POSITIVITY CONSTRAINT IMPLEMENTATIONS FOR MULTIFRAME BLIND DECONVOLUTION RECONSTRUCTIONS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

DECEMBER 2014

By

Paul A. Billings

Dissertation Committee:

Todd Reed, Chairperson
Kyungim Baek
Olga Boric-Lubecke
Yingfei Dong
Aaron Ohta
Gary Varner

Keywords: Parameterization, Inverse Problem, Deconvolution

ABSTRACT

Reconstruction of imagery degraded by atmospheric turbulence is inherently an ill-posed problem. Multiple solutions can be found which satisfy the measured data to the extent allowed by the noise statistics. The symmetric nature of a spatially invariant imaging system gives rise to an ambiguity between object and distortion. Furthermore, portions of the object spatial frequency information are often attenuated beyond usefulness. Research in this field is primarily concerned with resolving this ambiguity and making a quality estimate of the object in a timely fashion.

Many of the more advanced image reconstruction algorithms are iterative algorithms, seeking to minimize error/cost or maximize likelihood/conditional probability. The estimates, of course, are random variables, and regularization and constraints are often employed to guide solutions or reduce effects of overfitting. While positivity is an often employed constraint, one can conceive of various ways to achieve this beyond the typical approach of squaring the search parameters. We evaluate several functional parameterizations as well as the use of an optimizer employing search boundaries. We also consider the order of application in conjunction with a smoothness constraint. Performance is quantified by metrics, including RMS error (spatial domain), RMS phase error curves (spatial frequency domain), and timing to characterize computational demands.

We find the RMS phase error curves to be the most valuable as they assess reconstruction accuracy as a function of level of detail. Of the parameterizations evaluated, implementation of object positivity through squaring the parameters prior to application of a smoothness constraint achieved the best blend of accuracy and runtime execution speed.

Development of the theoretical background in a common framework contributed to reuse of constituent quantities, both in the development of the likelihood and many gradient expressions as well as implementation of those expressions into computer code. Such reuse promotes accuracy as well as ease of new development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Term/Symbol	Definition/Description
*	Denotes convolution in various equations
$\otimes$	Denotes correlation in various equations
A	Single parameter in the MLE discussion
AMOS	Air Force Maui Optical and Supercomputing observatory
B	Bias level (counts) of a detector
ASIC	Application-specific integrated circuit
$b(x)$	Blur kernel or smoothness constraint
c.c.	Complex conjugate
$\mathbf{d}$	Vector of all measured data
D	Aperture diameter of the imaging system
$D(\cdot)$	Distance from the origin (function)
$d_k(x)$	The k^{th} measured image
$\Delta\theta$	Angular sample spacing (subscripted to denote relevancy)
Δx	Linear sample spacing (subscripted to denote relevancy)
$\Phi_y(x)$	The y^{th} basis function
f	Focal length of the imaging system
$\Im(\cdot)$	Fourier transform
FPGA	Field programmable gate array
FWHM	Full width at half max
$g_k(x)$	The k^{th} noiseless (model) image
Γ	Covariance matrix
$H_k(x)$	The k^{th} pupil function (complex amplitude)

Term/Symbol	Definition/Description
$h_k(x)$	The k^{th} coherent point spread function
I	Identity matrix
$I(\cdot)$	Indicator function
$L(\cdot)$	Likelihood function
L-BFGS-B	An optimization algorithm with bounds constraint
λ	Penalty function weight Wavelength of illumination Mean rate in a Poisson PDF
μ	Mean (subscripted to denote specific context)
$M(\cdot)$	An indicator function with values 0 or 1
MAP	Maximum *a posteriori*
MFBD	Multi-frame blind deconvolution
ML (MLE)	Maximum likelihood (estimation)
$N\left(\mu_A, \sigma_A^2\right)$	Normal (Gaussian) distribution with specified mean and variance
N_A, N_o, N_p	Number of pixels across the image, object, and pupil arrays
$n_k(x)$	The noise distortion of the k^{th} image
$o(x)$	Object intensity values
O.T.	Other terms to be neglected
P	Peak value of the noiseless image
P(x)	Penalty function implementing a soft constraint
p(x)	Parameters describing the object
PDF	Probability density function
PSF	Point spread function
R	Pupil radius Range between object and imaging system pupil Cholesky factorization of a covariance matrix
r_0	Atmospheric seeing parameter
$r_k(x)$	The residual of the k^{th} frame

Term/Symbol	Definition/Description
RMS (RMSE)	Root mean square (error)
RMSPE	RMS phase error (spatial frequency domain)
RV	Random variable
σ_w^2	General noise variance
σ_d^2	Read noise of the detector
$s_k(x)$	The k^{th} point spread function
SNR	Signal to noise ratio
τ_0	Atmospheric decorrelation time
τ_0	Atmospheric decorrelation time
$\boldsymbol{\theta}$	Vector of all parameters
u	Two dimensional indexing variable, typically spatial frequency
W_p, W_o	Spatial width or extent of the pupil and object
WGN	White Gaussian noise
w.r.t.	With respect to
ω_0	Fundamental frequency for the discrete Fourier transform
x	Two dimensional indexing variable
y	Two dimensional indexing variable
ψ	Phase of the aberrated wavefront across the aperture
z	Two dimensional indexing variable

1 INTRODUCTION

1.1 Background

1.1.1 General Discussion

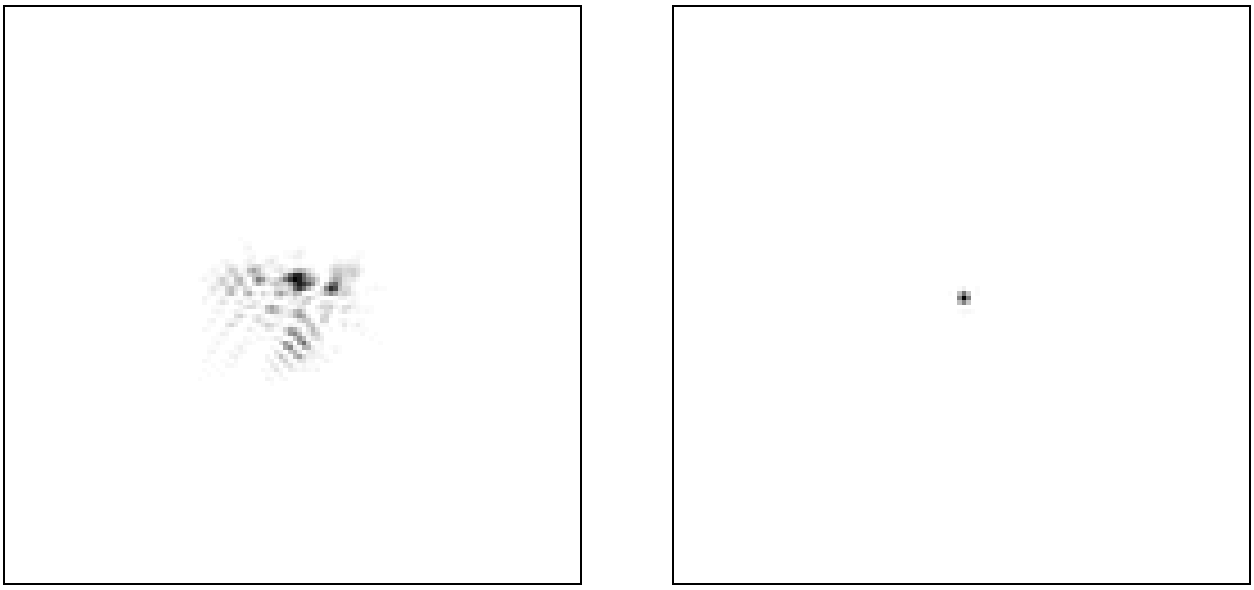
With any ground based incoherent imaging system, phase aberrations are often the limiting factor in achieving resolution.[1] These aberrations arise from a number of causes, with atmospheric turbulence being one of the largest contributors. As light propagates from the object to detector, minor variations in the index of refraction of air cause phase delays between various rays, potentially giving rise to a degraded image. Of primary interest is near-field turbulence, when the imaging system captures these rays very soon after they are distorted. This is very typical with ground-to-space imaging applications.

David Fried studied this issue in depth [2] and noted that there is spatial correlation to the temperature variations. He developed the concept of a characteristic size, termed the "atmospheric seeing parameter" and denoted by r_0, within which the index of refraction could be regarded as the "same". While varying with time of day, local terrain, and other factors, typical values of r_0 are 1 to 20 cm for visible wavelength illumination.

The aperture of an optical system gathers the light originating from the various points in the scene. As long as the aperture diameter is approximately less than r_0, all light rays captured in a given instant will have undergone about the same refraction by the atmosphere. Conveniently, little distortion will occur beyond possibly a uniform shift of the image. Unfortunately, the drive to examine dimmer objects and longer ranges requires the capture of more photons, which is achieved with increasingly larger apertures.

The drawback to larger apertures is that not all the rays that are captured have experienced the same refraction, having gone through different areas of air and turbulence. The rays from each pocket of air will experience a different phase delay which causes a local tilt of the wavefront, giving rise to a distorted image. The larger the aperture, the more pockets of air

are involved, and the more distorted the image becomes. Thus, the ratio of aperture to seeing size (D/ r_0) is a measure of how distorted an image will be. Example images of a single point source object are shown in Figure 1. (Brighter pixels are shown with a greater amount of black throughout this document.) By definition, these are the Point Spread Functions (PSFs) for those particular atmospheric realizations.



(a) (b)

Figure 1: Point source short exposure images through turbulence for (a) $D/ r_0 = 16$ and (b) $D/ r_0 = 1$

Furthermore, Fried noted a temporal correlation of the index of refraction, with the decorrelation time denoted by τ_0. Imaging sensors have finite exposure times. As long as the exposure time is approximately less than τ_0, the atmosphere is frozen and only spatial variations need to be considered. Long exposures from large apertures are extremely blurred due to the low-pass nature of this integration.

Several decades ago, Labeyrie noted that the spatial frequency content of short exposure imagery extended nearly out to the diffraction limit of the optics – well beyond that of long exposure images.[3] He pioneered the field of "speckle imaging", demonstrating that information was not lost, merely scrambled. The trick to unscrambling lay in the determination of the phase aberrations.

Measurement of these phase aberrations is possible (e.g., using a Shack-Hartmann wavefront sensor), and the extension of these ideas lead to the area of adaptive optics.[4,5] This is not a cure-all, however. Measurement errors, delays, imperfect correction, as well as greatly increased system complexity and cost often prohibit going down this road.

On the other hand, several methods exist to determine and correct these phase aberrations directly from the imagery itself using various transformations that are invariant to the effects of atmospheric turbulence [6], (e.g., Knox-Thompson [7] or triple-correlation/bispectrum [8]). One drawback to these methods is that they normally require knowledge of the average pupil amplitude function (commonly obtained from separate measurements of a point source with the assumption that the atmospheric characteristics are similar during the two times of observation).

Another technique to deal with the phase distortions is phase-diversity, which was first mentioned by Gonsalves [9,10] and generalized by Paxman *et al.* [11] to accommodate an arbitrary number of diversity measurements. Two (or more) images are recorded simultaneously, differing only by a known aberration (commonly defocus). Estimation theory is used to determine the most likely *common* object giving rise to the different images. This was later extended to phase-diverse speckle imaging by using multiple diversity images recorded at different times.[12] Each of these additional sets of images is made after the atmosphere has decorrelated.

Diversity is the key to these estimation approaches. As with most types of estimation, employing multiple uncorrelated measurements reduces the errors. In our case of incoherent imaging, there is a single common object that gives rise to each of these different images, which is a very powerful constraint. One of the drawbacks to phase-diverse speckle imaging is that the captured photons must be split among the various (simultaneous exposure) image planes. This reduces the signal to noise ratio of any given image, which can be a major problem for faint objects. The degenerate case is a “single-channel phase-diverse speckle” approach – in other words, it uses diversity in time only.

Others have derived this technique from a different direction, and we use their term in the remainder: *multi-frame blind deconvolution* (MFBD).[13-18] The phrase “multi-frame” is due to the use of multiple short exposure frames collected at different times through independent atmospheric realizations. To support the assumption of a common object, the scene must not

change appreciably throughout the image ensemble. "Deconvolution" is the process of estimating the object by observing the convolution of object and (known) point spread function (PSF). The PSF is said to be "deconvolved" from the measured image data, yielding the object. "Blind" is used to denote the fact that we do *not* know the point spread functions (i.e., the atmospheric distortions), and they are *jointly* estimated along with the object.

1.1.2 Multi-Frame Blind Deconvolution

We assume space-invariant incoherent imaging, regarding our images as the two-dimensional convolution of a single object and the various PSFs. Furthermore, we use the discrete representation to avoid the intractability of estimating analog waveforms. The k^{th} noiseless image in the ensemble is formed as:

$$g_k(x) = o(x) * s_k(x) = \sum_{x'} o(x') s_k(x - x') \qquad (1)$$

where $o(x)$ is the object, $s_k(x)$ is the k^{th} point spread function, and x is a two-dimensional coordinate variable. The operator * denotes convolution. The goal, of course, is to estimate $o(x)$ and $s_k(x)$ from noisy data $d_k(x)$, and the following form for the data is assumed:

$$d_k(x) = g_k(x) + n_k(x) \qquad (2)$$

(a) (b)

Figure 2: (a) typical raw frame and (b) 16-frame MFBD reconstruction, D/r_0=16, SNR=40 dB

There are two schools of thought with regard to an estimation approach. In one case, maximum likelihood (ML) estimation, the parameters of interest are regarded as deterministic quantities; the only source of randomness is the noise. The second approach, maximum *a posteriori* (MAP) estimation, regards each of the parameter values as a realization of an underlying random process. While the distinction is a bit subtle, MAP estimation has the potential for more accurate estimates if the underlying random processes can be statistically described.

The following sections contain a brief review of these two types of estimation [19], primarily to establish notation and draw comparisons that are commonly used in the image reconstruction field.

1.1.2.1 Maximum Likelihood Estimation

A simple example will be used to illustrate the maximum likelihood framework to estimate the unknown parameters of interest. The fundamental basis of MLE is that the parameters are non-random. The only source of randomness is the noise which corrupts the various data, denoted by *w*. For this problem, we assume a model with a single parameter A to describe the various measurements *d*(*n*):

$$d(n) = A + w(n) \tag{3}$$

The noise is assumed to be white Gaussian noise (WGN) or Gaussian distributed with zero mean, variance σ_w^2, and uncorrelated from sample to sample. As a result, the data will also be Gaussian distributed but with a mean of A. The probability density function (PDF) of *d* is parameterized by the variable A and will be denoted with a semicolon.

$$p(d(n);A) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left[-\frac{1}{2}\left(\frac{d(n)-A}{\sigma_w} \right)^2 \right] \tag{4}$$

Often independence of the noise samples is assumed which allows the PDF of a set of random variables to be formed as the product of their marginal distributions. This also holds true in this particular case of uncorrelated Gaussian RVs. Denoting the set of N measurements by **d**, its PDF is jointly Gaussian distributed (in N dimensions):

$$p(\mathbf{d};A) = \prod_n p(d(n);A) = \frac{1}{\left(2\pi\sigma_w^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma_w^2} \sum_n \left(d(n)-A\right)^2 \right] \tag{5}$$

For fixed values of $d(n)$, this can be regarded as a function of the parameter A, and this function is called the "likelihood function". The goal of maximum likelihood estimation is to identify the location of the peak in the likelihood function; i.e., identify the parameter value "most likely" to have given rise to the measured data:

$$\hat{A}_{ML} = \arg\max_{A} p(\mathbf{d}; A) \tag{6}$$

This location is unchanged by monotonic transformation, so in practice the logarithm of the PDF is used as the likelihood function. (This is more correctly called the "log-likelihood function", but is often shortened as it, too, satisfies the intuitive nature of "likelihood".) Further, terms which do not contain the parameters of interest are ignored as they do not affect the location of the peak (they merely shift the entire function up or down), allowing yet another formulation of the "likelihood function". These terms will be denoted "O.T." for clarity.

$$L\left(\hat{A} \mid \mathbf{d}\right) = \ln p(\mathbf{d}; \hat{A}) = -\frac{1}{2\sigma_w^2} \sum_n \left(d(n) - \hat{A}\right)^2 + \text{O.T.} \tag{7}$$

Note the dimensionality of this function is equal to the number of parameters (N_θ), one in this case. Also note the change in notation where the pipe (|) is used to denote "given the realization" of a particular random variable. Finding the location of the peak of the likelihood function is relatively straight forward, involving equating the derivative to zero and solving for the parameter of interest. (When multiple parameters are involved, the solution is the point in the N_θ-dimensional space where all partial derivatives equate to zero.) For this simple example, we have:

$$\begin{gathered} \frac{d}{d\hat{A}} L\left(\hat{A} \mid \mathbf{d}\right) = 0 \\ \frac{d}{d\hat{A}} L\left(\hat{A} \mid \mathbf{d}\right) = \frac{-1}{2\sigma_w^2} \sum_n \frac{d}{d\hat{A}} \left[\left(d(n) - \hat{A}\right)^2 + \text{O.T.} \right] = \frac{1}{\sigma_w^2} \sum_n \left(d(n) - \hat{A}\right) = 0 \end{gathered} \tag{8}$$

Solving for $\hat{A}$, we find the maximum likelihood estimate to be:

$$\hat{A}_{ML} = \frac{1}{N} \sum_n d(n) \tag{9}$$

1.1.2.2 Maximum *A Posteriori* Estimation

Maximum likelihood estimation is often utilized due to its simplicity. It needs no other information other than a model and data. As previously mentioned, however, it stands to reason that more accurate estimates may be possible if more information is available. In the maximum *a posteriori* framework, the parameters of interest are regarded as realizations of random variables, the possible values of which can be described in a statistical manner. For example, that in addition to the N samples of data, we also know that the random variable A is normally distributed as $N\left(\mu_A, \sigma_A^2\right)$.

As before, we wish to obtain our estimate as the solution to an equation involving our estimate. As can be guessed from the name of the framework, we seek to identify the location of the maximum of the *a posteriori* distribution:

$$\hat{A}_{MAP} = \arg\max_A p(A \mid \mathbf{d}) \tag{10}$$

We turn to Bayes' Theorem which relates the conditional probabilities of these random quantities:

$$p(A \mid \mathbf{d}) = p(\mathbf{d} \mid A)\frac{p(A)}{p(\mathbf{d})} \tag{11}$$

The first term $p(\mathbf{d} \mid A)$ was derived in the maximum likelihood section (though with the notation $p(\mathbf{d}; A)$), and the next term is the ratio of the unconditional densities. This is very intuitively satisfying, as it effectively amplifies or attenuates various parts of the maximum likelihood formulation in accordance with the *a priori* information.

As before, we can define several related "likelihood" functions based on $p(A \mid \mathbf{d})$, here using the logarithm and ignoring terms not involving A (since they do not influence the location of the function peak):

$$L\left(\hat{A} \mid \mathbf{d}\right) = \ln p(\hat{A} \mid \mathbf{d}) = \ln p\left(\mathbf{d} \mid \hat{A}\right) + \ln p\left(\hat{A}\right) + \text{O.T.} \tag{12}$$

Substitution of the densities gives a likelihood function with two useful terms:

$$L\left(\hat{A} \mid \mathbf{d}\right) = -\frac{1}{2\sigma_w^2}\sum_n \left(d(n) - \hat{A}\right)^2 - \frac{1}{2\sigma_A^2}\left(\hat{\mathrm{A}} - \mu_{\mathrm{A}}\right)^2 + \text{O.T.} \tag{13}$$

We note a data-driven term, which was present in the maximum likelihood formulation as well as a second term which reduces the likelihood as the estimate deviates from the *a priori* expected value. Proceeding as before, we equate the derivative of the likelihood function to zero in order to find the estimate. Kay addresses this derivation in Example 10.1 in [19] where he finds the maximum *a posteriori* estimate to be:

$$\hat{A}_{MAP} = \frac{\sigma_A^2 \hat{A}_{ML} + \sigma_{ML}^2 \mu_A}{\sigma_A^2 + \sigma_{ML}^2} \tag{14}$$

where $\sigma_{ML}^2 = \sigma_w^2 / N$ is the variance of the maximum likelihood estimate. From this, we see the MAP estimate is a weighted average of the data-driven (maximum likelihood) estimate ($\hat{A}_{ML}$) and the *a priori* estimate (the mean μ_A). Examination of Eq. 14 shows a natural fusion or blending of information as the number of samples grows. With few data points, the estimate is heavily influenced by the *a priori* information; with many points, the data-driven estimate dominates.

1.1.3 Constrained Estimation

We contend the primary difference in ML and MAP estimation is the injection of *a priori* information. The MAP estimation procedure can leverage more than the noisy measurements, and thus should be more accurate than ML. In practice, however, it is difficult to describe the statistical nature of the underlying random processes (i.e., specify the prior densities). Consider the case of the object brightness. We usually know little more than that the pixel values must be positive. While it is possible to determine an upper bound (requiring knowledge of the source intensity, attenuation of any intervening media, positions of source and object, the angularly-dependent reflectance of the (unknown) object, and so forth), this is usually intractable, computationally prohibitive, or simply not a tight enough bound to be useful.

1.1.3.1 ML with Soft Constraints

Neglecting the philosophical differences of random vs. deterministic parameters, we have seen the ML and MAP estimator equations are remarkably similar. Indeed, under the Gaussian noise assumption, the MAP estimator simply has an additional term, which serves to lower the likelihood "score" of the ML estimator. This has been generalized to the use of "penalty terms"

to reject undesirable solutions. We broaden our notation to include a number of parameters organized into a vector $\boldsymbol{\theta}$.

$$L_{soft}\left(\hat{\boldsymbol{\theta}} \mid \mathbf{d}\right) = L_{ML}\left(\hat{\boldsymbol{\theta}} \mid \mathbf{d}\right) - \lambda P\left(\hat{\boldsymbol{\theta}} \mid \mathbf{d}\right) \tag{15}$$

For instance a solution containing negative object pixel values can be made less desirable by augmenting the ML equation with a term that increases with more (or larger) negative values:

$$P\left(\hat{\boldsymbol{\theta}} \mid \mathbf{d}\right) = P\left(o(x)\right) = \sum_{x \in \{o(x)<0)\}} \left|o(x)\right|^2 \tag{16}$$

Smoothness in the intensity is usually a desirable property for many man-made objects (except around edges). One term which penalizes rapid spatial intensity variations is based on the Laplacian operator:

$$P\left(\hat{\boldsymbol{\theta}} \mid \mathbf{d}\right) = P\left(o(x)\right) = \sum_{x} \nabla^2 o(x) \tag{17}$$

(The Laplacian operator is often approximated by the difference of a point and the average of its neighbors.)

Multiple constraints can simply be tacked on as additional terms, each weighted with its own strength coefficient. On the surface, this is all very pleasing. In each of these, more information is present than the raw measurements, hopefully yielding more accurate estimates.

The choice of penalty strength λ, by design, will greatly influence the solution, but this is unfortunately a double-edged sword. An "optimal" value is usually problem specific and often must be determined – i.e., tuned – by trial and error or some variant of supervised learning techniques.[20,21] Multiple constraints can be especially problematic if there are undesired cancellation or amplification effects. Penalty terms are types of "soft" constraints because some violation is possible or even desired.

1.1.3.2 ML with Hard Constraints

In contrast, a hard constraint will always be satisfied. There is no "gray area", and therefore, no tuning constant (or constants) with which the researcher must fiddle (i.e., the various λs). To implement hard constraints, the problem is redefined in such a manner that the constraint must be satisfied. Consider for example, the positivity (technically, the non-

negativity) of the object intensity values. Following Biraud's approach [22], we can redefine the object values to be the square of the parameters of interest:

$$o(x) = p^2(x) \tag{18}$$

With this change, the object intensity is always non-negative. A hard constraint has been applied, hopefully yielding more accurate solutions.

1.1.4 Nonlinear Search

A side effect of many hard constraints is that the likelihood equation is obviously nonlinear (in terms of the parameters of interest). Indeed, by the vary nature of the incoherent imaging problem, the object and PSF parameters combine in a nonlinear manner through the convolution operator. Even further, the use of a realistic noise assumption (e.g., Poisson) also yields a nonlinear equation. Ultimately, closed form expressions for the parameter estimates are not typically available, mandating the use of numerical techniques (i.e., iterative) to identify the parameters which maximize the likelihood function.

Consider the nature of this likelihood function. If the recorded images are 128x128 pixels, a typical implementation requires some sixteen thousand variables (parameters) to describe the object intensity alone. While there are some techniques to reduce this (e.g., a size assumption, a.k.a. a "support constraint"), the likelihood function is nevertheless very multi-dimensional. Recall that the goal of the estimation process is to identify the location of the peak of a multi-dimensional hyper-surface.

The maximization of a multidimensional function is no trivial matter. The dependency of the function on the parameters plays a crucial role in how difficult that optimization is. It is natural to wonder if these hard constraints may have made it more difficult to find the solution (peak). Are there other parameterizations of positivity that give rise to a hyper-surface that are more easily searched? What about using the original formulation – directly searching for $o(x)$ – but using an optimizer with bounds restrictions (i.e., only searching in the area $o(x) \geq 0$)?

1.1.5 Other Approaches to Improve Performance

The majority of the effort in this field has been, in one form or another, an attempt to increase performance. "Performance" is a bit vague, but perhaps rightly so, as execution time and quality are easily traded. Consider an advancement that reduces the time per iteration. After

this improvement is implemented, the researcher can stop sooner for the same quality or they can run for the same length of time and achieve better quality. Conversely, an advancement that makes larger progress toward the solution need not run as long to be acceptable. Individual circumstances will usually dictate which aspect is the higher priority.

The approaches to improvement can be grouped into two categories (as described in Table 1): hardware/software development and those related to the image reconstruction algorithms themselves. The most obvious solution to improvement is simply to run on faster hardware. The performance differences when executed on computers that are just a few years more recent are amazing. In this category, we also lump improvements due more efficient software development methods such as reusing computed quantities or use of compiled languages.

Additionally, one can take advantage of parallel resources. MFBD reconstructions are extremely amenable to the “embarrassingly parallel” approach because individual reconstructions typically do not depend on each other. If latency can be tolerated, this often achieves maximum throughput. Sometimes, however, it is desired to make a single reconstruction as fast as possible (minimizing latency), and finer grained parallelism is warranted. MFBD is also quite amenable to this approach, as the individual frames are typically independent, with a bit of consolidation for the common object calculations, as will be seen later in the theoretical development.

Lastly, dedicated implementation of parts of the processing logic in hardware can pay dividends. For example, we will see that the number of Fourier transforms plays a key role in runtime per iteration. Specialized array processors, field programmable gate arrays (FPGAs), or application-specific integrated circuits (ASICs) are all approaches that can accelerate part of the processing.

The second category of improvement is to make improvements in the algorithms themselves. Many of these attempts involve the incorporation of *a priori* information. This research is one example of this category. Additional constraints serve to reduce the number of potential solutions which may make it easier for the search algorithm to find an acceptable solution.

Object positivity, smoothness, and support are all extremely common constraints. Beyond that, approaches are quite varied. Phase-diversity is common for imaging of the sun where reduced SNR due to the partitioning of photons onto multiple detectors is less of a concern.[23,24] Others have used Kolmogorov statistics as prior information for the point spread functions.[25] In the case of known object classes, incorporation of *a priori* information on the power spectrum of the object can improve reconstructions as well.[26] Other research has made use of the tendency for the high spatial frequency content of natural scenes to be sparse (i.e., many frequencies have a magnitude near zero).[27,28]

The method by which the smoothness constraint is achieved also varies. While many use convolution with a relatively smooth kernel, others have sought to minimize the total variation (roughly akin to path or arc length as a function varies).[29,30]

An additional constraint or assumption often imposed on the PSFs is that of phase-only distortions in the aperture. (We discuss this in section 2.4.1.) Originally adopted in order to reduce the number of degrees of freedom, reconstructions employing this assumption were found to often be fairly accurate even in the presence of magnitude distortions (scintillation) at the aperture.[31]

Particulars of the imaging system can also suggest constraints. For example, truncation of the object by the edge of the field of view can be explicitly modeled with a “guardband” [32,33] and object information outside of the field of view estimated from information that was blurred into the frame by the atmospheric distortion. Such a constraint is crucial to avoid ringing artifacts from the periodic assumption imposed by the use of the discrete Fourier transform while processing. Detector pixel saturation and grayscale quantization [34] of the raw imagery can be explicitly modeled and reconstructions improved as a result. The pixel sampling (or angular field of view) of the images can be modeled and a “de-aliased” object estimated at a higher spatial sampling rate.[35,36]

Another possibility is to start the search from a point closer to the final solution – possibly leveraging other solution techniques (e.g., bispectrum [8,37] or an MFBD estimate based on previous frames) in the formation of the initial estimate. Finally, improvements may be found through the use of different optimization algorithms. Many of the very early MFBD papers describe searching the parameter space through iteratively applying constraints in the spatial and

frequency domains, a specialized case of generalized projections.[15,38] Later approaches (e.g., Lane [16]) cast the problem in terms of minimizing a cost or error function using an iterative search algorithm such as steepest-descent, conjugate gradient, or expectation maximization search algorithms. Presently, conjugate gradient and limited memory quasi-Newton methods of optimization of a likelihood or cost function seem to dominate typical use, but these are certainly not the only approaches.[39]

In addition to the search technique, the quantities being sought have changed over time. The early formulations (circa mid-1970s and 1980s) alternated between the object and PSF parameters. However, in this research, like that of many others after that timeframe, we search for the object and PSF parameters simultaneously. Interestingly, some of the most recent research into search techniques has resurrected the notion of focusing on one set of parameters at a time.[40,41]

Table 1: Common approaches used to improve reconstruction performance

Hardware/Software Development	Algorithm Enhancements
Newer computers	Constraints / *a priori* information **Positivity**, smoothness, support, symmetry, phase-only distortions, etc.
Implementation choices	
Compiled languages	
Floating point coprocessors	Multiple measurements (frames or wavelength)
Multiple processors (parallelism)	Improved initial estimates
FPGAs / ASICs	Search or optimization technique

For context, this research seeks to increase performance by combining aspects of these two categories. From the Software Development side, there are many approaches to implementing or achieving a desired end result. From the Algorithm side, we adopt one of the most common constraints as that desired end result. These are bolded in Table 1. Specifically, we investigate the particular manner in which positivity is implemented in the reconstruction software. The hypothesis is that performance will vary across the different implementation approaches.

1.2 Research Description

In the Background section, we noted it is desirable to augment maximum likelihood estimation with constraints. In this study, we wish to investigate the implementation of a positivity constraint in an estimation procedure. While we consider four parameterizations for evaluation, the near universal use of a smoothness constraint by other researchers mandates its inclusion in this work as well. As will be described in section 2.3, performance is affected by the order of application of the positivity and smoothness constraints, and both orderings are evaluated for each type of positivity implementation. The following approaches are considered (with equation numbers taken from section 2.3):

Table 2: Object parameterizations to investigate

Description	Parameterization (Eq)
Square 1	$p^2(x)*b(x)$ (56)
Square 2	$[p(x)*b(x)]^2$ (57)
Exp 1	$\exp p(x)*b(x)$ (62)
Exp 2	$\exp[p(x)*b(x)]$ (63)
Abs 1	$\lvert p(x)\rvert*b(x)$ (71)
Abs 2	$\lvert p(x)*b(x)\rvert$ (72)
Bounds	$p(x)*b(x)$ (82)

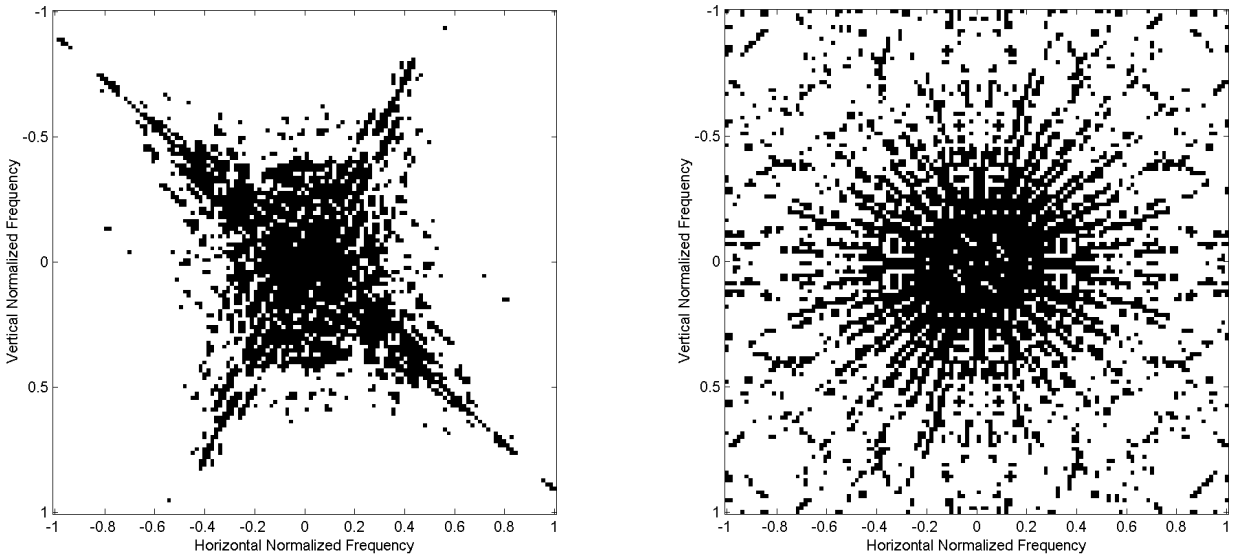
In the Bounds case, a bounds-constrained optimizer is used to search over the region where $p(x)$ is non-negative, and smoothness must be applied after positivity. In all cases, the smoothness constraint is achieved by convolution with a blur kernel $b(x)$. In the author's discussions with other researchers and literature search, the Square 2 parameterization is universally employed.

Criteria used for evaluation are each method's quality of reconstruction (root mean square or RMS spatial domain error and the RMS phase error in the spatial frequency domain) and computational demand (time required for a constant number of search iterations). Extensive attempts at code optimization were performed for more accurate comparison of the techniques (as opposed to their software implementations).

Performance of each object parameterization was evaluated across four data sets. These data sets are the combination of two different objects (or scenes) and two different levels of signal to noise ratio (corresponding to two levels of object brightness). The selected objects differed significantly in shape and spatial frequency content (shown in Figure 3 and Figure 4).

(a) (b)

Figure 3: True objects used in the image generation: (a) Satellite and (b) Spoke objects



(a) (b)

Figure 4: Areas of significant spatial frequency content based on the true spectrum of (a) Satellite and (b) Spoke objects

1.3 Significance of the Study

In the broadest sense, we wish to optimize reconstruction performance through the use of constraints. In our discussions with various researchers, any of them who employed a positivity constraint used the technique of squaring the smoothed parameters (i.e., the Square 2 parameterization), but something better may be available. The notion of "better" may be tied to reconstruction quality or speed of reconstruction.

Advances in sensor technology have given rise to increasingly larger arrays. While the computation capability of computers has also advanced, the computational burden of nonlinear image reconstruction is decidedly non-trivial. The state of the art has turned to massive parallelization to decrease the time spent on a given reconstruction, and near real-time results are now available. Even a modest savings in computational complexity will pay dividends due to the multiple iterations that are typically required. It may be possible that with a less complex likelihood surface, fewer iterations would be required to achieve the same likelihood or result.

We first developed a generalized, theoretical framework that is conducive to exploration of various implementations of hard constraints. This framework isolated the choice of noise model from the choice of object or PSF parameterization, yielding a consistent procedure for evaluating these constraints. While we applied this framework specifically to object positivity, it is by no means limited to that. Other constraints for either the object or PSF can be examined equally as well.

We then evaluated several implementations of object positivity (and their interaction with a common smoothness constraint) and found a clear benefit to using something other than the "standard" method. The benefit in this case was faster execution with no discernable degradation to the accuracy of reconstruction, as will be shown in Section 3 Results.

2 THEORETICAL DEVELOPMENT

2.1 Likelihood Gradients

We continue the development of Multi-frame Blind Deconvolution within the maximum likelihood framework by considering the nature of the noise corrupting our noiseless images. We assume a zero-mean noise distribution. The average value, then, at a particular pixel (x) will equal that of noiseless image $g_k(x)$. We also assume the noise is independent from pixel to pixel and from image to image. As a reminder, $g_k(x)$ is an implicit function of the object and k^{th} point spread function (atmospheric distortion).

Following the development of section 1.1.2.1, we seek to determine a likelihood function for our set of measurements. In the case of MFBD, those measurements comprise the set of pixel values in the images $\{d_k(x)\}$, so the likelihood function is:

$$L\left(\hat{o}(x), \{\hat{s}_k(x)\} \mid \{d_k(x)\}\right) = \ln p(\{d_k(x)\}; \{g_k(x)\}) = \sum_k \sum_x \ln p\left(d_k(x); g_k(x)\right) \qquad (19)$$

The ML estimates of the object and PSFs are found by identifying $\hat{o}(x)$ and $\{\hat{s}_k(x)\}$ that maximize the likelihood function. This is found by finding the point at which the partial derivatives of the likelihood function vanish. We use the chain rule extensively throughout this development and also drop the dependency and estimate/hat notation on the likelihood function. We start by developing an expression for the likelihood gradient with respect to (w.r.t.) the object parameters.

2.1.1 Gradients w.r.t. Object Parameters

Since we plan to discuss several object parameterizations (see section 1.1.5), we will first investigate the likelihood with respect to a particular object pixel value and then consider the effect of the object parameterizations on that pixel value. We recall the chain rule form of the derivative with respect to $o(y)$, where y is a 2 dimensional indexing variable:

$$\frac{\partial L}{\partial o(y)} = \sum_k \sum_x \frac{\partial L}{\partial g_k(x)} \frac{\partial g_k(x)}{\partial o(y)} \tag{20}$$

The last derivative is that of the spatially-invariant incoherent imaging equation:

$$\frac{\partial g_k(x)}{\partial o(y)} = \frac{\partial}{\partial o(y)} \left[o(x) * s_k(x) \right] = \frac{\partial}{\partial o(y)} \sum_z o(z) s_k(x-z) = s_k(x-y) \tag{21}$$

The first term depends on the particular noise model assumed. For now, denote this term by:

$$r_k(x) \equiv \frac{\partial L}{\partial g_k(x)} \tag{22}$$

It will soon be seen that this term acts as a "residual" term that describes how well the estimate matches the data. With this definition, we have the likelihood gradient with respect to the y[th] pixel in the object estimate as:

$$\frac{\partial L}{\partial o(y)} = \sum_k \sum_x r_k(x) s_k(x-y) \tag{23}$$

The summation over x is recognized as the definition of a correlation[◊] giving the likelihood gradient as an accumulation of the correlations between the residuals and PSFs:

$$\frac{\partial L}{\partial o(y)} = \sum_k r_k(y) \otimes s_k(y) \tag{24}$$

[◊] We remind the reader that unlike the convolution operator (*), the correlation operator ($\otimes$) is not commutative.

At this point, we recall that the y[th] pixel of the object estimate $o(y)$ is itself a function of our search parameters $p(x)$ discussed in section 1.1.5. Employing the chain rule one last time, we find the likelihood gradient with respect to our object parameters of interest as:

$$\frac{\partial L}{\partial p(x)} = \sum_y \frac{\partial L}{\partial o(y)} \frac{\partial o(y)}{\partial p(x)} = \sum_y \left[\sum_k r_k(y) \otimes s_k(y) \right] \frac{\partial o(y)}{\partial p(x)} \tag{25}$$

In the remaining development, we derive expressions for $r_k(y)$ under three different noise models (section 2.2). We also compute the gradients of the various object parameterizations under investigation (section 2.3). As a practical matter, these specific instances merely "plug into" the general expression given in Eq. 25.

2.1.2 Gradients w.r.t. Point Spread Function Parameters

In addition to the object parameters, blind deconvolution seeks to estimate the parameters describing the various point spread functions. Due to the commutative property of the convolution operator, the likelihood gradient with respect to the PSF parameters is similar to that for the object parameters. The fundamental difference is that a given PSF parameter is specific to a particular frame, and the likelihood gradient w.r.t. that parameter will only contain contributions from its associated frame. We are first concerned with the derivative of the noiseless imaging equation with respect to the y[th] pixel value in the m[th] PSF. (This is the analog of Eq. 21.)

$$\frac{\partial g_k(x)}{\partial s_m(y)} = \frac{\partial}{\partial s_m(y)} \sum_z s_k(z) o(x-z) = \sum_z \frac{\partial}{\partial s_m(y)} s_k(z) o(x-z) \tag{26}$$

While we implicitly interchanged the order of summation and derivative operators in the object development, here we show this expression to highlight the subscripts on the PSFs. This gradient is zero for $k \neq m$, and we can represent this with an indicator function $I(k\text{-}m)$ (also known as a unit sample sequence[52]).

$$I(w) = \begin{cases} 1, w = 0 \\ 0, w \neq 0 \end{cases} \tag{27}$$

Using this notation, we have:

$$\frac{\partial g_k(x)}{\partial s_m(y)} = \sum_z o(x-z) I(z-y) I(k-m) = o(x-y) I(k-m) \tag{28}$$

This indicator function will "filter" the summation over k that was kept in the previous development, and a correlation is recognized as before. The gradient with respect to the PSF pixel value is thus:

$$\frac{\partial L}{\partial s_m(y)} = \sum_k \sum_x r_k(x) o(x-y) I(k-m) = r_m(y) \otimes o(y) \tag{29}$$

The final step is to consider that the PSF pixel values could be represented by some function of the PSF parameters. As an example, positivity as well as normalization could be implemented. For now, we consider the gradient with respect to a particular parameter $q_k(j)$, which is the j[th] parameter of the k[th] PSF:

$$\frac{\partial L}{\partial q_k(j)} = \sum_m \sum_y \frac{\partial L}{\partial s_m(y)} \frac{\partial s_m(y)}{\partial q_k(j)} = \sum_m \sum_y r_m(y) \otimes o(y) \frac{\partial s_m(y)}{\partial q_k(j)} \tag{30}$$

2.2 Noise Models

We next consider the effect of several noise models. The particular choice of noise model will dictate the form of the likelihood function as well as the residual term $r_k(x)$ in the likelihood gradients.

2.2.1.1 Poisson Noise

The quantized nature of the incoming illumination gives rise to random arrival times of the photons. Equivalently, the number of photons landing on our detector during a fixed exposure time is random. The average number of photons to arrive on a particular pixel is described by the noiseless image $g_k(x)$. Any particular realization, however, is governed by a Poisson distribution. Recall the probability density function[◊] for a Poisson random variable n with an average number of events λ:

$$p_{Poisson}(n;\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{31}$$

[◊] Arguably a "probability mass function"; however, we will generalize to other non-discrete PDFs later and view "density function" as the more general term.

From (19) and ignoring terms that are irrelevant to maximization, a Poisson noise model gives a likelihood function of

$$L_{Poisson} = \sum_k \sum_x \ln p\big(d_k(x); g_k(x)\big) = \sum_k \sum_x d_k(x) \ln g_k(x) - g_k(x) \tag{32}$$

The gradient residual term is then:

$$r_k^{(Poisson)}(x) = \frac{\partial L}{\partial g_k(x)} = \frac{d_k(x)}{g_k(x)} - 1 \tag{33}$$

Now that we have a concrete expression for $r_k(x)$, we can see the residual nature that contributes to the overall gradient computations. When the object and PSF estimates cause the estimated noiseless or "rendered" image to be comparable to the measured image, the gradient becomes small. From that standpoint, those estimates are viewed as close to optimal. In practice, it is undesirable for this term to be exactly zero due to a tendency to "fit the noise" as described in section 2.3.2. Fortunately, overfitting is less of a problem as more frames are used.

One final comment is that these expressions are ill-defined as $g_k(x)$ approaches zero. Consequently, care must be taken during implementation to "avoid" numeric issues in such regions for reconstructions of space-based scenes.

2.2.1.2 Gaussian Noise

Most imaging detectors typically have noise associated with the read-out and associated electronics as well as noise due to thermal variations. This noise is very well described by a Gaussian probability density function with a constant variance. Use of this density is appropriate when read noise is relatively large (dominating the photon noise) or the scene has relatively low contrast (and the photon noise consequently has approximately constant variance).

$$p_{Gaussian}(d; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d-\mu}{\sigma}\right)^2} \tag{34}$$

With the realization that the mean value of a pixel equals that of the noiseless image, (19) can be used to find a tentative constant-variance Gaussian likelihood function of

$$L_{Gaussian,1} = \sum_k \sum_x \ln p\big(d_k(x); g_k(x)\big) = \sum_k \sum_x -\frac{1}{2} \frac{\left[d_k(x) - g_k(x)\right]^2}{\sigma^2} - \ln \sigma \tag{35}$$

We explicitly note here that the effect of a constant variance (σ) is a bias and scaling but does not alter the location of the peak of the likelihood function. Thus, an equivalent likelihood function that can be maximized is:

$$L_{Gaussian} = -\sum_k \sum_x \left[d_k(x) - g_k(x) \right]^2 \tag{36}$$

From this, we see that this noise model is intuitively satisfying from a "least-squared-error" interpretation. Maximizing this likelihood function is equivalent to minimizing the squared error. The gradient residual function easily found to be:

$$r_k^{(Gaussian)}(x) = \frac{\partial L}{\partial g_k(x)} = 2\left[d_k(x) - g_k(x) \right] \tag{37}$$

Finally, we see these two terms (L and $r_k(x)$) remain well defined as $g_k(x)$ tends toward zero.

2.2.1.3 Spatially Dependent Gaussian Noise

The previous noise models address two different sources of noise, and the effectiveness of each model will depend greatly on the relative contributions of those noise sources to the imagery. It is reasonable to expect that the Poisson noise model will be more accurate when read noise is negligible; but unfortunately, read noise is not always negligible. Also in many cases, pixel values are biased functions of the number of photons captured during the exposure time of the image; i.e., a pixel exposed to an effective intensity level of zero has an output value greater than zero counts. In the quest for a more faithful model of the entire imaging system, many imaging systems can afford to characterize their detectors and estimate the bias signal (B) and read noise variance (σ_d^2). Accounting for these effects separately from the object intensity will yield a more accurate photometric characterization of the object itself.

We present one final noise model that allows the incorporation of detector noise characteristics in a straightforward manner as well as preserving the signal-dependent (and thus usually spatially dependent) noise that is present due to photon arrivals. Due to the Central Limit Theorem [42], we note that the PDF of a Poisson random variable with an expected count (λ) of more than about 100 is very well approximated by a Gaussian curve with variance and mean equal to that expected count. We also can leverage the fact that the variance of the sum of two Gaussian random variables is equal to the sum of the variances, allowing easy combination of the

effects of detector read noise (variance σ_d^2) and photon noise (variance $g_k(x)$). (We assume equal exposure times for all frames in the ensemble and therefore expect equal read-noise bias and variance for each pixel.)

Looking back at (35), we can no longer eliminate the variance from the likelihood function since it is not constant to all terms of the summation. We instead consider a spatially dependent variance:

$$\sigma_k^2(x) = \text{var}(\text{readnoise}) + \text{var}(\text{photon noise}) = \sigma_d^2 + g_k(x) \tag{38}$$

Thus, the likelihood function with a spatially-dependent Gaussian noise model with constant bias signal B is given by:

$$L_{SVGaussian} = \sum_k \sum_x \ln p\big(d_k(x); g_k(x)\big) = -\sum_k \sum_x \frac{\big[d_k(x) - g_k(x) - B\big]^2}{\sigma_k^2(x)} + \ln \sigma_k^2(x) \tag{39}$$

To compute the likelihood gradient, we first substitute an intermediate variable $w_k(x)$:

$$w_k(x) = d_k(x) - g_k(x) - B \tag{40}$$

so that the likelihood function can be written as:

$$L_{SVGaussian} = -\sum_k \sum_x \frac{w_k^2(x)}{\sigma_k^2(x)} + \ln \sigma_k^2(x) \tag{41}$$

and the derivative with respect to an arbitrary noiseless image pixel $g_m(y)$ is then:

$$\frac{\partial L}{\partial g_m(y)} = -\sum_k \sum_x \frac{\sigma_k^2(x) 2 w_k(x) \dfrac{\partial}{\partial g_m(y)} w_k(x) - w_k^2(x) \dfrac{\partial}{\partial g_m(y)} \sigma_k^2(x)}{\sigma_k^4(x)} + \frac{1}{\sigma_k^2(x)} \frac{\partial}{\partial g_m(y)} \sigma_k^2(x) \tag{42}$$

We next note the derivatives of $w_k(x)$ and $\sigma_k^2(x)$ with respect to the arbitrary pixel are nonzero only for $k=m$ and $x=y$, which can be reflected with indicator functions (as in 2.1.2):

$$\frac{\partial}{\partial g_m(y)} w_k(x) = -I(k-m) I(x-y) \tag{43}$$

$$\frac{\partial}{\partial g_m(y)} \sigma_k^2(x) = +I(k-m) I(x-y) \tag{44}$$

These indicator functions will reduce the summations in the likelihood gradient to terms involving the indices *m* and *y* only, and we can distribute the negative in front of the summations to find:

$$\frac{\partial L}{\partial g_m(y)} = \frac{\sigma_m^2(y)2w_m(y) + w_m^2(y)}{\sigma_m^4(y)} - \frac{1}{\sigma_m^2(y)} \tag{45}$$

For a slightly simpler expression, we can factor $w_m(y)$ out of the first term:

$$\frac{\partial L}{\partial g_m(y)} = w_m(y)\left(\frac{2}{\sigma_m^2(y)} + \frac{w_m(y)}{\sigma_m^4(y)}\right) - \frac{1}{\sigma_m^2(y)} \tag{46}$$

and employ a common denominator of $\sigma_m^2(y)$:

$$\frac{\partial L}{\partial g_m(y)} = \frac{w_m(y)\left(2 + \frac{w_m(y)}{\sigma_m^2(y)}\right) - 1}{\sigma_m^2(y)} \tag{47}$$

We now realize that *m* and *y* are arbitrary indices and replace them with *k* and *x* to match the previous definition of the gradient residual expression (22). Additionally, we substitute the definition of $w_k(x)$ to find the residual function without that intermediate variable as:

$$r_k^{(SVGaussian)}(x) = \frac{\partial L}{\partial g_k(x)} = \frac{\left(d_k(x) - g_k(x) - B\right)\left(2 + \frac{d_k(x) - g_k(x) - B}{\sigma_k^2(x)}\right) - 1}{\sigma_k^2(x)} \tag{48}$$

Finally, we note that the majority of real-world detectors will have a non-zero read noise variance. In contrast to the Poisson formulation which had g_k(x) alone in the denominator of the residual function, inclusion of an estimate of the read noise variance ($\sigma_k^2 = \sigma_d^2 + g_k(x)$) will greatly increase the numerical stability of *L* and $r_k(x)$ for near-zero pixel values in the noiseless image g_k(x). (A similar limiting effect is seen on low-amplitude spectral terms when performing Wiener filtering.[43]) As one can imagine, near-zero values are likely to be common in g_k(x) when reconstructing images of orbiting objects.

2.3 Object Parameterizations

At this point, we have examined the likelihood function for several noise models and have expressions for the likelihood gradients which involve a residual term and the partial derivative of the object intensities with respect to the object parameters. In this section we

discuss several parameterizations of the object in order to achieve positivity. In addition, we also discuss a very important constraint, that of smoothness. The forward or "rendering" process will convert from the set of object parameter values $\{p(x)\}$ to the set of object intensity values $\{o(x)\}$ used in the calculation of the noiseless image estimates.

2.3.1 Pixelated Object

The first parameterization method is not much of a stretch of the imagination. The object pixel values are identically equal to the parameter values:

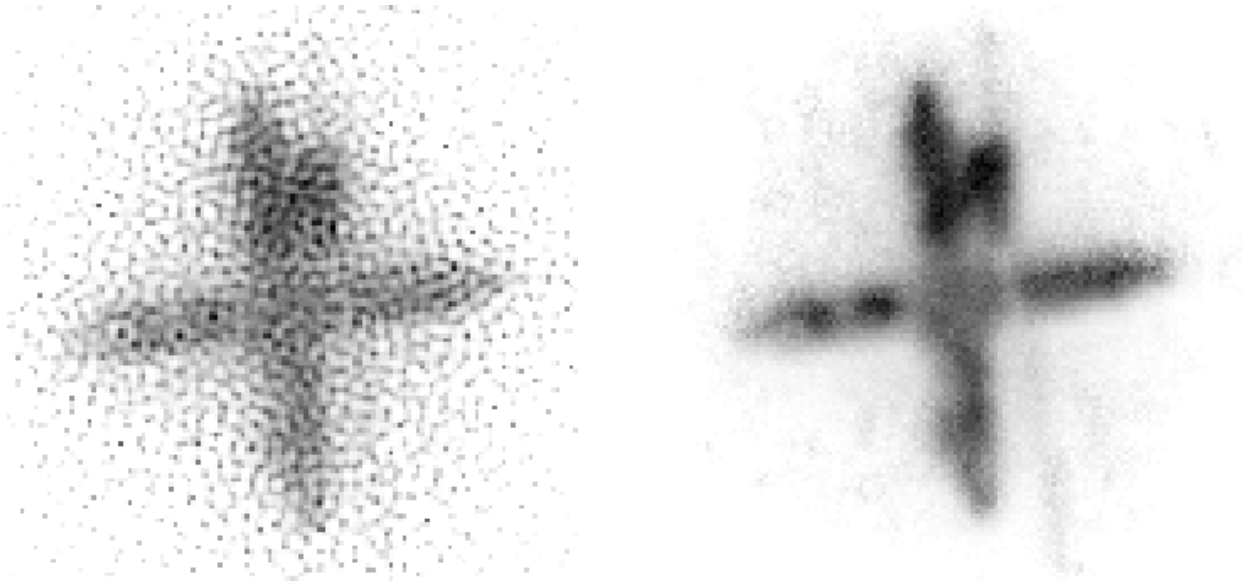
$$o(x) = p(x) \tag{49}$$

With an unconstrained optimization method, it therefore is possible for object pixels to take on negative values. This is undesirable considering that these pixels represent intensity values. In practice, many of the estimated pixel values will be positive by virtue of the likelihood function trying to match the (mostly positive) data. However, ringing near edges where the background intensity is very low or even zero (e.g., ground-based imagery of objects in orbit) will likely be present with estimated intensities below zero.

Secondly, such a parameterization has a great many degrees of freedom, and reconstructions have a tendency to have rapid variations between adjacent pixels. Such rapid variations are not typically present in the true object scene but are artifacts of estimating an answer from noisy data.[44,45] Historically, the most common "regularizing" technique to reduce these artifacts when using an iterative estimator was to abort the iterations before this occurred. Of course, choosing that stopping point was quite difficult and certainly data dependent. Another technique utilizes *a priori* knowledge of the object pdf or penalty terms in the likelihood function as discussed above. Including a term that is sensitive to rapid variations (e.g., gradient or localized variance) will guide the search away from the mottled solution.

2.3.2 Smoothness Constraint

A different technique, first developed in the area of tomography, uses the concept of sieves [46] or a small smoothing kernel [11,47]. As this is extremely common, we also employ this technique to reduce the tendency for the maximization procedure to overfitting the noise.



(a) (b)

Figure 5: Example object estimates that are (a) overfit and (b) regularized

From the *a priori* information viewpoint, a sieve enforces some amount of correlation between pixels within a given neighborhood. Abrupt transitions are reduced under the assumption that either most real objects do not have such sharp changes or that (in most instances) you're willing to sacrifice some of that sharpness to avoid the overfitting problem. The basic technique is to perform a weighted average or smoothing of the object parameters, which is often implemented as a convolution of the parameters and a blur kernel:

$$o(x) = p(x) * b(x) \tag{50}$$

Typically, a 2-dimensional, radially symmetric, Gaussian blur kernel is used as the kernel, and the full width at half max (FWHM) is one measure of the amount of blurring that will be induced. The width selected is based on the average signal-to-noise ratio (SNR) of the pixel values, with larger widths being used for imagery with lower SNR values.

2.3.2.1 Low-pass Filter Viewpoint

A second way to look at a smoothness constraint is as a low-pass filter operation. By taking the Fourier transform of (50), we see the object spectrum experiences a filtering of its spatial frequencies. A Gaussian spatial blur kernel, for example, will greatly attenuate the higher spatial frequencies and tend to pass the lower frequencies.

$$O(u) = P(u)B(u) \tag{51}$$

In fact, this is an excellent method by which to select the width of the blur kernel. In Figure 6, we have depicted the spectral amplitude of a hypothetical, one-dimensional object signal (blue) along with that of an example noise process realization (green).

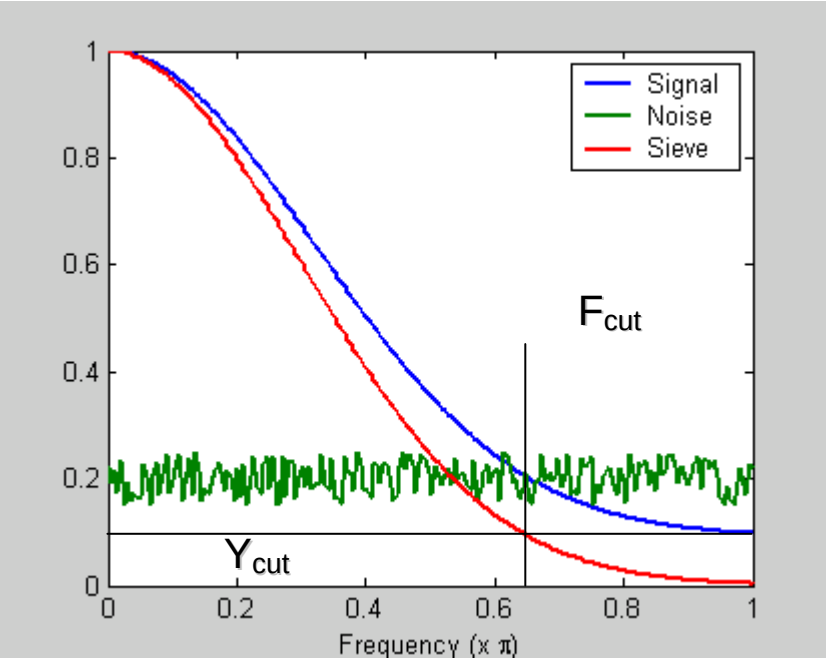


Figure 6: SNR-based sieve/filter design in the spatial frequency domain

The point at which the strength of the signal equals that of the noise (SNR = 1) makes for a natural cutoff frequency beyond which restoration of the signal is typically not attempted. A Gaussian curve can be constructed (red) such that it has some nominal value at that cutoff frequency. In this example, 0.1 was picked for the filter value, providing for at least 10 dB of attenuation in the noise-dominated areas of the spatial frequency domain. This selected cutoff frequency has a direct relation to the FWHM of the spatial kernel, and the filtering operation can

be performed in either the spatial domain (by convolution) or the frequency domain (by multiplication and then inverse transforming).

2.3.2.2 Basis Function Viewpoint

We have identified another way to view this smoothness constraint. In this alternate viewpoint, the object is regarded as a superposition of smooth functions. This can be seen by starting from the definition of the smoothed object:

$$o(x) = p(x)*b(x) = \sum_y p(y)b(x-y) \tag{52}$$

The argument to the blur function merely serves to reflect and shift it across the 2D array p(y). We can treat these reflected and shifted versions of $b(x)$ as a collection of functions indexed by a shift index (subscript):

$$\Phi_y(x) \equiv b(x-y) \tag{53}$$

Rewriting in this manner, it becomes obvious that the object is a weighted superposition of the functions in the set $\{\Phi_y(x)\}$, and the weights are given by the parameters we seek to estimate:

$$o(x) = \sum_y p(y)\Phi_y(x) \tag{54}$$

From here, it is a short leap to conceive of a different set of functions with other desirable properties (e.g., sharp edges, flat smooth regions, rotational symmetry, etc.) that would encompass more sophisticated regularization. Depending on the functions in the set, fewer parameters may be needed to adequately describe the object as well. One distinct advantage of the set of shifted copies, however, is that the speed of computation is potentially very fast due to implementation as a convolution (either spatially or via the Fourier domain).

2.3.3 Positivity through the Square

We now come to the heart of this research and explore a positivity constraint through the use of the square of the estimated parameters. This is a hard constraint, as it is impossible to obtain a negative value (for real $p(x)$):

$$o_{sq,0}(x) = p^2(x) \tag{55}$$

Due to the ubiquitous desire to regularize solutions with a smoothness constraint, it does little good to consider this particular parameterization. Instead, we consider two possibilities that differ in the order that positivity and the smoothness constraint are applied:

$$o_{sq,1}(x) = p^2(x) * b(x) \tag{56}$$

$$o_{sq,2}(x) = \left[p(x) * b(x)\right]^2 \tag{57}$$

Be aware, however, that the Square 1 formulation can technically permit negatives if $b(x)$ has negative values. Clearly when $b(x)$ is a Gaussian kernel this will not occur, but other choices of $b(x)$ could easily introduce negative values.

The above expressions are used to form $g_k(x)$ and compute the likelihood. As described in (25), the gradient residual function requires the partial derivative of these expressions with respect to the estimated object parameters. For $o_{sq,1}(y)$, we have

$$\frac{\partial o_{sq,1}(y)}{\partial p(x)} = \frac{\partial}{\partial p(x)} \sum_z p^2(z)\, b(y-z) = 2\, p(x) b(y-x) \tag{58}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{sq,1}(x)} = 2\, p(x) \left\{ \left[\sum_k r_k(x) \otimes s_k(x) \right] \otimes b(x) \right\} \tag{59}$$

For $o_{sq,2}(y)$, we have

$$\frac{\partial o_{sq,2}(y)}{\partial p(x)} = \frac{\partial}{\partial p(x)} \left[\sum_z p(z)\, b(z-y) \right]^2 = 2\left[p(y) * b(y)\right] b(y-x) \tag{60}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{sq,2}(x)} = 2 \left\{ \left[\sum_k r_k(x) \otimes s_k(x) \right] \left[p(x) * b(x)\right] \right\} \otimes b(x) \tag{61}$$

2.3.4 Positivity through the Exponential

The second method of implementing positivity is through the use of an exponential function, achieving a hard constraint (for real $p(x)$). As before, we examine this constraint in conjunction with a smoothness constraint due to widespread use and caution that Exp 1 may include negative values in the case where $b(x)$ contains negative values.

$$o_{\exp,1}(x) = \exp p(x) * b(x) \tag{62}$$

$$o_{\exp,2}(x) = \exp\left[p(x) * b(x)\right] \tag{63}$$

The above expressions are used to form $g_k(x)$ and compute the likelihood. As described in (25), the gradient residual function requires the partial derivative of these expressions with respect to the estimated object parameters. For $o_{\exp,1}(y)$, we have

$$\frac{\partial o_{\exp,1}(y)}{\partial p(x)} = \frac{\partial}{\partial p(x)} \sum_z \exp p(z)\ b(y-z) = \exp p(x) b(y-x) \tag{64}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{\exp,1}(x)} = \exp p(x) \left\{ \left[\sum_k r_k(x) \otimes s_k(x) \right] \otimes b(x) \right\} \tag{65}$$

For $o_{\exp,2}(y)$, we use an intermediate variable $w(z)$:

$$w(z) = p(z) * b(z) = \sum_y p(y) b(z-y) \tag{66}$$

so that the object can be written as:

$$o_{\exp,2}(y) = \exp w(y) \tag{67}$$

and the derivative of the object with respect to this variable is:

$$\frac{\partial o_{\exp,2}(y)}{\partial w(z)} = \exp w(y) I(y-z) \tag{68}$$

With the chain rule, we find

$$\frac{\partial o_{\exp,2}(y)}{\partial p(x)} = \sum_z \frac{\partial o_{\exp,2}(y)}{\partial w(z)} \frac{\partial w(z)}{\partial p(x)} = \sum_z \exp w(y) I(y-z) b(z-x) = \exp w(y) b(y-x) \tag{69}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{\exp,2}(x)} = \left\{ \left[\sum_k r_k(x) \otimes s_k(x) \right] o(x) \right\} \otimes b(x) \tag{70}$$

2.3.5 Positivity through the Absolute Value

The third method of implementing positivity is through the use of the absolute value function, achieving a hard constraint. As before, we examine this constraint in conjunction with a smoothness constraint due to widespread use and caution that Abs 1 may include negative values in the case where $b(x)$ contains negative values.

$$o_{\exp,1}(x) = \left|p(x)\right| * b(x) \tag{71}$$

$$o_{\exp,2}(x) = \left|p(x) * b(x)\right| \tag{72}$$

The above expressions are used to form $g_k(x)$ and compute the likelihood. As described in (25), the gradient residual function requires the partial derivative of these expressions with respect to the estimated object parameters. First, we recall the (modified) derivative of the absolute value function:

$$\frac{d}{du}\left|f(u)\right| = \operatorname{sgn} f(u)\left[\frac{d}{du} f(u)\right] \tag{73}$$

where the signum (or "sign") function is defined as:

$$\operatorname{sgn} x = \begin{cases} +1, x > 0 \\ 0\ , x = 0 \\ -1, x < 0 \end{cases} \tag{74}$$

In truth, the derivative of |f(u)| is undefined where f(u) = 0. However, we realize that if f(u) = 0^+ (a number slightly greater than 0), the optimization process will seek to reduce f(u), and if f(u) = 0^- (a number slightly less than 0) the gradient search will seek to increase f(u). Therefore, in such a case, the locally optimal value must be f(u) = 0, so we simply define the derivative for the purposes of the gradient search to be 0. (The signum function is a convenient method to achieve this in practice.)

For $o_{abs,1}(y)$, we have

$$\frac{\partial o_{abs,1}(y)}{\partial p(x)} = \frac{\partial}{\partial p(x)} \sum_z \left|p(z)\right| b(y-z) = \operatorname{sgn} p(x) b(y-x) \tag{75}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{abs,1}(x)} = \operatorname{sgn} p(x)\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x)\right\} \tag{76}$$

For $o_{abs,2}(y)$, we use an intermediate variable $w(z)$:

$$w(z) = p(z) * b(z) = \sum_y p(y) b(z-y) \tag{77}$$

so that the object can be written as:

$$o_{abs,2}(y) = |w(y)| \tag{78}$$

and the derivative of the object with respect to this variable is:

$$\frac{\partial o_{abs,2}(y)}{\partial w(z)} = \operatorname{sgn} w(y) I(y-z) \tag{79}$$

With the chain rule, we find

$$\frac{\partial o_{abs,2}(y)}{\partial p(x)} = \sum_z \frac{\partial o_{abs,2}(y)}{\partial w(z)} \frac{\partial w(z)}{\partial p(x)} = \sum_z \operatorname{sgn} w(y) I(y-z) b(z-x) = \operatorname{sgn} w(y) b(y-x) \tag{80}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{abs,2}(x)} = \left\{ \left[\sum_k r_k(x) \otimes s_k(x) \right] \operatorname{sgn}\left[p(x) * b(x) \right] \right\} \otimes b(x) \tag{81}$$

2.3.6 Positivity through Bounds Constraints

The final method of implementing positivity is through the use of an optimizer with bounds constraints. We have already touched upon this parameterization in conjunction with the smoothness constraint (see (50)), which is repeated here:

$$o_{bc}(x) = p(x) * b(x) \tag{82}$$

In this case, we cannot change the application order of smoothness and positivity, since the positivity is handled by the optimizer directly and is applied before smoothness. Care must be taken that $b(x)$ is positive.

The above expression is used to form $g_k(x)$ and compute the likelihood. As described in (25), the gradient residual function requires the partial derivative of this expression with respect to the estimated object parameters. This derivative is then:

$$\frac{\partial o_{bc}(y)}{\partial p(x)} = \frac{\partial}{\partial p(x)} \sum_z p(z)\ b(y-z) = b(y-x) \tag{83}$$

and substitution into (25) gives the likelihood gradient as

$$\frac{\partial L}{\partial p_{bc}(x)} = \left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x) \tag{84}$$

2.3.7 Gradient Summary – Object Parameters

Table 3 lists the various parameterizations along with their likelihood gradients:

Table 3: Likelihood gradient equation summary

Description	Parameterization (eq)	Likelihood Gradient (wrt. Object) (eq)
Square 1	$p^2(x)*b(x)$ (56)	$2\,p(x)\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x)\right\}$ (59)
Square 2	$[p(x)*b(x)]^2$ (57)	$2\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right][p(x)*b(x)]\right\} \otimes b(x)$ (61)
Exp 1	$\exp p(x)*b(x)$ (62)	$\exp p(x)\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x)\right\}$ (65)
Exp 2	$\exp[p(x)*b(x)]$ (63)	$\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] o(x)\right\} \otimes b(x)$ (70)
Abs 1	$\lvert p(x)\rvert*b(x)$ (71)	$\operatorname{sgn} p(x)\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x)\right\}$ (76)
Abs 2	$\lvert p(x)*b(x)\rvert$ (72)	$\left\{\left[\sum_k r_k(x) \otimes s_k(x)\right] \operatorname{sgn}[p(x)*b(x)]\right\} \otimes b(x)$ (81)
Bounds	$p(x)*b(x)$ (82)	$\left[\sum_k r_k(x) \otimes s_k(x)\right] \otimes b(x)$ (84)

2.4 Point Spread Function Parameterization

We now leave discussion of the object and turn to the point spread functions. In blind deconvolution, the PSF parameters are estimated along with the object parameters. Positivity and unit-energy are two constraints that are almost always employed. In addition, some correlation between pixels is also desirable in light of the physics that are present. (The aperture mask will

manifest itself as a convolution in the spatial domain.) Just as with the object, we have an expression for the likelihood gradient with respect to the PSF parameters which involve a residual term (from the noise model) and the partial derivative of the PSF intensities with respect to the PSF parameters. In this section we discuss one parameterization of the PSF that is often used and which achieves the above mentioned constraints of positivity, unit-energy, and pixel correlation.

2.4.1 Phase Errors at the Aperture, Zernike Expansion

This parameterization method directly leverages the fact that there are phase errors induced by the atmosphere in the near-field of the imaging system. Additionally, the optical parameters of the aperture such as cutoff frequency (diffraction limit) and central obscuration are naturally modeled. To further limit the degrees of freedom, the phase distortions are modeled (parameterized) as a superposition of classical optical aberrations. For example, tip, tilt, defocus, astigmatism, and so forth, with the set of PSF parameters $\{q_m(n)\}$ corresponding to the weighting of the n^{th} aberration in the m^{th} frame. Zernike polynomials are used to describe the various phase aberrations.

The point spread function for an incoherent imaging system is given by the magnitude-squared of the inverse Fourier transform $\left(\Im^{-1}[\cdot]\right)$ of the pupil function. Representing the pupil function for the m^{th} frame by $H_m(u)$, its inverse transform by $h_m(y)$, and a superscript "*" to denote the complex conjugate, we have the following relations for computing the PSFs from the pupil functions:

$$s_m(y) = \left|h_m(y)\right|^2 = h_m(y)h_m^*(y) \tag{85}$$

$$h_m(y) = \Im^{-1}\left[H_m(u)\right] = \frac{1}{N^2}\sum_u H_m(u)\exp(i\omega_0 ux) \tag{86}$$

As previously mentioned, we represent the phase of the complex pupil function as a superposition of Zernike polynomial functions, the set of which we denote as $\{\Phi_n(u)\}$ with weights $\{q_m(n)\}$. The magnitude of the pupil function $\left(\left|H_m(u)\right|\right)$ is assumed known, and typically this is just a mask with values of 1 to represent the pupil and 0 to represent the central obscuration and edges of the aperture. The pupil function is then:

$$H_m(u) = \left|H_m(u)\right| \exp\left[i \sum_n q_m(n) \Phi_n(u) \right] \tag{87}$$

Along with the object, the above expressions are used to form $g_k(x)$ and compute the likelihood. As described in (30), the gradient residual function requires the partial derivative of $s_m(y)$ expressions with respect to the estimated PSF parameters. Those PSF parameters are buried somewhat deep, but the chain rule will see us through. Using "c.c." to represent the complex conjugate of the previous terms,

$$\frac{\partial s_m(y)}{\partial q_k(j)} = \frac{\partial}{\partial q_k(j)} h_m(y) h_m^*(y) = h_m^*(y) \frac{\partial h_m(y)}{\partial q_k(j)} + c.c. \tag{88}$$

We now focus on the derivative of $h_m(y)$ via the chain rule:

$$\frac{\partial h_m(y)}{\partial q_k(j)} = \sum_r \sum_u \frac{\partial h_m(y)}{\partial H_r(u)} \frac{\partial H_r(u)}{\partial q_k(j)} \tag{89}$$

The first term is:

$$\frac{\partial h_m(y)}{\partial H_r(u)} = \frac{\partial}{\partial H_r(u)} \frac{1}{N^2} \sum_v H_m(v) \exp(i\omega_0 vy) = \frac{1}{N^2} \exp(i\omega_0 uy) I(m-r) \tag{90}$$

The second term is:

$$\frac{\partial H_r(u)}{\partial q_k(j)} = \frac{\partial}{\partial q_k(j)} \left|H_r(u)\right| \exp\left[i \sum_n q_r(n) \Phi_n(u) \right] = H_r(u) i \Phi_j(u) I(r-k) \tag{91}$$

Combining, the derivative of $h_m(y)$ is:

$$\begin{aligned} \frac{\partial h_m(y)}{\partial q_k(j)} &= \sum_r \sum_u \frac{1}{N^2} \exp(i\omega_0 uy) I(m-r) H_r(u) i \Phi_j(u) I(r-k) \\ &= \frac{i}{N^2} \sum_u H_k(u) \Phi_j(u) \exp(i\omega_0 uy) I(m-k) \end{aligned} \tag{92}$$

Substitution into (88) gives us the PSF derivative we seek:

$$\frac{\partial s_m(y)}{\partial q_k(j)} = \frac{i h_m^*(y)}{N^2} \sum_u H_k(u) \Phi_j(u) \exp(i\omega_0 uy) I(m-k) + c.c. \tag{93}$$

One last bit of simplification is in order before further substitution. Consider a general complex number can be represented as (a+ib). We note that our derivative expression is of the form $i(a+ib)+c.c = ia - b + (-ia - b) = -2b$, bringing our final PSF derivative to:

$$\frac{\partial s_m(y)}{\partial q_k(j)} = -2\,\mathrm{Im}\left\{ \frac{h_m^*(y)}{N^2} \sum_u H_k(u) \Phi_j(u) \exp(i\omega_0 uy) I(m-k) \right\} \tag{94}$$

This is then substituted into (30) to give the likelihood gradient as:

$$\frac{\partial L}{\partial q_k(j)} = -2 \sum_m \sum_y r_m(y) \otimes o(y)\, \mathrm{Im}\left\{ \frac{h_m^*(y)}{N^2} \sum_u H_k(u) \Phi_j(u) \exp(i\omega_0 uy) I(m-k) \right\} \tag{95}$$

We note that $\Phi_j(u)$ is a real factor and can be factored out, while the correlation is also real and can be inserted into the Im{} operator. Recognizing the summation over y is an inverse Fourier transform, our simplified likelihood gradient is finally:

$$\frac{\partial L}{\partial q_k(j)} = -2 \sum_u \Phi_j(u)\, \mathrm{Im}\left\{ H_k(u) \Im^{-1}\left\{ h_k^*(y) \left[r_k(y) \otimes o(y) \right] \right\} \right\} \tag{96}$$

2.5 Likelihood Maximization

With the expression for the likelihood function, the desire is to find the object and PSF parameters yielding the maximum value of that function. In order to identify the maximum likelihood, we turn to numerical optimization techniques. There is no “best” algorithm for searching this likelihood space, unfortunately. The fact we are able to analytically define the likelihood gradients, however, greatly improve the pace at which we can proceed and allows the use of the more powerful techniques.

From the standpoint of the solution, it should not matter what algorithm is used. After all, the location of the peak doesn’t depend on how we traverse the (hyper) surface (assuming a global peak, of course). However, since we care about finding the peak – or something reasonably close to it – in a timely fashion, we choose to use a relatively standard approach for

this type of problem. By and large, the two most commonly used approaches are conjugate gradient and limited memory quasi-Newton methods.

In our case, we require an optimizer that implements bounds constraints and leverage the work by academia with the use of the L-BFGS-B package from Northwestern University.[48] From their description:

> L-BFGS-B is a limited memory algorithm for solving large nonlinear optimization problems subject to simple bounds on the variables. It is intended for problems in which information on the Hessian matrix is difficult to obtain, or for large dense problems. L-BFGS-B can also be used for unconstrained problems, and in this case performs similarly to its predecessor, algorithm L-BFGS (Harwell routine VA15).

Additionally, while quasi-Newton methods require storage of additional values, they also tend to have faster convergence rates than the conjugate gradient methods.[49] We set the number of BFGS corrections that are saved to 10 (in the middle of the recommend range) but did not pursue evaluation at other settings given the success of reconstructions at that setting.

One additional consideration with numerical optimization is that of scaling. A problem is poorly scaled if the changes in the function value f(x) are very much larger for changes in x in a certain direction compared to another direction. For example, in a two dimensional function, a unit change along each dimension causes function values of 10 and 10^6. This arises in our case because we are searching for object parameters as well as the PSF parameters. Typical object parameter values could range upwards to several thousand, while the coefficients for a Zernike polynomial expansion are typically between zero and one. Discussions with other researchers indicate that the consensus is to normalize the data before searching and then scale the resultant object estimate. The common scale factor is the maximum pixel value contained within the ensemble of data. Consequently, the object parameter values are significantly better matched to those of the PSF parameters. Conveniently, quasi-Newton methods are also quite insensitive to any residual scaling issues.[49]

One additional technique that may be useful for more scale-sensitive search algorithms is to alternate the type of parameters being searched. For example, some number of iterations where the object parameters are held constant and PSF parameters are adjusted followed by some

number of iterations where the PSF parameters are held constant and object parameters are adjusted, then repeating the cycle.

In practice, the L-BFGS-B code we used requires the computation of the likelihood function and gradients once per iteration. In other words, the "current" object and PSF parameters are "rendered" into object and PSF estimates and convolved to generate the various noiseless image estimates. The specific choice of noise distribution (Poisson, Gaussian, or Spatially Varying Gaussian) dictates the likelihood and gradient equations, the evaluations of which are provided to the optimizer. In this work, we selected the Gaussian noise model for all reconstructions.

The iterative approach mandates an initial estimate for the object and PSF parameters. In this work, we set the initial object estimate to the average of the images in the ensemble after their centroids had been aligned. We then used the inverse of the parameterization expressions in order to initialize the actual parameters. Since the inverse of the smoothness constraint is ill-defined, this part of the inverse logic was omitted. The initial PSF parameters (Zernike polynomial coefficients) were set at zero with the exception of the tip and tilt terms (i.e., the vertical and horizontal shift) for each frame. These parameters were initialized based on the centroid of each frame.

2.6 Simulation of Data

2.6.1 Atmospheric turbulence

Much of the accuracy of the simulated imagery is a result of several decades of work by many researchers. We primarily leverage the work by Noll [50] who described a modal expansion of the phase perturbations due to light propagation through an inhomogeneous atmosphere. Zernike polynomials form an orthonormal basis for functions defined on the unit circle, which is especially *apropos* given that most optical systems have circular apertures. Using Zernike polynomials, we form a series expansion to describe the phase of the aberrated wavefront across an aperture of radius R as:

$$\psi(R\rho,\theta) = \sum_{i=1}^{N} a_i Z_i(\rho,\theta) \tag{97}$$

Because of the underlying Gaussian distributed nature of $\psi(R\rho,\theta)$ at every point in space, the random variables a_i are also Gaussian distributed with some covariance matrix Γ_a. The elements of the covariance matrix of these coefficients have been calculated for index of refraction fluctuations characterized by the Kolmogorov power spectral density and are given by: [50]

$$\overline{a_i a_j} = 0.0072\left(\frac{D}{r_0}\right)^{5/3}(-1)^{(n_i+n_j-2m_i)/2}\left[(n_i+1)(n_j+1)\right]^{1/2}\pi^{8/3}\delta_{m_i m_j} \times\frac{\Gamma(14/3)\Gamma[(n_i+n_j-5/3)/2]}{\Gamma[(n_i-n_j+17/3)/2]\Gamma[(n_j-n_i+17/3)/2]\Gamma[(n_i+n_j+23/3)/2]}, \text{ for even (i-j)} \quad (98)$$

$$\overline{a_i a_j} = 0, \text{ for odd (i-j)} \quad (99)$$

In order to generate a vector of random coefficients which have this covariance, we perform a Cholesky factorization of the covariance matrix Γ_a into the product of two square matrices:

$$\Gamma_a = \mathbf{R}\mathbf{R}^T \quad (100)$$

We form a vector **b** of random draws of a zero mean, unit variance, Gaussian random variable, which yields the covariance of **b** as the identity matrix:

$$E\{\mathbf{b}\mathbf{b}^T\} = \mathbf{I} \quad (101)$$

Now use **b** to generate a new vector **a** as:

$$\mathbf{a} = \mathbf{R}\mathbf{b} \quad (102)$$

We confirm the covariance of this vector **a** matches the desired covariance:

$$E\{\mathbf{a}\mathbf{a}^T\} = E\{\mathbf{R}\mathbf{b}(\mathbf{R}\mathbf{b})^T\} = E\{\mathbf{R}\mathbf{b}\mathbf{b}^T\mathbf{R}^T\} = E\{\mathbf{R}\mathbf{R}^T\} = \Gamma_a \quad (103)$$

Although the covariance expression is rather complicated, it is a fairly straightforward process to obtain coefficients that will generate statistically accurate phase variations (in the covariance sense). The resultant phase screen forms the basis of a point spread function as outlined in the first part of section 2.4.1. This process is repeated for every frame of imagery to be simulated, yielding a set of independent atmospheric realizations corresponding to Kolmogorov turbulence.

2.6.2 Sampling

Care must be taken to accurately simulate the imaging process, which is complicated by the fact the simulation is conducted in a discrete environment. There are three planes of interest with their associated pixilated arrays: 1) object, 2) pupil, and 3) image planes. There are two choices for the basis of the sampling: 1) detector pixel spacing and 2) object pixel spacing. Associated with these are the angular sample spacings at the detector and object.

For a linear sample spacing (pixel pitch) *d* on the detector and effective system focal length *f*, the angular sample spacing at the detector is:

$$\Delta\theta_d = \frac{d}{f} \tag{104}$$

The real-world object has an extent W, and it's a simple matter to define the pixel size (i.e., the sample spacing) in the object plane for an array size of N_o pixels as:

$$\Delta x_o = \frac{W}{N_o} \tag{105}$$

As seen from the imaging system pupil at a range of *R*, the object pixels have an angular extent of

$$\Delta\theta_o = \frac{\Delta x_o}{R} = \frac{W}{RN_o} \tag{106}$$

It is very likely that $\Delta\theta_d \neq \Delta\theta_o$, which mandates a choice of which sample spacing to use in the image formation process. If the detector pixel spacing is used as the basis, then the object must be resampled prior to use (in order to adjust the angular sampling at the object). If the detector pixel spacing is used as the basis, then the image must be resampled as a final step (to provide the sampling dictated by the detector). The resampling in each case will entail either 1) averaging across several smaller pixels (and dealing with partial pixel effects) or 2) interpolation, depending on choice of reference as shown in Table 4.

Additionally, the choice of basis for the angular sample spacing (detector or image) has ramifications for the sampling of the pupil function. If the detector angular spacing is used, the pupil diameter (in pixels within the array) is conveniently constant. If the object angular spacing is used, the pupil diameter in pixels will be a function of range to the object. The constant pupil

size is attractive for performance optimization (as well as simplicity), but we recommend choosing the smaller sample spacing as the reference.

Table 4: Sampling reference ramifications

	Sampling Reference	
	Detector/Image	Object
Ref Spacing < Other spacing	Interpolate object 1st	Average image last
Ref Spacing > Other Spacing	Average object 1st	Interpolate image last
Pupil Diameter (pixels)	Constant	Changes with range

Lastly, we note that the optical system with aperture diameter D effectively yields a low-pass filtering operation with a hard cutoff at a spatial frequency of D/λ [rad^{-1}], implying a minimum full-width at half-max (FWHM) resolution of λ/D [rad]. Nyquist sampling considerations would have the simulation angular sample spacing not to exceed half that or:

$$\Delta\theta_{Nyquist} = \frac{\lambda}{2D} \tag{107}$$

At this point, we proceed with the assumption that the simulation will use the angular sample spacing of the object, and will later revisit the case where the detector spacing is chosen as the reference.

Neglecting brightness effects and assuming we are in the “far-field” (i.e., the Fraunhofer region), Fourier optics tells us that the field amplitude at the pupil is a scaled (in lateral extent) version of the Fourier transform of the amplitude at the object plane.[51] Similarly, the field amplitude at the image plane is a scaled version of the Fourier transform of the amplitude at the pupil. The sampling at each of the planes must take this scaling into account. The following development uses one-dimensional notation for simplicity. We also regard a representative, single-lens imaging system in which “pupil” and “aperture” are interchangeable.

It is important throughout this that the reader separate the notion of “image/pupil” from “spatial/Fourier” domain. While this is typical in the computer vision fields, it will not serve us here. In point of fact, the aperture in a real imaging system (e.g., mirror or lens) has a spatial extent which must be considered. Further, it also has information that *after* a Fourier transform

must be viewed as *spatial* information, which is suggestive of being in the Fourier domain to start.

In the following, x_i is a coordinate in the image plane, x_p is a coordinate in the pupil plane, x_o is a coordinate in the object plane, λ is the wavelength and f is the effective focal length (distance from pupil to detector).

Denoting the Fourier transform of the pupil amplitude by $P(u)$ and neglecting brightness scaling factors, we start with the intensity at the image plane from Fourier optics theory as:

$$I(x_i) = \left|P(u)\right|^2\Big|_{u=\frac{x_i}{\lambda f}} = \left|P\left(\frac{x_i}{\lambda f}\right)\right|^2 \tag{108}$$

By inspection, we have an expression relating the Fourier domain sample spacing in the pupil to the spatial domain sample spacing in the image as:

$$\Delta u = \frac{\Delta x_i}{\lambda f} \tag{109}$$

The aperture is assumed to be circular with physical extent D and is represented in the pupil array by a circle with a diameter of N_p pixels. Thus, the spatial domain sample spacing in the pupil plane is:

$$\Delta x_p = \frac{D}{N_p} \tag{110}$$

We assume the pupil array is of size N_A, and therefore the spatial extent of the entire pupil array is:

$$W_p = N_A \Delta x_p \tag{111}$$

We also know that the Fourier domain sample spacing in the pupil is equal to the inverse of this spatial extent:

$$\Delta u = \frac{1}{W_p} \tag{112}$$

Equating (109) and (112), we have the following relation:

$$\frac{\Delta x_i}{\lambda f} = \frac{1}{N_A \Delta x_p} \tag{113}$$

Rearranging and combining with the spatial sampling of the pupil (110), we see the angular sample spacing of the image is:

$$\Delta\theta_i = \frac{\Delta x_i}{f} = \frac{\lambda}{N_A \Delta x_p} = \frac{\lambda}{D}\frac{N_p}{N_A} \tag{114}$$

For the purpose of the simulation, we equate the angular sample spacings of the object and image and also assume equal sized arrays for convenience ($N_o = N_A$). Recall that if the actual sample spacings are different, one is adjusted either prior or after image formation.

$$\frac{\lambda}{D}\frac{N_p}{N_A} = \frac{W}{RN_A} \tag{115}$$

We can rearrange this expression to find the size of the aperture – in pixels – that is required to accurately simulate this scenario:

$$N_p = \frac{D}{\lambda}\frac{W}{R} \tag{116}$$

In other words, the number of samples across the aperture is the product of the optical cutoff frequency and the angular extent of the scene. For example, to accurately simulate the image of a 6 m scene at a range of 500 km as seen by the 1.6 m telescope on Haleakala in the visible wavelength (λ=0.5 μm), one would need a circle of diameter (1.6*6/0.5e-6/500e3) ~ 38 pixels to represent the pupil.

Looking back at the original image expression in (108), we can view the transform of the image – the optical transfer function when the source is a point source – as proportional to the autocorrelation of the pupil amplitude. Recall that a correlation can be computed via the Fourier domain as:

$$f(x) \otimes g(x) \xleftrightarrow{\Im} F(u)G^*(u) \tag{117}$$

When the functions *f* and *g* are equal (i.e., an autocorrelation), the Fourier side of the relation reduces to $|F(u)|^2$, which matches the form of the imaging equation. To avoid wraparound effects from that autocorrelation corrupting the optical transfer function, the array size N_A should be at least twice the pupil extent (N_p).

$$N_A \geq 2N_p \tag{118}$$

An alternate view of this is that the Nyquist criterion is being preserved as substitution into (114) gives

$$\Delta\theta_i = \frac{\lambda}{D}\frac{N_p}{N_A} \le \frac{\lambda}{D}\frac{N_p}{2N_p} \Rightarrow \Delta\theta_i \le \frac{\lambda}{2D} \tag{119}$$

Given the above sampling relations are maintained, an image can be calculated with (1) with possible averaging if the detector angular sample spacing differs from the reference.

If, on the other hand, the angular sample spacing of the detector (and therefore the final, desired simulated image) is used as the reference, then the object must be resampled first. This effectively changes the value of the object extent (W) in the above development with the result that the "adjusted" object spacing will match that of the detector. (This is, in fact, exactly what happens in a real-world imaging system with a fixed field of view, in which the lateral extent of the scene within the field of view naturally changes as a function of range.)

Using the detector sample spacing as the reference (and thus $\Delta\theta_i = \Delta\theta_d$), (114) is manipulated to find the expression for the number of pixels across the pupil. Similar to the previous expression, the number of pixels across the pupil is seen to be the product of the optical cutoff frequency and the angular extent of the image.

$$N_p = \Delta\theta_d N_A \frac{D}{\lambda} \tag{120}$$

In the work here, we sidestep the majority of the sampling issues completely and simply set $N_p = N_A/2$, which yields critical sampled imagery and assume the detector angular sample spacing matches that of the object. Resampling of the object or image is not necessary. The specific values of scene extent, range, etc. is not unique, of course, and so specific values are immaterial to our purposes.

2.6.3 Noiseless Image Formation

We assume narrow-band, spatially-invariant, incoherent imaging. As a result, the noiseless image is the convolution of the object and system point spread function as denoted in (1). For the array sizes typically involved, this convolution is actually calculated via multiplication in the Fourier domain. Due to the discrete nature, this technically achieves a so-called "circular convolution."[52] However, we set the support of the pupil function at half that of the array size to avoid wraparound effects (i.e., yielding a "normal" convolution).

The notion of narrow-band imaging is not a well-defined notion. How “narrow” must the band be? Solar imaging, for instance, will commonly use extremely narrow filters, even less than a single angstrom wide. Some applications, are content with a hundred nanometers. Still others desire to use most of the visible spectrum with color imagery.

As we have seen from the sampling development, the wavelength of illumination plays a role in the size of the intensity pattern being sampled. Another overwhelming aspect is that the object intensities are rarely the same in different bands. From the reconstruction side, MFBD assumes a single, common object, which certainly does not match reality for wide band imagery. One of the simple approaches for color imagery, is to separately reconstruct the red, green, and blue channels. Unfortunately, it is very difficult to register these separate estimates and color-fringing will result. Results are often less than satisfactory because there are many more unknowns than for narrowband data (a factor of three for RGB channels). A more sophisticated approach will leverage the fact that the atmospheric aberrations and thus the point spread functions scale with wavelength and are not, in fact, independent between the channels. This is known in the industry as “wavelength diversity.” While these multiple looks at the same atmospheric distortion improve things considerably for the atmospheric parameter estimates, multiple wavelength-dependent objects usually have to be estimated.

As far as simulation is concerned, simulating wide(er) band imagery is a matter of combining the results of several (possibly many) narrow-band simulations. As long as the wavelength is considered in the sampling, it comes down to the tradeoff between accuracy and run-time (i.e., number of bands to simulate and then combine). That said, however, one is often able to obtain significant improvement even though the narrowband assumption is violated. The individual circumstance will dictate if the errors inherent in such an approach are tolerable. In the work here, we use a single wavelength in all simulations and reconstructions.

2.6.4 Noise

Once a noiseless image g(x) has been simulated, noise is added to each pixel. The random arrival time of photons (and thus the brightness during a fixed exposure interval) is assumed to have a Poisson distribution with a mean rate equal to that of the pixel values in the noiseless image. (Recall the variance of a Poisson random variable is equal to this mean rate.) Additionally, noise with a Gaussian distribution with a constant variance is applied to simulate

the effects of detector read noise. In this work, we assume a read noise standard deviation of 10 counts and desire a peak signal to noise ratio. Thus, the "signal" is the peak value (P) of the noiseless image.

The total noise variance is the sum of the constituent noise variances: P for the photon noise and σ_d^2 for the detector read noise. It is common to specify the desired signal to noise ratio in units of dB (SNRdb), so we have:

$$SNR = 10^{SNRdb/10} \tag{121}$$

By the definition of the signal to noise ratio:

$$SNR = \frac{P}{\sqrt{P + \sigma_d^2}} \tag{122}$$

This can be solved to give an expression for the brightness P:

$$P = \frac{1}{2}\left(SNR^2 + SNR\sqrt{SNR^2 + 4\sigma_d^2}\right) \tag{123}$$

Once the peak value (P) is computed, we scale the entire noiseless image g(x) appropriately. The noise in each pixel is assumed independent and random draws are made against a Poisson random variable with mean g(x) and a zero-mean Gaussian random variable with variance σ_d^2 for a combined noise term of:

$$n_k(x) = Poisson(g_k(x)) + Gaussian(0, \sigma_d^2) \tag{124}$$

and the final noisy image is given by:

$$d_k(x) = g_k(x) + n_k(x) \tag{125}$$

2.7 Accuracy Metrics

2.7.1 RMS Error (Spatial Domain)

The first metric we consider is the root mean square (RMS) error in the spatial domain. Since these data were simulated, we have the luxury of knowing the true answer a "perfect" reconstruction process would have generated. One issue is that the random shift that is present in each of the actual reconstructions complicates the accurate computation of this metric. While the reconstructions can be correlated against the true image for alignment, this approach merely leverages an integer-pixel estimate of the shift. Since the simulated atmospheric distortions were

not limited to integer tip & tilt aberrations (i.e., shifts), there is no reason to expect that the sampled representation of the reconstructed object necessarily be aligned with the original samples of the true object. Here we switch notation slightly, and refer to the reconstructed object as $\hat{o}(x)$ with a "hat" and the true object as $o(x)$.

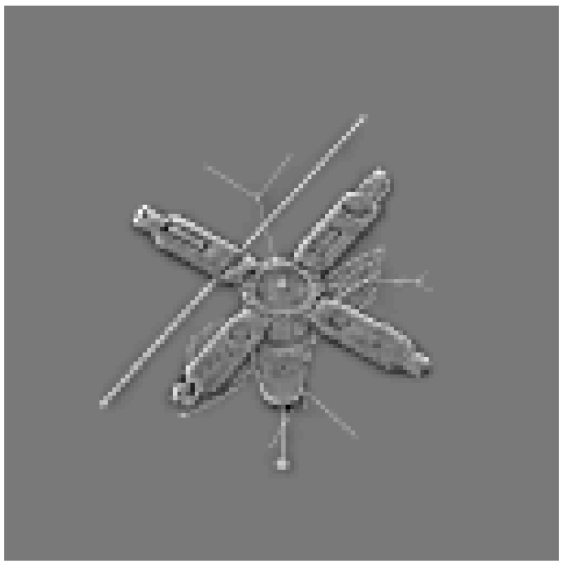


Figure 7: Example error of a single reconstruction shows a slight residual vertical shift

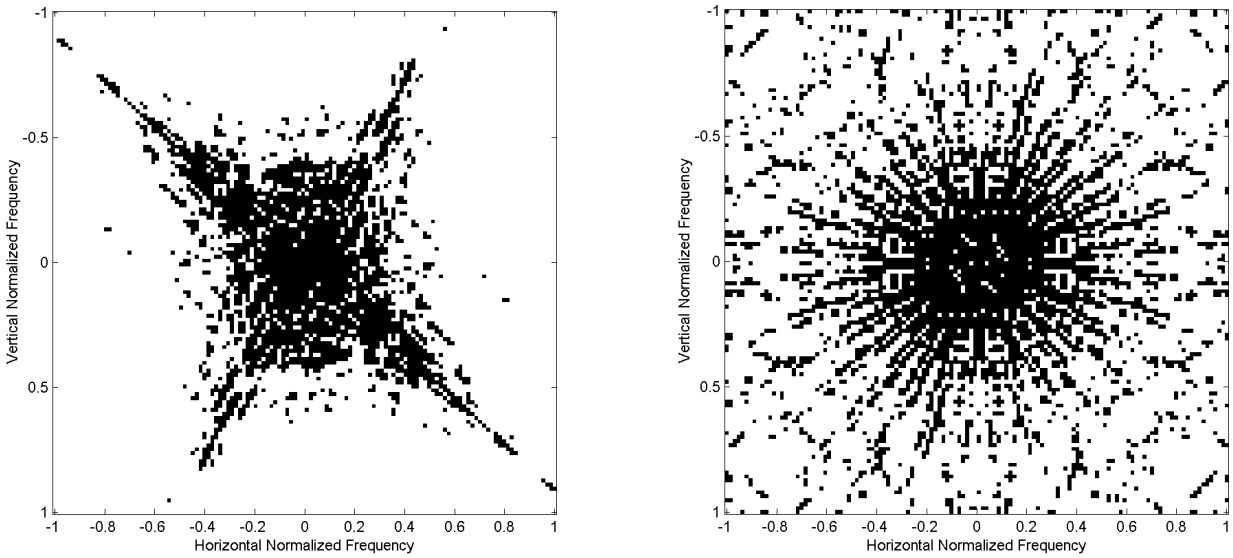
Figure 7 shows the error $\left[\hat{o}(x) - o(x)\right]$ of an example reconstruction after alignment by spatial correlation. A slight vertical shift is evident due to the white (negative values) at the top edge and black (positive values) at the bottom edge of the solar panels. In the aggregate, however, we expect similar errors to be present across all reconstruction methods, inflating all of them by approximately the same amount. As a gross figure of merit, we can compute the total RMS error as:

$$RMSE = \sqrt{\frac{1}{N_k N_x} \sum_{k,x} \left(\hat{o}_k(x) - o(x)\right)^2} \tag{126}$$

where N_k is the number of reconstructions, and N_x is the number of pixels in an image. The result is a scalar representing the reconstruction accuracy of each parameterization. Take note, however, that these results should not be compared across the two SNR cases since the object brightnesses are vastly different.

2.7.2 RMS Phase Error (Spatial Frequency Domain)

In addition to the RMS error in the spatial domain, we also consider the RMS phase error (RMSPE) computed in the spatial frequency domain. This metric emphasizes the shape of the reconstruction rather than the particular intensity levels. Additionally, we focus on the area of the spectrum that is meaningful to the true object by using an indicator mask $M(u)$ which is unity where the magnitude of the true spectrum is within 2 orders of magnitude of the peak magnitude and zero elsewhere. Figure 8 shows the spectral areas of interest for the two objects used in this study. The RMSPE metric is computed solely in these areas where the true object contains a reasonable level of signal.



(a) (b)

Figure 8: Indicator mask $M(u)$ of spectral locations of interest based on the true spectrum of (a) Satellite and (b) Spoke objects

At first blush, unfortunately, the phase is also corrupted by the random spatial domain shifts, as can be seen from the Fourier shift property: [51]

$$f(x - x_0) \xleftrightarrow{\mathfrak{I}} \exp(-j2\pi x_0 u) F(u) \tag{127}$$

Thus, a shift in the spatial domain manifests itself as a two-dimensional phase ramp in the frequency domain. Note the several black/white transitions in Figure 9(a) where the phase difference changes between $\pm\pi$. However, the signal to noise ratio of the frequency pixels around DC (i.e., the origin) is extremely high, and the phase ramp can be estimated from the first non-

zero frequency pixels in the vertical and horizontal directions. Of critical importance, however, is that this phase ramp characterizes any subpixel shift as well, and compensation provides insight into the phase error when the estimated and true objects are well aligned.

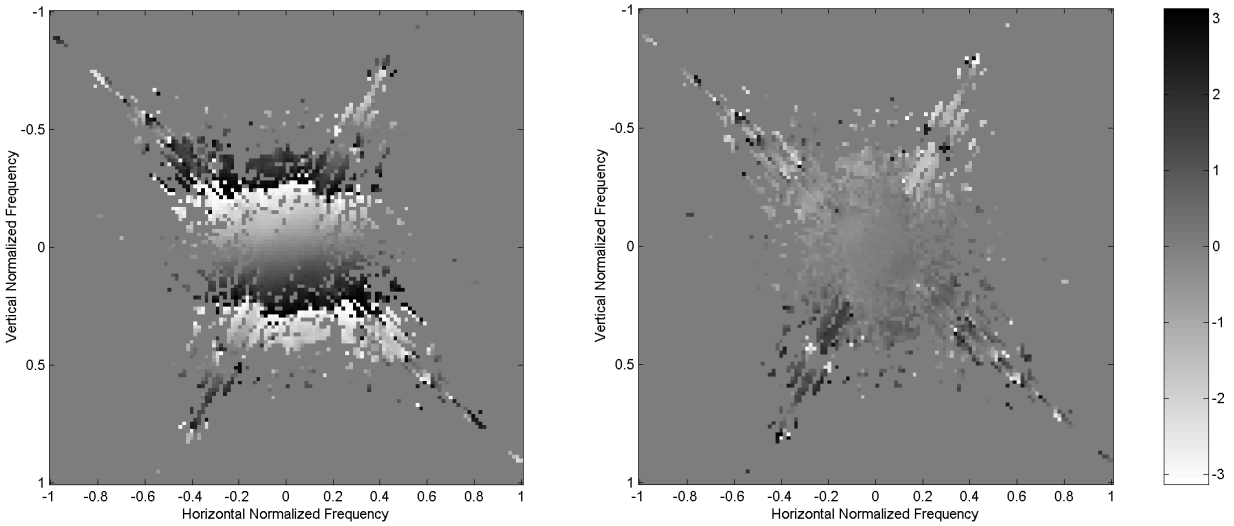


Figure 9: Example phase differences when the estimated and true objects are (a) misaligned and (b) aligned

Another aspect of this metric is the characterization of reconstructed resolution or level of detail, since resolution increases as accurate spatial frequency content increases. To maintain insight into the various spatial frequencies, we compute the metric within narrow annular regions of approximately constant spatial frequency magnitude. Using u as a 2D indexing variable (spatial domain coordinate), and the function $D(u)$ to denote the distance of that location from the origin, the r^{th} annular region is defined by $r < D(u) < r+1$. We compute the metric as:

$$RMSPE(r) = \frac{1}{N_k} \sum_k \left[\frac{\sum_{r<D(u)<r+1} \left(\arg \hat{O}_k(u) - \arg O(u) \right)^2 M(u)}{\sum_{r<D(u)<r+1} M(u)} \right]^{1/2} \tag{128}$$

In other words, we compute the RMS phase error within each annular region (subject to the masked area of interest) and then average across the reconstructions. The result is a curve for each algorithm relating the fidelity of reconstruction as a function of spatial frequency.

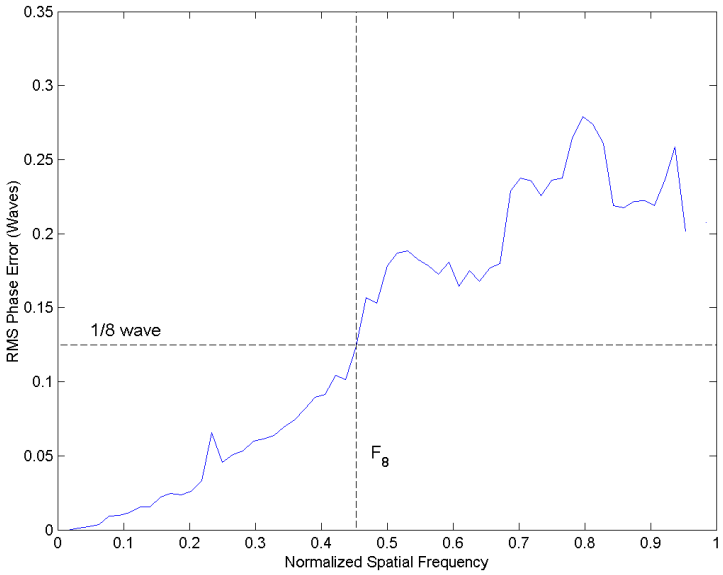


Figure 10: Example RMSPE curve relating fidelity of reconstruction to spatial frequency

Optical systems have been characterized in various ways over the years, and one of the fundamental measures of system quality is at what spatial frequency is the phase error greater than say, "1/8 of a wave". (A "wave" is simply an error of 2π radians.) Regarding our reconstruction process as part of the imaging "system", the RMSPE curves provide this information; the higher the frequency, the better. In our case, we normalize the frequency between 0 and 1, where 1 represents the highest frequency possible in our system r_{max} (the Nyquist frequency), and denote this metric by F_8 (for 1/8 wave error):

$$F_8 = \frac{\min_r RMSPE(r) > \frac{1}{8}(2\pi)}{r_{max}} \tag{129}$$

3 RESULTS

3.1 Simulation of image data

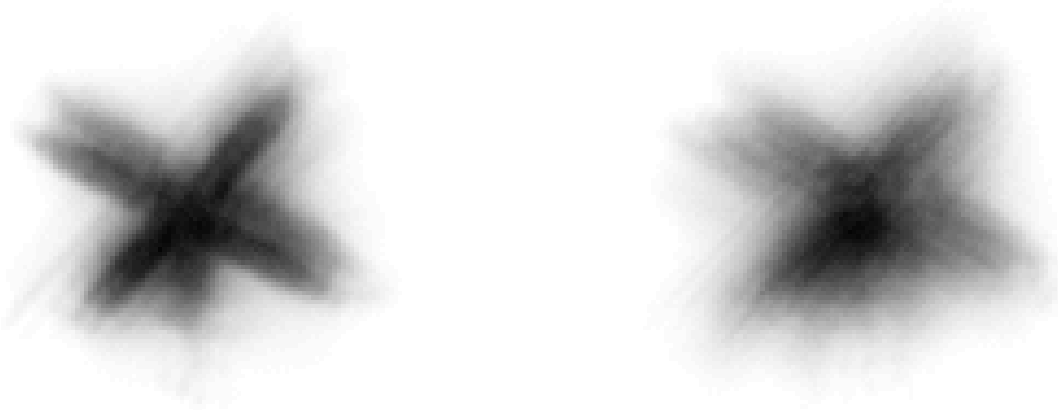
The image simulation process [53,54] leverages the fact that the primary, shape-distorting aberrations in an imaging system are due to phase errors in the near-field. For each image, the simulation draws a random realization of coefficients. These coefficients are distributed with a Kolmogorov spectrum which has been shown to model the statistics of the atmospheric distortions extremely well. These coefficients are then used to weight a set of basis functions described by Zernike polynomials [55]. The accumulation of weighted polynomials describes the phase distortions present at the aperture of the imaging system.

The turbulence "strength" – described by the relation of Fried's seeing parameter r_0 and the aperture diameter D – plays a critical role here as well. For all imagery, I used a D/r_0 ratio of 16 to reflect a common case experienced by the observatory on Halealaka (1.6 m telescope and 10 cm seeing). This is on the relatively poor side as will be seen in the imagery.

During the process of simulating the raw imagery, I was reminded of something Gonsalves said during a conference presentation: "Simulations are often doomed to success." While this study is primarily concerned with the performance improvements brought about by various parameterizations of positivity, we used well established theory in the simulation of data to minimize the chances of a false-success. In the final analysis, though, I did not want to recommend a parameterization if it would only work in a "sunny-day scenario" with a high signal to noise ratio (i.e., a bright object). As a result, we considered signal to noise ratios of 20 dB and 40 dB.

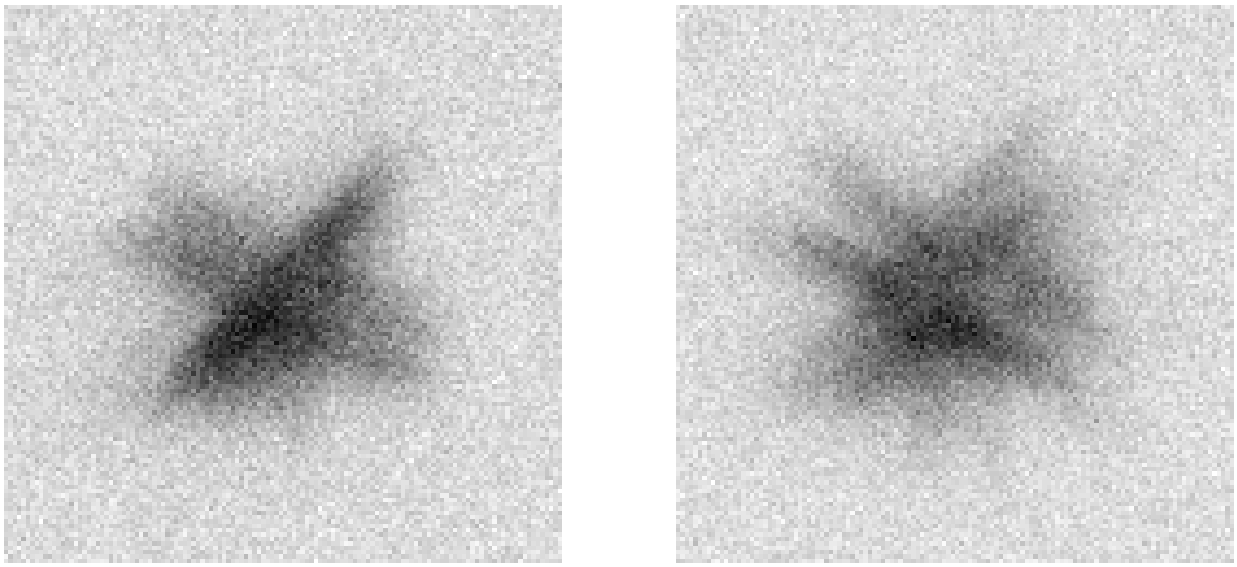
Gaussian read noise with a standard deviation of 10 counts was used for all imagery. In addition, signal-dependent, Poisson noise was applied to model the random arrival times of the photons during the effective integration time. Object brightness was set to yield a combined peak SNR of the atmospherically distorted signal of the aforementioned 40 dB or 20 dB.

An incoherent imaging system is assumed in the generation of the various point spread functions (PSFs) and resultant imagery [56]. Data were simulated for the two different objects shown in Figure 3. Each data set consisted of 1,600 independent realizations of the atmospheric turbulence. Figure 11 and Figure 12 show typical examples of the simulated raw imagery at various signal to noise ratios:



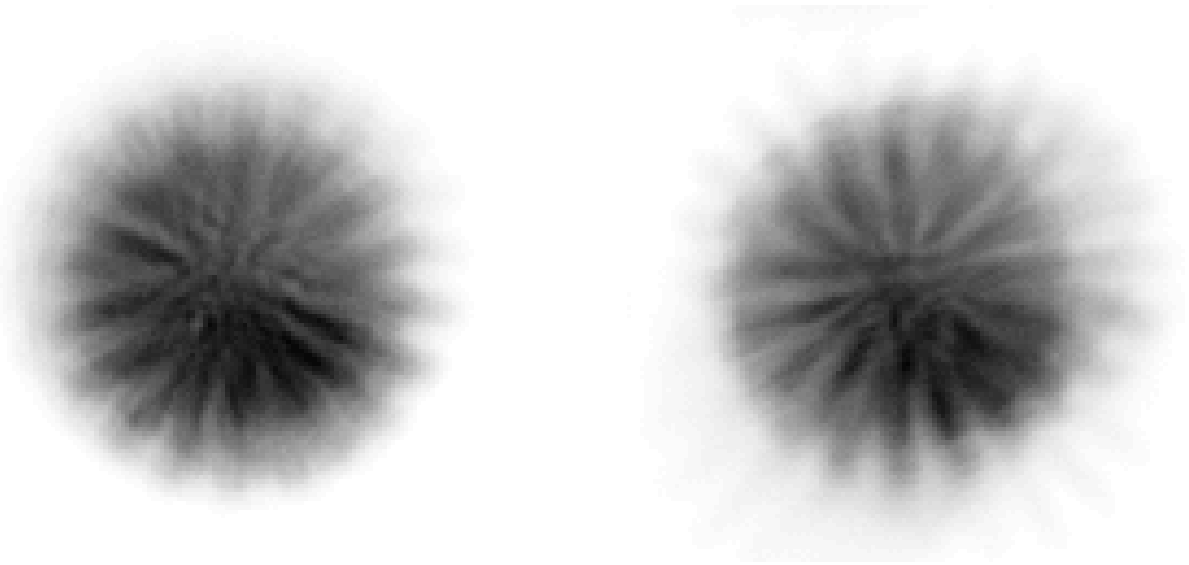
(a) (b)

Figure 11: Typical simulated imagery, Satellite object with SNR=40 dB



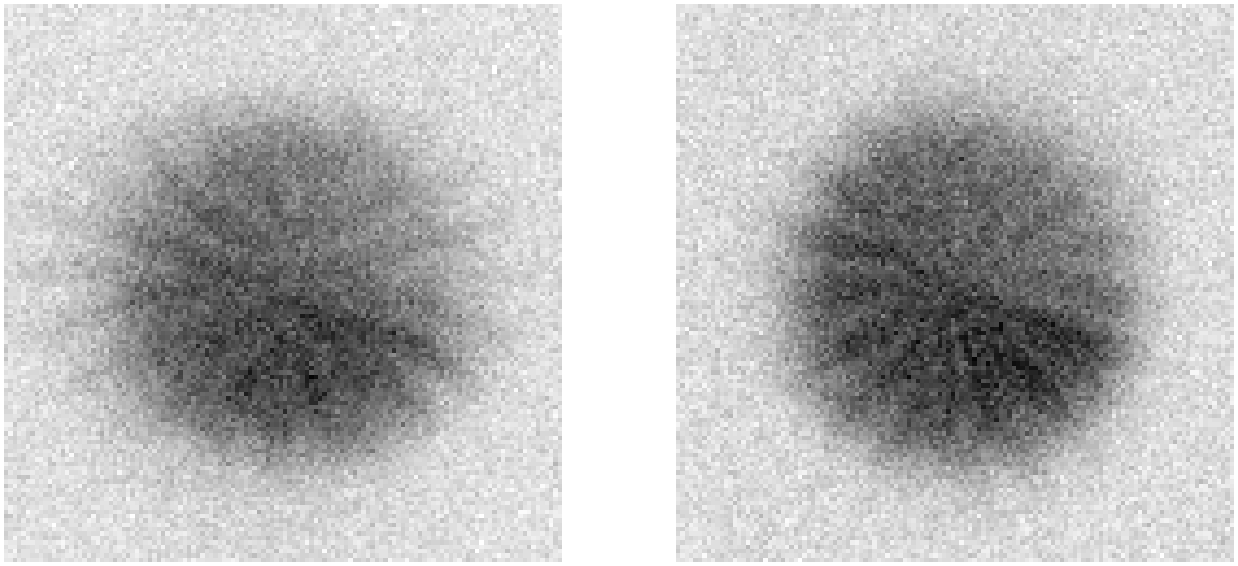
(a) (b)

Figure 12: Typical simulated imagery, Satellite object with SNR=20 dB



(a) (b)

Figure 13: Typical simulated imagery, Spoke object with SNR=40 dB



(a) (b)

Figure 14: Typical simulated imagery, Spoke object with SNR=20 dB

3.2 Implementation of constraints

Each of the equations in Table 3 was implemented in the reconstruction code. A great deal of effort was spent in the theoretical development to identify convolutions and correlations in the simplified expressions. For the array sizes involved, it is significantly faster to compute

these with the use of the fast Fourier transform. During implementation, further attention was spent in minimizing the number of calls to the transform routines, resulting in the extensive use of intermediate cache-like arrays. The transforms necessary for the PSFs ($s_k(x)$) or the residuals ($r_k(x)$) are immaterial to the purpose of this research and are excluded from the count.

Table 5 shows the number of object-related transforms and execution time required in the final implementation. The relative timing (as a percentage change) compared to that required by Square 1 and Square 2 are denoted by the two "% Ch" columns.

Table 5: Parameterization run-time metrics

Desc.	Render FFTs	Gradient FFTs Per Frame	Render Time (μs)	Gradient Time Per Frame (μs)	Combined Time Per 16 Frames (ms)	% Ch From Square1	% Ch From Square2	Rank
Square 1	2	1	1429 ± 9	1451 ± 10	24.6	0	-29%	4
Square 2	3	3	1765 ± 9	2057 ± 10	34.7	41%	0%	6
Exp 1	2	1	1578 ± 10	1413 ± 10	24.2	-2%	-30%	3
Exp 2	3	3	1946 ± 9	2121 ± 10	35.9	46%	3%	7
Abs 1	3	3	1453 ± 10	1069 ± 7	18.6	-25%	-46%	2
Abs 2	3	3	1796 ± 9	2017 ± 10	34.1	38%	-2%	5
Bounds	2	1	1246 ± 10	1037 ± 4	17.8	-28%	-49%	1

The timing data were estimated with repeated evaluations in conjunction with the computer's real-time clock. This is necessarily an asynchronous approach, of course, leading to slight differences in consecutive timing trials. The uncertainty quoted above is the 95% confidence interval about the estimate of the mean. Furthermore, the number of trials was adjusted such that that confidence interval was no more than 10 μs.

3.3 Image Reconstruction

With seven object parameterizations, two true objects, and two SNR settings, there are twenty-eight cases to consider. In all cases, a non-overlapping ensemble of 16 frames were used. In other words, frames 1-16 were used in the first reconstruction, frames 17-32 were used in the second, etc. With a data set of 1,600 frames, 100 reconstructions were generated for each case, and 150 iterations were used in the optimization. A smoothness constraint of 0.5 pixels and 2.0

pixels was used for the 40 dB and 20 dB SNR cases respectively. The initial object estimates were based on the centroid-aligned average of the raw frames in each ensemble.

The PSFs were parameterized by phase-only descriptions of the distortions at the aperture. These were expanded in a series expansion of 91 Zernike polynomial terms for each frame. The tip and tilt parameters were initialized based on the difference of the centroids of the initial object estimate and each raw frame. The remainder of the parameters were initialized to 0.

The resultant reconstructions were not required to be centered in the array. Indeed, the tip/tilt (horizontal and vertical shifts) required is part of the PSF estimation, and the exact position of the object centroid is relatively arbitrary. The relative misalignment between reconstructions, however, does pose a problem when comparing reconstructions. To mitigate this issue, all reconstructions in a set were aligned in the spatial domain by cross correlation using the true object as a reference.

The figures in the following sections show the average and the first individual reconstruction for each case. Visual assessment of quality in the spatial domain is subjective at best, and we are limited to broad comparisons with regard to perceived resolution and accuracy. Following these figures, however, we present RMS phase error plots, from which a quantitative resolution assessment is more easily made.

3.3.1 Satellite Reconstructions, 40 dB SNR

First we show the Satellite reconstructions at 40 dB SNR.

(a) (b)

Figure 15: Square 1 reconstructions, Satellite object, 40 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

Figure 16: Square 2 reconstructions, Satellite object, 40 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

Figure 17: Exp 1 reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 18: Exp 2 reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 19: Abs 1 reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 20: Abs 2 reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 21: Bounds reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

3.3.2 Satellite Reconstructions, 20 dB SNR

Next we show the Satellite reconstructions at the reduced SNR of 20 dB. We note additional blurring (compared to the 40 dB reconstructions).

(a) (b)

Figure 22: Square 1 reconstructions, Satellite object, 20 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

Figure 23: Square 2 reconstructions, Satellite object, 20 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

Figure 24: Exp 1 reconstructions, Satellite object, 20 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 25: Exp 2 reconstructions, Satellite object, 20 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 26: Abs 1 reconstructions, Satellite object, 20 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 27: Abs 2 reconstructions, Satellite object, 20 dB SNR

(a) set average, (b) individual reconstruction

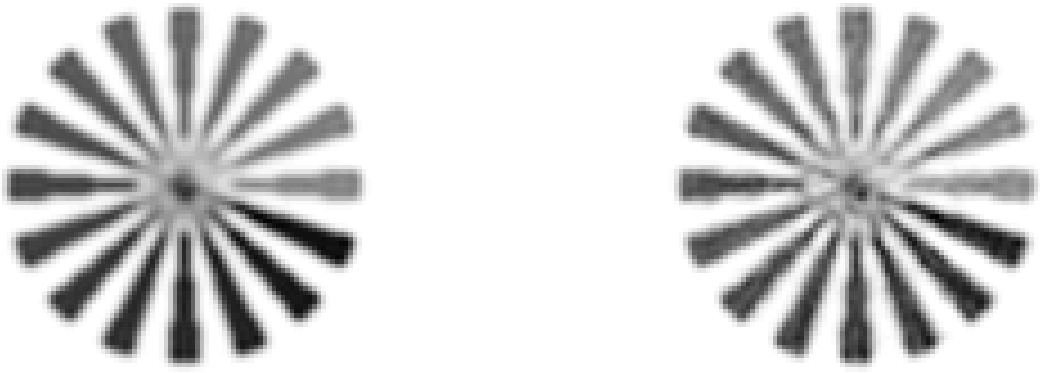
(a) (b)

Figure 28: Bounds reconstructions, Satellite object, 40 dB SNR

(a) set average, (b) individual reconstruction

3.3.3 Spoke Reconstructions, 40 dB SNR

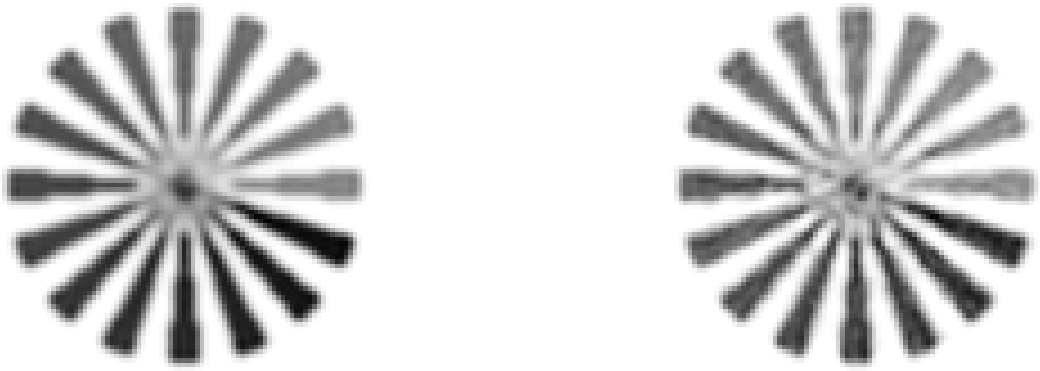
The next set of reconstructions illustrate performance on the Spoke object at 40 dB SNR.



(a) (b)

Figure 29: Square 1 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction



(a) (b)

Figure 30: Square 2 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 31: Exp 1 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 32: Exp 2 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

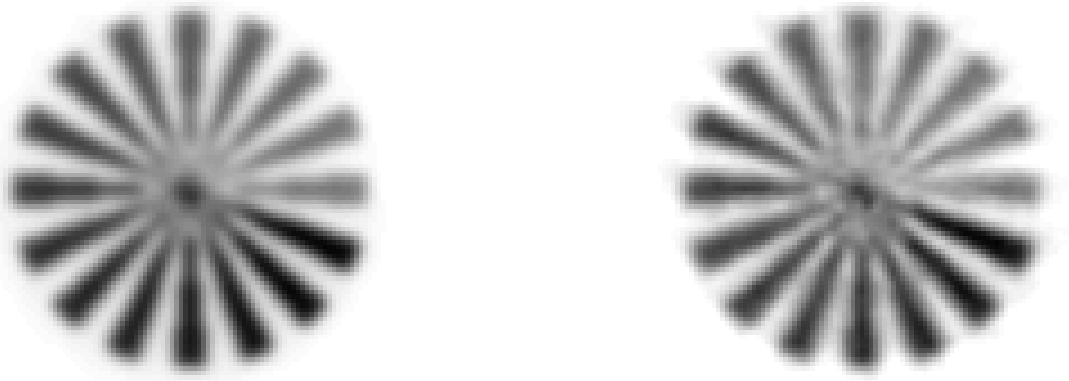
Figure 33: Abs 1 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 34: Abs 2 reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction



(a) (b)

Figure 35: Bounds reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

3.3.4 Spoke Reconstructions, 20 dB SNR

Next we show the Spoke reconstructions at the reduced SNR of 20 dB. As with the Satellite object, we note additional blurring (compared to the 40 dB reconstructions).

(a) (b)

Figure 36: Square 1 reconstructions, Spoke object, 20 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

Figure 37: Square 2 reconstructions, Spoke object, 20 dB SNR
(a) set average, (b) individual reconstruction

(a) (b)

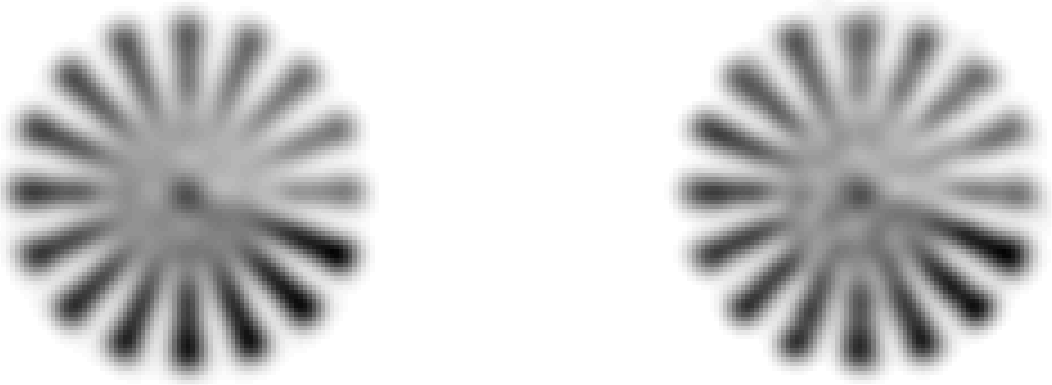
Figure 38: Exp 1 reconstructions, Spoke object, 20 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 39: Exp 2 reconstructions, Spoke object, 20 dB SNR

(a) set average, (b) individual reconstruction



(a) (b)

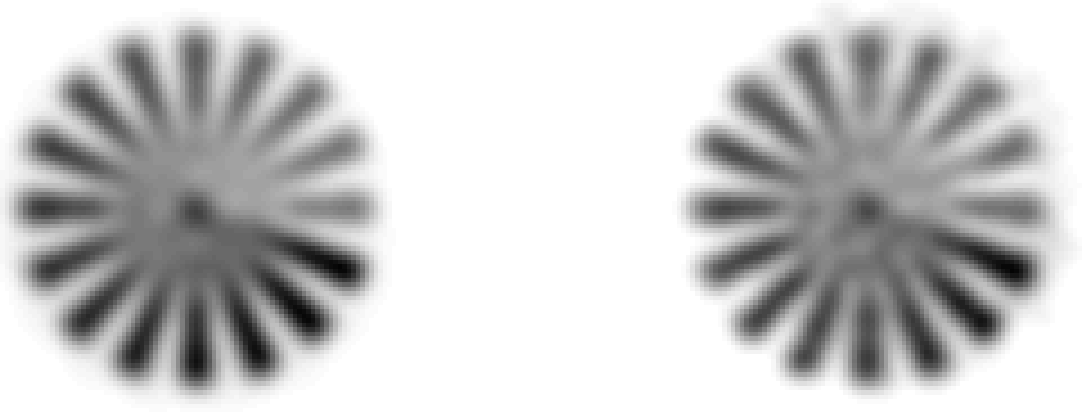
Figure 40: Abs 1 reconstructions, Spoke object, 20 dB SNR

(a) set average, (b) individual reconstruction

(a) (b)

Figure 41: Abs 2 reconstructions, Spoke object, 20 dB SNR

(a) set average, (b) individual reconstruction



(a) (b)

Figure 42: Bounds reconstructions, Spoke object, 40 dB SNR

(a) set average, (b) individual reconstruction

3.4 Observations and Metric Results

We offer the following broad observations regarding the spatial domain reconstructions:

- Most of the high-SNR average reconstructions look very similar, as one would hope.
- The Bounds reconstructions are substantially more blurry.
- The Square variants seem a bit better at reconstructing the fine details in the high-SNR individual reconstructions.
- The quality of the Exp variant reconstructions is noticeably lower for the low-SNR data.

However, we turn to the quantitative metrics (described in section 2.7) with which to further assess these parameterizations. Fundamentally, these are related to the true object and are possible by virtue of using simulated imagery. First we have the scalar metrics which are computed across all of the various reconstructions. Table 6 and Table 7 list the metrics for the Satellite and Spoke objects, respectively.

Table 6: Parameterization accuracy metrics, Satellite object

SNR (dB)	Parameterization	RMSE (counts)	F_8 (fraction of Nyquist)
40	Square 1	1001	0.47
40	Square 2	1000	0.47
40	Exp 1	1203	0.37
40	Exp 2	1236	0.36
40	Abs 1	1415	0.28
40	Abs 2	1230	0.34
40	Bounds	1489	0.30
20	Square 1	21.7	0.25
20	Square 2	21.2	0.25
20	Exp 1	24.2	0.22
20	Exp 2	27.3	0.21
20	Abs 1	25.5	0.21
20	Abs 2	22.7	0.23
20	Bounds	26.3	0.23

From these Satellite object metrics, we see three groups emerge. In order of reconstruction quality (highest to lowest), we have the Square variants, the Exp variants with Abs 2, and the Abs 1 with Bounds. Perhaps unexpectedly, there is not much variation within each family, with the exception of the two Abs implementations. Similar performance is seen on the Spoke object (Table 7), again with the Square family yielding the most accurate reconstructions.

Table 7: Parameterization accuracy metrics, Spoke object

SNR (dB)	Parameterization	RMSE (counts)	F_8 (fraction of Nyquist)
40	Square 1	2361	0.43
40	Square 2	2356	0.43
40	Exp 1	2735	0.28
40	Exp 2	2842	0.27
40	Abs 1	2733	0.30
40	Abs 2	2593	0.32
40	Bounds	3157	0.18
20	Square 1	48.2	0.26
20	Square 2	46.3	0.26
20	Exp 1	49.5	0.25
20	Exp 2	55.4	0.20
20	Abs 1	53.3	0.25
20	Abs 2	48.4	0.26
20	Bounds	56.9	0.18

Turning to the spatial frequency domain, the differences in each parameterization become a bit more obvious in the RMSPE curves, which highlight performance as a function of spatial frequency:

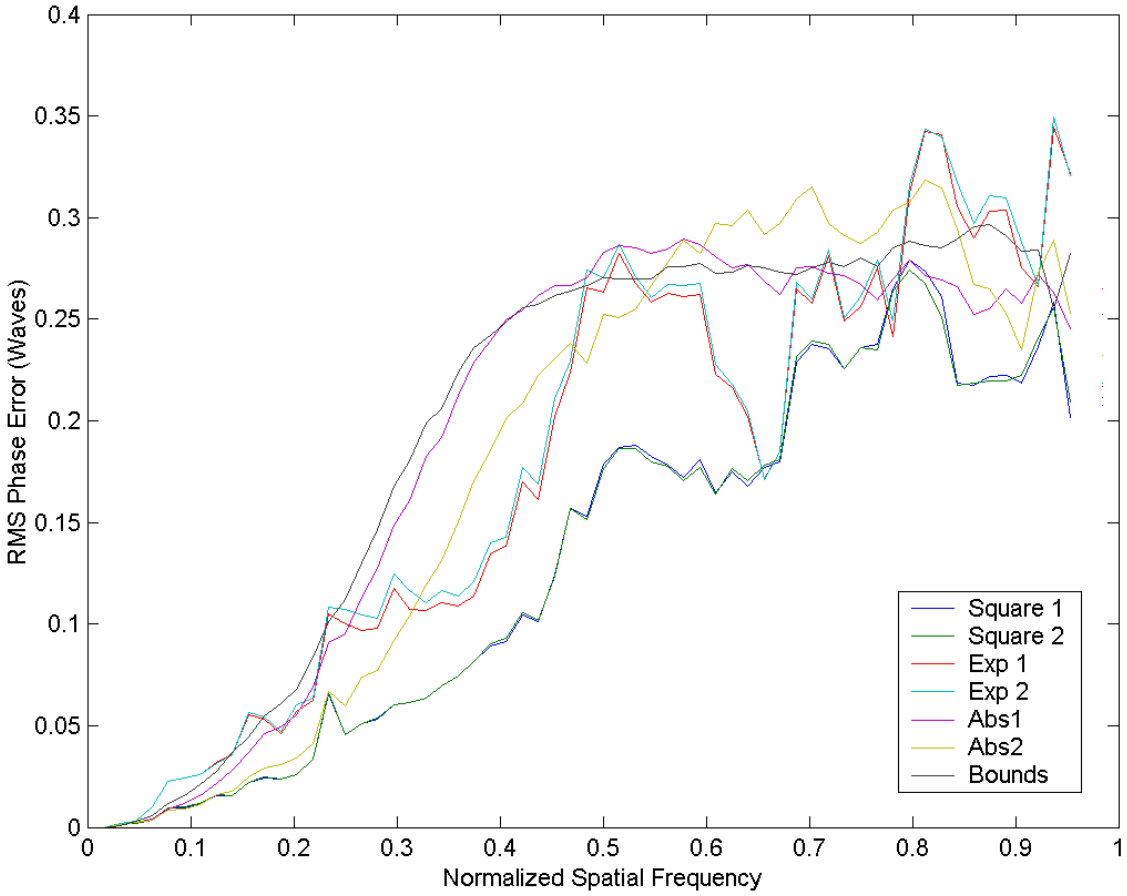


Figure 43: RMSPE for the Satellite object, 40 dB SNR data

In this high SNR case shown in Figure 43, we can see the three groups, with the most accurate spatial frequency information reconstructed by the Square parameterization family. We also see the similarity within each family (with the exception that Abs 2 is more accurate than Abs 1).

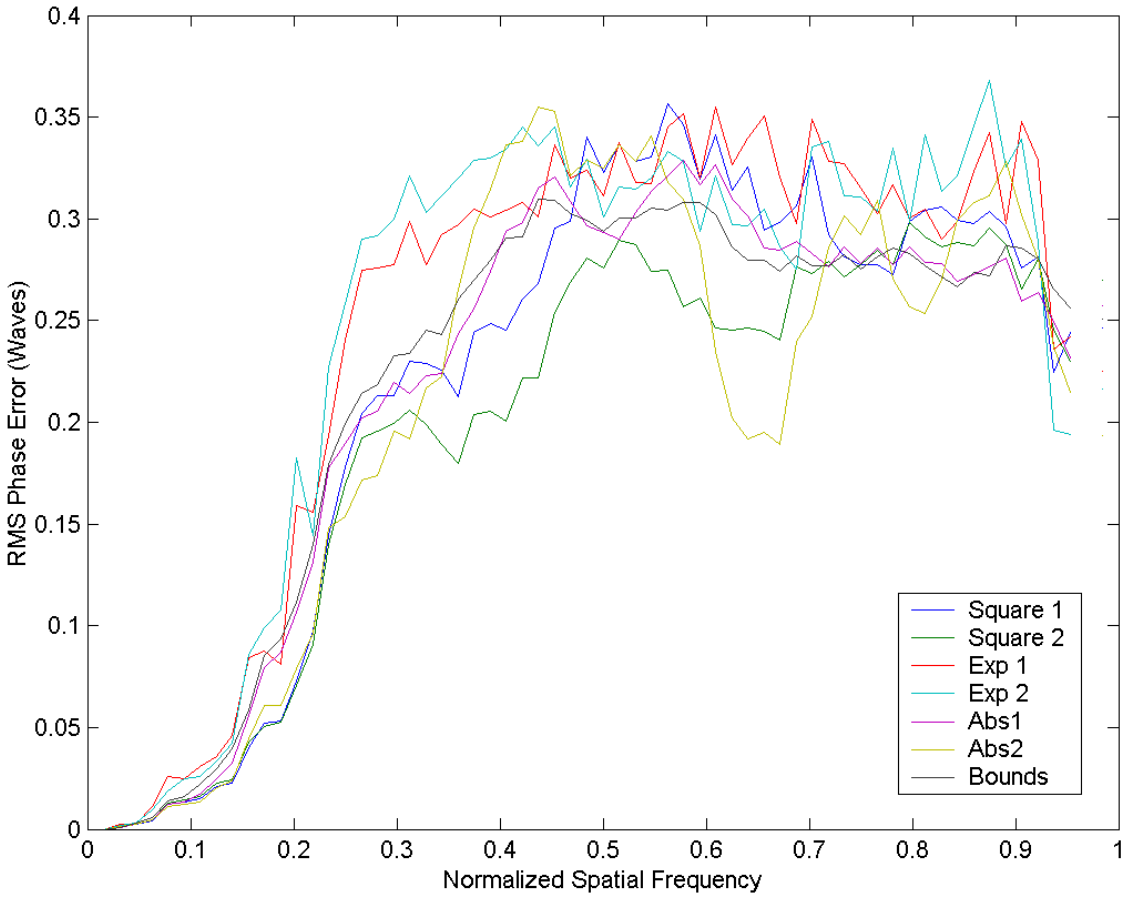


Figure 44: RMSPE for the Satellite object, 20 dB SNR data

In the low SNR case shown in Figure 44, we see the inability of all the algorithms to reconstruct accurately beyond the lowest frequencies, say a fourth of the Nyquist frequency. However, even here the Square variants provide the most accurate reconstructions. Beyond the low frequencies, the Square 2 variant delivers lower error than Square 1, but I would question how useful that is when the error is almost a quarter of a wave. An interesting reversal from the high SNR case is that the Bounds parameterization is better across the useful range of frequencies than the Exp variants.

Figure 45 shows the RMS phase error for the high SNR Spoke object, and we again see the superior performance of the Square variants. Unfortunately, we see a dismal performance for the Bounds parameterization on this target.

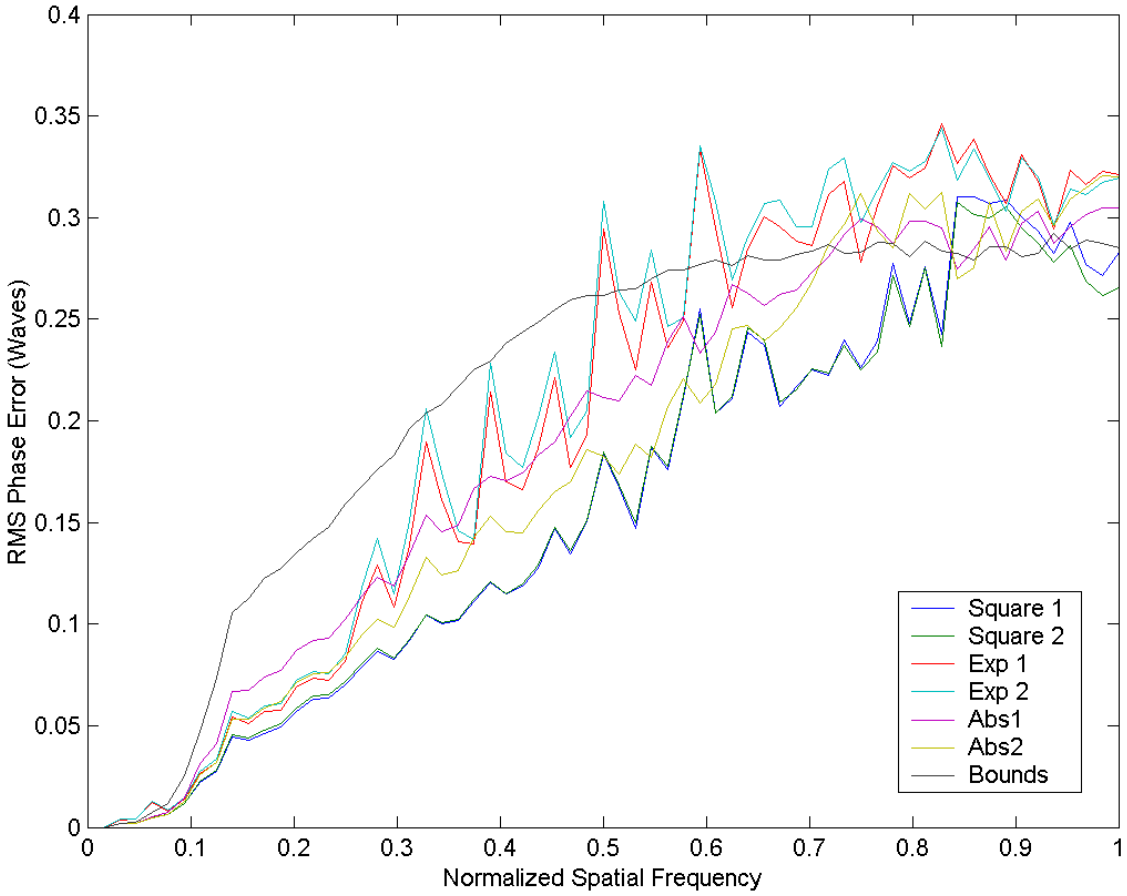


Figure 45: RMSPE for the Spoke object, 40 dB SNR data

Figure 46 shows the phase error performance for the low SNR Spoke data, which is mostly a repeat of the Satellite data. Namely, the Square family has the best performance of the lot.

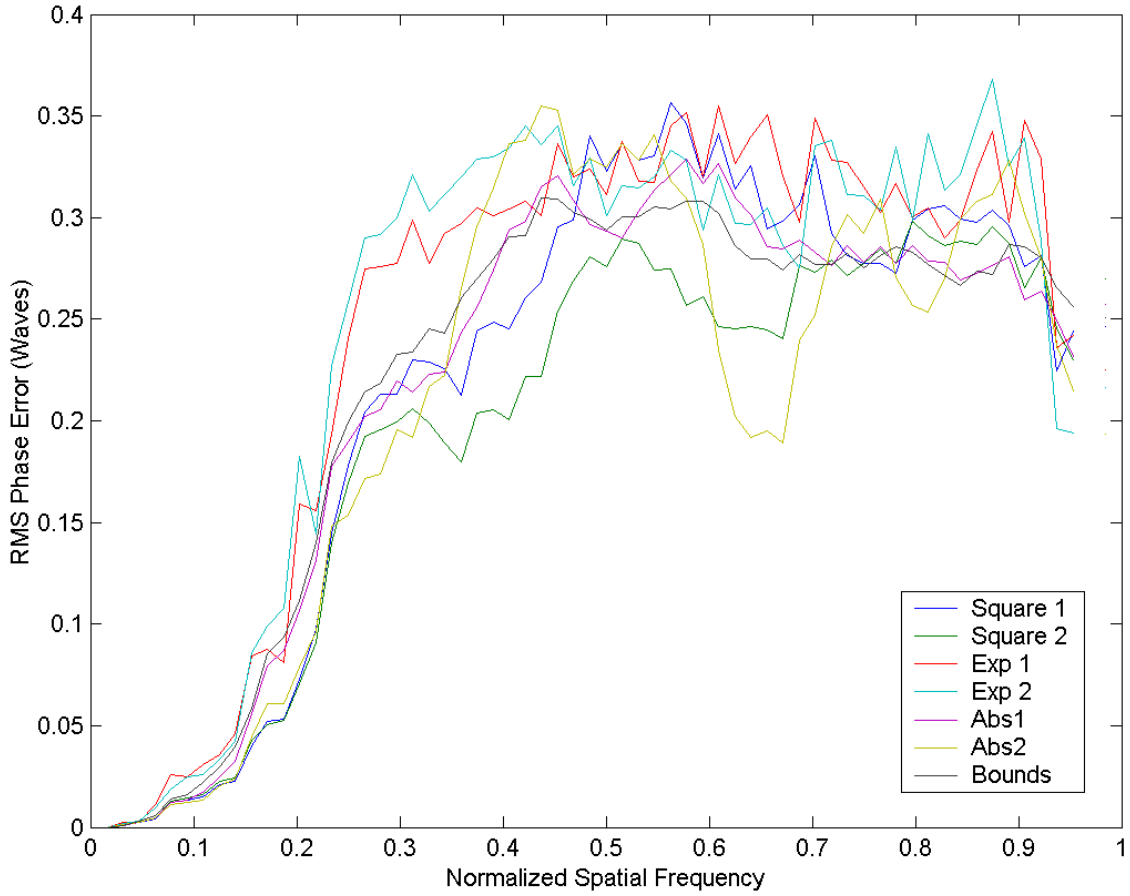


Figure 46: RMSPE for the Spoke object, 20 dB SNR data

A closer examination of the better reconstructed for this data region (i.e., the frequencies within the $1/8^{th}$ wave region or those below the F_8 cutoff) is shown in Figure 47. We note the Square 1 variant yielded slightly lower error than Square 2 for this object.

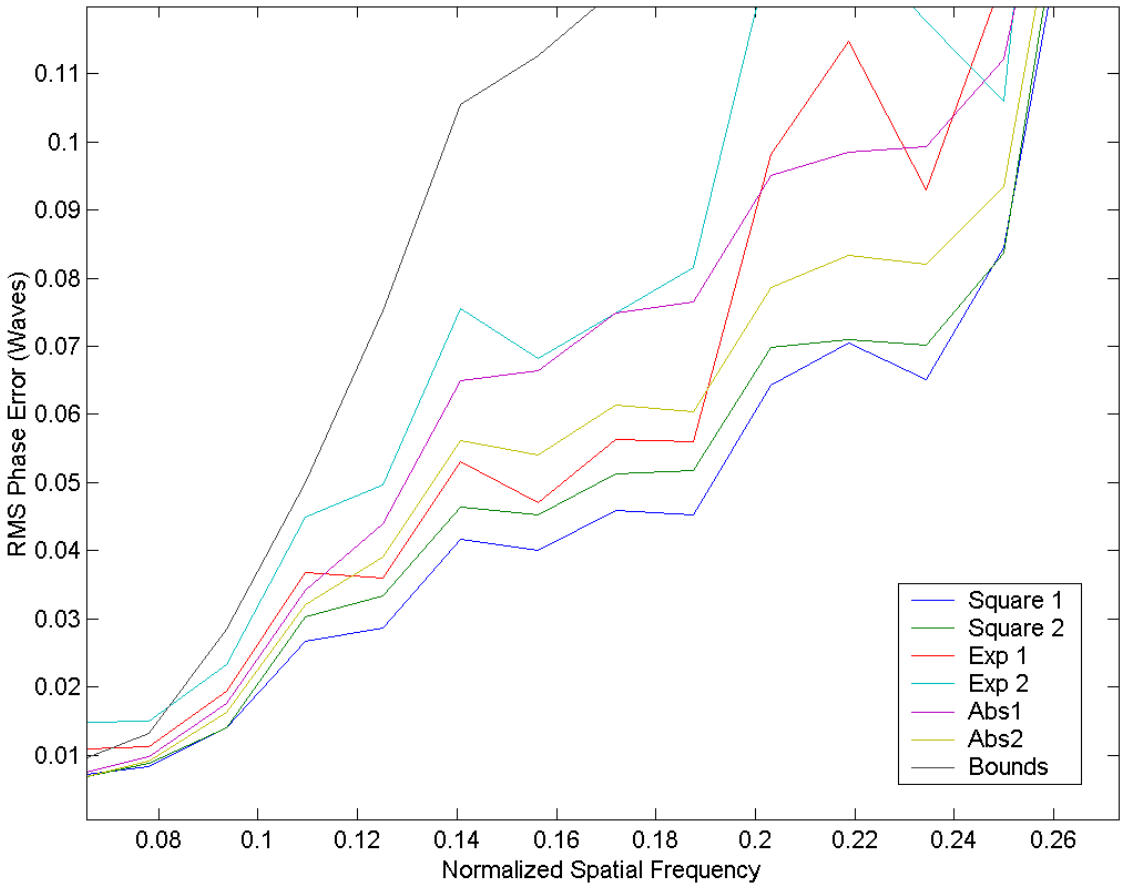


Figure 47: RMSPE for the Spoke object, 20 dB SNR data, F_8 region

4 SUMMARY/CONCLUSIONS

In this research, our desire was to investigate the implementation of a positivity constraint in an estimation procedure. We also favor the use of a hard constraint formulation rather than deal with the nebulous tuning that is required with soft constraint approaches. Our initial intent was to evaluate four approaches to implement the constraint, but this was later broadened to seven due to the interaction with a smoothness constraint that is applied by many researchers. Simulated imagery was used in order to provide quantitative assessment of reconstruction accuracy.

The most useful metric, by far, was the RMS phase error curves which were computed in the spatial frequency domain for each reconstruction. This metric provided insight into the accuracy of reconstruction at a finer level than is possible with a simple scalar metric. Specifically, these curves provide assessment of accuracy as a function of the level of detail present in the reconstruction. Surprisingly, even for the high signal to noise case, the errors for the Bounds, Exp, and Abs parameterizations were significantly higher than those of the Square variants in all but the very lowest levels of detail (low spatial frequencies). The Square 1 and Square 2 parameterizations were clearly the most accurate, and Square 2 seemed to just edge out Square 1 based on the phase error curves. The spatial domain RMS error confirmed the superiority of the Square parameterizations. Based on the F_8 metric, we felt both of these parameterizations yielded accurate details almost out to half of the diffraction limit in the high SNR case.

However, good reconstruction accuracy is not the whole story, and so we turn to the timing metrics. While one of the two most accurate parameterizations, Square 2, was not the slowest performer, it escaped that dubious honor only by about 3%. On the other hand, Square 1 had essentially the same accuracy but executed 29% faster. The time saved is likely not that meaningful for a single reconstruction, but when many are being performed it certainly adds up. As an example, it was not uncommon for the Air Force Maui Optical and Supercomputing

(AMOS) observatory on Haleakala to require image reconstructions for 10 satellite passes in a 24 hour period, each lasting perhaps 6 minutes, with ensembles covering 1/16 of a second. That's over fifty thousand reconstructions per day, which is one of the reasons they have access to a supercomputer!

As for the other parameterizations, Bounds had potential, having the fastest execution time (28% faster than Square 1), but it just didn't make the progress toward the solution that the others did in the given amount of iterations. Abs 1 was similar in this regard, and Abs 2 was both slower and less accurate in comparison to Square 1. The Exp variants really had issues with morphology for the low signal to noise ratio data. The "thinning" of the solar panels would be especially problematic for someone attempting to perform mensuration analysis on such reconstructions.

The development of a common theoretical framework greatly simplified the complexity of the development and contributed to accuracy of execution. The approach of treating the noise assumptions separately from the parameterization choices was especially useful. Additionally, the likelihood and gradient expressions were built on previously developed "layers", which avoided duplication and potential for error. Implementation was also improved since lower level layers could be used without modification once the first combination was implemented. As new parameterizations were added, the scrutiny of debugging could be focused on the newer, less tested code.

In this research, we examined several approaches to a commonly utilized constraint in the industry. However, the framework developed here is general enough to evaluate any potential hard constraint that one could dream up, and it has applicability to object as well as PSF constraints. This particular field of image reconstruction from atmospherically degraded imagery is constantly seeking improvement in either speed for a given quality or better quality for similar reconstruction times. The use of accurate and representative imagery and quantitative metrics based on known truth or "best case" performance is invaluable to assessment and improvement of the technology.

In particular, the recent attention of some researchers away from joint estimation of the object and PSF parameters is a complementary area of research. While the decrease in computational demand afforded by some of the parameterizations investigated in our work

would apply to their research as well, the reconstruction quality or accuracy may differ due to the alternate search methods. Additionally, many of those methods have developed improved methods of dealing with the complications due to the use of penalty terms in the likelihood function. Interaction of the parameterizations in this research with a penalty-based smoothness constraint as well as a comparison to penalty-based positivity could be very interesting.

5 REFERENCES

1 M. C. Roggemann and B. Welsh, *Imaging Through Turbulence,* CRC Press, Inc. (1996).

2 D. L. Fried, "Statistics of a Geometric Representation of Wavefront Distortion", *J. Opt. Soc. Am.*, vol. 55, Num. 11 (1965).

3 A. Labeyrie, "Attainment of diffraction limited resolution in large telescopes by fourier analyzing speckle patterns in star images," *Astron. Astrophys.*, vol. 6, p. 85 (1970).

4 H. W. Babcock, "The possibility of compensating astronomical seeing," *Publ. Astron. Soc. Pac.,* vol. 65, pp. 229-236 (1953).

5 J. W. Hardy, "Active optics: A new technology for the control of light," *Proc. IEEE,* vol. 66, pp. 651-697 (1978).

6 T. J. Schulz, B. E. Stribling, and Jason J. Miller, "Multiframe Blind Deconvolution with Real Data: Imagery of the Hubble Space Telescope," *Optics Express* , vol. 1, pp. 355-362 (1997).

7 K. T. Knox and B. J. Thompson, "Recovery of images from atmospherically degraded short exposure images," *Astrophys J. Letters*, vol. 193, L45-L48 (1974).

8 A. W. Lohmann, G. Weigelt, and B. Wirnitzer, "Speckle masking in astronomy: triple correlation theory and applications," *Appl. Opt.*, vol. 22, pp. 4028-4037 (1983).

9 R. A. Gonsalves and R. Childlaw, "Wavefront sensing by phase retrieval," *Applications of Digital Imaging Processing III*, A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng., vol. 207, pp. 32-39 (1979).

10 R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng., vol. 21, pp. 829-832 (1982).

11 R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Am. A*, vol. 9, pp. 1072-1085 (1992).

12 R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Phase-diverse speckle interferometry," in *Topical Meeting on Signal Recovery and Synthesis IV*, Technical Digest Series, vol. 11, (Optical Society of America, Washington DC, 1992), New Orleans, LA, April 1992.

13 W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.*, vol. 62, pp. 55-59 (1972).

14 L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, pp. 745-765 (1974).

15 G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, pp. 547-549 (1988).

16 R. G. Lane, "Blind deconvolution of speckle images," *J. Opt. Soc. Am. A,* vol. 9, pp. 1508-1514 (1992).

17 L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, pp. 745-765 (1974).

18 T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *J. Opt. Soc. Am. A,* vol. 10, pp. 1064-1073 (1993).

19 S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice Hall (1993).

20 S. J. Reeves and R. M. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Trans. Image Proc.*, vol. 1, pp. 301-311 (1992).

21 R. Vio, P. Ma, W. Zhong, J. Nagy, L. Tenorio, and W. Wamsteker, "Estimation of regularization parameters in multiple-image deblurring," *Astron. Astrophys.*, vol. 423, pp. 1179-1186 (2004).

22 Y. Biraud, "A new approach for increasing the resolving power by data processing," *Astron. Astrophys.*, vol. 1, pp 124-127 (1969).

23 R. Paxman, J. Seldin, M. Löfdahl, G. Scharmer, and C. Keller, "Evaluation of phase-diversity techniques for solar-image restoration", *Astrophys. Journal*, vol. 466, pp 1087-1099 (1996).

24 M. Noort, L. Voort, M. Löfdahl, “Solar image restoration by use of multi-frame blind deconvolution with multiple objects and phase diversity”, *Solar Physics*, Vol.228(1), pp 191-215 (2005).

25 W. Leung and R. Lane, “Blind deconvolution of images blurred by atmospheric speckle”, in Image Reconstruction from Incomplete Data, *Proc. SPIE 4123-73* (2000).

26 D. Gerwe, M. Jain, B. Calef, and C. Luna, “Regularization for non-linear image restoration using a prior on the object power spectrum”, *Proc. SPIE 5896* (2005).

27 S. Babacan, R. Molina, M. Do and A. Katsaggelos, "Bayesian Blind Deconvolution With General Sparse Image Priors," in *Proc. European Conference on Computer Vision*, Firenze, Italy, (2012).

28 D. Krishnan, T. Tay, R. Fergus, “Blind Deconvolution using a Normalized Sparsity Measure”, *Proc. IEEE Computer Vision and Pattern Recognition* (2011)

29 D. Strong, T. Chan, “Edge-preserving and scale-dependent properties of total variation regularization”, *Inverse Problems* vol. 19(6), S165 (2003).

30 D. Perrone, and P. Favaro, “Total variation blind deconvolution: The devil is in the details”, *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 2909-2916 (2014).

31 M. Roggemann, P. Billings, and J. Houchard, “Numerical studies of the value of including pupil intensity information in multi-frame blind deconvolution calculations for data measured in the presence of scintillation”, *Proc. AMOS Technical Conf.* (2007)

32 J. Seldin and R. Paxman, “Phase-diverse speckle reconstruction of solar data,” in Image Reconstruction and Restoration, *Proc. SPIE 2302-19* (1994).

33 P. Billings, M. Reiley, B. Stribling, “Mitigating turbulence-induced image blur using multiframe blind deconvolution”, *Proc. AMOS Technical Conf.* (2001).

34 M. Roggemann, P. Billings, and J. Houchard, “The effects of gray scale quantization and saturation on MFBD and bispectrum SSA image reconstructions”, *Proc. AMOS Technical Conf.* (2007).

35 B. Hunt, “Super-resolution of images: algorithms, principles, performance”, *Int. J. Imaging Sys. and Tech.*, vol. 6, pp. 297-304 (1995).

36 P. Billings, B. Hunt, K. Knox, and L. Roberts, Jr., "Resolving Closely-spaced objects with super-resolution", *Proc. AMOS Technical Conf.* (2004).

37 C. Matson, K. Schulze, P. Billings, and D. Tyler, "Multi-frame blind deconvolution and bispectrum processing of atmospherically-degraded data: a comparison", *Proc. SPIE* Vol. 4792 (2002).

38 A. Levi and H. Stark, "Image restoration by generalized projections with applications to restoration from magnitude," *J. Opt. Soc. Am. A*, vol. 1, pp. 932-943 (1984).

39 P. Ochs, Y. Chen, T. Brox, T. Pock, "IPIANO: Inertial proximal algorithm for nonconvex optimization", *SIAM J. Imaging Sciences*, vol. 7(2), pp. 1388-1419 (2014).

40 A. Levin, Y. Weiss, F. Durand, W. Freeman, "Efficient marginal likelihood optimization in blind deconvolution", *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 2657-2664 (2011).

41 D. Perrone, R. Diethelm, and P Favaro, "Blind Deconvolution via Lower-Bounded Logarithmic Image Priors", to be presented at International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (2015).

42 A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, Fourth Ed. McGraw Hill (2002).

43 L. Ludeman, *Random Processes: Filtering, Estimation, and Detection*, Wiley (2003).

44 T. R. Miller, J. W. Wallis, "Thoughts About Iterative Reconstruction," *Computer and Instrumentation Council Newsletter*, vol. 10-1 (1993).

45 D. L. Snyder, M. I. Miller, L. J. Thomas, Jr., and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstruction for emission tomography," *IEEE Trans Med Imaging*, vol. MI-6, pp. 228-237 (1987).

46 D.L. Snyder and M.I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," IEEE Trans. Nucl. Sci. **32(5)**, 3864-3872 (1985).

47 D. L. Snyder, M. I. Miller, L. J. Thomas, Jr., and D. G. Politte, "Noise and edge artifacts in maximum likelihood reconstructions for emission tomography," IEEE Trans. Med. Imag., **6(3)**, 228-238, (1987).

48 R. H. Byrd, P. Lu, J. Nocedal and C. Zhu, ``A limited memory algorithm for bound constrained optimization", *SIAM J. Scientific Computing*, vol. 16-5, pp. 1190-1208 (1995).

49 J. Nocedal and S. Wright, *Numerical Optimization*, Springer (1999).

50 R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.*, vol. 66, pp. 207-211 (1976).

51 J. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, Wiley (1978).

52 J. Proakis and D. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Third Ed., Prentice-Hall (1996).

53 M. C. Roggemann and B. Welsh, *Imaging Through Turbulence*, CRC Press, Inc. (1996).

54 N. Roddier, "Atmospheric wavefront simulation using Zernike polynomials", *Optical Engineering*, vol. ***29-10***, pp. 1174-1180 (1990)

55 V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils", *Engineering & Laboratory Notes* (August), Optical Society of America (1994).

56 R. L. Lucke, "The relation between physical and computer-generated point spread functions and optical transfer functions", *Am. J. Phys.*, vol. 69-12, pp. 1237-244 (2001).