

ALGEBRAIC RECONSTRUCTION ALGORITHMS FOR REMOTE  
SENSING IMAGE ENHANCEMENT

by

Matthew Willis

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

Brigham Young University

August 2000

Copyright © 2000 Matthew Willis

All Rights Reserved

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Matthew Willis

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

\_\_\_\_\_

Date

\_\_\_\_\_

Brian D. Jeffs, Chair

\_\_\_\_\_

Date

\_\_\_\_\_

David G. Long

\_\_\_\_\_

Date

\_\_\_\_\_

Randy Beard

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Matthew Willis in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

---

Date

---

Brian D. Jeffs  
Chair, Graduate Committee

Accepted for the Department

---

A. Lee Swindlehurst  
Graduate Coordinator

Accepted for the College

---

Douglas M. Chabries  
Dean, College of Engineering and Technology

## ABSTRACT

# ALGEBRAIC RECONSTRUCTION ALGORITHMS FOR REMOTE SENSING IMAGE ENHANCEMENT

Matthew Willis

Department of Electrical and Computer Engineering

Master of Science

This thesis explores algorithms used for satellite remote sensing image reconstruction. The primary aim is to develop and study algorithms for high resolution image reconstruction from orbital radiometer and scatterometer data. The discussion first focuses on the MART Maximum Entropy (ME) reconstruction algorithm and then extends the MART algorithm to include other forms of ME algorithms. These several ME algorithm forms are studied to compare stable convergence rates and performance in the presence of noise. Column normalized methods, a class of algorithms with good artifact suppression, are then introduced. The convergence properties of the Simultaneous Algebraic Reconstruction Technique (SART) algorithm are studied to understand the behavior of this class of algorithms. Structural comparisons of SART with the SIR algorithm give insight into the success of the SIR algorithm in remote sensing image reconstruction. This thesis concludes with experimental performance comparisons of several different algorithms for both real and simulated data.

## ACKNOWLEDGMENTS

I would first like to thank my advisor, Brian Jeffs, whose continual support, encouragement, and helpful insights have greatly aided me in this research. I would also like to thank David Long, for both his financial support of this research as well as his intellectual contributions to the field of image reconstruction.

And finally, I thank my family, especially my wife, Andra, for her understanding and patient disposition, and her unfailing, selfless support.

# Contents

<b>Acknowledgments</b>	<b>vi</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem Description . . . . .	2
1.2 Algebraic Reconstruction Techniques . . . . .	4
1.3 Research Contributions . . . . .	5
1.4 Thesis Outline . . . . .	6
<b>2 Imaging with Remote Sensing Instruments</b>	<b>9</b>
2.1 Radiometers . . . . .	9
2.2 Scatterometers . . . . .	11
<b>3 Row-Action Techniques</b>	<b>15</b>
3.1 Introduction . . . . .	15
3.2 Primal-dual Algorithms . . . . .	17
3.3 Maximum Entropy Algorithm Convergence . . . . .	19
3.3.1 Normal Implementation . . . . .	20
3.3.2 Block Implementation . . . . .	22
3.4 Maximum Entropy Algorithm Forms . . . . .	24
3.4.1 ME Algorithm Definitions . . . . .	25
3.4.2 Relaxation Factors . . . . .	26
3.5 Comparison of the Different ME Algorithms . . . . .	27

3.5.1	Fully Determined System . . . . .	27
3.5.2	Underdetermined System . . . . .	32
3.5.3	Inconsistent and Overdetermined Systems . . . . .	39
3.6	A Closed-form ME Approximation Algorithm . . . . .	41
3.6.1	A Fast Closed Form Implementation . . . . .	43
3.6.2	Experimental Results . . . . .	45
3.7	Problems with ART and MART . . . . .	45
<b>4</b>	<b>Column-normalized Algorithms</b>	<b>47</b>
4.1	Column-normalized Algorithms . . . . .	47
4.2	Simultaneous Algebraic Reconstruction Technique (SART) . . . . .	47
4.2.1	Finding the Solution . . . . .	48
4.2.2	SART Convergence . . . . .	52
4.2.3	Advantages of the Column Normalization . . . . .	56
4.3	BYU MART and SIR . . . . .	56
4.3.1	SIR . . . . .	59
4.3.2	BYU MART and SIR Convergence . . . . .	60
4.4	Additive SIR . . . . .	60
4.4.1	Additive SIR Convergence . . . . .	63
4.5	Conclusion . . . . .	66
<b>5</b>	<b>Algorithm Performance Comparison Using Simulated Data</b>	<b>69</b>
5.1	The Modified Median Filter . . . . .	70
5.2	SSM/I Geometry Simulations . . . . .	72
5.2.1	Noiseless Simulations . . . . .	73
5.2.2	Noisy Simulation . . . . .	76
5.3	NSCAT Geometry Simulations . . . . .	76
5.4	Conclusions . . . . .	84
<b>6</b>	<b>Algorithm Performance Comparison Using Real Data</b>	<b>85</b>
6.1	SSM/I Reconstructions . . . . .	85

6.2	NSCAT Reconstructions . . . . .	86
6.3	Conclusions . . . . .	89
<b>7</b>	<b>Conclusions</b>	<b>93</b>
7.1	Future Research . . . . .	94
<b>A</b>	<b>BYU MART and SIR Convergence Properties</b>	<b>97</b>
A.1	Convergence Properties . . . . .	97
A.1.1	Linear Case: BYU MART . . . . .	97
A.1.2	Non-Linear Case . . . . .	99
A.2	Conclusions . . . . .	104
A.3	Post Script . . . . .	104
	<b>Bibliography</b>	<b>109</b>



## List of Tables

3.1	Results of the noisy fully determined case, SNR=22.2 dB, where $k_{\min}$ is the iteration at which the solution error is minimized. . . . .	34
3.2	Results of the higher-dimensional noisy fully determined case, SNR=22.5 dB, where $k_{\min}$ is the iteration at which the solution error is minimized.	34
3.3	Typical Results for underdetermined, noisy reconstruction at 11.4 dB and 14.5 dB SNR. . . . .	35
3.4	Results of the higher-dimensional underdetermined case with 26.9 dB SNR, where $k_{\min}$ is the iteration at which the solution error is minimized. . . . .	37
3.5	Results of the higher-dimensional noisy over determined, inconsistent case, SNR=22.5 dB, where $k_{\min}$ is the iteration at which the solution error is minimized. . . . .	40



## List of Figures

1.1	Geometry of image reconstruction in remote sensing . . . . .	3
1.2	Geometry of image reconstruction in medical imaging . . . . .	4
2.1	Fourier transform of a typical SSM/I footprint (left), and a one-dimensional slice through the center (right). . . . .	10
2.2	Fourier transform of a typical NSCAT footprint (top), and one-dimensional slices taken diagonally through the center (bottom). . . . .	12
3.1	Plot of $\exp(y_i - \langle h_i, x^k \rangle)$ for $10 < y_i - \langle h_i, x^k \rangle < 10$ (left) and $\exp\left(\left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1\right)$ for $0 < \left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1 < 10$ (left). . . . .	27
3.2	Plot of $\exp((y_i - \langle h_i, x^k \rangle) * 0.05)$ for $10 < y_i - \langle h_i, x^k \rangle < 10$ (left) and $\exp\left(\left(\frac{y_i}{\langle h_i, x^k \rangle}\right)^{0.4} - 1\right)$ for $0 < \left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1 < 10$ (left). . . . .	28
3.3	Convergence of the three algorithms with a fully determined system—no noise (top) and noisy (bottom) with $\text{SNR} = 22.2$ dB. The actual trajectories are in solid lines (MART—blue, Elfving MART—green, Willis MART—red) and their projections onto the x, y, and z planes are dotted. . . . .	29
3.4	Solution error, shown at two different scales for clarity, of the various ME algorithms as a function of iteration for the fully determined, noisy case, $\text{SNR}=22.2$ dB. (top) Detail scale, 0–60 iterations, shows minimum error point is the same for both MART and Willis MART. (bottom) large scale, 0–1500 iterations—shows minimum error point for Elfving MART. . . . .	31

3.5	Reconstruction of the noisy, fully determined system: (a) Synthetic original image (b) Blurred image with 22.5 dB SNR noise added to the measurements (c) Algorithm 4 (MART) reconstruction (d) Algorithm 5 (Elfving MART) reconstruction (e) Algorithm 6 (Willis MART) reconstruction (f) The blurring function used to build the circular convolution matrix $\mathcal{H}$ . . . . .	33
3.6	Trajectories of the three algorithms for a noiseless, underdetermined case. . . . .	34
3.7	Results of noisy underdetermined system at 14.5 dB SNR (top) and 11.4 dB SNR (bottom). . . . .	36
3.8	Singular values for the three-dimensional $\mathcal{H}$ (left) and the higher-dimensional $\mathcal{H}$ (right). . . . .	37
3.9	Underdetermined 26.9 dB SNR noisy reconstructions of the synthetic truth image (d) for Algorithms (a) 4 (MART), (b) 5 (Elfving MART), and (c) 6 (Willis MART). . . . .	38
3.10	Overdetermined 22.5 dB SNR noisy reconstructions of the synthetic truth image (d) for Algorithms (a) 4 (MART), (b) 5 (Elfving MART), and (c) 6 (Willis MART). . . . .	40
3.11	Example results from the closed form maximum entropy algorithm. a) Original $256 \times 256$ pixel true image. b) Observed image, blurred with a circular psf low pass filter with cutoff frequency less than Nyquist / 4. No noise was added. c) Maximum entropy reconstruction using MART. d) Closed form approximation of maximum entropy. . . . .	46
4.1	Reconstruction from noiseless, randomly sampled data: a) ART b) MART c) SART d) Original Image . . . . .	57
4.2	Reconstruction from noiseless, regularly decimated sampling: a) ART b) MART c) SART d)Original Image . . . . .	58
4.3	Reconstruction from noiseless random sampling: a) SIR b) BYU MART c) MART d) Original Image . . . . .	61

4.4	Reconstruction from noiseless regularly decimated sampling: a) SIR b) BYU MART c) MART d) Original Image . . . . .	62
4.5	Reconstruction from noiseless random sampling: a) Additive SIR b) MART c) ART d)Original Image . . . . .	64
4.6	Reconstruction from noiseless regularly decimated sampling: a) Additive SIR b) MART c) ART d)Original Image . . . . .	65
4.7	Convergence trajectories for the SART and Additive SIR algorithms	67
5.1	The original synthetic “truth image” used for simulations . . . . .	70
5.2	The area in the black rectangle represents the region in the Amazon basin from which the measurement geometries are taken (the area is magnified at the bottom right hand corner of the image). . . . .	71
5.3	SSM/I sampling geometry—The dots in the left figure indicate the center of each footprint. The right figure shows a decimated version of the sampling to give a view of some representative footprints (both figures are on the same scale). . . . .	72
5.4	Noiseless reconstructions for SSM/I geometry using (a) MART (b) MART with filter (MARTF) (c) ART (d) ART with filter (ARTF) (e) SART (f) SART with filter (SARTF) . . . . .	74
5.5	Noiseless reconstructions for SSM/I geometry using (a) SIR (b) SIR with filter (SIRF) (c) AddSIR (d) AddSIR with filter (AddSIRF) . . .	75
5.6	45.0 dB SNR SSM/I reconstruction using (a) MART—398 iterations (b) MART with filter (MARTF)—42 iterations (c) ART—20 iterations (d) ART with filter (ARTF)—12 iterations (e) SART—1058 iterations (f) SART with filter (SARTF)—50iterations . . . . .	77
5.7	45.0 dB SNR SSM/I reconstruction using (a) SIR—694 iterations (b) SIR with filter (SIRF)—27 iterations (c) AddSIR—2098 iterations (d) AddSIR with filter (AddSIRF)—50 iterations . . . . .	78

5.8	NSCAT sampling geometry—The dots in the left figure indicate the center of each footprint. The right figure shows a decimated version of the sampling to give a view of some representative footprints (both figures are on the same scale. . . . .	79
5.9	Noiseless NSCAT reconstruction using (a) MART (b) MARTF (c) ART (d) ARTF (e) SART (f) SARTF . . . . .	80
5.10	Noiseless NSCAT reconstructions using (a) SIR (b) SIRF (c) AddSIR (d) AddSIRF . . . . .	81
5.11	19.7 dB SNR NSCAT reconstruction using (a) MART (b) MARTF (c) ART (d) ARTF (e) SART (f) SARTF . . . . .	82
5.12	19.7 dB SNR reconstruction using (a) SIR (b) SIRF (c) AddSIR (d) AddSIRF . . . . .	83
6.1	The result of using $x^\circ = 1/e$ for SSM/I data for MART after 15 iterations (left), and MARTF after 50 iterations (right). . . . .	86
6.2	SSM/I reconstruction of real data using (a) MART—30 iterations (b) MARTF—40 iterations (c) ART—5 iterations (d) ARTF—15 iterations (e) SART—30 iterations (f) SARTF—30 iterations . . . . .	87
6.3	SSM/I reconstruction of real data using (a) SIR—10 iterations (b) SIRF—10 iterations (c) AddSIR—10 iterations (d) AddSIRF—15 iterations . . . . .	88
6.4	NSCAT reconstruction of real data using (a) MART—60 iterations (b) MARTF—10 iterations (c) ART—15 iterations (d) ARTF—15 iterations (e) SART—20 iterations (f) SARTF—20 iterations . . . . .	90
6.5	NSCAT reconstruction of real data using (a) SIR—30 iterations (b) SIRF—45 iterations (c) AddSIR—25 iterations (d) AddSIRF—30 iterations . . . . .	91

# Chapter 1

## Introduction

With humankind's recent interest in global climatology and the impact that we have made on our environment, there is a pressing need for fast, accurate measurements of key global parameters such as ocean wind speed, ocean height, and land and ice characteristics. Some of the land and ice characteristics that are of interest include snow-cover classification, plant and soil moisture content, vegetation classification, and polar ice-extent mapping.

Microwave satellite instruments, such as radiometers and scatterometers, play an important role in helping determine these key global parameters. The advantage of microwave instruments is that they can take measurements of the earth's surface day or night, and are not affected by cloud cover. Another advantage of these instruments is their rapid repeat coverage. For example, the recent QuickScat scatterometer is able to cover almost the entire earth in one day.

The tradeoff for the fast coverage of radiometers and scatterometers is their low spatial resolution. Radiometer and scatterometer measurements over the ocean have been successfully used to provide data for numerical weather-prediction models and to measure ocean wind speed, but low resolution has limited their use in land and ice studies. However, by applying resolution enhancement algorithms to the data, images with sufficient detail for land and ice studies can be obtained.

## 1.1 Problem Description

Figure 1.1 illustrates the remote sensing imaging problem. First, the observed image is discretized into pixels. Each distinct measurement corresponds to the received microwave signal from an area on the ground, multiplied by the antenna aperture function. This single measurement ground area/aperture footprint usually covers a number of image pixels at the desired resolution for the restored image. In the case of a radiometer, this observation is the weighted average of the radiometric brightness temperature of the pixels in the aperture multiplied by the antenna response at each pixel. For a scatterometer, a weighted average over the illuminated region is observed as the microwave backscatter. The value of the  $i$ 'th measurement,  $y_i$ , then, is  $y_i = \langle \vec{x}, \vec{h}_i \rangle$  where  $\langle \cdot, \cdot \rangle$  denotes the standard vector inner product,  $\vec{x}$  is the row-scanned image vector, and  $\vec{h}_i$  contains information about the value and location of the antenna footprint, such that  $h_{ij}$  represents the contribution of the  $j$ 'th pixel to the  $i$ 'th measurement.

The measurements are stacked to yield the vector-matrix observation model

$$\vec{y} = \mathcal{H}\vec{x} \tag{1.1}$$

where  $y : \mathbb{R}^m$  is the measurement vector, and  $\mathcal{H} : \mathbb{R}^{m \times n}$  is the point spread matrix, or transfer matrix formed from the antenna footprint. The remote sensing imaging problem, then, is to estimate  $\vec{x}$ , given observations  $\vec{y}$  and prior knowledge of  $\mathcal{H}$ .

For the remainder of this thesis the vector symbol notation is dropped, with the understanding that variables without subscripts are understood to be vectors. The single exception to this rule is that  $h_i$  represents the  $i$ 'th row of  $\mathcal{H}$ .

Methods used to solve the remote sensing imaging problem were originally developed for image reconstruction in Computerized Tomography (CT) and electron microscopy. Figure 1.2 shows the geometry for image reconstruction in absorption tomography medical imaging for non diffracting ray propagation. This problem has many structural similarities to our remote sensing problem. The observed object (it could be a ribosome for electron microscopy, or a human head for X-ray C.T.) is discretized into pixels, and then an electromagnetic ray is sent through the object.

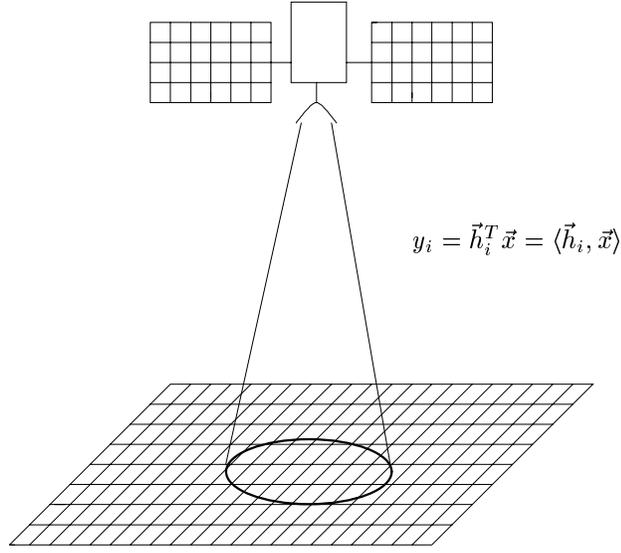


Figure 1.1: Geometry of image reconstruction in remote sensing

The measurement corresponding to the  $i$ 'th ray,  $y_i$ , is the attenuation of the ray. In the discretized image the attenuation is just the sum of the attenuation of the pixels transversed by the ray. As for remote sensing,  $y_i = \langle \vec{x}, \vec{h}_i \rangle$  where  $\vec{x}$  is the row-scanned image and  $\vec{h}_i$  contains the information about which pixels intersected the ray. Again, the measurements can be stacked to give Eq. (1.1).

Reconstructing an image from a set of measurements, as is done in medical imaging and remote sensing, is called *image reconstruction from projections*. There are several known approaches for reconstructing an image from projections. Some of the earliest methods were *Transform Based*, where the problem is worked in the Fourier domain. More recently, Alvarez-Perez et. al. used a transform based Weiner filter approach to improve resolution for ERS scatterometer data [1]. Some of the weaknesses of the Fourier methods is that they require regular (i.e., uniform) spatial sampling of the object and are computationally intensive [2],[3],[4]. Also, it must usually be assumed that the antenna aperture is spatially invariant. All of these restrictions are problematic in earth remote sensing reconstruction, where sampling

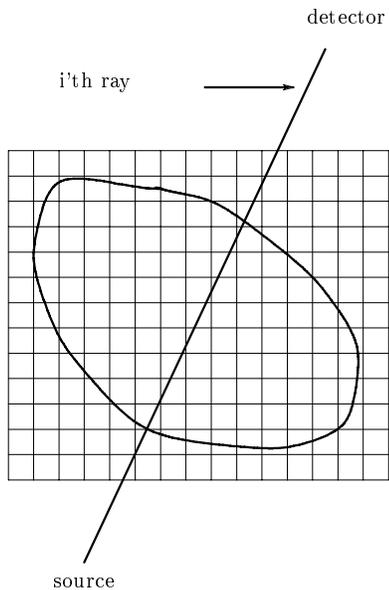


Figure 1.2: Geometry of image reconstruction in medical imaging

is irregular and the antenna aperture response is spatially varying. These difficulties motivate our interest in the algebraic methods discussed in the next section.

## 1.2 Algebraic Reconstruction Techniques

Another promising family of algorithms are the algebraic reconstruction techniques. These methods are iterative and can be used to reconstruct an image that is arbitrarily sampled with a spatially varying aperture. Two of the most well-known algorithms are the Algebraic Reconstruction Technique (ART) and the Multiplicative ART (MART), which appeared in [5] and [6] respectively for use in CT scanning reconstruction. The ART algorithm produces a least-squares solution satisfying the constrained minimization

$$\min_x \|x\|_2^2 \quad \text{such that} \quad Hx = y. \quad (1.2)$$

MART is a Maximum Entropy (ME) algorithm, with a solution satisfying

$$\min_x \sum_x x \ln x \quad \text{such that} \quad \mathcal{H}x = y. \quad (1.3)$$

The ART and MART algorithms will be discussed in more detail in Chapter 3.

There are important differences between medical imaging and remote sensing. Usually in medical imaging the observations are made at somewhat regular sample angles, and the image is oversampled. Remote sensing sampling, however, is irregular and, although sometimes oversampled, can also be undersampled. Because of these characteristics, there are shortcomings in the conventional algorithms (ART and MART) when applied to remote sensing reconstruction. These shortcomings led the Brigham Young University Microwave Earth Remote Sensing (BYU MERS) Lab to develop a new algebraic reconstruction technique, the Scatterometer Image Reconstruction algorithm (SIR) [7], for image reconstruction from remote sensing data. SIR not only gave superior results with fewer reconstruction artifacts than ART and MART, but it has been shown to actually increase resolution by enhancing the high frequency data contained in the side lobes of the antenna aperture response [8], [9]. Because of the success of SIR, there has been significant interest in understanding algebraic reconstruction algorithms in general and to understand and explain the success of SIR.

This thesis discusses and compares some of the algebraic techniques useful for creating enhanced resolution reconstructed images. Specifically, *column-normalized* algorithms, a new class of algorithms which includes SIR, is introduced. The characteristics of this class of algorithms that make them especially suited for microwave remote sensing image reconstruction are explored and evaluated.

### 1.3 Research Contributions

This thesis provides a firmer theoretical and practical understanding of algebraic reconstruction algorithms, specifically their properties as applied to remote sensing problems. Prior to this work there was little knowledge about why some algorithms performed better than others for this class of problems. Further, while

pursuing these goals, significant theoretical and algorithmic advances in the related ME problem were developed.

The research contributions presented here include several investigations into the theory and performance of iterative reconstruction methods. A new look at the convergence of Maximum Entropy (ME) iterative algorithms is developed in the context of a general primal-dual algorithm given by Lent in [10]. The proof of convergence given by Lent is expanded to include different forms of iterative update corrections. Different ME algorithms with several distinct update terms are then examined to see if any form has better performance in the presence of noise. Also, a fast, closed form algorithm that approximates the ME solution is derived and presented.

A new class of algorithms, called *column-normalized* algorithms, is then defined. As an example of a column-normalized algorithm, the Simultaneous ART (SART) algorithm, suggested by Andersen in [11] for CT scanning, is studied. It is found, and proved, that SART converges to a weighted minimum norm solution, where the weighting depends on the pixel sampling. While the precise type of solution to which SART converges is unknown, the SART analysis gives some insight into why the SART algorithm, as a column-normalized algorithm, performs better than the standard algorithms (ART and MART).

Several different column-normalized algorithms are compared experimentally to show that they give similar results and that SART is a viable algorithm for remote sensing reconstruction.

## 1.4 Thesis Outline

Chapter 2 gives a brief introduction to microwave scatterometers and radiometers. The chapter specifically treats the aspects of the systems that are important to imaging.

Chapter 3 discusses *row-action method* algorithms, the class of algorithms to which ART and MART belong. In this chapter the convergence proof of MART is expanded to include several forms of ME algorithms, and the different algorithms will be compared to see if any of them performs best in the presence of noise. A closed

form approximation to the MART algorithm is also presented. At the end of the chapter, some of the weaknesses of row-action algorithms are shown.

Chapter 4 introduces column-normalized algorithms as a way of compensating for the problems with row-action methods mentioned in Chapter 3. The convergence properties of the SART algorithm are explored, and its results are related to other column-normalized algorithms, such as SIR.

Chapters 5 and 6 compare the different algorithms for simulated and real data, respectively, and in Chapter 7 important conclusions are given and the original contributions reiterated.



## Chapter 2

### Imaging with Remote Sensing Instruments

Microwave instruments such as Synthetic Aperture Radar (SAR) are radars designed to image the earth's surface. This thesis is motivated by the desire to generate images from microwave scatterometer and radiometer data, which were not originally designed for imaging. Scatterometers and radiometers were designed mostly for indirect ocean and climatological measurements, where high resolution isn't necessary. In exchange for lower resolution (as compared to SAR) they provide much more rapid, full coverage of the earth.

For the image reconstruction experiments presented in Chapters 5 and 6, sampling geometry from the SSM/I radiometer and NSCAT scatterometer is used. To help explain the results from the different instruments, this chapter discusses the basic operation and characteristics of microwave radiometers and scatterometers that affect the quality of image reconstruction. A very high level discussion of instrument operation will be presented to explain how instrument operation affects the quality of the reconstructed image. For more details on radiometers and scatterometers the reader is referred to [12].

#### 2.1 Radiometers

A radiometer is a passive microwave instrument that detects incoherent radiometric energy at a given frequency. To ensure that a radiometer receives signals only from the area of interest, radiometer antennas are designed to have a very tapered aperture amplitude spatial response, or footprint, to keep the antenna sidelobes low.

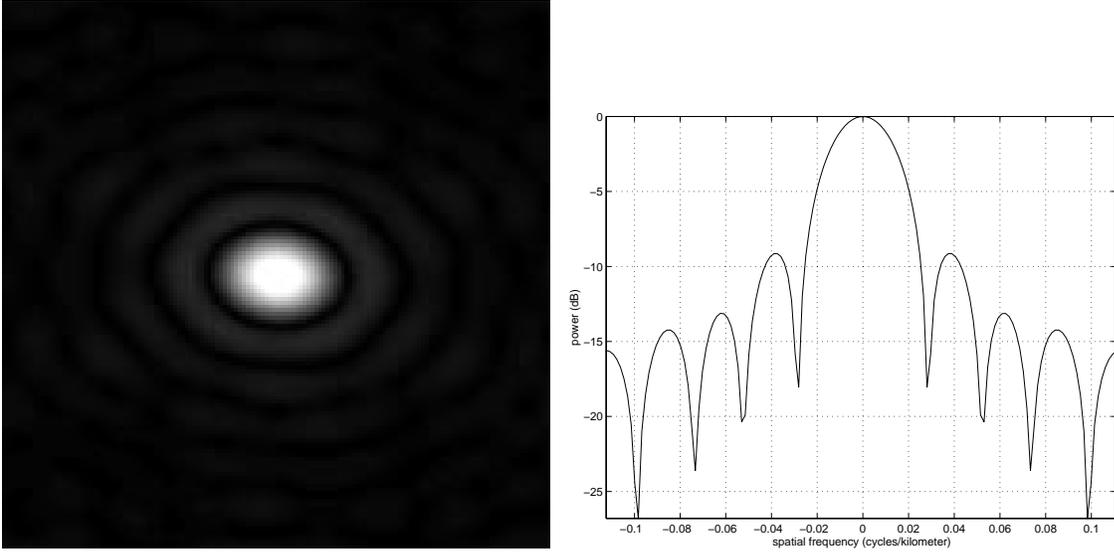


Figure 2.1: Fourier transform of a typical SSM/I footprint (left), and a one-dimensional slice through the center (right).

The frequency response of the antenna footprint on the ground is important in resolution enhancement. As was mentioned in Chapter 1, algebraic reconstruction algorithms create enhanced resolution images by emphasizing higher frequency information in the sidelobes of the footprint response.

Figure 2.1 shows the frequency response of a typical SSM/I footprint. Notice that the response is in units of spatial frequency, cycles per unit kilometer. A conventional reconstruction may only recover those frequencies within the 3-dB bandwidth of the aperture response [9]. The goal of the resolution enhancement algorithm is to increase the high frequency signal power contained in the sidelobes. To understand how an image reconstruction algorithm works, consider the ideal case of an image being fully sampled (that is, as many independent samples as pixels in the reconstructed image) with a spatially invariant footprint. In this case the transfer matrix  $\mathcal{H}$  in Eq. (1.1) is a linear convolution matrix. The image reconstruction algorithm iteratively inverts the matrix  $\mathcal{H}$ , deconvolving the image and recovering all the frequencies up to the Nyquist frequency, with the exception of those frequencies in the

null space of the aperture. The Nyquist frequency is inversely proportional to the sample grid spacing.

In practice, the range of recoverable frequency content is limited by the sampling pattern and noise. Because the image is generally irregularly and not fully sampled, the highest recoverable frequency decreases (see [8]). Also, any higher frequency contained in sidelobes below the noise floor will also be unrecoverable, further decreasing the range of attainable spatial frequencies.

## 2.2 Scatterometers

Scatterometers are active instruments, transmitting a coherent pulse and then measuring the normalized electromagnetic backscatter according to the monostatic radar equation [12]. Because a scatterometer is a coherent detector, the properties of its footprint are different than for radiometers. First, techniques such as Doppler range gating can be used to increase the resolution (decrease the size) of the footprint, which will increase the width of the main lobe in the frequency domain. Doppler range gating also increases the number of measurements the instrument can take on a given pass. Secondly, the scatterometer's antenna footprint is not as tapered as a radiometer's. Such a footprint leads to higher sidelobes in the footprint frequency response, which facilitates reconstructing higher spatial frequencies.

Figure 2.2 shows the Fourier transform for a typical NSCAT scatterometer footprint. NSCAT's footprints are narrow and elongated, giving very good resolution in one direction. Good resolution in all directions is achieved by overlapping measurements which have different orientations due to orbital ground path variations from different passes of the satellite. This suggests that NSCAT should give better resolution reconstructions than SSM/I. Chapters 5 and 6 confirm this.

Thus, an important issue when designing a microwave instrument to be used for imaging is the size and shape of the sensor antenna footprint. It is desirable to increase the spatial sampling, that is, obtain as many measurements per pass of the instrument as possible. However, there is a tradeoff between the number of measurements and the signal-to-noise ratio. Increasing the number of measurements

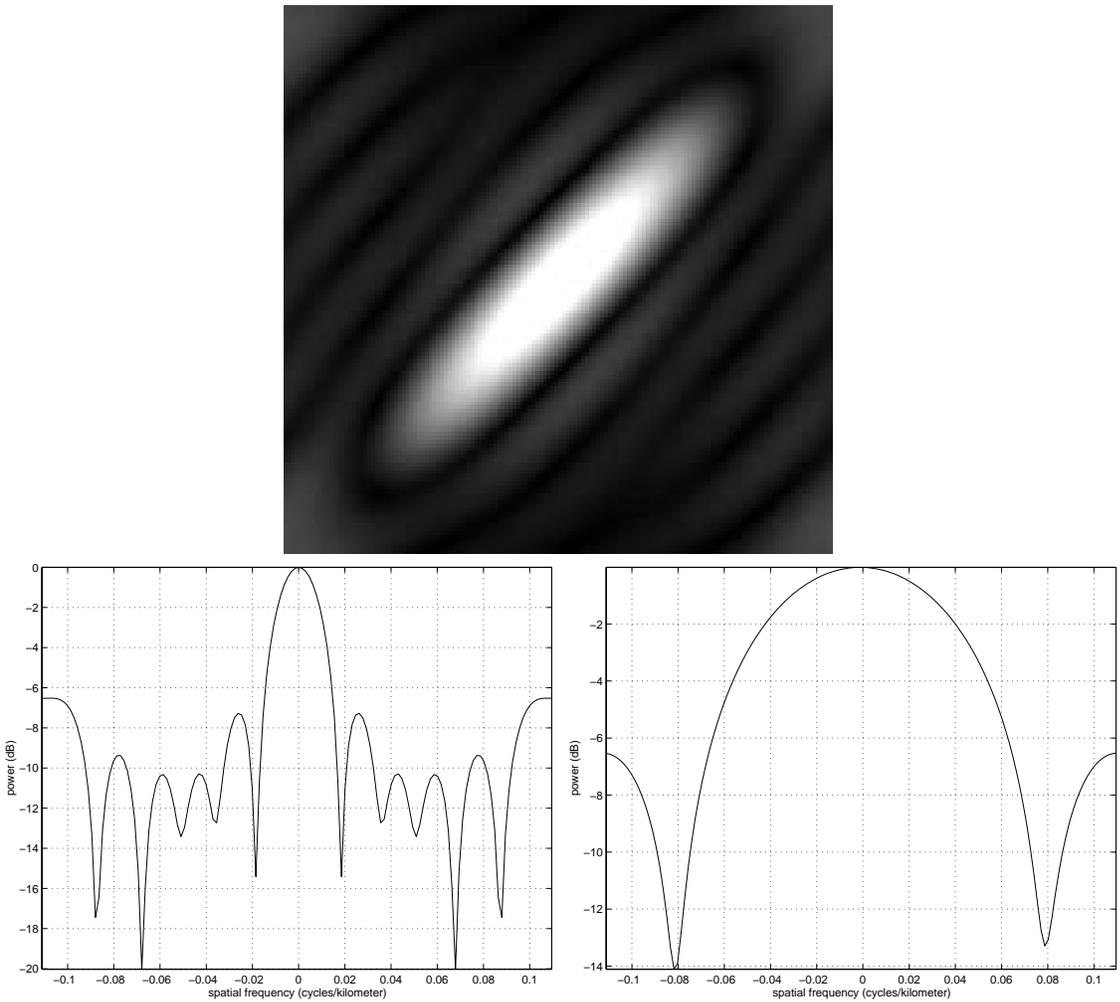


Figure 2.2: Fourier transform of a typical NSCAT footprint (top), and one-dimensional slices taken diagonally through the center (bottom).

by decreasing the size of the footprint via techniques such as Doppler range gating decreases the time-bandwidth product of the instrument, effectively decreasing the signal-to-noise ratio (SNR). Hence, when designing a system for imaging, there is a compromise between collecting more measurements to obtain a higher Nyquist frequency and corrupting the signal in the sidelobes with noise.



## Chapter 3

### Row-Action Techniques

#### 3.1 Introduction

A common and successful class of iterative algorithms used for image reconstruction is *row-action methods* (RA-methods), which were defined by Censor in [13].

##### **Definition 1 (Row-action Method)**

A row-action method is an iterative algorithm which has the following properties:

1. No changes are made to the original matrix,  $\mathcal{H}$ ,
2. No operations are performed on the matrix as a whole,
3. In a single iterative step, access is required to only one row of the matrix,
4. In a single iterative step, say when  $x^{k+1}$  is calculated, the only iterate needed is the immediate predecessor,  $x^k$ .

Because of the third definition constraint there is no need to store the entire matrix  $\mathcal{H}$  in the computer's memory. This is of practical importance in remote sensing where there are millions of pixels in the image ( $n \approx 1e6$ ) and hundreds of thousands of measurements ( $m \approx 1e5$ ), with a resulting  $H$  that may be much too big to fit in computer memory. This chapter will discuss in detail the properties of some row action algorithms.

Two very common row-action algorithms are the Algebraic Reconstruction Technique (ART) and the Multiplicative ART (MART) algorithms.

The ART algorithm was suggested for medical CT reconstruction in 1970 by Gordon *et al* [5], and was later proved to converge to the minimum norm solution [14], satisfying the constrained minimization problem

$$\min_x \|x\|_2^2 \quad \text{such that} \quad \mathcal{H}x = y. \quad (3.1)$$

The algorithm update is

**Algorithm 1 (ART)**

*Initialization:*  $x^\circ = 0$

*Iterative step:*

$$x_j^{k+1} = x_j^k + \frac{y_i - \langle h_i, x^k \rangle}{\|h_i\|} h_{ij} \quad (3.2)$$

where  $x_j^k$  is the  $j$ th element of the vector  $x$  at the  $k$ th iteration,  $h_i$  is the  $i$ th row of  $\mathcal{H}$ , and  $h_{ij}$  is the  $(i,j)$ th element of  $\mathcal{H}$ .

ART is a simple, intuitive algorithm. If the current guess  $x^k$  is too large, then the fidelity term,  $d_i^k = y_i - \langle h_i, x^k \rangle$  will be negative and will drive down the value of the pixel. ART is a Projection Onto Convex Sets (POCS) algorithm, where the iterations project the solution on the hyperplanes defined by the rows of  $\mathcal{H}$  and observations  $y$ .

Minimum norm regularization, although common, is not the only possible regularization that can be used to find a unique solution to 1.1. Another regularization, maximum entropy, was suggested for use in radio astronomy image reconstruction by Gull in 1978 [15]. Before that, in 1970, Gordon *et al.* suggested using a Multiplicative ART (MART) algorithm for CT image reconstruction. The MART algorithm was later proved to be a maximum entropy algorithm, satisfying the constrained optimization

$$\min_x \sum_i x_i \ln x_i \quad \text{such that} \quad \mathcal{H}x = y. \quad (3.3)$$

The algorithm update is

**Algorithm 2 (MART)**

*Initialization:*  $x^\circ = \exp(-1)$

*Iterative step:*

$$x_j^{k+1} = x_j^k \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^{\lambda_1 h_{ij}} \quad (3.4)$$

where  $\lambda_1$  is a relaxation parameter.

The maximum entropy solution is often favored for being the “least committal” solution. Maximum entropy reconstruction is an information theoretic method which treats the image itself as a probability density function, or pdf. Since maximum entropy distributions are often uniform intensity, maximum entropy reconstructions tend to have uniform areas. This explains why maximum entropy reconstruction has been so successful in radio astronomy with its uniform bright and dark areas.

An area of important research, when studying iterative algorithms, is convergence properties. The following section will introduce the general *primal-dual* algorithm, which will aid in understanding the convergence properties for some of the algorithms presented in this thesis.

### 3.2 Primal-dual Algorithms

This section relies on the work of Lent and Censor ([10]) who proved convergence for a general primal-dual algorithm.

Primal-dual algorithms can be used to solve the optimization problem

$$\arg \min_{x \in \mathbb{R}_+^n} f(x) \quad \text{such that} \quad y = \mathcal{H}x \quad (3.5)$$

where  $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$  is a real-valued, continuous, convex function that maps an  $n$  dimensional vector to a scalar, and the constraint on  $x$  is just the equality constraint from Eq. (1.1). The constraint that  $x \geq 0$  is necessary for maximum entropy algorithms, since entropy is defined only for positive values.

The first step in a primal-dual algorithm is to define a *dual* functional,  $g : \mathbb{R}^m \rightarrow \mathbb{R}$ , which is the minimum with respect to  $x$  of the Lagrangian:

$$g(u) = \min_{x \in \mathbb{R}^n} (f(x) + \langle u, y - \mathcal{H}x \rangle). \quad (3.6)$$

Note that the dual function,  $g$ , is a function of the dual variable  $u$  only, which is actually the vector of Lagrange multipliers.

A useful result, which is important to show convergence of primal-dual algorithm, is the weak Lagrange duality theorem [16], which guarantees that

$$\min_x (f(x) | \{x : \mathcal{H}x = y, x \geq 0\}) \geq \sup_u (g(u) | u \in \mathbb{R}^m). \quad (3.7)$$

In an iterative primal-dual algorithm, primal iterates  $\{x^k\}$  and dual iterates  $\{u^k\}$  are simultaneously generated, which in the limit as  $k$  goes to infinity converge to the solution of Eq. (3.5). Lent and Censor consider primal-dual algorithms of the following form:

**Algorithm 3 (primal-dual algorithm)**

*Initialization:*  $u^0 \in U$ .

*Iterative step.* Given  $u^k$ , solve the minimization problem (3.6) to find a minimizer  $\tilde{x}^k$ , i.e.,

$$g(u^k) = f(\tilde{x}^k) + \langle u^k, y - \mathcal{H}\tilde{x}^k \rangle. \quad (3.8)$$

Then, calculate  $u^{k+1}$  from

$$u^{k+1} = \mathcal{D}(u^k). \quad (3.9)$$

$U \subseteq \mathbb{R}^m$  is the permissible initialization set for the dual iterates, and  $\mathcal{D}$  is a *dual algorithmic operator*, which maps

$$\mathcal{D} : \mathbb{R}^m \rightarrow \mathbb{R}^m. \quad (3.10)$$

To continue we need a definition:

**Definition 2 (AF)**

(i) A sequence  $\{x^k\}_{k=0}^{\infty}$  will be called *asymptotically feasible* (AF, for short) with respect to  $\mathcal{H} \in \mathbb{R}^{m \times n}$  and  $y \in \mathbb{R}^m$ , if

$$\lim_{k \rightarrow \infty} \mathcal{H}x^k = y. \quad (3.11)$$

(ii) A dual algorithmic operator  $\mathcal{D}$  will be called *implying asymptotic feasibility* (IAF, for short) with respect to  $\mathcal{H}$ ,  $y$ , and  $f(x)$  of problem (3.5), if any sequence  $\{x^k\}_{k=0}^{\infty}$  generated by Algorithm (3) with  $\mathcal{D}$  is AF w.r.t.  $\mathcal{H}$  and  $y$ .

The workhorse theorem in this thesis used to show convergence of primal-dual algorithms is Theorem (12) from [10], which will be restated here.

**Theorem 1**

Define  $\{x^*\}$  to be

$$\arg \min_{x \in \mathbb{R}_+^n} f(x) \quad \text{such that} \quad y = \mathcal{H}x \quad (3.12)$$

and if  $\mathcal{D}$  is IAF w.r.t.  $\mathcal{H}$ ,  $y$ , and  $f$  of problem (3.5), then any sequence  $\{x^k\}_{k=0}^\infty$  generated by Algorithm (3) (the primal-dual algorithm), converges to  $\{x^*\}$ .

The reader is referred to [10] for the proof of Theorem (1) and other corresponding theorems and corollaries.

Given Theorem (1), to show convergence to the unique global minimum it is necessary that:

1. The function  $f(x)$  be convex, and
2. The dual algorithmic operator  $\mathcal{D}$  be IAF w.r.t.  $\mathcal{H}$ ,  $y$ , and  $f$ .

The following sections demonstrate how the convergence of maximum entropy algorithms, including but not limited to MART, can be proved using Theorem (1).

### 3.3 Maximum Entropy Algorithm Convergence

As stated in Section 3.1, iterative maximum entropy (ME) algorithms solve the constrained minimization problem

$$\min_x \sum_i x_i \ln x_i \quad \text{such that} \quad \mathcal{H}x = y, \quad (3.13)$$

where

$$\sum_i x_i \ln x_i \quad (3.14)$$

is the negative of the information theoretic entropy<sup>1</sup>. The dual functional  $g(u)$  is now

$$g(u) = \min_{x \in \mathbb{R}_+^n} \left( \sum_i x_i \ln x_i + \langle u, y - \mathcal{H}x \rangle \right). \quad (3.15)$$

---

<sup>1</sup>To properly form an entropy measure,  $x$  must be a probability density function. Here,  $x$  is an image intensity, so the interpretation as entropy is by analogy only.

The following sections show how Theorem 1 can be used to prove convergence of ME algorithms. Convergence is discussed for two algorithmic implementations: *normal* implementation, where each iteration uses only a single measurement, and *block* implementation, where every measurement is used for each iterative update. This development follows Lent and Censor closely (with a correction) [10], and will be extended in Sections 3.4 and 4.2.2 to prove convergence of alternative form ME algorithms and SART, respectively.

### 3.3.1 Normal Implementation

A *normal* implementation is where the estimate is affected by only one measurement per iteration. Thus, if  $k$  is the iteration and  $i$  is the measurement, then  $i$  is updated as  $i = k(\text{mod}(m)) + 1$ . Direct minimization of Eq. (3.15) with respect to  $x$  yields a minimizer  $\tilde{x}_j$ :

$$\tilde{x}_j = \exp(-1) \exp\left(\sum_{i=1}^m u_i h_{ij}\right), \quad j = 1, 2, \dots, n. \quad (3.16)$$

Putting this minimizing  $\tilde{x}_j$  into Eq. (3.15) gives the minimized form of the dual functional as

$$g(u) = -\sum_{j=1}^n \tilde{x}_j + \langle u, y \rangle. \quad (3.17)$$

Written in iterative form (i.e., as a function of the iteration  $k$ ), Eqs. (3.16) and (3.17) become

$$\tilde{x}_j^k = \exp(-1) \exp\left(\sum_{i=1}^m u_i^k h_{ij}\right), \quad j = 1, 2, \dots, n. \quad (3.18)$$

and

$$g(u^k) = -\sum_{j=1}^n \tilde{x}_j^k + \langle u^k, y \rangle. \quad (3.19)$$

Now define the dual algorithmic operator  $\mathcal{D}$  as

$$u^{k+1} = u^k + \lambda d^k e^i \quad (3.20)$$

where  $e^i$  is a vector of zeros with one at the  $i$ 'th place,  $\lambda$  is a relaxation parameter, and  $d^k$  is the fidelity term after the  $k$ th iteration. Using Eq. (3.20) in Eq. (3.18) gives

the algorithm update

$$\tilde{x}_j^{k+1} = \tilde{x}_j^k \exp(\lambda d_i^k h_{ij}) \quad (3.21)$$

where  $\lambda$  is a relaxation parameter.

Since entropy is convex, to prove convergence it is necessary only to show that  $\mathcal{D}$  is IAF w.r.t.  $\mathcal{H}$ ,  $y$ , and  $f$ . The first step taken in [10] to show that  $\mathcal{D}$  is IAF is to show that the dual functional is monotonically increasing. The following proof closely follows the proof of Proposition 16 in [10], with one modification for completeness.

**Proposition 2**

$g(u^k)$  is monotonically increasing in  $k$ .

*Proof:* For convenience, the tilde notation has been dropped from  $x$ . Consider the difference:

$$\begin{aligned} g(u^{k+1}) - g(u^k) &= \sum_{j=1}^n (x_j^k - x_j^{k+1}) + \langle u^{k+1} - u^k, y \rangle \\ &= \sum_{j=1}^n x_j^k (1 - \exp(d_i^k \lambda h_{ij})) + \langle d_i^k \lambda e^i, y \rangle \\ &= \sum_{j=1}^n x_j^k (1 - \exp(d_i^k \lambda h_{ij})) + d_i^k \lambda y_i. \end{aligned} \quad (3.22)$$

We now employ the inequality

$$1 - w^\alpha \geq \alpha(1 - w) \quad \text{for } w \geq 0, 0 \leq \alpha \leq 1. \quad (3.23)$$

Employing this relation gives

$$\begin{aligned} g(u^{k+1}) - g(u^k) &= \sum_{j=1}^n x_j^k (1 - \exp(d_i^k \lambda h_{ij})) + d_i^k \lambda y_i \\ &\geq \sum_{j=1}^n x_j^k (-\lambda h_{ij} (1 - \exp(-d_i^k))) + d_i^k \lambda y_i \\ &= -\lambda(y_i + \epsilon_i) (1 - \exp(-d_i^k)) + d_i^k \lambda y_i \\ &= \lambda y_i \underbrace{(\exp(-d_i^k) - 1 + d_i^k)}_{s(d^k)} + \underbrace{\epsilon_i^k \lambda (\exp(-d_i^k) - 1)}_{t(d^k, \epsilon^k)} \end{aligned} \quad (3.24)$$

where the error term,  $\epsilon_i^k = h_i x^k - y_i$ . The first additive term in Eq. (3.24),  $s(d^k)$ , is easily shown to be a non-negative convex function in  $d$ . Therefore

$$g(u^{k+1}) - g(u^k) \geq t(d^k, \epsilon^k). \quad (3.25)$$

It is also easily demonstrated that when the signs of  $\epsilon_i^k$  and  $d_i^k$  are opposite,  $t(d^k, \epsilon^k) \geq 0$ , so Eq. (3.25) becomes

$$g(u^{k+1}) - g(u^k) \geq 0 \quad (3.26)$$

for

$$\text{sgn}\{\epsilon^k\} = -\text{sgn}\{d^k\} \quad (3.27)$$

Thus we conclude that  $g(u^k)$  is monotonically increasing if  $d_i^k$  is chosen to have opposite sign from  $\langle h_i, x^k \rangle - y_i$ .

*q. e. d*

The proof in [10] failed to include the second additive term of Eq. (3.24),  $t(d^k, \epsilon^k)$ , which contains critical information about the form that  $d$  must have for convergence, and which will be used to develop alternative forms of the ME iteration. One fidelity term  $d$  that satisfies the condition necessary for convergence is

$$d_i = \log \left( \frac{y_i}{\langle h_i, x \rangle} \right) \quad (3.28)$$

which, when substituted into Eq. (3.21) gives the MART algorithm:

$$x_j^{k+1} = x_j^k \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^{\lambda h_{ij}} \quad (3.29)$$

More about the variety of forms that maximum entropy algorithms can have are discussed in Section 3.4.

With Prop 2 proved (which corresponds to Prop. 16 in [10]), Propositions 17 and 18 in [10] directly follow, proving  $\mathcal{D}$  to be IAF. Thus, by Theorem 1 (Theorem 12 in [10]) the algorithm is guaranteed to converge to the maximum entropy solution.

### 3.3.2 Block Implementation

The block implementation algorithm is similar to the non-block implementation, except that in the block form every measurement is used in computing each

update. The minimizing  $\tilde{x}^k$  and dual function  $g(u^k)$  are the same as in the normal implementation, given by Eqs. (3.18) and (3.19).

Now the update term for the dual variable is

$$u^{k+1} = u^k + \lambda \odot d^k, \quad (3.30)$$

where  $\odot$  represents the Schur matrix product, i.e. element by element multiplication. Notice how Eq. (3.30) differs from Eq. (3.20): every measurement is included in each update.

Substituting Eq. (3.30) into Eq. (3.18) gives the algorithm update

$$\begin{aligned} \tilde{x}_j^{k+1} &= \tilde{x}_j^k \exp \left( \sum_{i=1}^m \lambda_i d_i^k h_{ij} \right) \\ &= \tilde{x}_j^k \prod_{i=1}^m \exp(\lambda_i d_i^k h_{ij}). \end{aligned} \quad (3.31)$$

As in the normal implementation case, the first step in demonstrating that  $\mathcal{D}$  is IAF is to show that  $g(u^k)$  is monotonically increasing.

**Proposition 3**

$g(u^k)$  is monotonically increasing in  $k$ .

*Proof:*

This proof closely follows the proof given by Segman and Censor in [17], with some minor changes for completeness, similar as to what was done in Section 3.3.1. Once again the tilde notation has been dropped for convenience.

To show that the dual variable  $g(u)$  is monotonically increasing we need additional inequalities. Taking Eq. (3.23) with  $w = \exp(-v_i)$ , multiplying by  $\lambda_i$  and summing gives

$$1 - \sum_i \lambda_i \exp(-v_i^k h_{ij}) \geq - \sum_i \lambda_i h_{ij} (1 - \exp(v_i)) \quad (3.32)$$

where we assume that  $\lambda_i$  sums to one. The other relation is

$$\sum_i z_i^{r_i} \leq \sum_i r_i z_i \quad \text{if} \quad \sum_i r_i = 1. \quad (3.33)$$

Now consider the difference

$$\begin{aligned}
g(u^{k+1}) - g(u^k) &= \sum_{j+1}^n x_j^k \left[ 1 - \prod_{i=1}^m \exp(\lambda_i d_i h_{ij}) \right] + \sum_{i=1}^m y_i \lambda_i d_i \\
&\geq \sum_{j+1}^n x_j^k \left[ 1 - \sum_{i=1}^m \lambda_i \exp(d_i h_{ij}) \right] + \sum_{i=1}^m y_i \lambda_i d_i \\
&= \sum_{j+1}^n x_j^k \left[ 1 - \sum_{i=1}^m \lambda_i \exp(d_i h_{ij}) \right] + \sum_{i=1}^m y_i \lambda_i d_i \\
&\geq \sum_{j+1}^n x_j^k \sum_{i=1}^m -\lambda_i h_{ij} (1 - \exp(-d_i)) + \sum_{i=1}^m y_i \lambda_i d_i \\
&= \sum_{i=1}^m \lambda_i (y_i (\exp(-d_i) - 1 + d_i) + \epsilon_i^k (\exp(-d^k) - 1)) \\
&\geq 0
\end{aligned} \tag{3.34}$$

where  $\epsilon_i^k$  is defined as before, and we have to assume that  $\sum_i \lambda_i = 1$ .

*q. e. d*

The only difference between this proof and the proof given in [17] is the second additive term in Eq. (3.34),  $\epsilon_i^k (\exp(-d^k) - 1)$  which again gives information about the form of  $d$  necessary for convergence. Once again, this will be used to develop other forms of the ME algorithm. One can follow the rest of the convergence proof given in [17] to see that  $\mathcal{D}$  is IAF and that the block algorithm also converges to the maximum entropy solution.

### 3.4 Maximum Entropy Algorithm Forms

As noted in Sections 3.3.2 and 3.3.1, there are constraints on the form of the fidelity term  $d$  for the algorithm to converge. Namely, if  $\epsilon_i^k = h_i x^k - y_i < 0$ , which implies that  $y_i > \langle h_i, x^k \rangle$ , then the fidelity term  $d$  has to be greater than zero.

This suggests that there are many possible different fidelity terms that can be used in conjunction with the general maximum entropy algorithm. While these different algorithms converge to the same solution for a noiseless and consistent system, it is possible that some forms of  $d$  may perform better than others in the presence of noise, or that faster convergence could be achieved. The following sections explore

the behavior of the general maximum entropy with three different fidelity terms and present a comparative evaluation of each of these algorithms to determine if any such advantages can be identified.

### 3.4.1 ME Algorithm Definitions

One form for  $d$  already suggested is

$$d_i^k = \log \left( \frac{y_i}{\langle h_i, x^k \rangle} \right) \quad (3.35)$$

which results in the well-known MART algorithm, repeated here:

#### Algorithm 4 (MART)

*Initialization:*  $x^\circ = \exp(-1)$

*Iterative step:*

$$x_j^{k+1} = x_j^k \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^{\lambda_1 h_{ij}} \quad (3.36)$$

Two other possible forms of  $d$  are

$$d_i^k = y_i - \langle h_i, x^k \rangle \quad (3.37)$$

and

$$d_i^k = \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^\alpha - 1. \quad (3.38)$$

Equation (3.37) was suggested by Elfving in [18] as a possible fidelity term, and Eq. (3.38) is original. The algorithms corresponding to these two fidelity terms are

#### Algorithm 5 (Elfving MART)

*Initialization:*  $x^\circ = \exp(-1)$

*Iterative step:*

$$x_j^{k+1} = x_j^k \exp [\lambda_2 (y_i - \langle h_i, x^k \rangle) h_{ij}] \quad (3.39)$$

and

**Algorithm 6 (Willis MART)***Initialization:*  $x^\circ = \exp(-1)$ *Iterative step:*

$$x_j^{k+1} = x_j^k \exp \left\{ \lambda_3 \left[ \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^\alpha - 1 \right] h_{ij} \right\}. \quad (3.40)$$

The forms of  $d$  in Eqs. (3.37) and (3.38) both satisfy the conditions necessary for convergence as explained in Section 3.3. Hence, the ME proof of convergence in Section 3.3 is sufficient to prove that Elfving MART and Willis MART converge to the ME solution.

**3.4.2 Relaxation Factors**

While any algorithm that uses a fidelity term  $d$  with the necessary form should, in theory, converge, it was found that in practice not every form of  $d$  results in the algorithm convergence for all possible values of  $y$  and  $\mathcal{H}$ . The reason some forms do not work is because of the exponentiation in the update. For example, using

$$d_i^k = \log \left( \frac{y_i}{\langle h_i, x^k \rangle} \right) \quad (3.41)$$

results in an update that is linear with respect to the ratio  $\frac{y}{\langle h_i, x \rangle}$  (since the exponential and logarithm are inverse operations) and is perfectly antisymmetric about one (i.e.,  $f(1+x) = -f(1-x)$ ). These properties (linear and antisymmetric) make the algorithm robust.

The forms of  $d$  given by Eqs. (3.37) and (3.38) are not linear when put in the exponential. Figure 3.1 shows a plot of  $\exp(d)$  over a range of  $d$ . Notice how non-linear the update is. In running experiments with  $d_i^k = y_i - \langle h_i, x^k \rangle$ , the algorithm often “got stuck” at either very high or very low values (depending on the data range and initial condition) and not converge.

Some experimenting was required with the relaxation factors  $\lambda_2$  and  $\alpha$  to get Algorithms 5 and 6 (Elfving MART and Willis MART) to converge consistently over a wide range of data values. The exponentiation of the fidelity terms of the second two algorithms with  $\lambda_2 = 0.05$  and  $\alpha = 0.4$  is plotted in Figure 3.2 . With these

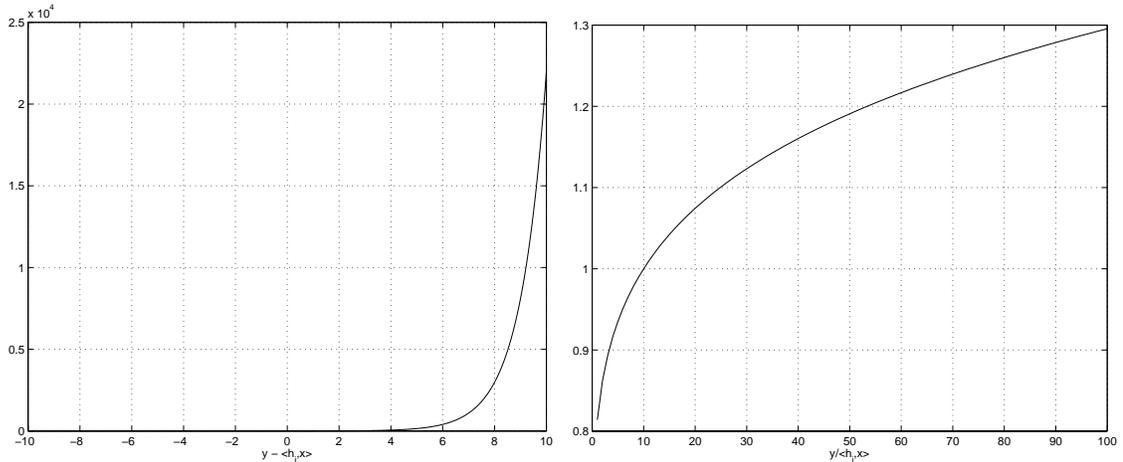


Figure 3.1: Plot of  $\exp(y_i - \langle h_i, x^k \rangle)$  for  $10 < y_i - \langle h_i, x^k \rangle < 10$  (left) and  $\exp\left(\left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1\right)$  for  $0 < \left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1 < 10$  (left).

relaxation factors the exponentiation of the fidelity terms are more linear over a larger range of  $d$ . In the actual simulation each of the updates are again multiplied by an additional relaxation factor  $\lambda = 0.5$ , such that the effective relaxation factors used in the following examples, unless stated otherwise, are  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.025$ ,  $\lambda_3 = 0.5$ , and  $\alpha = 0.4$ .

### 3.5 Comparison of the Different ME Algorithms

The properties of Algorithms 4, 5, and 6 (MART, Elfving MART, and Willis MART, respectively) are explored in the following sections to evaluate the algorithms' performance in a noisy system.

#### 3.5.1 Fully Determined System

In a consistent, fully determined system (i.e.  $\mathcal{H}$  is square and full rank) all three algorithms converge to the unique, correct solution.

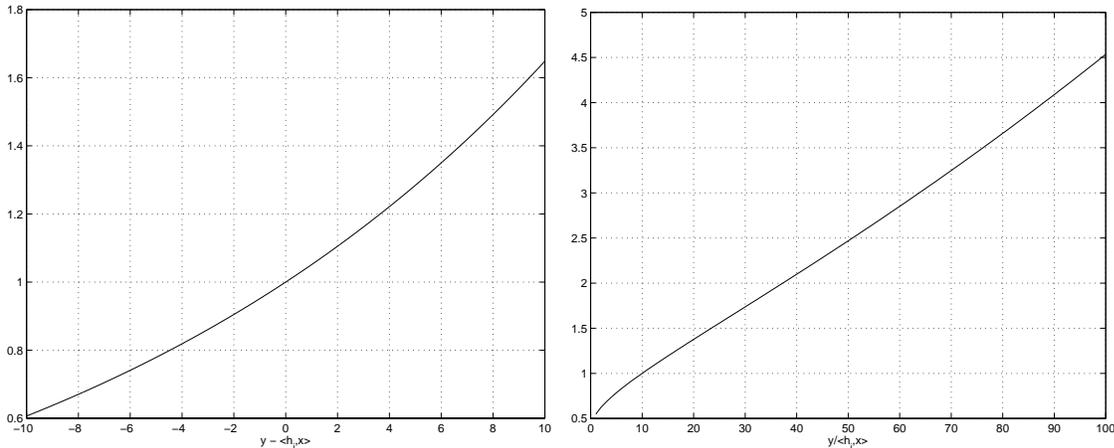


Figure 3.2: Plot of  $\exp((y_i - \langle h_i, x^k \rangle) * 0.05)$  for  $10 < y_i - \langle h_i, x^k \rangle < 10$  (left) and  $\exp\left(\left(\frac{y_i}{\langle h_i, x^k \rangle}\right)^{0.4} - 1\right)$  for  $0 < \left(\frac{y_i}{\langle h_i, x^k \rangle}\right) - 1 < 10$  (left).

Figure 3.3 shows plots of the trajectories of the three algorithms for three-dimensional fully determined noiseless and noisy cases. For this experiment

$$\mathcal{H} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{bmatrix} \quad \text{and} \quad x = [0.2 \ 0.4 \ 0.5]^T. \quad (3.42)$$

As expected, in the noise-free case all of the algorithms converge to the correct solution.

The more interesting case is where noise has been added to the measurements. For the noisy, consistent case (i.e. when the noise lies in the column space of  $\mathcal{H}$ ), the algorithm will converge, but not to the true solution. In this case, the solution at convergence (i.e. with infinite solutions) is usually not the closest iterate in the sequence to the desired solution. A better solution can often be formed by truncating the iterations prior to convergence. This truncation can actually regularize the solution to reduce noise amplification effects in the inverse problem, and is used commonly in iterative restoration [19]. For the fully determined case the “best” result is the iteration which is closest to the true solution in a squared error sense. Though

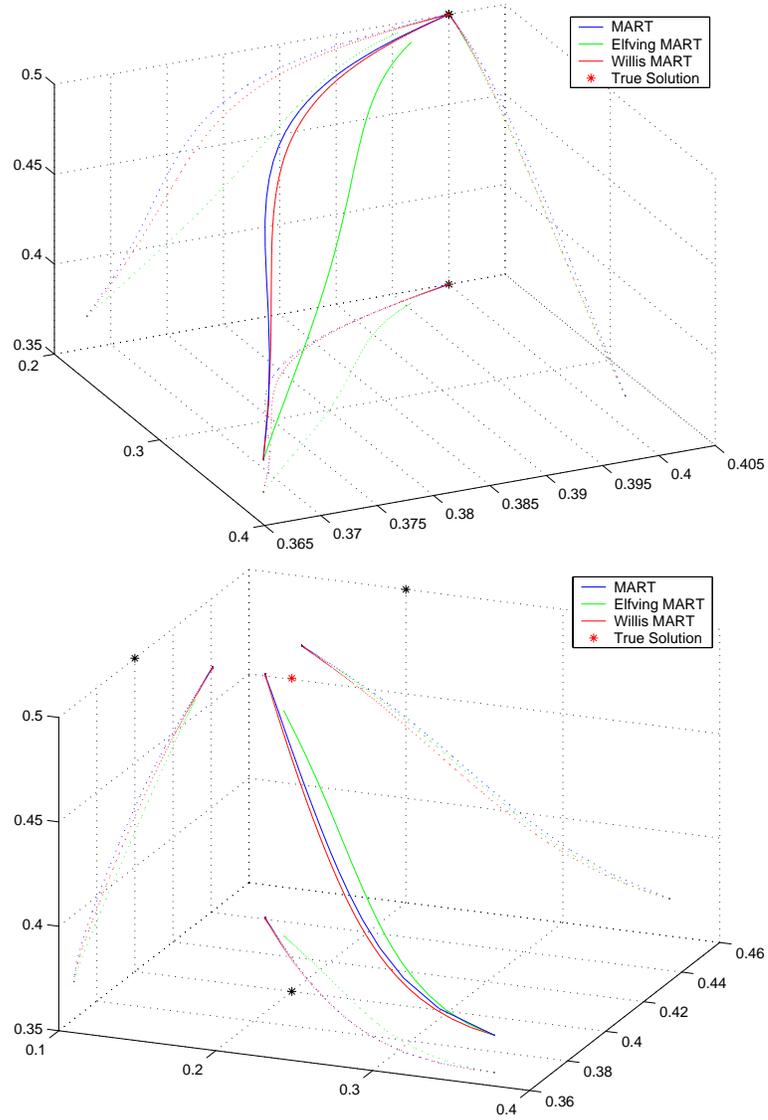


Figure 3.3: Convergence of the three algorithms with a fully determined system—no noise (top) and noisy (bottom) with  $\text{SNR} = 22.2$  dB. The actual trajectories are in solid lines (MART—blue, Elfving MART—green, Willis MART—red) and their projections onto the  $x$ ,  $y$ , and  $z$  planes are dotted.

it is impossible to know which iterate represents the “best” solution without having the true solution to compare with, we present an analysis here based on this distance measure to quantify the best possible performance achievable by each candidate algorithm to make comparative evaluations. In practice, operators commonly intervene and stop the iterations at the image deemed to be of highest perceived quality. This is usually very close to the “best” solution. Geometrically, in terms of Figure 3.3, it is the point on the trajectory of solutions that is closest (in Euclidean distance) to the true solution.

There are actually two errors of interest in these experiments:

**Solution Error** The error that has just been described—the distance between the current estimate and the true solution.

**Forward Projection Error** The squared error between the forward projection  $\mathcal{H}x^k$  and the actual measurements  $y$ .

The solution error as a function of iteration is plotted for the three algorithms in Figure 3.4, and the results of the noisy experiment are summarized in Table 3.1. Because of its small relaxation parameter, Elfving MART converges much slower than the other algorithms. Interestingly, although the three algorithms reach the minimum solution error at different numbers of iterations, the minimum solution error is very similar for the three algorithms. The forward projection errors at the iteration which minimizes the solution error is also very close for the three algorithms.

### Higher-Dimensional Examples

For the higher-dimensional cases  $\mathcal{H}$  was chosen to be a circular convolution matrix. Because circular convolution is a shift invariant linear operation, it can be expressed in matrix form, where the rows of the circular convolution matrix are made up of shifted versions of a blurring function. In this case, the blurring function, or footprint, was chosen so that  $\mathcal{H}$  is full rank. Figure 3.5 shows a synthetic image example used to compare algorithm performance. The results for higher-dimensional simulations using images are similar to the three-dimensional case. Like before, the

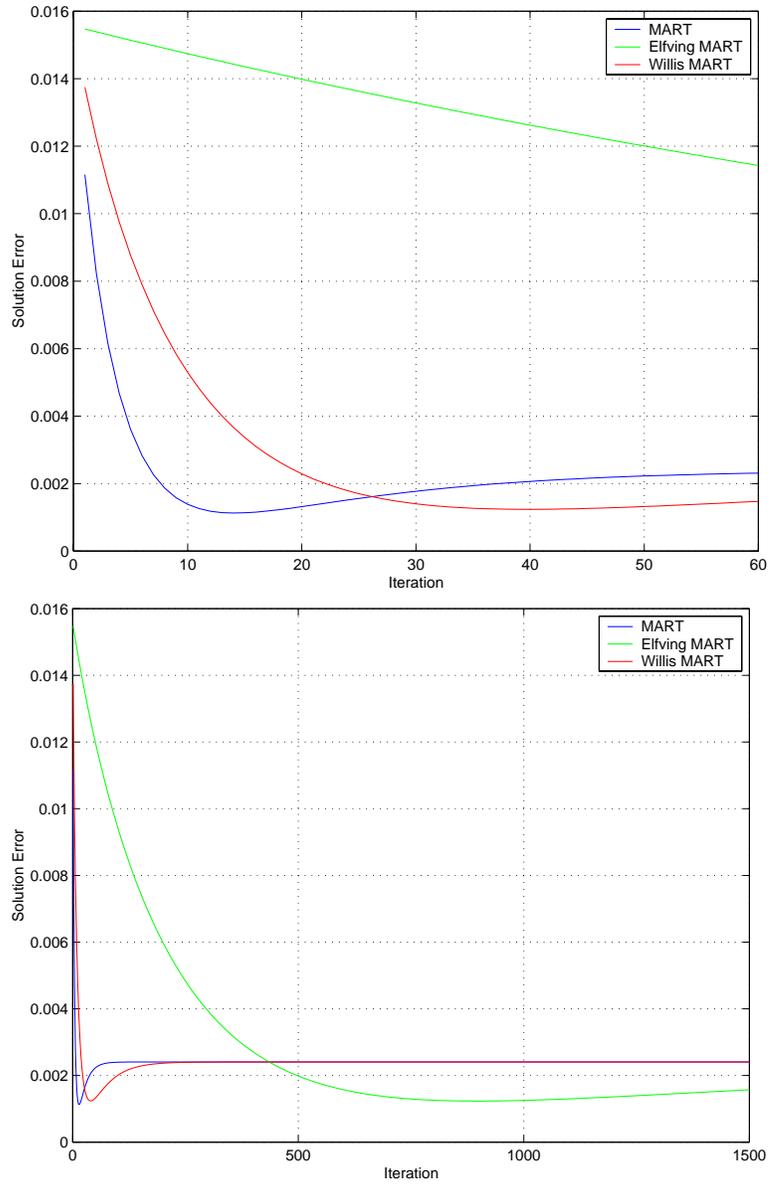


Figure 3.4: Solution error, shown at two different scales for clarity, of the various ME algorithms as a function of iteration for the fully determined, noisy case, SNR=22.2 dB. (top) Detail scale, 0–60 iterations, shows minimum error point is the same for both MART and Willis MART. (bottom) large scale, 0–1500 iterations—shows minimum error point for Elfving MART.

minimum solution errors are the same, and there is little difference between the forward projection errors. Referring to Figure 3.5, there seems to be little qualitative difference between the algorithms as well. Hence we conclude that, at least for the consistent, fully determined case, all of the algorithms have similar “best” solutions.

### 3.5.2 Underdetermined System

It is sometimes more realistic to model the remote sensing problem as an underdetermined system, since there are generally fewer measurements than pixels. For the three-dimensional case used in the following examples the parameters of the system are

$$\mathcal{H} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix} \quad \text{and} \quad x = [0.2 \ 0.4 \ 0.5]^T. \quad (3.43)$$

Since the system is underdetermined there is no unique solution, making the concept of a “solution error” a bit more nebulous. The forward projection error is defined as before, while the “maximum entropy error” takes the place of the solution error. The maximum entropy error is the squared error between the current estimate and the true maximum entropy solution which would have been found in the absence of observation noise.

Figure 3.6 shows the convergence of the three algorithms for the three-dimensional noiseless case. Because there are three pixels in the solution and only two measurements, there is one degree of freedom which forms a one-dimensional solution space, represented by the solid black line. Any solution along this line will satisfy the constraint  $y = \mathcal{H}x$ . Adding the maximum entropy criterion results in the unique solution, represented by the black asterisk. To aid in visualizing the algorithms’ progress in three-space the trajectories and solution space have been projected onto the  $(x, y)$ ,  $(y, z)$ , and  $(x, z)$  planes, represented by the dots.

Although the three algorithms take different paths, they all arrive at the maximum entropy solution. The biggest difference between the algorithms, as before, is the rate of convergence, with Algorithm 4 (MART) converging the fastest.

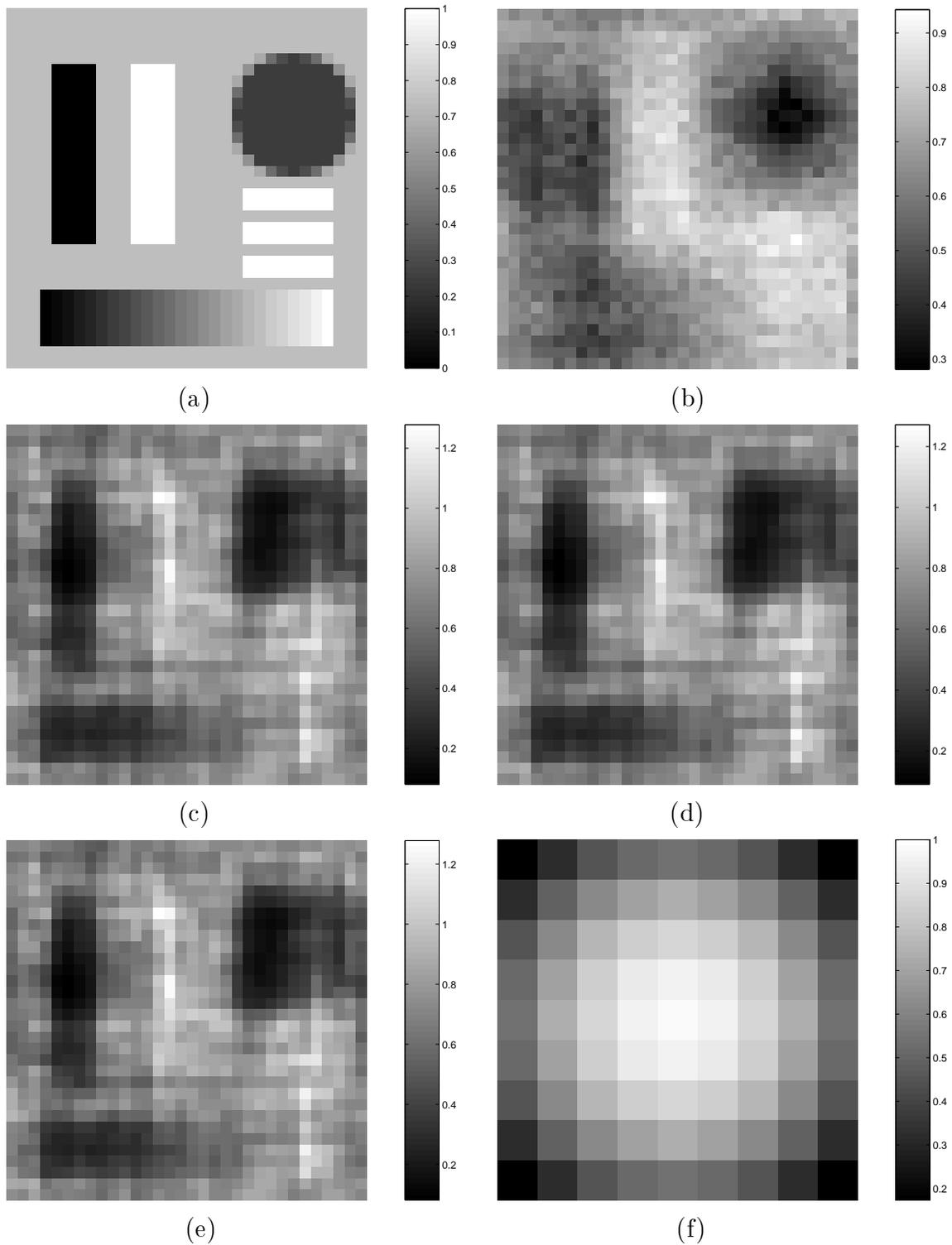


Figure 3.5: Reconstruction of the noisy, fully determined system: (a) Synthetic original image (b) Blurred image with 22.5 dB SNR noise added to the measurements (c) Algorithm 4 (MART) reconstruction (d) Algorithm 5 (Elfving MART) reconstruction (e) Algorithm 6 (Willis MART) reconstruction (f) The blurring function used to build the circular convolution matrix  $\mathcal{H}$

Table 3.1: Results of the noisy fully determined case, SNR=22.2 dB, where  $k_{\min}$  is the iteration at which the solution error is minimized.

Algorithm	Minimum Solution Error	$k_{\min}$	Forward Projection Error at $k_{\min}$
MART	.011	14	3.014e-4
Elfving MART	.012	897	2.736e-4
Willis MART	.012	40	2.659e-4

Table 3.2: Results of the higher-dimensional noisy fully determined case, SNR=22.5 dB, where  $k_{\min}$  is the iteration at which the solution error is minimized.

Algorithm	Minimum Solution Error	$k_{\min}$	Forward Projection Error at $k_{\min}$
MART	.0239	104	7.217e-4
Elfving MART	.0239	3000	7.336e-4
Willis MART	.0239	261	7.208e-4

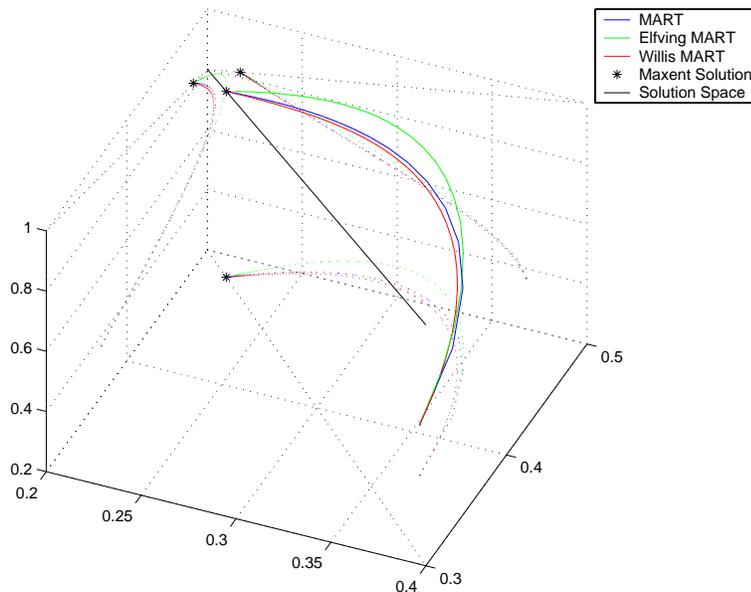


Figure 3.6: Trajectories of the three algorithms for a noiseless, underdetermined case.

Typical results for the three-dimensional experiments with additive white noise to give 11.4 dB and 14.5 dB SNR measurements are shown in Figure 3.7 and summarized in Table 3.3.

Table 3.3: Typical Results for underdetermined, noisy reconstruction at 11.4 dB and 14.5 dB SNR.

Algorithm	Minimum Solution Error	$k_{\min}$	Forward Projection Error at $k_{\min}$
11.4 dB SNR			
MART	6.53e-4	3	.0153
Elfving MART	1.12e-4	207	.0138
Willis MART	2.84e-4	9	.0140
14.5 dB SNR			
MART	5.42e-4	N/A	N/A
Elfving MART	6.62e-4	N/A	N/A
Willis MART	7.14e-4	N/A	N/A

The graphs and table indicate that all three algorithms follow similar trajectories, and have comparable errors. As in the three-dimensional fully determined case, no algorithm is more desirable as far as noise suppression.

There is an important caveat when using this three-dimensional example to study the effect of noise in image reconstruction. Figure 3.8 plots the singular values for the three dimensional  $\mathcal{H}$  in Eq. (3.43) and also the  $\mathcal{H}$  used for the higher-dimensional decimated experiments. The ratio of the smallest to largest singular values for  $\mathcal{H}$  in the three-dimensional experiment is 1.73, while the ratio of the smallest to largest singular values for the higher-dimensional case is 40.5, 23 times bigger. In an iterative reconstruction algorithm, the amount of noise amplification is a function of the range of singular values, with a larger span of singular values giving greater noise amplification [19]. A remote sensing application is likely to have a singular value span larger than the span of singular values for the three-dimensional case, so noise amplification will likely be more severe.

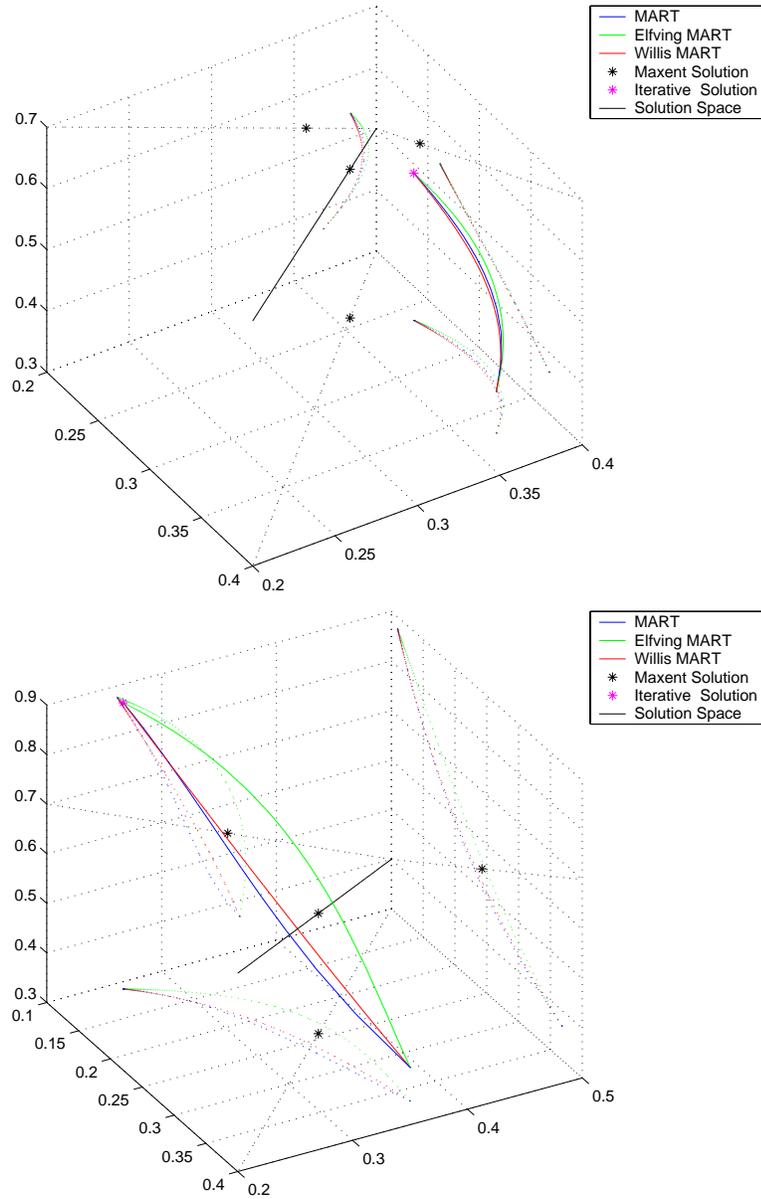


Figure 3.7: Results of noisy underdetermined system at 14.5 dB SNR (top) and 11.4 dB SNR (bottom).

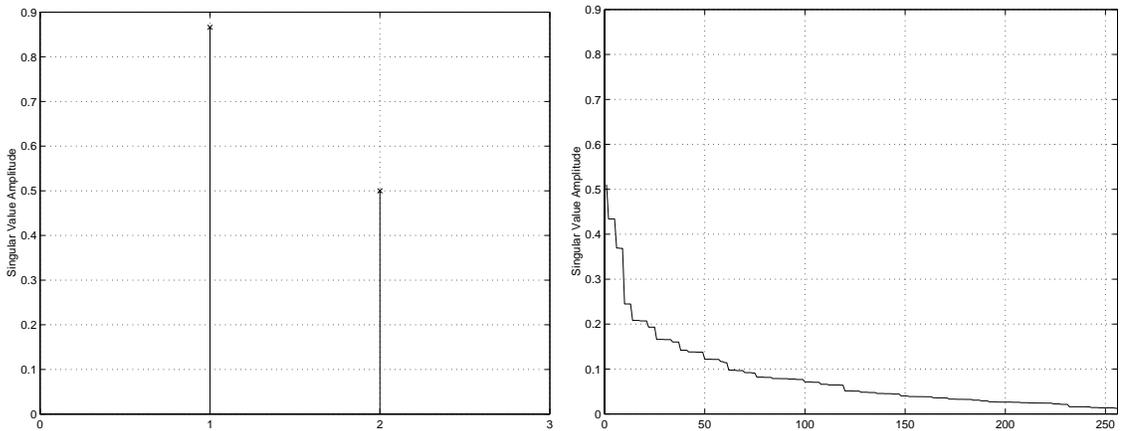


Figure 3.8: Singular values for the three-dimensional  $\mathcal{H}$  (left) and the higher-dimensional  $\mathcal{H}$  (right).

Table 3.4: Results of the higher-dimensional underdetermined case with 26.9 dB SNR, where  $k_{\min}$  is the iteration at which the solution error is minimized.

Algorithm	Minimum Solution Error	$k_{\min}$	Forward Projection Error at $k_{\min}$
MART	.0378	177	5.24e-4
Elfving MART	.0381	495	5.22e-4
Willis MART	.0385	5915	5.07e-4

Because of noise amplification, at some point it is necessary to truncate iterations before convergence. In [19], it is proved that truncating iterations is a valid regularization for suppressing noise amplification, the point at which to truncate the iterations depending on the singular values of  $\mathcal{H}$ .

### Higher-Dimensional Underdetermined System

The results of the higher-dimensional underdetermined case with white noise added to give a 26.9 dB SNR is shown in Figure 3.9 and summarized in Table 3.4.

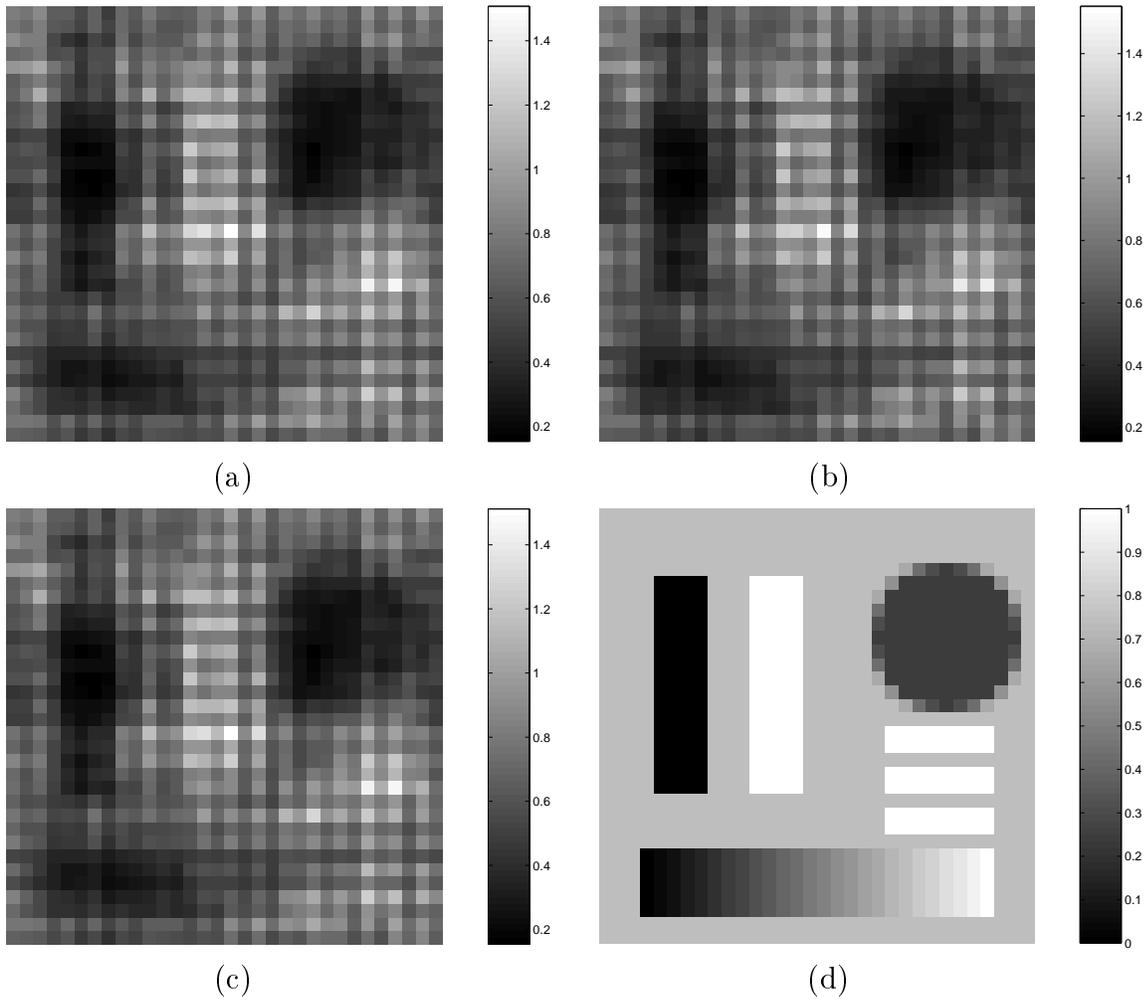


Figure 3.9: Underdetermined 26.9 dB SNR noisy reconstructions of the synthetic truth image (d) for Algorithms (a) 4 (MART), (b) 5 (Elfving MART), and (c) 6 (Willis MART).

The results are similar to the previous cases, with no one algorithm showing much better noise suppression performance than any other, but with MART having faster convergence.

### 3.5.3 Inconsistent and Overdetermined Systems

So far we have considered only the case where  $\mathcal{H}$  has been full row rank, so that adding noise to the measurements does not make the system inconsistent. Two other important cases are when  $\mathcal{H}$  is not full row rank and when  $\mathcal{H}$  is overdetermined but full column rank, both of which are possible remote sensing situations.

One difference between the behavior of the algorithms with an inconsistent system is the asymptotic behavior for many iterations. A consistent system will have all non-zero left singular values, and since the noise amplification is a function of the singular value spread there will be some bound on the noise amplification. An inconsistent system, however, will have singular values that are zero, creating an infinite singular value spread. Thus the iterative algorithms will infinitely amplify the noise in the limit. In practice, an underdetermined inconsistent system results when there are fewer measurements than pixels and some measurements are linear combinations of other measurements.

An overdetermined system, which is becoming more the case with the new scatterometers, such as QSCAT, has all non-zero left singular values, so it will converge to a finite solution in the limit of iterations. In either case, when there is measurement noise, the iterations can be interrupted before convergence to avoid noise amplification.

We are interested in the behavior of the algorithms when the maximum entropy error is minimized. The results of a typical simulation with an inconsistent, overdetermined system are shown in Table 3.5 and Fig. 3.10. As with the underdetermined and consistent cases, the only marked difference between the algorithms is the convergence rate.

Table 3.5: Results of the higher-dimensional noisy over determined, inconsistent case, SNR=22.5 dB, where  $k_{\min}$  is the iteration at which the solution error is minimized.

Algorithm	Minimum Solution Error	$k_{\min}$	Forward Projection Error at $k_{\min}$
MART	0.0107	541	$8.17e-5$
Elfving MART	0.0107	11000	$8.17e-5$
Willis MART	0.0106	1356	$8.17e-5$

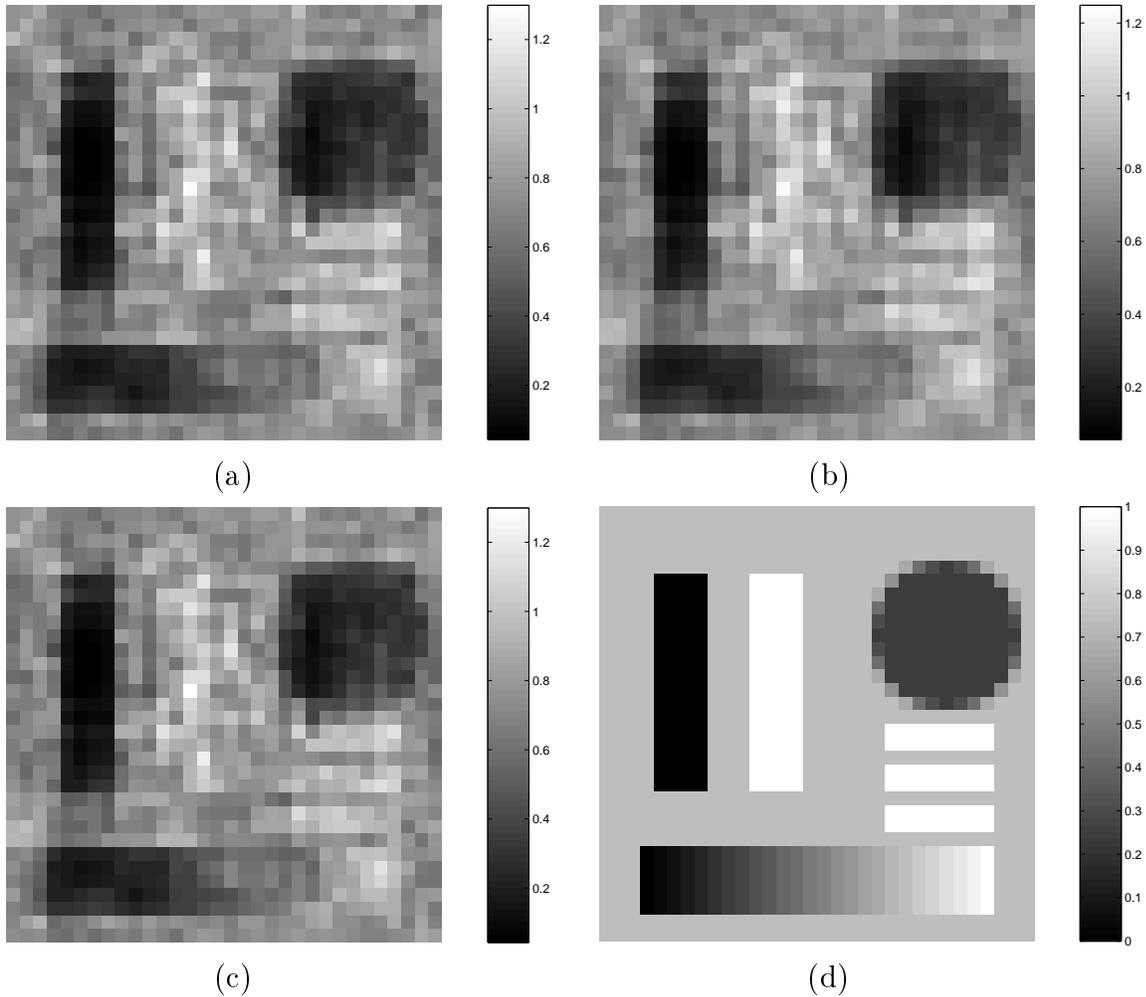


Figure 3.10: Overdetermined 22.5 dB SNR noisy reconstructions of the synthetic truth image (d) for Algorithms (a) 4 (MART), (b) 5 (Elfving MART), and (c) 6 (Willis MART).

The final conclusion, then, is that while no algorithm appears to perform better in the presence of noise, Algorithm 4 (MART), is the best algorithm to use, since it has the fastest stable convergence over a wider range of data.

### 3.6 A Closed-form ME Approximation Algorithm

An advantage of minimum norm reconstruction is the existence of a closed-form solution, the Moore-Penrose pseudo-inverse. One of our research goals was to find a closed-form ME solution that could be faster than iterative techniques. While our study into closed-form ME algorithms did not yield an exact ME solution algorithm, it did lead to a fast, closed-form approximation of the ME algorithm. The closed-form ME approximation is derived by projecting the solution onto the null space of  $\mathcal{H}$  to formulate an unconstrained optimization problem of lower dimension. A finite series expansion of the entropy expression leads to an algebraic solution to Eq. (3.3). This approach will be useful in theoretical analysis of the ME problem and is dramatically faster than MART.

All inverse problems with linear equality constraints  $y = Hx$  have admissible solutions of the form  $x = x_o + x_e$ , where  $x_o$  belongs to the row space of  $H$ , (i.e.  $x_o \in \Re\{H^T\}$ ) and  $x_e$  is drawn from the right null space of  $H$ , (i.e.  $x_e \in \mathcal{N}\{H^T\}$ ).  $x_o$  is the unique minimum norm solution (known as the minimum norm solution in the presence of measurement error) given by

$$x_o = H^\dagger y \tag{3.44}$$

where  $^\dagger$  indicates pseudo inverse. On the other hand,  $x_e$  is different for each optimization criterion that may be chosen, for example maximum entropy, minimum norm, minimum  $l_p$  norm, etc.  $x_e$  is typically a small perturbation from  $x_o$ , but as long as  $x_e \in \mathcal{N}\{H^T\}$ ,  $x$  will satisfy the constraint equation. Selecting an optimization criterion in effect determines  $x_e$  and thus determines a particular unique solution for  $x$ .

The approach of the proposed algorithm is to first find  $x_o$  and then to perturb this solution in the direction of the  $x_e$  given by the entropy criterion.  $x_o$  may be

computed using one of several standard algorithms. We can decompose  $\mathcal{H}$  into its range and null spaces using the SVD,

$$\mathcal{H} = [U_R|U_N] \Sigma [V_R|V_N]^H \quad (3.45)$$

where superscript  $H$  denotes conjugate transpose and  $\Sigma$  is the diagonal matrix of singular values.  $U_R$  and  $V_R$  are partitions of left and right singular matrices  $U$  and  $V$  respectively, which correspond to the non-zero singular values of  $H$ . Likewise,  $U_N$  and  $V_N$  contain the singular vectors corresponding to the zero singular values.  $U_N$  and  $V_N$  span the left and right null spaces of  $H$  respectively.

By construction,  $\mathcal{N}\{H^T\} = V_N$ . Therefore  $x_e = V_N z$  for some  $z$ . The vector  $z$  that leads to a maximum entropy solution of Eq. (3.3) is

$$z = \arg \max_z \left\{ - \sum_{i=1}^M (x_{oi} + [V_N z]_i) \ln(x_{oi} + [V_N z]_i) \right\}, \quad (3.46)$$

and the ME solution for  $x$  is simply

$$x_{ME1} = x_o + V_N z. \quad (3.47)$$

Note that this optimization is unconstrained in  $z$ . Also,  $z$  is length  $p = N - \text{rank}\{\mathcal{H}\} < N$ , so expressing  $x_{ME1}$  in terms of  $z$  dramatically reduces both the complexity of the minimization and the number of parameters to be estimated. Consider the entropy expression from the right hand side of equation (3.46). Using a finite series expansion approximation yields

$$\begin{aligned} \mathcal{E}(x) &= - \sum_i (x_{oi} + x_{ei}) \ln(x_{oi} + x_{ei}) \\ &\approx - \sum_i \left( x_{oi} \ln x_{oi} + (1 + \ln x_{oi}) x_{ei} + \frac{x_{ei}^2}{x_{oi}} \right) \\ &\approx -(c + r^H x_e + x_e^H B x_e) \\ &\approx -(c + r^H V_N z + z^H V_N^H B V_N z). \end{aligned} \quad (3.48)$$

where the approximation comes from taking the first term of the Taylor series expansion of  $\ln(1 + \frac{x_{ei}}{x_{oi}})$ .  $r = [1 + \ln x_{o1}, \dots, 1 + \ln x_{on}]^T$ ,  $B = \text{Diag}\{[x_{o1}^{-1}, \dots, x_{on}^{-1}]\}$ , and

scalar  $c = \sum_i x_{oi} \ln x_{oi}$ . To maximize entropy we take the derivative with respect to  $z$  and set it to zero, which yields

$$V_N^H r + 2V_N^H B V_N z = 0. \quad (3.49)$$

Therefore, noting that  $V_N^H B V_N$  is a full rank square matrix, equations (3.49) may be solved algebraically for  $z$ , which is substituted into equation (3.47) to yield our final closed form approximation

$$x_{ME1} \approx H^\dagger y - \frac{1}{2} V_N (V_N^H B V_N)^{-1} V_N^H r. \quad (3.50)$$

We note that  $B$  and  $r$  are direct functions of  $y$  through  $x_o$ , and  $V_N$  is a direct function of  $\mathcal{H}$ . The required matrix inverse is on a relatively small  $p \times p$  matrix. The most significant computation is the singular value decomposition used to form  $V_N$ .

### 3.6.1 A Fast Closed Form Implementation

If the blur represented by  $\mathcal{H}$  is due to circular convolution, the singular value decomposition of  $\mathcal{H}$  becomes trivial using the 2-D FFT, and an extremely fast implementation of equation (3.50) is possible. For non-circular FIR convolutional blur, the image frame can be extended using a zero fill border with a width equal to the psf region of support. The conventional convolution can then be exactly embedded in a slightly larger circular convolution representation. In this way, some practical restoration problems can use the following fast implementation. However, this restriction to regular sampling and shift invariant psf does not make it useful in many earth remote sensing problems.

For 2-D circular convolution,  $\mathcal{H}$  is  $N \times N$  doubly block circulant and is therefore diagonalized by the 2-D unitary DFT. Thus the SVD of equation (3.45) is given by  $\mathcal{H} = \mathcal{F}^H \Sigma \mathcal{F}$ , where  $\mathcal{F} = \frac{1}{\sqrt{N}} F \otimes F$ ,  $F$  is the 1-D DFT unitary transform matrix,  $\otimes$  indicates Kronecker matrix product, and  $\Sigma = \text{Diag}\{\text{vec}\{\text{FFT}_{2D}\{h[m, n]\}\}\}$  [20]. Here the elements of  $\Sigma$  are not ordered by magnitude. Subspace partitioning of  $\mathcal{H}$  as in (3.45) yields

$$V_N = \mathcal{F} P^T, \quad V_R = \mathcal{F} Q^T,$$

$$P = \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ & & 1 & 0 & 0 \\ & \cdots & 0 & 0 & 1 & 0 & \cdots \\ \vdots & & & & \ddots & & 0 \\ 0 & & & & \cdots & 0 & 1 \end{bmatrix}. \quad (3.51)$$

$P$  is a  $p \times N$  selection matrix formed by deleting all rows from  $I$  which correspond to non-zero singular values in  $\Sigma$ .  $Q$  is  $(N - p) \times N$  and contains the rows of  $I$  not in  $P$ , such that  $P^T P + Q^T Q = I$ . Equation (3.51) alone yields tremendous efficiencies in computing  $V_N$  and  $x_o$ .  $\mathcal{F}$  is known,  $P$  and  $Q$  are constructed easily by thresholding frequency bin magnitudes in  $\text{FFT}_{2D}\{h[m, n]\}$  to identify zero singular values, and  $x_o = \mathcal{F}^H Q^T (Q \Sigma Q^T)^{-1} Q \mathcal{F} y$ .

Additional computational savings come from recognizing that all products involving  $V_N^H$  in equation (3.50) are simply 2-D FFTs followed by masking out frequency bins corresponding to non-zero singular values. Thus for an arbitrary vector  $g = \text{vec}\{g[m, n]\}$ ,  $P \mathcal{F} g$  can be interpreted as operator notation for  $\text{MASK}_P\{\text{FFT}_{2D}\{g[m, n]\}\}$ . Likewise,  $\mathcal{F}^H P^H g \Omega$  IFFT $_{2D}\{\text{MASK}_P\{g[m, n]\}\}$ . Using FFTs also eliminates the need to store huge  $N \times N$  matrices by operating directly on the original  $\sqrt{N} \times \sqrt{N}$  images. Without this reduction, the problem is completely intractable for even modest sized images of  $256 \times 256$  pixels.

The only remaining computational difficulty in equation (3.50) is the matrix inverse,  $(V_N^H B V_N)^{-1}$ , which cannot be computed directly in the frequency domain. Instead, we find  $z$  with a steepest descent least-squares solution to equation (3.49). The following algorithm efficiently approximates  $z$  in about five iterations using FFTs and no matrices larger than the original images.

1.  $x_o = \mathcal{F}^H Q^T (Q \Sigma Q^T)^{-1} Q \mathcal{F} y$
2.  $z_{k=0} = -\frac{1}{2} P \mathcal{F} [\text{Diag}\{x_o\}] \mathcal{F}^H P^T P \mathcal{F} r$
3.  $z_{k+1} = z_k - \mu P \mathcal{F} B \mathcal{F}^H P^T P \mathcal{F} (2 B \mathcal{F}^H P^T z_k + r)$

4.  $k = k + 1$ . If  $k \leq K$  go to 3

5.  $x_{ME1} = x_o + \mathcal{F}^H P^T z_k$

where  $\mu$  is the iteration step size, and  $K \approx 5$  is the desired number of iterations.

### 3.6.2 Experimental Results

Figure 3.11 illustrates an example using the closed-form ME approximation. The original Hubbel Space Telescope image of planetary nebula NGC 6543 was circularly convolved with a low pass filter to produce the output image. Comparing the true maximum entropy solution as produced by MART with the closed form approximation shows striking similarity. The only apparent difference is an increased low level ringing in Figure 3.11d, which is difficult to see in this reproduction. This suggests the algorithm will be promising for a variety of ME image restoration applications. It is noteworthy that the closed form result was computed on a 400 MHz PC in approximately 40 seconds, while the MART computation required 24 hours and is still not fully converged.

### 3.7 Problems with ART and MART

One problem with row-action algorithms like ART and MART is manifest in Figure 3.9. The dark criss crossing reconstruction artifacts are a result of under-sampling. ART and MART don't appear to suppress sampling artifacts well when the image is undersampled. Artifact noise can be severe in some remote sensing applications where the sampling is neither very dense nor uniform, such as with SSM/I radiometer data. This problem has led to the use of column normalized algorithms in remote sensing reconstruction. Column normalized algorithms will be discussed in Chapter 4.

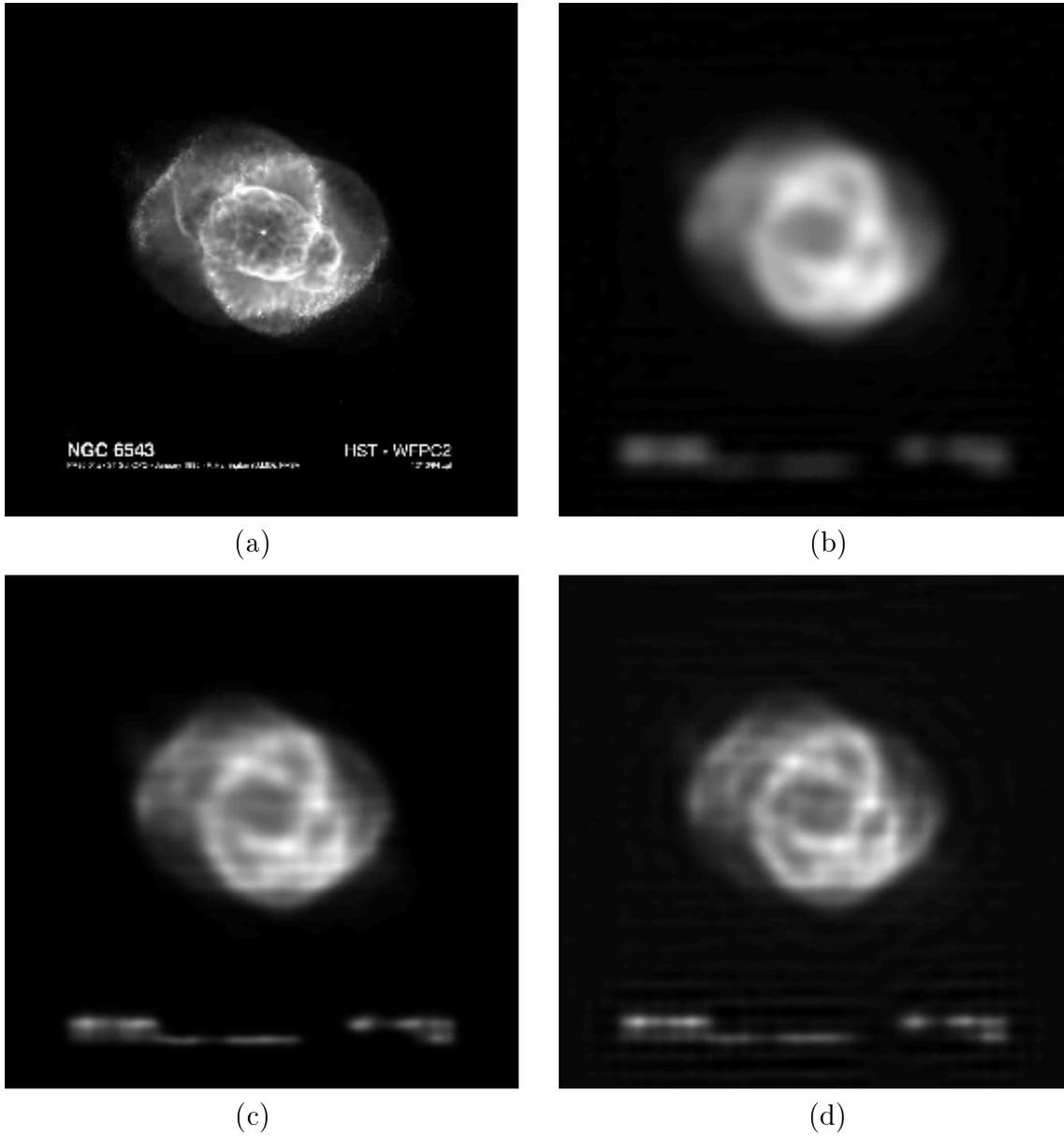


Figure 3.11: Example results from the closed form maximum entropy algorithm. a) Original  $256 \times 256$  pixel true image. b) Observed image, blurred with a circular psf low pass filter with cutoff frequency less than Nyquist / 4. No noise was added. c) Maximum entropy reconstruction using MART. d) Closed form approximation of maximum entropy.

## Chapter 4

### Column-normalized Algorithms

Chapter 3 studied row action algorithms, with an emphasis on the MART algorithm. Section 3.7 discussed some of the weaknesses of row action algorithms for an underdetermined system. This chapter will discuss column-normalized algorithms, which have good sampling artifact suppression.

#### 4.1 Column-normalized Algorithms

A *column-normalized* algorithm normalizes each update  $x_j$  by the sum of the elements in the  $j$ th column of  $\mathcal{H}$ . The  $j$ th column of  $\mathcal{H}$  contains important sampling information, including the number of observation samples  $y_i$  influenced by the  $j$ th pixel in  $x$ , and the relative weighting given to that pixel in the  $i$ th sample. Column-normalized algorithms differ from row-action algorithms in that a single pixel update needs information from every measurement. However, implementation of column-normalized algorithms require memory access to only one row of  $\mathcal{H}$  at a time, making it practical for remote sensing reconstruction where very large transfer matrices cannot be loaded into memory in their entirety. Section 3.7 pointed out some of the difficulties that standard row-action methods have with reconstruction artifacts. The following section studies in detail the SART algorithm, giving insight into how column-normalized algorithms can reduce reconstruction artifacts.

#### 4.2 Simultaneous Algebraic Reconstruction Technique (SART)

The Simultaneous Algebraic Reconstruction Technique (SART) algorithm was proposed for Computerized Tomography reconstruction by Andersen and Kak in 1984

(see [11], [21]) and was found to be useful in reducing imaging artifacts. The form of SART is similar to ART and can be considered a block ART algorithm. The iteration is as follows:

**Algorithm 7 (SART)**

*Initialization:*  $x^\circ = 0$ .

*Iterative step*

$$x_j^{k+1} = x_j^k + \frac{\sum_{i=1}^m h_{ij} d_i^k}{\sum_{i=1}^m h_{ij}} \quad (4.1)$$

where

$$d_i^k = y_i - \langle h_i, x^k \rangle. \quad (4.2)$$

It is sometimes more helpful for analysis to put the algorithm in matrix form, thus Eq. (4.1) becomes

$$x_j^{k+1} = x_j^k + \frac{\mathcal{H}_j^T d^k}{\mathcal{H}_j^T \mathbf{1}}. \quad (4.3)$$

where  $\mathcal{H}_j$  is the  $j$ th column of  $\mathcal{H}$ , and  $\mathbf{1}$  is a vector of ones. The  $\mathcal{H}_j^T \mathbf{1}$  term serves to normalize the update relative to the total weighting from all the observations.

An important aspect of an iterative algorithm is its convergence properties. To the author’s knowledge, no work has been published that has shown SART’s convergence properties—neither a proof of convergence nor what type of solution it converges to. One way to show convergence is to cast the algorithm as a primal-dual algorithm as was done for ME algorithms in Section 3.3 and apply Theorem 1. However, to use this approach it is first necessary to know what type of solution the algorithm converges to.

### 4.2.1 Finding the Solution

To determine the solution to which an algorithm converges we must return to the constrained optimization problem

$$\min_x f(x) \quad \text{such that} \quad \mathcal{H}x = y. \quad (4.4)$$

The first step in solving the constrained optimization problem is to find the minimum of the Lagrangian,

$$g(u) = \min_{x \in \mathbb{R}^n} (f(x) + \langle u, y - \mathcal{H}x \rangle). \quad (4.5)$$

Setting the derivative of the Lagrangian to zero and keeping the equality constraint results in the Kuhn-Tucker conditions:

$$\nabla f(x_j) = H_j^T u \quad \text{and} \quad \mathcal{H}x = y. \quad (4.6)$$

The derivative relation in Eq. (4.6) can be used to “work backwards” from the algorithm to find the  $f(x)$  that is being minimized.

### The MART Solution

As an example of how one can “work backwards” to find the minimizing function, we will show how one can start at the block MART algorithm and work back to the maximum entropy constraint. The block MART is updated as

$$x_j^{k+1} = x_j^k \prod_{i=1}^m \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^{h_{i,j} \lambda} \quad (4.7)$$

where each iteration calculates the update due every measurement. The  $\lambda$  term is simply an arbitrary relaxation factor and will be set to  $\lambda = 1$  for the rest of the analysis without loss of generality. MART can more generally be expressed in matrix form as

$$x_j^{k+1} = x_j^k \exp(\mathcal{H}_j^T d^k) \quad (4.8)$$

where

$$d_i^k = \log \left( \frac{y_i}{\langle h_i, x^k \rangle} \right) \quad (4.9)$$

and  $\mathcal{H}_j$  is the  $j$ th column of  $\mathcal{H}$ .

Now some assumption must be made regarding the form of the dual algebraic operator  $\mathcal{D}(u^k)$ ,  $\mathcal{D} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ . For MART, it is assumed that the dual variable is updated by

$$u^{k+1} = u^k + d^k. \quad (4.10)$$

Equation (4.10) is not an unreasonable assumption, since it satisfies the condition that if  $d^k$  goes to zero, then  $u$  goes to zero, and the  $x$  updates stop changing. Substituting Eq. (4.10) into Eq. (4.8) gives

$$x_j^{k+1} = x_j^k \exp(H_j^T(u^{k+1} - u^k)) \quad (4.11)$$

$$= \exp(H_j^T u^{k+1}) x_j^k \exp(-H_j^T u^k) \quad (4.12)$$

$$= \exp(H_j^T u^{k+1}) [x_j^{k-1} \exp(H_j^T(u^k - u^{k-1})) \exp(-H_j^T u^k)] \quad (4.13)$$

$$= \exp(H_j^T u^{k+1}) [x_j^{k-1} \exp(-H_j^T u^{k-1})] \quad (4.14)$$

$\Rightarrow$

$$x_j^{k+1} = \exp(H_j^T u^{k+1}) [x^0 \exp(-H_j^T u^0)] \quad (4.15)$$

$$= x_j^0 \exp(H_j^T u^{k+1}) \quad (4.16)$$

where we have assumed that  $u^0$  is zero. Dropping the iteration notation and solving Eq. (4.16) for  $\mathcal{H}_j^T u$  gives

$$\ln x_j - \ln x^0 = H_j^T u \quad (4.17)$$

Comparing Eq. (4.17) to Eq. (4.6) shows that

$$\nabla f(x_j) = \ln x_j - \ln x^0. \quad (4.18)$$

Integrating to find  $f(x)$  gives

$$f(x_j) = x_j \ln x_j - x_j - x_j \ln x^0 \quad (4.19)$$

$$= x_j \ln x_j - x_j(1 + \ln x_j^0) \quad (4.20)$$

$$(4.21)$$

which equals  $x_j \ln x_j$  if  $x^0 = \exp(-1)$ . So, if we assume that  $u^0 = 0$ , setting  $x^0 = \exp(-1)$  yields the maximum entropy solution.

## The SART Solution

Now the goal is to “work backwards” from the SART algorithm to find what kind of solution it converges to, as we have done for MART.

First, we assume that the dual variable,  $u$ , is updated the same as for MART,

$$u^{k+1} = u^k + d^k. \quad (4.22)$$

Once again, this is a reasonable assumption, since when  $d^k = 0$  both the primal and the dual variables will converge. Substituting Eq. (4.22) into Eq. (4.3) gives

$$\begin{aligned} x_j^{k+1} &= x_j^k + \frac{H_j^T (u^{k+1} - u^k)}{H_j^T \mathbf{1}} \\ &= x_j^k + \frac{H_j^T u^{k+1}}{H_j^T \mathbf{1}} - \frac{H_j^T u^k}{H_j^T \mathbf{1}} \\ &= x_j^{k-1} + \frac{H_j^T u^k}{H_j^T \mathbf{1}} - \frac{H_j^T u^{k-1}}{H_j^T \mathbf{1}} + \frac{H_j^T u^{k+1}}{H_j^T \mathbf{1}} - \frac{H_j^T u^k}{H_j^T \mathbf{1}} \\ &= x_j^{k-1} - \frac{H_j^T u^{k-1}}{H_j^T \mathbf{1}} + \frac{H_j^T u^{k+1}}{H_j^T \mathbf{1}} \\ &\Rightarrow \\ x_j^{k+1} &= x_j^0 - \frac{H_j^T u^0}{H_j^T \mathbf{1}} + \frac{H_j^T u^{k+1}}{H_j^T \mathbf{1}}. \end{aligned} \quad (4.23)$$

If  $u^0$  is set to zero, then Eq. (4.23) reduces to

$$(x_j - x_j^0) H_j^T \mathbf{1} = H_j^T u \quad (4.24)$$

Comparing Eq. (4.24) to Eq. (4.6) shows that SART minimizes the function

$$f(x_j) = \frac{x_j^2 H_j^T \mathbf{1}}{2} + x_j x_j^0 H_j^T \mathbf{1} \quad (4.25)$$

which reduces to

$$f(x_j) = \frac{x_j^2 H_j^T \mathbf{1}}{2} \quad (4.26)$$

if  $x^0$  is set to zero. SART, then, minimizes a *weighted minimum norm* function, where the weighting accounts for the pixel sampling.

For the sequel it will be convenient to write Eq. (4.26) as

$$f(x_j) = \frac{x_j^2 H_j^T \mathbf{1}}{\alpha}, \quad (4.27)$$

which is allowable because a multiplicative constant does not change the minimization solution.

### 4.2.2 SART Convergence

In Section 4.2.1 it was shown that the SART algorithm solves the constrained minimization problem

$$\min_x \frac{1}{\alpha} \sum_{j=1}^n x_j^2 \mathcal{H}_j^T \mathbf{1} \quad \text{such that} \quad y = \mathcal{H}x. \quad (4.28)$$

To prove convergence to this solution the techniques of Section 3.2 are employed. Again, Theorem 1 is used, and since the function  $f(x)$  is convex, all that is necessary to show convergence is to demonstrate that the SART iteration is IAF with respect to  $\mathcal{H}$ ,  $y$ , and the minimizing function. The following proof is analogous to proving Propositions 16 and 17 in [10]. Because SART is a block algorithm, Proposition 18 in [10] is not necessary to show convergence (see [17]).

To set up the problem, first define the Lagrangian minimum,

$$g(u) = \min_{x \in \mathbb{R}^n} (f(x) + \langle u, y - Hx \rangle) \quad (4.29)$$

$$= \min_{x \in \mathbb{R}^n} \left( \frac{1}{\alpha} \sum_{j=1}^n x_j^2 \sigma_j + \langle u, y - \mathcal{H}x \rangle \right) \quad (4.30)$$

where  $\sigma_j \equiv \mathcal{H}_j^T \mathbf{1}$ . Setting the derivative of the Lagrangian to zero (the Kuhn-Tucker condition) gives the minimizing  $\tilde{x}$  as

$$\tilde{x}_j = \frac{H_j^T u \alpha}{2\sigma_j}. \quad (4.31)$$

Setting  $\alpha = 2$  as in Eq. (4.26) and using the dual variable update suggested in Eq. (4.22), Eq. (4.31), written in iteration notation, becomes

$$\begin{aligned} \tilde{x}_j^{k+1} &= \frac{H_j^T (u^k + d^k)}{\sigma_j} \\ &= \tilde{x}_j^k + \frac{H_j^T d^k}{\sigma_j}, \end{aligned} \quad (4.32)$$

which is the SART iteration.

With the minimizing  $\tilde{x}$ ,  $g(u)$  can be written as

$$g(u) = \frac{1}{\alpha} \sum_{j=1}^n \tilde{x}_j^2 \sigma_j + \langle u, y \rangle - \sum_{j=1}^n \sum_{i=1}^m u_i h_{ij} \tilde{x}_j. \quad (4.33)$$

The first step in proving convergence is to show that  $g(u^k)$  is monotonically increasing in  $k$ :

**Proposition 4**

$g(u^k)$  is monotonically increasing in  $k$ .

*Proof:* For convenience, the tilde notation is dropped from  $x$ . Consider the difference:

$$\begin{aligned} g(u^{k+1}) - g(u^k) &= \frac{1}{\alpha} \sum_{j=1}^n [(x_j^{k+1})^2 - (x_j^k)^2] \sigma_j + \langle u^{k+1} - u^k, y \rangle \\ &\quad + \sum_{i=1}^m \sum_{j=1}^n (u_i^k h_{ij} x_j^k - u_i^{k+1} h_{ij} x_j^{k+1}) \end{aligned} \quad (4.34)$$

$$\begin{aligned} &= \frac{1}{\alpha} \sum_{j=1}^n \left[ \left( x_j^k + \frac{H_j^T d^k}{\sigma_j} \right)^2 - (x_j^k)^2 \right] \sigma_j + \langle u^{k+1} - u^k, y \rangle \\ &\quad + \sum_{i=1}^m \sum_{j=1}^n \left[ u_i^k h_{ij} x_j^k - u_i^{k+1} h_{ij} \left( x_j^k + \frac{H_j^T d^k}{\sigma_j} \right) \right] \end{aligned} \quad (4.35)$$

where we used Eq. (4.3). This yields

$$\begin{aligned} g(u^{k+1}) - g(u^k) &= \frac{1}{\alpha} \sum_{j=1}^n \left[ (x_j^k)^2 + 2x_j^k \frac{H_j^T d^k}{\sigma_j} + \left( \frac{H_j^T d^k}{\sigma_j} \right)^2 - (x_j^k)^2 \right] \sigma_j \\ &\quad + \langle u^{k+1} - u^k, y \rangle + \sum_{i=1}^m \sum_{j=1}^n u_i^k h_{ij} x_j^k - \sum_{i=1}^m \sum_{j=1}^n u_i^{k+1} h_{ij} x_j^k \\ &\quad - \sum_{i=1}^m \sum_{j=1}^n u_i^{k+1} h_{ij} \frac{H_j^T d^k}{\sigma_j} \end{aligned} \quad (4.36)$$

$$\begin{aligned} &= \frac{1}{\alpha} \sum_{j=1}^n \left[ 2x_j^k \frac{H_j^T d^k}{\sigma_j} + \left( \frac{H_j^T d^k}{\sigma_j} \right)^2 \right] \sigma_j + \langle u^{k+1} - u^k, y \rangle \\ &\quad + \langle u^k - u^{k+1}, y \rangle + \langle u^k - u^{k+1}, -d^k \rangle \\ &\quad - \sum_{i=1}^m \sum_{j=1}^n u_i^{k+1} h_{ij} \frac{H_j^T d^k}{\sigma_j}. \end{aligned} \quad (4.37)$$

Continuing to simplify,

$$g(u^{k+1}) - g(u^k) = \frac{1}{\alpha} \sum_{j=1}^n \left[ 2x_j^k \frac{H_j^T d^k}{\sigma_j} + \left( \frac{H_j^T d^k}{\sigma_j} \right)^2 \right] \sigma_j + \langle u^{k+1} - u^k, y \rangle$$

$$+\langle u^k - u^{k+1}, y \rangle + \|d^k\|_2^2 - \sum_{i=1}^m \sum_{j=1}^n u_i^{k+1} h_{ij} \frac{H_j^T d^k}{\sigma_j} \quad (4.38)$$

$$= \sum_{j=1}^n \left[ \left( \frac{2}{\alpha} x_j^k \frac{H_j^T d^k}{\sigma_j} + \frac{1}{\alpha} \left( \frac{H_j^T d^k}{\sigma_j} \right)^2 \right) \sigma_j - \frac{H_j^T d^k}{\sigma_j} \sum_{i=1}^m u_i^{k+1} h_{ij} \right] + \|d^k\|_2^2 \quad (4.39)$$

$$= \sum_{j=1}^n \left[ \left( \frac{2}{\alpha} x_j^k \frac{H_j^T d^k}{\sigma_j} + \frac{1}{\alpha} \left( \frac{H_j^T d^k}{\sigma_j} \right)^2 \right) \sigma_j - \frac{2}{\alpha} \frac{H_j^T d^k}{\sigma_j} x_j^{k+1} \sigma_j \right] + \|d^k\|_2^2 \quad (4.40)$$

$$= \sum_{j=1}^n \left[ \frac{2}{\alpha} H_j^T d^k (-x_j^{k+1} + x_j^k) + \frac{1}{\alpha} \left( \frac{(H_j^T d^k)^2}{\sigma_j} \right) \right] + \|d^k\|_2^2 \quad (4.41)$$

$$= \sum_{j=1}^n \left[ \frac{2}{\alpha} H_j^T d^k \left( \frac{-H_j^T d^k}{\sigma_j} \right) + \frac{1}{\alpha} \left( \frac{(H_j^T d^k)^2}{\sigma_j} \right) \right] + \|d^k\|_2^2 \quad (4.42)$$

$$= -\frac{1}{\alpha} \sum_{j=1}^n \left( \frac{(H_j^T d^k)^2}{\sigma_j} \right) + \|d^k\|_2^2. \quad (4.43)$$

Equation (4.43) can be written in matrix notation as

$$-\frac{1}{\alpha} (d^k)^T \mathcal{H} K K^T \mathcal{H}^T (d^k) + (d^k)^T (d^k) \quad (4.44)$$

where

$$K = \text{diag}\{\mathcal{H}^T \mathbf{1}\}^{-\frac{1}{2}}. \quad (4.45)$$

For  $g(u)$  to be monotonically increasing the variable  $\alpha$  must be picked so that Eq. (4.44) is strictly non-negative. This is guaranteed for  $\alpha$  greater than or equal to the largest singular value of  $\mathcal{H}K$ . Thus, an  $\alpha$  which depends only on  $\mathcal{H}$  can be picked such that  $g(u)$  is monotonically increasing.

*q. e. d*

Proposition 5 shows that the algorithm is IAF.

**Proposition 5**

If  $\{x^k\}_{k=0}^\infty$  and  $\{u^k\}_{k=0}^\infty$  are generated by SART then

$$(i) \quad \lim_{k \rightarrow \infty} [g(u^{k+1}) - g(u^k)] = 0, \quad (4.46)$$

$$(ii) \quad \lim_{k \rightarrow \infty} (\langle h_i^k, x^k \rangle - y_i) = 0, \quad (4.47)$$

$$(iii) \quad \lim_{k \rightarrow \infty} (u^{k+1} - u^k) = 0, \quad (4.48)$$

$$(iv) \quad \lim_{k \rightarrow \infty} \langle h_i^k, x^{k+1} \rangle - y_i = 0. \quad (4.49)$$

*Proof*

(i)  $\{g(u^k)\}_{k=0}^\infty$  is monotonically increasing by Proposition 4, and by Eq. (3.7) is bounded from above.

(ii) Proposition 4 and Eq. (4.46) imply that

$$\lim_{k \rightarrow \infty} d^k = 0. \quad (4.50)$$

(iii) Follows from Eqs. (4.22) and (4.50).

(iv) Use (4.22) to write

$$\begin{aligned} \langle h_i^k, x^{k+1} \rangle - y_i &= \sum_{j=1}^n h_{ij} \left( x_j^k + \frac{(u^{k+1} - u^k)}{\mathcal{H}_j^T \mathbf{1}} \right) - y_i \\ &= \langle h_i, x^k \rangle - y_i + \sum_{j=1}^n \frac{(u^{k+1} - u^k)}{\mathcal{H}_j^T \mathbf{1}} \end{aligned} \quad (4.51)$$

which goes to zero by Eqs. (4.47) and (4.48).

*q. e. d*

Because the algorithm is IAF as per Proposition 5, by Theorem 1 it converges to the solution Eq. (4.28). In summary, we have proven that the SART iteration converges to the weighed minimum norm solution in Eq. (4.27).

### 4.2.3 Advantages of the Column Normalization

The advantages of using a column-normalized algorithm, like SART, are shown in Figures 4.1 and 4.2. The weighting in the minimization function, which is a result of the column normalization, appears to help reduce sampling artifacts. For the randomly decimated case (Figure 4.1) the figure arguably appears less “noisy,” and the dark cross-hatch marks are not as pronounced in the case where the image was regularly decimated by 2 in the  $x$  and  $y$  directions (Figure 4.2). This characteristic is very important when using the images for land-type classification, since reconstruction artifacts can lead to erroneous classifications.

### 4.3 BYU MART and SIR

The BYU MERS lab has been using column-normalized algorithms for several years. In his thesis, Peter Whiting suggested using an algorithm that he called “Block ART” (which I have renamed “BYU MART”) for remote sensing image reconstruction [7]. Note the characteristic update column normalization by  $\mathcal{H}_j^T \mathbf{1}$ .

#### Algorithm 8 (BYU MART)

*Initialization:*  $x^\circ = \text{arbitrary}$ .

*Iterative step*

$$x_j^{k+1} = x_j^k \frac{\sum_{i=1}^M \left[ \frac{y_i}{\langle h_i, x^k \rangle} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}}, \quad (4.52)$$

or

$$x_j^{k+1} = x_j^k \frac{\mathcal{H}_j^T d_i^k}{\mathcal{H}_j^T \mathbf{1}} \quad (4.53)$$

in matrix form, where

$$d_i^k = \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^\lambda. \quad (4.54)$$

The reason the algorithm was originally called Block ART is because the update term from each measurement is added to give the final update term for the iteration.

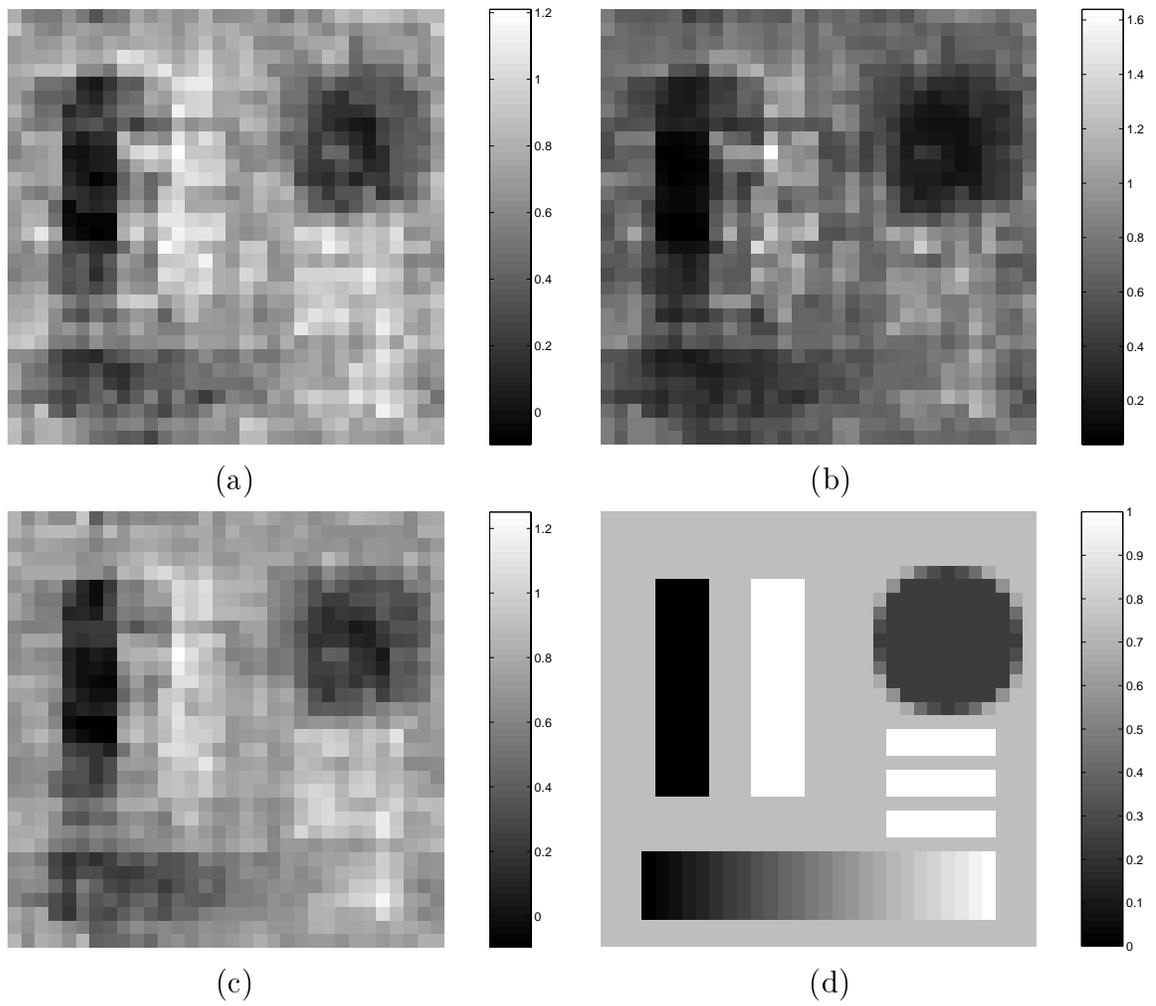


Figure 4.1: Reconstruction from noiseless, randomly sampled data: a) ART b) MART c) SART d) Original Image

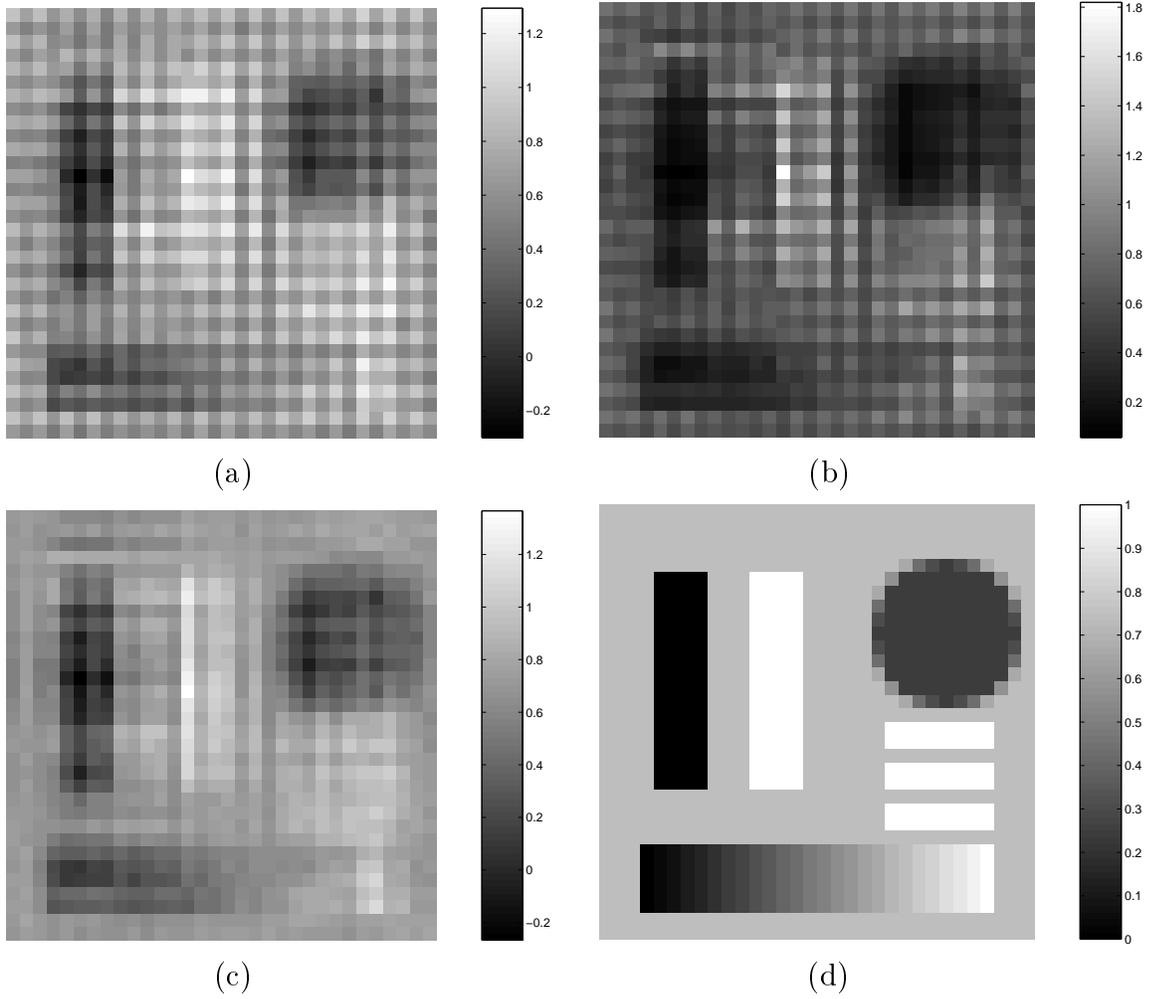


Figure 4.2: Reconstruction from noiseless, regularly decimated sampling: a) ART b) MART c) SART d)Original Image

However, I feel that BYU MART is a more appropriate name since the the current estimate is multiplied by the update term to get a new estimate, and the fidelity term  $d$  is the same form as in MART.

However, it should be noted that, despite its name, BYU MART *does not* converge to a maximum entropy solution as MART does. Currently, it is not known what solution BYU MART converges to, but we have experimentally demonstrated by comparison with MART that it is not a maximum entropy algorithm. Since most algorithms in this algebraic class yield solutions which depend on initialization,  $x^\circ$ , the arbitrary initialization specified for BYU MART suggests the solution form is indeterminate. It is probable that changing the initial condition will change the solution, but since the solution is not known there is no reason to use any particular initial condition.

#### 4.3.1 SIR

The Scatterometer Image Reconstruction (SIR) algorithm, which was also developed at BYU for remote sensing reconstruction, is closely related to MART (See [22],[23],[8]).

##### Algorithm 9 (SIR)

*Initialization:*  $x^\circ = \text{arbitrary}$ .

*Iterative step*

$$x_j^{k+1} = \frac{\sum_{i=1}^M u_{ij}^k h_{ij}}{\sum_{i=1}^M h_{ij}}, \quad (4.55)$$

or

$$x_j^{k+1} = \frac{\mathcal{H}_j^T u_{ij}}{\mathcal{H}_j^T \mathbf{1}} \quad (4.56)$$

in matrix form. The nonlinear update term  $u_{ij}$  is given by

$$u_{ij}^k = \begin{cases} \left[ \frac{1}{2f_i^k} \left( 1 - \frac{1}{d_i^k} \right) + \frac{1}{x_j^k d_i^k} \right]^{-1} & d_i^k \geq 1 \\ \left[ \frac{1}{2} f_i^k (1 - d_i^k) + x_j^k d_i^k \right] & d_i^k < 1. \end{cases} \quad (4.57)$$

where  $d_i^k$  is the fidelity term

$$d_i^k = \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)$$

and  $f_i^k$  is the forward projection  $\langle h_i, x^k \rangle$ .

The purpose of the nonlinear update  $u_{ij}$  is to limit the amount that the estimate can change at each iteration. This limit helps smooth inconsistent data and has been found in practice to reduce noise more effectively than BYU MART (see [7]). SIR's good performance in noise has made it the algorithm of choice for image reconstruction at the BYU MERS lab.

Figures 4.3 and 4.4 compare the performance of BYU MART and SIR against standard MART. Notice that, like SART, they do a better job of suppressing sampling artifacts than MART, especially in the regularly decimated case. Even though the type of solution is unknown, it is now believed that this artifact suppression is at least partly due to the normalization by the columns of  $\mathcal{H}$  in the iterative update. This understanding of the functioning of SIR is only now being realized by structural comparisons we have made with the SART algorithm based on our convergence analysis. By using appropriate parameter settings, SART and SIR can be demonstrated to yield remarkably similar results.

### 4.3.2 BYU MART and SIR Convergence

Specific convergence properties of BYU MART and SIR are still unknown. Experimental evidence indicates that they converge to different solutions, but what form of solution they converge to is unknown. Appendix A, however, shows an analysis by David Long which sheds light on some of the convergence properties of these two algorithms.

## 4.4 Additive SIR

The fact that adding a nonlinear update term to BYU MART creates a more robust algorithm in the presence of noise gives motivation to modify the SART algorithm in a similar way. Thus I have added a nonlinear update to SART and called the

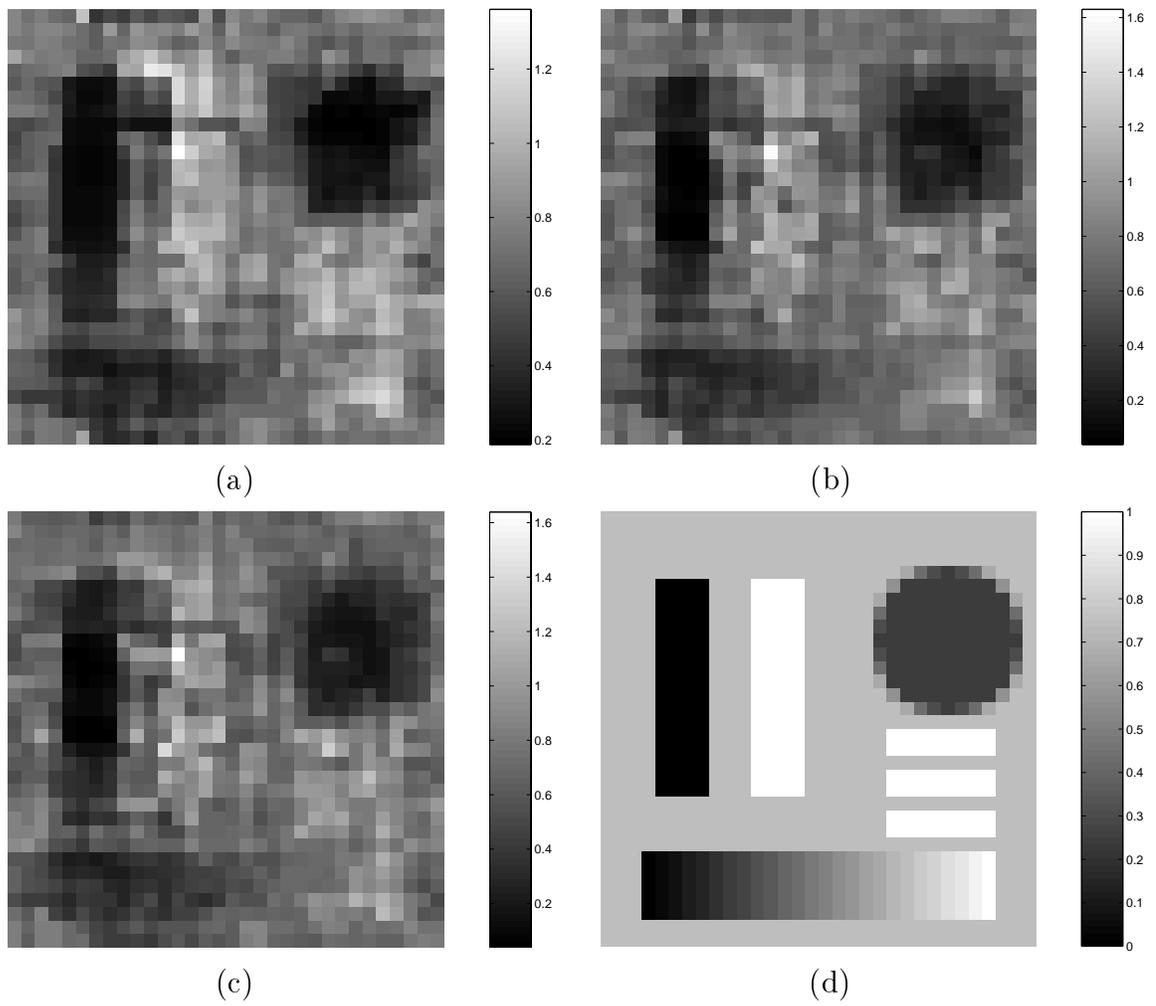


Figure 4.3: Reconstruction from noiseless random sampling: a) SIR b) BYU MART c) MART d) Original Image

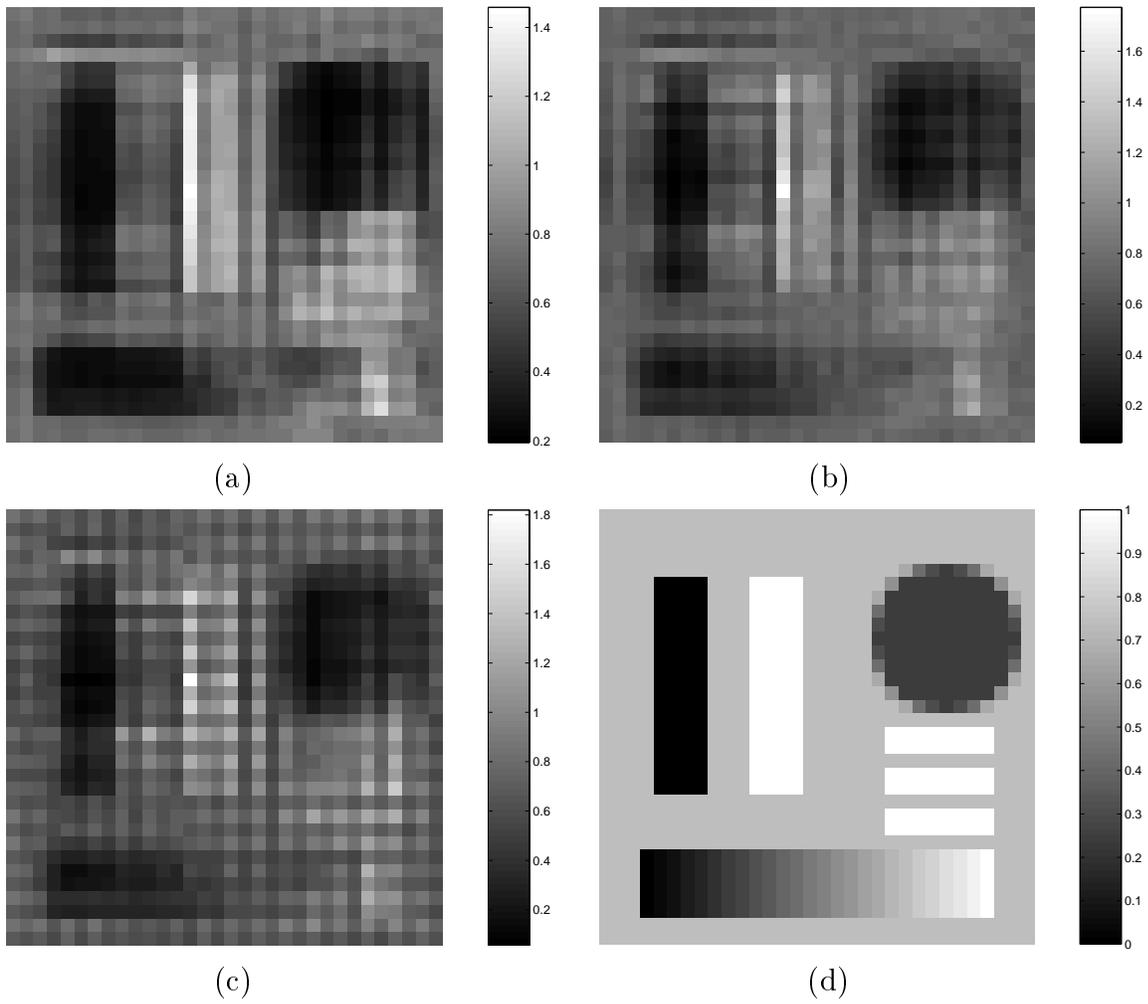


Figure 4.4: Reconstruction from noiseless regularly decimated sampling: a) SIR b) BYU MART c) MART d) Original Image

algorithm Additive SIR. The form of the Additive SIR algorithm is similar to that of SART, except with a non linear function of the fidelity term  $d$ .

**Algorithm 10 (Additive SIR)**

*Initialization:*  $x^0 = 0$ .

*Iterative step*

$$x_j^{k+1} = x_j^k + \frac{\sum_{i=1}^m h_{ij} d_i^k}{\sum_{i=1}^m h_{ij}}, \quad (4.58)$$

or

$$x_j^{k+1} = x_j^k + \frac{H_j^T d^k}{H_j^T \mathbf{1}} \quad (4.59)$$

in matrix form, where

$$f_i^k = \langle h_i, x^k \rangle, \quad (4.60)$$

$$\sigma_i^k = (y_i - f_i^k) \quad (4.61)$$

and

$$d_i^k = \begin{cases} f_i^k (1 - \exp(-\alpha \sigma_i^k)) + \beta \sigma_i^k \exp(-\alpha \sigma_i^k) & d_i^k \geq 0 \\ -f_i^k (1 - \exp(\alpha \sigma_i^k)) + \beta \sigma_i^k \exp(\alpha \sigma_i^k) & d_i^k < 0 \end{cases} \quad (4.62)$$

where  $\alpha$  and  $\beta$  are other relaxation factors that are determined empirically to give the best results in the presence of noise. The values of  $\alpha$  and  $\beta$  need to be determined experimentally and optimized for each particular instrument. For the SSM/I radiometer, I found  $\alpha = .001$  and  $\beta = .1$  to give stable convergence, while  $\alpha = .025$  and  $\beta = 1$  worked well for NSCAT.

Figures 4.5 and 4.6 compare the Additive SIR algorithm to ART and MART for decimated data. Once again, especially for the regularly decimated case, Additive SIR is much better than MART or ART at reducing sampling artifacts.

**4.4.1 Additive SIR Convergence**

In Section 4.3.2 it was shown that although BYU MART and SIR have similar forms they converge to different solutions. It is fair to ask, then, whether Additive SIR, with its nonlinear update, converges to a different solution than SART.

Figure 4.7 shows the convergence trajectories for the SART and Additive SIR algorithms for the three-dimensional system defined in (3.43). The two algorithms

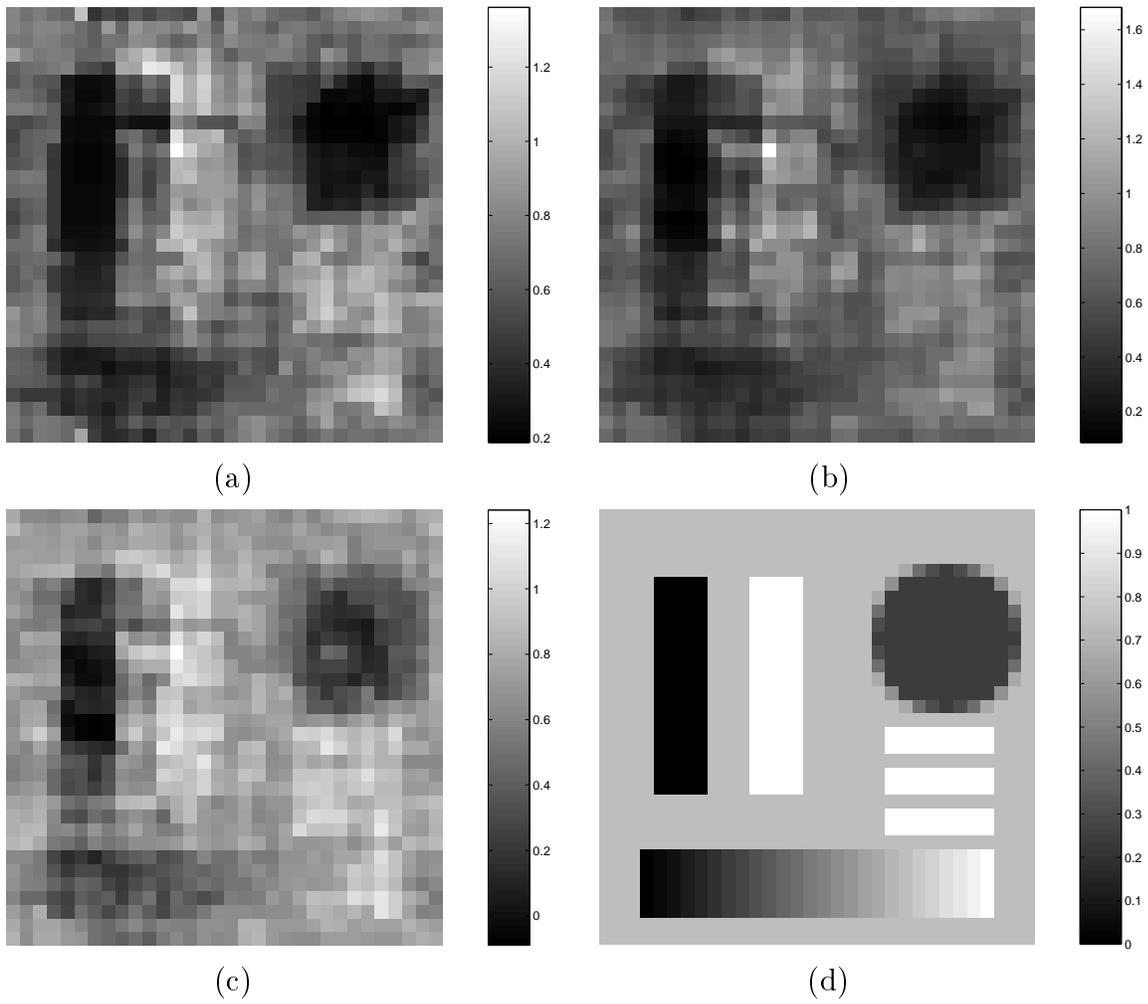


Figure 4.5: Reconstruction from noiseless random sampling: a) Additive SIR b) MART c) ART d)Original Image

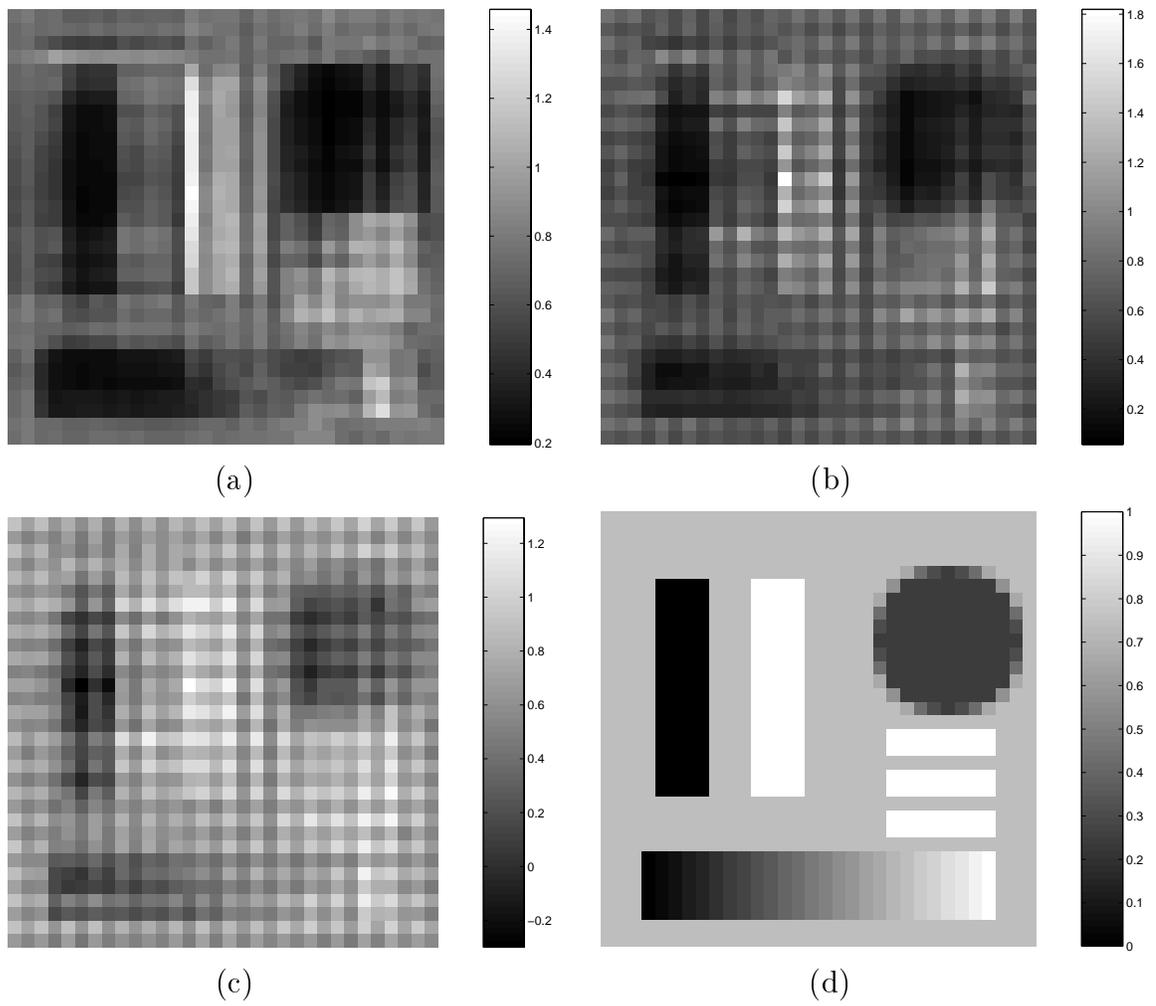


Figure 4.6: Reconstruction from noiseless regularly decimated sampling: a) Additive SIR b) MART c) ART d)Original Image

follow very similar trajectories and converge to the same solution. Numerically, therefore, they appear to be equivalent in convergence. This may make sense in the light of the proof of convergence of SART in Section 4.2.2, since there no restrictions are placed on the form of  $d$  needed for convergence.

However, as stated Section 3.5.2, care should be taken when using the three-dimensional examples. While the trajectories of the two algorithms appear to be the same for this three-dimensional case, they may not be the same for higher dimensional cases. Since with real, noisy data the algorithms are usually stopped well before convergence, the convergence path is important in determining the final image.

## 4.5 Conclusion

This chapter introduced column-normalized algorithms. The convergence properties of the SART algorithm indicate that column normalization helps reduce sampling artifacts. Chapters 5 and 6 will show how several column-normalized algorithms, as well as ART and MART, compare in reconstructing simulated and real data.

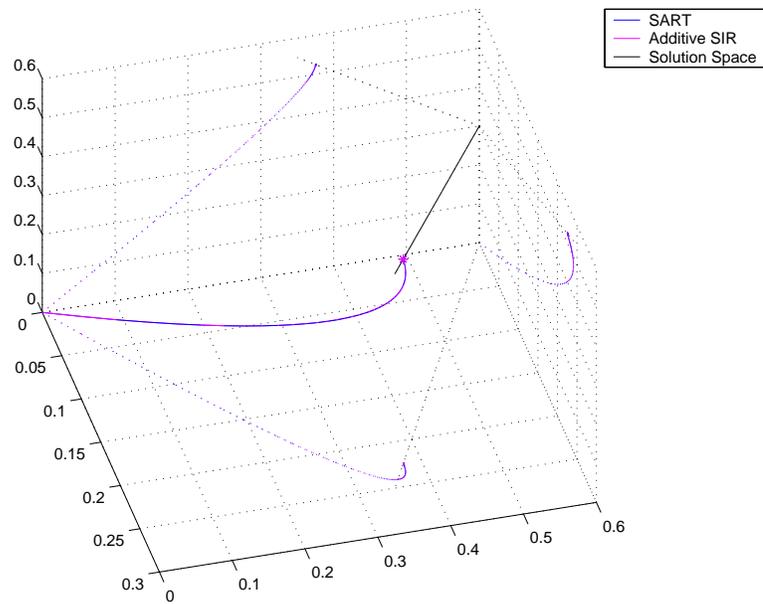


Figure 4.7: Convergence trajectories for the SART and Additive SIR algorithms



## Chapter 5

### Algorithm Performance Comparison Using Simulated Data

This chapter compares the performance of the several previously described reconstruction algorithms using simulated data. Remote sensing observations of the image in Figure 5.1 are simulated, and the image is reconstructed using actual sampling geometries and antenna response patterns from the SSM/I and NSCAT satellites. The sample patterns used are the same as those for an actual data set of an area over the Amazon, pictured in Figure 5.2. The sampled region covers from  $-58^\circ$  to  $-54^\circ$  longitude, and  $-6^\circ$  latitude to the equator. The simulated ground truth image is used with realistic sensor system modeling so that reconstruction performance can be evaluated against a known true reference.

When considering the reconstruction results which are presented in the sequel, we emphasize the following issues:

1. This is not meant to be a comprehensive comparison of the different algorithms. It would be irresponsible to look at the results of the following two chapters and try to make overly specific conclusions. Rather, the point of these simulations is to see generally how the algorithms perform.
2. This is not meant to be a comparison of the two instruments. Chapter 2 was provided to give an introduction to some of the issues involved in reconstruction from radiometer and scatterometer data sets. This is not a complete treatment on the tradeoffs between using the two different instruments. The purpose of this simulation is to compare how the different algorithms perform for each instrument separately.

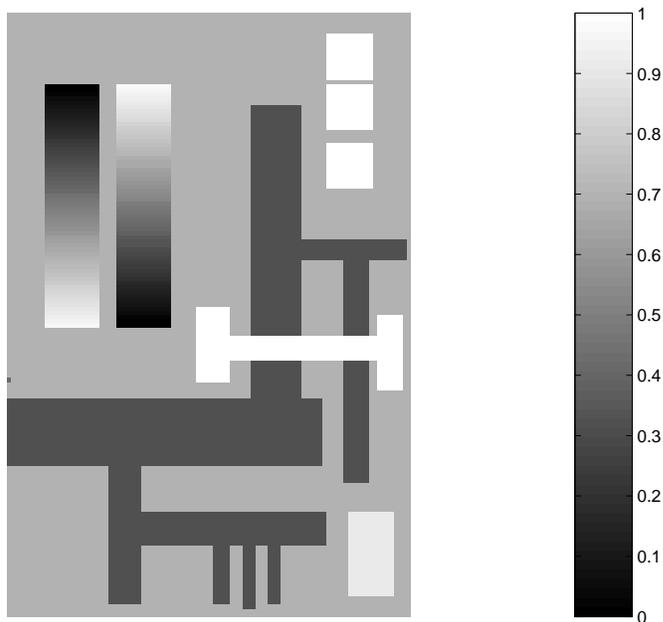


Figure 5.1: The original synthetic “truth image” used for simulations

## 5.1 The Modified Median Filter

When the SIR algorithm was initially developed, it was observed that using a modified median filter helped control the noise level in the image [7]. This filter selects the median intensity when there is a wide range of pixel values in the window, and computes an average when there is a narrow range of values. The modified median filter used in this simulation has a 3 pixel by 3 pixel window, and computes an average value if the difference between the second lowest and second highest intensity values is less than  $9.8e-4$ . Because of the complexity of analyzing such a nonlinear and non-stationary filter, we have not presented a theoretical development discussion on the modified median filter as used in image reconstruction. Algorithms with this filter are included in the sequel for performance comparison purposes. The algorithms compared in this section are Algebraic Reconstruction Technique (ART), ART with the modified median filter (ARTF), Multiplicative ART (MART), MART with filter (MARTF), Simultaneous ART (SART), SART with filter (SARTF), Additive SIR



Figure 5.2: The area in the black rectangle represents the region in the Amazon basin from which the measurement geometries are taken (the area is magnified at the bottom right hand corner of the image).

(AddSIR), AddSIR with filter (AddSIRF), the Scatterometer Image Reconstruction Technique (SIR), and SIR with filter (SIRF).

## 5.2 SSM/I Geometry Simulations

The first set of simulations uses a sampling grid matched to an SSM/I sampling over the Amazon region from Julian day 190 to 194 in 1999, corresponding to 2604 measurements. Figure 5.3 shows the instrument sampling over the region of interest for the specified day range.

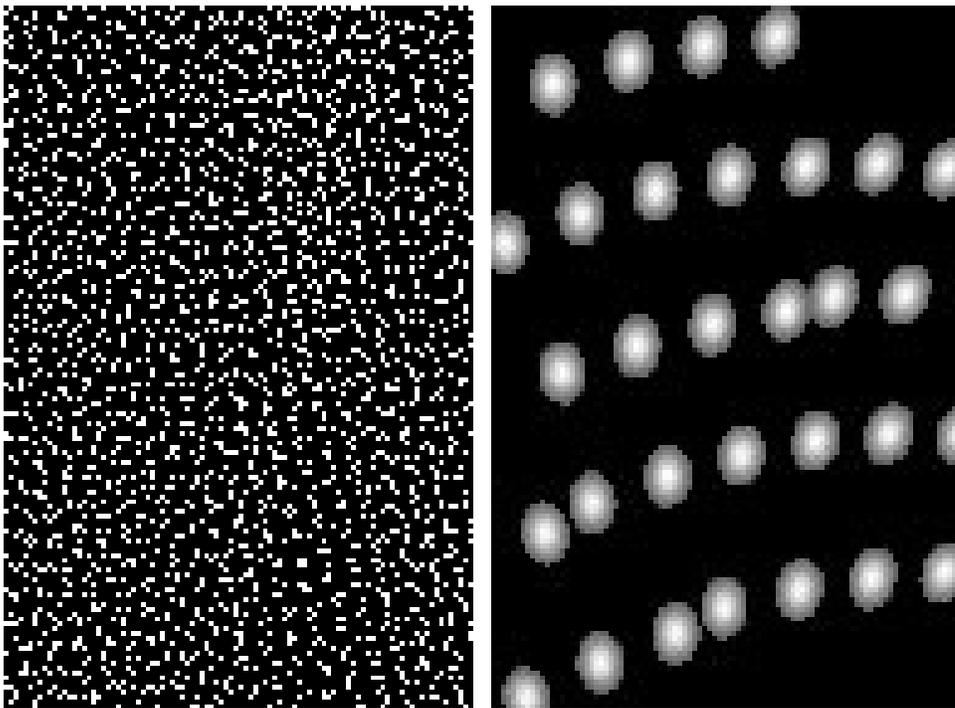


Figure 5.3: SSM/I sampling geometry—The dots in the left figure indicate the center of each footprint. The right figure shows a decimated version of the sampling to give a view of some representative footprints (both figures are on the same scale).

### 5.2.1 Noiseless Simulations

Although unrealistic as a model of actual sensor signals, noiseless simulations can provide insight into the strengths and weaknesses of the different algorithms. Figures 5.4 and 5.5 show the results of applying reconstruction algorithms to simulated data.

The column-normalized algorithms all seem to behave well, but there are differences between the algorithms: for example, the SIR algorithm does a better job of picking out edges than the other algorithms, but the tradeoff is that it has more ringing, as can be seen in the reconstruction of the two gradients in the upper left hand corner of the image.

When conducting these simulation experiments, we found the convergence of MART to be strongly dependent upon the initial condition. When MART was introduced in Chapter 3, the initial estimate  $x^\circ$  was given as  $x^\circ = \exp(-1)$ . Other initial conditions also guarantee ME convergence for MART, namely any  $x^\circ$  satisfying

$$1 + x_j^\circ = -\mathcal{H}_j^T u^\circ \quad (5.1)$$

where  $u^\circ \in \mathbb{R}^m$  is arbitrary [10]. An obvious choice for  $u^\circ$  is zero, which gives the familiar initial condition  $x^\circ = \exp(-1)$ . Experimental evidence shows, however, that picking an initial condition close to the expected mean value of the image greatly increases MART's convergence rate. As long as the the initial condition satisfies Eq. (5.1), ME convergence is guaranteed. In Chapter 6 we will show an example of image reconstruction from real data which emphasizes the importance of picking good initial conditions when using MART.

It is also important to remember that when using MART it is assumed that all of the pixel intensity values sum to 1, i.e.  $\sum_j x_j = 1$ . Since this condition is not realistically satisfied in remote sensing, MART reconstructions are not scale invariant, i.e., a measurement scaling does not imply the same reconstructed pixel scaling.

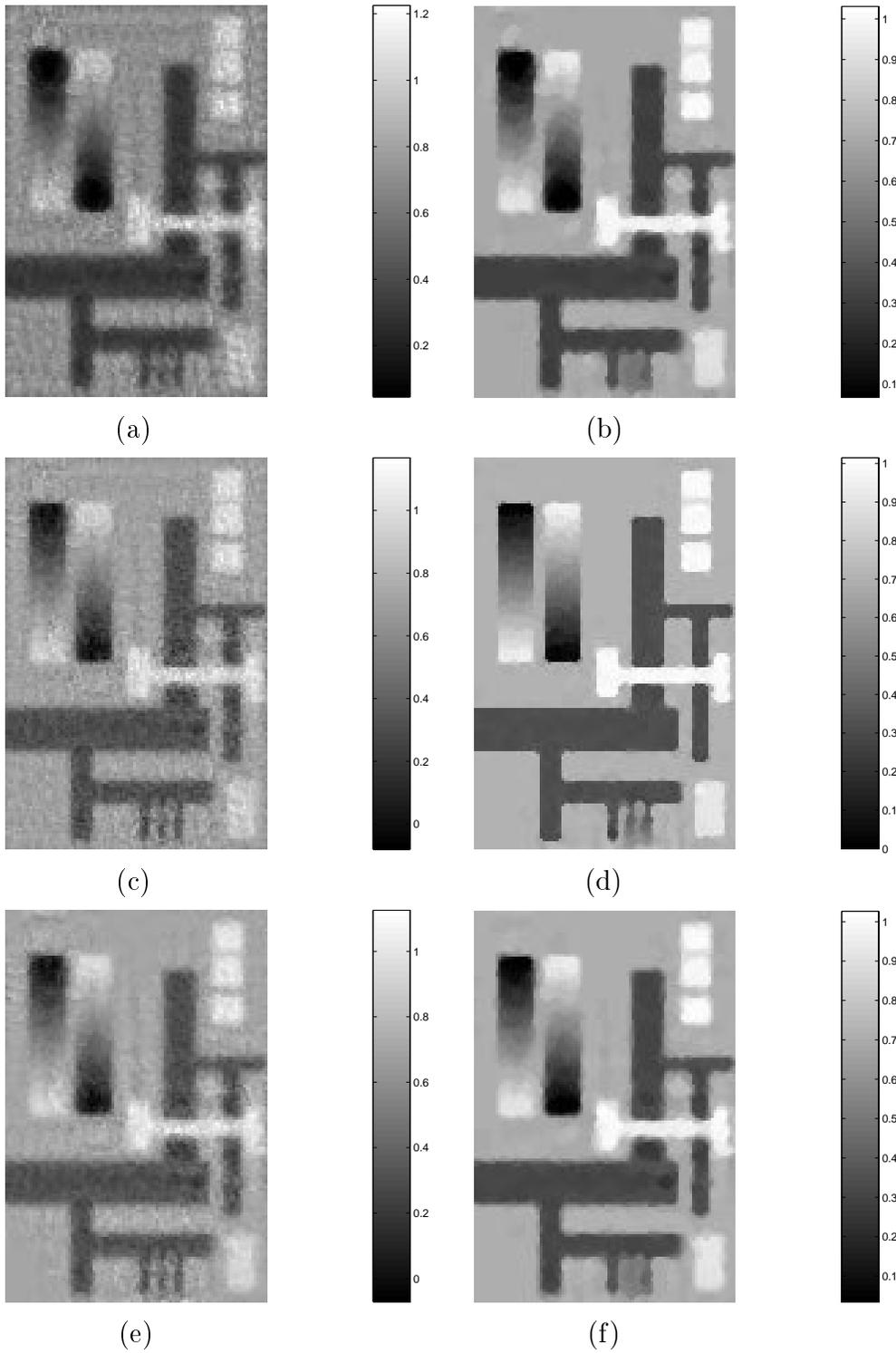


Figure 5.4: Noiseless reconstructions for SSM/I geometry using (a) MART (b) MART with filter (MARTF) (c) ART (d) ART with filter (ARTF) (e) SART (f) SART with filter (SARTF)

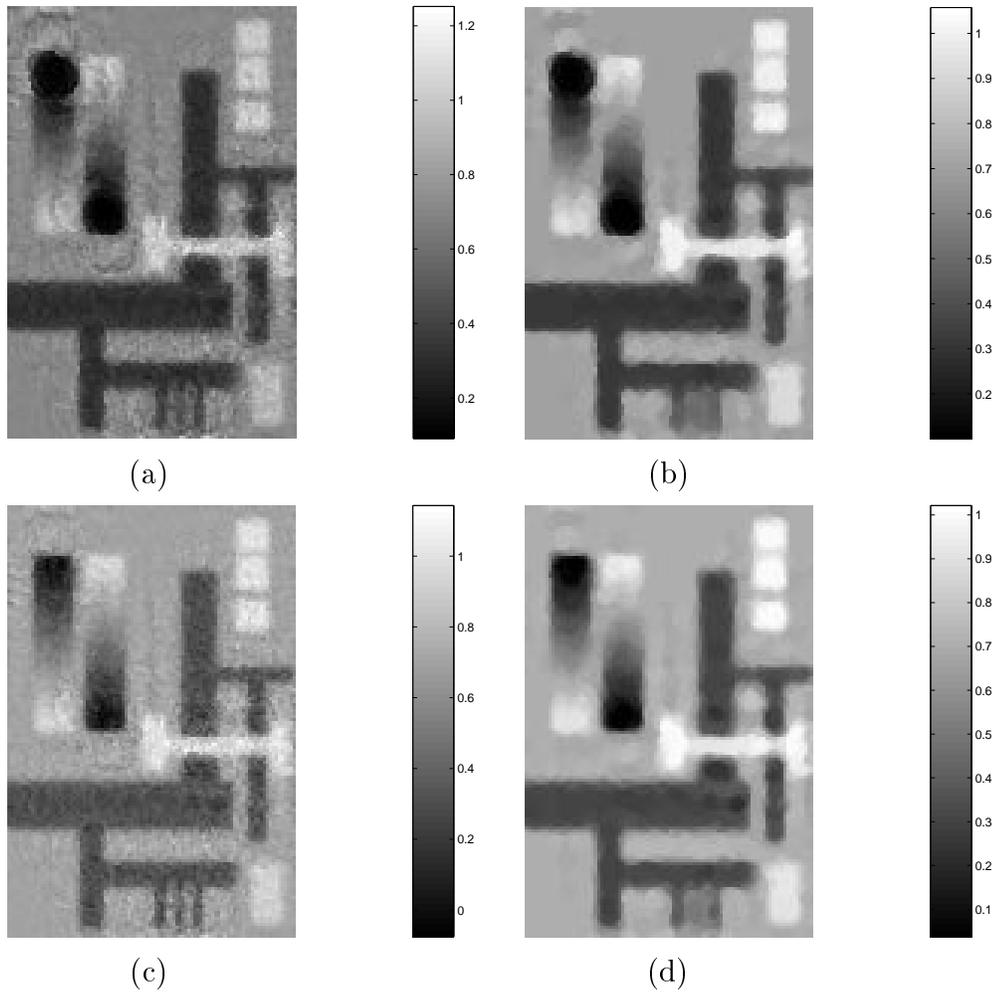


Figure 5.5: Noiseless reconstructions for SSM/I geometry using (a) SIR (b) SIR with filter (SIRF) (c) AddSIR (d) AddSIR with filter (AddSIRF)

### 5.2.2 Noisy Simulation

For the noisy SSM/I simulations, noise characteristics modeled after the instrument noise parameters are used. One model for SSM/I's noise is 1 Kelvin for a maximum ground radiometric brightness temperature of 280 Kelvin. Because the synthetic truth image is normalized such that its highest value is 1, the noise is normalized to  $1/280 = .0036$ . For this simulation, adding i.i.d, white, Gaussian noise with  $\sigma = .0036$  to the measurements to simulate measurement noise results in a signal to noise ratio of 45.0 dB. To aid in comparing the different algorithms in the presence of noise, Figures 5.6 and 5.7 show the resulting image at the iteration with the lowest solution error, as was done in Chapter 3.

Figures 5.6 and 5.7 suggest that ART and MART do not perform as well as the column-normalized algorithms in the presence of noise. This attribute lessens their utility for remote sensing reconstruction.

### 5.3 NSCAT Geometry Simulations

This section compares the various algorithms using the NSCAT sampling geometry. Figure 5.8 shows the measurement locations for NSCAT over the region of interest for the 8-day period between Julian day 156 and Julian day 164, 1997. Note that, while still undersampled relative to the reconstruction image pixel grid, the data set is more densely sampled than SSM/I data. Because of the denser sampling and smaller footprint, one would expect better reconstructions.

Figures 5.9 and 5.10 show the noiseless reconstructions, while Figures 5.11 and 5.12 show noisy reconstructions. The noise model for NSCAT is a standard deviation of 1% of the backscatter. Given the maximum backscatter intensity of 1 for the simulations, the additive noise would have a standard deviation of 0.01. However, to compare the effect of noise more easily, i.i.d, white, Gaussian noise with  $\sigma = 0.07$  was instead added to the measurements, resulting in a 19.7 dB SNR. For the noiseless simulations, all of the algorithms performed very well, which is expected for a highly sampled system, while the column-normalized algorithms appear to have better noise suppression properties than ART and MART.

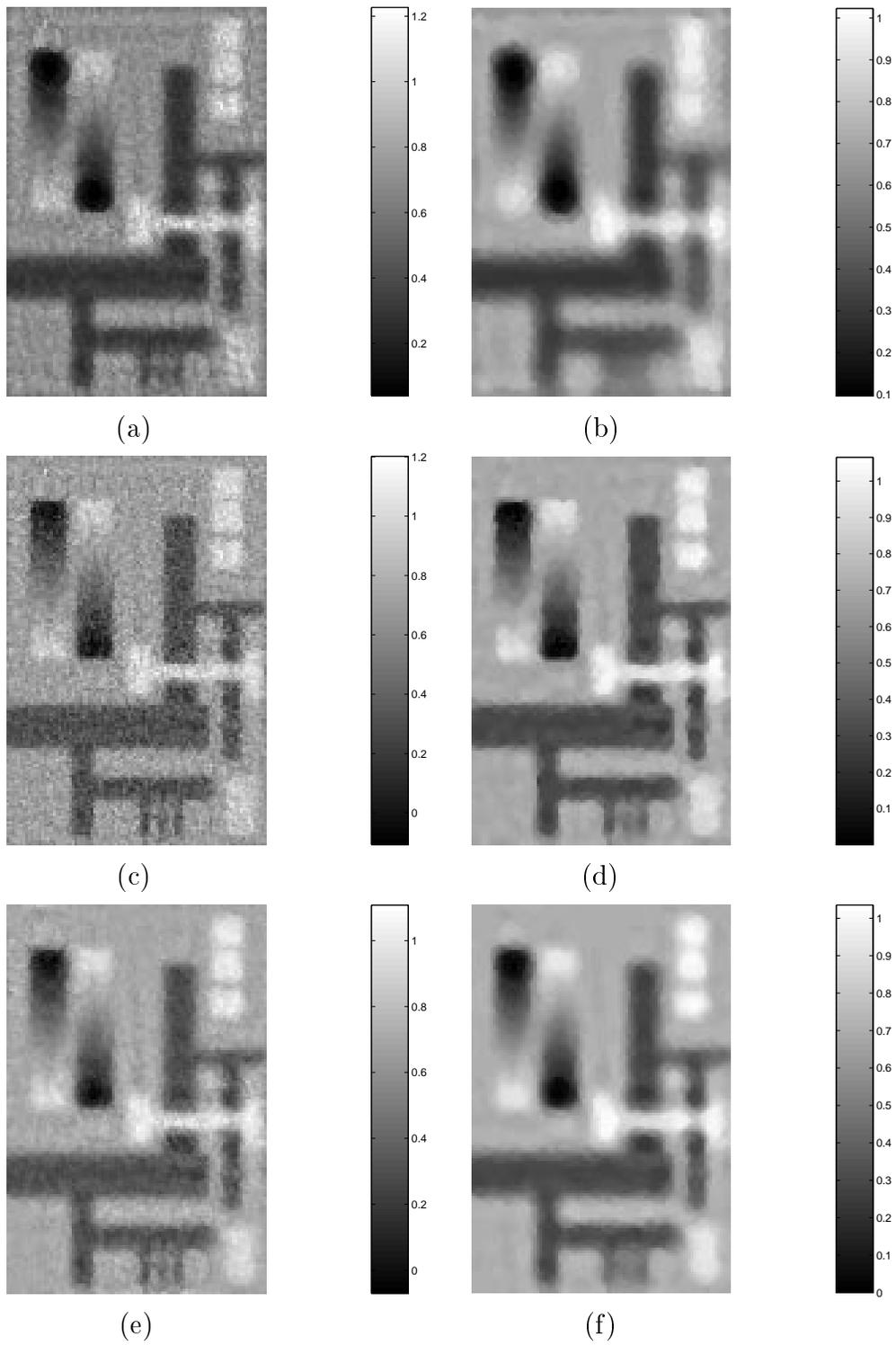


Figure 5.6: 45.0 dB SNR SSM/I reconstruction using (a) MART—398 iterations (b) MART with filter (MARTF)—42 iterations (c) ART—20 iterations (d) ART with filter (ARTF)—12 iterations (e) SART—1058 iterations (f) SART with filter (SARTF)—50iterations

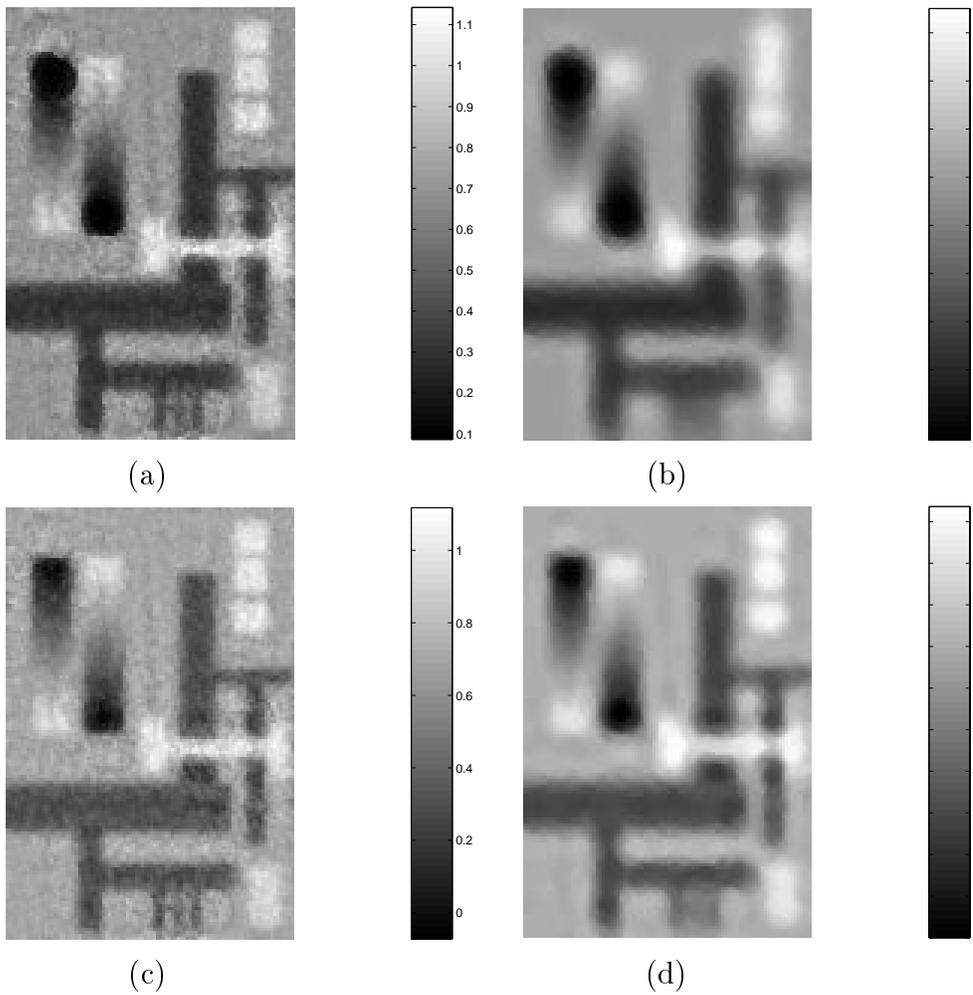


Figure 5.7: 45.0 dB SNR SSM/I reconstruction using (a) SIR—694 iterations (b) SIR with filter (SIRF)—27 iterations (c) AddSIR—2098 iterations (d) AddSIR with filter (AddSIRF)—50 iterations

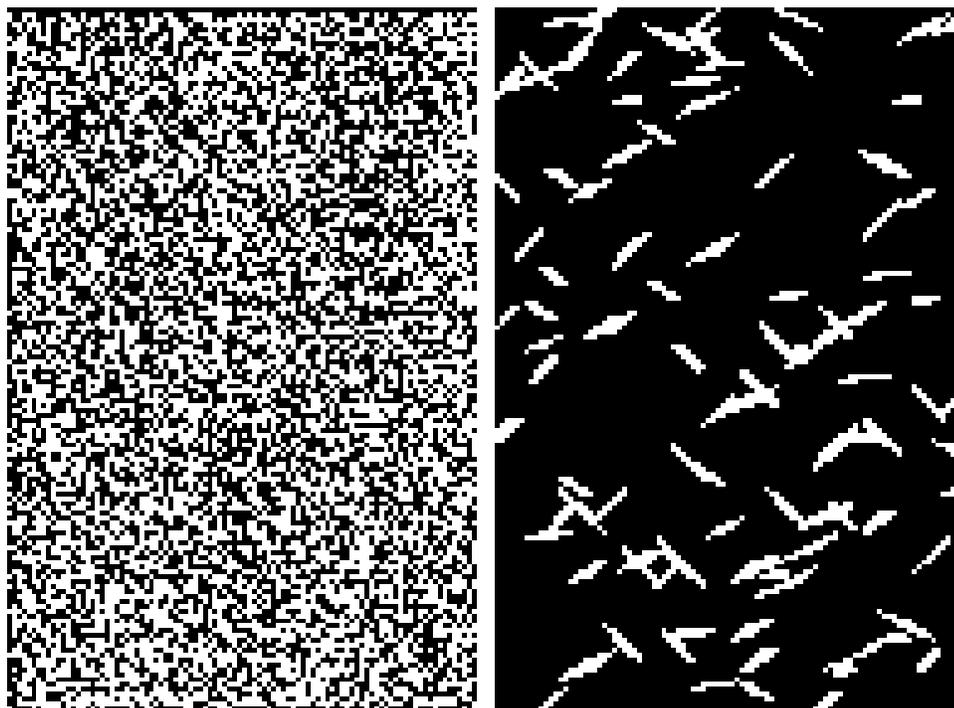


Figure 5.8: NSCAT sampling geometry—The dots in the left figure indicate the center of each footprint. The right figure shows a decimated version of the sampling to give a view of some representative footprints (both figures are on the same scale).

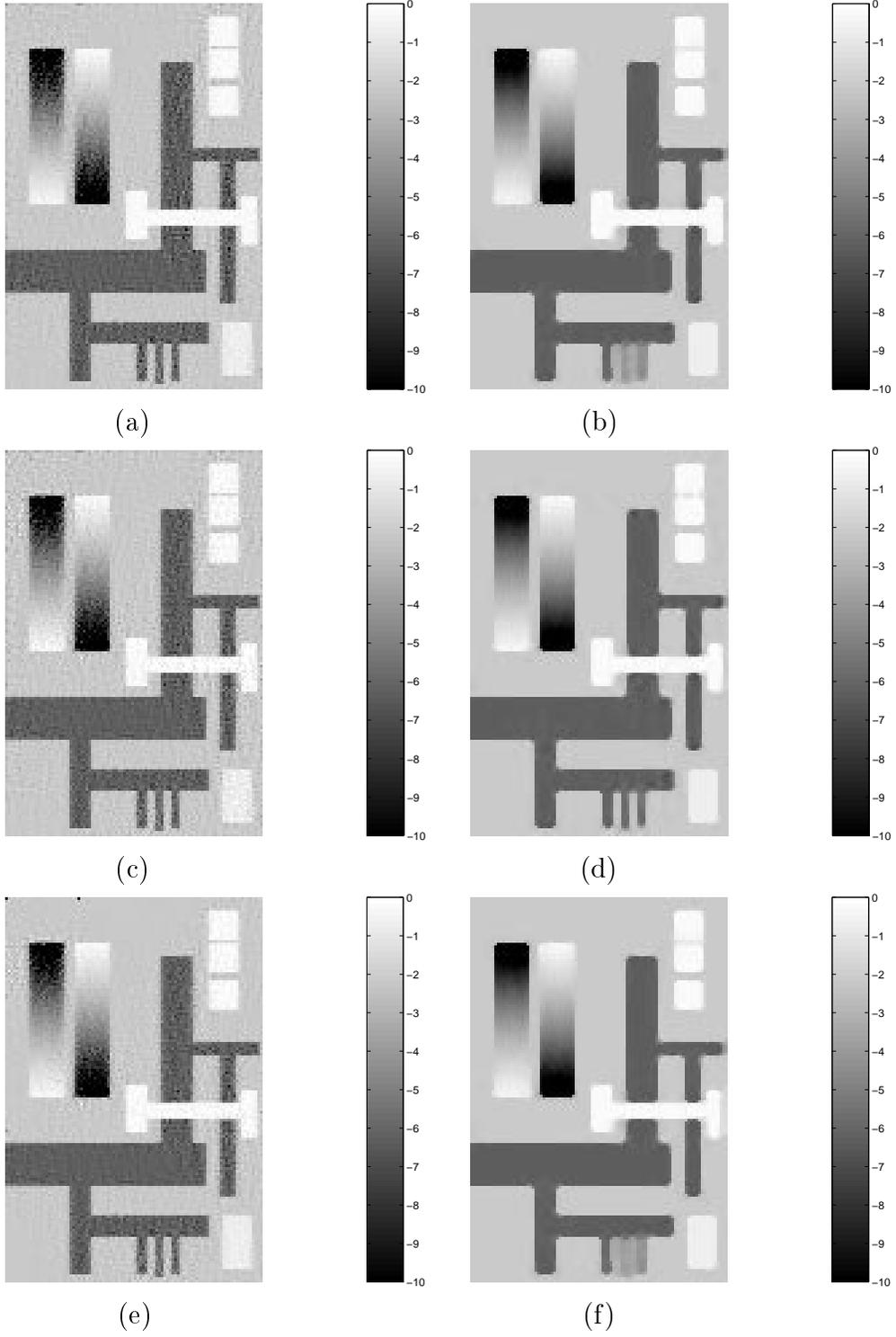


Figure 5.9: Noiseless NSCAT reconstruction using (a) MART (b) MARTF (c) ART (d) ARTF (e) SART (f) SARTF

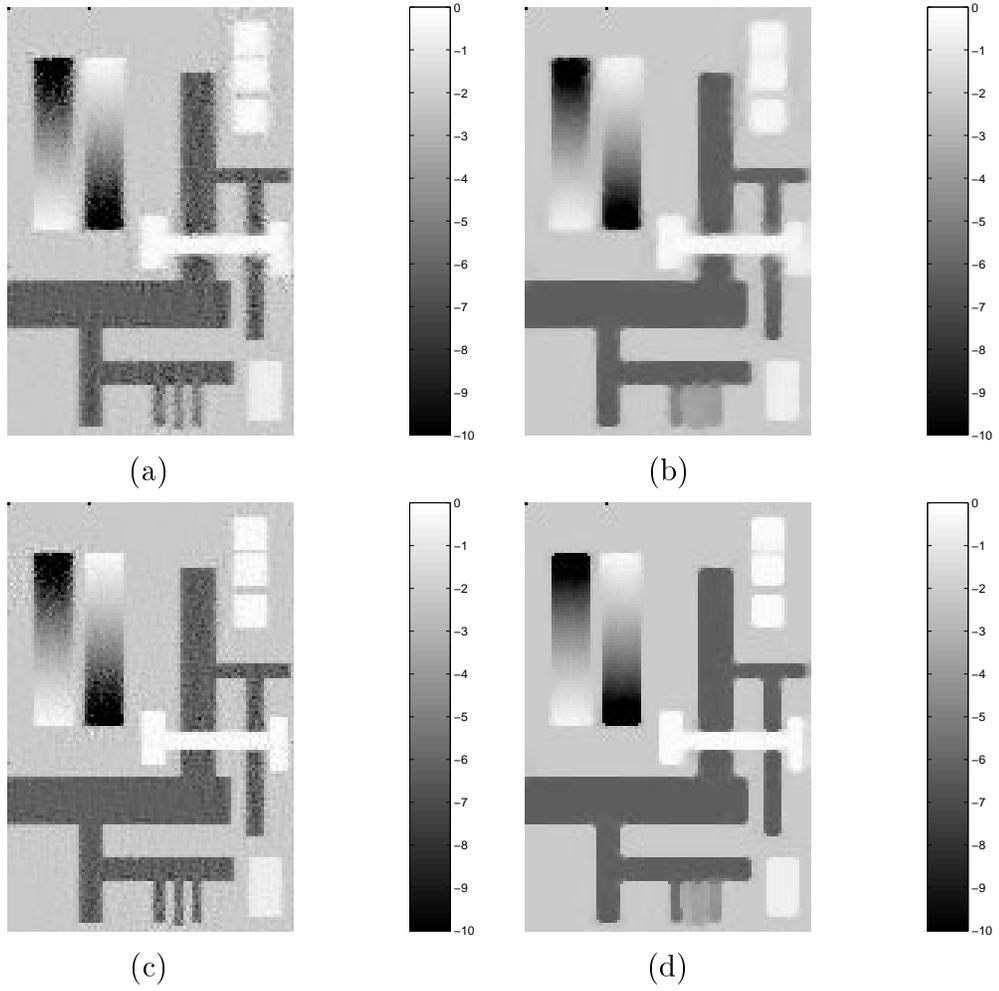


Figure 5.10: Noiseless NSCAT reconstructions using (a) SIR (b) SIRF (c) AddSIR (d) AddSIRF

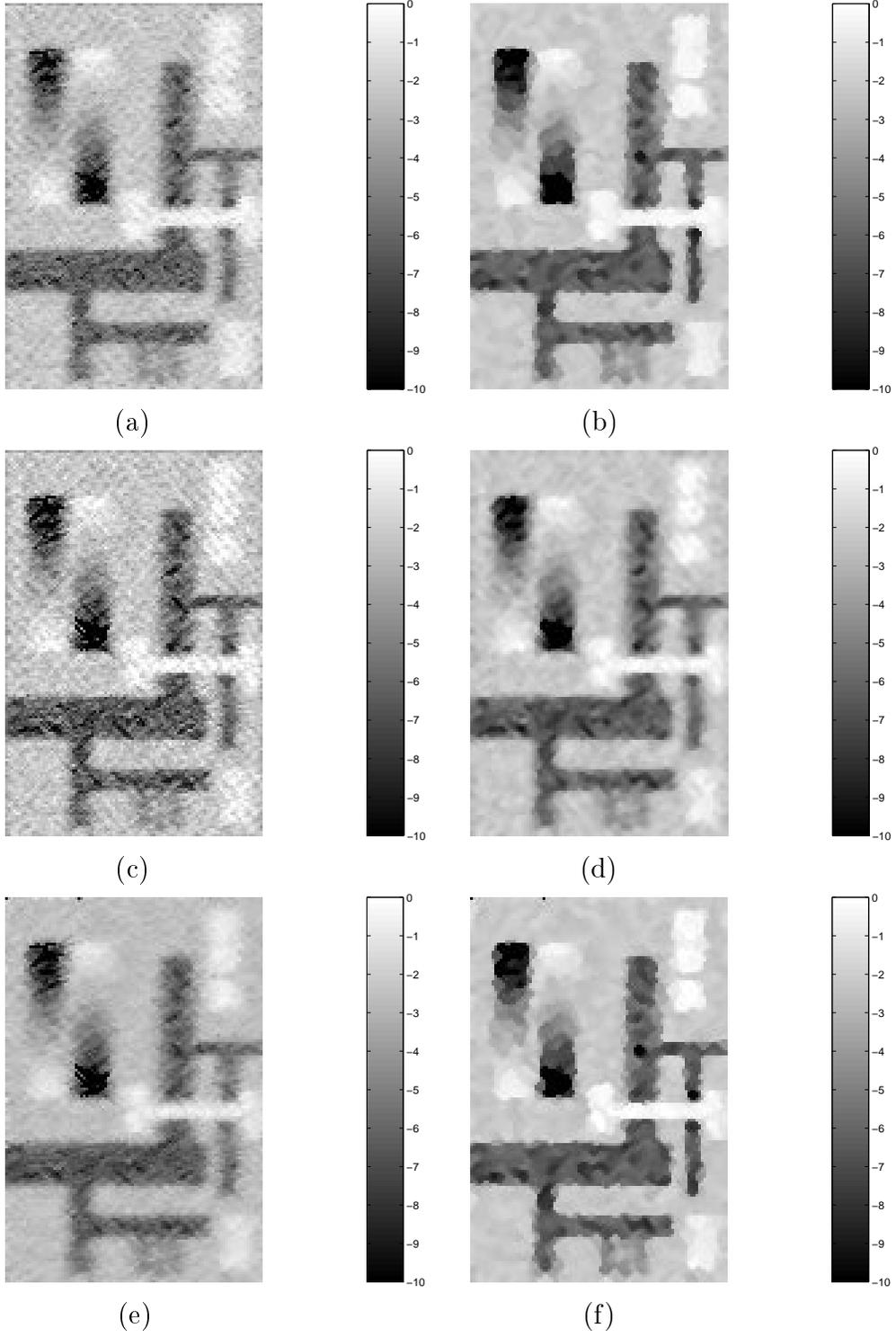


Figure 5.11: 19.7 dB SNR NSCAT reconstruction using (a) MART (b) MARTF (c) ART (d) ARTF (e) SART (f) SARTF

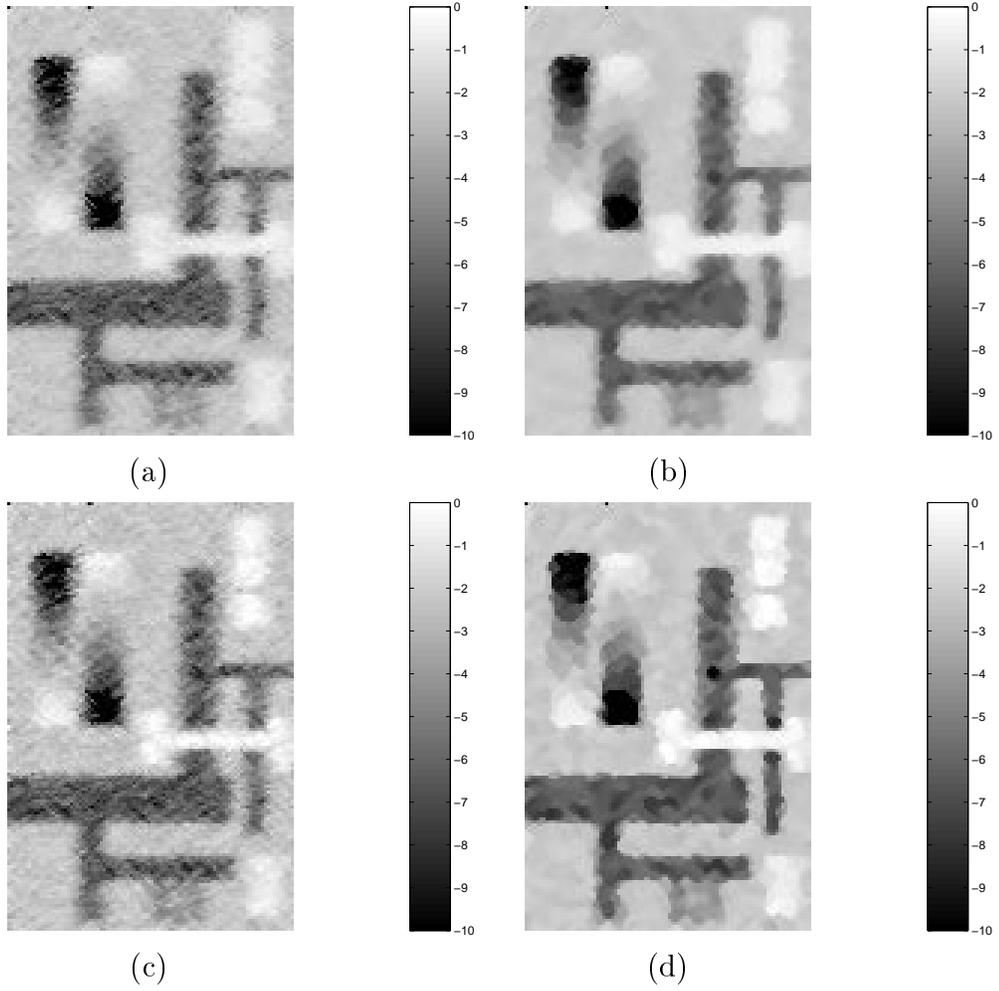


Figure 5.12: 19.7 dB SNR reconstruction using (a) SIR (b) SIRF (c) AddSIR (d) AddSIRF

## 5.4 Conclusions

The simulations in this chapter show how well the various algorithms reconstruct data from actual remote sensing measurement geometries. While all of the algorithms perform well, the column-normalized algorithms suppress artifacts and noise most effectively. The similarities between the SART and SIR reconstructions suggest that the primary advantage of SIR, as it has been used over several years, is the column normalization.

## Chapter 6

### Algorithm Performance Comparison Using Real Data

In this chapter, the various reconstruction algorithms are compared using actual SSM/I and NSCAT data. The measurements for the real data are taken from the same area (shown in Figure 5.2) and day range as for the simulated data in Chapter 5.

When reconstructing noisy data there is a tradeoff between resolution enhancement and noise amplification. Because the “truth” image is not known for real data, the iteration at which to stop for best results is somewhat subjective. Usually the algorithm cannot be left to iterate until convergence because of noise amplification, but stopping too early will result in a lower resolution image. The images in the following sections are those that the author considers “best,” so there is subjectivity in the choice.

#### 6.1 SSM/I Reconstructions

Figures 6.2 and 6.3 show the reconstruction results for real SSM/I data. As in Chapter 5, all of the algorithms perform reasonably well. MART and ART seem to have the noisiest reconstructions, ART appearing to be the worse of the two. The elliptical pattern of the noise for ART and MART suggest the need for the sampling artifact suppression that the column-normalized algorithms provide.

Figure 6.1 shows reconstruction for MART using the initial condition  $x^{\circ} = 1/e$ . While for the noiseless case this initial condition will lead to the maximum entropy solution, because of how the convergence properties depend on the initial condition, using this initial condition for SSM/I data leads to unacceptable results. Thus, a

uniform image at the expected mean value of the final image was used as the starting value to obtain the images shown in Fig. 6.2. This strong dependence on the starting point may be a disadvantage to using MART for remote sensing reconstruction.

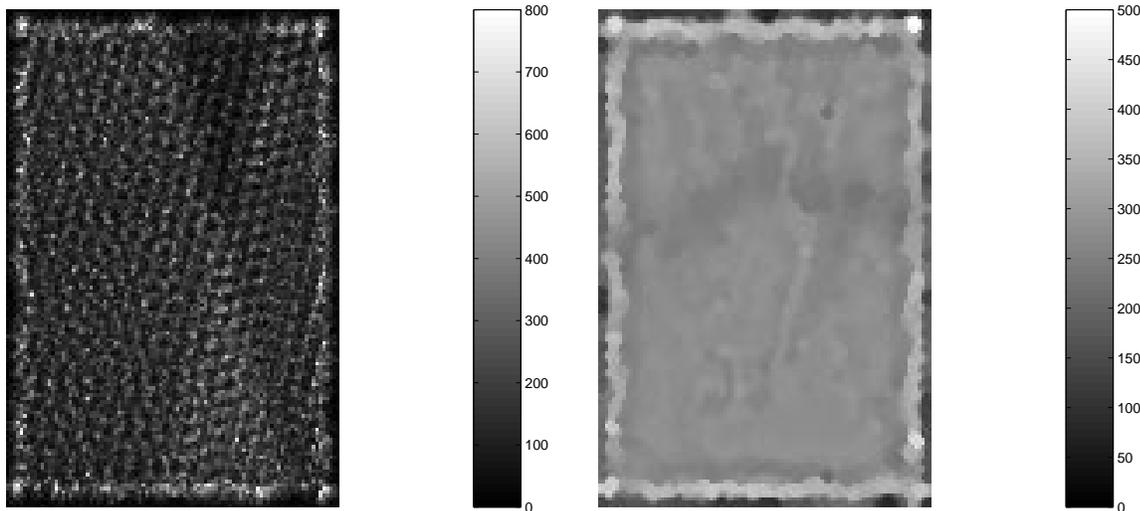


Figure 6.1: The result of using  $x^\circ = 1/e$  for SSM/I data for MART after 15 iterations (left), and MARTF after 50 iterations (right).

## 6.2 NSCAT Reconstructions

Scatterometers, unlike radiometers, have an incidence angle dependence, that is, the measurement depends on the angle at which the instrument looks at the ground. This incidence angle dependence is generally nonlinear, but over a certain range of incidence angles a linear model can be used for the backscatter measurements:

$$\sigma^\circ = \mathcal{A} + \mathcal{B}(\theta - \phi) \tag{6.1}$$

where  $\mathcal{A}$  represents the incidence angle-normalized backscatter coefficient, and  $\mathcal{B}$  represents the dependence of the backscatter on the incidence angle  $\theta$ .  $\phi$  is the angle to which  $\phi$  is normalized [7]. NSCAT reconstruction can be called *dual-variable* reconstruction because there are two parameters to estimate,  $\mathcal{A}$  and  $\mathcal{B}$ . More details

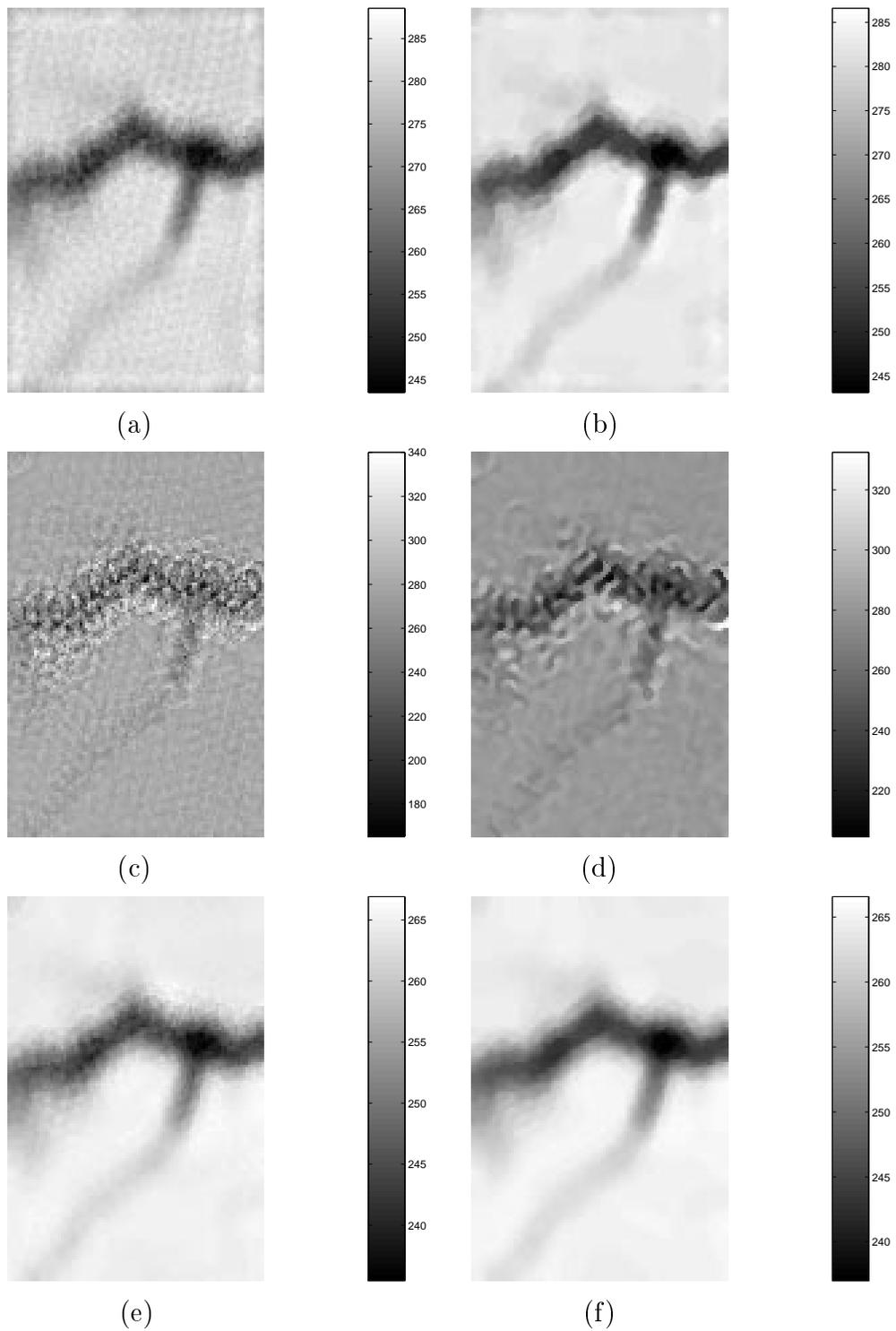


Figure 6.2: SSM/I reconstruction of real data using (a) MART—30 iterations (b) MARTF—40 iterations (c) ART—5 iterations (d) ARTF—15 iterations (e) SART—30 iterations (f) SARTF—30 iterations

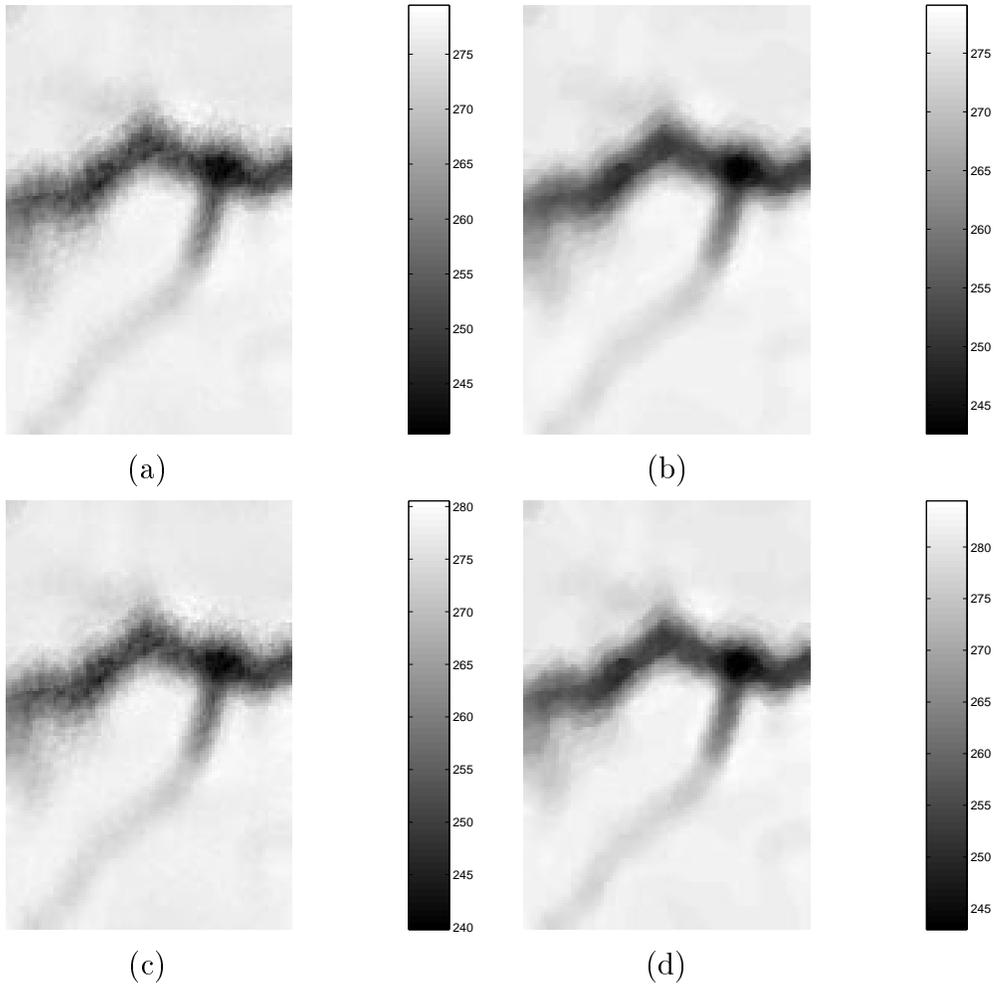


Figure 6.3: SSM/I reconstruction of real data using (a) SIR—10 iterations (b) SIRF—10 iterations (c) AddSIR—10 iterations (d) AddSIRF—15 iterations

on how to estimate  $\mathcal{B}$  are given in [7]. While there is information contained in the  $\mathcal{B}$  image, the primary resolution enhancement potential is in the  $\mathcal{A}$  image, so only these images are shown in the following comparisons.

Figures 6.4 and 6.5 show the results of using the several algorithms. As explained in Section 6.1, the choice of which iteration to use is subjective. Notice how much better ART and MART perform for the NSCAT data, as compared to the SSM/I data. One reason why ART and MART may perform better is that the sampling is denser (see Fig. 5.8) and the footprint is smaller. ART and MART may also perform better with NSCAT data is because of the shape of the footprint: The tapered SSM/I footprint has lower sidelobes, placing high frequency content closer to the noise floor and making it more difficult for ART and MART, which don't perform well in noise, to reconstruct.

### 6.3 Conclusions

This section compared the reconstruction results for different row action and column-normalized algorithms using real SSM/I and NSCAT data. While all of the algorithms performed well with the NSCAT data, the artifact suppression qualities of the column-normalized algorithms are necessary for good SSM/I reconstruction.

It should be remembered that this chapter is not meant to be an exhaustive comparison of the different algorithms. It remains as future research to do a more comprehensive comparison of the several algorithms, including optimizing the algorithms for different instruments.

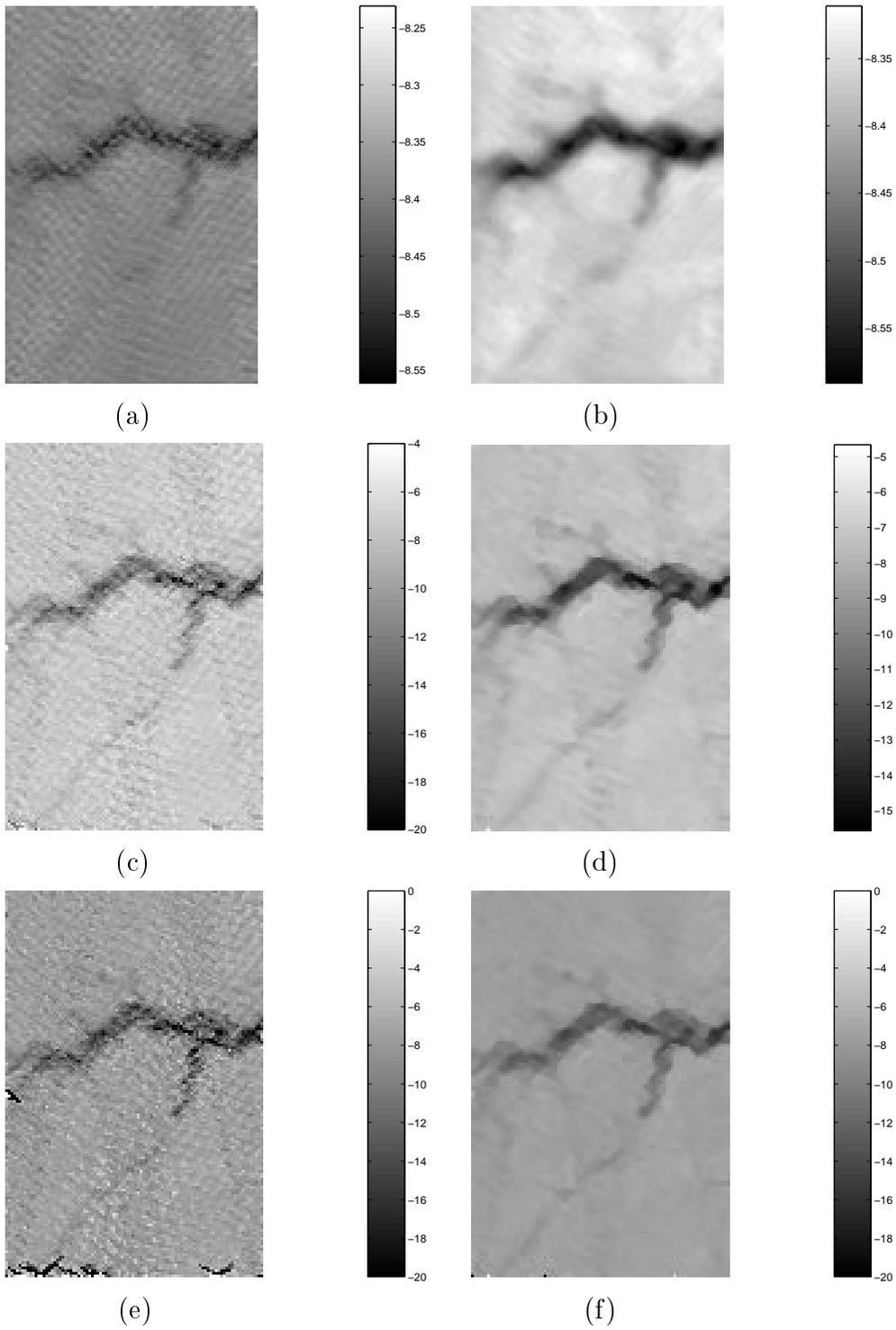


Figure 6.4: NSCAT reconstruction of real data using (a) MART—60 iterations (b) MARTF—10 iterations (c) ART—15 iterations (d) ARTF—15 iterations (e) SART—20 iterations (f) SARTF—20 iterations

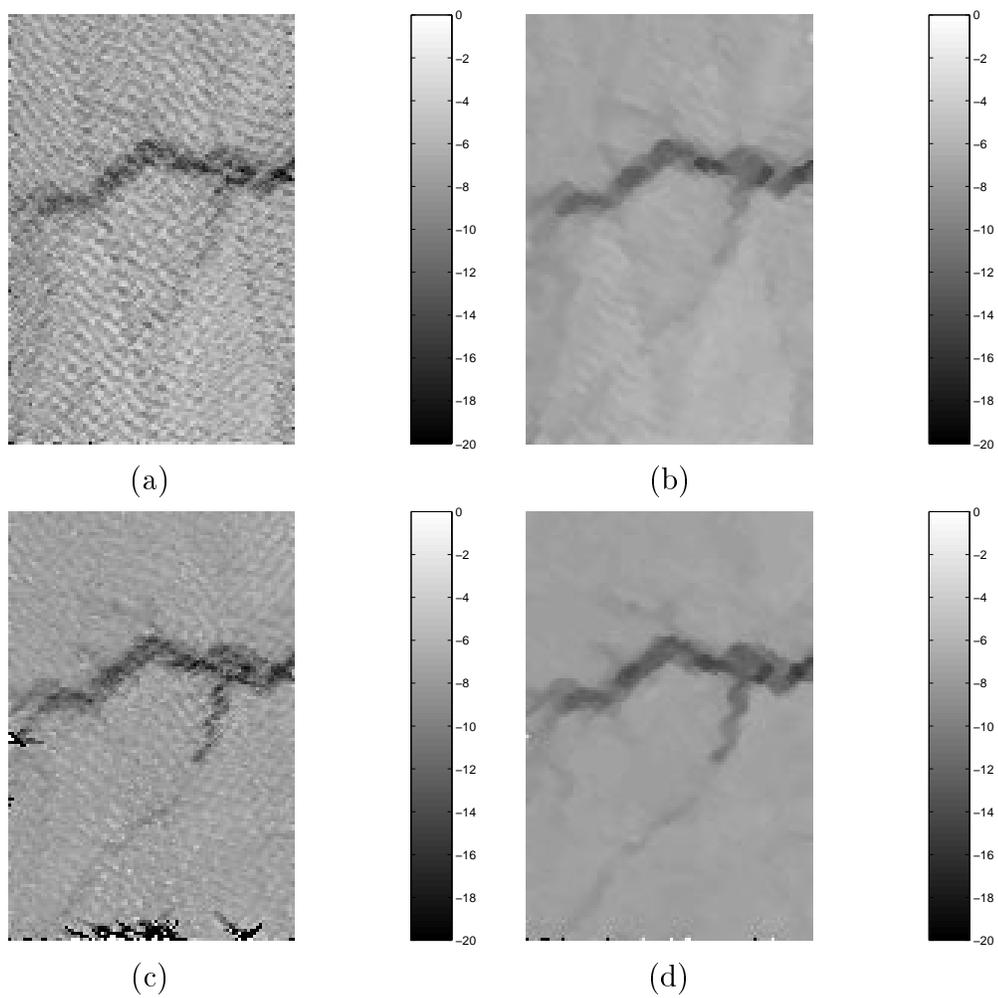


Figure 6.5: NSCAT reconstruction of real data using (a) SIR—30 iterations (b) SIRF—45 iterations (c) AddSIR—25 iterations (d) AddSIRF—30 iterations



## Chapter 7

### Conclusions

Image reconstruction algorithms have been used for several years to create enhanced resolution images of the earth's surface from radiometer and scatterometer data. These images have helped researchers monitor the planet's health. The BYU MERS lab has implemented the SIR algorithm with good results to create these enhanced resolution images. The SIR algorithm appeared to give better results than standard reconstruction algorithms like ART and MART, but it was not previously understood why.

This thesis has closely examined and compared several image reconstruction algorithms, with the goal of gaining more insight into the SIR algorithm. As part of this examination, convergence properties of Maximum Entropy (ME) algorithms were studied, and the convergence proof for the MART algorithm given by Lent and Censor in [10] was expanded to include varying forms of ME algorithms. Because ME algorithms are an important class of reconstruction methods, several different ME algorithms were studied to compare both performance in the presence of noise and convergence rates. It was found that the three ME forms considered all had comparable noise performance, but the MART algorithm has the fastest stable convergence over the widest range of values, suggesting that it is a good and viable algorithm for ME reconstruction. A fast algorithm for ME approximation was also introduced. None of the forms of MART, old or new, were found to compare favorably with SIR, or SART, or other column-normalized algorithms when used with NSCAT or SSM/I data.

The study of various reconstruction algorithms led to defining a new class of algorithms termed column-normalized algorithms, in which the iterative update is normalized by the columns of the transfer matrix  $\mathcal{H}$ . The SART algorithm was studied as an example of a column-normalized algorithm and was found and proved to converge to a weighed minimum norm solution where the weighting is the sum of the columns of  $\mathcal{H}$ , which contain sampling information for each pixel. This normalization appears to reduce the artifacts resultant from irregularly sampled and undersampled data. While the specific solution to which the SIR algorithm theoretically converges to has not been found, SIR is a column-normalized algorithm, and the normalization makes the algorithm robust for remote sensing reconstruction.

Chapters 5 and 6 presented a short comparison of the algorithms for simulated and real data, respectively. From the results, it appears that column-normalized algorithms are more robust than ART and MART for highly underdetermined and noisy systems. While ART and MART appear to be viable remote sensing reconstruction algorithms for highly sampled and low noise systems, column-normalized algorithms offer a better, more robust, option for a wider range of systems. Further, the now well-understood SART algorithm yielded results very similar to those from SIR.

## 7.1 Future Research

Although the introduction of column-normalized algorithms gives more insight into the nature of SIR, and Appendix A shows some convergence properties of SIR, it still is unknown what solution SIR converges to, nor is there a formal proof of convergence. The step-size nonlinearity in SIR makes it difficult to analyze, and perhaps a more fruitful course is to study the convergence properties of the BYU MART algorithm first, and then apply those results to SIR.

Chapters 5 and 6 provide only a brief comparison of the different algorithms, and more research is required for a better comparison. Further research into using the SART algorithm for remote sensing image reconstruction is also needed. To date, only the SIR algorithm has been generally used for image reconstruction in the BYU MERS Lab. There are, however, advantages to using SART, including the fact that

the solution is known, and, it is computationally less demanding because of being an additive algorithm. Research will be required to optimize the SART algorithm for different microwave instruments and to compare its results with SIR to see if using SART is warranted.



## Appendix A

### BYU MART and SIR Convergence Properties

#### A.1 Convergence Properties

While no formal proof of the convergence of SIR exists, empirical evidence suggests that SIR converges to a unique value [8]. We note that at the point of convergence  $x^{k+1} = x^k = x$ . We assume that  $x$  is unipolar, that is, either  $x_j > 0 \forall j$  or  $x_j < 0 \forall j$ . We note that in the noise-free case, with convergence to the true value for all  $j$ ,

$$y_i = \langle h_i, x \rangle = \sum_{n=1}^N x_n h_{in}$$

when  $\mathcal{H}$  is normalized.

##### A.1.1 Linear Case: BYU MART

Assuming convergence of  $x^k$  to  $x$  has occurred for the linearized SIR or BYU MART algorithm, we obtain the following:

$$x_j = x_j \frac{\sum_{i=1}^M \left[ \frac{y_i}{\langle h_i, x \rangle} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}}. \quad (\text{A.1})$$

Then,

$$x_j \left( 1 - \frac{\sum_{i=1}^M \left[ \frac{y_i}{\langle h_i, x \rangle} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}} \right) = 0, \quad (\text{A.2})$$

which we can write as

$$x_j A_j = 0 \quad (\text{A.3})$$

where

$$A_j = 1 - \frac{\sum_{i=1}^M \left[ \frac{y_i}{\langle h_i, x \rangle} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}} = 1 - \frac{\sum_{i=1}^M \left[ \frac{y_i}{\sum_{n=1}^N x_n h_{in}} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}}. \quad (\text{A.4})$$

Since  $x \neq 0$ , for Eq. (A.3) to hold we must have  $A_j = 0 \forall j$  which implies that

$$\frac{\sum_{i=1}^M \left[ \frac{y_i}{\sum_{n=1}^N x_n h_{in}} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}} = 1 \quad \forall j. \quad (\text{A.5})$$

This is equivalent to

$$\sum_{i=1}^M \left[ \frac{y_i}{\langle h_i, x \rangle} \right]^\lambda h_{ij} = \sum_{i=1}^M h_{ij} \quad \forall j, \quad (\text{A.6})$$

or

$$\sum_{i=1}^M d_i^\lambda h_{ij} = \sum_{i=1}^M h_{ij} \quad \forall j \quad (\text{A.7})$$

if we substitute

$$d_i = \frac{y_i}{\langle h_i, x \rangle}. \quad (\text{A.8})$$

Making  $\lambda = 1$  and writing (A.7) in matrix form gives

$$\mathcal{H}^T d = \mathcal{H}^T \mathbf{1}. \quad (\text{A.9})$$

Because  $\mathcal{H}$  is generally underdetermined, then  $\mathcal{H}^T$  will be overdetermined, so that the only way Eq. (A.9) will be satisfied is if

$$d_i = 1 \forall i, \quad (\text{A.10})$$

which implies that  $\langle h_i, x \rangle = y_i \forall i$ , which is the desired convergence result.

### A.1.2 Non-Linear Case

The nonlinear case is much more difficult to analyze than the linear case because of the duo-case definition of the scale factor in Eq. (4.57). To analyze the non-linear case, we will make some simplifying assumptions that  $\lambda = 1$  (no damping). With this simplification, Eq. (4.57) can be written as

$$u_{ij}^k = \begin{cases} \left[ \frac{1}{2} \left( \frac{1}{f_i^k} - \frac{1}{y_i} \right) + \frac{f_i^k}{x_j^k y_i} \right]^{-1} & \frac{y_i}{f_i^k} \geq 1 \\ \left[ \frac{1}{2} (f_i^k - y_i) + \frac{x_j^k y_i}{f_i^k} \right] & \frac{y_i}{f_i^k} < 1. \end{cases} \quad (\text{A.11})$$

Note that both cases give the same result for  $y_i/f_i^k = 1$  so the equality case split can be adapted to have equality for either case. Assuming convergence,  $x^{k+1} = x^k = x$ , we can write,

$$u_{ij} = \begin{cases} \left[ \frac{1}{2} \left( \frac{1}{f_i} - \frac{1}{y_i} \right) + \frac{f_i}{x_j y_i} \right]^{-1} & \frac{y_i}{f_i} > 1 \\ \left[ \frac{1}{2} (f_i - y_i) + \frac{x_j y_i}{f_i} \right] & \frac{y_i}{f_i} \leq 1. \end{cases} \quad (\text{A.12})$$

#### Lower Branch

In order to have a tractable problem to analyze, we now assume that all  $y_i/f_i \leq 1$  for all  $j$  so that only the lower case occurs:

$$u_{ij} = \frac{1}{2} (f_i - y_i) + \frac{x_j y_i}{f_i}. \quad (\text{A.13})$$

Putting this into Eq. (4.55) at convergence we obtain

$$x_j = \frac{\sum_{i=1}^M \left[ \frac{1}{2} (f_i - y_i) + \frac{x_j y_i}{f_i} \right] h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.14})$$

$$= \frac{\frac{1}{2} \sum_{i=1}^M (f_i - y_i) h_{ij}}{\sum_{i=1}^M h_{ij}} + x_j \frac{\sum_{i=1}^M \frac{y_i}{f_i} h_{ij}}{\sum_{i=1}^M h_{ij}}. \quad (\text{A.15})$$

Thus,

$$x_j \left( 1 - \frac{\sum_{i=1}^M \frac{y_i}{f_i} h_{ij}}{\sum_{i=1}^M h_{ij}} \right) = \frac{\frac{1}{2} \sum_{i=1}^M (f_i - y_i) h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.16})$$

which we can express as

$$x_j A_j = B_j \quad (\text{A.17})$$

where

$$A_j = 1 - \frac{\sum_{i=1}^M \frac{y_i}{f_i} h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.18})$$

$$B_j = \frac{\frac{1}{2} \sum_{i=1}^M (f_i - y_i) h_{ij}}{\sum_{i=1}^M h_{ij}}. \quad (\text{A.19})$$

Keeping the  $\lambda$  it can be shown that

$$A_j = 1 - \frac{\sum_{i=1}^M \left[ \frac{y_i}{f_i} \right]^\lambda h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.20})$$

$$= \frac{\sum_{i=1}^M \left( 1 - \left[ \frac{y_i}{f_i} \right]^\lambda \right) h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.21})$$

$$B_j = \frac{\frac{1}{2} \sum_{i=1}^M f_i \left( 1 - \left[ \frac{y_i}{f_i} \right]^\lambda \right) h_{ij}}{\sum_{i=1}^M h_{ij}}. \quad (\text{A.22})$$

We note that in the linear SIR case, we obtained the solution equation  $x_j A_j = 0$  [Eq. (A.3)] while in the non-linear SIR we have  $x_j A_j = B_j$  [Eq. (A.17)]. This linear

case requires that  $A_j = 0 \forall j$ . A similar condition for the non-linear case is more difficult to arrive at.

First, we note that in the noise-free (desired) case,  $f_i = y_i$ . Then,  $B_j = 0$  and  $A_j = 0$ , satisfying Eq. (A.3). Thus, non-linear SIR has a solution point at the desired solution which matches linear SIR. However, a solution at  $x_j = B_j/A_j$  may also be possible. Note that (for  $\lambda = 1$ )

$$x_j = \frac{B_j}{A_j} = \frac{\frac{1}{2} \sum_{i=1}^M (f_i - y_i) h_{ij}}{\sum_{i=1}^M h_{ij} - \sum_{i=1}^M \frac{y_i}{f_i} h_{ij}} \quad (\text{A.23})$$

$$= \frac{\sum_{i=1}^M (f_i - y_i) h_{ij}}{2 \sum_{i=1}^M \left( \frac{f_i - y_i}{f_i} \right) h_{ij}}, \quad (\text{A.24})$$

i.e.,  $B_j/A_j$  is the ratio of the point spread weighted normalized projection difference and the point spread weighted projection difference. For a general  $\lambda$

$$\frac{B_j}{A_j} = \frac{1 \sum_{i=1}^M f_i \left( 1 - \left[ \frac{y_i}{f_i} \right]^\lambda \right) h_{ij}}{2 \sum_{i=1}^M \left( 1 - \left[ \frac{y_i}{f_i} \right]^\lambda \right) h_{ij}}. \quad (\text{A.25})$$

In the ideal, noise-free case, both  $A_j$  and  $B_j$  are zero, so the ratio is undefined. Using L'Hospital's rule we can compute the ratio in the limit. Note that to do this  $y_i$  and  $h_{ij}$  are fixed and that  $f_i$  ( $= \langle h_i, x \rangle$ ) is converging to  $y_i$ . We thus take the derivatives with respect to  $f_i$ . The algebra can be simplified by defining a vector  $F$  with elements  $f_i$ . Then

$$\frac{dA_j}{dF} = \sum_{i=1}^M \lambda y_i^\lambda f_i^{-(\lambda+1)} h_{ij} \quad (\text{A.26})$$

$$= \sum_{i=1}^M D_{ij} \quad (\text{A.27})$$

$$\frac{dB_j}{dF} = \sum_{i=1}^M \left( 1 - \left[ \frac{y_i}{f_i} \right]^\lambda + f_i \lambda y_i^\lambda f_i^{-(\lambda+1)} \right) h_{ij} \quad (\text{A.28})$$

$$= \sum_{i=1}^M f_i D_{ij} + \sum_{i=1}^M \left(1 - \left[\frac{y_i}{f_i}\right]^\lambda\right) h_{ij} \quad (\text{A.29})$$

$$= \sum_{i=1}^M f_i D_{ij} + A_j \quad (\text{A.30})$$

where

$$D_{ij} = \lambda y_i^\lambda f_i^{-(\lambda+1)} h_{ij}. \quad (\text{A.31})$$

The solution for  $x_j = B_j/A_j$  when  $f_i = y_i$  is then

$$x_j = \frac{B_j}{A_j} = \frac{\sum_{i=1}^M f_i D_{ij} + A_j}{\frac{dA_j}{dF}} \quad (\text{A.32})$$

$$= \frac{\frac{1}{2} \sum_{i=1}^M \lambda \left[\frac{y_i}{f_i}\right]^\lambda h_{ij} + \sum_{i=1}^M \left(1 - \left[\frac{y_i}{f_i}\right]^\lambda\right) h_{ij}}{\sum_{i=1}^M \lambda \frac{1}{f_i} \left[\frac{y_i}{f_i}\right]^\lambda h_{ij}}. \quad (\text{A.33})$$

Letting  $f_i = y_i$  in this expression and simplifying, we obtain

$$x_j = \frac{\frac{1}{2} \sum_{i=1}^M \lambda h_{ij}}{\sum_{i=1}^M \frac{\lambda}{f_i} h_{ij}} \quad (\text{A.34})$$

$$= \frac{\frac{1}{2} \sum_{i=1}^M \lambda h_{ij}}{\sum_{i=1}^M \frac{\lambda}{y_i} h_{ij}} \quad (\text{A.35})$$

$$= \frac{\frac{1}{2} \sum_{i=1}^M h_{ij}}{\sum_{i=1}^M \frac{h_{ij}}{y_i}}. \quad (\text{A.36})$$

This is a key result. While it initially may suggest that it may be able to compute the converged solution without iteration, in fact this result suggests that the solution must take this form. Since, in general, it will not, the solution may be biased or the

analysis may be flawed. We also note that the value of  $\lambda$  does not affect the final solution.

## Upper Branch

We considered the lowercase for  $u_{ij}$  in Eq. (A.12). We now consider the other case, that is, when  $y_i/f_i \geq 0$  for all  $j$ . In this case (using  $\lambda = 1$ )

$$u_{ij} = \left[ \frac{1}{2} \left( \frac{1}{f_i} - \frac{1}{y_i} \right) + \frac{f_i}{x_j y_i} \right]^{-1} \quad (\text{A.37})$$

which can be written as

$$u_{ij} = \frac{x_j y_i}{f_i + \frac{1}{2} \frac{x_j y_i}{f_i} - \frac{1}{2} x_j} \quad (\text{A.38})$$

$$= \frac{x_j y_i}{f_i + C_i x_j} \quad (\text{A.39})$$

where

$$C_i = \frac{1}{2} \left( \frac{y_i}{f_i} - 1 \right) = \frac{1}{2} \frac{y_i - f_i}{f_i}. \quad (\text{A.40})$$

In the ideal normalized case,  $y_i = f_i$  so  $C_i = 0$ . We can then write

$$u_{ij} = \frac{x_j y_i}{f_i} = x_j d_i. \quad (\text{A.41})$$

The analysis from this point is identical to the linearized SIR equation. It follows that the non-linear case has a solution at the desired point.

If we plug Eq. (A.39) into Eq. (4.55) we obtain at convergence

$$x_j = x_j \frac{\sum_{i=1}^M \frac{y_i}{f_i + C_i x_j} h_{ij}}{\sum_{i=1}^M h_{ij}} \quad (\text{A.42})$$

which implies that convergence occurs at

$$\frac{\sum_{i=1}^M \frac{y_i}{f_i + C_i x_j} h_{ij}}{\sum_{i=1}^M h_{ij}} = 1 \quad \forall j \quad (\text{A.43})$$

which we can write as

$$\sum_{i=1}^M \left(1 - \frac{y_i}{f_i + C_i x_j}\right) h_{ij} = 0 \quad \forall j. \quad (\text{A.44})$$

In the ideal normalized case  $C_i = 0$  so this becomes

$$\sum_{i=1}^M \left(1 - \frac{y_i}{f_i}\right) h_{ij} = 0 \quad \forall j. \quad (\text{A.45})$$

We note that this is the same condition at convergence that BYU MART meets.

## A.2 Conclusions

Using simple analysis techniques, we have shown that

1. The BYU MART algorithm has a convergent point at a solution compatible with the constraint  $y = \mathcal{H}x$ .
2. The damping factor  $\lambda$  used in the algorithm does not affect the ideal solution.
3. The SIR algorithm has a solution point compatible with the constraint  $y = \mathcal{H}x$ .
4. Other convergent but biased solutions may be possible.

## A.3 Post Script

In Section 4.2 it was shown how one could “work backwards” from an algorithm to find what kind of solution it converges to. The following paragraphs show how this was tried without success:

Remember that the update in matrix form is

$$x_j^{k+1} = x_j^k \left( \frac{(H^T d^k)_j}{H_j^T \mathbf{1}} \right) \quad \text{where} \quad d_i^k = \left( \frac{y_i}{\langle h_i, x^k \rangle} \right)^\lambda. \quad (\text{A.46})$$

We need an assumption about the dual variable update. Here we notice one of the differences between MART and SART, and BYU MART. One way for BYU MART to converge would be for  $d^k$  to go to one, not zero like MART. Thus the dual variable

update would be

$$u^{k+1} = u^k \odot d^k \tag{A.47}$$

where  $\odot$  denotes an element-by-element Schur product.

Substituting A.47 into A.46 gives

$$x_j^{k+1} = x_j^k \left( \frac{H_j^T(u^{k+1} ./ u^k)}{H_j^T \mathbf{1}} \right) \tag{A.48}$$

$\Rightarrow$

$$= x_j^0 \prod_{l=0}^k \left( \frac{H_j^T(u^{l+1} ./ u^l)}{H_j^T \mathbf{1}} \right) \tag{A.49}$$

where  $./$  denotes element-by-element division. There does not appear to be a way (to the author) to reduce this to be just in terms of  $x^0$  and  $u^0$  like with MART. The difficulty lies in the summation over all of the measurements, which above is represented by matrix multiplication. Several other forms of Eq. (A.47) were examined, but to no avail. It still remains an area of future research to see what kind of solutions the BYU MART and SIR algorithms converge to.



## Bibliography

- [1] J.L. Alvarez-Perez, S.J. Marshall, and K. Gregson, “Resolution improvement of ERS scatterometer data over land by Weiner filtering”, *Remote Sens. Environ.*, vol. 71, pp. 261–271, 2000.
- [2] R.A. Crowther, L.A. Amos, J.T. Finch, D.J. DeRosier, and A. Klug, “Three dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs”, *Nature*, vol. 226, pp. 421, 1970.
- [3] D.J. DeRosier and A. Klug, “Reconstruction of three dimensional structures from electron micrographs”, *Nature*, vol. 217, pp. 130, 1968.
- [4] R.M. Lewitt, “Reconstruction algorithms: Transform methods”, *Proc. IEEE*, vol. 71, pp. 390–408, March 1983.
- [5] R. Gordon, R. Bender, and G.T. Herman, “Algebraic reconstruction technique (ART) for three-dimensional electron microscopy and x-ray photography”, *J. Theoret. Biol.*, vol. 29, pp. 471–481, 1970.
- [6] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections”, *J. Theoret. Biol.*, vol. 36, pp. 105–117, 1972.
- [7] P. Whiting, “Resolution enhancement of SEASAT scatterometer data”, Master’s thesis, Brigham Young University, 1992.
- [8] D. Early, *A Study of the Scatterometer Image Reconstruction Algorithm and its Applications to Polar Ice Studies*, PhD thesis, Brigham Young University, May 1998.

- [9] D. Early and D.G. Long, “Image reconstruction and enhanced resolution imaging from irregular samples”, *submitted to TGARRS*, 2000.
- [10] A. Lent and Y. Censor, “The primal-dual algorithm as a constraint-set-manipulation device”, *Mathematical Programming*, 1991.
- [11] A.H. Andersen and A.C. Kak, “Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm”, *Ultrasonic Imaging*, 1984.
- [12] F. Ulaby, R. Moore, and A. Fung, *Microwave Earth Remote Sensing*, Artech House, 1981.
- [13] Y. Censor, “Row-action methods for huge and sparse systems and their applications”, *SIAM Review*, 1981.
- [14] G.T. Herman and A. Lent, “ART, mathematics and applications”, *J. Theo. Biol.*, vol. 42, pp. 24–32, 1973.
- [15] S.F. Gull and G.J. Daniell, “Image reconstruction from incomplete and noisy data”, *Nature*, 1978.
- [16] D.G. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1969.
- [17] Y. Censor and J. Segman, “On block-iterative entropy maximization”, *Journal of Information and Optimization Sciences*, 1987.
- [18] T. Elfving, “On some methods for entropy maximization and matrix scaling”, *Linear Algebra and its Applications*, 1980.
- [19] R.L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers, 1991.
- [20] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.
- [21] A.H. Andersen, “Algebraic reconstruction in CT from limited views”, *IEEE Trans. on Med. Img.*, vol. 8, no. 1, pp. 50–55, March 1989.

- [22] D.G. Long and D.L. Daum, “Spatial resolution enhancement of SSM/I data”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 407–417, 1998.
- [23] D.G. Long, P.J. Harding, and P. Whiting, “Resolution enhancement of spaceborne scatterometer data”, *IEEE Transactions of Geoscience and Remote Sensing*, vol. 31, pp. 700–715, May 1993.