under that condition. Observe further that the minima of the two methods are different. The Cramèr-Rao bound method predicts an optimal defocus value of 1.30 waves whereas equation (4.43) predicts an optimal defocus value of 1.75 waves. The Monte Carlo simulations yield an optimal defocus of 2.20 waves. The predicted optimal defocus value using approximation (4.43) is closer to the Monte Carlo result than that predicted by the known-object Cramèr-Rao bound. This makes intuitive sense as (4.43) incorporates the uncertainty in the object whereas the Cramèr-Rao method does not.

Table 4.1: Predictions of optimal defocus strength using (4.43), (4.30), and (4.15)

method	**extended scene**	**point object**
Monte Carlo: Eq. (4.15)	2.20 waves	1.05 waves
Variance approximation: Eq. (4.43)	1.75 waves	0.95 waves
Known-object CRB: Eq. (4.30)	1.30 waves	0.85 waves

Figure 4.9 shows comparisons of the average of the estimate variances using (4.43) and (4.30), as well as the RMSE of the Monte Carlo simulation, for the point object over a range of defocus values. As with the extended scene, the Cramèr-Rao bound is lower than the variance predicted by (4.43). We see also that the optimal defocus

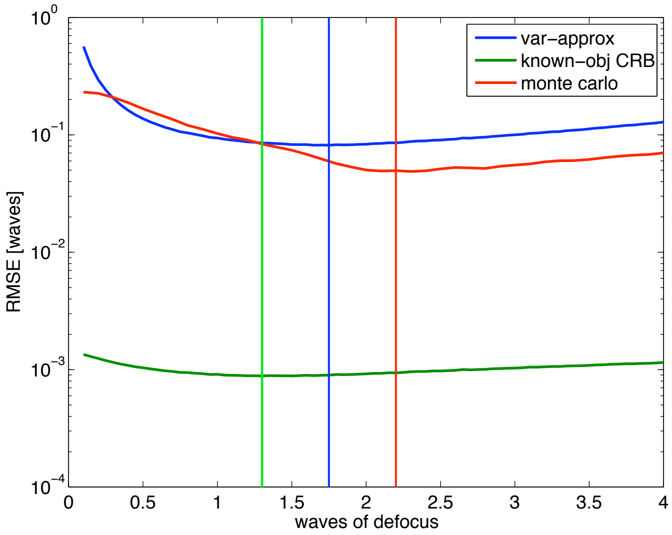


Figure 4.8: Comparison of the average of the estimate standard deviations calculated via both (4.43) and (4.30), as well as the RMSE of the Monte Carlo simulation, for the extended scene.

values for the two methods are very nearly equal; the minimum for the Cramèr-Rao bound occurs at 0.85 waves and the minimum for equation (4.43) is at 0.95 waves. The optimal defocus value for (4.15) occurs at 1.05 waves. Although the minima for the three methods are very near one another we note that the minimum given by (4.43) is closer to the minimum obtained through Monte Carlo simulations.

It is important to note that both the Cramèr-Rao bound and the variance approximation (4.43) are useful as *guides* in system parameter selection; one would be remiss to use either of these methods as a complete substitute for full simulation using (4.15). The primary benefit of (4.43) is that it can provide reasonable bounds on parameters of interest for far less computational expense than performing full Monte Carlo simulations. The ratio of computation time, for a single realization and a [64 × 64] object, of (4.15) to (4.43) is ≈ 12; a sizable time savings.

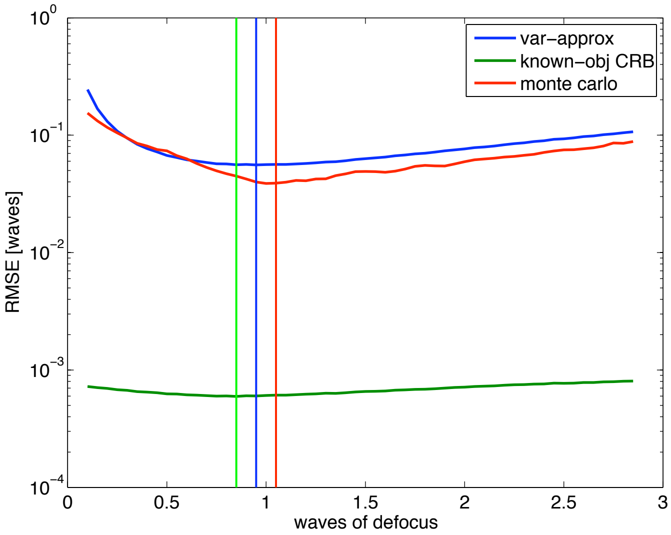


Figure 4.9: Comparison of the average of the estimate standard deviations calculated via both (4.43) and (4.30), as well as the RMSE of the Monte Carlo simulation, for the point object.

4.7 Conclusion and future work

A method for approximating the variance of aberration estimates for quadratically regularized weighted least squares estimators in the phase diversity context have been presented. The approximation has been shown to be a fairly accurate guide in selecting an appropriate amount of defocus for the diversity channel. The benefit of using (4.43) over (4.30) is delineated when the object is an extended scene; however the two methods do not differ substantially when the object is a point source.

The next steps in this research are to explore approximations to the bias of (4.15). An approximation to the bias could be used with (4.43) to approximate the entire mean squared error for aberration estimation. These expressions could also be used to explore how regularization of the object affects the aberration estimates. A natural next step is also to investigate why the known-object Cramèr-Rao bound differs from the approximation (4.43) less for point objects than for extended scenes.

CHAPTER V

Conclusion and future work

5.1 Summary

In this dissertation we have analyzed penalized-likehood estimation techniques for polarimetric imagery. We have explored the question of which space is the appropriate space to estimate in: intensity space or Stokes space. It was found that estimation of Stokes vectors directly provides estimates with lower overall RMS error as compared with restoring the intensity images and then transforming to Stokes space for interpretation. We have also explored how the addition of a cross-channel regularization term affects estimation accuracy. It was found that the addition of a cross-channel regularization term improves interpretability of Stokes parameter estimates when estimating the Stokes parameters directly and when using a traditional estimator.

While the addition of a cross-channel regularization term has improved interpretability for both estimators it has also added another set of regularization "tuning" parameters. For practical implementation the added computation cost of additional regularization penalty functions should be considered. The proposed Stokes space estimator has been shown to provide lower RMS reconstruction errors, however, the traditional estimator has only one tuning parameter to be adjusted as opposed to the Stokes estimator which has three. Further study into the computational cost verses reconstruction error between the proposed estimator and the traditional estimator is warranted for implementation decisions.

We have also developed a unified framework for joint estimation of Stokes images and aberrations from polarimetric measurements that contain phase diversity. We explored two methods, (3.18) and (3.22), for joint estimation of Stokes images and aberrations. It was found that estimation accuracy follows a task-based hierarchy, i.e., in a joint-estimation framework the choice of algorithm is task dependent. When the task is image restoration (aberrations are nuisance parameters) an algorithm that jointly estimates object and aberrations while incorporating *a priori* knowledge of the *object* is appropriate. However, if the aberration parameters are of interest and the object is a nuisance parameter then a reduced-parameter algorithm should be chosen.

The magnitudes of the regularization tuning parameters are very different depending on which parameters are of importance. When the aberration parameters are of interest and the object is nuisance the regularization parameter is quite small and the estimates are weakly dependent on the parameter over a wide range of values. In the limit as the regularization parameter approaches zero (3.22) becomes a maximum-likelihood estimator (that cannot be evaluated for circular apertures). It seems then, that the primary role of regularization in (3.22) is to ensure invertibility of the matrix in (3.21).

We have developed of a method for approximating the variance of aberration estimates for quadratically penalized weighted least squares estimators in the phase diversity context. Comparisons for system-parameter selection were made between the commonly used known-object Cramèr-Rao bound and our variance approximation that takes into account an unknown object. The impact including the uncertainty of the object in the formulation has been shown to be significant when selecting the diversity defocus parameter when imaging extended scenes. When imaging a point object the known-object Cramèr-Rao bound yields a minimum variance defocus value nearly equal to our approximation that incorporates the uncertainty in the object. From a practical point of view our approximation to the variance has significant utility; evaluation of the variance approximation (4.43) for 15 aberration parameters is 12 times faster than a single evaluation of (4.15).

5.2 Future Work

While this dissertation has focused primarily on polarimetric image reconstruction algorithms the framework is general and may be applied to other imaging modalities. The framework developed may be immediately applied to any imaging modality where there is a linear relationship between the estimation space and measurement space. An example of this is multi/hyper-spectral imaging. Spectral measurements are made using optical filters that pass a range of wavelengths. Real scenes under solar illumination contain information across a continuum of wavelengths, each spectral measurement is a sum of contributions from different wavelengths. In the polarimetric case our linear transformation $\mathbf{T}_{J\times 3}$ specified $J \geq 3$, but with spectral imaging we have $\mathbf{T}_{J\times K}$ where $K \geq J$. A model of this type highlights the possibility of estimating fine spectral components from measurements that are coarsely sampled in wavelength. This could be used, for example, for detection of materials with narrow spectral signatures. For example, a typical multispectral imager may have four spectral bands each with bandwidths of $\bar{\lambda}/5$ where $\bar{\lambda}$ is the mean wavelength of a particular band. In our proposed framework one could estimate spectral bands with bandwidths of $\bar{\lambda}/10$ thereby effectively increasing the spectral resolution of the imager. Consequently, there is the potential to detect finer spectral signatures. Moreover, improvement in detection of fine spectral signatures is a motivation for regularization design.

There are subtleties that have to be addressed when applying these methods to spectral imaging. The wavelength dependence of the point-spread-function as well as the sample rate at the detector should be modeled appropriately. Research into the impact of aliasing, due to wavelength variation, is of practical importance for spectral imager design. Also, regularization strategies for recovery of aliased information in the spectral imaging context has yet to be explored.

A more general extension of the Stokes estimation framework would be to consider estimation spaces that are nonlinearly related to the measurements. Some related work has been done already in the area spectral anomaly detection [51, 50, 76]. However, development of image reconstruction algorithms in this area has received little attention. Exploration of penalized-likelihood estimators that perform estimation in a space that is nonlinearly related to the measurements offers the potential for dimensionality reduction opens up new motivations for regularization penalty design.

All of the algorithms in this work have assumed a monochromatic object, in reality the world polychromatic. These algorithms can be generalized to accommodate polychromatic objects. There are two areas where the polychromatic nature of light will enter: (1) point-spread function, and (2) detector sampling. The PSF scales with wavelength to first order so that can be dealt with in a straightforward manner; the detector sampling also scales with wavelength but one must be mindful of aliasing effects at the blue end of the spectrum.

The polarimetric-phase-diverse wavefront sensing algorithm (4.15) can be explored further by switching paradigms to a Bayesian framework. In the Bayesian frame work the regularization penalty is viewed as a statistical prior on the object. In our framework the only requirement on our quadratic regularization penalty was that $\mathbf{C}'\mathbf{C}$ be positive definite. Using Parseval's theorem (3.22) can be written in the Fourier domain. In the Bayesian paradigm, in the Fourier domain, the regularization penalty now takes on the meaning of the inverse power-spectral density of the object. For extended scenes different object PSDs can be explored. For example, an image of Manhattan, New York, will have strong spatial frequencies corresponding to the gridlike street structure of the city, whereas farm land may not exhibit any preferential direction in the frequency domain.

It would be natural to extend the expression for the approximate joint covariance matrix to include other system parameters. For example, in this dissertation we have assumed that the measurements were all perfectly registered and sampled at the Nyquist rate, however, this is seldom the case for real phase-diversity systems. Generalization of (4.43) to include multiple system parameters would provide a means for studying the effects of channel misregistration, aliasing, or other system parameters on wavefront estimation. Similarly, in our polarimetric work we have assumed the polarimetric channels were perfectly aligned and sampled at the Nyquist rate. Again, a generalization of (4.43) to polarimetric-phase-diverse systems would provide a starting point for analyzing how system nonidealities effect wavefront estimation.

Analytical analysis of our covariance approximation under the consideration of a random $\boldsymbol{\alpha}$ is forthcoming. To complement this we plan to compare variations of our analytic predictions over an aberration ensemble with the variations of the Monte Carlo simulations over the same ensemble. Moreover, we plan to tailor the aberration ensemble to more accurately represent atmospheric turbulence using Kolmogorov statistics.

In our joint estimation of object and aberrations we have found that one can either have better object estimates *or* better aberration estimates but not both. We plan to thoroughly investigate joint estimation versus staged estimation. Intuitively it seems that the full joint model would give the best estimates for object and aberration but it is not clear that this is the case and it deserves further attention.

Further analysis of phase-diverse wavefront sensor performance would benefit from an expression for the approximate bias of the aberration estimates. This calculation involves a third order Taylor expansion of an implicit estimator[29] and so simplifying assumptions seem imperative. However, once armed with approximations for the covariance and bias of the aberrations estimates it would be of great utility in analyzing how regularization of the object *or* aberrations, affects the estimation of the aberration parameters. That is, one could analyze the effect of a statistical prior on the aberrations in addition to the object regularization.

As mentioned above, higher fidelity system models are practical importance. Approximations for the covariance and bias of (4.15) can be generalized to accommodate registration parameters, channel transmission, pixel aggregation, etc. That is, if the joint parameter to be estimated, $\mathbf{\Theta}$, consists of L types of parameters, $\mathbf{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L)$, that enter into the estimation problem both linearly and nonlinearly, one could analyze their joint impact on reconstruction quality using approximations to the bias and covariance.

In appendix B we demonstrated graphically, for a three-channel polarimeter, that the set of angles $\{\theta_i\} = \frac{\pi(i-1)}{J}$ for $i = 1, \ldots, J$ was an optimal in the Cramèr-Rao sense. We also conjectured that the same set of angles would be optimal for a J-channel polarimeter and showed that those angles are indeed an extreme point of the J-channel objective function. A path forward for proving that the set of angles is indeed a minimum variance set includes showing that the set of angels is a local minimum and not a maximum. This can be accomplished by showing that the Hessian evaluated at the prescribed set is positive definite. The next, more difficult step, is to show that the angles are a global minimum and not a local minimum. A second approach would be to explore a single-value decomposition of the FIM . The FIM has a large amount of structure and so may have a "nice" SVD. Once the SVD is in hand one can use the fact that the trace is similarity invariant and obtain the inverse without too much difficulty. From there one could work towards a closed-form expression for the minimizing set of angles.

APPENDICES

APPENDIX A

Numerical minimization of (3.22)

Numerical minimization of equation (3.22)

The minimization of (3.22) is most easily accomplished by invoking Parseval's theorem. If we ignore inconsequential constants our objective function can be written in the Fourier domain as

$$\Phi(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{j=1}^{J}\big\|\mathbf{Y}_j - \boldsymbol{\Omega}_j\Lambda_j\boldsymbol{\Theta}(\boldsymbol{\alpha})\big\|^2 + \frac{1}{2}\sum_{n=0}^{2}\beta_n\big\|\mathbf{D}\Theta_n(\boldsymbol{\alpha})\big\|^2, \tag{A.1}$$

where $\mathbf{Y}_j = \mathbf{Q}'\mathbf{y}_j$, $\Lambda_j = [\mathbf{T}_{J\times 3}]_{(j,1:3)} \otimes \mathbf{I}_{n_p}$, $\boldsymbol{\Theta}(\boldsymbol{\alpha}) = \mathbf{Q}'\mathbf{S}(\boldsymbol{\alpha})$, $\boldsymbol{\Omega}(\boldsymbol{\alpha}) = \mathbf{Q}'\mathbf{B}(\boldsymbol{\alpha})$, and $\mathbf{D}$ is a matrix such that $\mathbf{D}'\mathbf{D} = \mathbf{QCC}'\mathbf{Q}'$. The partial derivatives of (A.1) with respect to the ith aberration parameter can be written

$$\frac{\partial\Phi}{\partial\alpha_i} = -\frac{1}{2}\sum_{j=1}^{J}\left\langle \frac{\partial\boldsymbol{\Omega}_j}{\partial\alpha_i}\Lambda_j\boldsymbol{\Theta} + \boldsymbol{\Omega}_j\Lambda_j\frac{\partial\boldsymbol{\Theta}}{\partial\alpha_i}\ ,\ \mathbf{Y}_j - \boldsymbol{\Omega}_j\Lambda_j\boldsymbol{\Theta}\right\rangle + \frac{1}{2}\sum_{n=0}^{2}\beta_n\left\langle \mathbf{D}\frac{\partial\Theta_n}{\partial\alpha_i}\ ,\ \mathbf{D}\Theta_n\right\rangle + c.c,$$

where the dependencies on α_i have been suppressed for brevity, $\langle\cdot\ ,\ \cdot\rangle$ is the Euclidean inner product over $\mathbb{C}^{3n_p}$, and $c.c$ denotes complex conjugate. To evaluate this expression, the partial derivatives of $\boldsymbol{\Omega}_j$ and $\boldsymbol{\Theta}(\boldsymbol{\alpha})$ must be calculated.

Expression for $\frac{\partial \boldsymbol{\Theta}}{\partial \alpha_i}$

Recall, equation (3.21) expresses the Stokes parameters in terms of the aberration parameters:

$$\mathbf{S}(\boldsymbol{\alpha}) = \Bigg[\underbrace{\left(\mathbf{T}'_{J\times 3} \otimes \mathbf{I}_{n_p}\right) \mathbf{B}\left(\boldsymbol{\alpha}\right)' \mathbf{B}\left(\boldsymbol{\alpha}\right) \left(\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p}\right) + \sigma^2 \boldsymbol{\beta}_3 \otimes \mathbf{C}'\mathbf{C}}_{\mathbf{P}}\Bigg]^{-1} \times \underbrace{\left(\mathbf{T}'_{J\times 3} \otimes \mathbf{I}_{n_p}\right) \mathbf{B}\left(\boldsymbol{\alpha}\right)' \mathbf{y}}_{\mathbf{V}}. \tag{A.2}$$

This expression may be written as a matrix-vector product: $\mathbf{S}(\boldsymbol{\alpha}) = \mathbf{P}^{-1}(\boldsymbol{\alpha};\boldsymbol{\beta}_3)\mathbf{V}(\boldsymbol{\alpha};\mathbf{y})$, and a similar expression may be written in the Fourier domain

$$\boldsymbol{\Theta}(\boldsymbol{\alpha}) = \mathbf{A}^{-1}(\boldsymbol{\alpha};\boldsymbol{\beta}_3)\mathbf{X}(\boldsymbol{\alpha};\mathbf{Y}). \tag{A.3}$$

The partial derivative of (A.1) with respect to α_i is then

$$\frac{\partial \boldsymbol{\Theta}}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i}\left(\mathbf{A}^{-1}\mathbf{X}\right) = \frac{\partial \mathbf{A}^{-1}}{\partial \alpha_i}\mathbf{X} + \mathbf{A}^{-1}\frac{\partial \mathbf{X}}{\partial \alpha_i} \tag{A.4}$$

$$= -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \alpha_i}\mathbf{A}^{-1}\mathbf{X} + \mathbf{A}^{-1}\frac{\partial \mathbf{X}}{\partial \alpha_i} \tag{A.5}$$

$$= -\mathbf{A}^{-1}\left(2\sum_{j=1}^{J}[\mathbf{T}_{J\times 3}]'_j[\mathbf{T}_{J\times 3}]_j \otimes \operatorname{Re}\left\{\boldsymbol{\Omega}'_j\frac{\partial \boldsymbol{\Omega}_j}{\partial \alpha_i}\right\}\right)\mathbf{A}^{-1}\sum_{j=1}^{J}\Lambda'_j\boldsymbol{\Omega}'_j\mathbf{Y}_j + \mathbf{A}^{-1}\sum_{j=1}^{J}\Lambda'_j\frac{\partial \boldsymbol{\Omega}'_j}{\partial \alpha_i}\mathbf{Y}_j \quad . \tag{A.6}$$

With this expression in place it remains to write down an expression for $\mathbf{A}^{-1}$ and compute the partial derivatives of $\boldsymbol{\Omega}_j$ with respect to α_i.

Expression for $\frac{\partial \mathbf{\Omega}_j}{\partial \alpha_i}$

$$\frac{\partial}{\partial \alpha_i} \mathbf{\Omega}_j = \frac{\partial}{\partial \alpha_i} c \ \mathcal{F}\left[\left|\mathcal{F}^{-1}\left[P_j\right]\right|^2\right] \tag{A.7}$$

$$= c \ \mathcal{F}\left[\frac{\partial}{\partial \alpha_i}\left|\mathcal{F}^{-1}\left[P_j\right]\right|^2\right] \tag{A.8}$$

$$= 2c \ \mathcal{F}\left[\mathrm{Re}\left\{\left(\mathcal{F}^{-1}\left[\frac{\partial P_j}{\partial \alpha_i}\right]\right)\left(\mathcal{F}^{-1}\left[P_j\right]\right)^*\right\}\right] \tag{A.9}$$

$$= 2c \ \mathcal{F}\left[\mathrm{Re}\left\{\left(\mathcal{F}^{-1}\left[\imath \psi_i P_j\right]\right)\left(\mathcal{F}^{-1}\left[P_j\right]\right)^*\right\}\right] \tag{A.10}$$

$$= 2c \ \mathcal{F}\left[-\mathrm{Im}\left\{\left(\mathcal{F}^{-1}\left[\psi_i P_j\right]\right)\left(\mathcal{F}^{-1}\left[P_j\right]\right)^*\right\}\right] \tag{A.11}$$

$$= 2c \ \mathcal{F}\left[\mathrm{Im}\left\{\mathcal{F}^{-1}\left[P_j\right]\left(\mathcal{F}^{-1}\left[\psi_i P_j\right]\right)^*\right\}\right], \tag{A.12}$$

where c normalizes the PSF to unit volume.

Expression for $\mathbf{A}^{-1}$ for the case of polarization angles $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$

For a J channel polarimeter the matrix $\mathbf{A}$ has a (3×3) block structure and so can be inverted by hand. For our specific case, the expression for $\mathbf{A}^{-1}$ can be written as horizontally concatenated (3×1) block-column vectors in the following way:

$$\mathbf{A}^{-1} = \frac{1}{16\det[\mathbf{A}]} \begin{bmatrix} \left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right)\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right) \\ -\left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right) \\ -\left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right)\left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right) \end{bmatrix}$$

$$\begin{bmatrix} -\left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right) \\ \left(\sum_{j=1}^{4}|\boldsymbol{\Omega}_j|^2 + 4\beta_0|\mathbf{D}|^2\right)\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right) - \left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right)^2 \\ \left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)\left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right) \end{bmatrix}$$

$$\begin{bmatrix} -\left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right)\left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right) \\ \left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)\left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right) \\ \left(\sum_{j=1}^{4}|\boldsymbol{\Omega}_j|^2 + 4\beta_0|\mathbf{D}|^2\right)\left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right) - \left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)^2 \end{bmatrix},$$

where

$$\begin{aligned} 64\det[\mathbf{A}] = {} & \left(\sum_{j=1}^{4}|\boldsymbol{\Omega}_j|^2 + 4\beta_0|\mathbf{D}|^2\right)\left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right)\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right) \\ & - \left(|\boldsymbol{\Omega}_2|^2 - |\boldsymbol{\Omega}_4|^2\right)^2\left(|\boldsymbol{\Omega}_1|^2 + |\boldsymbol{\Omega}_3|^2 + 4\beta_1|\mathbf{D}|^2\right) \\ & - \left(|\boldsymbol{\Omega}_1|^2 - |\boldsymbol{\Omega}_3|^2\right)^2\left(|\boldsymbol{\Omega}_2|^2 + |\boldsymbol{\Omega}_4|^2 + 4\beta_2|\mathbf{D}|^2\right). \end{aligned}$$

With the component expressions for the partial derivatives for the reduced-parameter objective function in place gradient-based search methods may be used. It is worthy to note that search methods that do not use gradient information are futile for the objective function (3.22). We examined the standard derivative-free simplex method (the MATLAB *fminsearch* algorithm) and found that reaching the global minimum of (3.22) would take more than 3 days of computation whereas the gradient-based L-BFGS method requires less than 20 minutes to reach the global minimum.

APPENDIX B

Optimal angles for a three-channel linear polarimeter

Optimal angles for a three-channel polarimeter

In this appendix we use the Cramèr-Rao bound as a metric for determining the optimal polarization angles for a three-channel linear polarimeter; we also provide a conjecture for a set of optimal polarization angles for J-channel linear polarimeter. Consider the following model for a J-channel linear polarimeter

$$\mathbf{y} = (\mathbf{T}_{J\times 3} \otimes \mathbf{I}_{n_p})\mathbf{S} + \boldsymbol{\varepsilon}, \tag{B.1}$$

where,

$$\mathbf{T}_{J\times 3} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\cos(2\theta_1) & \frac{1}{2}\sin(2\theta_1) \\ \frac{1}{2} & \frac{1}{2}\cos(2\theta_2) & \frac{1}{2}\sin(2\theta_2) \\ \vdots & \vdots & \vdots \\ \frac{1}{2} & \frac{1}{2}\cos(2\theta_J) & \frac{1}{2}\sin(2\theta_J) \end{bmatrix} \tag{B.2}$$

This model ignores the effects of optical blur as polarization is independent of phase in the visible regime. Under the noise model $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I}_{n_p})$ we may write down the following log-likelihood function

$$\mathrm{L}(\mathbf{S};\boldsymbol{\theta}) = \frac{1}{2}\left\|\mathbf{y} - (\mathbf{T}'_{J\times 3}(\boldsymbol{\theta})\mathbf{T}_{J\times 3}(\boldsymbol{\theta}) \otimes \mathbf{I}_{n_p})\mathbf{S}\right\|^2, \tag{B.3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$. The Fisher-Information Matrix (FIM) for (B.3) is given by

$$\mathrm{F} = \mathbf{T}'_{J\times 3}(\boldsymbol{\theta})\mathbf{T}_{J\times 3}(\boldsymbol{\theta}) \otimes \mathbf{I}_{n_p}. \tag{B.4}$$

At this point we may consider the FIM as a function of $\boldsymbol{\theta}$. The variances in the estimates of the Stokes parameters are then a function of the polarizer angles and are equal to the diagonal elements of the inverse FIM :

$$\mathrm{F}^{-1} = \left[\mathbf{T}'_{J\times 3}(\boldsymbol{\theta})\mathbf{T}_{J\times 3}(\boldsymbol{\theta})\right]^{-1} \otimes \mathbf{I}_{n_p}, \tag{B.5}$$

where

$$\mathbf{T}'_{J\times 3}(\boldsymbol{\theta})\mathbf{T}_{J\times 3}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{\imath=1}^{J} \frac{1}{4} & \sum_{\imath=1}^{J} \frac{1}{4}\cos(2\theta_\imath) & \sum_{\imath=1}^{J} \frac{1}{4}\sin(2\theta_\imath) \\ \sum_{\imath=1}^{J} \frac{1}{4}\cos(2\theta_\imath) & \sum_{\imath=1}^{J} \frac{1}{4}\cos^2(2\theta_\imath) & \sum_{\imath=1}^{J} \frac{1}{8}\sin(4\theta_\imath) \\ \sum_{\imath=1}^{J} \frac{1}{4}\sin(2\theta_\imath) & \sum_{\imath=1}^{J} \frac{1}{8}\sin(4\theta_\imath) & \sum_{\imath=1}^{J} \frac{1}{4}\sin^2(2\theta_\imath) \end{bmatrix}. \tag{B.6}$$

The inverse of any invertible 3 × 3 matrix F is written

$$\mathrm{F}^{-1} = \frac{1}{\det(\mathrm{F})}\begin{bmatrix} \mathrm{F}_{22}\mathrm{F}_{33} - \mathrm{F}_{23}^2 & \mathrm{F}_{13}\mathrm{F}_{32} - \mathrm{F}_{12}\mathrm{F}_{33} & \mathrm{F}_{12}\mathrm{F}_{23} - \mathrm{F}_{13}\mathrm{F}_{22} \\ \mathrm{F}_{23}\mathrm{F}_{31} - \mathrm{F}_{21}\mathrm{F}_{33} & \mathrm{F}_{11}\mathrm{F}_{33} - \mathrm{F}_{13}^2 & \mathrm{F}_{13}\mathrm{F}_{21} - \mathrm{F}_{11}\mathrm{F}_{23} \\ \mathrm{F}_{21}\mathrm{F}_{32} - \mathrm{F}_{22}\mathrm{F}_{31} & \mathrm{F}_{12}\mathrm{F}_{31} - \mathrm{F}_{11}\mathrm{F}_{32} & \mathrm{F}_{11}\mathrm{F}_{22} - \mathrm{F}_{12}^2 \end{bmatrix}. \tag{B.7}$$

We define our performance metric to be the trace of the FIM :

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; = \operatorname{Tr}[\mathrm{F}^{-1}(\boldsymbol{\theta})] \overset{\triangle}{=} \boldsymbol{\Psi}(\boldsymbol{\theta}). \tag{B.8}$$

Without loss of generality we can set $\theta_1 = 0$. Now, specializing to three channels we have $\boldsymbol{\theta} = (\theta_2, \theta_3)$. The objective function $\boldsymbol{\Psi}(\theta_2, \theta_3)$ becomes

$$\begin{aligned}\boldsymbol{\Psi}(\theta_2,\theta_3) = -\frac{1}{8}&\frac{4\cos(2\theta_2)+\cos(4\theta_2)+4\cos(2(\theta_2-\theta_3))+\cos(4(\theta_2-\theta_3))}{\sin^2(\theta_2)\sin^2(\theta_2-\theta_3)\sin^2(\theta_3)} \\ &+\frac{4\cos(2\theta_3)+\cos(4\theta_3)-15}{\sin^2(\theta_2)\sin^2(\theta_2-\theta_3)\sin^2(\theta_3)}\end{aligned} \tag{B.9}$$

Figures B.1 and B.2 show density plots of $\log[\boldsymbol{\Psi}(\theta_2, \theta_3)]$ $\{\theta_2, \theta_3\} \in (0, \pi)$; figure B.2 has been thresholded at log(25) for better color representation. Observe that there are two convex regions separated by the line $\theta_2 = \theta_3$. If we demand that the angles are ordered in a monotonically increasing fashion, the valid solution space lies above the line $\theta_2 = \theta_3$. We solve (B.8) numerically using a constrained Nelder-Meade numerical optimization routine. The results of the search are

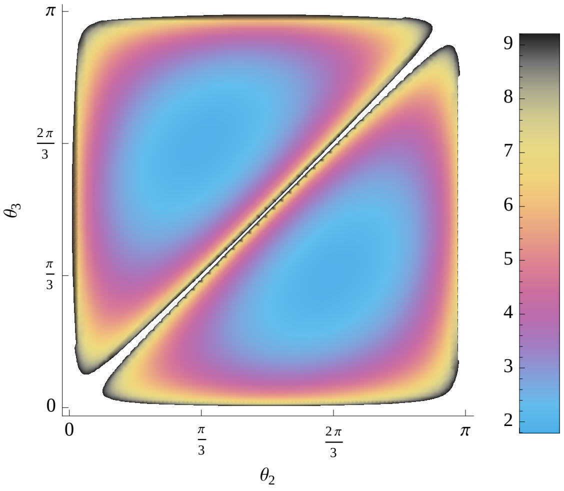


Figure B.1: Density plot of $\log(\boldsymbol{\Psi})$.

$$\theta_2 \rightarrow 1.0471975511966241732$$

$$\theta_3 \rightarrow 2.0943951023933713478.$$

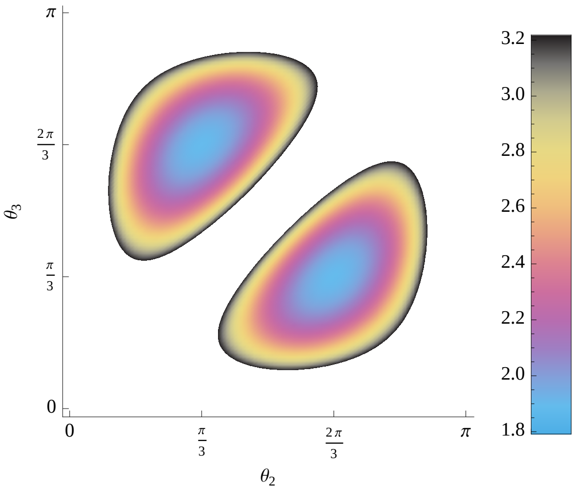


Figure B.2: Density plot of $\log(\mathbf{\Psi})$ thresholded for enhanced color representation.

These numbers are suspiciously close to $\pi/3$ and $2\pi/3$ respectively. Although (B.9) does not appear to lend itself to an analytic minimum, we can evaluate the gradient of our objective function at the intuitive solution of $(\pi/3, 2\pi/3)$. If the gradient vanishes we conjecture that $(\pi/3, 2\pi/3)$ is the global minimum. The partial derivatives of (B.9) are

$$\begin{aligned}\frac{\partial \mathbf{\Psi}}{\partial \theta_2} &= \frac{\sin(2\theta_2 - 5\theta_3) + \sin(4\theta_2 - 5\theta_3) + 4\sin(2\theta_2 - 3\theta_3) + 3\sin(4\theta_2 - 3\theta_3)}{8\sin^3(\theta_2)\sin^3(\theta_2 - \theta_3)\sin^2(\theta_3)} \\ &+ \frac{-26\sin(2\theta_2 - \theta_3) + 3\sin(4\theta_2 - \theta_3) + 4\sin(2\theta_2 + \theta_3)}{8\sin^3(\theta_2)\sin^3(\theta_2 - \theta_3)\sin^2(\theta_3)} \\ &+ \frac{\sin(4\theta_2 + \theta_3) + \sin(2\theta_2 + 3\theta_3)}{8\sin^3(\theta_2)\sin^3(\theta_2 - \theta_3)\sin^2(\theta_3)}\end{aligned} \tag{B.10}$$

$$\begin{aligned}\frac{\partial \mathbf{\Psi}}{\partial \theta_3} &= \frac{4\cos(\theta_2)\sin(2(\theta_2 - 2\theta_3))\cot^2(\theta_2)}{4\sin^3(\theta_2 - \theta_3)\sin^3(\theta_3)} \\ &+ \frac{(4\cos(2\theta_2)\cos(4\theta_2) - 13)\csc^2(\theta_2)\sin(\theta_2 - 2\theta_3)}{4\sin^3(\theta_2 - \theta_3)\sin^3(\theta_3)}.\end{aligned} \tag{B.11}$$

Evaluation the partial derivatives at $(\pi/3, 2\pi/3)$ verifies that indeed the gradient of the objective vanishes.

Conjecture

We conjecture that a set of minimum-variance angles, in the Cramè-Rao sense, for a J-channel linear polarimeter is $\{\theta_l\} = \{\pi \frac{l-1}{J}\}$ for $l = 1, \ldots, J$. We show next that this set of angles is an extreme point of $\boldsymbol{\Psi}(\boldsymbol{\theta})$. The diagonal elements of the J-channel inverse FIM can be written as a ratio of functions, $f(\boldsymbol{\theta})/g(\boldsymbol{\theta})$:

$$[\mathrm{F}^{-1}]_{(1,1)} = \frac{f_1(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}, \quad [\mathrm{F}^{-1}]_{(2,2)} = \frac{f_2(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}, \quad [\mathrm{F}^{-1}]_{(3,3)} = \frac{f_3(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}, \tag{B.12}$$

where the functions have been defined to be consistent with (B.7). Consider the derivative of our metric with respect to the lth component of $\boldsymbol{\theta}$:

$$\frac{\partial}{\partial \theta_l} \mathrm{Tr}[\mathrm{F}^{-1}] = \frac{\partial}{\partial \theta_l} \frac{f_1(\boldsymbol{\theta}) + f_2(\boldsymbol{\theta}) + f_3(\boldsymbol{\theta})}{g(\boldsymbol{\theta})} \tag{B.13}$$

The general form of the first derivative of a ratio of functions is given by the quotient rule

$$\left(\frac{f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right)' = \frac{f'(\boldsymbol{\theta})g(\boldsymbol{\theta}) - g'(\boldsymbol{\theta})f(\boldsymbol{\theta})}{g(\boldsymbol{\theta})^2}, \tag{B.14}$$

this function vanishes if $f'(\boldsymbol{\theta})$ and $g'(\boldsymbol{\theta})$ are both zero ($g(\boldsymbol{\theta})$ is nonzero by assumption).

We will differentiate each term in the numerator and show that the resulting sum is zero for the conjectured set of angles, we will then differentiate the denominator and show that it vanishes for the same set of angles.

$$\frac{\partial f_1}{\partial \theta_l} = \frac{\partial}{\partial \theta_l}\left(\mathrm{F}_{22}\mathrm{F}_{33} - \mathrm{F}_{23}^2)\right) \tag{B.15}$$

$$= \frac{\partial}{\partial \theta_l}\left(\sum_{\imath=1}^{J} \frac{1}{4}\cos^2(2\theta_\imath)\right)\left(\sum_{\jmath=1}^{J} \frac{1}{4}\sin^2(2\theta_\jmath)\right) - \frac{\partial}{\partial \theta_l}\left(\sum_{k=1}^{J} \frac{1}{8}\sin(4\theta_k)\right)^2 \tag{B.16}$$

$$= \begin{array}{l} (\cos(2\theta_l)\sin(2\theta_l))\left(\left(\sum_{\imath=1}^{J} \frac{1}{4}\cos^2(2\theta_\imath)\right) - \left(\sum_{\jmath=1}^{J} \frac{1}{4}\sin^2(2\theta_\jmath)\right)\right) \\ -\ \cos(4\theta_l)\sum_{k=1}^{J} \frac{1}{8}\sin(4\theta_k) \end{array}. \tag{B.17}$$

We evaluate (B.17) by invoking the following trigonometric identities

$$\sum_{\imath=1}^{J} \cos^2(L\pi\frac{\imath-1}{J}) = \sum_{\imath=1}^{J} \sin^2(L\pi\frac{\imath-1}{J}) = \frac{J}{2}, \quad L = \text{even} \tag{B.18}$$

$$\sum_{k=1}^{J} \cos(L\pi\frac{k-1}{J}) = \sum_{k=1}^{J} \sin(L\pi\frac{k-1}{J}) = 0, \quad L = \text{even}, \tag{B.19}$$

using these in (B.17) yields

$$\frac{\partial f_1}{\partial \theta_l} = (1/4)\cos(2\pi\frac{l-1}{J})\sin(2\pi\frac{l-1}{J})(\frac{J}{2} - \frac{J}{2}) - (1/8)\cos(4\pi\frac{l-1}{J})(0) = 0. \tag{B.20}$$

Continuing with the next two terms we have

$$\frac{\partial f_2}{\partial \theta_l} = \frac{\partial}{\partial \theta_l}\left(\mathrm{F}_{11}\mathrm{F}_{33} - \mathrm{F}_{13}^2)\right) \tag{B.21}$$

$$= \frac{J}{4}\frac{\partial}{\partial \theta_l}\sum_{\jmath=1}^{J} \frac{1}{4}\sin^2(2\theta_\jmath) - \frac{\partial}{\partial \theta_l}\left(\sum_{k=1}^{J} \frac{1}{4}\sin(2\theta_k)\right)^2 \tag{B.22}$$

$$= \frac{J}{4}\cos(2\theta_l)\sin(2\theta_l) - \frac{1}{4}\cos(2\theta_l)\sum_{k=1}^{J} \frac{1}{4}\sin(2\theta_k), \tag{B.23}$$

evaluation over the prescribed set yields

$$\frac{\partial f_2}{\partial \theta_l} = \frac{J}{4} \cos\left(2\pi \frac{l-1}{J}\right) \sin\left(2\pi \frac{l-1}{J}\right). \tag{B.24}$$

The third term is

$$\begin{align}
\frac{\partial f_3}{\partial \theta_l} &= \frac{\partial}{\partial \theta_l} \left(\mathrm{F}_{11}\mathrm{F}_{22} - \mathrm{F}_{12}^2)\right) \tag{B.25}\\
&= \frac{J}{4} \frac{\partial}{\partial \theta_l} \sum_{\imath=1}^{J} \frac{1}{4} \cos^2(2\theta_\imath) - \frac{\partial}{\partial \theta_l} \left(\sum_{k=1}^{J} \frac{1}{4} \cos(2\theta_k) \right)^2 \tag{B.26}\\
&= -\frac{J}{4} \cos(2\theta_l) \sin(2\theta_l) - \frac{1}{4} \sin(2\theta_l) \sum_{k=1}^{J} \frac{1}{4} \sin(2\theta_k), \tag{B.27}
\end{align}$$

evaluation over the prescribed set yields

$$\frac{\partial f_3}{\partial \theta_l} = -\frac{J}{4} \cos\left(2\pi \frac{l-1}{J}\right) \sin\left(2\pi \frac{l-1}{J}\right). \tag{B.28}$$

Putting the three terms together we have

$$\left.\frac{\partial (f_1 + f_2 + f_3)}{\partial \theta_l}\right|_{\theta_l = \pi(l-1)/J} = 0, \tag{B.29}$$

since this expression is independent of l, it holds for all components of $\boldsymbol{\theta}$. It remains to show that $\frac{\partial}{\partial \theta l} g(\boldsymbol{\theta})\big|_{\theta_l} = 0,\ l = 1, \ldots, J.$

Recall, $g(\boldsymbol{\theta}) = \det(\mathrm{F})$. We will use the following identity in our work:

$$\frac{\partial \det(\mathrm{F})}{\partial \theta} = \det(\mathrm{F}) \operatorname{Tr}\left(\mathrm{F}^{-1} \frac{\partial \mathrm{F}}{\partial \theta}\right). \tag{B.30}$$

Equation (B.30) vanishes if $\operatorname{Tr}\left(\mathrm{F}^{-1} \frac{\partial \mathrm{F}}{\partial \theta_l}\right) = 0$. We now write out the elements of the trace

$$\begin{aligned}\left[\mathrm{F}^{-1}\frac{\partial \mathrm{F}}{\partial\theta_l}\right]_{11} &= \left(\mathrm{F}_{22}\mathrm{F}_{33} - \mathrm{F}_{23}^2\right)\frac{\partial \mathrm{F}_{11}}{\partial\theta_l} \\ &+ \left(\mathrm{F}_{13}\mathrm{F}_{32} - \mathrm{F}_{12}\mathrm{F}_{33}\right)\frac{\partial \mathrm{F}_{21}}{\partial\theta_l}, \\ &+ \left(\mathrm{F}_{12}\mathrm{F}_{23} - \mathrm{F}_{13}\mathrm{F}_{22}\right)\frac{\partial \mathrm{F}_{31}}{\partial\theta_l}\end{aligned} \tag{B.31}$$

$$\begin{aligned}\left[\mathrm{F}^{-1}\frac{\partial \mathrm{F}}{\partial\theta_l}\right]_{22} &= \left(\mathrm{F}_{23}\mathrm{F}_{31} - \mathrm{F}_{21}\mathrm{F}_{33}\right)\frac{\partial \mathrm{F}_{12}}{\partial\theta_l} \\ &+ \left(\mathrm{F}_{11}\mathrm{F}_{33} - \mathrm{F}_{13}^2\right)\frac{\partial \mathrm{F}_{22}}{\partial\theta_l} \quad , \\ &+ \left(\mathrm{F}_{13}\mathrm{F}_{21} - \mathrm{F}_{11}\mathrm{F}_{23}\right)\frac{\partial \mathrm{F}_{32}}{\partial\theta_l}\end{aligned} \tag{B.32}$$

$$\begin{aligned}\left[\mathrm{F}^{-1}\frac{\partial \mathrm{F}}{\partial\theta_l}\right]_{33} &= \left(\mathrm{F}_{21}\mathrm{F}_{32} - \mathrm{F}_{22}\mathrm{F}_{31}\right)\frac{\partial \mathrm{F}_{13}}{\partial\theta_l} \\ &+ \left(\mathrm{F}_{12}\mathrm{F}_{31} - \mathrm{F}_{11}\mathrm{F}_{32}\right)\frac{\partial \mathrm{F}_{23}}{\partial\theta_l}. \\ &+ \left(\mathrm{F}_{11}\mathrm{F}_{22} - \mathrm{F}_{12}^2\right)\frac{\partial \mathrm{F}_{33}}{\partial\theta_l}\end{aligned} \tag{B.33}$$

Each of these expressions must be evaluated at $\theta_l = \pi\frac{l-1}{J}$. Consider (B.31); the first term is zero due to F_{11} being independent of angle, the second and third terms also vanish since $\mathrm{F}_{12}\big|_{\theta_l=\pi\frac{l-1}{J}} = \mathrm{F}_{13}\big|_{\theta_l=\pi\frac{l-1}{J}} = 0$. Equation (B.32) has two vanishing terms since $\mathrm{F}_{23}\big|_{\theta_l=\pi\frac{l-1}{J}} = 0$ as well. Lastly, using the previous identities, equation (B.32) also has two vanishing terms. There are two nonzero terms left in the trace:

$$\mathrm{Tr}\left(\mathrm{F}^{-1}\frac{\partial \mathrm{F}}{\partial\theta_l}\right)\Bigg|_{\theta_l=\pi\frac{l-1}{J}} = \left[\left(\mathrm{F}_{11}\mathrm{F}_{33} - \mathrm{F}_{13}^2\right)\frac{\partial \mathrm{F}_{22}}{\partial\theta_l} + \left(\mathrm{F}_{11}\mathrm{F}_{22} - \mathrm{F}_{12}^2\right)\frac{\partial \mathrm{F}_{33}}{\partial\theta_l}\right]\Bigg|_{\theta_l=\pi\frac{l-1}{J}}. \tag{B.34}$$

Observe that $\frac{\partial \mathrm{F}_{22}}{\partial\theta_l} = -\frac{\partial \mathrm{F}_{33}}{\partial\theta_l}$, $\mathrm{F}_{12}\big|_{\theta_l=\pi\frac{l-1}{J}} = \mathrm{F}_{12}\big|_{\theta_l=\pi\frac{l-1}{J}} = 0$, and that $\mathrm{F}_{22}\big|_{\theta_l=\pi\frac{l-1}{J}} =$

$\mathrm{F}_{33}\big|_{\theta_l=\pi\frac{l-1}{J}} = \frac{J}{8}$. Putting these together we have

$$\mathrm{Tr}\left(\mathrm{F}^{-1}\frac{\partial \mathrm{F}}{\partial \theta_l}\right)\bigg|_{\theta_l=\pi\frac{l-1}{J}} = \frac{J^2}{16} - \frac{J^2}{16} = 0. \tag{B.35}$$

It has been demonstrated that the set polarization angles $\{\theta_l\} = \{\pi\frac{l-1}{J}\}$ $l = 1, \ldots, J$ is an extreme point of $\mathbf{\Psi}$. Further research is needed to prove that the set of angles $\{\theta_l\} = \{\pi\frac{l-1}{J}\}$ $l = 1, \ldots, J$ is a global minimum of $\mathbf{\Psi}$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Ahn and R.M. Leahy. Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET. *IEEE Transactions on Medical Imaging*, 27(3):413, 2008.

[2] S. Ainouz, J. Zallat, A. de Martino, and C. Collet. Physical interpretation of polarization-encoded images by color preview. *Optics express*, 14:5916–5927, 2006.

[3] A. Ambirajan and D.C. Look Jr. Optimum angles for a polarimeter: part I (Journal Paper). *Optical Engineering*, 34(06):1651–1655, 1995.

[4] A. Ambirajan and D.C. Look Jr. Optimum angles for a polarimeter: part I (Journal Paper). *Optical Engineering*, 34(06):1656–1658, 1995.

[5] I.B. Ayed, A. Mitiche, and Z. Belhadj. Polarimetric image segmentation via maximum-likelihood approximation and efficient multiphase level-sets. *IEEE transactions on pattern analysis and machine intelligence*, 28(9):1493, 2006.

[6] L. Bar, A. Brook, N. Sochen, and N. Kiryati. Deblurring of color images corrupted by impulsive noise. *IEEE Transactions on Image Processing*, 16(4):1101, 2007.

[7] H.H. Barrett, K.J. Myers, and S. Dhurjaty. Foundations of image science. *Journal of Electronic Imaging*, 14:029901, 2005.

[8] J.D. Barter, H.R. Thompson, and C.L. Richardson. Visible-regime polarimetric imager: a fully polarimetric, real-time imaging system. *Appl. Opt*, 42:1620–1628, 2003.

[9] I. Berezhnyy and A. Dogariu. Polarimetric description of superposing random electromagnetic fields with different spectral composition. *Journal of the Optical Society of America A*, 21(2):218–222, 2004.

[10] A. Blanc, L.M. Mugnier, and J. Idier. Marginal estimation of aberrations and image restoration by use of phase diversity. *Journal of the Optical Society of America A*, 20(6):1035–1045, 2003.

[11] P. Blomgren and T.F. Chan. Color TV: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[12] M.R. Bolcar and J.R. Fienup. Method of phase diversity in multi-aperture systems utilizing individual sub-aperture control. In *Proc. SPIE*, volume 5896, pages 126–133, 2005.

[13] M.R. Bolcar and J.R. Fienup. Sub-aperture piston phase diversity for segmented and multi-aperture systems. *Applied Optics*, 48(1):5–12, 2009.

[14] T.F. Chan and C.K.T. Wong. Multichannel image deconvolution by total variation regularization. In *Proceedings of SPIE*, volume 3162, page 358, 1997.

[15] R.A. Chipman. *"Polarimetry" in Handbook of Optics*, volume 2. McGraw-Hill, 2 edition, 1995.

[16] G. Cincotti. Polarization gratings: design and applications. *IEEE Journal of Quantum Electronics*, 39(12):1645–1652, 2003.

[17] F. Cremer, W. De Jong, and K. Schutte. Processing of polarimetric infrared images for landmine detection. In *2nd International Workshop on Advanced Ground Penetrating Radar (IWAGPR*. Citeseer, 2003.

[18] B.H. Dean, D.L. Aronstein, J.S. Smith, R. Shiri, and D.S. Acton. Phase retrieval algorithm for JWST flight and testbed telescope. In *Proceedings of SPIE*, volume 6265, page 626511, 2006.

[19] B.H. Dean and C.W. Bowers. Diversity selection for phase-diverse phase retrieval. *Journal of the Optical Society of America A*, 20(8):1490–1504, 2003.

[20] J.J. Dolne and H.B. Schall. Information theoretic bounds of phase diversity for diversity polynomials and noise statistics. In *Proceedings of SPIE*, volume 5553, page 127, 2004.

[21] J.J. Dolne and H.B. Schall. Cramer-Rao bound and phase-diversity blind deconvolution performance versus diversity polynomials. *Applied optics*, 44(29):6220–6227, 2005.

[22] J.J. Dolne and H.B. Schall. Fundamental performance bounds of phase diversity blind deconvolution algorithm for various diversity polynomials, noise statistics, and scene size. In *Atmospheric Propagation II. Edited by Young, Cynthia Y.; Gilbreath, G. Charmaine. Proceedings of the SPIE*, volume 5793, pages 118–128, 2005.

[23] M.J. Duggin. Factors controlling discrimination in imaging polarimetry. In *Proceedings of SPIE*, volume 5432, page 43. Citeseer, 2004.

[24] M.J. Duggin. Factors controlling the manual and automated extraction of image information using imaging polarimetry. In *Proceedings of SPIE*, volume 5432, page 85. Citeseer, 2004.

[25] M.J. Duggin and R. Loe. Algorithms for target discrimination and contrast enhancement using narrowband polarimetric image data. In *Proc. SPIE*, volume 4481. Citeseer, 2002.

[26] W.G. Egan. Polarization and surface roughness. In *Proceedings of SPIE*, volume 3426, page 144, 1998.

[27] YC Eldar. Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation. *IEEE Transactions on Signal Processing*, 54(8):2943–2956, 2006.

[28] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Kluwer Academic Pub, 1996.

[29] J.A. Fessler. Mean and variance of implicitly defined biased estimators (such aspenalized maximum likelihood): applications to tomography. *IEEE Transactions on Image Processing*, 5(3):493–506, 1996.

[30] J.A. Fessler. *Image Reconstruction: Algorithms and Analysis.* TBD, TBD.

[31] J.A. Fessler and W.L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs. *IEEE Transactions on Image processing*, 5(9):1346–1358, 1996.

[32] J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin. Hubble space telescope characterized by using phase-retrieval algoritihms. *Applied Optics*, 32:1747–1768, 1993.

[33] JR Fienup, JC Marron, TJ Schulz, and JH Seldin. Hubble Space Telescope characterized by using phase-retrieval algorithms. *Applied Optics*, 32(10):1747–1767, 1993.

[34] B. Formwalt and S. Cain. Optimized phase screen modeling for optical turbulence. *Appl. Opt*, 45:5657–5668, 2006.

[35] D.L. Fried. Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements. *J. Opt. Soc. Am*, 67:370–375, 1977.

[36] V.L. Gamiz and J.F. Belsher. Performance limitations of a four-channel polarimeter in the presence of detection noise. *Optical Engineering*, 41:973, 2002.

[37] G.C. Giakos. Multifusion, Multispectral, Optical Polarimetric Imaging Sensing Principles. *IEEE transactions on Instrumentation and Measurement*, 55(5):1628–1633, 2006.

[38] G.C. Giakos, L. Fraiwan, N. Patnekar, S. Sumrain, G.B. Mertzios, and S. Periyathamby. A sensitive optical polarimetric imaging technique for surface defects detection of aircraft turbine engines. *IEEE Transactions on instrumentation and measurement*, 53(1):216–222, 2004.

[39] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, 19:R1, 2003.

[40] J.D. Gonglewski, C.R. de Hainaut, C.M. Lampkin, and R.C. Dymale. System design of a wavefront sensing package for a wide field of view optical phased array. *Optical Engineering*, 27:785–792, 1988.

[41] R.A. Gonsalves. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, 21:829–832, 1982.

[42] R.A. Gonsalves and R. Chidlaw. Wavefront sensing by phase retrieval. In *Proceedings of the SPIE*, volume 207, pages 32–39, 1979.

[43] J.W. Goodman. *Introduction to Fourier optics 3rd edition*. Roberts & Company Publishers, 2005.

[44] E. Hecht. Note on an operational definition of the Stokes parameters. *American Journal of Physics*, 38:1156–1158, 1970.

[45] E. Hecht. *Hecht optics*. Addison-Wesley, 1998.

[46] AO Hero and MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB. A Cramer-Rao type lower bound for essentially unbiased parameter estimation. 1992.

[47] AO Hero III, JA Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform CRbound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, 1996.

[48] M. Houde, R.L. Akeson, J.E. Carlstrom, J.W. Lamb, D.A. Schleuning, and D.P. Woody. Polarizing grids, their assemblies, and beams of radiation. *Publications of the Astronomical Society of the Pacific*, 113(783):622–638, 2001.

[49] B. R. Hunt. The application of constrained least squares estimation to image restoration by digital computers. *IEEE Transactions on Computation*, 22(9):805–12, 1973.

[50] H. Kwon and NM Nasrabadi. Kernel orthogonal subspace projection for hyperspectral signal classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43(12):2952–2962, 2005.

[51] H. Kwon and N.M. Nasrabadi. Kernel matched subspace detectors for hyperspectral target detection. *IEEE transactions on pattern analysis and machine intelligence*, pages 178–194, 2006.

[52] L. Le Hors, P. Hartemann, D. Dolfi, and S. Breugnot. Phenomenological model of paints for multispectral polarimetric imaging. In *Proceedings of SPIE*, volume 4370, page 94, 2001.

[53] D.J. Lee, M.C. Roggemann, and B.M. Welsh. Cramer-Rao analysis of phase-diverse wave-front sensing. *Journal of the Optical Society of America A*, 16(5):1005–1015, 1999.

[54] L. Lee, G. Vasudevan, and E.H. Smith. Point-by-point approach to phase-diverse phase retrieval. In *Proceedings of SPIE*, volume 4850, page 441, 2003.

[55] D.A. LeMaster and S.C. Cain. Multichannel blind deconvolution of polarimetric imagery. *Journal of the Optical Society of America A*, 25(9):2170–2176, 2008.

[56] E.A. Lewis and J.P. Casey Jr. Electromagnetic reflection and transmission by gratings of resistive wires. *Journal of Applied Physics*, 23:605, 1952.

[57] D.C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

[58] S.M.F. Nee, C. Yoo, T. Cole, and D. Burge. Characterization for imperfect polarizers under imperfect conditions. *Applied optics*, 37(1), 1998.

[59] J.A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308, 1965.

[60] R.J. Noll. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3):207–211, 1976.

[61] R.J. Noll. Zernike polynomials and atmospheric turbulence. *J. Opt. Soc. Am*, 66(3):207–211, 1976.

[62] R. G. Paxman. Roles for phase diversity in compensated imaging. In *Adaptive Optics*, volume 13 of *Technical Digest Series*, pages 136–138, Maui, Hawaii, July 8-12 1996.

[63] R. G. Paxman and J. R. Fienup. Optical misalignment sensing and image reconstruction using phase diversity. 5:914–923, 1988.

[64] R. G. Paxman, J. R. Fienup, and S. L. Crippen. Aberration correction for phase-array telescopes using phase diversity. In A. F. Gmitro, P. S. Idell, and I. J. LaHaie, editors, *Digital Image Synthesis and Inverse Optics*, volume 1351 of *Proc. SPIE*, pages 787–797, San Diego, CA, July 1990.

[65] R. G. Paxman and J. H. Seldin. Fine-resolution astronomical imaging using phase-diverse speckle. In *Wavefront Supported Post-Facto Image Correction, Torben Anderson, ed.*, Workshop hosted by the Nordic Telescope Group, Risϕ, Roskilde, Denmark, November 1992.

[66] R. G. Paxman and J. H. Seldin. Fine-resolution imaging of solar features using phase-diverse speckle imaging. In R. R. Raddick, editor, *Real Time and Post-Facto Solar Image Correction*, volume 13 of *National Solar Observatory/Sacramento Peak Summer Workshop*, pages 112–118, Sunspot, NM, September 1992.

[67] R. G. Paxman and J.H. Seldin. Fine-resolution astronomical imaging with phase-diverse speckle. In *Digital Recovery and Synthesis II*, volume 2029, pages 287–298, San Diego, CA, 1993.

[68] RG Paxman and JR Fienup. Optical misalignment sensing and image reconstruction using phase diversity. *Journal of the Optical Society of America A*, 5(6):914–923, 1988.

[69] R.G. Paxman, T.J. Schulz, and J.R. Fienup. Joint estimation of object and aberrations by using phase diversity. *Journal of the Optical Society of America A*, 9(7):1072–1085, 1992.

[70] R.G. Paxman, J.H. Seldin, M.G. Loefdahl, G.B. Scharmer, and C.U. Keller. Evaluation of phase-diversity techniques for solar-image restoration. *The Astrophysical Journal*, 466:1087, 1996.

[71] R.G. Paxman, B.J. Thelen, R.J. Murphy, K.W. Gleichman, and J.A. Georges III. Phase-diverse adaptive optics for future telescopes. In *Proceedings of SPIE*, volume 6711, page 671103, 2007.

[72] R.G. Paxman, B.J. Thelen, and J.H. Seldin. Phase-diversity correction of turbulence-induced space-variant blur. *Optics letters*, 19(16):1231–1233, 1994.

[73] C.M. Persons, D.B. Chenault, M.W. Jones, K.D. Spradley, M.G. Gulley, and C.A. Farlow. Automated registration of polarimetric imagery using Fourier transform techniques. In *Proc. SPIE*, volume 4819, pages 107–117, 2002.

[74] J. Peterson, G. Jensen, and J. Kristl. Imaging polarimetry capabilities and measurement uncertainties in remote sensing applications. In *SPIE*, volume 4133, pages 221–232. Citeseer, 2000.

[75] N. Playle, M. Gaskell, R. Deas, and R. Rutherford. Feasibility of tripwire detection using infrared polarimetric sensor. In *Proceedings of SPIE*, volume 5415, page 745. Citeseer, 2004.

[76] KI Ranney and M. Soumekh. Hyperspectral anomaly detection within the signal subspace. *IEEE Geoscience and Remote Sensing Letters*, 3(3):312–316, 2006.

[77] R.B. Reid, M.E. Oxley, M.T. Eismann, and M.E. Goda. Quantifying surface normal estimation. In *Proceedings of SPIE*, volume 6240, page 624001, 2006.

[78] N.A. Roddier. Atmospheric wavefront simulation and Zernike polynomials. In *Proceedings of SPIE*, volume 1237, page 668, 1990.

[79] DS Sabatke, MR Descour, EL Dereniak, WC Sweatt, SA Kemme, and GS Phipps. Optimization of retardance for a complete Stokes polarimeter. *Optics Letters*, 25(11):802–804, 2000.

[80] D.S. Sabatke, A.M. Locke, E.L. Dereniak, and R.W. McMillan. Linear operator theory of channeled spectropolarimetry. *Journal of the Optical Society of America A*, 22(8):1567–1576, 2005.

[81] F.A. Sadjadi. Passive three-dimensional imaging using polarimetric diversity. *Optics letters*, 32(3):229–231, 2007.

[82] F.A Sadjadi and C.S.L Chun. Passive polarimetric IR target classification. *IEEE Transactions on Aerospace and Electronic Systems*, 37(2):740–751, 2001.

[83] F.A. Sadjadi and C.S.L. Chun. Automatic detection of small objects from their infrared state-of-polarization vectors. *Optics letters*, 28(7):531–533, 2003.

[84] I.K. Sendur, J.T. Johnson, and B.A. Baertlein. Analysis of polarimetric IR phenomena for detection of surface mines. In *Proceedings of SPIE*, volume 4394, page 153, 2001.

[85] B. J. Thelen, R. G. Paxman, D. A. Carrara, and J. H. Seldin. Map estimation of fixed aberrations, dynamic aberrations, and the object from phase-diverse speckle data. 1999.

[86] B. J. Thelen, R. G. Paxman, J. H. Seldin, and D. R. Rice. Bayes estimation of dynamic and fixed aberrations, and object from phase-diverse speckle data. In *Digital Image Recovery and Synthesis III*, volume 2827 of *Proc. SPIE*, pages 32–43, Denver, CO, August 1996.

[87] B.J. Thelen and R.G. Paxman. Parameter dimension of turbulence-induced phase errors and its effects on estimation in phase diversity. In *Proceedings of SPIE*, volume 2302, page 300, 1994.

[88] B.J. Thelen, R.G. Paxman, D.A. Carrara, and J.H. Seldin. Overcoming turbulence-induced space-variant blur by using phase-diverse speckle. *Journal of the Optical Society of America A*, 26(1):206–218, 2009.

[89] J.S. Tyo. Relation between system optimization and systematic errors in Stokes vector polarimeters. *Polarization Analysis, Measurement, and Remote Sensing*, 4481:22.

[90] J.S. Tyo. Optimum linear combination strategy for an N-channel polarization-sensitive imaging or vision system. *Journal of the Optical Society of America A*, 15(2):359–366, 1998.

[91] J.S. Tyo. Maximization of information content in polarimetric measurements. In *IEEE Antennas and Propagation Society International Symposium, 2000*, volume 2, 2000.

[92] J.S. Tyo. Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error. *Applied optics*, 41(4):619–630, 2002.

[93] J.S. Tyo, D.L. Goldstein, D.B. Chenault, and J.A. Shaw. Review of passive imaging polarimetry for remote sensing applications. *Appl. Opt*, 45:5453–5469, 2006.

[94] J.S. Tyo and H. Wei. Optimizing imaging polarimeters constructed with imperfect optics. *Applied optics*, 45(22):5497–5503, 2006.

[95] J.R. Valenzuela and J.A. Fessler. Regularized estimation of Stokes images from polarimetric measurements. In *Proceedings of SPIE*, volume 6814, page 681403, 2008.

[96] J.R. Valenzuela and J.A. Fessler. Joint estimation of stokes images and aberrations from phase diverse polarimetric measurements. In *Computational Optical Sensing and Imaging*. Optical Society of America, 2009.

[97] J.R. Valenzuela and J.A. Fessler. Joint reconstruction of Stokes images from polarimetric measurements. *Journal of the Optical Society of America A*, 26(4):962–968, April 2009.

[98] J.R. Valenzuela, J.A. Fessler, and R.G. Paxman. Joint estimation of stokes images and aberrations from phase-diverse polarimetric measurements. *Journal of the Optical Society of America A*, 27(5), May 2010.

[99] E.P. Wallner. Optimal wave-front correction using slope measurements. *SPIE MILESTONE SERIES MS*, 92:413–413, 1994.

[100] J. Zallat, C. Collet, and Y. Takakura. Clustering of polarization-encoded images. *Applied optics*, 43(2):283–292, 2004.

[101] J. Zallat et al. Optimal configurations for imaging polarimeters: impact of image noise and systematic errors. *Journal of Optics A: Pure and Applied Optics*, 8:807–814, 2006.

[102] J. Zallat, P. Graebling, and Y. Takakura. Using polarimetric imaging for material classification. In *Proceedings of IEEE International Conference on Image Processing*, pages 827–830.

[103] J. Zallat and C. Heinrich. Polarimetric data reduction: a Bayesian approach. *Optics express*, 15(1):83, 2007.

[104] Fuzhen Zhang. *Matrix Theory*. Springer, 1999.