5-2010

A Novel Technique for CTIS Image-Reconstruction

Mitchel Dewayne Horton
mhorton5@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Mitchel Dewayne Horton entitled "A Novel Technique for CTIS Image-Reconstruction." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Michael D. Vose, Major Professor

We have read this dissertation and recommend its acceptance:

Bruce MacLennan, Hairong Qi, Sergey Gavrilets

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Mitchel DeWayne Horton entitled "A Novel Technique for CTIS Image-Reconstruction." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Michael D. Vose, Major Professor

We have read this dissertation
and recommend its acceptance:

Bruce MacLennan

Hairong Qi

Sergey Gavrilets

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Novel Technique for CTIS Image-Reconstruction

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Mitchel DeWayne Horton

May 2010

Dedication

This dissertation is dedicated to Bob Garrett.

Acknowledgments

I would like to thank my advisor, Dr. Vose. For the past 4 1/2 years, he has set forth a matchless paradigm for how an advisor should give selflessly of himself, acting at every turn in the highest interests of his student.

I would like to thank Dr. Jessee Poore and the Tennessee Science Alliance for my first year of funding. Thank you to the Department of Electrical Engineering and Computer Science for 3 1/2 years of Teaching Assistant funding. Thank you to Jacob Barhen and ORNL LDRD for the problem with which this dissertation is concerned. Thank you to my committee: Dr. Bruce MacLennan, Dr. Hairong Qi, and Dr. Sergey Gavrilets.

Abstract

Computed tomography imaging spectrometer (CTIS) technology is introduced and its use is discussed. An iterative method is presented for CTIS image-reconstruction in the presence of both photon noise in the image and post-detection Gaussian system noise. The new algorithm assumes the transfer matrix of the system has a particular structure. Error analysis, performance evaluation, and parallelization of the algorithm is done. Complexity analysis is performed for the proof of concept code developed. Future work is discussed relating to potential improvements to the algorithm.

An intuitive explanation for the success of the new algorithm is that it reformulates the image reconstruction problem as a constrained problem such that an explicit closed form solution can be computed when the constraint is ignored. Incorporating the constraint leads to an inverse matrix problem which can be dealt with using a conjugate gradient method. A weighted iterative refinement technique is employed because the conjugate gradient solver is terminated prematurely.

This dissertation makes the following contributions to the state of the art. First, our method is several orders of magnitude faster that the previous industry best (multiplicative algebraic reconstruction technique (MART) and mixed-expectation reconstruction technique (MERT)). Second, error bounds are established. Third, open source proof of concept code is made available.

Preface

This dissertation is organized roughly into three major parts: Chapters 1 and 2 introduce the CTIS device; Chapter 3 describes our method; Chapters 4, 5, 6 and 7 provide analysis of the method. There is a List of Symbols on page xiii and a List of Equations on page xvi. The Appendix contains proofs of lemmas developed by the author and standard results used in the body of the text.

The following format is followed where the justifications for steps (equalities or inequalities) are provided as footnotes (the following is simply an example of the formatting; no attempt is made to provide the assumptions necessary for the equalities and inequalities seen to hold):

$$\begin{aligned}
\rho(Z'\phi Z') &= u^T Z'\phi Z' u \\
&= v^T \phi v \\
&=_1 ||v||^2 \left(\frac{v}{||v||}\right)^T \phi \left(\frac{v}{||v||}\right) \\
&\leq ||v||^2 \sup_{||w||=1} w^T \phi w \\
&=_2 ||v||^2 \rho(\phi) \\
&<_3 ||v||^2
\end{aligned}$$

The footnote associated with subscripts 1,2, and 3 provide justification for the steps indicated.

[1] If $v = 0$ then $\rho(Z'\phi Z') = 0$ which is ≤ 1.

[2] from Lemma (A.28)

[3] It was shown that $\rho(\phi) < 1$.

Contents

List of Figures

List of Symbols

ω	entry in Fourier transform; see Equation (A.320)
β	Poisson distribution expected value; see Equation (A.145)
g	measured image on the focal plane array
H	transfer matrix representing the optical processing of the system
f	object vector representing the flash hyperspectral image
n_1, n_2	noise terms
λ	eigenvalue
e	natural logarithm
p	probability mass function; probability density function
x	noise removed measured image on the focal plane array
σ^2	variance
μ	regularization parameter
π	pi
T	rectangular circulant block; see Equation (3.20)
n	number of rows in H
m	number of columns in H
w	number of wavelengths H is calibrated at
α	number of rectangular circulant blocks in each partition of H
a	number of columns in each rectangular circulant block
γ	shift amount; see Equation (3.22)
$\mathbb{Z}$	set of all integers
I	identity matrix
C	circulant matrix

c	first column of circulant matrix
e_0	first column of identity matrix
F	Fourier transform matrix
D	diagonal matrix
Q	embedded matrix; see Equation (3.48)
E	embedded matrix; see Equation (3.50)
A, B, U	arbitrary matrix
$\mathcal{H}$	expanded system matrix; see Equation (3.64)
y	expanded f; see Equation (3.67)
h	modified left hand side; see Equation (3.67)
M	renamed inverse; see Equation (3.88)
P	constituent term of M; see Equation (3.90)
W	inverse square of factor of P; see Equation (3.99)
Λ	arbitrary diagonal matrix of eigenvalues
Z	enforces zero pattern; see Equation (3.132)
Z'	enforces zero pattern; see Equation (3.131)
$\vec{y_i}$	part of approximate solution to Equation (3.78); see Equation (3.133)
$\vec{v_i}$	part of approximate solution to Equation (3.78); see Equation (3.133)
η	vector of weights; see Equation (3.188)
V	partitioned matrix ($V_i = H\vec{f_i}$); see Equation (3.200)
s	solution to $Hf = Hs$; see Equation (**??**)
ρ	spectral radius
ϵ	error
ϕ	another name for $M\mathcal{H}^T\mathcal{H}$; see Equation (4.42)
θ	another name for $Z'\phi$; see Equation (4.43)
V', V''	subspaces
Θ	upper bound
S	another name for FWF^*; see Equation (3.120)
G, X, J, O	matrices in Sherman-Morrison-Woodbury identity
φ	another name for $\mu(Z'P)(Z'P)^T$; see Equation (4.36)

Y	term in scalar product; see Equation (4.47)
ϵ_f	relative error in reconstructed image
$\vec{f_\infty}$	expanded f; see Equation (4.103)
$\hat{f}$	recovered image
r_g	maps x, y position in focal plane array to position in the vector g
r_f	maps x, y, u position in discretized object cube to position in the vector f
ξ	width of focal plane array

List of Equations

$M\mathcal{H}^T\mathcal{H} = \mu PP^T$	useful identity (3.88)
$Z = \text{diag}((I_w \otimes E)\mathbf{1}_m)$	definition of Z (3.132)
$Z' = I - Z$	definition of Z' (3.131)
$\vec{y_0} = ZMh$	
$\vec{y_{i+1}} = \vec{y_i} + ZM\mathcal{H}^T\mathcal{H}\vec{v_i}$	definition of $\vec{y_i}$ (3.133)
$\vec{v_0} = Z'Mh$	
$\vec{v_{i+1}} = Z'M\mathcal{H}^T\mathcal{H}\vec{v_i}$	definition of $\vec{v_i}$ (3.200)
$\vec{x_i} = \vec{y_i} + \vec{v_i}$	definition of $\vec{x_i}$ (A.39)
$\vec{y_\infty} = \lim_{i\to\infty} \vec{y_i}$	definition of $\vec{y_\infty}$ (3.150)
$\vec{y_\infty} = Z(I + \mu PP^T Z'\{I - \mu(Z'P)(Z'P)^T\}^{-1}Z')Mh$	equivalent definition of $\vec{y_\infty}$ (3.150)
$\vec{y_\infty} = Z(I + \mu PP^T Z'\{I + P(\mu^{-1}I - P^T Z'P)^{-1}P^T Z'\})Mh$	equivalent definition of $\vec{y_\infty}$ (3.181)
$\vec{f_0} = (I_w \otimes E^T)\,\vec{y}_\infty$	first approximation to solving $x = Hf$ (3.185)
$x - H\sum \eta_i \vec{f_i} = Hf$	problem associated with $\vec{f_{i+1}}$ (3.189)
$\phi = M\mathcal{H}^T\mathcal{H}$	definition of ϕ (4.42)
$\theta = Z'\phi$	definition of θ (4.43)
$\vec{x_\infty} = \lim_{i\to\infty} \vec{x_i} = Mh + (\phi - I)(I - \theta)^{-1}\vec{v_0}$	definition of $\vec{x_\infty}$ (A.2)
$\mu M = I - \phi$	useful identity; Lemma (A.2)

Chapter 1

Introduction

The goal of imaging spectrometry is to obtain an image of an object at several different wavelengths [3]. In many different applications, a deeper understanding of the nature of a scene can be obtained with the availability of a spectrum at each pixel of the image. According to Garcia and Dereniak, pioneers in the field, "spectral data can reveal the condition or identity of object-space elements in vegetation monitoring, mineral identification, detection of chemical agents, temperature and emissivity estimation, and medical imaging applications" [2, 4].

The computed tomography imaging spectrometer (CTIS) is a non-scanning flash hyperspectral imaging (FHI) snapshot instrument which simultaneously captures the spectral information for each pixel in its field of view in the shortest possible time (typically several milliseconds) [3, 5, 6, 8, 9]. The adjective *snapshot* refers to the fact that all data is collected in a single integration [9]. The adjective *hyperspectral* indicates multiple frequencies (dozens to hundreds) and the fact that those frequencies are continuous. The snapshot property is a desirable property when, without it, a short dynamic event could be missed or a long dynamic event could be corrupted (because the scene is changing while information is being collected) [11]. The CTIS records spatial and spectral information about a scene by sampling a tomographic dispersion pattern formed by a computer generated hologram disperser (see Figure (1.1) [13]) [6, 10]. The raw data collected a by CTIS instrument (what is on the focal plane array) must be processed in order to extract the set of images (each at a different wavelength) which the camera captures [3, 11]. A mathematical model

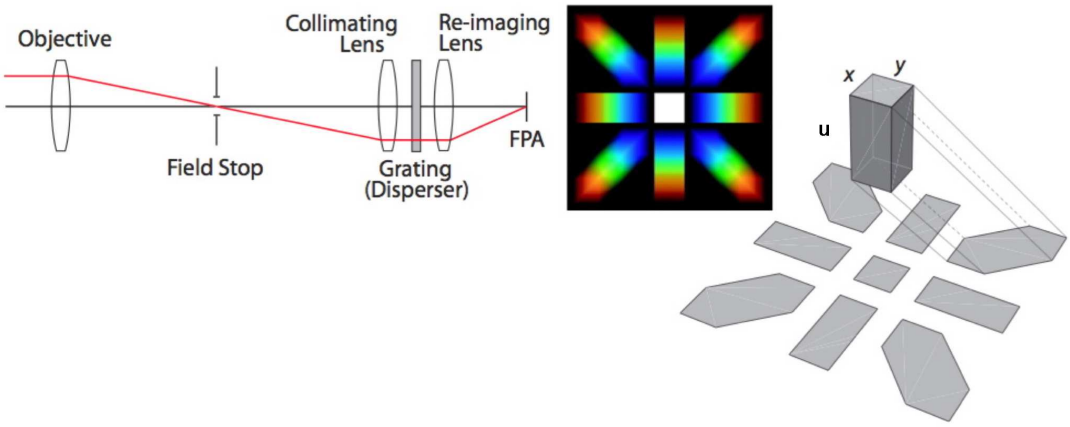

Figure 1.1: The CTIS optical layout, showing an example FPA image [13].

describing the system is

$$g = Hf + n_1 + n_2 \tag{1.1}$$

where the transfer matrix H represents the optical processing of the CTIS system, g is the measured image obtained on the focal plane array, f is the object vector representing the flash hyperspectral image, n_1 is photon noise, and n_2 is post detection system noise [10, 49].

People typically refer to the spatial and spectral region of interest as the *object cube* or the *continuous object cube*. The (x,y,u) space is, conceptually, a cube in that the area in the field stop (x,y) makes up two dimensions and the wavelength (u) makes up the third dimension (see Figure (1.1) [13]) [11].

Each point in the continuous object cube has an associated brightness, referred to as the spectral radiance. The SI unit of spectral radiance is watts per steradian per square meter per nanometer. One way to make sense of the units is that spectral radiance is a name for the amount of light that is emitted from a given area within a given solid angle within a range centered around a given wavelength. Because a CTIS device is not always used in the visible region of the spectrum, we measure the amount of light in watts as opposed to photons. Because a CTIS device is a discrete digital system, the (x,y,u) space is divided into a set of contiguous volumes, known as *voxels*. The resulting assembly of voxels in referred to as the *discretized data cube*. Each voxel in the discretized data cube

has an associated brightness, referred to as intensity. The units of intensity are watts per steradian since it is the integral of the continuous object cube over area and wavelength. The estimation of the intensity of the voxels of the discretized object cube is referred to as the *data cube*. It is the data cube that a CTIS device measures in an indirect way as an image g on the focal plane array. To obtain what we are interested in — an estimate f of the brightness of each pixel in the field of view at a number of discrete wavelengths — the image g on the focal plane array must be processed to reconstruct f [11, 12].

The vector g contains the measurements obtained from the pixels on the focal plane array (FPA) contiguously ordered into a vector. The function r_g maps from an x, y position on the focal plane array to a position in the vector g:

$$r_g(x, y) = y\xi + x \tag{1.2}$$

where ξ is the width of the focal plane array in pixels. The vector f is an estimate of the intensity of the voxels within the discretized object cube also lexicographically ordered into a vector. The function r_f maps from an x, y, u position in the discretized object cube to a position in the vector f:

$$r_f(x, y, u) = ua\alpha + ya + x \tag{1.3}$$

where a and α are the width (in pixels) and depth (the number of distinct wavelengths of interest) of the discretized object cube respectively. The reconstruction problem is to solve Equation (1.1) for f.

Each column of the system matrix H contains the response of the system to the illumination of a single voxel. Said another way, the (i, j)-th position in H describes the number of electrons stored by the i-th pixel of the FPA (represented by g_i) per Watt/Steradian at the j-th voxel in the discretized object cube (represented by f_j). The system matrix could be obtained by placing a small-diameter optical fiber at each pixel in the field stop and illuminating that fiber at each wavelength of interest. Then, at each pixel and at each wavelength, what is on the focal plane array is scaled and becomes a column of the system matrix. In practice, however, the fiber is placed at a single pixel only and illuminated at each wavelength of interest. This is because of the shift invariant assumption. The shift

invariance assumption is that: were the optical fiber to be moved by a single pixel, the image on the focal plane array would remain unchanged other than to be shifted by one pixel in the same direction [11, 13]. See Equations (3.19), (3.20), (3.22), and (3.25).

In summary, the CTIS records spatial and spectral information about a scene in a short time interval (typically several milliseconds) in an indirect way which requires post processing of the raw data g collected. The post processing "image reconstruction" problem corresponds to solving Equation (1.1) for f. That image reconstruction problem is nearly intractable due to the large size of the system matrix H.

Chapter 2

Where CTIS is used Today

According to Scholl, Hege, O'Connell, and Dereniak, "the basic principles of hyperspectral imaging systems like the CTIS had been known by a number of researchers since the 1980's, but it was Descour, Dereniak, and Volin that had worked out the practical issues and built real working systems" [2, 11, 48]. It is clear that CTIS technology has flourished since that time. Every significant aspect of the original CTIS design has seen massive amounts of attention and improvement. CTIS systems have become a vital part of a myriad of disparate applications from astronomy to cell physiology. Were it not for the snapshot capability of the CTIS device, progress in many areas would be languishing. CTIS technology has been combined with other technologies to achieve essential capabilities heretofore unavailable.

2.1 Traditional Spectrometry

As previously stated, the objective of imaging spectrometry is to obtain in image of an object at several different wavelengths [3]. According to Luo, Liao, Wang, Liang, Fen, and Yang, "imaging spectrometers currently produce a nearly continual stream of multi-dimensional image data, containing a wealth of spatial and spectral information, which has opened groundbreaking perspectives in many applications" [43]. Measurements of the spectral content are used in a broad range of diverse fields [2, 20–24, 29, 39]. Application examples are numerous [3], including both defense and commercial applications [35].

For example, hyperspectral imaging devices have been used in remote sensing [6, 36], remote sensing reconnaissance [8, 20–23, 25], and airborne remote sensing [2, 29, 39]. They have been used to track targets or threats in space (ballistic missiles and near-Earth asteroids and comets, for example) [35]. Hyperspectral imaging devices are used in vegetation monitoring, mineral identification, and molecular physics [2, 4, 20–24]. They are used for the detection of chemical agents and temperature and emissivity estimates. [2, 4, 35, 43]. Hyperspectral imaging can assist in examination and surgical tasks [2–4, 8, 25] and has proven applicable in optical testing and biomedical imaging [2, 29, 39]. Measurements of the spectral content are used in astronomy [20–24].

Traditional hyperspectral imaging spectrometers collect (x, y, u) data by employing some form of scanning, such as pushbroom scanning or whiskbroom scanning [6, 37, 38]. Imaging Fourier-transform spectrometers require scanning of the optical path difference between the two arms of a Michelson interferometer. According to Johnson, Wilson, Fink, Humayun, and Bear, "Hadamard encoding slit spectrometers [8, 26], liquid crystal and acousto-optic tunable filters [8, 27], spectral-temporal scanning [8, 29], and more recently volume holographic methods [8, 30]; all scan to build up the spatial-spectral cube of information" [8]. When a scene is stationary, scanning is acceptable, but for dynamic scenes, artifacts can occur [6]. Not only that, but in the area of cell physiology, traditional hyperspectral imaging limits the ability to view rapidly occurring events [5, 19]. In the area of retinal imaging, because of rapid eye movements, traditional hyperspectral imaging systems are plagued with motion artifacts [8, 33] and patient discomfort can become an issue [8]. In the case of astronomy, traditional hyperspectral imaging spectrometers do not perform well in the presence of rapidly evolving scenes. Not only that, platform vibration can cause imaging irregularities [9]. The elimination of scanning removes the problems stated above; in addition, a nonscanning device is more rugged due to the absence of moving parts [2].

2.2 CTIS

Recall that the computed tomography imaging spectrometer (CTIS) is a non-scanning flash hyperspectral imaging (FHI) snapshot instrument which simultaneously captures the spectral information for each pixel in its field of view in the shortest possible time (typically

several milliseconds) [3, 5, 6, 8, 9]. Because of this, the CTIS overcomes all of the problems mentioned above associated with traditional hyperspectral imaging spectrometers.

CTIS instruments have been used for optical testing, biomedical imaging, and airborne remote sensing [2, 29, 39]. They have been used successfully for stellar classification [8, 31], fluorescence microscopy [8, 16], and geology measurements [8, 32]. CTIS technology has been used extensively in military [44, 45] and space surveillance [9, 44]. It has been demonstrated that CTIS instruments can be used to observe the response of individual cells multivariate cellular response [19]. The CTIS microscope can be used to study the effects of drugs and toxic agents on the physiology of the liver [17]. In the area of clinical diagnosis, CTIS instruments can be used for optical biopsy, providing a less invasive alternative to traditional approaches [17, 18]. CTIS systems have been tested, using an adaptive optics (AO) compensated telescope, for astronomical multispectral imaging [9]. CTIS instruments have been used to detect threat objects [2, 35]. In missile defense, the kinetic kill of a Hera drone by a PAC-III interceptor was detected by a CTIS instrument, operating at 15 frames per second [41]. Using a CTIS device, mine-like objects have been successfully detected [45]. Research has been done to study the feasibility of using CTIS systems for scene and target characterization from space-based platforms [39].

2.3 Improvements to the Original CTIS Design

In the case of cell physiology, a second-generation system is under development that includes the use a larger focal plane array (resulting in increased spatial resolution) and higher system transmission efficiency (resulting in increased temporal resolution) [47]. Depending on the size of the field stop, larger focal plane arrays can also result in higher spectral resolution. A larger field stop results in less room on the focal plane array for diffraction orders. A smaller field stop results in more room on the focal plane array for diffraction orders. More diffraction orders implies greater spectral resolution [40]. Advancements in both focal plane array technology and computer-generated-hologram design have led to significant increases in the number of voxels within reconstructed CTIS data cubes [14, 19].

2.4 Summary

Because the CTIS is a nonscanning device, it overcomes all of the problems with traditional spectrometry that stem from scanning. CTIS technology is pervasive. In addition, changes to the original CTIS design have resulted in increased spatial, spectral, and temporal resolution which only serve to exacerbate what is already a nearly intractable post processing reconstruction problem. Increased spatial and spectral resolution implies an increase in the size of the data cube. Increased temporal resolution implies an increase in the number of data cubes per unit time.

Chapter 3

A Heuristic Technique for CTIS Image-Reconstruction

This chapter is a revised version of a paper by the same name published in the journal *Applied Optics* in 2007 by Michael D. Vose and Mitchel D. Horton:

M. D. Vose and M. D. Horton, *A heuristic technique for CTIS image-reconstruction*, Appl. Opt. **46** (2007), 6498-6503.

My use of the word "we" in this chapter refers to my co-author and myself. My primary contributions to this paper include (1) implementing and testing code, (2) research, and (3) proofreading.

3.1 Introduction

Maximum-likelihood image reconstruction techniques are important tools in the restoration of noise corrupted images. They have widespread use in areas from imaging spectroscopy to emission tomography [3, 50, 51]. Our technique first removes noise via maximum-likelihood, and then subsequently performs image reconstruction by exploiting the particular structure of the matrix modeling the noise-free system. It is therefore specific to particular systems, like the computed-tomography imaging spectrometer, whose noise-free system matrix may be acceptably approximated by one having the required structure.

A computed-tomography imaging spectrometer (CTIS) records spatial and spectral

information about a scene by sampling a tomographic dispersion pattern formed by a computer generated hologram disperser [6]. A mathematical model describing the system is

$$g = Hf + n_1 + n_2 \tag{3.1}$$

where the transfer matrix H represents the optical processing of the CTIS system, g is the measured image obtained on the focal plane array, f is the object vector representing the flash hyperspectral image, and the n_i are noise terms.

Garcia and Dereniak [49] take n_1 to be Poisson distributed photon noise, and n_2 to be zero mean post-detection Gaussian system noise. They assume the standard deviation σ_s of n_2 is known, and that each Poisson distributed term $(Hf)_k + (n_1)_k$ is well approximated by a normal distribution with mean and variance $(Hf)_k$. Assuming statistical independence of all components g_m, they present an iterative technique to recover f from the measurement g, based on a maximum likelihood estimator.

The mean and variance are the same for a Poisson distribution. The Poisson probability mass function is given by

$$p(n) = \frac{\beta^n e^{-\beta}}{n!} \tag{3.2}$$

where β is the expected value. The expected value (or mean) is given by β (see (A.9) in the Appendix). The variance is given by β (see A.10 in the Appendix).

This paper extends the results of Garcia and Dereniak in two ways. First, the result of their maximum likelihood estimator is obtained in explicit closed form, which transforms the stochastic problem (3.1) into a deterministic problem of the form

$$x = Hf \tag{3.3}$$

Second, a heuristic method for recovering f from x is presented which for larger problems is significantly better with respect to both accuracy and computational time (the object vector f used by Garcia and Dereniak [49] has 289 components; in our experiments f has $174,960$ components). The improvement, however, comes at the cost of requiring the transfer matrix H to have special structure. The precise details of that structure are explained after discussing noise removal.

3.2 Removing Noise

The normal probability density function is given by

$$p(u) = \exp\{-(u-\beta)^2/(2\sigma^2)\}/\sqrt{2\pi\sigma^2} \tag{3.4}$$

where σ^2 is the variance and β is the mean. The assumptions described in the introduction imply that g_k is approximated by a normal random variable with mean

$$x_k = (Hf)_k \tag{3.5}$$

and variance $\sigma_s^2 + (Hf)_k$. The probability of measuring a specific element g_k given f is then approximated by

$$\text{Prob}(g_k|f) = \exp\{-(g_k - x_k)^2/(2x_k + 2\sigma_s^2)\}/\sqrt{2\pi(x_k + \sigma_s^2)} \tag{3.6}$$

where

$$x = Hf \tag{3.7}$$

Assuming statistical independence of the components of g, the probability of measuring the entire vector g given f is approximated by

$$\text{Prob}(g|f) = \prod_k \exp\{-(g_k - x_k)^2/(2x_k + 2\sigma_s^2)\}/\sqrt{2\pi(x_k + \sigma_s^2)} \tag{3.8}$$

This is essentially Garcia and Dereniak's problem formulation [49]. Rather than basing an iterative technique on Equation (3.8), we solve explicitly as follows.

The logarithm of Equation (3.8) is

$$\sum_k -(g_k - x_k)^2/(2x_k + 2\sigma_s^2) - \log(\sqrt{2\pi(x_k + \sigma_s^2)}) \tag{3.9}$$

Differentiating with respect to x_k yields

$$\frac{4(g_k - x_k)(x_k + \sigma_s^2) + 2(g_k - x_k)^2}{4(x_k + \sigma_s^2)^2} - \frac{1}{2(x_k + \sigma_s^2)} \tag{3.10}$$

and leads to the optimization condition

$$0 = \frac{(x_k + \sigma_s^2) - 2(g_k - x_k)(x_k + \sigma_s^2) - (g_k - x_k)^2}{(x_k + \sigma_s^2)^2} \tag{3.11}$$

Since x_k is assumed to be a variance – of $(Hf)_k + (n_1)_k$ – it is the nonnegative root of

$$x_k^2 + (1 + 2\sigma_s^2)x_k + (\sigma_s^2 - 2\sigma_s^2 g_k - g_k^2) = 0 \tag{3.12}$$

Namely

$$x_k = \frac{\sqrt{(1 + 2\sigma_s^2)^2 + 4(g_k(g_k + 2\sigma_s^2) - \sigma_s^2)} - (1 + 2\sigma_s^2)}{2} \tag{3.13}$$

In order that x_k is indeed non negative, it must happen that

$$g_k(g_k + 2\sigma_s^2) - \sigma_s^2 \geq 0 \tag{3.14}$$

hence

$$g_k^2 + 2\sigma_s^2 g_k - \sigma_s^2 \geq 0 \tag{3.15}$$

and therefore

$$g_k \geq \sqrt{\sigma_s^4 + \sigma_s^2} - \sigma_s^2 \tag{3.16}$$

Hence by redefining g_k where necessary to ensure the above constraint (3.16), then

$$x = Hf \tag{3.17}$$

may be regarded as known – it is obtained via Equation (3.13) – and it remains to determine f.

3.3 Recovering the Hyperspectral Image

Our heuristic method of obtaining an approximate solution to

$$x = Hf \tag{3.18}$$

is based on two assumptions concerning the structure of the $n \times m$ matrix H.

1. Assume the transfer matrix can be partitioned as

$$H = (H_0 \mid \cdots \mid H_{w-1}) \tag{3.19}$$

where each H_k is comprised of (the same number of) *rectangular* circulant blocks

$$H_k = (T_{k,0} \mid \cdots \mid T_{k,\alpha-1}) \tag{3.20}$$

(i.e., the i, j th entry of the rectangular matrix $T_{k,\ell}$ is some function of $i - j \mod n$).

2. Assume each $T_{k,\ell}$ is $n \times a$ Thus H has

$$m = a\alpha w \tag{3.21}$$

columns. Assume also that

$$T_{k,\ell} \;=\; R_\gamma^\ell \, T_{k,0} \tag{3.22}$$

where

$$\gamma \geq a, \tag{3.23}$$

$$n \geq \alpha\gamma, \tag{3.24}$$

and R_γ is the $n \times n$ circulant matrix

$$(R_\gamma)_{i,j} \;=\; [\gamma = i - j \mod n] \tag{3.25}$$

Here the notation [*expression*] denotes 1 if *expression* is true, and 0 otherwise.

Two integers u and v are said to be congruent modulo n if their difference $(u - v)$ is an integer multiple of n. If this is the case, it is expressed as:

$$u \equiv u \mod n \tag{3.26}$$

Congruence modulo n is an equivalence relation, and the equivalence class of the integer v is the set $\{v + n\mathbb{Z}\}$. This set, consisting of the integers congruent to v modulo n, is called the congruence class of v modulo n. One way to interpret

$$u = v \bmod n \tag{3.27}$$

is to say that

$$\{u + n\mathbb{Z}\} = \{v + n\mathbb{Z}\} \tag{3.28}$$

Equivalently,

$$u \in v + n\mathbb{Z} \tag{3.29}$$

Although it is known that the CTIS process is *not* shift invariant [2, 15, 52] – i.e., the second assumption above does not hold – past study [11] has shown the assumption to yield a reasonable approximation. This paper is not concerned with that approximation, but considers the problem (defined by Equation (3.3)) of recovering f from x based on the assumptions (enumerated above) which are typically made for CTIS (noise removal was dealt with in the previous section). In other words, we are concerned with a *mathematical abstraction* rather than with the actual physical system.

It is easily proved by induction that

$$(R_\gamma^\ell)_{i,j} \;=\; [\ell\gamma = i - j \bmod n] \tag{3.30}$$

Base case ($\ell = 0$):

$$(R_\gamma^0)_{i,j} \;=\; I_{i,j} \;=\; [0 = i - j \bmod n] \;=\; [0\gamma = i - j \bmod n] \tag{3.31}$$

Inductive step:

$$(R_\gamma^\ell)_{i,j} \;=\; [\ell\gamma = i - j \bmod n] \Longrightarrow (R_\gamma^{\ell+1})_{i,j} \;=\; [(\ell + 1)\gamma = i - j \bmod n] \tag{3.32}$$

Let

$$nh \in n\mathbb{Z} \tag{3.33}$$

be such that

$$\gamma = k - j + nh \tag{3.34}$$

assuming that

$$\gamma = k - j \bmod n \tag{3.35}$$

Then by the inductive hypothesis and the definition of matrix multiplication

$$\begin{aligned}
(R_\gamma^{\ell+1})_{i,j} &= \sum_k (R_\gamma^\ell)_{i,k}(R_\gamma)_{k,j} && (3.36)\\
&= \sum_k [\ell\gamma = i - k \bmod n][\gamma = k - j \bmod n] && (3.37)\\
&= \sum_k [\ell\gamma \in i - k + n\mathbb{Z}][\gamma = k - j + nh] && (3.38)\\
&= [\ell\gamma \in i - (\gamma + j - nh) + n\mathbb{Z}] && (3.39)\\
&= [\ell\gamma \in i - \gamma - j + n\mathbb{Z}] && (3.40)\\
&= [(\ell+1)\gamma \in i - j + n\mathbb{Z}] && (3.41)\\
&= [(\ell+1)\gamma = i - j \bmod n]. && (3.42)
\end{aligned}$$

Any $n \times n$ circulant matrix C is characterized by the existence of a vector c such that

$$C_{i,j} = c_{i-j \bmod n} \tag{3.43}$$

and thus is determined by its first column

$$c = Ce_0 \tag{3.44}$$

where e_0 is the first column of the identity matrix (a circulant matrix is differentiated from a *rectangular* circulant matrix by the requirement that a circulant matrix is square). Moreover, from Lemma A.16, C is circulant if and only if

$$C = F^* D F \tag{3.45}$$

where D is diagonal; in fact,

$$D = \sqrt{n}\,\mathrm{diag}(Fc) \tag{3.46}$$

(see Appendix), where F^* is the conjugate transpose of F, and F is the Fourier transform matrix

$$F_{i,j} \;=\; \frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}\,ij/n} \tag{3.47}$$

These facts may be used to computational advantage when implementing our method, since the circulant matrices involved may be diagonalized (and computations may be conducted in the diagonalizing basis).

Define the $\gamma \times a$ matrix Q by

$$Q = \begin{pmatrix} I_a \\ \hline 0 \end{pmatrix} \tag{3.48}$$

where I_k is the $k \times k$ identity matrix. Hence the previous constraint

$$\gamma \geq a \tag{3.49}$$

Let E be the $n \times a\alpha$ matrix

$$E = \begin{pmatrix} I_\alpha \otimes Q \\ \hline 0 \end{pmatrix} \tag{3.50}$$

where $\otimes$ denotes the Kronecker product [53]. Hence the previous constraint

$$n \geq \alpha\gamma \tag{3.51}$$

If A is an $i \times j$ matrix and B is a $u \times v$ matrix, then the Kronecker product $A \otimes B$ is the $iu \times jv$ block matrix

$$\begin{pmatrix} a_{11}B & \dots & a_{1j}B \\ \vdots & \ddots & \vdots \\ a_{i1}B & \cdots & a_{ij}B \end{pmatrix} \tag{3.52}$$

It follows (by partitioning the vector f below) that

$$Hf \;=\; (H_0 \mid \cdots \mid H_{w-1}) \begin{pmatrix} f_0 \\ \vdots \\ f_{w-1} \end{pmatrix} \;=\; \sum H_k f_k \;=\; \sum C_k E f_k \tag{3.53}$$

where C_k is the circulant matrix whose first column is that of $T_{k,0}$. This is a consequence of the equality

$$H_k b = C_k E b \tag{3.54}$$

(for all k and any $a\alpha \times 1$ vector b), which is verified by partitioning b. Note that

$$(C_k E b)_i = \sum_j (C_k)_{i,j} \begin{pmatrix} Qb_0 \\ \vdots \\ \hline Qb_{\alpha-1} \\ 0 \end{pmatrix}_j \tag{3.55}$$

from the definition of matrix multiplication, E, Q, and b. The right hand side of Equation (3.55) is equal to

$$\sum_{r<\alpha} \sum_{\ell<\gamma} (C_k)_{i,r\gamma+\ell} \, (Qb_r)_\ell \tag{3.56}$$

from the fact that Eb is zero filled after $\gamma\alpha$ entries, Q has γ rows, and the fact that $n > \alpha\gamma$. Expression (3.56) is equal to

$$\sum_{r<\alpha} \sum_{\ell<\gamma} (T_{k,0})_{i-r\gamma-\ell \bmod n,\, 0} \begin{pmatrix} b_r \\ \hline 0 \end{pmatrix}_\ell \tag{3.57}$$

from the fact that C_k is the circulant matrix whose first column is that of $T_{k,0}$, the fact that any $n \times n$ circulant matrix C is characterized by the existence of a vector c such that $C_{i,j} = c_{i-j \bmod n}$ and thus is determined by its first column $c = Ce_0$, and the definition of Q. Expression (3.57) is equal to

$$\sum_{r<\alpha} \sum_{\ell<a} (T_{k,0})_{i-r\gamma \bmod n,\ell} \, (b_r)_\ell \tag{3.58}$$

from the fact that $\gamma > a$, the i, j th entry of $T_{k,\ell}$ is some function of $i - j \bmod n$ the fact that $i - r\gamma + n\mathbb{Z} - \ell + n\mathbb{Z} = i - r\gamma - \ell + n\mathbb{Z} + n\mathbb{Z}$, and the fact that b_r has a rows. Expression (3.58) is equal to

$$\sum_{r<\alpha} \sum_{\ell<a} \sum_{j<n} [r\gamma = i - j \bmod n] \, (T_{k,0})_{j,\ell} \, (b_r)_\ell \tag{3.59}$$

since $[r\gamma = i - j \bmod n]$ is true when $j = i - r\gamma \bmod n$. Expression (3.59) is equal to

$$\sum_{r<\alpha} \sum_{\ell<a} \sum_{j<n} (R_\gamma^r)_{i,j} (T_{k,0})_{j,\ell} \, (b_r)_\ell \tag{3.60}$$

from the definition of R_γ^r. Expression (3.60) is equal to

$$(\sum_r R_\gamma^r \, T_{k,0} b_r)_i \tag{3.61}$$

from the definition of matrix multiplication. Expression (3.61) is equal to

$$(\sum_r T_{k,r} b_r)_i \tag{3.62}$$

from the definition of $T_{k,r}$. Expression (3.62) is equal to

$$(H_k b)_i \tag{3.63}$$

from the definition of H_k.

Define the partitioned matrix $\mathcal{H}$ by

$$\mathcal{H} = (C_0 | \cdots | C_{w-1}) \tag{3.64}$$

and note that by what has been shown above

$$x \;=\; Hf \;=\; \mathcal{H}(I_w \otimes E) f \tag{3.65}$$

This is because

$$\mathcal{H}(I_w \otimes E)f = (C_0|\cdots|C_{w-1}) \begin{pmatrix} E & & \\ & \ddots & \\ & & E \end{pmatrix} \begin{pmatrix} f_0 \\ \vdots \\ f_{w-1} \end{pmatrix} = \sum C_k E f_k \tag{3.66}$$

Since $\mathcal{H}$ is composed of circulant blocks and because $\mathcal{H}^T\mathcal{H}$ is symmetric, one might therefore hope the system

$$h = \mathcal{H}^T x = \mathcal{H}^T\mathcal{H}\, y \tag{3.67}$$

$$y = (I_w \otimes E)f \tag{3.68}$$

is a reasonable alternative to $x = Hf$, since

$$f = (I_w \otimes E^T)y \tag{3.69}$$

follows from Equation (3.68), because $(I_w \otimes E^T)(I_w \otimes E) = I_m$. From Horn [53] and Equations (3.50), (3.48)

$$(I_w \otimes E^T)(I_w \otimes E) = I_w \otimes (E^T E) \tag{3.70}$$

$$= I_w \otimes (I_\alpha \otimes Q^T)(I_\alpha \otimes Q) \tag{3.71}$$

$$= I_w \otimes I_\alpha \otimes Q^T Q \tag{3.72}$$

$$= I_w \otimes I_\alpha \otimes I_a \tag{3.73}$$

$$= I_{w\alpha a} \tag{3.74}$$

$$= I_m \tag{3.75}$$

Hence $y = (I_w \otimes E)f \Rightarrow$

$$(I_w \otimes E^T)y = (I_w \otimes E^T)(I_w \otimes E)f = f \tag{3.76}$$

Note that Equation (3.68) and (3.69) imply that

$$y = (I_w \otimes E)(I_w \otimes E^T)y \tag{3.77}$$

Attempting to solve Equation (3.67) is problematic, since $\mathcal{H}^T\mathcal{H}$ may be singular, and a solution y satisfying Equation (3.67) might not satisfy Equation (3.77).

The first difficulty is addressed by considering the regularized problem

$$h \;=\; (\mu I + \mathcal{H}^T\mathcal{H})\, y \tag{3.78}$$

Choosing

$$G = \mu I \tag{3.79}$$

$$X = \mathcal{H}^T \tag{3.80}$$

$$J = I \tag{3.81}$$

and

$$O = \mathcal{H} \tag{3.82}$$

in the Sherman-Morrison-Woodbury identity[55]

$$(G + XJO)^{-1} = G^{-1} - G^{-1}X\left(J^{-1} + OG^{-1}X\right)^{-1}OG^{-1} \tag{3.83}$$

yields the following solution to 3.78

$$y \;=\; (\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H})\, h \tag{3.84}$$

In our numerical experiments

$$\mu = 0.01 \tag{3.85}$$

What is particularly nice about this is that the matrix

$$I + \mu^{-1}\sum C_k C_k^T \tag{3.86}$$

is circulant, and therefore its inverse is trivial to compute. From Lemma A.17, circulant matrices are closed under transpose, addition, multiplication, scaling (i.e., multiplication by scalars), powers, roots, and are trivially inverted (when nonsingular) [56] (see Appendix).

Define the $nw \times nw$ matrix M and the $nw \times n$ matrix P by naming the right hand side of Equation (3.84) to express it as

$$y \;=\; Mh \;=\; (\mu^{-1}I - PP^T)h \tag{3.87}$$

and note that since

$$M = (\mu I + \mathcal{H}^T\mathcal{H})^{-1} = \mu^{-1}I - PP^T \tag{3.88}$$

it follows that

$$M\mathcal{H}^T\mathcal{H} \;=\; \mu PP^T \tag{3.89}$$

where

$$P \;=\; \mu^{-1}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1/2} \tag{3.90}$$

Equation (3.89) follows from

$$\begin{aligned} M\mathcal{H}^T\mathcal{H} &= M(\mu I + \mathcal{H}^T\mathcal{H} - \mu I) && (3.91)\\ &= M(\mu I + \mathcal{H}^T\mathcal{H}) - M\mu I && (3.92)\\ &= I - M\mu I && (3.93)\\ &= I - M\mu && (3.94)\\ &= I - (I\mu^{-1} - PP^T)\mu && (3.95)\\ &= \mu PP^T && (3.96) \end{aligned}$$

The expression given above for P is justified by the following lemma

Lemma 3.3.1. *It is true that*

$$(\mu^{-1}I - PP^T) = (\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H}) \tag{3.97}$$

Proof. Note that

$$P = \mu^{-1}\mathcal{H}^T W^{-1/2} \tag{3.98}$$

where

$$W = I + \mu^{-1}\sum C_k C_k^T \tag{3.99}$$

Assuming that $W^{-1/2}$ is symmetric (which will be justified later),

$$\begin{aligned}
\mu^{-1}I - PP^T &= \mu^{-1}I - (\mu^{-1}\mathcal{H}^T W^{-1/2})(\mu^{-1}\mathcal{H}^T W^{-1/2})^T && (3.100)\\
&= \mu^{-1}I - (\mu^{-1}\mathcal{H}^T W^{-1/2})(\mu^{-1}W^{-1/2}\mathcal{H}) && (3.101)\\
&= \mu^{-1}I - \mu^{-2}\mathcal{H}^T(W^{-1/2}W^{-1/2})\mathcal{H} && (3.102)\\
&= \mu^{-1}I - \mu^{-2}\mathcal{H}^T W^{-1}\mathcal{H} && (3.103)\\
&= \mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H} && (3.104)
\end{aligned}$$

□

A sanity check at this point is that $\mu I + \mathcal{H}^T\mathcal{H}$ acts like the inverse of $\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H}$.

$$\begin{aligned}
& (\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H})(\mu I + \mathcal{H}^T\mathcal{H}) && (3.105)\\
= \;& I - \mu^{-1}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H} + \mu^{-1}\mathcal{H}^T\mathcal{H} && (3.106)\\
& -\mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H}\mathcal{H}^T\mathcal{H} && (3.107)\\
= \;& I - \mu^{-1}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H} + \mu^{-1}\mathcal{H}^T\mathcal{H} && (3.108)\\
& -\mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}((I + \mu^{-1}\mathcal{H}\mathcal{H}^T) - I)\mathcal{H}\mu && (3.109)\\
= \;& I - \mu^{-1}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H} + \mu^{-1}\mathcal{H}^T\mathcal{H} && (3.110)\\
& -\mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}((I + \mu^{-1}\sum C_k C_k^T) - I)\mathcal{H}\mu && (3.111)\\
= \;& I - \mu^{-1}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H} + \mu^{-1}\mathcal{H}^T\mathcal{H} && (3.112)\\
& -\mu^{-2}\mathcal{H}^T\mathcal{H}\mu + \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H}\mu && (3.113)\\
= \;& I && (3.114)
\end{aligned}$$

Note that from Lemmas A.12, A.13, and A.14,

$$F = F^T \tag{3.115}$$

$$F^* = \overline{F} \tag{3.116}$$

and

$$F^*F = I \tag{3.117}$$

(see Appendix).

Let W be the real symmetric matrix $I + \mu^{-1} \sum C_k C_k^T$. Because each C_k is circulant

$$C_k = F^* D_k F \tag{3.118}$$

where D_k is diagonal. Hence

$$D_k = F C_k F^* \tag{3.119}$$

Note that

$$\begin{aligned} FWF^* &= I + \mu^{-1} \sum F C_k F^* F C_k^T F^* && (3.120)\\ &= I + \mu^{-1} \sum D_k (F C_k F^*)^* && (3.121)\\ &= I + \mu^{-1} \sum D_k D_k^* && (3.122)\\ &= S && (3.123) \end{aligned}$$

where S is a positive real diagonal matrix. Let $W^{-1/2}$ denote the matrix $F^* S^{-1/2} F$ where the positive real diagonal matrix $S^{-1/2}$ is defined by

$$(S^{-1/2})_{i,i} = 1/\sqrt{S_{i,i}} \tag{3.124}$$

where we take the nonnegative square root. This is reasonable provided

$$W(W^{-1/2})^2 = I \tag{3.125}$$

Note that

$$(F^*SF)(F^*S^{-1/2}F)^2 = F^*SFF^*S^{-1/2}FF^*S^{-1/2}F \quad (3.126)$$

$$= F^*SS^{-1/2}S^{-1/2}F \quad (3.127)$$

$$= I \quad (3.128)$$

It will be shown that

$$W^{-1/2} = (W^{-1/2})^T \quad (3.129)$$

which was assumed in the proof of Lemma 3.3.1. Note that $S^{-1/2}$ is a real symmetric matrix, hence

$$(W^{-1/2})^* = (F^*S^{-1/2}F)^* = F^*S^{-1/2}F = W^{-1/2} \quad (3.130)$$

Therefore, it suffices to show that $F^*S^{-1/2}F$ is real. By Lemma (A.8), $F^*S^{-1/2}F$ is real. Hence $W^{-1/2}$ is symmetric as desired.

Equation (3.89) will be used below. The second difficulty with using the system (3.67) (a solution y satisfying Equation (3.67) might not satisfy Equation (3.77)) is dealt with in a heuristic manner as follows. Let

$$Z' = I - Z \quad (3.131)$$

where Z is defined as

$$Z = \mathrm{diag}((I_w \otimes E)\mathbf{1}_m) \quad (3.132)$$

and where $\mathbf{1}_m$ is the $m \times 1$ vector all of whose entries are 1. Define the vectors $\vec{y_i}$ and $\vec{v_i}$ as follows (these vectors are decorated with "hats" so as to distinguish them from scalars)

$$\vec{y_0} = ZMh \quad (3.133)$$

$$\vec{y_{i+1}} = \vec{y_i} + ZM\mathcal{H}^T\mathcal{H}\vec{v_i} \quad (3.134)$$

$$\vec{v_0} = Z'Mh \quad (3.135)$$

$$\vec{v_{i+1}} = Z'M\mathcal{H}^T\mathcal{H}\vec{v_i} \quad (3.136)$$

The following illustrates the effect of moving from i to $i+1$.

$$\begin{aligned}
\vec{v_i} &= I\vec{v_i} && (3.137)\\
&= MM^{-1}\vec{v_i} && (3.138)\\
&= M(\mu I + \mathcal{H}^T\mathcal{H})\vec{v_i} && (3.139)\\
&\approx M\mathcal{H}^T\mathcal{H}\vec{v_i} && (3.140)\\
&= IM\mathcal{H}^T\mathcal{H}\vec{v_i} && (3.141)\\
&= (Z+Z')M\mathcal{H}^T\mathcal{H}\vec{v_i} && (3.142)\\
&= ZM\mathcal{H}^T\mathcal{H}\vec{v_i} + Z'M\mathcal{H}^T\mathcal{H}\vec{v_i} && (3.143)\\
&= \vec{y_{i+1}} - \vec{y_i} + \vec{v_{i+1}} && (3.144)
\end{aligned}$$

Hence

$$\vec{y_i} + \vec{v_i} \approx \vec{y_{i+1}} + \vec{v_{i+1}} \tag{3.145}$$

The solution to Equation (3.78) (which is the regularized version of 3.67) is

$$Mh = (Z+Z')Mh = \vec{y_0} + \vec{v_0}. \tag{3.146}$$

The part of that solution which solves Equation (3.77) is $\vec{y_0}$ and the remainder is $\vec{v_0}$. The hope is that transitioning from $\vec{y_i} + \vec{v_i}$ to $\vec{y_{i+1}} + \vec{v_{i+1}}$ would approximately preserve the solution since

$$\vec{y_i} + \vec{v_i} \approx \vec{y_{i+1}} + \vec{v_{i+1}} \tag{3.147}$$

while simultaneously moving toward an approximate solution more nearly satisfying Equation (3.77). That would be the case if the sequence of remainders $v_0, v_1, \ldots$ converged to zero. If $\lim_{i\to\infty} v_i = 0$, then the sequence $\vec{y_i} + \vec{v_i}$ of approximate solutions would converge (if it *did* converge) to

$$\lim_{i\to\infty} \vec{y_i} + \vec{v_i} = \lim_{i\to\infty} \vec{y_i} = \vec{y_\infty} \tag{3.148}$$

Moreover, that limit is not only an approximate solution to Equation (3.67), it also satisfies Equation (3.77) because every $\vec{y_i}$ does from Lemma (A.1).

Even though our method works and is comparatively fast for the problems we have tried it on, one might wonder whether $\lim_{i\to\infty} \vec{y_i} + \vec{v_i}$ and $\lim_{i\to\infty} \vec{y_i}$ exist and if they are equal. Another worry is that the series below in the expression for $\vec{y_\infty}$ might not converge and the inverse below in the expressions for $\vec{y_\infty}$ might not exist. Moreover, the error per iteration from the approximation

$$\vec{y_i} + \vec{v_i} \approx \vec{y_{i+1}} + \vec{v_{i+1}} \tag{3.149}$$

while perhaps small, might accumulate over an infinite number of iterations to an unacceptable value. These details are sorted out in Chapter 4.

It follows from Equation (3.89) and the recursive equations above (assuming of course that the series below converges and that the inverse below exists) that

$$\begin{aligned}
\vec{y_\infty} &= \lim_{i\to\infty} \vec{y_i} && (3.150)\\
&= ZMh + \mu ZPP^T Z' \sum_{i=0}^{\infty} \{\mu(Z'P)(Z'P)^T\}^i \vec{v_0} && (3.151)\\
&= Z(I + \mu PP^T Z'\{I - \mu(Z'P)(Z'P)^T\}^{-1} Z')Mh && (3.152)
\end{aligned}$$

Equation (3.152) will be established in three steps. First, it will be shown by induction that

$$\vec{v_n} = \{\mu(Z'P)(Z'P)^T\}^n \vec{v_0} \tag{3.153}$$

Base case ($n = 0$):

$$\vec{v_0} = \{\mu(Z'P)(Z'P)^T\}^0 \vec{v_0} \tag{3.154}$$

Inductive step: using $M\mathcal{H}^T\mathcal{H} = \mu PP^T$ (which is Equation (3.89))

$$\begin{aligned}
\vec{v_n} &= \{\mu(Z'P)(Z'P)^T\}^n \vec{v_0} \Longrightarrow && (3.155)\\
\vec{v_{n+1}} &= Z'M\mathcal{H}^T\mathcal{H}\{\mu(Z'P)(Z'P)^T\}^n \vec{v_0} && (3.156)\\
&= Z'\mu PP^T\{\mu(Z'P)(Z'P)^T\}^n \vec{v_0} && (3.157)\\
&= Z'\mu PP^T Z'\{\mu(Z'P)(Z'P)^T\}^n \vec{v_0} && (3.158)
\end{aligned}$$

$$= \mu(Z'P)(Z'P)^T\{\mu(Z'P)(Z'P)^T\}^n\vec{v_0} \quad (3.159)$$

$$= \{\mu(Z'P)(Z'P)^T\}^{n+1}\vec{v_0} \quad (3.160)$$

Second, it will be shown by induction that:

$$\vec{y_n} = ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \quad (3.161)$$

Base case ($n = 0$):

$$\vec{y_0} = ZMh \quad (3.162)$$

$$= ZMh + \mu ZPP^TZ'\Big(0\Big)\vec{v_0} \quad (3.163)$$

$$= ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \quad (3.164)$$

Inductive step: using Equation (3.89) and $\vec{v_n} = \{\mu(Z'P)(Z'P)^T\}^n\vec{v_0}$,

$$\vec{y_n} = ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \Longrightarrow \quad (3.165)$$

$$\vec{y_{n+1}} = ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \quad (3.166)$$

$$+ ZM\mathcal{H}^T\mathcal{H}\vec{v_n} \quad (3.167)$$

$$= ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \quad (3.168)$$

$$+ \mu ZPP^TZ'\{\mu(Z'P)(Z'P)^T\}^n\vec{v_0} \quad (3.169)$$

$$= ZMh + \mu ZPP^TZ'\Big(\sum_{i=0}^{n}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \quad (3.170)$$

Finally, assuming the series below converges and the inverse below exists,

$$\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i = I + \{\mu(Z'P)(Z'P)^T\}\Big(\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i\Big) \quad (3.171)$$

$$\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i - \{\mu(Z'P)(Z'P)^T\}\Big(\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i\Big) \;=\; I \quad (3.172)$$

$$\Big(I - \{\mu(Z'P)(Z'P)^T\}\Big)\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i \;=\; I \quad (3.173)$$

$$\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i \;=\; \Big(I - \{\mu(Z'P)(Z'P)^T\}\Big)^{-1} \quad (3.174)$$

Therefore

$$\vec{y_\infty} \;=\; ZMh + \mu ZPP^TZ'\sum_{i=0}^{\infty}\{\mu(Z'P)(Z'P)^T\}^i\vec{v_0} \quad (3.175)$$

$$\;=\; Z(I + \mu PP^TZ'\{I - \mu(Z'P)(Z'P)^T\}^{-1}Z')Mh \quad (3.176)$$

Choosing

$$G = I \quad (3.177)$$

$$X = -Z'P \quad (3.178)$$

$$J = \mu I \quad (3.179)$$

and

$$O = P^TZ' \quad (3.180)$$

in the Sherman-Morrison-Woodbury identity to evaluate the inverse occurring in Equation (3.150) yields

$$\vec{y_\infty} \;=\; Z(I + \mu PP^TZ'\{I + P(\mu^{-1}I - P^TZ'P)^{-1}P^TZ'\})Mh \quad (3.181)$$

Note that evaluating $\vec{y_\infty}$ requires *only* that the inverse $(\mu^{-1}I - P^T Z' P)^{-1}$ can be multiplied by vectors on the right. Hence products of the form

$$(\mu^{-1}I - P^T Z' P)^{-1} v \tag{3.182}$$

can be obtained by applying a conjugate gradient method [55] to solve

$$v = (\mu^{-1}I - P^T Z' P) u \tag{3.183}$$

The application of the Sherman-Morrison-Woodbury identity above results in the inverse of a smaller object which results in a more efficient algorithm.

An alternate approach might be to precompute a partial singular value decomposition approximation to $(\mu^{-1}I - P^T Z' P)^{-1}$ since the objects involved depend only upon the system matrix. This alternate approach is one of the things that might be addressed in the future.

The first step in our heuristic method is to take the first approximation $\vec{f_0}$ to solving

$$x = Hf \tag{3.184}$$

as

$$\vec{f_0} \;=\; (I_w \otimes E^T)\, \vec{y}_\infty \tag{3.185}$$

where y_∞ is determined by Equation (3.181) subject to the condition that the embedded conjugate gradient method – which solves expressions of the form (3.183) – is terminated at the first local minimum of error (in solving Equation (3.183); see Section (3.5)) or at LIM number of conjugate gradient steps (LIM is a parameter, typically chosen less than 5).

Subsequent steps in our heuristic method proceed as follows. Let the vectors produced by previous steps be

$$\vec{f_0}, \ldots, \vec{f_i} \tag{3.186}$$

and let

$$\sum \eta_i \vec{f_i} \tag{3.187}$$

be the least-squares solution [55] to minimizing

$$\| x - H \sum \eta_j \vec{f_j} \| \tag{3.188}$$

Obtain $\vec{f_{i+1}}$ in the same manner as $\vec{f_0}$, except with respect to the problem

$$x - H \sum \eta_j \vec{f_j} = Hf \tag{3.189}$$

Note that

$$\begin{aligned}
& ||x - H \sum \eta_j \vec{f_j}||^2 && (3.190)\\
= \; & (x - H \sum \eta_j \vec{f_j})^T (x - H \sum \eta_j \vec{f_j}) && (3.191)\\
= \; & x^T x - x^T H \sum \eta_j \vec{f_j} - \sum \eta_j \vec{f_j}^T H^T x + \sum \eta_j \vec{f_j}^T H^T H \sum \eta_j \vec{f_j} && (3.192)\\
= \; & x^T x - 2x^T H \sum \eta_j \vec{f_j} - \sum \eta_j \vec{f_j}^T H^T H \sum \eta_j \vec{f_j} && (3.193)
\end{aligned}$$

Therefore

$$\begin{aligned}
& \frac{\partial}{\partial \eta_k} (||x - H \sum \eta_j \vec{f_j}||^2) && (3.194)\\
= \; & -2x^T H \vec{f_k} + \sum_{h,j} \vec{f_h}^T H^T H \vec{f_j} ([h = k]\eta_j + \eta_h [j = k]) && (3.195)\\
= \; & -2x^T H \vec{f_k} + \sum_{j} \vec{f_k}^T H^T H \vec{f_j} \eta_j + \sum_{h} \vec{f_h}^T H^T H \vec{f_k} \eta_h && (3.196)\\
= \; & -2x^T \vec{v_k} + 2 \sum_{j} \vec{v_j}^T \vec{v_k} \eta_j && (3.197)
\end{aligned}$$

where

$$\vec{v_i} = H \vec{f_i} \tag{3.198}$$

Setting the derivative to zero gives

$$\vec{v_k}^T x = \sum_i \vec{v_k}^T \vec{v_i} \eta_i \tag{3.199}$$

Let V be the partitioned matrix

$$V = (\vec{v_0}, \ldots, \vec{v_i}) \tag{3.200}$$

and let η be the vector

$$\eta = \begin{pmatrix} \eta_0 \\ \vdots \\ \eta_i \end{pmatrix} \tag{3.201}$$

Then the minimization condition (3.199) is equivalent to

$$V^T V \eta = V^T x \tag{3.202}$$

or

$$\eta = (V^T V)^{-1} V^T x \tag{3.203}$$

3.4 Experimental Results

We used a 4194304×174960 shift invariant matrix

$$H = (H_0 | \cdots | H_{23}) \tag{3.204}$$

approximating a calibrated CTIS system. Each H_i has 90 rectangular circulant blocks, each of which is 4194304×81. Whereas f would typically be interpreted as encoding 24 images of the same object at various wavelengths, that interpretation is completely artificial and totally irrelevant to the mathematical problem which is represented by Equation (3.1). The image recovery problem has pertinent characteristics which are:

- Information goes in, as represented by some vector f.
- That information becomes encoded in the vector

$$g = Hf + n_1 + n_2 \tag{3.205}$$

- One attempts to recover the original information f from the encoding g.

It serendipitously happens that there are a total of

$$24 \cdot 90 \cdot 81 = 174960 \tag{3.206}$$

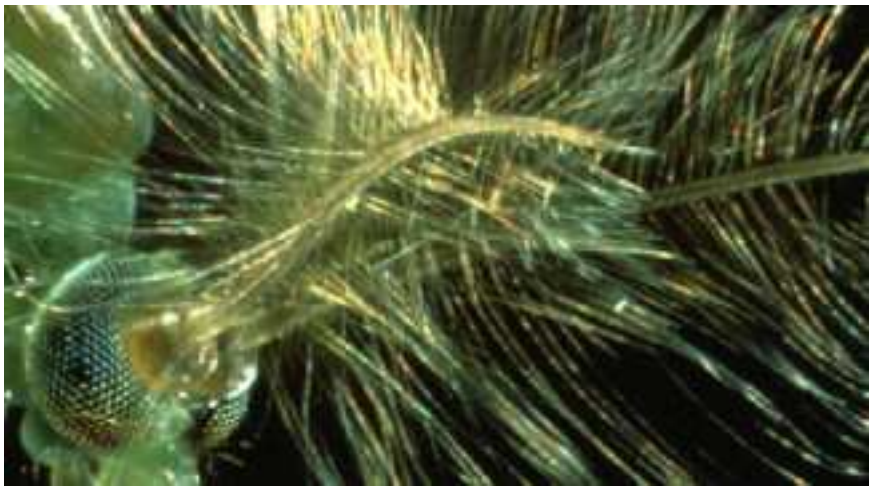

Figure 3.1: Initial image f: close-up of mosquito.

components of f, exactly enough to accommodate a 324 by 180 color ppm image (PPM Format Specification: http://netpbm.sourceforge.net/doc/ppm.html). For purposes of evaluating our method, this interpretation of f – as a single 324 by 180 color ppm image – is far superior to dealing with 24 monochromatic images (each 90 by 81 pixels) because they would be much too small to realistically portray a complicated image suitable to visually demonstrate the power of our method. The ppm image we used is a close-up of a mosquito (see Figure (3.1)). Figure (3.2) shows the focal plane array. The information in the ppm image shown in Figure (3.1) was assembled into f and multiplied by the system matrix, H to produce g in accordance with Equation (3.205) for the noise free case. The focal plane array corresponds to the image which g encodes in accordance with Equation (1.2).

Figures (3.2), (3.3), (3.4), (3.5), and (3.6) correspond to the noise-free case. Figure (3.3) shows average pixel error plotted against "iterations" together with standard deviation (plotted as vertical lines) for the following methods

- Multiplicative Algebraic Reconstruction Technique (MART).[16]
- Our heuristic method (Vose-Horton method).

Average pixel error is the average over all components of the absolute deviation between the image (i.e., components of f) and the recovered image. The meaning of "iteration" for

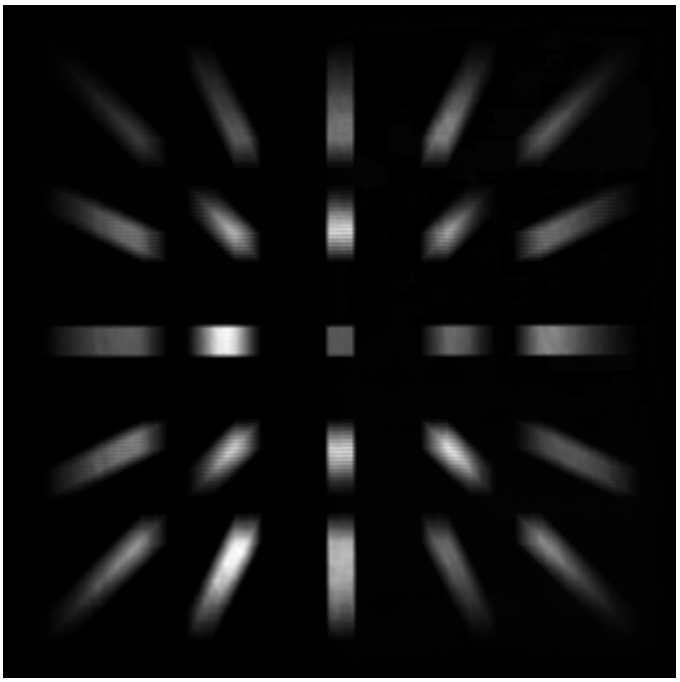

Figure 3.2: Focal plane, mosquito (no noise).

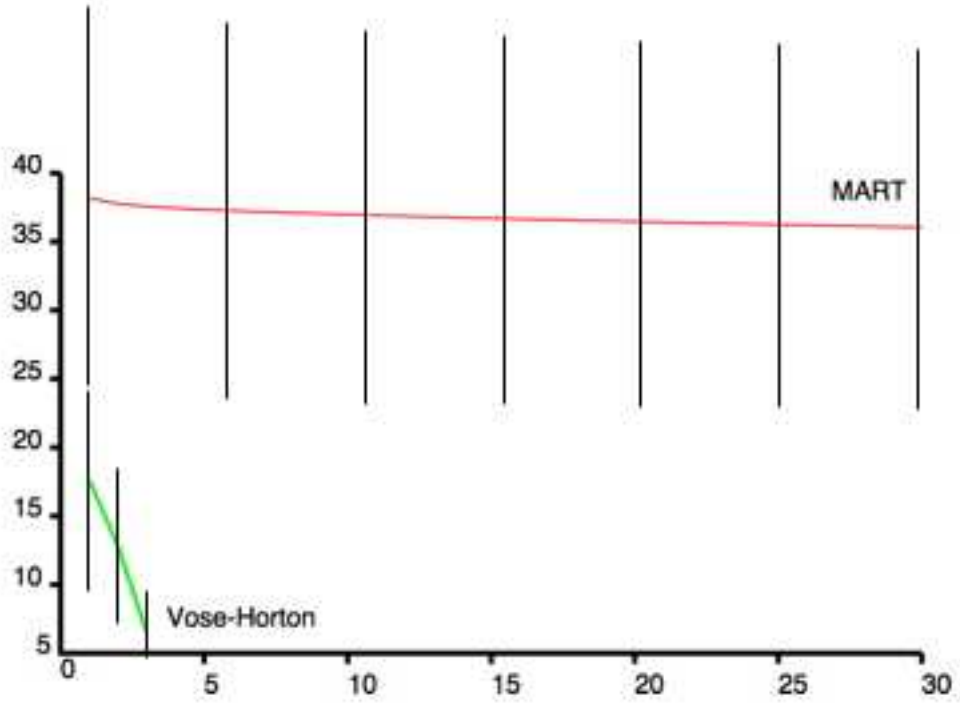

Figure 3.3: Average pixel error vs. iteration with standard deviation.

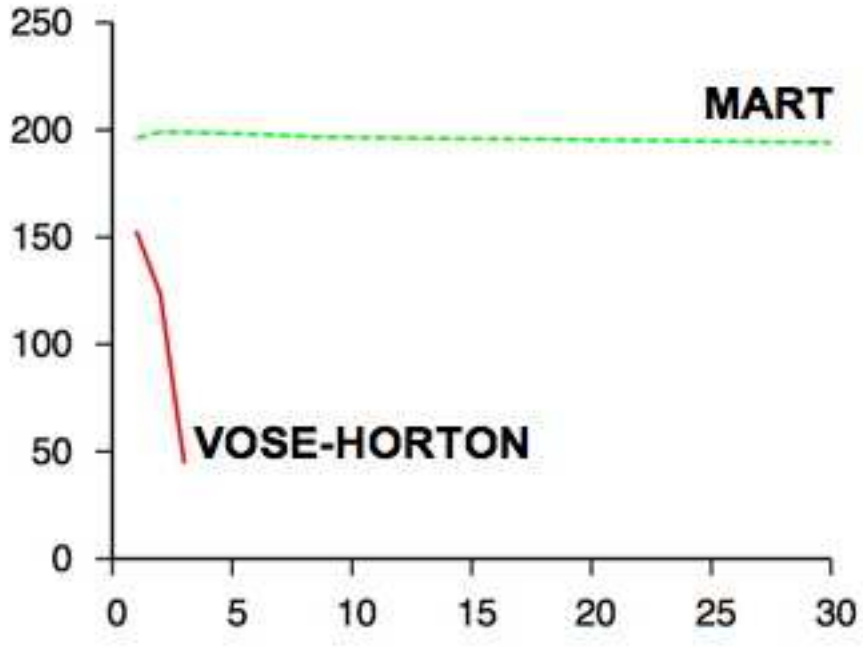

Figure 3.4: Max pixel error vs. iteration.

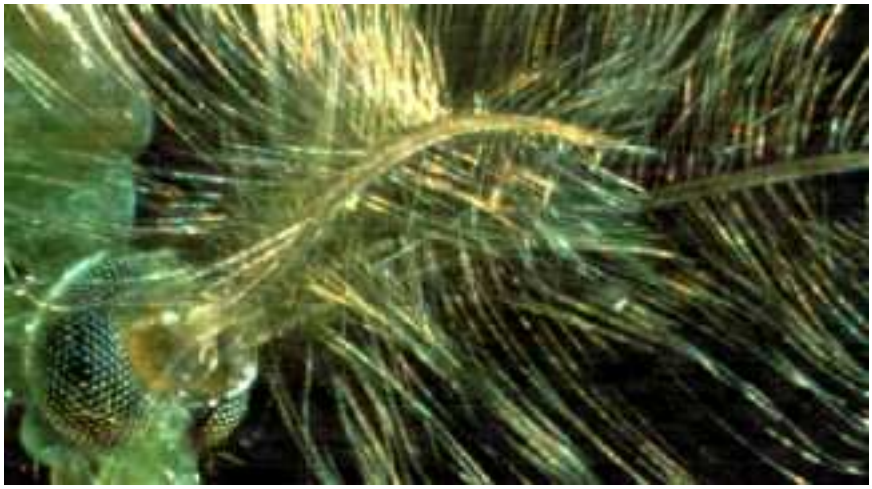

Figure 3.5: Vose-Horton method (3 iterations; 457 seconds).

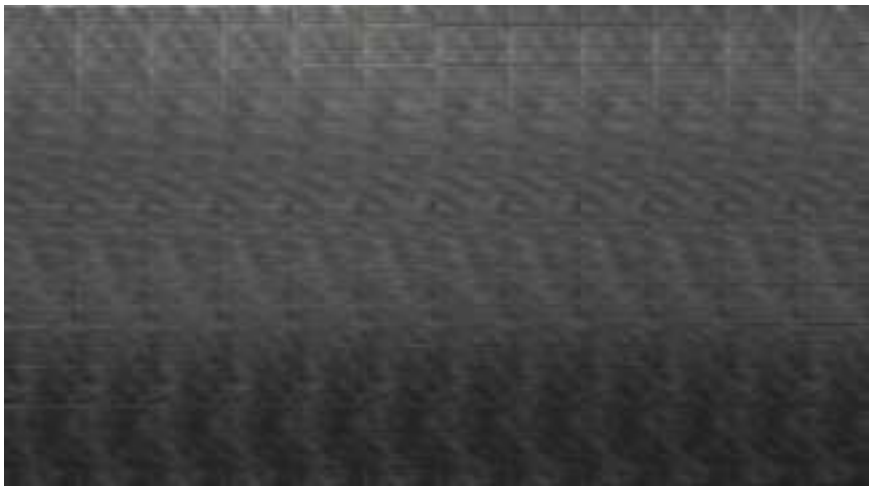

Figure 3.6: MART (30 iterations; 457 seconds).

MART is a transition from the current to the next estimate

$$f^{k+1} = f^k \frac{H^T g}{H^T H f^k} \tag{3.207}$$

The meaning of "iteration" for our method is a step as described in the previous section. Since an iteration of Vose-Horton takes longer than an iteration of MART, *both methods were run for an equal amount of time* (457 seconds executing on a gentoo gnu/linux dual-socket dual-core amd64 workstation), which corresponds to 30 iterations for MART, and 3 iterations for Vose-Horton. Note that precomputing objects which only rely on the system matrix H and work done by the fftw Fourier transform code to optimize its work is not included in our accounting for time because in both cases the results of that work can be stored on disk and used for any reconstruction problem for the same CTIS device. It is clear from Figure (3.3) that the Vose-Horton method results in a much lower average pixel error than mart. Note that the curve for MART is initially slightly concave up. This means that the improvement per iteration is decreasing. The curve for the Vose-Horton method, by contrast, is slightly concave down. This means that the improvement for iteration is increasing. For both MART and Vose-Horton, the standard deviation decreases with each iteration. However, for Vose-Horton, the standard deviation is decreasing at a faster rate. Figure (3.4) shows max pixel error plotted against iteration for MART and Vose-Horton.

Clearly the max pixel error drops sharply for Vose-Horton and remains large for MART. This agrees with what is seen in the reconstructed images. See Figure (3.5) and Figure (3.6). Comparing Figures (3.1) and (3.5), there is a clear visual indication that the Vose-Horton method has done a good job of reconstructing the original image in the amount of time allotted. Comparing Figures (3.1) and (3.6), there is a clear visual indication that MART has done a poor job of reconstructing the original image in the amount of time allotted. These visual indications agree with the plot in Figure (3.3).

Figures (3.7), (3.8), (3.9), (3.10), and (3.11) correspond to the case where noise corrupts the image. As was done by Garcia and Dereniak [49], we take

$$\sigma_s = 2 \tag{3.208}$$

The Vose-Horton method is compared with the mixed-expectation image-reconstruction technique (MERT) [49]. *Both methods were run for an equal amount of time (470 seconds)*, which corresponds to 21 iterations for MERT, and 4 iterations for Vose-Horton. The meaning of "iteration" for MERT is a transition from the current to the next estimate

$$f_i^{k+1} = f_i^k \left(\frac{\displaystyle\sum_{j=1}^{n} \frac{g_j^2 + 2g_j\sigma_s^2)H_{ji}}{[(Hf^k)_j + \sigma_s^2]^2}}{\displaystyle\sum_{j=1}^{n} \frac{[(Hf^k)_j^2 + 2(Hf^k)_j\sigma_s^2]H_{ji}}{[(Hf^k)_j + \sigma_s^2]^2}} \right) \tag{3.209}$$

Figure (3.7) shows the focal plane array. As in Figure (3.2), the information in the ppm image shown in Figure (3.1) was assembled into f and multiplied by the system matrix, H to produce g in accordance with Equation (3.205). Because this is the case where noise corrupts the image, noise is added to the resulting image in accordance with the noise model discussed in Section (3.2). The focal plane array corresponds to the image which g encodes in accordance with Equation (1.2). Figure (3.9) shows average pixel error plotted against "iterations" for (MERT) and Vose-Horton together with standard deviation (plotted as vertical lines). It is clear from Figure (3.9) that the Vose-Horton method results in a much lower average pixel error than MERT. Note that in contrast to the curve for MART in Figure (3.3), the curve for MERT oscillates. However, the shape

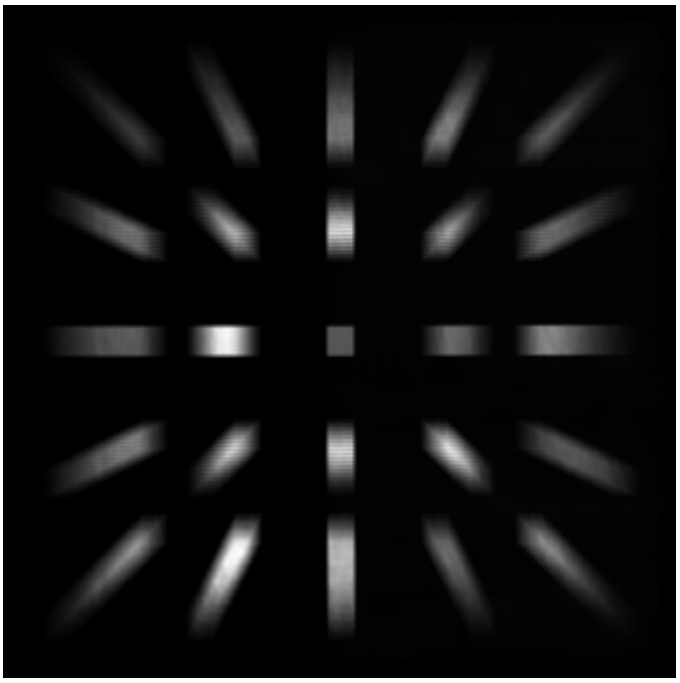

Figure 3.7: Focal plane, mosquito (noise).

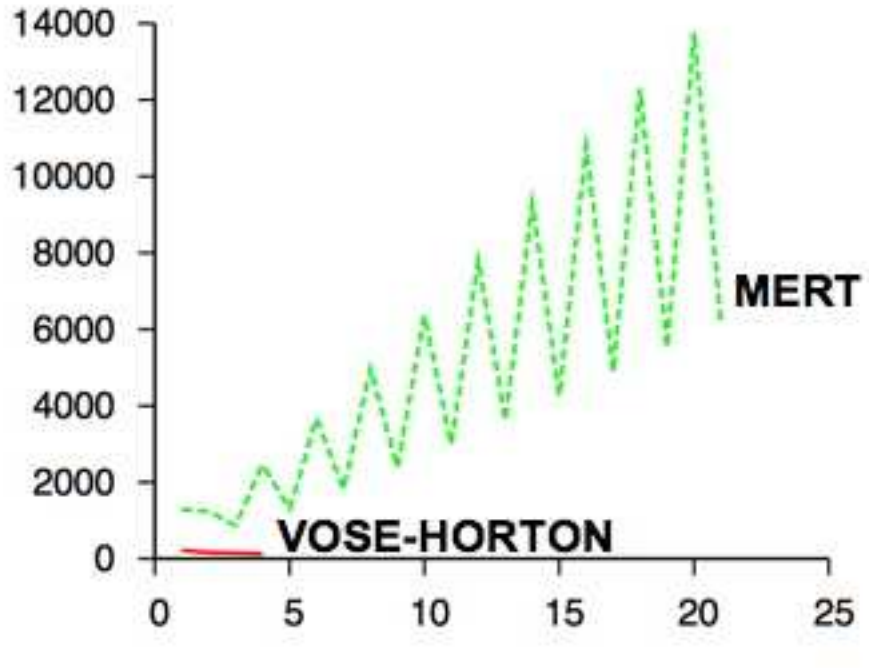

Figure 3.8: Max pixel error vs. iteration (noise level $\sigma_s = 2$).

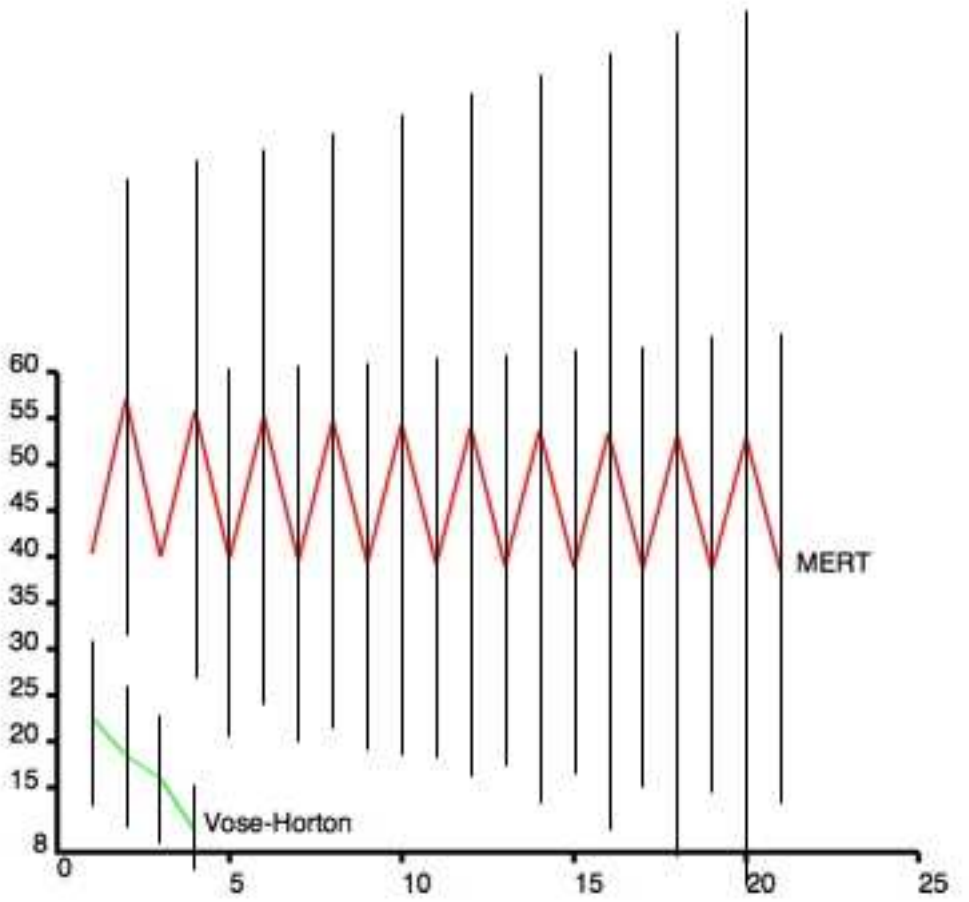

Figure 3.9: Average pixel error vs. iteration (noise level $\sigma_s = 2$) with standard deviation.

Figure 3.10: Vose-Horton method (4 iterations, noise level $\sigma_s = 2$; 470 seconds).

Figure 3.11: MERT (21 iterations, noise level $\sigma_s = 2$; 470 seconds).

of the oscillation decreases slowly. The curve for the Vose-Horton method, by contrast decreases much more quickly. For Vose-Horton, the standard deviation decreases over time and for MERT the standard deviation increases over time. Figure (3.8) shows max pixel error plotted against iteration for MERT and Vose-Horton. The max pixel error steadily decreases with iteration for Vose-Horton and steadily increases for MERT. Figures (3.10) and (3.11) show the reconstructed images. Comparing Figures (3.1) and (3.10), there is a clear visual indication that the Vose-Horton method has done a good job of reconstructing the original image in the amount of time allotted. Comparing Figures (3.1) and (3.11), there is a clear visual indication that MART has done a poor job of reconstructing the original image in the amount of time allotted. These visual indications agree with the plot in Figure (3.9).

3.5 Conjugate Gradient Method

The following iterative scheme to compute a solution to

$$Ax = b \tag{3.210}$$

(where A is symmetric positive definite, and the notation $(\cdot\,,\cdot)$ in the algorithm below denotes inner product) is what we mean by a "conjugate gradient" method.

```
x = initial
p = r = b - Ax
e = (r,r)
while (e > epsilon){
  v = Ap
  a = e/(p,v)
  x = x + a*p
  r = r - a*v
  a = (r,r)
  p = r + (a/e)*p
  e = a
}
```

The definition of "local minimum of error" (referred to in the description of our method) is the first local minimum in the sequence e_1, e_2, e_3, ... where e_i is the value of the variable `e` at the ith iteration of the while loop in the code above.

3.6 Summary

An iterative method is presented for CTIS image-reconstruction in the presence of both photon noise in the image and post-detection Gaussian system noise. The new algorithm, which assumes the transfer matrix of the system has a particular structure, is evaluated through computer simulations which demonstrate that, for larger problems, it is significantly better than both MART and MERT with respect to accuracy and computation time.

Chapter 4

Error Analysis

In this chapter it will be shown that the sequence of remainders $v_0, v_1, \ldots$ (see Equations (3.200), (3.136)) converges to zero. It will be shown that the limit (3.150) and the series (3.151) exist and are equal. It will be shown that the inverse in expression (3.152) exists. An expression for the error ϵ (the amount by which the approximate solution $\vec{y_\infty}$ deviates from satisfying equality (3.78)) will be developed. A geometric interpretation of that error will be shown. Finally, an expression for the relative error ϵ_f, in the reconstructed image will be given.

4.1 Object Existence Proofs

Recall that

$$Z' = I - Z \tag{4.1}$$

where Z is defined as

$$Z = \text{diag}((I_w \otimes E)\mathbf{1}_m) \tag{4.2}$$

and where $\mathbf{1}_m$ is the $m \times 1$ vector all of whose entries are 1.

$$\vec{y_\infty} = \lim_{i \to \infty} \vec{y_i} \tag{4.3}$$

$$= ZMh + \mu ZPP^T Z' \sum_{i=0}^{\infty} \{\mu(Z'P)(Z'P)^T\}^i \vec{v_0} \tag{4.4}$$

$$= Z(I + \mu PP^T Z' \{I - \mu(Z'P)(Z'P)^T\}^{-1} Z')Mh \tag{4.5}$$

It will be shown that the limit (4.3), series (4.4), and inverse in expression (4.5) exist.

Recall that

$$\vec{v_n} \;=\; \mu^n\{(Z'P)(Z'P)^T\}^n\vec{v_0} \tag{4.6}$$

The conditions under which $\vec{v_n}$ converges to 0 will be discussed. The object $\{(Z'P)(Z'P)^T\}$ is symmetric and is real if P is real. Recall that

$$P \;=\; \mu^{-1}\mathcal{H}^T(I+\mu^{-1}\sum C_kC_k^T)^{-1/2} \tag{4.7}$$

and W is the real symmetric matrix $I+\mu^{-1}\sum C_kC_k^T$. It was shown earlier that $W^{-1/2}$ is real. Therefore P is real and $\{(Z'P)(Z'P)^T\}$ can be written in the form UDU^T where D is a real diagonal matrix and U is real unitary. Furthermore,

$$\begin{aligned}\vec{v_n} = \mu^n\{(Z'P)(Z'P)^T\}^n\vec{v_0} &= \mu^n(UDU^T)^n\vec{v_0}\\ &= U(\mu D)^nU^T\vec{v_0}\end{aligned} \tag{4.8}$$

The elements of the matrix D are the eigenvalues of $\{(Z'P)(Z'P)^T\}$, and $\max_i(|D_{ii}|)$ is the spectral radius, $\rho(\{(Z'P)(Z'P)^T\})$ of $\{(Z'P)(Z'P)^T\}$. If

$$\mu\rho(\{(Z'P)(Z'P)^T\}) < 1 \tag{4.9}$$

then μD is a diagonal matrix all of whose entries have absolute value less than 1. It is immediate from Equation (4.6) and Equation (4.8) that $\vec{v_n}$ converges to 0 in that case.

Inequality (4.9) will now be established. Recall that

$$\mu PP^T = (\mu I+\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T\mathcal{H} \tag{4.10}$$

It follows that

$$\mu\rho(Z'PP^TZ') = \rho(Z'(\mu I+\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T\mathcal{H}Z') = \rho(Z'\phi Z') \tag{4.11}$$

where

$$\phi = (\mu I + \mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T\mathcal{H} \tag{4.12}$$

If $\mu > 0$ then by Lemmas (A.21) and (A.22) $\rho(\phi) < 1$. By Lemma A.24,

$$x \in \text{ spec}(Z'\phi Z') \Rightarrow x = u^T Z'\phi Z' u \tag{4.13}$$

where

$$||u|| = 1 \tag{4.14}$$

and spec(A) denotes the spectrum of A. Let

$$v = Z'u \tag{4.15}$$

and note that

$$||v||^2 \leq 1 \tag{4.16}$$

(by Lemma (A.25)) and

$$x = v^T\phi v \tag{4.17}$$

From Lemma A.22, $Z'\phi Z'$ has nonnegative real eigenvalues. Therefore the spectral radius is the maximal eigenvalue; let it be x.

$$\begin{aligned}
\rho(Z'\phi Z') &= u^T Z'\phi Z' u && (4.18)\\
&= v^T\phi v && (4.19)\\
&=_1 ||v||^2 \left(\frac{v}{||v||}\right)^T \phi \left(\frac{v}{||v||}\right) && (4.20)\\
&\leq ||v||^2 \sup_{||w||=1} w^T\phi w && (4.21)\\
&=_2 ||v||^2\rho(\phi) && (4.22)\\
&< ||v||^2 && (4.23)\\
&\leq 1 && (4.24)
\end{aligned}$$

[1] If $v = 0$ then $\rho(Z'\phi Z') = 0$ which is < 1

Now the conditions under which $\vec{y}_n$ converges will be discussed. It has been shown that for all n:

$$\vec{y_n} = ZMh + \mu ZPP^T\Big(\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i\Big)\vec{v_0} \tag{4.25}$$

Hence, $\vec{y_n}$ converges if

$$\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i \tag{4.26}$$

converges. Note that

$$\begin{aligned}
\sum_{i=0}^{n-1}\{\mu(Z'P)(Z'P)^T\}^i &= \sum_{i=0}^{n-1}\mu^i U D^i U^T && (4.27)\\
&= U(\sum_{i=0}^{n-1}\mu^i D^i)U^T && (4.28)\\
&= U(\sum_{i=0}^{n-1}\mu^i \sum_j e_j e_j^T D_{jj}^i)U^T && (4.29)\\
&= U(\sum_j \sum_{i=0}^{n-1} e_j e_j^T (\mu D_{jj})^i)U^T && (4.30)\\
&= U(\sum_j e_j (\sum_{i=0}^{n-1}(\mu D_{jj})^i) e_j^T)U^T && (4.31)
\end{aligned}$$

It was shown that if $\mu > 0$ then

$$\mu\rho(\{(Z'P)(Z'P)^T\}) < 1 \tag{4.32}$$

Therefore, by Lemma (A.22)

$$0 \leq \mu D_{jj} < 1 \tag{4.33}$$

for all j and as $n \to \infty$,

$$\sum_{i=0}^{n-1}(\mu D_{jj})^i = \frac{1-(\mu D_{jj})^n}{1-\mu D_{jj}} \longrightarrow \frac{1}{1-\mu D_{jj}} \tag{4.34}$$

[2] from Lemma (A.28)

Hence

$$\vec{y_n} \longrightarrow ZMh + \mu ZPP^T U(\sum_j e_j \frac{1}{1-\mu D_{jj}} e_j^T) U^T \vec{v_0} \tag{4.35}$$

Therefore the limit (4.3) and series (4.4) exist and are equal, provided $\mu > 0$.

To see that the inverse in expression (4.5) exists, let

$$\varphi = \mu(Z'P)(Z'P)^T \tag{4.36}$$

and assume that the inverse does not exist. Then for some

$$v \neq 0 \tag{4.37}$$

$$(I - \varphi)v = 0 \tag{4.38}$$

$$\Rightarrow \quad v = \varphi v \tag{4.39}$$

which implies that φ has eigenvalue 1 which is a contradiction because it was shown that if $\mu > 0$ then

$$\rho(\varphi) < 1 \tag{4.40}$$

Therefore the inverse (4.5) exists, provided $\mu > 0$.

4.2 Error in $\vec{y_\infty}$

It will be shown that the error ϵ (the amount by which the approximate solution $\vec{y_\infty}$ deviates from satisfying equality (3.78)) is given by

$$\epsilon = ||(I - \theta)^{-1} Z'(I - \theta)h|| \tag{4.41}$$

where

$$\phi = M\mathcal{H}^T\mathcal{H} \tag{4.42}$$

and

$$\theta = Z'\phi \tag{4.43}$$

It will be shown that the subspace

$$V' = \{(I-\theta)^{-1}Z'(I-\theta)h \ : \ h \in \mathbb{R}^{nw}\} \tag{4.44}$$

and the subspace

$$V'' = \{(I-\theta)^{-1}Z(I-\theta)h \ : \ h \in \mathbb{R}^{nw}\} \tag{4.45}$$

form a direct sum;

$$V' \oplus V'' = \mathbb{R}^{nw} \tag{4.46}$$

Not only that, V', V'' are orthogonal with respect to the scalar product

$$\langle u, v\rangle = u^T Y v \tag{4.47}$$

where

$$Y = (I-\theta^T)(I-\theta) \tag{4.48}$$

Moreover, the scalar product is positive definite, so that the angle between u and v can be defined as

$$\cos^{-1}\langle \frac{u}{||u||}, \frac{v}{||v||}\rangle \tag{4.49}$$

where

$$||u|| = \sqrt{\langle u, u\rangle} \tag{4.50}$$

Recall that $\vec{y_\infty}$ is an approximate solution to the regularized problem

$$(\mu I + \mathcal{H}^T\mathcal{H})y = M^{-1}y = h \tag{4.51}$$

such that

$$y = (I_w \otimes E)(I_w \otimes E^T)y \tag{4.52}$$

A solution y to Equation (4.51) satisfies

$$M^{-1}y = h \tag{4.53}$$

Define the error ϵ (the amount by which the approximate solution $\vec{y_\infty}$ deviates from satisfying that equality) by

$$\begin{aligned} \epsilon &= ||M^{-1}\vec{y_\infty} - h|| && (4.54)\\ &=_3 ||h - \mu(I-\theta)^{-1}Z'Mh - h|| && (4.55)\\ &= ||\mu(I-\theta)^{-1}Z'Mh|| && (4.56)\\ &= ||\mu(I-\theta)^{-1}\vec{v_0}|| && (4.57)\\ &=_4 ||(I-\theta)^{-1}Z'(I-\theta)h|| && (4.58) \end{aligned}$$

where

$$\phi = M\mathcal{H}^T\mathcal{H} \tag{4.59}$$

and

$$\theta = Z'\phi \tag{4.60}$$

The subspace

$$V' = \{(I-\theta)^{-1}Z'(I-\theta)h \ : \ h \in \mathbb{R}^{nw}\} \tag{4.61}$$

and the subspace

$$V'' = \{(I-\theta)^{-1}Z(I-\theta)h \ : \ h \in \mathbb{R}^{nw}\} \tag{4.62}$$

form a direct sum;

$$V' \oplus V'' = \mathbb{R}^{nw} \tag{4.63}$$

The equality

$$V' + V'' = \mathbb{R}^{nw} \tag{4.64}$$

follows from

$$(I-\theta)^{-1}Z'(I-\theta)h + (I-\theta)^{-1}Z(I-\theta)h = (I-\theta)^{-1}(Z'+Z)(I-\theta)h = h \tag{4.65}$$

[3] from Lemma (A.3)

[4] from Lemma (A.4)

The equality $V' \cap V'' = 0$ follows from the following. Suppose $q \in V' \cap V''$, then

$$\begin{aligned} q &= (I-\theta)^{-1} Z (I-\theta) u && (4.66)\\ &= (I-\theta)^{-1} Z' (I-\theta) v && (4.67)\\ &= (I-\theta)^{-1} (I-Z)(I-\theta) v && (4.68)\\ &= (I-\theta)^{-1} (I-\theta) v - (I-\theta)^{-1} Z (I-\theta) v && (4.69)\\ &= v - (I-\theta)^{-1} Z (I-\theta) v && (4.70) \end{aligned}$$

$$\Longrightarrow$$

$$\begin{aligned} v &= q + (I-\theta)^{-1} Z (I-\theta) v && (4.71)\\ &= (I-\theta)^{-1} Z (I-\theta) u + (I-\theta)^{-1} Z (I-\theta) v && (4.72)\\ &= (I-\theta)^{-1} Z (I-\theta)(u+v) && (4.73) \end{aligned}$$

$$\Longrightarrow$$

$$\begin{aligned} q &= (I-\theta)^{-1} Z' (I-\theta) v && (4.74)\\ &= (I-\theta)^{-1} Z' (I-\theta)(I-\theta)^{-1} Z (I-\theta)(u+v) && (4.75)\\ &= (I-\theta)^{-1} Z' Z (I-\theta)(u+v) && (4.76)\\ &= 0 && (4.77) \end{aligned}$$

Moreover, V', V'' are orthogonal with respect to the scalar product

$$\langle u, v \rangle = u^T Y v \tag{4.78}$$

where

$$Y = (I - \theta^T)(I - \theta) \tag{4.79}$$

Since

$$a^T (I-\theta)^T Z' (I-\theta)^{-T} Y (I-\theta)^{-1} Z (I-\theta) b \;=\; a^T (I-\theta)^T Z' Z (I-\theta) b \tag{4.80}$$

$$= \quad a^T(I-\theta)^T 0(I-\theta)b \tag{4.81}$$

$$= \quad 0 \tag{4.82}$$

By Theorem (A.18) and (A.19), Y is positive definite iff $(I-\theta)$ is nonsingular. Note that $(I-\theta)$ is nonsingular if

$$\rho(\theta) = \rho(Z'\mu PP^T) < 1 \tag{4.83}$$

It was shown that if $\mu > 0$ then

$$\rho(\mu PP^T) = \rho(\phi) < 1 \tag{4.84}$$

Therefore

$$\rho(\phi Z') \quad \leq_5 \quad \sup_{||v||=1} ||\phi Z'v|| \tag{4.85}$$

$$= \quad \sup_{w, w=Z'v, ||v||=1} ||\phi w|| \tag{4.86}$$

$$= \quad \sup_{w, w=Z'v, ||v||=1} ||w|| \frac{||\phi w||}{||w||} \tag{4.87}$$

$$\leq_6 \quad \sup_{w \neq 0} \frac{||\phi w||}{||w||} \tag{4.88}$$

$$<_7 \quad 1 \tag{4.89}$$

Hence

$$\rho(\theta) = \rho(Z'\phi) = \rho((Z'\phi)^T) = \rho(\phi Z') < 1 \tag{4.90}$$

The angle between u and v is defined to be

$$\cos^{-1}\langle \frac{u}{||u||}, \frac{v}{||v||} \rangle \tag{4.91}$$

where

$$||u|| = \sqrt{\langle u, u \rangle} \tag{4.92}$$

[5] from Lemma (A.29)

[6] from Lemma (A.25)

[7] from Theorem (A.20)

This makes sense (the arccos above has appropriate argument) provided

$$-1 \leq \langle \frac{u}{||u||}, \frac{v}{||v||} \rangle \leq 1 \tag{4.93}$$

Let

$$r = (I - \theta)u \text{ and } b = (I - \theta)v \tag{4.94}$$

Then

$$\begin{aligned} \frac{u^T Y v}{\sqrt{u^T Y u}\sqrt{v^T Y v}} &= \frac{((I-\theta)u)^T((I-\theta)v)}{\sqrt{((I-\theta)u)^T(I-\theta)u\ ((I-\theta)v)^T(I-\theta)v}} && (4.95) \\ &= \frac{r^T b}{\sqrt{r^T r\ b^T b}} && (4.96) \\ &\in [-1, 1] && (4.97) \end{aligned}$$

by the Schwartz inequality

$$|v^T w| \leq \sqrt{v^T v}\sqrt{w^T w} \tag{4.98}$$

To summarize, the error

$$\epsilon = ||M^{-1}\vec{y_\infty} - h|| \tag{4.99}$$

is the norm of the orthogonal projection of

$$h = \mathcal{H}^T x \tag{4.100}$$

to the subspace V' with respect to the inner product

$$\langle u, v \rangle = u^T(1 - \theta^T)(1 - \theta)v \tag{4.101}$$

See Figures (4.1) and (4.2). Not only that, the subspaces V' and V'' are orthogonal with respect to the inner product above (4.101).

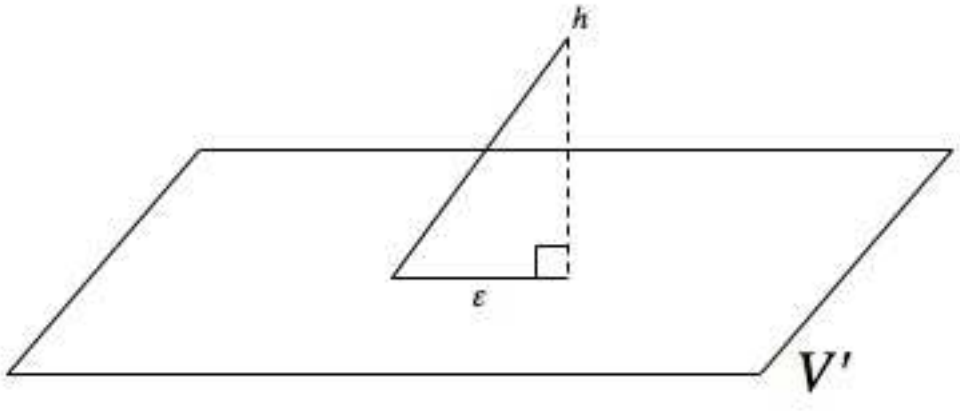

Figure 4.1: Geometric interpretation of error, ϵ

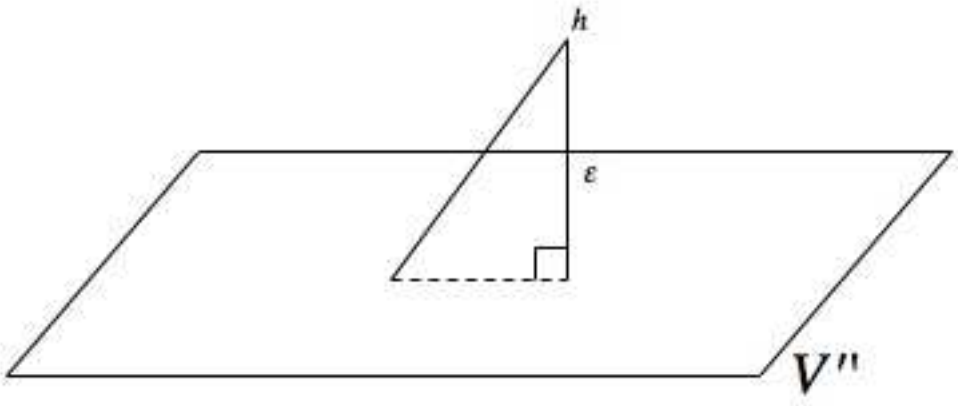

Figure 4.2: Geometric interpretation of error, ϵ

4.3 Error in $\vec{f_\infty}$

It will be shown that the relative error ϵ_f, in the reconstructed image $\vec{f_\infty}$ is given by

$$\epsilon_f = \frac{||f - \vec{f_\infty}||}{||f||} \leq ||(I - \phi)(I - \theta)^{-1}(I_w \otimes E)|| \tag{4.102}$$

where

$$\vec{f_\infty} = (I_w \otimes E^T)\vec{y_\infty} \tag{4.103}$$

Note that

$$\begin{aligned}
\epsilon_f &= \frac{||f - \vec{f_\infty}||}{||f||} && (4.104)\\
&=_8 \frac{||(I_w \otimes E)(f - \vec{f_\infty})||}{||f||} && (4.105)\\
&=_9 \frac{||(\mu M + (I - \phi)(I - \theta)^{-1}Z'(I - \mu M))(I_w \otimes E)f||}{||f||} && (4.106)\\
&=_{10} \frac{||(I - \phi)(I - \theta)^{-1}(I_w \otimes E)f||}{||f||} && (4.107)\\
&\leq_{11} ||(I - \phi)(I - \theta)^{-1}(I_w \otimes E)|| && (4.108)
\end{aligned}$$

as desired. Note that the error estimate is independent of the reconstructed image. Moreover, since the estimate is a function of the regularization parameter μ (see Equations (4.12) and (4.108)), the estimate might be optimized with respect to μ.

4.4 Summary

It was shown that the sequence of remainders $v_0, v_1, \ldots$ (see Equations (3.200), (3.136)) converges to zero. It was shown that the limit (3.150) and the series (3.151) exist and are equal. It was shown that the inverse in expression (3.152) exists. An expression for the error ϵ (the amount by which the approximate solution $\vec{y_\infty}$ deviates from satisfying equality (3.78)) was developed. A geometric interpretation of that error was shown. Finally, an

[8] from Lemma (A.6)

[9] from Lemma (A.5)

[10] from Lemma (A.7), (A.2)

[11] The matrix norm we use is $||A|| = \sup_{x \neq 0} ||Ax||/||x||$. It follows that $||Ax|| \leq ||A||\ ||x||$

expression for the relative error ϵ_f, in the reconstructed image was given which is uniform in the sense that it is independent of the reconstructed image.

Chapter 5

Performance Evaluation

5.1 Additional Testing

Figure (5.1) shows average pixel error plotted against iterations (the standard deviation is plotted as vertical lines) for MART and Vose-Horton for the noise free case when f is the image in Figure (5.3). Both methods were run for the same amount of time. It is clear from Figure (5.1) that the Vose-Horton method results in a much lower average pixel error than MART. Note that both the curve for MART and the curve for the Vose-Horton method is concave up. This means that the improvement per iteration is decreasing. The standard deviation decreases sharply for Vose-Horton and gradually for MART. Figure (5.2) shows the max pixel error. Clearly the max pixel error drops sharply for Vose-Horton but is somewhat flat for MART. Figures (5.4) and (5.5) show the reconstructed images. Comparing Figures (5.3) and (5.4), there is a clear visual indication that the Vose-Horton method has done a good job of reconstructing the original image in the amount of time allotted. Comparing Figures (5.3), (5.4), and (5.5), there is a clear visual indication that MART has done a fair job of reconstructed the original image in the time allotted but not as good of a job as the Vose-Horton method. Figure (5.6) shows the reconstructed image when MART was run for 42300 seconds (100 time longer than the Vose-Horton method). Comparing Figures (5.3), (5.4), and (5.6), there is a clear visual indication that MART performs as well as the Vose-Horton method when given 100 times as long.

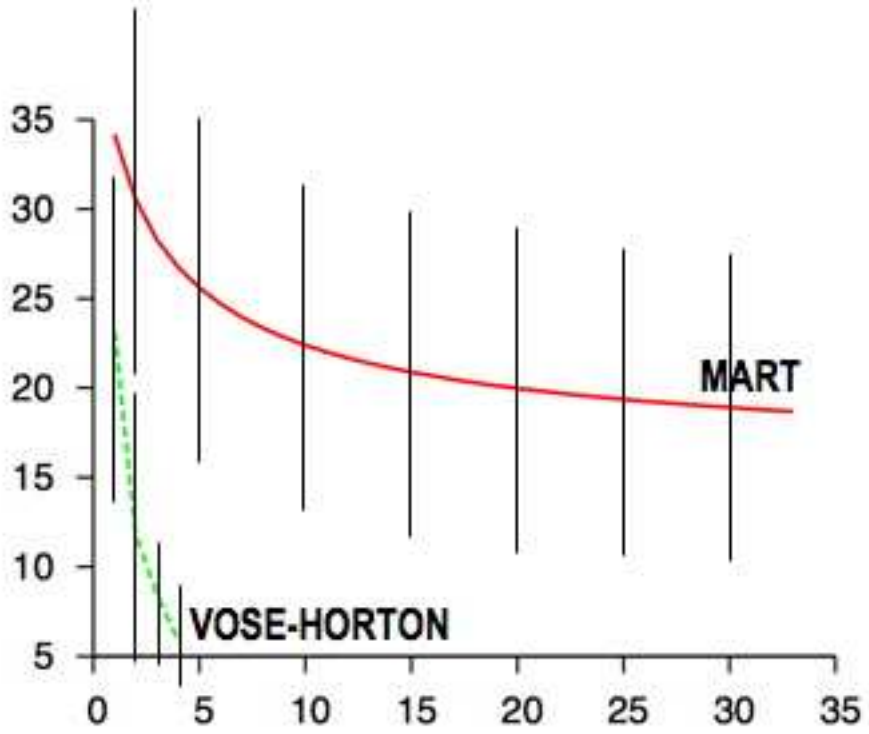

Figure 5.1: Average pixel error vs. iteration with standard deviation

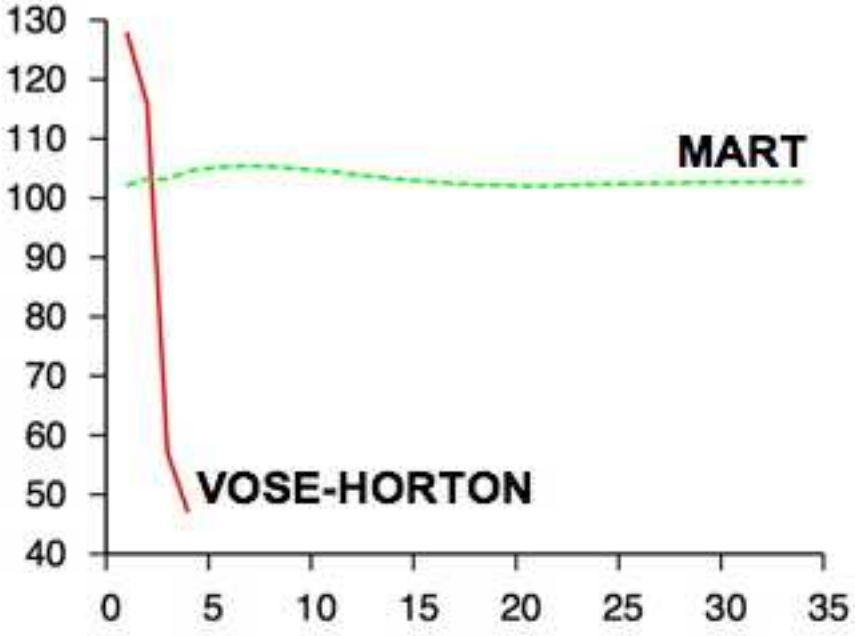

Figure 5.2: Max pixel error vs. iteration

Figure 5.3: Initial image f: 24 identical images

Figure 5.4: Vose-Horton method (4 iterations; 423 seconds).

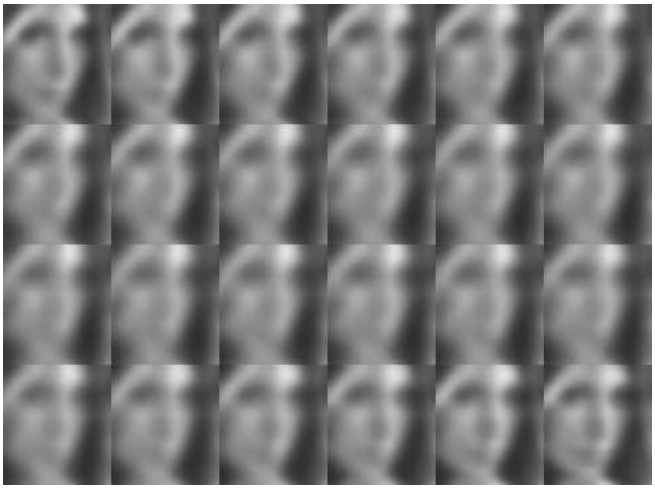

Figure 5.5: MART (33 iterations; 423 seconds).

Figure 5.6: MART (3100 iterations; 42300 seconds).

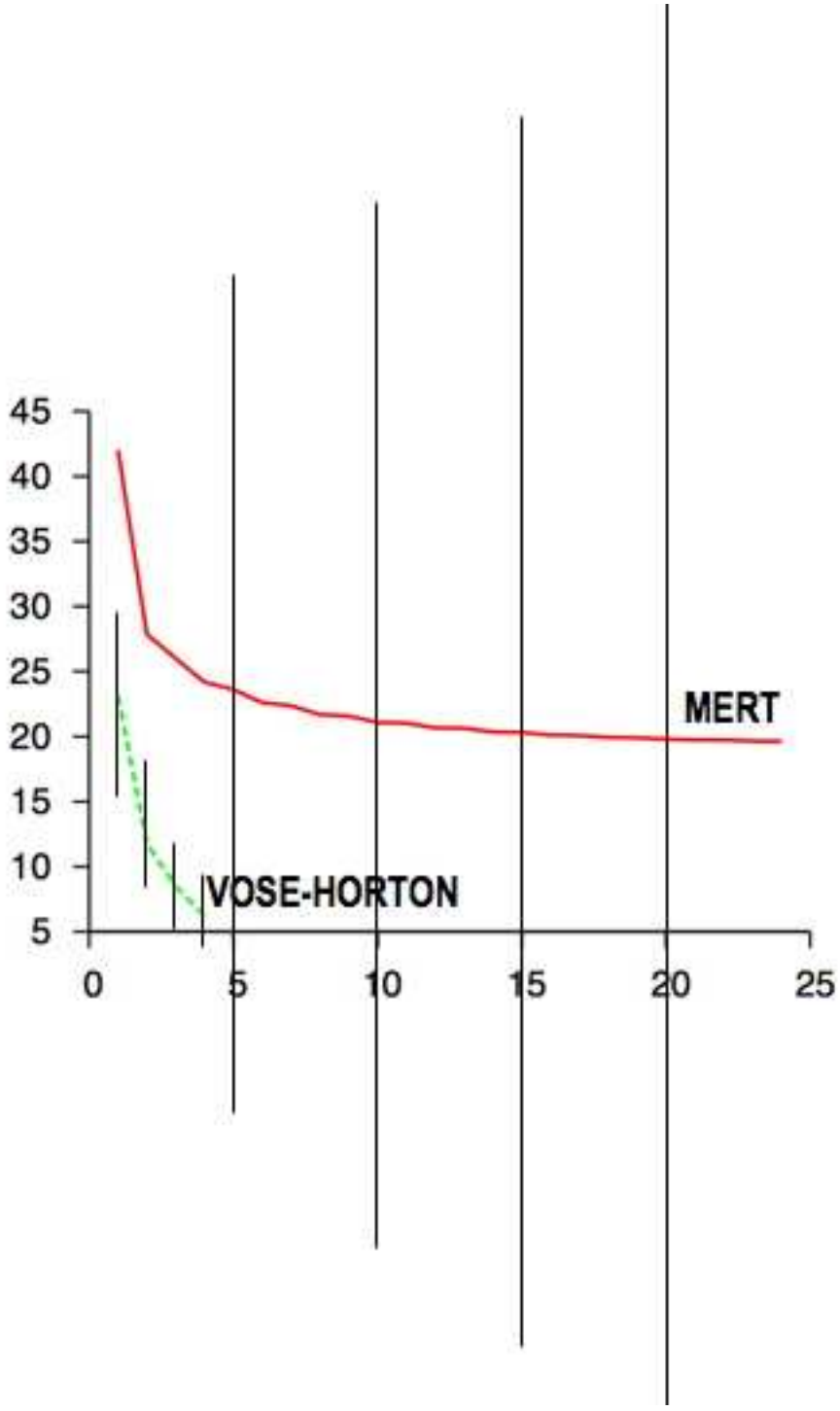

Figure 5.7: Average pixel error vs. iteration (noise level $\sigma_s = 2$) with standard deviation.

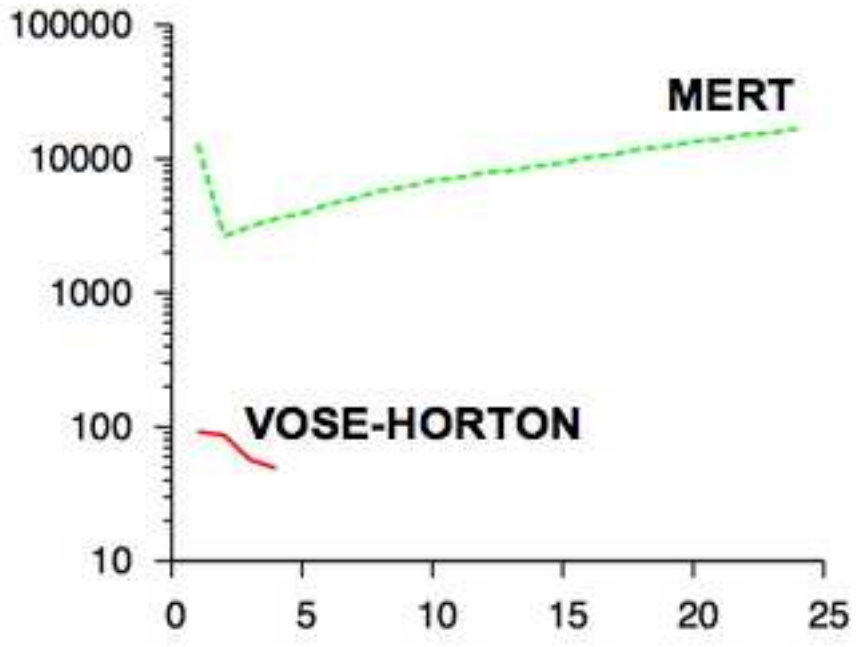

Figure 5.8: Max pixel error vs. iteration (noise level $\sigma_s = 2$).

Figure (5.7) shows average pixel error plotted against iterations (standard deviation is plotted as vertical lines) for MERT and Vose-Horton for the case where noise corrupts the image when f is the image in Figure (5.3). Both methods were run for the same amount of time. It is clear from Figure (5.7) that the Vose-Horton method results in a much lower average pixel error than MERT. Note that both the curve for MERT and the curve for the Vose-Horton method is concave up. This means that the improvement per iteration is decreasing. The standard deviation decreases sharply for Vose-Horton and increases steadily for MERT. Figure (5.8) shows the max pixel error. The max pixel error for Vose-Horton decreases over time. For MERT the max pixel error decreases sharply at first then steadily increases with time. Figures (5.9) and (5.10) show the reconstructed images. Comparing Figures (5.3) and (5.9), there is a clear visual indication that the Vose-Horton method has done a good job of reconstructing the original image in the amount of time allotted. Comparing Figures (5.3) and (5.10), there is a clear visual indication that MERT has done a poor job of reconstructing the original image in the time allotted. Figure (5.11)

Figure 5.9: Vose-Horton method (4 iterations, noise level $\sigma_s = 2$; 423 seconds).

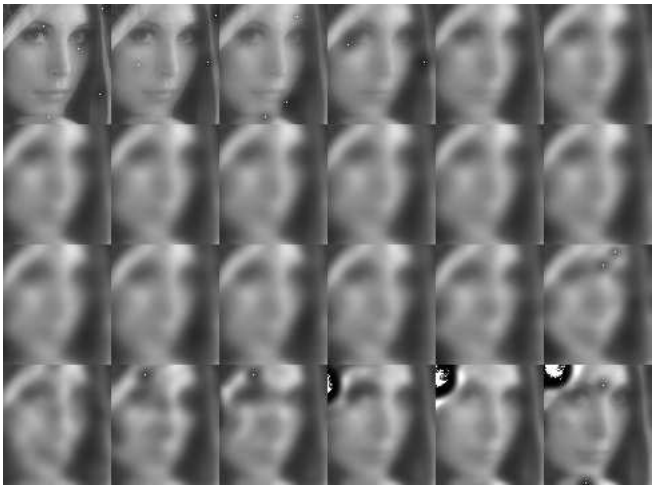

Figure 5.10: MERT (24 iterations, noise level $\sigma_s = 2$; 423 seconds).

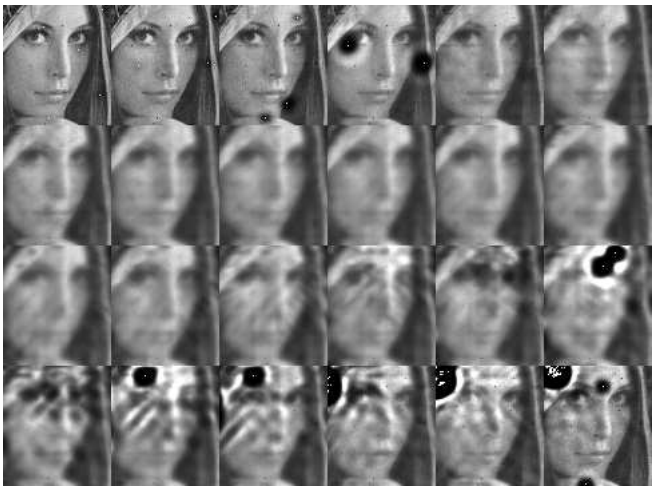

Figure 5.11: MERT (804 iterations, noise level $\sigma_s = 2$; 14170 seconds).

shows the reconstructed image when MERT was run for 14170 seconds (30 times longer than the Vose-Horton method). Comparing Figures (5.3), (5.10), and (5.11), there is a clear visual indication that MERT does a poor job of reconstructing the original image when run 30 times longer than the Vose-Horton method. Not only that, it has diverged in places.

5.2 Method Accuracy Metrics

We use two metrics to judge the accuracy of our method. The first metric is pixel error (the absolute deviation between the image and the recovered image). We look at the average pixel error, the standard deviation of the pixel error, and the max pixel error. The previous Figures (3.3), (3.4), (3.9), (3.8), (5.1), (5.2), (5.7), and (5.8) are examples of this metric. The second metric is peak signal to noise ratio [34] and is given by

$$PSNR = 10 \log_{10}[(L-1)/\sqrt{MSE}]$$

where L is the maximum value a pixel can have and MSE is the mean square error and is approximated [54] by:

$$MSE = \frac{1}{m} \sum_{i=0}^{m-1} [f(i) - \hat{f}(i)]^2$$

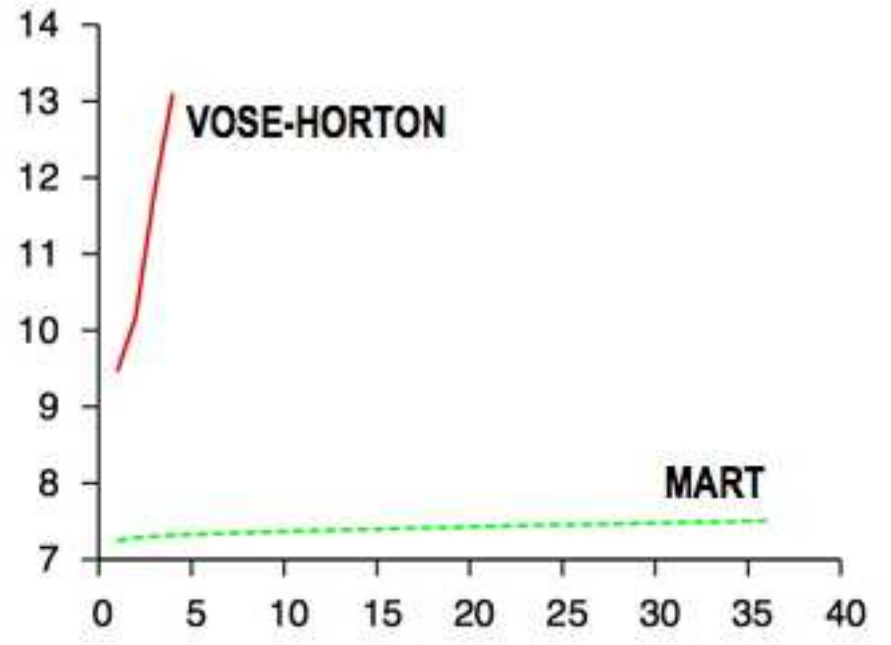

Figure 5.12: Peak signal to noise ratio vs. iteration.

where $\hat{f}$ represents the recovered image. Figures (5.12) and (5.13) show this metric. The figures are graphs plotting peak signal to noise ratio against iteration for the evaluated methods. It is clear from Figure (5.12) that the Vose-Horton method has a higher peak signal to noise ratio than MART when both methods are run for the same amount of time. This is consistent with what is seen in Figure (3.3). In other words, one might expect that less average pixel error might result in a higher peak signal to noise ratio. It is clear from Figure (5.13) that the Vose-Horton method has a higher peak signal to noise ratio than MERT when both methods are run for the same amount of time. This is consistent with what is seen in Figure (3.9).

5.3 Method Robustness

The robustness of the Vose-Horton method is compared to MART for an incorrectly calibrated system matrix. The miscalibrated matrix was obtained by adding Gaussian noise (zero mean, variance σ) to the correctly calibrated system matrix and replacing resulting

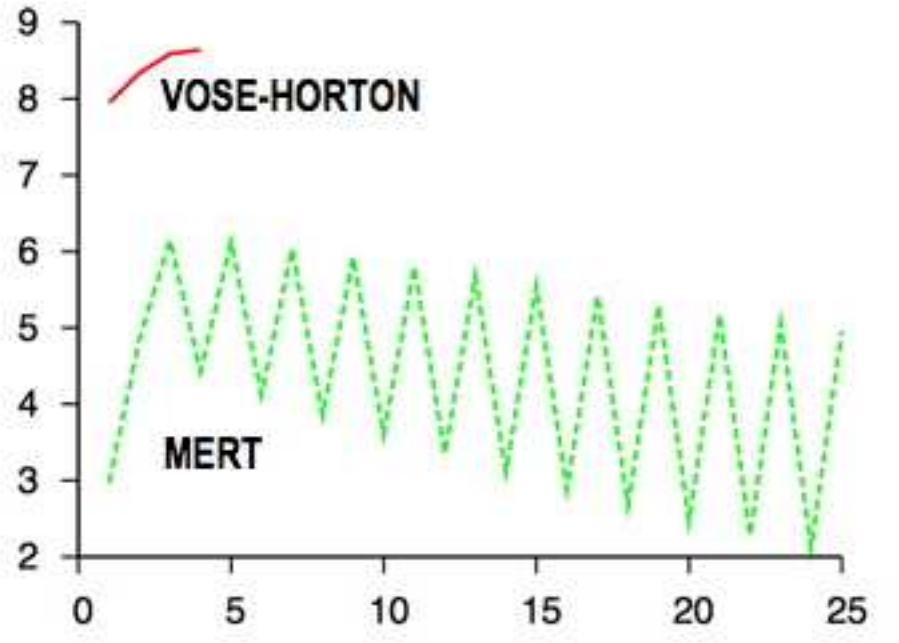

Figure 5.13: Peak signal to noise ratio vs. iteration (noise level $\sigma_s = 2$).

negative values with zero. Figure (5.14) shows the average pixel error (the average pixel error at the final iteration) vs. noise level for Vose-Horton and MART together with standard deviation (plotted as vertical lines). Both methods were run for the same amount of time. Figure (5.15) shows the max pixel error vs. noise level for Vose-Horton and MART. Figures (5.16), (5.17), (5.18), (5.19), (5.20), and (5.21) show the reconstructed images for Vose-Horton. Figures (5.22), (5.23), (5.24), (5.25), (5.26), and (5.27) show the reconstructed images for MART. Comparing Figures (3.1), (5.16), (5.17), (5.18), (5.19), (5.20), and (5.21), it is clear that for Vose-Horton, with the exception of the reconstructed image at noise level ($\sigma = .6$), the reconstructed image becomes visually farther away from the original image as the noise in the system matrix increases. This is consistent with what is seen in Figures (5.14) and (5.15). Comparing Figures (5.22), (5.23), (5.24), (5.25), (5.26), and (5.27), it is clear that for MART, the reconstructed image becomes steadily worse as the noise in the system matrix increases. This is consistent with what is seen in Figure (5.14). Note that the standard deviation appears to be increasing as noise in the system

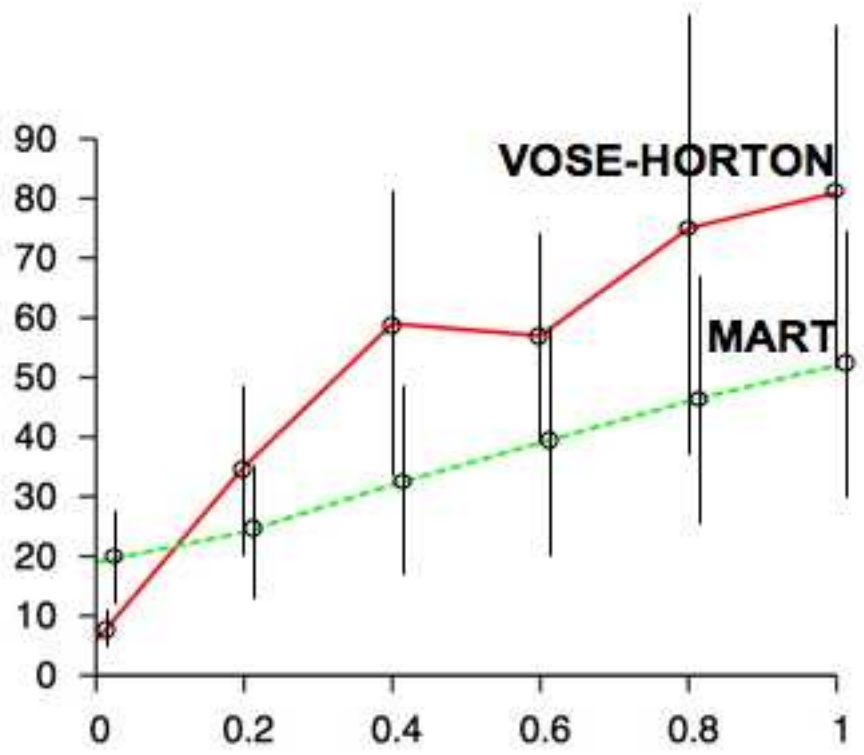

Figure 5.14: Average pixel error vs. system miscalibration with standard deviation.

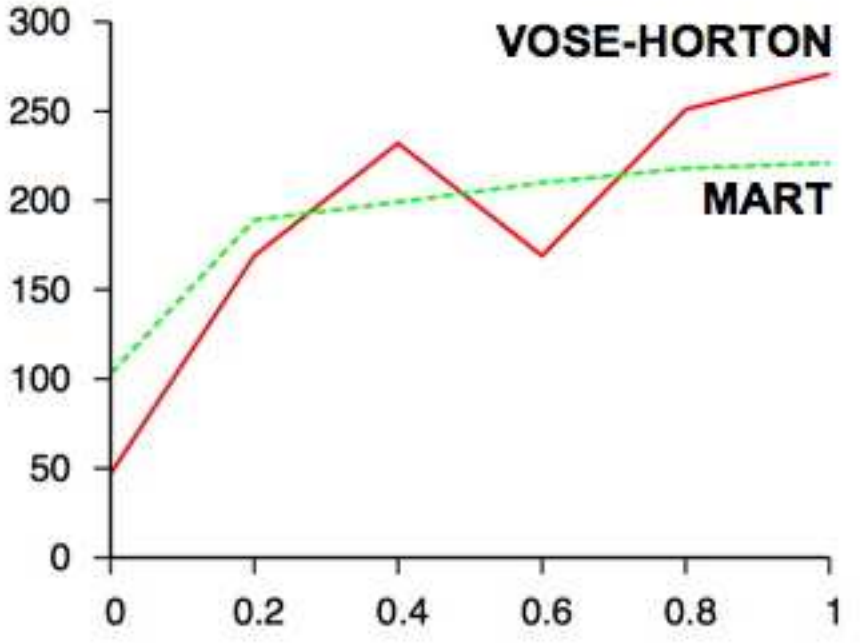

Figure 5.15: Max pixel error vs. system miscalibration.

Figure 5.16: Vose-Horton method, system miscalibration level $\sigma = 0$

Figure 5.17: Vose-Horton method, system miscalibration level $\sigma = .2$.

Figure 5.18: Vose-Horton method, system miscalibration level $\sigma = .4$.

Figure 5.19: Vose-Horton method, system miscalibration level $\sigma = .6$.

Figure 5.20: Vose-Horton method, system miscalibration level $\sigma = .8$.

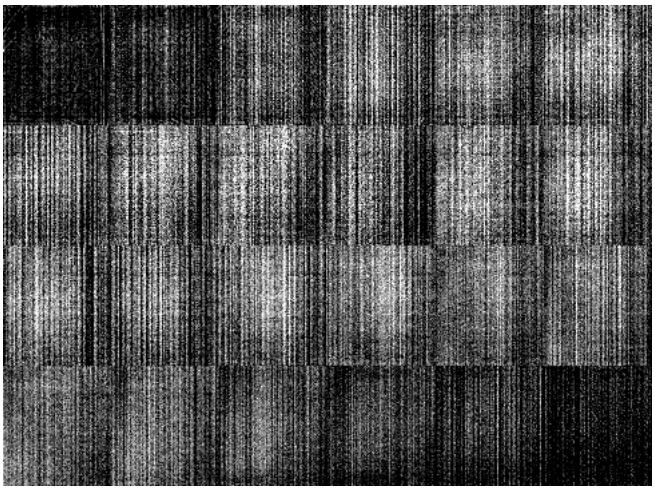

Figure 5.21: Vose-Horton method, system miscalibration level $\sigma = 1$.

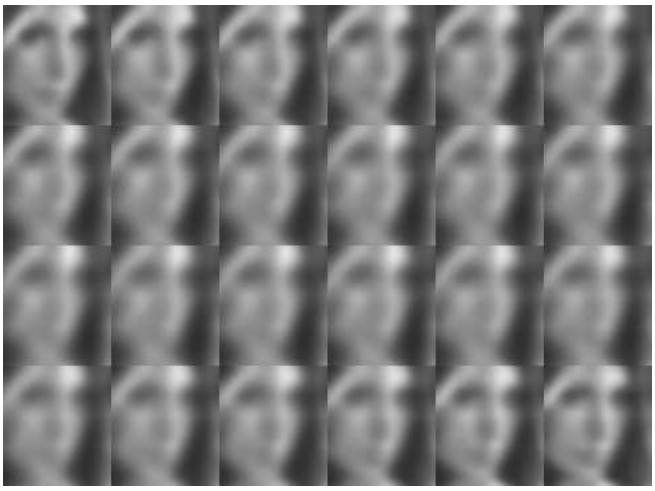

Figure 5.22: MART, system miscalibration level $\sigma = 0$

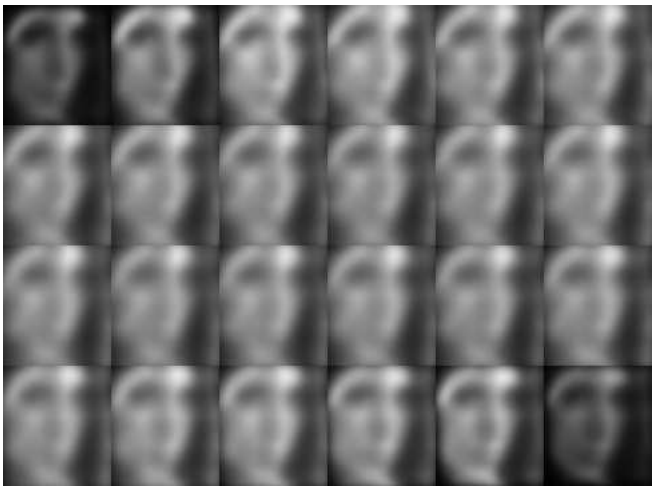

Figure 5.23: MART, system miscalibration level $\sigma = .2$.

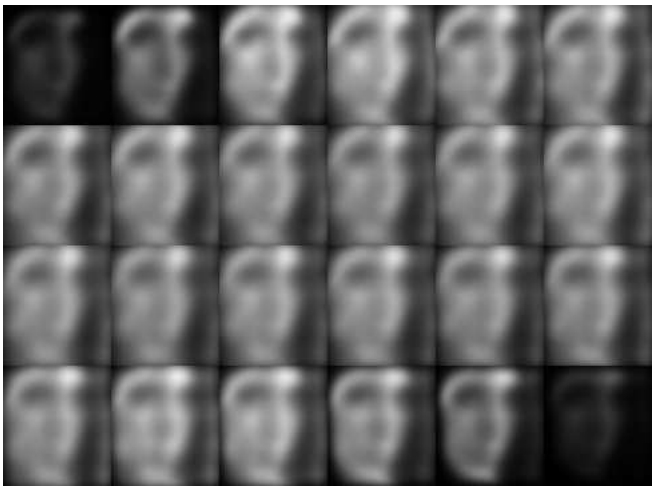

Figure 5.24: MART, system miscalibration level $\sigma = .4$.

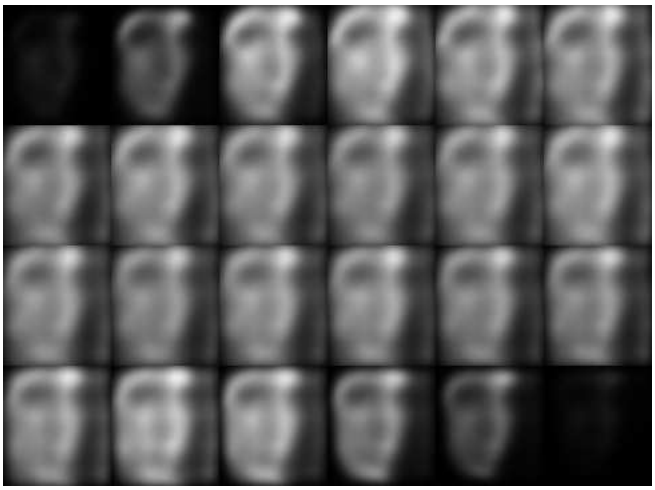

Figure 5.25: MART, system miscalibration level $\sigma = .6$.

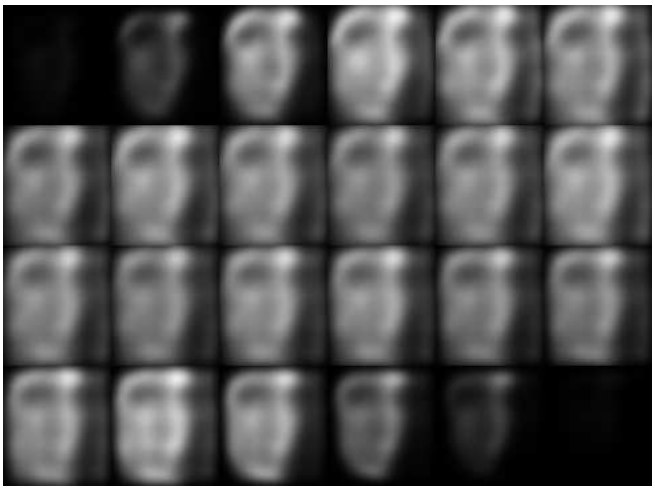

Figure 5.26: MART, system miscalibration level $\sigma = .8$.

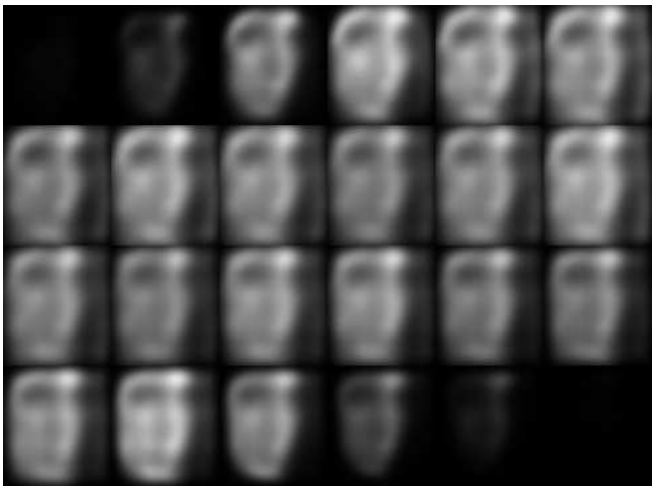

Figure 5.27: MART, system miscalibration level $\sigma = 1$.

matrix increases for both Vose-Horton and MART.

One way to interpret the robustness results so far is to note that reconstructed images for MART exhibit less contrast than those for Vose-Horton and to suggest that it might be the case that a miscalibrated system matrix favors reconstruction methods for which the reconstructed images exhibit less contrast.

The robustness of the Vose-Horton method is compared to MERT in the presence of noise in the system matrix. Figure (5.28) shows the average pixel error (the average pixel error at the final iteration) vs. noise level for Vose-Horton and MERT together with standard deviation. Both methods were run for the same amount of time. Figure (5.29) shows the max pixel error vs. noise level for Vose-Horton and MERT. Figures (5.30), (5.31), (5.32), (5.33), (5.34), and (5.35) show the reconstructed images for MERT. Comparing Figures (5.30), (5.31), (5.32), (5.33), (5.34), and (5.35), it is clear that for MERT, the reconstructed image becomes steadily worse as the noise in the system matrix increases. This is consistent with what is seen in Figure (5.28). Note that the standard deviation appears to be increasing as noise in the system matrix increases for Vose-Horton but decreasing for MERT. Note that in Figure (5.15), MART has a ceiling max pixel error of 220 and that in Figure (5.29), MERT results in much higher max pixel errors. Also note that the max pixel error increases for MART as noise in the system matrix increases, but for MERT the max pixel error decreases.

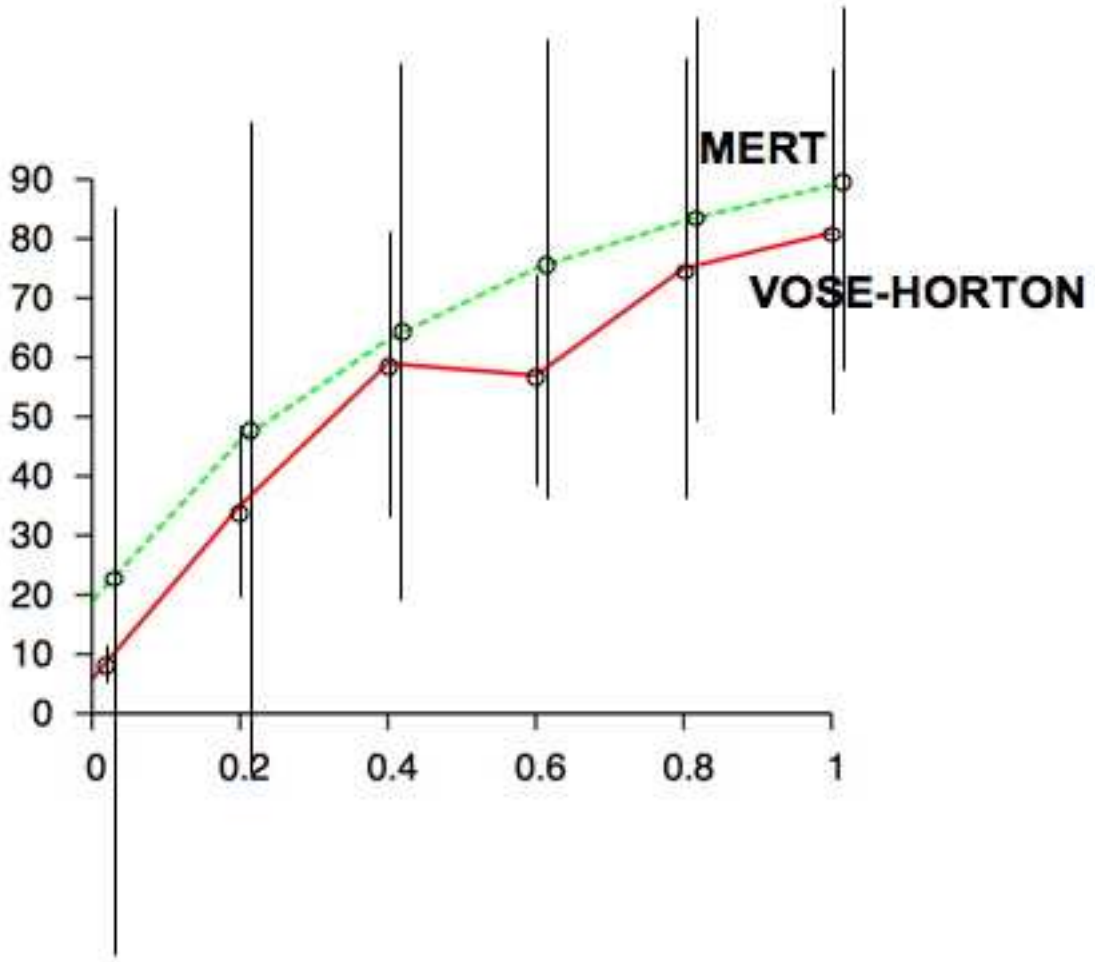

Figure 5.28: Average pixel error vs. system miscalibration with standard deviation.

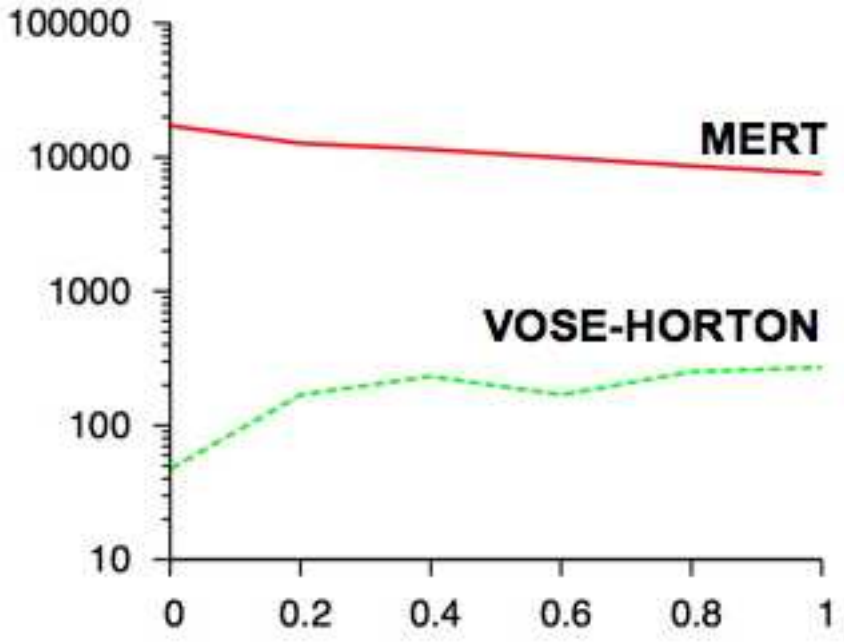

Figure 5.29: Max pixel error vs. system miscalibration.

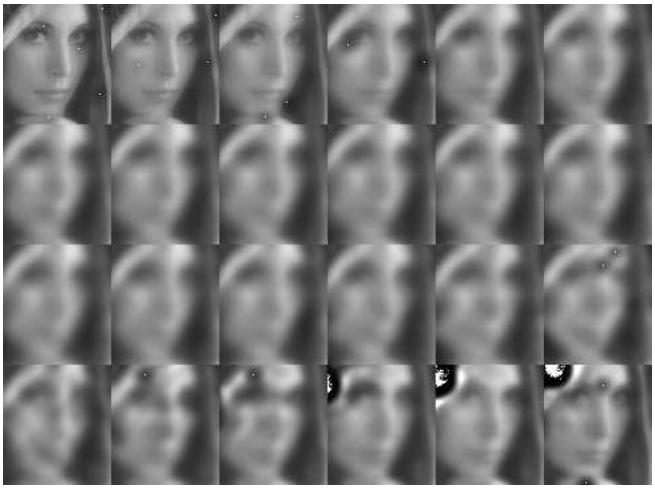

Figure 5.30: MERT, system miscalibration level $\sigma = 0$

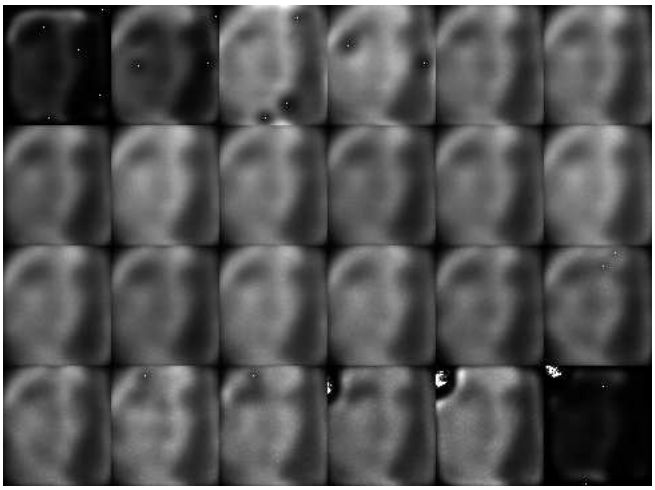

Figure 5.31: MERT, system miscalibration level $\sigma = .2$.

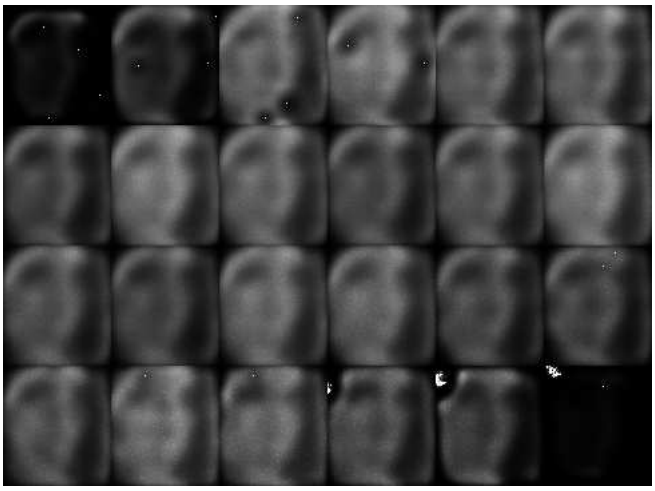

Figure 5.32: MERT, system miscalibration level $\sigma = .4$.

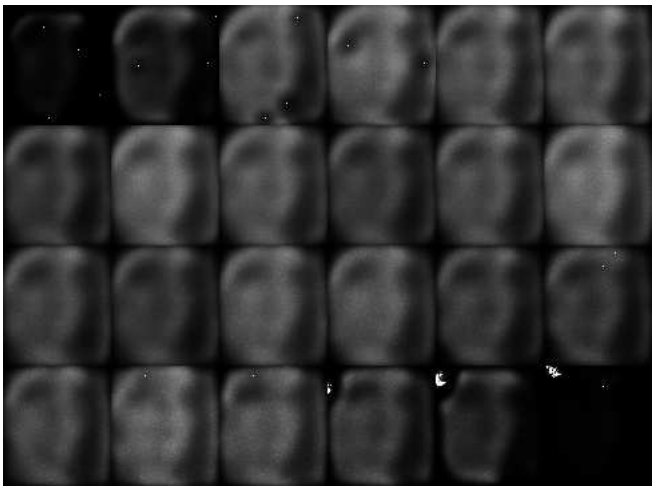

Figure 5.33: MERT, system miscalibration level $\sigma = .6$.

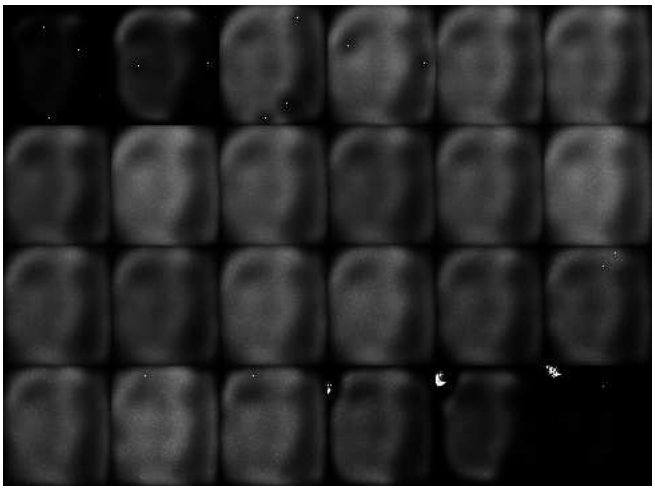

Figure 5.34: MERT, system miscalibration level $\sigma = .8$.

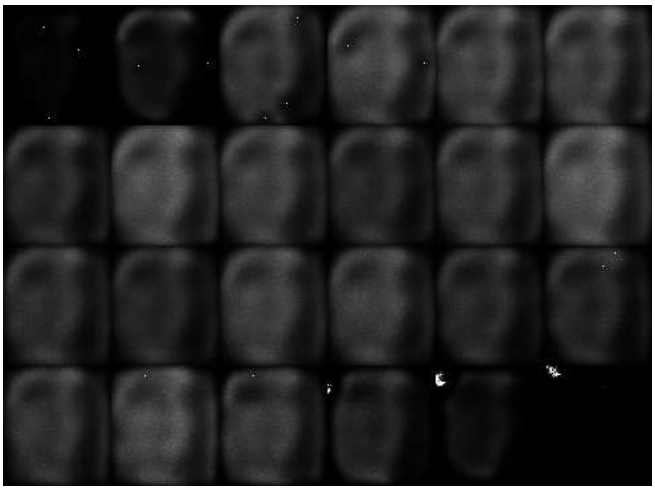

Figure 5.35: MERT, system miscalibration level $\sigma = 1$.

The robustness results for MERT do not support the suggestion made earlier that a miscalibrated system matrix favors reconstruction methods for which the reconstructed images exhibit less contrast.

5.4 Summary

Additional testing was done. The additional testing showed how our method compares against MART and MERT for an easier problem than the mosquito. Two method accuracy method metrics were discussed: pixel error and peak signal to noise ratio. Method robustness in the presence of noise in the system matrix was explored.

Chapter 6

Parallelization

Our method was parallelized in three ways: coarse grain parallelization, fftw library parallelization, and fine grain parallelization.

6.1 Coarse Grain Parallelization

At startup, a number of coarse grain parallelization threads are created and put to sleep. When a piece of code that has been coarse grain parallelized runs, it wakes up the other threads, does its piece of the work, then waits for the other threads to finish. Here is an example of a piece of code that has been parallelized in this fashion. The other threads are running code that is similar to the code below.

```
// wake other threads up
for (i = 0; i < (THREADS-1); i++){
  sem_post(S+i);
}

// do only my part
for (j = 0; j < w; j++){
  if (j%THREADS == 0){
    for (i = 0; i < n; i++) c_cjm(C[j][i],Tc3[0][i],Tc0[i]);
    ifft(n,reverse_cpy(n,Tc0,Tc2));
```

```
    for (i = 0; i < n; i++){
      OUT(j,i) = Tc2[i][0];
    }
  }
}

// wait for other threads to finish
for (i = 0; i < (THREADS-1); i++){
  sem_wait(F+i);
}
```

The number of threads is a `#define` and can be reset at compile time. Coarse grain parallelization occurs, in general, at the function level.

6.2 fftw Library Parallelization

The fftw library (used to perform fast Fourier transforms) can be parallelized by linking to the threaded fftw library and telling fftw at runtime how many threads are available for parallelization. The number of threads available to the fftw library is a `#define` and can be reset at compile time.

6.3 Fine Grain Parallelization

At startup, a number of fine grain parallelization threads are created and put to sleep. When a piece of code that has been fine grain parallelized runs, it wakes up the other threads, does its piece of the work, then waits for the other threads to finish. Here is an example of a piece of code that has been parallelized in this fashion. The other threads are running code that is similar to the code below.

```
// wake other threads up
for (i = 0; i < ((STHREADS)-1); i++){
  sem_post(SS+i);
}
```

```
// do only my part
for (i = 0; i < n; i++){
  if (i%STHREADS == 0){
     c_cjm(A[j][i],Tc0[i],Tc1[i]);
  }
}

// wait for other threads to finish
for (i = 0; i < ((STHREADS)-1); i++){
  sem_wait(SF+i);
}
```

The number of threads is a `#define` and can be reset at compile time. Fine grain parallelization occurs, in general, at the statement level, whereas coarse grain parallelization involves each thread performing a significantly larger chunk of work.

6.4 Experimental Results

Figure (6.1) shows wall clock time in seconds vs. number of threads for independent fftw and coarse grain parallelization; runs were done on a gentoo gnu/linux twelve-core amd opteron workstation. By independent fftw and coarse grain parallelization, it is meant that the fftw and coarse grain parallelization were not done simultaneously. Both fftw and coarse grain parallelization show a reduction in wall clock time independently. However, due to the overhead in thread context switching and memory bandwidth, execution time is not inversely proportional to the number of compute threads. Note that for both fftw parallelization and course grain parallelization, the curves are concave up which means that the improvement per available compute nodes is decreasing. Figure (6.2) shows time in seconds vs. number of threads for simultaneous fftw and coarse grain parallelization. The optimal number of threads for simultaneous fftw and coarse grain parallelization was found to occur at 11 fftw threads and 12 coarse grain parallelization threads for a time of 169 seconds.

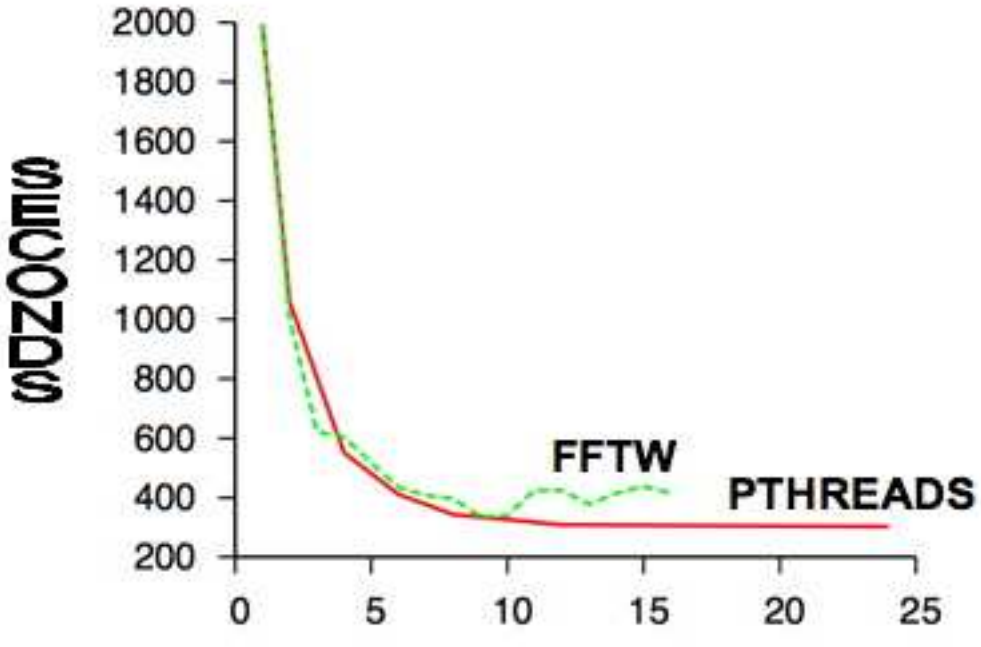

Figure 6.1: Time (seconds) vs. number of threads for fftw library(FFTW) and number of threads for coarse grain parallelization (PTHREADS); see Section(6.4)

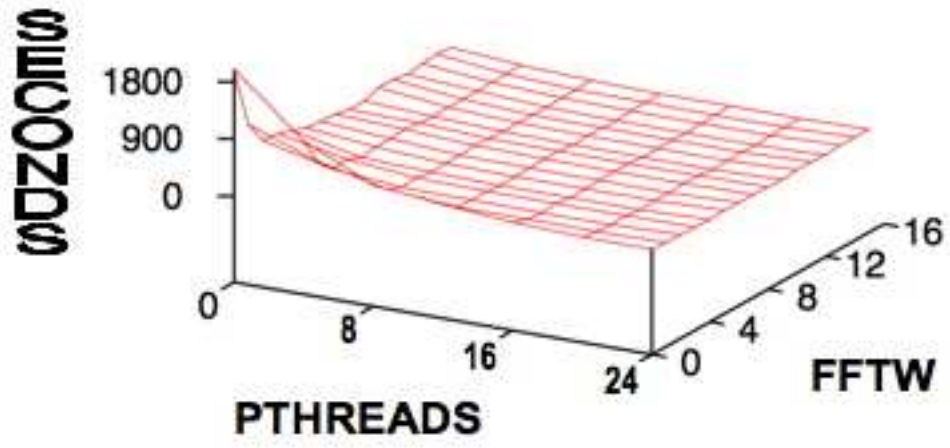

Figure 6.2: Time (seconds) vs. number of threads for fftw library(FFTW) and number of threads for coarse grain parallelization (PTHREADS); see Section(6.4)

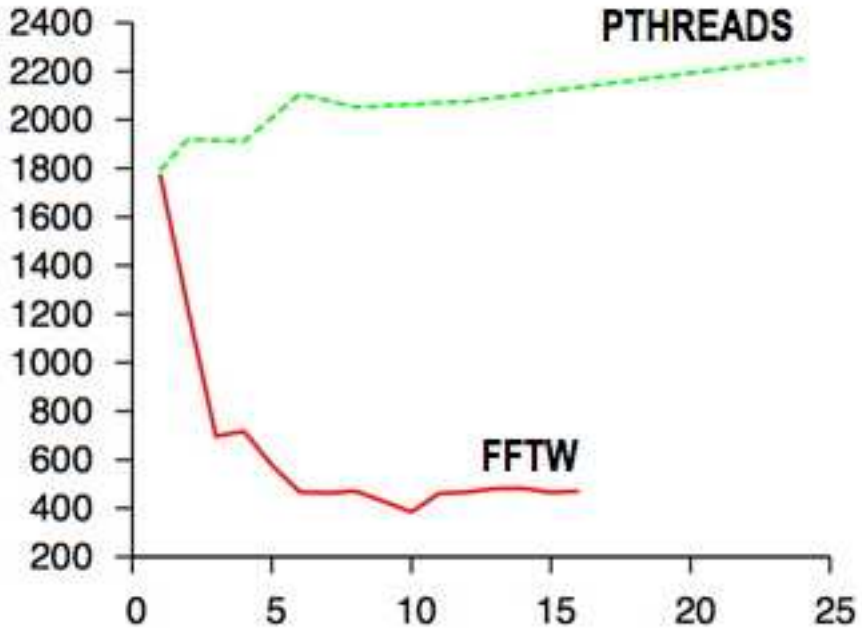

Figure 6.3: Time (seconds) vs. number of threads for fftw library(FFTW) and number of threads for fine grain parallelization (PTHREADS); see Section(6.4)

Figure (6.3) shows time in seconds vs. number of threads for independent fftw and fine grain parallelization. Fine grain parallelization shows an overall trend of an increase in time, most likely due to the fact that the parallelization is so fine grain that the benefit gained by running in parallel can be diminished by the overhead of thread context switching and thread synchronization. Figure (6.4) shows time in seconds vs. number of threads for simultaneous fftw and fine grain parallelization. The optimal number of threads for simultaneous fftw and fine grain parallelization was found to occur at 10 fftw threads and 2 fine grain parallelization threads for a time of 340 seconds.

6.5 Summary

Three types of parallelization was done. The first is course grain, in which a large blocks of code are executed in parallel. The second is fftw library parallelization in which the

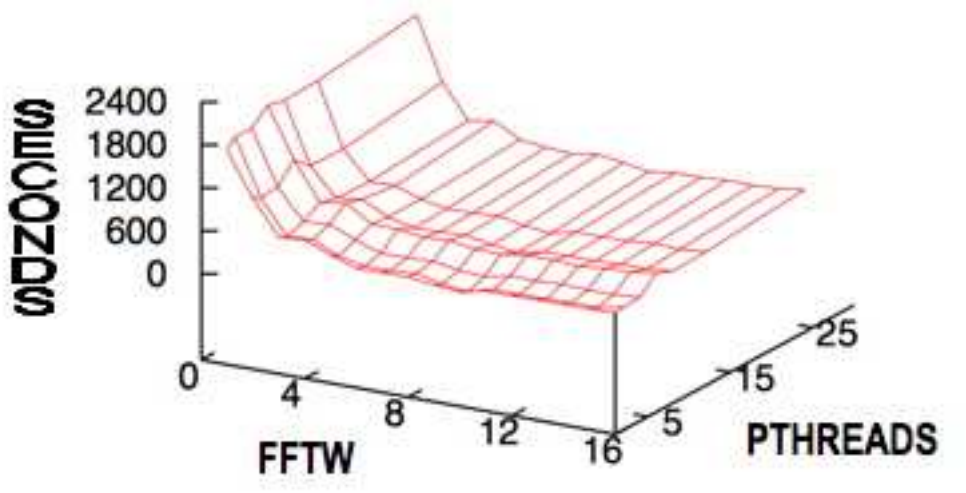

Figure 6.4: Time (seconds) vs. number of threads for fftw library(FFTW) and number of threads for fine grain parallelization (PTHREADS); see Section(6.4)

library we use to compute fast Fourier transforms is run in parallel mode. The third is fine grain, in which much smaller blocks of code are executed in parallel. It was observed that course grain parallelization and fftw library parallelization result in improvements to execution time but fine grain parallelization does not.

Chapter 7

Complexity Analysis

$$\vec{y_\infty} = Z(I + \mu P P^T Z' \{I + P(\mu^{-1} I - P^T Z' P)^{-1} P^T Z'\}) M h \tag{7.1}$$

How many operations do we use to compute $\vec{y_\infty}$? Define an operation as either an addition, a subtraction, a multiplication, a division, a logical evaluation, or an assignment. Recall that the first step in our heuristic method is to take the first approximation $\vec{f_0}$ to solving

$$x = Hf \tag{7.2}$$

as

$$\vec{f_0} = (I_w \otimes E^T)\vec{y_\infty} \tag{7.3}$$

Recall that subsequent steps proceed as follows. Let the vectors produced by previous steps be

$$\vec{f_0}, \ldots, \vec{f_i} \tag{7.4}$$

and let

$$\sum \eta_i \vec{f_i} \tag{7.5}$$

be the least-squares solution to minimizing

$$||x - H \sum \eta_i \vec{f_i}|| \tag{7.6}$$

Obtain $\vec{f_{i+1}}$ in the same manner as $\vec{f_0}$ except with respect to the problem

$$x - H \sum \eta_i \vec{f_i} = Hf \tag{7.7}$$

The first time $\vec{y_\infty}$ is computed,

$$h = \mathcal{H}^T x \tag{7.8}$$

where

$$\mathcal{H} = (C_0 | \ldots | C_{w-1}), \tag{7.9}$$

C_k is the $n \times n$ circulant matrix whose first column is the first column of the first rectangular circulant block associated with the k-th wavelength of the system matrix, and x is the noise removed measured image obtained on the focal plane array (it is read from memory). When $\vec{y_\infty}$ is computed in subsequent steps,

$$h = \mathcal{H}^T (x - H \sum \eta_i \vec{f_i}) \tag{7.10}$$

For the initial computation of $\vec{y_\infty}$, the first step is to compute Equation (7.8) For subsequent computations of $\vec{y_\infty}$, the first step is to compute Equation (7.10). The number of operations used to compute

$$x - H \sum \eta_i \vec{f_i} \tag{7.11}$$

is

$$\Theta(wn \log n) + \Theta(i^3) + \Theta(im) \tag{7.12}$$

where m is the number of columns of H. Justification for this number of operations is given in Section (7.3). In either case, $\mathcal{H}^T$ is a column of w $n \times n$ circulant matrices and is multiplied by an $n \times 1$ column vector on the right.

7.1 Agglomerate Operations

There are three composite operations that occur several times during the course of our algorithm. Each of these agglomerate operations will be discussed, named, and measured in turn. Later, a given agglomerate operation will be referred to by name and its associated

cost used where appropriate.

7.1.1 Circulant Multiply

Whenever a circulant matrix is multiplied by a column vector, the fact that the matrix is circulant is leveraged. A circulant matrix C can always be diagonalized:

$$C = F^* D F \tag{7.13}$$

where F^* is the conjugate transpose of F,

$$D = \sqrt{n}\text{diag}(Fc) \tag{7.14}$$

c is the first column of C, and F is the Fourier transform matrix

$$F_{i,j} \;=\; (1/\sqrt{n})e^{2\pi\sqrt{-1}\,ij/n} \tag{7.15}$$

Because C is real and F is symmetric, we obtain

$$C = F\overline{D}F^* \tag{7.16}$$

(by conjugation) where

$$\overline{D} = \sqrt{n}\text{diag}(F^*c) \tag{7.17}$$

So the multiplication of a circulant matrix by a vector can be computed by a 3 step process. Step 1 is to take the inverse Fourier transform (F^*) of the vector, using

$$\Theta(n \log n) \tag{7.18}$$

operations. Step 2 is to multiply the result of step 2 by $\sqrt{n}\text{diag}(F^*c)$ and uses $8n$ operations for an $n \times n$ circulant matrix ($\sqrt{n}\text{diag}(F^*c)$ is precomputed and read from memory. Its components and the components of the result of step 2 are complex. Our multiplication of a complex number by a complex number uses 4 multiplications, 1 addition, 1 subtraction,

and 2 assignments.) Step 3 is to take the Fourier transform of the result of step 2 and uses

$$\Theta(n \log n) \tag{7.19}$$

operations. So the number of operations to multiply an $n \times n$ circulant matrix by an $n \times 1$ column vector is

$$2\Theta(n \log n) + 8n = \Theta(n \log n) \tag{7.20}$$

7.1.2 Circulant Column Multiply

Whenever a column comprised of w circulant matrices is multiplied by an $n \times 1$ column vector, each $n \times n$ circulant matrix is multiplied by the $n \times 1$ column vector. The resulting w column vectors are assembled into a $wn \times 1$ column vector. Since the time for a circulant multiply is

$$\Theta(n \log n) \tag{7.21}$$

the number of operations to multiply a column of w circulant matrices by an column vector is

$$\Theta(wn \log n) \tag{7.22}$$

7.1.3 Circulant Row Multiply

Whenever a row comprised of w circulant matrices is multiplied by a $wn \times 1$ column vector, the column vector is partitioned into w equally sized pieces and the i-th circulant matrix is multiplied by the i-th partition. The resulting w column vectors are added together to produce a single $n \times 1$ column vector. Since the time for a circulant multiply is

$$\Theta(n \log n) \tag{7.23}$$

the number of operations to multiply a row of w circulant matrices by a column vector is

$$\Theta(wn \log n) \tag{7.24}$$

7.2 Operation Count

The first two steps in computing $\vec{y_\infty}$ are to compute h and then Mh. Because

$$Mh = \mu^{-1}h - PP^Th \tag{7.25}$$

we compute Mh as follows

$$u_1 = P^Th \tag{7.26}$$

$$u_2 = Pu_1 \tag{7.27}$$

$$u_3 = \mu^{-1}h - u_2 \tag{7.28}$$

The remaining steps in computing $\vec{y_\infty}$ are

$$u_4 = Z'u_3 \tag{7.29}$$

$$u_5 = P^Tu_4 \tag{7.30}$$

$$u_6 = (\mu^{-1}I - P^TZ'P)^{-1}u_5 \tag{7.31}$$

$$u_7 = Pu_6 \tag{7.32}$$

$$u_8 = Z'(u_3 + u_7) \tag{7.33}$$

$$u_9 = PP^Tu_8 \tag{7.34}$$

and the final step is to compute $Z(u_3 + \mu u_9)$.

Computing h is an instance of a circulant column multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.35}$$

operations. Because

$$P^T = (K_0^T, \ldots, K_{w-1}^T) \tag{7.36}$$

where the K_k are circulant matrices, computing P^Th is an instance of a circulant row

multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.37}$$

operations. Computing Pu_1 is an instance of a circulant column multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.38}$$

operations. Computing $\mu^{-1}h - u_2$ takes

$$\Theta(wn) \tag{7.39}$$

operations. So the number of operations to compute Mh is

$$\Theta(wn \log n) \tag{7.40}$$

Z' is a $wn \times wn$ diagonal matrix with either 0 or 1 on the diagonal. Therefore, the number of operations to compute $Z'u_3$ is

$$\Theta(wn) \tag{7.41}$$

Computing $P^T u_4$ is an instance of a circulant row multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.42}$$

operations. Computing

$$u_6 = (\mu^{-1}I - P^T Z' P)^{-1} u_5 \tag{7.43}$$

is done by applying a prematurely terminated conjugate gradient method to approximately solve

$$u_5 = (\mu^{-1}I - P^T Z' P) u_6 \tag{7.44}$$

The following iterative scheme to compute a solution to

$$Ax = b \tag{7.45}$$

(where the notation $(\cdot\,,\cdot)$ in the algorithm below denotes inner product) is what we mean

by a conjugate gradient method (In this case, A is the $n \times n$ matrix $(\mu^{-1}I - P^T Z' P)$, x and b are $n \times 1$ column vectors). The number of operations to compute each step is indicated.

```
x = initial              Θ(n)
p = r = g                Θ(n)
e = (r,r)                Θ(n)
while (e > epsilon){
  v = Ap                 Θ(wn log n)
  a = e/(p,v)            Θ(n)
  x = x + a*p            Θ(n)
  r = r - a*v            Θ(n)
  a = (r,r)              Θ(n)
  p = r + (a/e)*p        Θ(n)
  e = a                  Θ(n)
}
```

So the number of computations to compute $(\mu^{-1}I - P^T Z' P)^{-1} u_5$ by applying a conjugate gradient method to solve

$$u_5 = (\mu^{-1}I - P^T Z' P) u_6 \tag{7.46}$$

is

$$\Theta(pwn \log n) \tag{7.47}$$

where p is the number of iterations the conjugate gradient method runs. Computing Pu_6 is an instance of a circulant column multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.48}$$

operations. Computing $Z'(u_3 + u_7)$ takes

$$\Theta(wn) \tag{7.49}$$

operations. Computing $PP^T u_8$ is an instance of a circulant row multiply and an instance

of circulant column multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.50}$$

operations. The final step in computing $\vec{y_\infty}$ is to compute $Z(u_3 + \mu u_9)$ which takes

$$\Theta(wn) \tag{7.51}$$

operations.

So the first time $\vec{y_\infty}$ is computed, the number of operations is

$$\begin{aligned}
\Theta(wn \log n) &+ && (7.52)\\
\Theta(wn \log n) &+ && (7.53)\\
\Theta(wn) &+ && (7.54)\\
\Theta(wn \log n) &+ && (7.55)\\
\Theta(pwn \log n) &+ && (7.56)\\
\Theta(wn \log n) &+ && (7.57)\\
\Theta(wn) &+ && (7.58)\\
\Theta(wn \log n) &+ && (7.59)\\
\Theta(wn) &= \Theta(pwn \log n) && (7.60)
\end{aligned}$$

where p is the number of conjugate gradient iterations.

For subsequent computations of $\vec{y_\infty}$, the number of operations is identical to the above except for the addition of the number of operations to compute $x - H \sum \eta_i \vec{f_i}$

$$\Theta(wn \log n) + \Theta(i^3) + \Theta(im) \tag{7.61}$$

and is, therefore

$$\Theta(pwn \log n) + \Theta(i^3) + \Theta(im) \tag{7.62}$$

where p is the number of conjugate gradient iterations and i corresponds to the iteration number.

So the number of operations to compute $\vec{y_\infty}$ i times is

$$\Theta(p_1 wn \log n) + \sum_{j=2}^{i} \big(\Theta(p_j wn \log n) + \Theta(j^3) + \Theta(jm)\big) \tag{7.63}$$

$$= \Theta(p_1 wn \log n) + \sum_{j=2}^{i} \Theta(p_j wn \log n) + \sum_{j=2}^{i} \Theta(j^3) + \sum_{j=2}^{i} \Theta(jm) \tag{7.64}$$

$$= \Theta(p_1 wn \log n) + \Theta(wn \log n \sum_{j=2}^{i} p_j) + \Theta(i^4) + \Theta(i^2 m) \tag{7.65}$$

$$= \Theta(i^4 + i^2 m + wn \log n \sum_{j=1}^{i} p_j) \tag{7.66}$$

where p_j is the number of conjugate gradient steps for the j-th time $\vec{y_\infty}$ is computed.

7.3 Least-Squares Operation Count

How many operations do we use to compute $x - H \sum \eta_i f_i$ the jth time? Recall from Equation (3.187)

$$\sum \eta_i \vec{f_i} \tag{7.67}$$

is the least-squares solution to minimizing

$$||x - H \sum \eta_i \vec{f_i}|| \tag{7.68}$$

which has the minimization condition:

$$\vec{f_k}^T H^T x = \sum_{i=0}^{j-1} \vec{f_k}^T H^T H \vec{f_i} \eta_i \tag{7.69}$$

The first step is to store f_{j-1} (which is the most recently computed f) to memory; that takes

$$\Theta(m) \tag{7.70}$$

operations where m is the number of columns in the system matrix H. The next step is to compute $H^T H f_{j-1}$ and takes

$$\Theta(wn \log n) \tag{7.71}$$

operations. The next step is to compute $\vec{f_{j-1}}^T H^T x$ and requires

$$\Theta(m) \tag{7.72}$$

operations. (it amounts to taking a dot product because $H^T x$ is available from a previous computation). The next step is to compute $\vec{f_k}^T H^T H \vec{f_i}$ where k and i range from 0 to $j-1$ and requires

$$\Theta(jm) \tag{7.73}$$

operations. It amounts to computing a dot product j times, because the $H^T H f_i$ have already been computed (some from previous Vose-Horton steps). The next step is to solve for η. Recall that by setting

$$\vec{v_i} = H\vec{f_i} \tag{7.74}$$

we can rewrite the minimization condition as

$$V^T V \eta = V^T x \tag{7.75}$$

Hence

$$\eta = (V^T V)^{-1} V^T x \tag{7.76}$$

Solving for η using a handwritten iterative solver takes

$$\Theta(j^3) \tag{7.77}$$

operations (the solver inverts a $j \times j$ matrix.).

The next step is to compute $x - H \sum \eta_i f_i$ from the constituent pieces. First the sum is computed and takes

$$\Theta(jm) \tag{7.78}$$

operations. Next, the multiplication by H is performed. This multiplication is carried out in a higher dimension so it is actually a multiplication by $\mathcal{H}$. This is an instance of a circulant row multiply and, therefore, takes

$$\Theta(wn \log n) \tag{7.79}$$

operations. The final step is to perform the subtraction and takes

$$\Theta(n) \tag{7.80}$$

operations. So the number of operations to compute $x - H \sum \eta_i f_i$ from the constituent pieces is

$$\Theta(wn \log n) + \Theta(jm) \tag{7.81}$$

operations.

The total number of operations is therefore

$$\begin{aligned}
\Theta(m) &+ && (7.82)\\
\Theta(wn \log n) &+ && (7.83)\\
\Theta(m) &+ && (7.84)\\
\Theta(jm) &+ && (7.85)\\
\Theta(m) &+ && (7.86)\\
\Theta(j^3) &+ && (7.87)\\
\Theta(wn \log n) + \Theta(jm) &= \Theta(wn \log n) + \Theta(j^3) + \Theta(jm) && (7.88)
\end{aligned}$$

7.4 Summary

It was shown that the number of operations to compute $y_\infty^{\rightarrow}$ a total of i times is

$$\Theta(i^4 + i^2 m + wn \log n \sum_{j=1}^{i} p_j) \tag{7.89}$$

where p_j is the number of conjugate gradient steps for the j-th time $\vec{y_\infty}$ is computed, w is the number of wavelengths the system matrix is calibrated at, m is the number of columns of the system matrix, and n is the number of rows of the system matrix. The method would, therefore, seem to be the most sensitive to the number of rows of the system matrix since n is governed by the size of the focal plane array and is typically greater than four-million and i is typically less than five.

Chapter 8

Source Code

Open source proof of concept code implementing our algorithm is available from

http://www.cs.utk.edu/~horton/ctis/ctis.tar.

The code is documented with internal comments and there is an electronic document that describes its use.

Our method was implemented on a gentoo gnu/linux twelve core amd64 workstation with 32 GB/RAM. It relies on the fact that there is sufficient RAM available to hold various matrices in memory (at run time, the executable is using approximately 16 GB of RAM for the image reconstruction using the system matrix described in this dissertation). When using our code on another machine, if there is insufficient RAM available to hold the computation in memory, the code will run (assuming sufficient swap space), but much slower.

An estimate for the memory required is $(128 * 2 * \text{sizeof(double)} * n + 8 * m)$ where n is the number of rows in the system matrix H and m is the number of columns in H.

Before running the code that implements our method, the fftw library must be installed. Next, the system matrix must be written to disk in the representation our code expects. Code to write the system matrix to disk is provided and the format that code expects to see is documented. Next a parameter file must be created containing the sizes associated with the problem instance. Documentation is provided for this parameter file. The code contains various `#defines` that must be set by the user (documentation exists for these

settings). Finally the code must be compiled (documentation is provided for compilation).

Chapter 9

Future Work

9.1 Partial Eigen Decomposition

Algorithm enhancements can be investigated to improve upon the reduction in reconstruction time our method has already achieved. The most obvious possibility is to focus on the solution of (3.182). It is expected that there are more cost-effective alternatives to the prematurely terminated conjugate gradient method presently used. This expectation makes sense, because the operators involved depend only upon the system matrix. Any initial computational investment (such as precomputing an approximate eigen decomposition) has zero cost at image reconstruction time, and can reasonably be regarded as amortized over arbitrarily many image-reconstruction instances. This represents a potential for further speedup on a per device basis. In addition, there are probably better initial states for iterative solvers than we are currently using, and it certainly makes sense to explore the use of matrix preconditioning [1].

9.2 GPU parallelization

Our method is particularly well suited to efficient parallelization because (unlike MART and MERT) few iterations are required. It makes sense to look into mapping our technique to a Graphical Processing Unit (GPU) architecture. An affordable and massively parallel open-source implementation can be produced (a teraflop of GPU computing power can now be brought to a desktop machine for the price of a graphics card).

9.3 More Efficient Formulas

A byproduct of the work in Chapter 4 is the development of previously unknown and unappreciated relationships between basic mathematical objects relevant to the Vose-Horton method. In addition, skill was developed in manipulating, rearranging and transforming expressions. A consequence is the discovery of improved formulas for implementing the Vose-Horton method which are significantly faster to compute.

9.3.1 More Efficient $\vec{y_0}$

Recall from Equations (3.133) and (3.78) that

$$\vec{y_0} = ZMh = Zy \tag{9.1}$$

One thing to try might be to compute

$$\lim_{\mu \to 0} \vec{y_0} \tag{9.2}$$

and use it in place of $\vec{y_\infty}$ in Equation (3.185),

$$\vec{f_0} \;=\; (I_w \otimes E^T)\, \vec{y}_\infty \tag{9.3}$$

This would have the advantage of being much faster due to avoiding the embedded conjugate gradient method used to compute the inverse in Equation (3.181) which gives $\vec{y_\infty}$; that equation is

$$\vec{y_\infty} \;=\; Z(I + \mu P P^T Z' \{I + P(\mu^{-1} I - P^T Z' P)^{-1} P^T Z'\}) Mh \tag{9.4}$$

If it turns out that the associated error is acceptable, then this faster approach would be an option. Note that from Equations (3.84) and (3.67)

$$\begin{aligned}
\lim_{\mu\to 0}\vec{y_0} &= \lim_{\mu\to 0} Z(\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I+\mu^{-1}\sum C_kC_k^T)^{-1}\mathcal{H})\mathcal{H}^T x && (9.5)\\
&= \lim_{\mu\to 0} Z\frac{(I-\mu^{-1}\mathcal{H}^T(\mu^{-1}\{\mu I+\sum C_kC_k^T\})^{-1}\mathcal{H})\mathcal{H}^T x}{\mu} && (9.6)\\
&= \lim_{\mu\to 0} Z\frac{\mathcal{H}^T x-\mathcal{H}^T(\mu I+\sum C_kC_k^T)^{-1}(\mu I+\sum C_kC_k^T-\mu I)x}{\mu} && (9.7)\\
&= \lim_{\mu\to 0} Z\frac{(\mathcal{H}^T-\mathcal{H}^T\{I-\mu(\mu I+\sum C_kC_k^T)^{-1}\})x}{\mu} && (9.8)\\
&= \lim_{\mu\to 0} Z\frac{\mu\mathcal{H}^T(\mu I+\sum C_kC_k^T)^{-1}x}{\mu} && (9.9)\\
&= Z\mathcal{H}^T\lim_{\mu\to 0}(\mu I+\sum C_kC_k^T)^{-1}x && (9.10)\\
&= Z\mathcal{H}^T(\sum C_kC_k^T)^{-1}x && (9.11)
\end{aligned}$$

It might turn out that the inverse in (9.11) above does not exist. Because computations are carried out in the diagonalized basis, one way to address this problem is, for each diagonal entry that is within τ of zero, set the inverse of that entry to zero. Another way to address this problem is to use

$$Z\mathcal{H}^T(\mu I+\sum C_kC_k^T)^{-1}x \tag{9.12}$$

in place of (9.11). One might expect that, for the ideas above, the error could be examined in much the way it has been done in Chapter 4.

9.3.2 Simplified $\vec{y_\infty}$

Note that from Equations (9.5) through (9.10) it follows that

$$\vec{y_0} = Z\mathcal{H}^T(\mu I+\sum C_kC_k^T)^{-1}x \tag{9.13}$$

Equation (9.13) can be used to obtain a simpler expression for $\vec{y_\infty}$ as follows. Recall Equations (3.88), (4.42), (4.43) and Lemma (A.2)

$$M = (\mu I+\mathcal{H}^T\mathcal{H})^{-1} \tag{9.14}$$

$$\phi = M\mathcal{H}^T\mathcal{H} \tag{9.15}$$

$$\theta = Z'\phi \tag{9.16}$$

$$\mu M = I - \phi \tag{9.17}$$

From Equations (9.14) and (9.17) it follows that

$$(I-\phi)^{-1} = \mu^{-1}M^{-1} = \mu^{-1}(\mu I + \mathcal{H}^T\mathcal{H}) = I + \mu^{-1}\mathcal{H}^T\mathcal{H} \tag{9.18}$$

Recall Lemma (A.3) and Equation (3.67)

$$M^{-1}\vec{y_\infty} = h - \mu(I-\theta)^{-1}Z'Mh \tag{9.19}$$

$$h = \mathcal{H}^T x \tag{9.20}$$

From Equations (9.1), (3.84), (9.13) it follows that

$$\begin{aligned} y &= (\mu^{-1}I - \mu^{-2}\mathcal{H}^T(I + \mu^{-1}\sum C_k C_k^T)^{-1}\mathcal{H})h = Mh & (9.21)\\ &= \mathcal{H}^T(\mu I + \sum C_k C_k^T)^{-1}x & (9.22) \end{aligned}$$

From Equation (9.19) it follows that

$$\begin{aligned} \vec{y_\infty} &= Mh - \mu M(I-\theta)^{-1}Z'Mh & (9.23)\\ &= (I - \mu M(I-\theta)^{-1}Z')Mh & (9.24)\\ &= (I - \mu M(I-\theta)^{-1}Z')y & (9.25) \end{aligned}$$

Note that from Equation (9.17) and the definition of the inverse of the product of matrices it follows that

$$\mu M(I-\theta)^{-1} = (I-\phi)(I-\theta)^{-1} = \{(I-\theta)(I-\phi)^{-1}\}^{-1} \tag{9.26}$$

From Equations (9.18), (9.15), (9.16), (9.26) it follows that

$$\begin{aligned} \mu M(I-\theta)^{-1} &= \{(I - Z'(\mu I + \mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T\mathcal{H})(I + \mu^{-1}\mathcal{H}^T\mathcal{H})\}^{-1} && (9.27) \\ &= (I + \mu^{-1}\mathcal{H}^T\mathcal{H} - \mu^{-1}Z'\mathcal{H}^T\mathcal{H})^{-1} && (9.28) \\ &= (I + \mu^{-1}Z\mathcal{H}^T\mathcal{H})^{-1} && (9.29) \end{aligned}$$

From Equations (9.25), (9.29) it follows that

$$\vec{y_\infty} = y - (I + \mu^{-1}Z\mathcal{H}^T\mathcal{H})^{-1}Z'y \tag{9.30}$$

Comparing the above equation to Equation (9.4), it is clear that there are fewer operations. Not only that, the inverse occurring in (9.30) is easier to deal with.

9.3.3 Another Representation for $\vec{y_\infty}$

Another alternate expression for $\vec{y_\infty}$ can be obtained as follows. Define the object L_η by

$$L_\eta = (\eta I + \mathcal{H}\mathcal{H}^T)^{-1} \tag{9.31}$$

Note from Equation (9.22) that

$$y = \mathcal{H}^T L_\mu x \tag{9.32}$$

From Equation (9.31), the following four equations follow.

$$\mathcal{H}\mathcal{H}^T = L_\eta^{-1} - \eta I \tag{9.33}$$

$$\mathcal{H}^T\mathcal{H}\mathcal{H}^T = \mathcal{H}^T L_\eta^{-1} - \eta\mathcal{H}^T \tag{9.34}$$

$$\eta\mathcal{H}^T + \mathcal{H}^T\mathcal{H}\mathcal{H}^T = \mathcal{H}^T L_\eta^{-1} \tag{9.35}$$

$$(\eta I + \mathcal{H}^T\mathcal{H})\mathcal{H}^T = \mathcal{H}^T L_\eta^{-1} \tag{9.36}$$

Note that by choosing

$$G = \eta I \tag{9.37}$$

$$X = \mathcal{H}^T \tag{9.38}$$

$$J = I \tag{9.39}$$

and

$$O = \mathcal{H} \tag{9.40}$$

in the Sherman-Morrison-Woodbury Identity, Equation (3.83), yields the following equality:

$$(\eta I + \mathcal{H}^T\mathcal{H})^{-1} = \eta^{-1}(I - \mathcal{H}^T L_\eta \mathcal{H}) \tag{9.41}$$

From Equations (9.41) and (9.33), the following equation follows:

$$(I - \mathcal{H}^T L_\eta \mathcal{H})\mathcal{H}^T = \eta \mathcal{H}^T L_\eta \tag{9.42}$$

Recall Equation (9.30)

$$\vec{y_\infty} = y - (I + \mu^{-1} Z\mathcal{H}^T\mathcal{H})^{-1} Z'y \tag{9.43}$$

Note that by choosing

$$G = I \tag{9.44}$$

$$X = Z \tag{9.45}$$

$$J = \mu^{-1}(\eta I + \mathcal{H}^T\mathcal{H}) \tag{9.46}$$

and

$$O = I \tag{9.47}$$

in the Sherman-Morrison-Woodbury Identity, Equation (3.83), yields the following equality:

$$(I + \mu^{-1} Z(\eta I + \mathcal{H}^T\mathcal{H}))^{-1} = (I - Z(\mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1} + Z)^{-1} = (*) \tag{9.48}$$

It follows from Equations (9.43) and (9.48) that for small η

$$\vec{y_\infty} \approx y - (*)Z'y = Zy + Z(\mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1} + Z)^{-1} Z'y \tag{9.49}$$

Let

$$w = Z'y \tag{9.50}$$

and let

$$v = (\mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1} + Z)^{-1}w \tag{9.51}$$

Then

$$w = \mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1}v + Zv \tag{9.52}$$

Because $w = Z'y$, it follows that

$$w = Z'w \tag{9.53}$$

and

$$w = Z'w = \mu Z'(\eta I + \mathcal{H}^T\mathcal{H})^{-1}v \tag{9.54}$$

Moreover, Equation (9.52) yields

$$0 = Zw = \mu Z(\eta I + \mathcal{H}^T\mathcal{H})^{-1}v + Zv \tag{9.55}$$

Define the vector $t = y + Z\alpha$ and note that

$$Z't = Z'y = w \tag{9.56}$$

and

$$Zt = Z(y + \alpha) \tag{9.57}$$

Then by constraining t as follows

$$t = \mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1}v \tag{9.58}$$

it follows that

$$w = Z't = \mu Z'(\eta I + \mathcal{H}^T\mathcal{H})^{-1}v \tag{9.59}$$

Hence Equation (9.58) implies Equation (9.54). From Equations (9.58), (9.32), (9.36), and the definition $t = y + Z\alpha$, it follows that

$$\begin{aligned} v &= \mu^{-1}(\eta I + \mathcal{H}^T\mathcal{H})t = \mu^{-1}(\eta I + \mathcal{H}^T\mathcal{H})(\mathcal{H}^T L_\mu x + Z\alpha) && (9.60)\\ &= \mu^{-1}(\eta I + \mathcal{H}^T\mathcal{H})Z\alpha + \mu^{-1}\mathcal{H}^T L_\eta^{-1} L_\mu x && (9.61)\end{aligned}$$

From Equations (9.55), (9.61) it follows that

$$0 = Z\alpha + Z(\eta I + \mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T L_\eta^{-1} L_\mu x + \mu^{-1} Z(\eta I + \mathcal{H}^T\mathcal{H})Z\alpha + \mu^{-1} Z\mathcal{H}^T L_\eta^{-1} L_\mu x \quad (9.62)$$

Rearranging terms and multiplying both sides by μ yields

$$\begin{aligned} (\mu I + Z(\eta I + \mathcal{H}^T\mathcal{H}))Z\alpha &= -Z(\mu(\eta I + \mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T L_\eta^{-1} + \mathcal{H}^T L_\eta^{-1})L_\mu x && (9.63)\\ & && (9.64)\end{aligned}$$

From equations (9.42), (9.41), and (9.31), it follows that

$$\begin{aligned} (\mu I + Z(\eta I + \mathcal{H}^T\mathcal{H}))Z\alpha &= -Z(\frac{\mu}{\eta}(I - \mathcal{H}^T L_\eta \mathcal{H})\mathcal{H}^T L_\eta^{-1} + \mathcal{H}^T L_\eta^{-1})L_\mu x && (9.65)\\ &= -Z(\frac{\mu}{\eta}\eta\mathcal{H}^T L_\eta L_\eta^{-1} + \mathcal{H}^T L_\eta^{-1})L_\mu x && (9.66)\\ &= -Z(\mu\mathcal{H}^T + \mathcal{H}^T L_\eta^{-1})L_\mu x && (9.67)\\ &= -Z\mathcal{H}^T(\mu I + L_\eta^{-1})L_\mu x && (9.68)\\ &= -Z\mathcal{H}^T(\mu I + \eta I + \mathcal{H}\mathcal{H}^T)L_\mu x && (9.69)\\ &= -Z\mathcal{H}^T L_{\mu+\eta}^{-1} L_\mu x && (9.70)\end{aligned}$$

Choose

$$Z\alpha = Z\mathcal{H}^T\beta x \quad (9.71)$$

Then the left-hand side of Equation (9.65) is given by

$$\begin{aligned}(\mu I + Z(\eta I + \mathcal{H}^T\mathcal{H}))Z\alpha &= \mu Z\mathcal{H}^T\beta x + Z\eta\mathcal{H}^T\beta x + Z\mathcal{H}^T\mathcal{H}Z\mathcal{H}^T\beta x && (9.72)\\ &= ((\mu+\eta)Z\mathcal{H}^T + Z\mathcal{H}^T\mathcal{H}Z\mathcal{H}^T)\beta x && (9.73)\\ &= Z\mathcal{H}^T((\mu+\eta)I + \mathcal{H}Z\mathcal{H}^T)\beta x && (9.74)\end{aligned}$$

Hence a solution for α which satisfies Equation (9.70) is given by the following choice for β (Equations (9.31), (9.33) are used to simplify the right-hand side).

$$\begin{aligned}\beta &= -((\mu+\eta)I + \mathcal{H}Z\mathcal{H}^T)^{-1}L_{\mu+\eta}^{-1}L_\mu && (9.75)\\ &= -((\mu+\eta)I + \mathcal{H}Z\mathcal{H}^T)^{-1}I + \eta L_\mu && (9.76)\end{aligned}$$

Therefore from Equations (9.71), (9.76), it follows that:

$$Z\alpha = -Z\mathcal{H}^T((\mu+\eta)I + \mathcal{H}Z\mathcal{H}^T)^{-1}(I + \eta L_\mu)x \tag{9.77}$$

From Equations (9.33), (9.77), and (9.61), it follows that:

$$\begin{aligned}v &= -\mu^{-1}(\eta I + \mathcal{H}^T\mathcal{H})Z\mathcal{H}^T((\mu+\eta)I + \mathcal{H}Z\mathcal{H}^T)^{-1}(I + \eta L_\mu)x && (9.78)\\ &+ \mu^{-1}\mathcal{H}^T L_\eta^{-1}L_\mu x && (9.79)\end{aligned}$$

Hence

$$\begin{aligned}\lim_{\eta\to 0} v &= -\mu^{-1}\mathcal{H}^T\mathcal{H}Z\mathcal{H}^T(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1}x + \mu^{-1}\mathcal{H}^T\mathcal{H}\mathcal{H}^T L_\mu x && (9.80)\\ &= \mu^{-1}\mathcal{H}^T(\mathcal{H}\mathcal{H}^T L_\mu - (\mu I + \mathcal{H}Z\mathcal{H}^T - \mu I)(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1})x && (9.81)\\ &= \mu^{-1}\mathcal{H}^T(\mathcal{H}\mathcal{H}^T L_\mu - I + \mu(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1})x && (9.82)\\ &= \mu^{-1}\mathcal{H}^T(\mu(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1} - \mu L_\mu)x && (9.83)\\ &= -\mathcal{H}^T(L_\mu - (\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1})x && (9.84)\\ &= -\mathcal{H}^T L_\mu x + \mathcal{H}^T(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1}x && (9.85)\\ &= -y + \mathcal{H}^T(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1}x && (9.86)\end{aligned}$$

It follows from Equations (9.49), (9.51), (9.50), and (9.86) that:

$$\vec{y_\infty} = Zy + Zv = Z\mathcal{H}^T(\mu I + \mathcal{H}Z\mathcal{H}^T)^{-1}x \tag{9.87}$$

It follows that:

$$\vec{y_\infty} = Z\mathcal{H}^T\{L_\mu^{-1/2}L_\mu^{1/2}(\mu I + \mathcal{H}Z\mathcal{H}^T)L_\mu^{1/2}L_\mu^{-1/2}\}^{-1}x \tag{9.88}$$

Note that

$$L_\mu^{1/2}(\mu I + \mathcal{H}Z\mathcal{H}^T)L_\mu^{1/2} = \mu L_\mu + (L_\mu^{1/2}\mathcal{H})Z(L_\mu^{1/2}\mathcal{H})^T \tag{9.89}$$

Using Equation (9.33) and the fact that $\mathcal{H}\mathcal{H}^T L_\mu^{1/2} = L_\mu^{1/2}\mathcal{H}\mathcal{H}^T$ (the matrices involved are block-diagonal in the Fourier basis)

$$\begin{aligned} \mu L_\mu &= (\mu I - L_\mu^{-1})L_\mu + I & (9.90)\\ &= I - \mathcal{H}\mathcal{H}^T L_\mu & (9.91)\\ &= I - (L_\mu^{1/2}\mathcal{H})(L_\mu^{1/2}\mathcal{H})^T & (9.92) \end{aligned}$$

Since

$$\begin{aligned} (L_\mu^{1/2}\mathcal{H})Z(L_\mu^{1/2}\mathcal{H})^T &= (L_\mu^{1/2}\mathcal{H})(I - Z')(L_\mu^{1/2}\mathcal{H})^T & (9.93)\\ &= (L_\mu^{1/2}\mathcal{H})(L_\mu^{1/2}\mathcal{H})^T - (L_\mu^{1/2}\mathcal{H})Z'(L_\mu^{1/2}\mathcal{H})^T & (9.94) \end{aligned}$$

it follows from Equations (9.88), (9.89), (9.92), and (9.94) that

$$\vec{y_\infty} = Z\mathcal{H}^T L_\mu^{1/2}(I - (L_\mu^{1/2}\mathcal{H})Z'(L_\mu^{1/2}\mathcal{H})^T)^{-1}L_\mu^{1/2}x \tag{9.95}$$

9.3.4 Leveraging the kernel of $\mathcal{H}$

The following is another thing that might be done. Let

$$u \in \text{ Ker } (\mathcal{H}) \tag{9.96}$$

so that

$$\mathcal{H}u = 0 \tag{9.97}$$

Given any y

$$\mathcal{H}y = \mathcal{H}(y \pm u) = \mathcal{H}(y') \tag{9.98}$$

where

$$y' = y \pm u \tag{9.99}$$

Let $\{b_0, b_1, \ldots\}$ be a basis for $\mathrm{Ker}(\mathcal{H})$ and find

$$\eta = \begin{pmatrix} \eta_0 \\ \vdots \end{pmatrix} \tag{9.100}$$

which minimizes

$$||Z'y - Z' \sum_{i=0}^{a} b_i \eta_i||^2 \tag{9.101}$$

where a is small. Then

$$Z'y \approx Z'b_0\eta_0 + Z'b_1\eta_1 \ldots \tag{9.102}$$

Let

$$q = \sum b_i \eta_i \tag{9.103}$$

Then

$$Z'(y - q) \approx 0 \tag{9.104}$$

Let

$$y' = y - q \tag{9.105}$$

and use Zy' in place of $\vec{y_\infty}$.

9.4 Summary

Future directions were outlined: replacing the embedded conjugate gradient solver with an SVD based solver, using a better starting guess for the conjugate gradient solver, matrix preconditioning, and mapping the algorithm to GPU hardware. A consequence of the work

in Chapter 4 was the discovery of improved formulas for implementing the Vose-Horton method which have the potential to significantly reduce computation time. Moreover, variations of the Vose-Horton method were identified, such as computing with $\vec{y_0}$ as opposed to $\vec{y_\infty}$.

Chapter 10

Conclusion

The computed tomography imaging spectrometer (CTIS) is a non-scanning flash hyperspectral imaging (FHI) snapshot instrument which simultaneously captures the spectral information for each pixel in its field of view in the shortest possible time (typically several milliseconds) [3, 5, 6, 8, 9].

CTIS instruments have been used for optical testing, biomedical imaging, and airborne remote sensing [2, 29, 39]. They have been used for mine detection [45] and geology measurements [8, 32]. CTIS technology has been evaluated for astronomical multispectral imaging [9]. The desire has been expressed to use CTIS instruments to detect eruptions and impact explosions on planetary bodies [46]. CTIS technology has been coupled with fluorescence microscopy for the analysis of physiological reaction within living tissue [5, 16, 17, 19]. CTIS instruments have been tested for use in the in vivo functional mapping of the human retina [8], and they have the potential for use in time-critical biomedical applications in which patient comfort is a concern [5]. CTIS technology has been used extensively for space surveillance [9, 44], military [44, 45], and scene and target characterization [39] applications. In missile defense, the kinetic kill of a Hera drone by a PAC-III interceptor was detected by a CTIS instrument [41]. High-speed spectral imaging by CTIS of a missile in flight has been demonstrated [7]. CTIS technology has been examined for missile detection in the presence of laser countermeasures [42].

An iterative method was presented for CTIS image-reconstruction in the presence of both photon noise in the image and post-detection Gaussian system noise. The new al-

gorithm, which assumes the transfer matrix of the system has a particular structure, was evaluated through computer simulations which demonstrate that, for larger problems, it is significantly better than both MART and MERT with respect to accuracy and computation time.

Our method was parallelized in three ways: coarse grain parallelization, fftw library parallelization, and fine grain parallelization.

A theoretical characterization of the experimentally verified Vose-Horton method was presented and a provable theoretical explanation for its demonstrated success was produced. Closed-form error bounds were obtained which make possible advance predictions for image reconstruction performance.

An operation count was provided. The operation count was given as a function of the problem size, the number of Vose-Horton steps incurred, and the number of embedded conjugate gradient steps for each Vose-Horton step.

Documented open-source code implementing our method has been made available. The memory requirements for running the implementation were clearly stated.

Future directions were outlined: replacing the embedded conjugate gradient solver with an SVD based solver, using a better starting guess, matrix preconditioning, and mapping the algorithm to GPU hardware. A consequence of the work in Chapter 4 was the discovery of improved formulas for implementing the Vose-Horton method which are significantly faster to compute: computing with $\vec{y_0}$ as opposed to $\vec{y_\infty}$, and more efficient computation of y, $\vec{y_0}$ and $\vec{y_\infty}$.

A significant outcome of this dissertation was the discovery of a rich collection of ideas and directions for future research which will undoubtedly lead to significant improvements to the Vose-Horton method.

Bibliography

[1] K. Chen, *Matrix Preconditioning Techniques and Applications*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2005.

[2] M. R. Descour and E. L. Dereniak, *Computed-tomography imaging spectrometer: experimental calibration and reconstruction results*, Applied Optics **34** (1995), 4817-4826.

[3] M. R. Descour, *Non-scanning imaging spectrometry*, Ph.D. Thesis, University of Arizona, Tucson, Arizona, 1994.

[4] G. Vane and A. F. H. Goetz, *Terestrial imaging spectroscopy*, Remote Sensing Environ. **24** (1988), 1-29.

[5] C. E. Volin, B. K. Ford, M. R. Descour, J. P. Garcia, D. W. Wilson, P. D. Maker, and G. H. Bearman, *High-speed spectral imager for imaging transient fluorescence phenomena*, Applied Optics **37** (1998), 8112-8119.

[6] M. R. Descour, C. E. Volin, T. M. Gleeson, E. L. Dereniak, M. F. Hopkins, D. W. Wilson, and P. D. Maker, *Demonstration of a computed-tomography imaging spectrometer using a computer-generated hologram disperser*, Applied Optics **36** (1997), 3694-3698.

[7] M. R. Descour, C. E. Volin, E. Dereniak, K. J. Thome, A. B. Schumacher, D. W. Wilson, and P. D. Maker, *Demonstration of a high-speed nonscanning imaging spectrometer*, Optical Society of America **22** (1997), no. 16, 1271-1273.

[8] W. R. Johnson, D. W. Wilson, W. Fink, M. Humayun, and G. Bearman, *Snapshot hyperspectral imaging in ophthalmology*, Journal of Biomedical Optics **12** (2007), no. 1. 014036.

[9] K. Hege, D. O'Connell, W. Johnson, S. Basty, and E. L. Dereniak, *Hyperspectral Imaging for Astronomy and Space Surviellance*, Imaging Spectrometry IX (S. S. Shen and P. E. Lewis, eds.), Proc. SPIE, 2003.

[10] M. D. Vose and M. D. Horton, *A heuristic technique for CTIS image-reconstruction*, Appl. Opt. **46** (2007), 6498-6503.

[11] C. E. Volin, *Portable snapshot imaging spectrometer*, Ph.D. Thesis, University of Arizona, Tucson, Arizona, 2000.

[12] *Glossary of Meteorology*, American Meteorological Society, 1989.

[13] N. H. Hagen, E. L.Dereniak, and D. T. Sass, *Fourier methods of improving reconstruction speed for CTIS imaging spectrometers*, Imaging Spectrometry XII (S. S. Shen and P. E. Lewis, eds.), Proc. SPIE, 2007.

[14] _____, *Maximizing the resolution of a CTIS instrument*, Imaging Spectrometry XI (S. S. Shen and P. E. Lewis, eds.), Proc. SPIE, 1995.

[15] W. R. Johnson, E. K. Hege, D. O'Connell, and E. L. Dereniak, *Novel calibration recovery technique for an EM tomographic reconstruction*, Optical Engineering **43** (2004), no. 01.

[16] B. K. Ford and M. R. Descour, *Large-image-format computed tomography imaging spectrometer for fluorescence microscopy*, Optics Express **9** (2001), no. 9, 444-453.

[17] B. K. Ford, C. E. Volin, S. M. Murphy, R. M. Lynch, and M. R. Descour, *Computed tomography-based spectral imaging for fluorescence microscopy*, Biophysical Journal **80** (2001), no. 2, 986-993.

[18] Y. S. Sabharwal, A. R. Rouse, L. Donaldson, M. F. Hopkins, and A. F. Gmitro, *Slit-scanning confocal microendoscope for high resolution in vivo imaging*, Applied Optics **38** (1999), 7133-7144.

[19] M. R. Descour, T. S. Tkaczyk, B. K. Ford, R. M. Lynch, A. Locke, and E. Dereniak, *The Computed Tomography Imaging Spectrometer*, Lasers and Electro-Optics Society, LEOS, the 16th Annual Meeting of the IEEE (2003), 460-461.

[20] D. Sabatke, A. Locke, E. L. Derenia, M. Descour, J. Garcia, T. Hamilton, and R. W. McMillan, *Snapshot imaging spectropolarimeter*, Opt. Eng. (Bellingham) **41** (2002), no. 5, 1048-1054.

[21] R. M. A. Azzam, *Ellipsometry*, 2nd ed. (M. Bass, ed.), Handbook of Optics, McGraw Hill, New York, 1995, pp. 27.1-27.

[22] R. A. Chipman, *Polarimetry*, 2nd ed. (M. Bass, ed.), Handbook of Optics, McGraw Hill, New York, 1995, pp. 22.1-37.

[23] B. Henderson, *Optical spectrometers*, 2nd ed. (M. Bass, ed.), Handbook of Optics, McGraw Hill, New York, 1995, pp. 20.1-32.

[24] ———, *Spectroscopic Measurements*, 2nd ed. (M. Bass, ed.), Handbook of Optics, McGraw Hill, New York, 1995, pp. 28.1-27.

[25] T. Vo-Dinh, *Biomedical Photonics Handbook*, CRC Press, Boca Raton, FL, 2002.

[26] R. Riesenberg and U. Dillner, *HADAMARD imaging spectrometer with micro slit matrix*, Imaging Spectrometry V (M. R. Descour and S. S. Shen, eds.), Proc. SPIE, 1999, pp. 203-213.

[27] H. R. Morris, Hoyt C. C, and P. J. Treado, *Imaging spectrometers for fluorescence and Raman microscopy: acousto-optic and liquid-crystal-tunable filters*, Appl. Spectrosc. **48** (1994), 857-866.

[28] R. Riesenberg and U. Dillner, *HADAMARD imaging spectrometer with micro slit matrix*, Imaging Spectrometry V (M. R. Descour and S. S. Shen, eds.), Proc. SPIE, 1999, pp. 203-213.

[29] J. E. Maurguia, T. D. Reeves, J. M. Mooney, W. S. Ewing, and F. D. Shepherd, *A compact visible/near infrared hyperspectral imager*, Infrared Detectors and Focal Plane Arrays VI (E. L. Dereniak and R. E. Sampson, eds.), Proc. SPIE, 2000, pp. 457-468.

[30] W. Liu, G. Barbastathis, and D. Psaltis, *Volume holographic hyperspectral imaging*, Appl. Opt. **43** (2004), 3581-3599.

[31] J. F. Scholl, E. K. Hege, M. L. Hart, D. O'Connell, W. R. Johnson, and E. L. Dereniak, *Evaluations of classification and spectral unmixing algorithms using ground based satellite imaging*, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII (S. S. Shen and P. E. Lewis, eds.), Proc. SPIE, 2006, pp. 807-819.

[32] W. R. Johnson, D. W. Wilson, and G. Bearman, *MEMS integration with an all-reflective SWIR snapshot hyperspectral imager*, Imaging Spectrometry XI (San Francisco, CA), Proc. SPIE, 2006.

[33] J. M. Beach, K. J. Schwenzer, and S. Srinivas, *Oximetry of retinal vessels by dual-wavelength imaging: calibration and influence of pigmentation*, J. Appl. Physiol. **86** (1999), no. 2, 748-758.

[34] 2010. `http://www.ece.utk.edu/~qi/ece472-572/lecture07_restoration_denoise.pdf`.

[35] J. F. Scholl, E. L. Dereniak, M. R. Descour, C. P. Tebow, and C. E. Volin, *Phase grating design for a dual-band snapshot imaging spectrometer*, Applied Optics **42** (2003), no. 1.

[36] A. F. H. Goetz, *Imaging spectrometry for earth remote sensing*, Science **228** (1985), 1147-1153.

[37] R. W. Basedow, D. C. Carmer, and M. E. Anderson, *HYDICE system, implementation, and performance*, Imaging Spectrometry (M. R. Descour, J. .M. Mooney, D. L. Perry, and L. R. B. Illing, eds.), Proc. SPIE, 1995, pp. 258-267.

[38] W. M. Porter and H. T. Enmark, *A system overview of the Airborne Visible/Infrared Imaging Spectrometer (A VIRIS)*, Imaging Spectroscopy II (G. Vane, ed.), Proc. SPIE, 1987, pp. 22-31.

[39] B. K. Ford, M. P. Wilson, and J. S. Salazar, *Reconstruction Algorithm Development and Assessment for a Computed Tomography Based-Spectral Imager*, Sandia National Laboratories, 2005. Sandia Report, SAND2005-6385.

[40] B. A. Kinder and E. L. Dereniak, *Spectral LADAR Receiver*, U. S. Army Research Office, 2004. 41485.4-PH.

[41] Jr. J. D. George, E. L. Dereniak, and J. P. Garcia, *Visible Spectrometer Measurements of the PAC III Intercept of Hera at WSMR on 31-Mar-2001: Final Report*, U. S. Army Research Office, 2001. FRS # 302810.

[42] H. C. Schau, R. E. Soulon, G. L. Hall, and D. J. Garrood, *Restoration, Target Restoration, and Countermeasure Removal using the Computed Tomographic Imaging Spectrometer (CTIS)*, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII (S. S. Shen and P. E. Lewis, eds.), Proc. SPIE, 2002.

[43] Y. Luo, N. Liao, X. Wang, M. Liang, J. Fen, and P. Yang, *Fast processing of imaging spectrometer data cube based on FPGA design*, Mulitispectral Image Processing (H. Maitre, H. Sun, J. Lui, and E. Song, eds.), Proc. SPIE, 2007.

[44] W. R. Johnson, D. W. Wilson, and G. Bearman, *Spatial-spectral modulating snapshot hyperspectral imager*, Applied Optics **45** (2006), no. 9.

[45] M. R. Descour, E. L. Derenia, and A. C. Dubey, *Mine detection using instantaneous spectral imaging*, Detection Technologies for Mines and Minelike Targets (A. C. Dubey, I. Cindrich, J. Ralston, and K. Rigano, eds.), Proc. SPIE, 1995, pp. 286-304.

[46] D. W. Wilson, P. D. Maker, and R. E. Muller, *Reconstructions of Computed-Tomography Imaging Spectrometer Image*, 1997. JPL TRS 1992+.

[47] B. K. Ford, *Computed tomography-based spectral imaging for fluorescence microscopy*, Biophysics Journal (2001).

[48] J. F. Scholl, E. K. Hege, D. O'Connell, and E. L. Dereniak, *Model based compression of the calibration matrix for hyperspectral imaging systems*, Mathematics of Data/Image Pattern Recognition, Compression, Coding, and Encryption X, with Applications (G. X. Ritter, M. S. Schmalz, J. Barrera, and J. T. Astola, eds.), Proc. SPIE, 2007.

[49] J. P. Garcia and E. L. Dereniak, *Mixed-expectation image-reconstruction technique*, Applied Optics **38** (1999), no. 17, 3745-3748.

[50] H. H. Barrett, *Image reconstruction and the solution of inverse problems in medical imaging* (A. Todd-Pokropek and M. A. Viergever, eds.), The Formation, Handling, and Evaluation of Medical Images, M. A. Viergever, Berlin, 1991, pp. 33-39.

[51] T. R. Miller and J. W. Wallis, *Clinically Important Characteristics of Maximum-Likelihood Reconstruction*, Journal of Nuclear Medicine **33** (1995), 1678-1684.

[52] C. E. Volin, J. P. Garcia, E. L. Dereniak, M. R. Descour, D. T. Sass, and C. G. Simi, *MWIR Computed Tomography Imaging Spectrometer: Calibration and Imaging Experiments*, Proc. SPIE, 1999, pp. 192-202.

[53] R. A. Horn, *Topics in Matrix Analysis*, Cambridge University Press, 1994.

[54] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2008.

[55] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Sherman-Morrison and Woodbury Formulas*, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 1986, pp. 66-68.

[56] P. J. Davis, *Circulant Matrices*, Chelsea Publishing Company, 1994.

[57] S. Lang, *Linear Algebra*, Springer-Verlag, 1987.

[58] H. B. Maidinsah, *Properites of Circulant Matrices*, Master's Thesis, University of Tennessee, 1987.

Appendix

Lemma A.1. *Recall that*

$$\vec{y_0} = ZMh \tag{A.1}$$

and

$$\vec{y_{i+1}} = \vec{y_i} + ZM\mathcal{H}^T\mathcal{H}\vec{v_i} \tag{A.2}$$

For all i, $\vec{y_i}$ satisfies the following constraint:

$$y \;=\; (I_w \otimes E)(I_w \otimes E^T)y \tag{A.3}$$

Proof. For all i there exists a vector w such that

$$\vec{y_i} = Zw$$

Base case ($n = 0$)

$$\vec{y_0} = ZMh = Z(Mh) = Zw \tag{A.4}$$

Inductive hypothesis

$$\vec{y_i} = Zw$$

$$\begin{aligned} \vec{y_{i+1}} &= \vec{y_i} + ZM\mathcal{H}^T\mathcal{H}\vec{v_i} && \text{(A.5)} \\ &= Zw + ZM\mathcal{H}^T\mathcal{H}\vec{v_i} && \text{(A.6)} \\ &= Z(w + M\mathcal{H}^T\mathcal{H}\vec{v_i}) && \text{(A.7)} \end{aligned}$$

From Lemma (A.11),

$$ZZ = Z \tag{A.8}$$

Next it will be shown that

$$Z = (I_w \otimes E)(I_w \otimes E^T) \tag{A.9}$$

Note that Z and $(I_w \otimes E)(I_w \otimes E^T)$ are both $nw \times nw$ matrices. Recall from Equation

(3.48) that Q is the $\gamma \times a$ matrix

$$Q = \left(\frac{I_a}{0} \right) \tag{A.10}$$

Recall from Equation (3.50) that E is the $n \times a\alpha$ matrix

$$E = \left(\frac{I_\alpha \otimes Q}{0} \right) \tag{A.11}$$

Recall from Equation (3.132) that Z is the $nw \times nw$ matrix

$$Z \;=\; \mathrm{diag}((I_w \otimes E)\mathbf{1}_m) \tag{A.12}$$

and where $\mathbf{1}_m$ is the $m \times 1$ vector all of whose entries are 1. Then

$$Z_{i,j} = [i = j][(i \bmod n) < \alpha\gamma][((i \bmod n) \bmod \gamma) < a]] \tag{A.13}$$

Now

$$\begin{aligned}
& ((I_w \otimes E)(I_w \otimes E^T))_{i,j} && \text{(A.14)}\\
= \;& \sum_k (I_w \otimes E)_{ik}(I_w \otimes E^T)_{k,j} && \text{(A.15)}\\
= \;& \sum_k [(i \bmod n) < \alpha\gamma][((i \bmod n) \bmod \gamma) < a][k = d(i)] && \text{(A.16)}\\
& [(j \bmod n) < \alpha\gamma][((j \bmod n) \bmod \gamma) < a][k = d(j)] && \text{(A.17)}\\
= \;& [i = j][(i \bmod n) < \alpha\gamma][((i \bmod n) \bmod \gamma) < a] && \text{(A.18)}\\
= \;& Z_{i,j} && \text{(A.19)}
\end{aligned}$$

The function d encodes the fact that whether or not there is a 1 in a particular column of E is a function of the row. There can be at most a single column of E with a value of 1 in a given row. Hence d is injective. □

Lemma A.2. *Let*

$$\phi = M\mathcal{H}^T\mathcal{H} \tag{A.20}$$

and

$$M = (\mu I + \mathcal{H}^T\mathcal{H})^{-1} \tag{A.21}$$

Then

$$\mu M = I - \phi \tag{A.22}$$

Proof.

$$\begin{aligned} I &= (\mu I + \mathcal{H}^T\mathcal{H})^{-1}(\mu I + \mathcal{H}^T\mathcal{H}) && \text{(A.23)}\\ &= M(\mu I + \mathcal{H}^T\mathcal{H}) && \text{(A.24)}\\ &= \mu M + M\mathcal{H}^T\mathcal{H} && \text{(A.25)} \end{aligned}$$

Hence

$$\begin{aligned} \mu M &= I - M\mathcal{H}^T\mathcal{H} && \text{(A.26)}\\ &= I - \phi && \text{(A.27)} \end{aligned}$$

as was to be shown. □

Lemma A.3.

$$M^{-1}\vec{y_\infty} = h - \mu(I - \theta)^{-1}Z'Mh \tag{A.28}$$

Proof. Recall that

$$\vec{y_0} = ZMh \tag{A.29}$$

$$\vec{y_{i+1}} = \vec{y_i} + ZM\mathcal{H}^T\mathcal{H}\vec{v_i} \tag{A.30}$$

$$\vec{v_0} = Z'Mh \tag{A.31}$$

$$\vec{v_{i+1}} = Z'M\mathcal{H}^T\mathcal{H}\vec{v_i} \tag{A.32}$$

and that

$$Z' = I - Z \tag{A.33}$$

Also recall that

$$M = (\mu I + \mathcal{H}^T\mathcal{H})^{-1} \tag{A.34}$$

Note that

$$\begin{aligned} \vec{y_{i+1}} + \vec{v_{i+1}} &= \vec{y_i} + (Z + Z')M\mathcal{H}^T\mathcal{H}\vec{v_i} && \text{(A.35)} \\ &= \vec{y_i} + M\mathcal{H}^T\mathcal{H}\vec{v_i} && \text{(A.36)} \\ &= \vec{y_i} + M\{(\mu I + \mathcal{H}^T\mathcal{H}) - \mu I\}\vec{v_i} && \text{(A.37)} \\ &= \vec{y_i} + \vec{v_i} - \mu M\vec{v_i} && \text{(A.38)} \end{aligned}$$

Let

$$\vec{x_i} = \vec{y_i} + \vec{v_i} \qquad \text{(A.39)}$$

Then

$$\vec{x_{i+1}} = \vec{x_i} - \mu M\vec{v_i} \qquad \text{(A.40)}$$

It will be shown that for all n

$$\vec{x_n} = \vec{x_0} - \sum_{i=0}^{n-1} \mu M\vec{v_i} \qquad \text{(A.41)}$$

Base case (n=1):

$$\begin{aligned} \vec{x_1} &= \vec{x_0} - \mu M\vec{v_0} \quad \text{(by Equation (A.40))} && \text{(A.42)} \\ &= \vec{x_0} - \sum_{i=0}^{n-1} \mu M\vec{v_i} && \text{(A.43)} \end{aligned}$$

Inductive step: The inductive hypothesis is

$$\vec{x_n} = \vec{x_0} - \sum_{i=0}^{n-1} \mu M\vec{v_i} \qquad \text{(A.44)}$$

By Equation (A.40) and the inductive hypothesis

$$\begin{aligned} \vec{x_{n+1}} &= \vec{x_0} - \sum_{i=0}^{n-1} \mu M\vec{v_i} - \mu M\vec{v_n} && \text{(A.45)} \\ &= \vec{x_0} - \sum_{i=0}^{n} \mu M\vec{v_i} && \text{(A.46)} \end{aligned}$$

as desired. Let

$$\vec{x_\infty} = \lim_{i\to\infty} \vec{x_i} \tag{A.47}$$

By Equation (A.41)

$$\begin{aligned}
\vec{x_\infty} &= \vec{x_0} - \sum_{i=0}^{\infty} \mu M \vec{v_i} && \text{(A.48)}\\
&= \vec{y_0} + \vec{v_0} - \sum_{i=0}^{\infty} \mu M \vec{v_i} && \text{(A.49)}\\
&= ZMh + Z'Mh - \sum_{i=0}^{\infty} \mu M \vec{v_i} && \text{(A.50)}\\
&= (Z + Z')Mh - \sum_{i=0}^{\infty} \mu M \vec{v_i} && \text{(A.51)}\\
&= Mh - \sum_{i=0}^{\infty} \mu M \vec{v_i} && \text{(A.52)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \vec{v_i} && \text{(A.53)}
\end{aligned}$$

Moreover, it was previously shown that

$$\vec{y_\infty} = \lim_{i\to\infty} \vec{y_i} \tag{A.54}$$

exists and

$$\lim_{i\to\infty} \vec{v_i} = 0 \tag{A.55}$$

Hence

$$\vec{x_\infty} = \lim_{i\to\infty} \vec{x_i} = \lim_{i\to\infty} \vec{y_i} + \vec{v_i} = \vec{y_\infty} \tag{A.56}$$

Next, recall that

$$\theta = Z'\phi \tag{A.57}$$

From Lemma (A.11)

$$Z'Z' = Z' \tag{A.58}$$

Now by Equation (A.56)

$$\begin{aligned}
\vec{y_\infty} = \vec{x_\infty} &= Mh - \mu M \sum_{i=0}^{\infty} \vec{v_i} && \text{(A.59)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \{\mu(Z'P)(Z'P)^T\}^i \vec{v_0} && \text{(A.60)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \{\mu(Z'P)(P^TZ')\}^i \vec{v_0} && \text{(A.61)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \{\mu(Z'P)(P^TZ')\}^i Z'Mh && \text{(A.62)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \{\mu(Z'P)(P^T)\}^i Z'Mh && \text{(A.63)}\\
&= Mh - \mu M \sum_{i=0}^{\infty} \{\mu(Z'P)(P^T)\}^i \vec{v_0} && \text{(A.64)}\\
&= Mh - \mu M (I - \mu Z'PP^T)^{-1} \vec{v_0} && \text{(A.65)}\\
&= Mh - \mu M (I - \theta)^{-1} \vec{v_0} && \text{(A.66)}
\end{aligned}$$

Now

$$\begin{aligned}
M^{-1}\vec{y_\infty} &= M^{-1}Mh + M^{-1}\mu M(I-\theta)^{-1}\vec{v_0} && \text{(A.67)}\\
&= h - \mu(I-\theta)^{-1}\vec{v_0} && \text{(A.68)}\\
&= h - \mu(I-\theta)^{-1}Z'Mh && \text{(A.69)}
\end{aligned}$$

□

Lemma A.4.

$$\mu(I-\theta)^{-1}\vec{v_0} = (I-\theta)^{-1}Z'(I-\theta)h \qquad \text{(A.70)}$$

Proof.

$$\begin{aligned}
\mu(I-\theta)^{-1}\vec{v_0} &= \mu(I-\theta)^{-1}Z'Mh && \text{(A.71)}\\
&= (I-\theta)^{-1}Z'\mu Mh && \text{(A.72)}\\
&=_1 (I-\theta)^{-1}Z'(I-\phi)h && \text{(A.73)}
\end{aligned}$$

$$
\begin{aligned}
&= (I-\theta)^{-1}Z'Z'(I-\phi)h && \text{(A.74)}\\
&= (I-\theta)^{-1}Z'(Z'-Z'\phi)h && \text{(A.75)}\\
&= (I-\theta)^{-1}Z'(Z'-Z'Z'\phi)h && \text{(A.76)}\\
&= (I-\theta)^{-1}Z'Z'(I-Z'\phi)h && \text{(A.77)}\\
&= (I-\theta)^{-1}Z'Z'(I-\theta)h && \text{(A.78)}\\
&= (I-\theta)^{-1}Z'(I-\theta)h && \text{(A.79)}
\end{aligned}
$$

□

Lemma A.5.

$$(I_w \otimes E)(f - \vec{f_\infty}) = (\mu M + (I-\phi)(I-\theta)^{-1}Z'(I-\mu M))(I_w \otimes E)f \qquad \text{(A.80)}$$

Proof. Recall Equation (3.88)

$$M = (\mu I + \mathcal{H}^T\mathcal{H})^{-1} \qquad \text{(A.81)}$$

It follows that

$$M^{-1} = \mu I + \mathcal{H}^T\mathcal{H} \qquad \text{(A.82)}$$

and

$$\mathcal{H}^T\mathcal{H} = M^{-1} - \mu I \qquad \text{(A.83)}$$

Recall Equation (4.42) and Equation (4.43)

$$\phi = M\mathcal{H}^T\mathcal{H} \qquad \text{(A.84)}$$

$$\theta = Z'\phi \qquad \text{(A.85)}$$

From Equations (4.103) and (3.77), it follows that

$$\vec{y_\infty} = (I_w \otimes E)\vec{f_\infty} \qquad \text{(A.86)}$$

[1] from Lemma (A.2)

and

$$M^{-1}(I_w \otimes E)\vec{f_\infty} = M^{-1}\vec{y_\infty} \tag{A.87}$$

Recall Equations (3.67), (3.65)

$$h = \mathcal{H}^T x = \mathcal{H}^T\mathcal{H}(I_w \otimes E)f \tag{A.88}$$

It follows that

$$M^{-1}(I_w \otimes E)f = (\mu I + \mathcal{H}^T\mathcal{H})(I_w \otimes E)f = \mu(I_w \otimes E)f + h \tag{A.89}$$

Subtracting Equation (A.87) from Equation (A.89) yields

$$M^{-1}(I_w \otimes E)(f - \vec{f_\infty}) = \mu(I_w \otimes E)f + h - M^{-1}\vec{y_\infty} \tag{A.90}$$

It follows that

$$(I_w \otimes E)(f - \vec{f_\infty}) = \mu M(I_w \otimes E)f + Mh - \vec{y_\infty} \tag{A.91}$$

Note that from Equations (A.59) through (A.66), and Lemma (A.2)

$$\vec{x_\infty} = \lim_{i\to\infty} \vec{x_i} = Mh + (\phi - I)(I - \theta)^{-1}\vec{v_0} \tag{A.92}$$

It follows that

$$-\vec{x_0} + \vec{x_\infty} = (\phi - I)(I - \theta)^{-1}Z'Mh \tag{A.93}$$

Note that

$$\begin{aligned}
Mh - \vec{y_\infty} &= (Z + Z')(Mh - \vec{y_\infty}) && \text{(A.94)}\\
&= Z'Mh + Z(Mh - \vec{y_\infty}) && \text{(A.95)}\\
&= Z'Mh + \vec{y_0} - \vec{y_\infty} && \text{(A.96)}\\
&= \vec{v_0} + \vec{y_0} - \vec{y_\infty} && \text{(A.97)}\\
&= \vec{x_0} - \vec{x_\infty} && \text{(A.98)}\\
&= (I - \phi)(I - \theta)^{-1}Z'Mh && \text{(A.99)}
\end{aligned}$$

$$= (I-\phi)(I-\theta)^{-1}Z'M\mathcal{H}^T\mathcal{H}(I_w \otimes E)f \quad \text{(A.100)}$$

$$= (I-\phi)(I-\theta)^{-1}Z'M(M^{-1}-\mu I)(I_w \otimes E)f \quad \text{(A.101)}$$

$$= (I-\phi)(I-\theta)^{-1}Z'(I-\mu M)(I_w \otimes E)f \quad \text{(A.102)}$$

Hence

$$(I_w \otimes E)(f - \vec{f_\infty}) = (\mu M + (I-\phi)(I-\theta)^{-1}Z'(I-\mu M))(I_w \otimes E)f \quad \text{(A.103)}$$

as desired. □

Lemma A.6. *For vector v*

$$||(I_w \otimes E)v|| \ = ||v|| \quad \text{(A.104)}$$

Proof. Recall from Equation (3.48) that Q is the $\gamma \times a$ matrix

$$Q = \left(\frac{I_a}{0}\right) \quad \text{(A.105)}$$

Recall from Equation (3.50) that E is the $n \times a\alpha$ matrix

$$E = \left(\frac{I_\alpha \otimes Q}{0}\right) \quad \text{(A.106)}$$

Then

$$(I_w \otimes E^T)(I_w \otimes E) = I_w \otimes (E^T E) \quad \text{(A.107)}$$

$$= I_w \otimes (I_\alpha \otimes Q^T)(I_\alpha \otimes Q) \quad \text{(A.108)}$$

$$= I_w \otimes I_\alpha \otimes Q^T Q \quad \text{(A.109)}$$

$$= I_w \otimes I_\alpha \otimes I_a \quad \text{(A.110)}$$

$$= I_{w\alpha a} \quad \text{(A.111)}$$

$$= I_m \quad \text{(A.112)}$$

Hence

$$\begin{aligned} ||(I_w \otimes E)v|| &= v^T(I_w \otimes E^T)(I_w \otimes E)v && \text{(A.113)} \\ &= v^T I v && \text{(A.114)} \\ &= v^T v && \text{(A.115)} \\ &= ||v|| && \text{(A.116)} \end{aligned}$$

□

Lemma A.7.

$$\mu M + (I - \phi)(I - \theta)^{-1} Z' \phi = (I - \phi)(I - \theta)^{-1} \tag{A.117}$$

Proof. Note that

$$(I - \theta)^{-1} = \sum_{k=0}^{\infty} \theta^k \tag{A.118}$$

It follows that

$$(I - \theta)^{-1}\theta = \theta + \theta^2 + \dots \tag{A.119}$$

and

$$I + (I - \theta)^{-1}\theta = \sum_{k=0}^{\infty} \theta^k = (I - \theta)^{-1} \tag{A.120}$$

It follows that

$$\begin{aligned} \mu M + (I - \phi)(I - \theta)^{-1} Z' \phi &= I - \phi + (I - \phi)(I - \theta)^{-1}\theta && \text{(A.121)} \\ &= (I - \phi)(I + (I - \theta)^{-1}\theta) && \text{(A.122)} \\ &= (I - \phi)(I - \theta)^{-1} && \text{(A.123)} \end{aligned}$$

as desired. □

Lemma A.8. *Let*

$$S = FWF^* \tag{A.124}$$

where

$$W = I + \mu^{-1} \sum C_k C_k^T \tag{A.125}$$

Then $F^* S^{-1/2} F$ is real.

Proof. Choose

$$\Lambda = S \tag{A.126}$$

in Theorem (A.30); since

$$W = F^* S F \tag{A.127}$$

is a real symmetric matrix

$$0 = \sum_v \omega(v(i-j))(\lambda_v - \lambda_{-v}) \tag{A.128}$$

where

$$\lambda_i = S_{ii} \tag{A.129}$$

Let

$$k = i - j \tag{A.130}$$

Then since

$$F_{kv} = F_{vk} = \frac{1}{\sqrt{n}} \omega(vk) \tag{A.131}$$

$$0 = \frac{1}{\sqrt{n}} \sum_v \omega(vk)(\lambda_v - \lambda_{-v}) \quad = \quad \sum_v F_{kv}(\lambda_v - \lambda_{-v}) \tag{A.132}$$

$$= \quad \left(F \begin{pmatrix} \lambda_0 - \lambda_{-0} \\ \lambda_1 - \lambda_{-1} \\ \cdots \\ \lambda_{n-1} - \lambda_{-(n-1)} \end{pmatrix} \right)_k \tag{A.133}$$

Hence

$$F \begin{pmatrix} \lambda_0 - \lambda_{-0} \\ \lambda_1 - \lambda_{-1} \\ \cdots \\ \lambda_{n-1} - \lambda_{-(n-1)} \end{pmatrix} = 0 \tag{A.134}$$

Since F is invertible, it follows that

$$\begin{pmatrix} \lambda_0 - \lambda_{-0} \\ \lambda_1 - \lambda_{-1} \\ \cdots \\ \lambda_{n-1} - \lambda_{-(n-1)} \end{pmatrix} = 0. \tag{A.135}$$

Hence

$$\lambda_k = \lambda_{-k} \tag{A.136}$$

and since the λ_k are positive,

$$\frac{1}{\lambda_k^{1/2}} = \frac{1}{\lambda_{-k}^{1/2}} \tag{A.137}$$

Therefore

$$0 = \sum_v \omega(v(i-j))(\frac{1}{\lambda_v^{1/2}} - \frac{1}{\lambda_{-v}^{1/2}}) \tag{A.138}$$

and hence $F^* S^{-1/2} F$ is real as desired. □

Standard Results

Lemma A.9. *Let the Poisson probability mass function be given by*

$$p(n) = \frac{\beta^n e^{-\beta}}{n!} \tag{A.139}$$

Then the expected value is given by β.

Proof.

$$\begin{aligned} E(n) &= \sum_{n=0}^{\infty} np(n) = \sum_{n=0}^{\infty} n\frac{\beta^n e^{-\beta}}{n!} && \text{(A.140)} \\ &= \sum_{n=1}^{\infty} \frac{\beta^n e^{-\beta}}{(n-1)!} && \text{(A.141)} \\ &= \sum_{m=0}^{\infty} \frac{\beta^{m+1} e^{-\beta}}{(m)!} && \text{(A.142)} \\ &= \beta \sum_{m=0}^{\infty} \frac{\beta^m e^{-\beta}}{(m)!} && \text{(A.143)} \end{aligned}$$

$$= \beta \tag{A.144}$$

□

Lemma A.10. *Let the Poisson probability mass function be given by*

$$p(n) = \frac{\beta^n e^{-\beta}}{n!} \tag{A.145}$$

Then the variance is given by β.

Proof.

$$\begin{aligned}
\sigma^2 &= \sum_{n=0}^{\infty}(n-\beta)^2 p(n) && \text{(A.146)}\\
&= \sum_{n=0}^{\infty}(n-\beta)^2 \frac{\beta^n e^{-\beta}}{n!} && \text{(A.147)}\\
&= e^{-\beta}\sum_{n=0}^{\infty}(n^2+\beta^2-2n\beta)\frac{\beta^n}{n!} && \text{(A.148)}\\
&= e^{-\beta}\left(\sum_{n=0}^{\infty}\frac{\beta n\beta^{n-1}}{(n-1)!} + \beta^2\sum_{n=0}^{\infty}\frac{\beta^n}{n!} - 2\beta^2\sum_{n=0}^{\infty}\frac{\beta^{n-1}}{(n-1)!}\right) && \text{(A.149)}\\
&= e^{-\beta}\left(\sum_{m=0}^{\infty}\frac{\beta(m+1)\beta^m}{m!} + \beta^2 e^{\beta} - 2\beta^2 e^{\beta}\right) && \text{(A.150)}\\
&= e^{-\beta}\left(\beta\sum_{m=0}^{\infty}\frac{m\beta^m}{m!} + \beta\sum_{m=0}^{\infty}\frac{\beta^m}{m!} + \beta^2 e^{\beta} - 2\beta^2 e^{\beta}\right) && \text{(A.151)}\\
&= e^{-\beta}(\beta^2 e^{\beta} + \beta e^{\beta} + \beta^2 e^{\beta} - 2\beta^2 e^{\beta}) = \beta && \text{(A.152)}
\end{aligned}$$

□

Lemma A.11. *For any binary diagonal matrix A,*

$$AA = A \tag{A.153}$$

Proof.

$$(AA)_{i,j} = \sum_u A_{iu}A_{uj} \tag{A.154}$$

$$= \sum_u [i = u] A_{iu} [u = j] A_{uj} \tag{A.155}$$

$$= \sum_u [i = u \wedge u = j] A_{iu} A_{uj} \tag{A.156}$$

$$= [i = j] A_{ii} A_{ii} \tag{A.157}$$

$$= [i = j] A_{ii} \tag{A.158}$$

$$= [i = j] A_{ij} \tag{A.159}$$

$$= A_{ij} \tag{A.160}$$

□

Lemma A.12.

$$F = F^T \tag{A.161}$$

where F is the Fourier transform matrix.

Proof.

$$F_{i,j} = \frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}\, ij/n} \tag{A.162}$$

$$= \frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}\, ji/n} \tag{A.163}$$

$$= F_{j,i} \tag{A.164}$$

$$= F^T_{i,j} \tag{A.165}$$

□

Lemma A.13.

$$F^* = \overline{F} \tag{A.166}$$

where F is the Fourier transform matrix.

Proof.

$$F^* = (\overline{F})^T \tag{A.167}$$

$$= \overline{F^T} \tag{A.168}$$

$$= \overline{F} \tag{A.169}$$

□

Lemma A.14.

$$F^*F = I \tag{A.170}$$

where F is the Fourier transform matrix.

Proof. Note that

$$\begin{aligned}
\overline{e^{b\sqrt{-1}}} &= \overline{\cos(b) + i\sin(b)} && \text{(A.171)}\\
&= \cos(b) - i\sin(b) && \text{(A.172)}\\
&= \cos(-b) + i\sin(-b) && \text{(A.173)}\\
&= e^{-b\sqrt{-1}} && \text{(A.174)}
\end{aligned}$$

Now

$$\begin{aligned}
(F^*F)_{i,j} &= \sum_{k=0}^{n-1}(F^*)_{i,k}F_{k,j} && \text{(A.175)}\\
&= \sum_{k=0}^{n-1}(\overline{F})_{k,i}F_{k,j} && \text{(A.176)}\\
&= \sum_{k=0}^{n-1}\overline{(\frac{1}{\sqrt{n}}e^{2\pi\sqrt{-1}\frac{ki}{n}})}(\frac{1}{\sqrt{n}}e^{2\pi\sqrt{-1}\frac{kj}{n}}) && \text{(A.177)}\\
&= \frac{1}{n}\sum_{k=0}^{n-1}e^{2\pi\sqrt{-1}\frac{k(j-i)}{n}} && \text{(A.178)}\\
&\quad \text{The expression above is 1 if } i=j. \text{ Otherwise, it is} && \text{(A.179)}\\
&= \left(\frac{1}{n}\right)\frac{(e^{2\pi\sqrt{-1}\frac{(j-i)}{n}})^n - 1}{(e^{2\pi\sqrt{-1}\frac{(j-i)}{n}}) - 1} && \text{(A.180)}\\
&= 0 && \text{(A.181)}
\end{aligned}$$

□

Lemma A.15.

$$FF = F^*F^* \quad \textit{is a permutation matrix} \tag{A.182}$$

where F is the Fourier transform matrix.

Proof.

$$
\begin{aligned}
(FF)_{i,j} &= \sum_{k=0}^{n-1} F_{i,k}F_{k,j} && \text{(A.183)}\\
&= \sum_{k=0}^{n-1} (\frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}\frac{ki}{n}})(\frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}\frac{kj}{n}}) && \text{(A.184)}\\
&= \frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi\sqrt{-1}\frac{k(i+j)}{n}} && \text{(A.185)}\\
&\quad \text{The expression above is 1 if } i+j=0 \text{ or } i+j=n. && \text{(A.186)}\\
&\quad \text{Otherwise, it is} && \text{(A.187)}\\
&= \left(\frac{1}{n}\right)\frac{(e^{2\pi\sqrt{-1}\frac{(i+j)}{n}})^n - 1}{(e^{2\pi\sqrt{-1}\frac{(i+j)}{n}}) - 1} && \text{(A.188)}\\
&= 0 && \text{(A.189)}
\end{aligned}
$$

Therefore

$$(FF)_{ij} = [i + j = 0 \bmod n] = [j = -i \bmod n]$$

Hence, for each row there is a unique column containing a 1.

$$F^*F^* = (FF)^* = (\overline{FF}) = FF \tag{A.190}$$

since FF is real. □

Lemma A.16. *A matrix C is circulant if and only if*

$$C = F^*DF \tag{A.191}$$

where

$$c = Ce_0 \tag{A.192}$$

e_0 is the first column of the identity matrix,

$$D = \sqrt{n}\ diag(Fc) \tag{A.193}$$

F^* is the conjugate transpose of F, and F is the Fourier transform matrix

$$F_{i,j} = \frac{1}{\sqrt{n}} e^{2\pi\sqrt{-1}ij/n} \tag{A.194}$$

Proof. Recall that any $n \times n$ circulant matrix C is characterized by the existence of a vector c such that

$$C_{i,j} = c_{i-j \bmod n} \tag{A.195}$$

and thus is determined by its first column, c. Let

$$\omega = e^{2\pi\sqrt{-1}/n} \tag{A.196}$$

Recall that

$$\sum_{j=0}^{n-1} \omega^{ij} = \begin{cases} n & \text{if } i \bmod n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.197}$$

Then the i, ℓ-th element of FCF^* is (interpreting subscripts mod n)

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \omega^{ij} c_{j-k} \omega^{-k\ell} &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{m=0}^{n-1} \omega^{ij} c_m \omega^{-(j-m)\ell} && \text{(A.198)} \\ &= \sum_{m=0}^{n-1} c_m \omega^{\ell m} \big(\frac{1}{n} \sum_{j=0}^{n-1} \omega^{(i-\ell)j} \big) && \text{(A.199)} \\ &= \begin{cases} \sum_{m=0}^{n-1} c_m \omega^{\ell m} & \text{if } i = \ell \\ 0 & \text{if } i \neq \ell \end{cases} && \text{(A.200)} \end{aligned}$$

which is the i, ℓ-th element of $\sqrt{n}\,\mathrm{diag}(Fc)$. Conversely, if $C = F^* \sqrt{n}\,\mathrm{diag}(Fc) F$ for some vector c, then the circulant equalities above show that

$$FCF^* = FAF^* \tag{A.201}$$

where

$$A_{jk} = c_{j-k} \tag{A.202}$$

interpreting subscripts mod n. Therefore

$$C_{ij} = A_{ij} = c_{i-j} \tag{A.203}$$

□

Lemma A.17. *Circulant matrices are closed under transpose, addition, multiplication, scaling (i.e., multiplication by scalars), powers, roots, and are trivially inverted (when nonsingular).*

Proof. Note that from Lemma (A.15)

$$FF = F^*F^* \tag{A.204}$$

is a permutation matrix. It follows from the proof of Lemma (A.16) that a matrix C is circulant if and only if

$$C = F^*DF \tag{A.205}$$

$$(F^*DF)^T = F^T D^T F^{*^T} = FDF^* = F^*(FFDF^*F^*)F = F^*D'F \ \ [58] \tag{A.206}$$

$$F^*DF + F^*D'F = F^*(D + D')F = F^*D''F \tag{A.207}$$

$$(F^*DF)(F^*D'F) = (F^*D)(FF^*)(D'F) = F^*(DD')F = F^*D''F \tag{A.208}$$

$$\alpha F^*DF = F^*\alpha DF = F^*D'F \tag{A.209}$$

$$(F^*DF)^n = \overbrace{(F^*DF)(F^*DF)\dots(F^*DF)}^{n \text{ times}} = F^*D^nF = F^*D'F \tag{A.210}$$

$$((F^*DF)^{1/n})^n = \overbrace{(F^*D^{1/n}F)(F^*D^{1/n}F)\dots(F^*D^{1/n}F)}^{n \text{ times}} = F^*DF \tag{A.211}$$

$$(F^*DF)^{-1} = F^{-1}D^{-1}F^{*^{-1}} = F^*D^{-1}F \tag{A.212}$$

□

Theorem A.18. *Any real symmetric matrix A is positive definite iff all its eigenvalues are positive.*

Proof. A real symmetric matrix is said to be positive definite if

$$\forall x \ . \ x \neq 0 \Rightarrow x^T A x > 0 \tag{A.213}$$

Let v be an eigenvector for the eigenvalue λ. Then

$$v \neq 0 \tag{A.214}$$

and

$$Av = \lambda v \tag{A.215}$$

Let A be positive definite. Then

$$v^T A v = \lambda v^T v = \lambda ||v||^2 > 0 \Rightarrow \lambda > 0 \tag{A.216}$$

Let

$$\text{spec}(A) \subset \mathbb{R}^+ \tag{A.217}$$

Because A is real symmetric with positive eigenvalues,

$$A = U^T D U \tag{A.218}$$

where D is a positive diagonal matrix; let its square root be R. Because

$$x \neq 0 \tag{A.219}$$

and U is invertible,

$$Ux \neq 0 \tag{A.220}$$

Because

$$Ux \neq 0 \tag{A.221}$$

and R is invertible,

$$RUx \neq 0 \tag{A.222}$$

Then

$$\begin{aligned} x^T Ax &= x^T U^T DUx && \text{(A.223)} \\ &= (Ux)^T D(Ux) && \text{(A.224)} \\ &= (Ux)^T R^T R(Ux) && \text{(A.225)} \\ &= (RUx)^T (RUx) && \text{(A.226)} \\ &= ||RUx||^2 && \text{(A.227)} \\ &> 0 && \text{(A.228)} \end{aligned}$$

□

Theorem A.19. *If B is a real non singular matrix, then $B^T B$ has positive eigenvalues.*

Proof. This is a proof by contradiction. Let v be an eigenvector for $B^T B$ with eigenvalue λ and assume

$$\lambda \leq 0 \tag{A.229}$$

Then

$$v \neq 0 \tag{A.230}$$

and

$$B^T Bv = \lambda v \tag{A.231}$$

Hence

$$v^T B^T Bv = (Bv)^T (Bv) = ||Bv||^2 = v^T \lambda v = \lambda ||v||^2 \tag{A.232}$$

Because

$$v \neq 0 \tag{A.233}$$

and B is non singular,

$$Bv \neq 0 \tag{A.234}$$

Hence the above provides the following contradiction:

$$\text{(posititve)} = \text{(non-positive)(positive)} \tag{A.235}$$

□

Theorem A.20. *Let B be a real symmetric matrix with real nonnegative eigenvalues such that*

$$\rho(B) < 1 \tag{A.236}$$

Then

$$\sup_{v \neq 0} \frac{||Bv||}{||v||} < 1 \tag{A.237}$$

Proof.

$$\begin{aligned}
\sup_{v \neq 0} \frac{||Bv||}{||v||} &= \sup_{v \neq 0} \frac{||B\frac{v}{||v||}||}{||\frac{v}{||v||}||} && \text{(A.238)} \\
&= \sup_{||w||=1} \frac{||Bw||}{||w||} && \text{(A.239)} \\
&= \sup_{||w||=1} ||Bw|| && \text{(A.240)} \\
&= \sup_{||w||=1} w^T B^T B w && \text{(A.241)} \\
&=_2 \sup_{||w||=1} w^T A w && \text{(A.242)} \\
&=_3 \rho(A) && \text{(A.243)} \\
&=_4 \rho(B^2) && \text{(A.244)} \\
&=_5 \rho(B)^2 && \text{(A.245)} \\
&< 1^2 && \text{(A.246)}
\end{aligned}$$

□

Lemma A.21. *Let A be a matrix. If A has nonnegative real eigenvalues, and $\mu > 0$ then*

$$\rho((\mu I + A)^{-1} A) < 1 \tag{A.247}$$

Proof. Let $\text{spec}(A)$ be the spectrum of A. By the spectral mapping theorem, the spectrum

[2] $A = B^T B$

[3] from Lemmas (A.22), (A.28)

[4] $A = B^2$

[5] from Lemma (A.26)

of $(\mu I + A)^{-1}A$ is $\Lambda = \{(\mu + \lambda)^{-1}\lambda \ : \ \lambda \in \text{spec}(A)\}$. Let

$$f(\lambda) = |\frac{\lambda}{\mu + \lambda}| \tag{A.248}$$

It follows that as long as A has nonnegative eigenvalues,

$$f(\lambda) = \frac{|\lambda|}{|\mu + \lambda|} = \frac{\lambda}{\mu + \lambda} \tag{A.249}$$

Moreover,

$$f'(\lambda) = \frac{\mu + \lambda - \lambda}{(\mu + \lambda)^2} = \frac{\mu}{(\mu + \lambda)^2} > 0 \tag{A.250}$$

Hence $f(\lambda)$ is an increasing function. It follows that,

$$\rho((\mu I + A)^{-1}A) = \max_{\lambda \in \Lambda} f(\lambda) = \frac{\rho(A)}{\mu + \rho(A)} \tag{A.251}$$

and is, therefore, bounded by 1. □

Lemma A.22. *Let A be a matrix such that*

$$A = B^T B \tag{A.252}$$

If A is real, then A has nonnegative real eigenvalues

Proof. Consider the bilinear form:

$$\langle u, v \rangle = u^T B^T B v \tag{A.253}$$

Note that

$$\langle u, u \rangle = u^T B^T B u = (Bu)^T (Bu) \tag{A.254}$$

Call Bu by the name ν, then

$$\langle u, u \rangle = \sum_i \nu_i^2 \geq 0. \tag{A.255}$$

Because B^TB is real symmetric, it can be diagonalized so that

$$B^TB = UDU^T \tag{A.256}$$

and

$$u^T(B^TB)u = u^TUDU^Tu \tag{A.257}$$

where U is orthonormal and D is diagonal. Consider columns of U

$$U = \begin{pmatrix} | & & | \\ U_1 & \dots & U_n \\ | & & | \end{pmatrix} \tag{A.258}$$

Then

$$\langle U_i, U_i\rangle = U_i^TUDU^TU_i = U_i^T \begin{pmatrix} | & & | \\ U_1 & \dots & U_n \\ | & & | \end{pmatrix} \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \begin{pmatrix} — & U_1 & — \\ & \vdots & \\ — & U_n & — \end{pmatrix} U_i = d_i. \tag{A.259}$$

It was shown that a particular bilinear form is always nonnegative. It was shown that all eigenvalues of B^TB are the result of evaluating that bilinear form with a suitable argument. Therefore, all eigenvalues of B^TB are nonnegative. □

Lemma A.23. *For any matrix A, if $\mu I + A$ is invertible, then*

$$(\mu I + A)^{-1}A = A(\mu I + A)^{-1} \tag{A.260}$$

Proof. Multiplying Equation (A.260) by $(\mu I + A)$ gives

$$\begin{aligned} A &= (\mu I + A)A(\mu I + A)^{-1} && \text{(A.261)} \\ &= (\mu A + A^2)(\mu I + A)^{-1} && \text{(A.262)} \\ &= A(\mu I + A)(\mu I + A)^{-1} && \text{(A.263)} \\ &= A && \text{(A.264)} \end{aligned}$$

□

Lemma A.24. *Let B be a real matrix with real eigenvalue λ. Then there exists a real eigenvector v such that*

$$\lambda = v^T B v \tag{A.265}$$

and

$$||v|| = 1 \tag{A.266}$$

Proof. Let w be an eigenvector for the eigenvalue λ. Then

$$Bw = \lambda w \tag{A.267}$$

Let

$$v = w/||w|| \tag{A.268}$$

Note that

$$||v|| = 1 \tag{A.269}$$

Hence

$$v^T B v = \lambda v^T v = \lambda ||v||^2 = \lambda \tag{A.270}$$

□

Lemma A.25. *Let*

$$v = Au \tag{A.271}$$

where u is a real vector,

$$||u|| = 1 \tag{A.272}$$

and A is a binary diagonal matrix. Then

$$||v||^2 \leq 1 \tag{A.273}$$

Proof.

$$||v||^2 = v^T v = \sum_i (Au)_i^2 = \sum_i u_i^2 A_{ii} \leq \sum_i u_i^2 = 1 \tag{A.274}$$

□

Lemma A.26. *For real symmetric matrix A,*

$$\rho(A^2) = \rho(A)^2 \tag{A.275}$$

Proof. Because A is real symmetric,

$$A = UDU^T \tag{A.276}$$

Moreover,

$$A^2 = UD^2U^T \tag{A.277}$$

$$\xi \in \text{spec}(A^2) \leftrightarrow \xi = D_{ii}^2 = \zeta^2 \text{ where } \zeta = D_{ii} \in \text{spec}(A) \tag{A.278}$$

$$\begin{aligned}
\rho(A^2) &= \max\{|\lambda| \,:\, \lambda \in \text{ spec } (A^2)\} && \text{(A.279)}\\
&= \max\{|\lambda| \,:\, \lambda \in \{\zeta^2 \,:\, \zeta \in \text{ spec } (A)\}\} && \text{(A.280)}\\
&= \max\{|\zeta^2| \,:\, \zeta \in \text{ spec } (A)\} && \text{(A.281)}\\
&= \max\{|\zeta|^{\,2} : \zeta \in \text{ spec } (A)\} && \text{(A.282)}\\
&= \rho(A)^2 && \text{(A.283)}
\end{aligned}$$

□

Lemma A.27. *If B is a real symmetric matrix with real nonnegative eigenvalues, then there exists a real symmetric matrix A having real nonnegative eigenvalues such that*

$$B = A^2 \tag{A.284}$$

Lemma A.28. *For real symmetric matrix B with nonnegative real spectrum,*

$$\rho(B) = \sup_{||u||=1} u^T B u \tag{A.285}$$

Proof. Note that $\rho(B)$ is some function of the roots of $\det(B - \lambda I)$. For all invertible

matrices A,

$$\begin{aligned}
\det(ABA^{-1} - \lambda I) &= \det(A(B - \lambda I)A^{-1}) && \text{(A.286)}\\
&= \det(A)\det(B - \lambda I)\det(A^{-1}) && \text{(A.287)}\\
&= \det(A)\det(A^{-1})\det(B - \lambda I) && \text{(A.288)}\\
&= \det(B - \lambda I) && \text{(A.289)}
\end{aligned}$$

Hence

$$\rho(B) = \rho(ABA^{-1}) \tag{A.290}$$

From (A.290) and the fact that B is real symmetric,

$$\rho(B) = \rho(ABA^{-1}) = \rho(D) \tag{A.291}$$

where D is a diagonal matrix all of whose entries are nonnegative reals. Note that

$$\begin{aligned}
\sup_{||u||=1} u^T D u &= \sup_{||u||=1} (\sum e_j u_j)^T D (\sum e_i u_i) && \text{(A.292)}\\
&= \sup_{||u||=1} \sum_{i,j} u_j u_i e_j^T D e_i && \text{(A.293)}\\
&= \sup_{||u||=1} \sum_{i,j} u_j u_i d_i e_j^T e_i && \text{(A.294)}\\
&= \sup_{||u||=1} \sum_{i} u_i^2 d_i && \text{(A.295)}\\
&\leq \sup_{||u||=1} \sum_{i} u_i^2 \rho(D) && \text{(A.296)}\\
&= \rho(D) && \text{(A.297)}
\end{aligned}$$

Hence,

$$\sup_{||u||=1} u^T D u \leq \rho(D) \tag{A.298}$$

Now let

$$\lambda = \rho(D) \tag{A.299}$$

and let

$$\lambda v = Dv \tag{A.300}$$

where

$$||v|| = 1 \tag{A.301}$$

Then

$$\sup_{||u||=1} u^T D u \geq v^T D v = v^T \lambda v = \lambda ||v||^2 = \lambda = \rho(D) \tag{A.302}$$

Hence,

$$\sup_{||u||=1} u^T D u \geq \rho(D) \tag{A.303}$$

From (A.291), (A.298), and (A.303)

$$\rho(B) = \rho(D) = \sup_{||u||=1} u^T D u \tag{A.304}$$

Because B is real symmetric, there exists an orthogonal matrix U such that

$$B = U^T D U \tag{A.305}$$

where $U^T = U^{-1}$. Suppose $||u|| = 1$, then

$$||Uu|| = \sqrt{(Uu)^T (Uu)} = \sqrt{u^T U^T U u} = \sqrt{u^T u} = \sqrt{1} = 1 \tag{A.306}$$

Then by choosing u appropriately,

$$\sup_{||v||=1} v^T D v = (Uu)^T D (Uu) = u^T (U^T D U) u = u^T B u \leq \sup_{||v||=1} v^T B v \tag{A.307}$$

Note that

$$\sup_{||u||=1} u^T B u = \sup_{||u||=1} u^T U^T D U u = \sup_{||u||=1} (Uu)^T D (Uu) \tag{A.308}$$

Also, note that if

$$v = Uu \tag{A.309}$$

then

$$||v||^2 = v^T v = (Uu)^T U u = u^T U^T U u = ||u||^2 \tag{A.310}$$

Then, by choosing v appropriately,

$$\sup_{||u||=1} (Uu)^T D(Uu) = (Uu)^T D(Uu) = v^T Dv \leq \sup_{||v||=1} v^T Dv \tag{A.311}$$

Hence by (A.307) and (A.311),

$$\sup_{||u||=1} u^T Du = \sup_{||u||=1} u^T Bu \tag{A.312}$$

And hence by (A.304) and (A.312),

$$\rho(B) = \sup_{||u||=1} u^T Bu \tag{A.313}$$

as desired. □

Lemma A.29. *For square (complex) matrix A,*

$$\rho(A) \leq \sup_{||u||=1} ||Au|| \tag{A.314}$$

Proof. Let λ and v be such that

$$\rho(A) = |\lambda| \tag{A.315}$$

and

$$Av = \lambda v \tag{A.316}$$

where

$$||v|| = 1 \tag{A.317}$$

Then

$$\rho(A) = |\lambda| = |\lambda| \ \ ||v|| = ||\lambda v|| = ||Av|| \leq \sup_{||u||=1} ||Au|| \tag{A.318}$$

as desired. □

Theorem A.30. *Let*

$$\Lambda = \begin{pmatrix} \lambda_0 & & 0 \\ & \ddots & \\ 0 & & \lambda_{n-1} \end{pmatrix} \tag{A.319}$$

be a real diagonal matrix,

$$\omega(u) = e^{2\pi\sqrt{-1}\frac{u}{n}} \tag{A.320}$$

and extend λ_v to all integer subscripts by

$$i = v \ mod \ n \Rightarrow \lambda_i = \lambda_v \tag{A.321}$$

Then $F^*\Lambda F$ is real iff for all i and j

$$0 = \sum_v \omega(v(i-j))(\lambda_v - \lambda_{-v}) \tag{A.322}$$

Proof. Note that $F^*\Lambda F$ is real iff

$$F^*\Lambda F = F^T\Lambda\overline{F}. \tag{A.323}$$

The i,jth entry of the left-hand side of Equation (A.323) is

$$\begin{aligned} (F^*\Lambda F)_{i,j} &= (\overline{F}\Lambda F)_{i,j} && \text{(A.324)} \\ &= \sum_{u,v} \overline{F_{i,u}}[u=v]F_{v,j}\lambda_u && \text{(A.325)} \\ &= \sum_u \overline{F_{i,u}}F_{u,j}\lambda_u && \text{(A.326)} \\ &= \frac{1}{n}\sum_u \omega(-iu)\omega(uj)\lambda_u && \text{(A.327)} \\ &= \frac{1}{n}\sum_{u=0}^{n-1} \omega(u(j-i))\lambda_u && \text{(A.328)} \\ & && \text{(A.329)} \end{aligned}$$

Let

$$v = -u \tag{A.330}$$

Then

$$\frac{1}{n}\sum_{u=0}^{n-1}\omega(u(j-i))\lambda_u \;=\; \frac{1}{n}\sum_{v=1-n}^{0}\omega(v(i-j))\lambda_{-v} \tag{A.331}$$

Let

$$q = n + v \tag{A.332}$$

Then

$$\frac{1}{n}\sum_{v=1-n}^{0}\omega(v(i-j))\lambda_{-v} \;=\; \frac{1}{n}\sum_{q=1}^{n}\omega(q(i-j))\lambda_{n-q} \tag{A.333}$$

Standardize the range by interpreting subscripts of λ mod n and trading n for 0. Then

$$\frac{1}{n}\sum_{q=1}^{n}\omega(q(i-j))\lambda_{n-q} \;=\; \frac{1}{n}\sum_{q=0}^{n-1}\omega(q(i-j))\lambda_{-q} \tag{A.334}$$

The i, j-th entry of the right-hand side of Equation (A.323) is

$$\begin{aligned}
(F^T\Lambda\overline{F})_{i,j} &= (F\Lambda\overline{F})_{i,j} && \text{(A.335)}\\
&= \sum_{u,v} F_{i,u}[u=v]\overline{F_{v,j}}\lambda_u && \text{(A.336)}\\
&= \sum_{u} F_{i,u}\overline{F_{u,j}}\lambda_u && \text{(A.337)}\\
&= \frac{1}{n}\sum_{u}\omega(iu)\omega(-uj)\lambda_u && \text{(A.338)}\\
&= \frac{1}{n}\sum_{u=0}^{n-1}\omega(u(i-j))\lambda_u && \text{(A.339)}
\end{aligned}$$

Hence Equation (A.323) is equivalent to

$$0 = \sum_{v}\omega(v(i-j))(\lambda_v - \lambda_{-v}) \tag{A.340}$$

□

Vita

Mitchel DeWayne Horton was born in Kingsport, TN on May 24, 1959. He was raised in Kingsport, TN, went to grade school at Andrew Johnson Elementary School, and went to junior high school at Ross N. Robinson Junior High School. He graduated from Tri-Cities Christian High School in 1977. He received a B.S. in mechanical engineering from the University of Tennessee in 1984 and an M.S. in mechanical engineering from Duke University in 1985. He spent 19 years working in industry as a software development engineer.

Mitch completed his doctorate in computer science at the University of Tennessee in May, 2010.