June 2006 **LIDS Publication # 2705**

Research supported in part by:

National Science Foundation Graduate Research Fellowship Program, Air Force Research Laboratory Grants FA8650-04-1-1719 and FA8650-04-C-1703

Joint Anisotropy Characterization and Image Formation in Wide-Angle Synthetic Aperture Radar

Kush R. Varshney

Joint Anisotropy Characterization and Image Formation in Wide-Angle Synthetic Aperture Radar

by

Kush R. Varshney

B.S., Electrical and Computer Engineering
Cornell University, 2004

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

Author . . . .
Department of Electrical Engineering and Computer Science
May 10, 2006

Certified by . . . .
Müjdat Çetin
Research Scientist
Thesis Supervisor

Certified by . . . .
John W. Fisher III
Principal Research Scientist
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Joint Anisotropy Characterization and Image Formation in Wide-Angle Synthetic Aperture Radar

by

Kush R. Varshney

Submitted to the Department of Electrical Engineering and Computer Science
on May 10, 2006, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Imagery formed from wide-angle synthetic aperture radar (SAR) measurements has fine cross-range resolution in principle. However, conventional SAR image formation techniques assume isotropic scattering, which is not valid with wide-angle apertures. Also, the spatial location of scattering centers may migrate as a function of viewing angle across the aperture. The problem of jointly forming images and characterizing anisotropy as well as characterizing scattering center migration in wide-angle SAR is considered in the thesis. The approach not only compensates for anisotropy and migration in the image formation process, but gives much more information, useful for scene interpretation, than a simple image would.

A method based on a sparse representation of anisotropic scattering with an overcomplete basis composed of basis vectors with varying levels of angular persistence is presented. Solved as an inverse problem, the result is a complex-valued, aspect-dependent response for each scatterer in a scene. The non-parametric approach jointly considers all scatterers within one system of equations. The choice of the overcomplete basis set incorporates prior knowledge of anisotropy, but the method is flexible enough to admit solutions that may not match a family of parametric functions. Sparsity is enforced through regularization based on the ℓ_p quasi-norm, $p < 1$, leading to a non-convex minimization problem. A quasi-Newton method is applied to the problem and a novel greedy graph-structured algorithm is developed and also applied. Results are demonstrated on synthetic examples and realistic examples with XPatch data, including the backhoe public release dataset.

Thesis Supervisor: Müjdat Çetin
Title: Research Scientist

Thesis Supervisor: John W. Fisher III
Title: Principal Research Scientist

Acknowledgments

> मातापित्रोर्गुरूणां च पूजा बहुमता मम ।
> अत्र युक्तो नरो लोकान्यशश्च महदश्नुते ॥ ३॥
>
> एत एव त्रयो लोका एत एवाश्रमास्त्रयः ।
> एत एव त्रयो वेदा एत एव त्रयोऽग्नयः ॥ ६॥
>
> — भीष्म उवाच, अध्याय १०९, शांतिपर्वं, महाभारत

I thank all the innumerable people who have helped me reach this point. Without the support of so many well-wishers, I don't know where I would be.

I thank all of my teachers, past and present, for shaping my intellectual outlook and equipping me with the skills that have gone into writing this thesis. I would especially like to thank Müjdat Çetin, John Fisher, and Alan Willsky. Müjdat is the source of many of the ideas in the work. Whether eight meters or eight thousand kilometers away, he has been a superb advisor and mentor. This thesis really would not have been possible without his steadfast efforts. John has given good guidance throughout the process and provided a different perspective on the research. I am grateful to him for filling in the void as thesis advisor. Alan has been an invaluable resource, always amazing me with his ability to grasp, process, and expand upon things I am discussing even before the ideas have completely formed in my mind, let alone left my mouth.

Others I would like to acknowledge for their assistance are Randy Moses and Rajan Bhalla, who furnished ideas, data, and code. I also acknowledge the financial support of the National Science Foundation Graduate Research Fellowship Program.

It has been an honor interacting with the fellow students of the Stochastic Systems Group every day. Also, friends, roommates, and others have been indispensable. Taking my mind out of the game some of the time, they have allowed me to be productive the rest of the time. Thank you. I would also like to thank Al Oppenheim, my graduate counselor, for going way above and beyond the call.

I would like to thank my family all over the world for wishing me the best. I thank my brother, Lav, for supporting me and pushing me every step of the way. Finally, I would like to thank my parents, to whom I dedicate this thesis. Thank you for always being there.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Mathematical problems are formulated and motivated by specific classes of applications and by the methods of solution that one already knows. Throughout this process, questions are asked: How can we extend existing mathematical methodologies? How can we use existing methodologies in the context of a specific physical problem to obtain a tractable formulation which addresses the issues of interest in the more ill-defined physical problem?
>
> — *Alan S. Willsky (1982), in IEEE Control Systems Magazine*

The wealth of human experience begins with our senses. Whether telescopes for observing phenomena too far to see, microscopes for phenomena too small, or stroboscopes for phenomena too fast, we have developed technologies to extend our sensory capabilities. Radar, an acronym for ‘radio detection and ranging,’ which allows observation at night-time, as it actively produces its own energy, and allows observations through obstacles, such as clouds and foliage, and at great distances, due to the propagation properties of electromagnetic waves, is another such technology. In this thesis, advanced image formation techniques are developed for synthetic aperture radar, further enriching our ability to sense.

1.1 Overview

The principles underlying radar have been known since the development of electromagnetic theory in the nineteenth century [39]. However, development of radar technology began in earnest in the 1920s [38]. In its basic operation, radar works by transmitting pulses of electromagnetic radiation. If a pulse encounters an object, it is reflected, or in radar terminology, is *scattered*. Scattered pulses can then be measured to detect objects and determine how far they are, based on the time difference between transmission and reception, and the known speed of propagation. This distance is known as range, hence the phrase radio detection and ranging. Conventional radar is limited in its ability to provide information in the direction perpendicular to the direction of range measurements, known as the cross-range direction. Synthetic aperture radar (SAR), first proposed in the 1950s, collects measurements that allow the formation of ground images, range vs. cross-range.

SAR synthetically forms a long antenna, or aperture, using pulses transmitted and received over a distance. The physical layout of the SAR imaging system uses a radar set mounted on an aircraft in flight. As the aircraft proceeds along its flight path, it transmits radar pulses towards the ground and receives the reflected pulses. The length of the flight path, the length of the synthetic aperture, is inversely proportional to resolution in the cross-range direction of formed images: the longer the aperture, the finer the resolution. In spotlight-mode SAR, the radar is steered so that it always looks towards a single patch on the ground from every point along its flight path. Accurate knowledge of aircraft location is essential for SAR image formation.

Advances in navigation technologies and avionics over the past several years now allow the synthesis of longer apertures than ever before, and consequently more finely resolved images. A problem that arises with long, or wide-angle, apertures is that dependence of scattering behavior on aspect angle, termed *anisotropy*, becomes prominent because objects are viewed from different sides rather than from nearly the same point of view. For example, a mirror or flat metal sheet may reflect strongly when viewed straight on, but barely reflect at all from an oblique angle. This is in opposition to narrow-angle imaging, where it is a fairly reasonable assumption that scattering amplitude is constant over the aperture. Conventional SAR image formation techniques implicitly make this isotropic scattering assumption. Characterized anisotropy may be used as a feature for automatic target recognition, such as to discriminate an isotropic tree trunk, reflecting the same from all directions, and the anisotropic flat metal side of a truck. It may also be used for improved image formation. Therefore, for effective use of data from wide-angle apertures, it is of interest to develop data processing techniques that characterize anisotropy and account for it in image formation.

1.2 Contributions

This thesis puts forth an inverse problem formulation utilizing an overcomplete basis and sparsifying regularization for joint image formation and anisotropy characterization in wide-angle SAR. Sparsifying regularization has been applied to inverse problems including acoustic source localization [36] and isotropic SAR imaging [10], but has not previously been applied to the SAR anisotropy characterization problem. One general principle in previous work on the anisotropy problem has been to divide the full wide-angle aperture into smaller subapertures and form a sequence of subaperture images with inherently reduced cross-range resolution for use in further processing [2, 13, 23, 50, 32, 34, 22, 41, 12]. Another general principle has been to develop parametric models for angle-dependent scattering behavior [47, 54, 26, 25, 17]. The proposed methodology does not suffer a reduction in resolution because the entire available aperture is used and is more flexible than parametric models. The proposed framework solves for multiple spatial locations jointly, ameliorating the ill-effects of close proximity neighboring scatterers [32].

In wide-angle imaging, besides anisotropy becoming prominent, certain scattering mechanisms appear to migrate in their spatial location as a function of aspect angle. This type of scattering, which has not been given much heed in past work, is well-incorporated into the overcomplete basis formulation.

A graph-structured interpretation leading towards novel approximate algorithms to solve the inverse problem is developed in the thesis. These algorithms, having reduced memory requirements, may well find application in a wide variety of sparse signal representation settings beyond the specific problem of anisotropy in SAR.

The interrelationship between points, as parts of larger objects, is broached in the penultimate chapter. Through extensions to the sparsifying regularization cost function, certain object-level preferences are encoded within the image formation process.

Overall, this thesis aims to take the application wide-angle SAR imaging and formulate it in the mathematics of inverse problems and sparse signal representation. In this process, previously unaccounted for physical phenomena are dealt with, new methods of solution are developed, and the enhancement of physically interesting or meaningful objects is explored.

1.3 Organization of Thesis

The contribution of this thesis is situated within the SAR community; the first part of Chapter 2 is a welcome basket of sorts — a primer on the application area. The geometry of spotlight-mode SAR data collection is described along with different domains in which measurements may be viewed. The discussion is from the perspective of tomography, avoiding the specialized terminology of radar [42]. The point-scatterer model of radar, a key premise in this work, is given next. Also, data collected over wide-angle apertures are contrasted with data collected over narrow-angle apertures, the main point being that anisotropy is prominent in wide-angle imaging. The second part of Chapter 2 is a figurative backhoe loader. In preparation for the remainder of the thesis, important nuggets from the related methodologies of sparse signal representation and inverse problems are extracted. The section concludes with the pouring of the main mathematical foundation for the thesis: a sparsifying regularization approach to solving inverse problems is discussed. The third part of Chapter 2 discusses prior art, summarizing and discussing previous work done on the characterization of anisotropic scattering, including parametric methods and image-domain methods.

Chapter 3 lays out the blueprint and erects the framework for bringing the SAR image formation and anisotropy characterization application together with the inverse problem–sparsity mathematical formalism. Specifically, an overcomplete expansion of the point-scattering observation model is proposed, along with a discussion on the choice of vectors for the expansion. Chapter 4 builds upon the framework of the previous chapter, describing methods of solving the inverse problem while imposing sparsity. First, the quasi-Newton method of [9] is applied to the problem; examples are presented with synthetic data and with scenes containing realistic canonical scatterers. A nonlinear least-squares method for characterizing migratory scattering is developed and examples are given as well. Also, greedy algorithms are developed based on a graph-structured interpretation of the overcomplete expansion. Examples with synthetic data, with scenes composed of realistic canonical scatterers, and with a scene containing a literal backhoe loader are given.

Chapter 5 is concerned with practical issues in the methods of solution of Chapter 4, such as parameter selection, preprocessing, and performance in the presence of noise. Extensions to the framework are proposed in Chapter 6. Specifically, variations are introduced that, in addition to sparsity in the overcomplete expansion, put object-level preferences into the image formation and anisotropy characterization. Also, the method of characterizing migratory scattering is generalized. The final chapter provides a brief summary and articulates future directions of research.

Chapter 2

Background

> Bernard of Chartres used to say that we were like dwarfs on the shoulders of giants, so that we can see more than they, and things at a greater distance, not by virtue of any sharpness of sight on our part, or any physical distinction, but because we are carried high and raised by their giant size.
>
> — *John of Salisbury (1159), in Metalogicon*[1]

Both the application area and class of mathematical methodologies investigated in this thesis are vast as general disciplines. Consequently, a deep coverage of background material is not attempted. It is hoped that this chapter at least pecks a little hole into that vast realm.

2.1 Spotlight-Mode Synthetic Aperture Radar

Synthetic aperture radar is often explained in specialized radar terminology such as fast time, slow time, Doppler shifts, etc. Spotlight-mode SAR, however, also has an interpretation as a tomographic observation process [42], and can be modeled similarly to other imaging modalities, such as magnetic resonance imaging, acoustic microscopy, and positron emission tomography. The rudiments of SAR observation are presented below via the tomographic interpretation; a comprehensive non-tomographic treatment may be found in [8], among many others.

---

[1] Translation from Oxford Dictionary of Quotations.

2.1.1 Observation Geometry

As mentioned in Section 1.1, SAR uses a radar set mounted on an aircraft to collect measurements. The same antenna is used for both transmission and reception of electromagnetic pulses. From one point along the aircraft's flight path, the radar transmits a pulse in a certain direction, illuminating a certain portion of the ground — the ground patch, and receives back the scattered pulse, which depends on the characteristics of the ground patch. Pulses are similarly transmitted and received at many points along the flight path, with the direction of illumination varying depending on the mode of operation.

SAR has distinct modes of operation: stripmap, spotlight, and scan, distinguished by the illumination direction. In stripmap-mode, the radar antenna remains fixed with respect to the flight path, illuminating a sequence of ground patches that form a strip parallel to the flight path. In spotlight-mode, the focus of this thesis, the radar antenna continually changes its look direction to always illuminate the same ground patch. Scan-mode is a generalization with stripmap-mode and spotlight-mode as special cases that allows the look direction and illuminated ground patch to both vary.

The geometry of data collection in spotlight-mode SAR is illustrated in Fig. 2-1. Coordinates on the ground plane x, range, and y, cross-range, are centered in the ground patch. Measurements are taken at equally spaced aspect angles θ as the aircraft traverses the flight path. The aircraft is at an altitude, and consequently forms an elevation angle with the ground plane. The work in this thesis is concerned with two-dimensional imaging; thus, the three-dimensional geometry is neglected. The geometry reduces to the ground plane geometry shown in Fig. 2-2. The ground patch, with radius L, is shaded.

2.1.2 Data Domains

The scattering from the ground patch under observation is manifested as an amplitude scaling and phase shift that can be expressed as a complex number at each point.

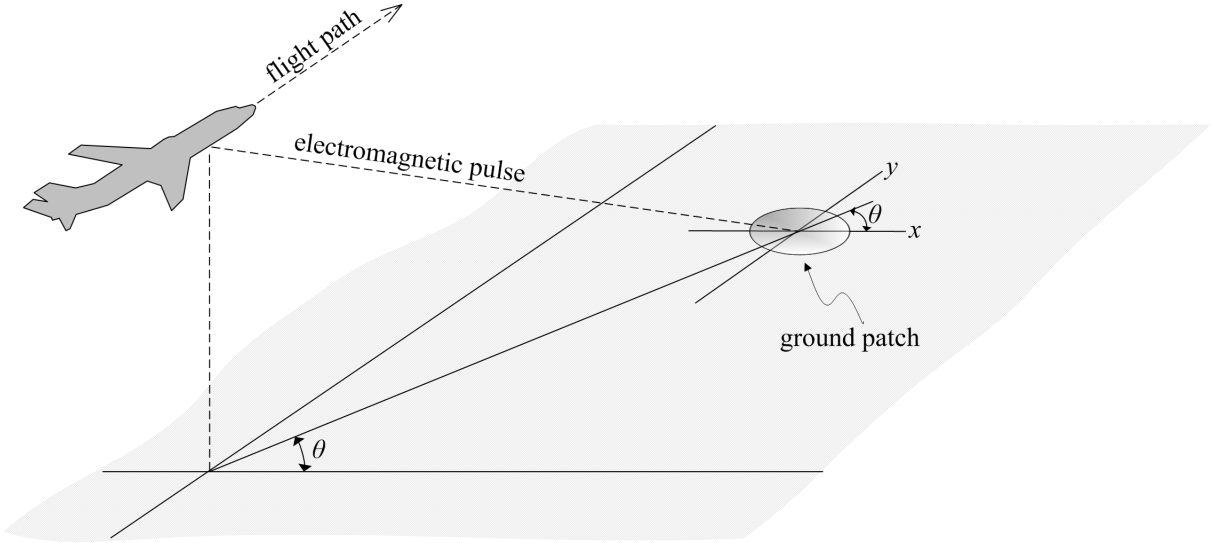


Figure 2-1: Spotlight-mode SAR observation geometry.

Thus, scattering from the entire ground patch can be characterized by a complex-valued function of two spatial variables $s(x, y)$, which is referred to in some instances as the *scattering function* and in others as the *image*. Due to the design of the radar pulse and the physics of the observation process, the collection of received signals is not $s(x, y)$ directly. Procedures for obtaining the image from the measurements are known as image formation. The relationship between the measurements, obtained over a finite bandwidth of frequencies and over a range of aspect angles, and the scattering function is given by:

$$r(f, \theta) = \iint_{x^2+y^2 \leq L^2} s\left(x, y\right) \exp\left\{-j\frac{4\pi f}{c}\left(x\cos\theta + y\sin\theta\right)\right\} dx\, dy. \tag{2.1}$$

The set of aspect angles θ is inherently discrete, because pulses are transmitted from a discrete set of points along the flight path. The measurements are sampled in frequency f to allow digital processing. The collection of measurements $r(f, \theta)$ is known as the *phase history*. A derivation of the relationship between the scattering function and phase history is delayed to Sec. 2.1.3, in which a derivation is given for the point-scatterer model. An example of a phase history is given in Fig. 2-3, which shows $|r(f, \theta)|$ for frequencies from approximately 7 GHz to 13 GHz and aspect angles

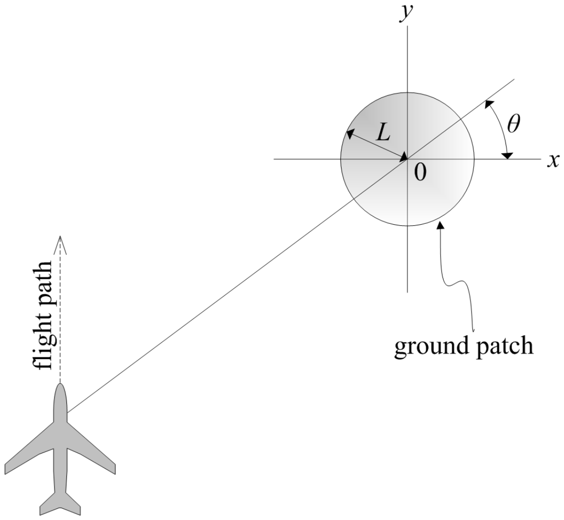


Figure 2-2: Ground plane geometry in spotlight-mode SAR.

from -55° to $+55^\circ$ for a particular scene. The shading is in decibels.

Equation (2.1) may be recognized as a Fourier-like transform. Let us investigate this in more detail. Introducing an auxiliary variable ρ and using the sifting property of the Dirac delta, the phase history expression is rewritten as:

$$r(f,\theta) = \int\limits_{|\rho|\le L} \iint\limits_{x^2+y^2\le L^2} \delta\left(\rho - x\cos\theta - y\sin\theta\right) s\left(x,y\right) \exp\left\{-j\frac{4\pi f}{c}\rho\right\} dx\,dy\,d\rho. \quad (2.2)$$

The projection of the scattering function in the range direction at aspect angle θ, parameterized by ρ is:

$$\hat{R}(\rho,\theta) = \iint\limits_{x^2+y^2\le L^2} \delta\left(\rho - x\cos\theta - y\sin\theta\right) s\left(x,y\right) dx\,dy. \quad (2.3)$$

These projections are also known as range profiles. Thus, for each aspect angle the phase history is the one-dimensional Fourier transform of the range profile at that angle:

$$r(f,\theta) = \int_{-L}^{L} \hat{R}(\rho,\theta) \exp\left\{-j\frac{4\pi f}{c}\rho\right\} d\rho. \quad (2.4)$$

This result is a consequence of the projection slice theorem [7]. Since the measure-

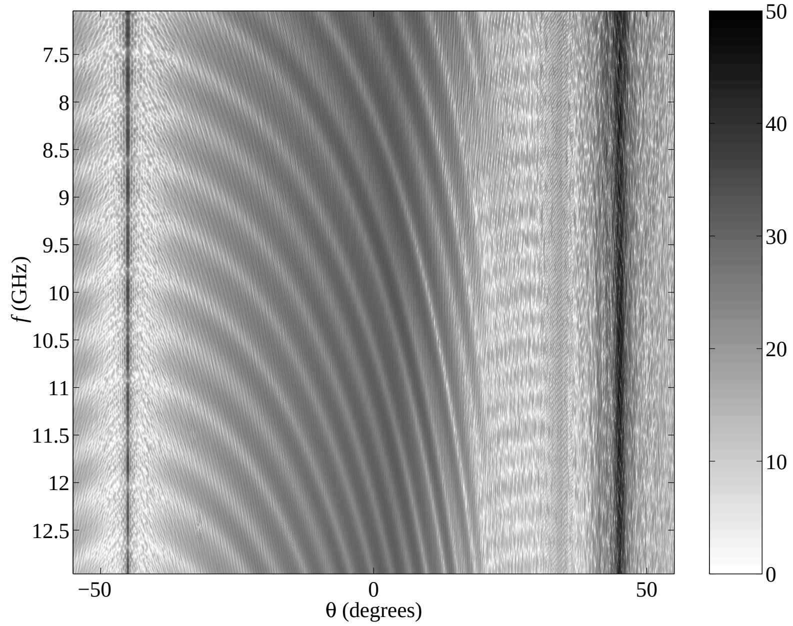


Figure 2-3: A typical wide-angle SAR phase history with large bandwidth.

ments are limited to a band in frequency, performing one-dimensional inverse transforms results in range profiles convolved with the inverse transform of the rectangle function, the periodic sinc or Dirichlet function. Also, because the measurements are sampled, the inverse transform used is the inverse discrete Fourier transform calculated using a fast algorithm such as the fast Fourier transform (FFT). As an example, the collection of range profile magnitudes for the phase history of Fig. 2-3 is shown in Fig. 2-4. The shading in this figure is also in decibels.

The phase history expression (2.1) is the two-dimensional Fourier transform of the scattering function, but in a special way. The phase history measurements for a single aspect angle form a radial slice in the Fourier domain. The collection of all measurements for all aspect angles traces out a keystone-shaped region in the Fourier plane, as illustrated in Fig. 2-5. Note that the region of support shown in Fig. 2-5 does not correspond to the data of Figs. 2-3, 2-4, and 2-6. Due to the samples lying on a polar lattice, difficulties in inverting this two-dimensional Fourier transform

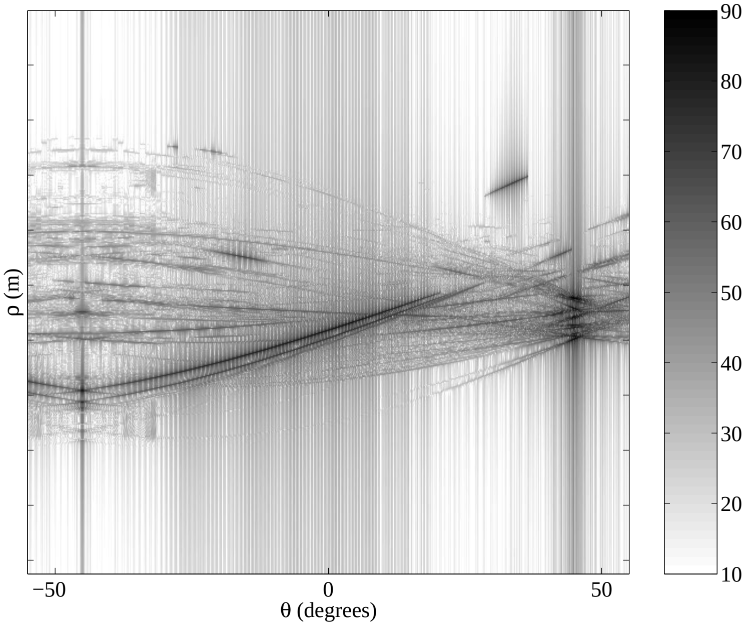


Figure 2-4: A typical wide-angle SAR range profile.

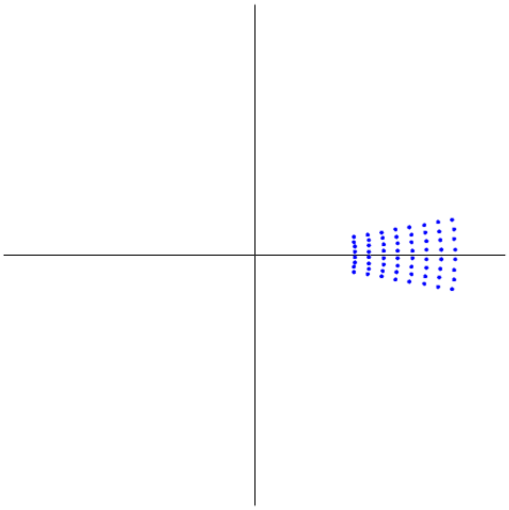


Figure 2-5: Support of measurements on Fourier plane.

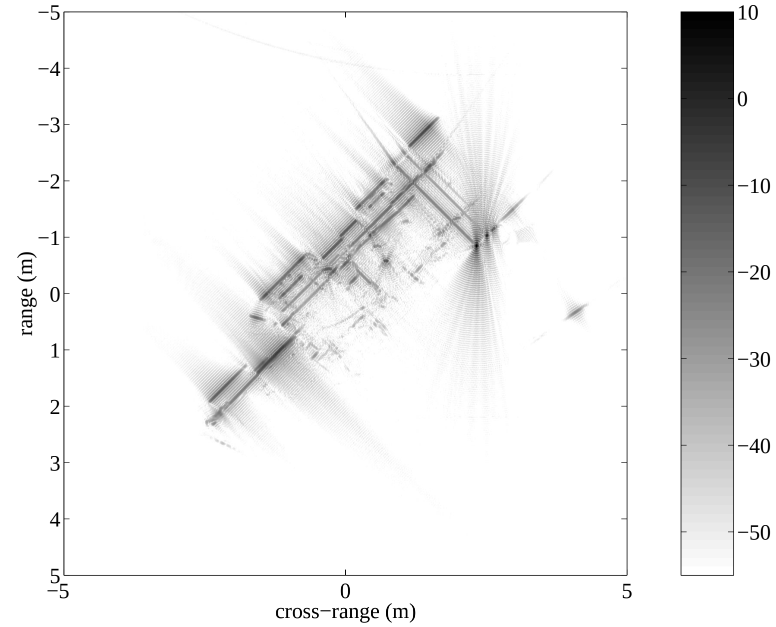


Figure 2-6: A typical conventionally-formed wide-angle SAR image.

arise. Specifically, fast algorithms applicable to Cartesian lattices do not transfer to polar format data directly. Polar FFT algorithms generally involve interpolation and resampling onto Cartesian grids before applying the standard FFT; the topic is still an active area of research [3]. Polar format algorithms for conventional SAR image formation apply the inverse FFT to interpolated, resampled phase history data [30]. The image magnitude reconstructed from the phase history measurements of Fig. 2-3 using conventional image formation is shown in Fig. 2-6 with decibel shading.

Having discussed the observation model for SAR as well as transforms taking data from the measured phase history domain to the range profile domain and to the image domain, the next section makes a modeling assumption that the scattering function is decomposable into a sum of point scatterers. By making such an assumption, the problem moves from the uncountably infinite realm to the finite realm.

2.1.3 Point-Scatterer Model

The scattering response of objects such as vehicles on the ground is well-approximated by the superposition of responses from point scattering centers when using ranges of frequencies commonly employed in SAR [31]. Therefore, under the assumption of point scattering, the image $s(x, y)$ is defined over a finite set of spatial locations and the relationship between measurements and underlying image transits from an integral, i.e. equation (2.1), to a sum, see equation (2.13). The relationship is derived below, starting with the radar pulse.

Consider the time-limited linear frequency modulated chirp signal:

$$\gamma(t) = \begin{cases} e^{j\left(2\pi f_c t + \alpha t^2\right)}, & -\frac{T}{2} \leq t \leq \frac{T}{2} \\ 0, & \text{otherwise} \end{cases}, \tag{2.5}$$

with carrier frequency f_c and chirp rate 2α. A common pulse used in SAR for illumination is $\cos(2\pi f_c t + \alpha t^2), |t| \leq \frac{T}{2}$, which equals $\Re\{\gamma(t)\}$.

Now consider obtaining measurements from a single point scatterer located at spatial location (x_0, y_0), with aspect angle fixed at θ, the center of the ground patch $(0, 0)$ at distance D from the radar, and the point (x_0, y_0) at distance $D + \rho$. When a pulse is transmitted towards the ground patch, the echo emanating from a differential area around (x_0, y_0) is delayed in time by $\frac{2(D+\rho)}{c}$, taking into account the round trip distance traveled and the speed of propagation c. The pulse has its magnitude scaled by $|s(x_0, y_0)|$, incurs a phase shift of $\angle s(x_0, y_0)$, and has an overall attenuation by a factor A. The received signal, $\bar{r}_0(t)$, coming from the differential area is:

$$\begin{aligned} &\bar{r}_0(t) \\ &= A|s(x_0, y_0)| \cos\left(2\pi f_c \left(t - \frac{2(D+\rho)}{c}\right) + \alpha\left(t - \frac{2(D+\rho)}{c}\right)^2 + \angle s(x_0, y_0)\right) dx\, dy \end{aligned} \tag{2.6}$$

over the time interval $-\frac{T}{2} + \frac{2(D+\rho)}{c} \leq t \leq \frac{T}{2} + \frac{2(D+\rho)}{c}$. Defining $\tau = \frac{2D}{c}$ and manipulating equation (2.6) using complex envelope analysis:

$$\bar{r}_0(t) = A\Re\left\{ s(x_0, y_0)\gamma\left(t - \tau - \frac{2\rho}{c}\right)\right\} dx\, dy. \tag{2.7}$$

As part of SAR processing, the received signal is downconverted from f_c by first mixing it with $\gamma^*(t-\tau)$, the complex conjugate of $\gamma(t-\tau)$, giving:

$$\hat{r}_0(t) = A\Re\left\{ s(x_0, y_0)e^{j\left(2\pi f_c\left(t-\tau-\frac{2\rho}{c}\right)+\alpha\left(t-\tau-\frac{2\rho}{c}\right)^2\right)}\right\} e^{-j\left(2\pi f_c(t-\tau)+\alpha(t-\tau)^2\right)} dx\, dy \tag{2.8}$$

$$\begin{aligned} &= \frac{A}{2}s(x_0, y_0)e^{j\left(2\pi f_c\left(2t-2\tau-\frac{2\rho}{c}\right)+\alpha(t-\tau)^2+\alpha\left(t-\tau-\frac{2\rho}{c}\right)^2\right)} dx\, dy \\ &+ \frac{A}{2}s(x_0, y_0)e^{j\left(\frac{4\alpha\rho^2}{c^2}-\frac{2\rho}{c}(2\pi f_c+2\alpha(t-\tau))\right)} dx\, dy. \end{aligned} \tag{2.9}$$

The augend in equation (2.9) is centered around frequency $2f_c$, whereas the addend is centered around frequency 0. In the second step of downconversion, $\hat{r}_0(t)$ is lowpass filtered, leaving:

$$\tilde{r}_0(t) = \frac{A\, dx\, dy}{2} s(x_0, y_0)e^{j\left(\frac{4\alpha\rho^2}{c^2}-\frac{2\rho}{c}(2\pi f_c+2\alpha(t-\tau))\right)}. \tag{2.10}$$

Under normal SAR operating conditions, $c^2 \gg 4\alpha\rho^2$, so $e^{j\frac{4\alpha\rho^2}{c^2}} \approx 1$. Also, the overall attenuation along with the differentials is removed for the purposes of the model. Thus:

$$r_0(t) = s(x_0, y_0)e^{-2j\left(\frac{2\pi}{c}f_c+\frac{2\alpha(t-\tau)}{c}\right)\rho}. \tag{2.11}$$

It should be remembered that the definition of $r_0(t)$ in equation (2.11) is over a time interval of length T and zero outside the interval. Also, it should be noted that in this interval of time, the spatial frequency varies from $\left(\frac{2\pi}{c}f_c - \frac{\alpha T}{c}\right)$ to $\left(\frac{2\pi}{c}f_c + \frac{\alpha T}{c}\right)$. Thus taking samples in time is tantamount to acquiring measurements over a frequency band. The quantity ρ, the offset in the distance to the radar due to the point (x_0, y_0) being offset from the center of the ground patch, is directly related by trigonometry to x_0, y_0, and θ by $\rho = x_0 \cos\theta + y_0 \sin\theta$. Thus the expression giving the measurement of a single point scatterer is:

$$r_0(f,\theta) = s\left(x_0, y_0\right) \exp\left\{-j\frac{4\pi f}{c}\left(x_0\cos\theta + y_0\sin\theta\right)\right\}. \tag{2.12}$$

Then, by superposition, if a scene contains N_s point scattering centers, the total measured phase history is:

$$r(f,\theta) = \sum_{p=1}^{N_s} s\left(x_p, y_p\right) \exp\left\{-j\frac{4\pi f}{c}\left(x_p\cos\theta + y_p\sin\theta\right)\right\}. \tag{2.13}$$

The point-scatterer model has been derived; the steps leading to the integral equation (2.1) are very similar. The measurement model does not yet include anisotropy; the model is broadened in the next section. However, before continuing to the next section, the range profile of point scattering centers is discussed.

To begin the discussion, let frequency be continuous and of infinite bandwidth. With a single point scatterer of unit amplitude, $r(f,\theta) = e^{-j\frac{4\pi f}{c}(x_0\cos\theta + y_0\sin\theta)}$. As a common transform pair it is known that $\mathscr{F}^{-1}\left\{e^{-j\omega\rho_0}\right\} = \delta\left(\rho - \rho_0\right)$. Thus taking $\omega = \frac{4\pi f}{c}$, the range profile for a point scatterer is:

$$\hat{R}_0(\rho,\theta) = \delta(\rho - x_0\cos\theta - y_0\sin\theta), \tag{2.14}$$

which can also be seen by setting $s(x,y) = \delta(x_0, y_0)$ in equation (2.3). The range profile is non-zero at a sinusoid:

$$\rho(\theta) = x_0\cos\theta + y_0\sin\theta. \tag{2.15}$$

For this reason, the range profile domain is known as the sinogram domain in other tomographic applications. It is also known as the Radon transform domain [7]. It should be noted that in applications other than SAR, data are primarily real and positive. The sinusoidal result may also be obtained directly from the observation geometry. With samples and bandlimited frequency, instead of the pure Dirac delta, the range profile samples a periodic sinc function with its main lobe centered at the

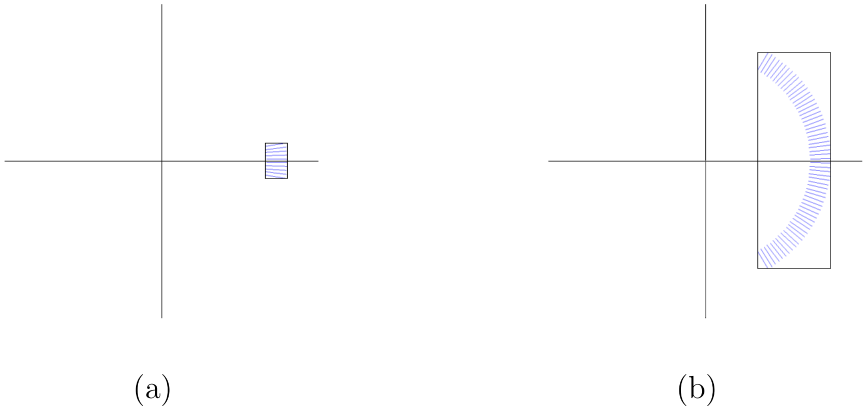

(a) (b)

Figure 2-7: Measurements in Fourier plane under (a) narrow-angle imaging, and (b) wide-angle imaging.

same point. With large enough bandwidth, the sidelobes are nearly indiscernible.

2.1.4 Wide-Angle Imaging and Anisotropy

Typical parameters for a SAR system given in [8] include an aperture covering 1.12° of aspect. In wide-angle SAR, however, the range of aspect angles is twenty or one hundred times as large. Imaging over wide-angle apertures has many effects, including measurements covering an annular region of the frequency plane, measurements exhibiting anisotropy, and scattering centers migrating in spatial location as a function of aspect angle.

Although not the main focus of the thesis, an issue that arises in wide-angle imaging relates to the region in the Fourier domain covered by the measurements. As seen in Fig. 2-7, when a narrow angle is used, the support in the Fourier plane is nearly rectangular, but for the same bandwidth, when a wide angle is subtended, the region of support is more clearly annular. As discussed in Sec. 2.1.2, conventional polar format image formation algorithms perform the inverse Fourier transform on the frequency domain samples by interpolating the samples onto a rectangular grid and then performing the inverse transform. Because the bounding rectangle is nearly filled in the narrow-angle case, interpolation is valid. However, in the wide-angle case, the bounding rectangle is sparsely populated and thus interpolation error may be significant, especially in the top and bottom right corners and central left portion of the bounding rectangle. The main consequence of wide-angle data collection and the focus of this thesis, anisotropy, is covered next.

In the exposition of Sec. 2.1.2 and Sec. 2.1.3, an implicit assumption is made that the scattering function $s(x, y)$ is not dependent on aspect angle. As introduced in Sec. 1.1, the assumption is not valid in wide-angle imaging. In fact, scattering behaviors are dependent both on the aspect angle and the frequencies in the radar pulse. The scattering function should be written explicitly as $s(x, y, f, \theta)$ to capture the dependencies. In conventional image formation techniques, the failure to model frequency dependence and angle dependence results in an averaging over those variables, leading to inaccurate scattering estimates. In wide-angle imaging, angle-dependence is more prominent than frequency-dependence, so it is the focus of the research. Methodologies developed for the characterization of angular anisotropy ought to be useful for characterizing frequency dependence as well. In this work, it is assumed that the scattering function is not dependent on frequency. With anisotropy, phase history is modeled as:

$$r\left(f, \theta\right) = \sum_{p=1}^{N_s} s\left(x_p, y_p, \theta\right) \exp\left\{-j\frac{4\pi f}{c}\left(x_p \cos\theta + y_p \sin\theta\right)\right\}. \tag{2.16}$$

In the range profile domain, the amplitudes of the sinusoids vary with $s(\theta)$. Sinusoidal segments with varying amplitude are visible in Fig. 2-4.

One other effect that is prominent in wide-angle imaging much more so than in narrow-angle imaging is that certain scattering mechanisms migrate as a function of aspect angle. That is to say that a scattering center appears to be in different spatial locations depending on the viewing direction. For example, consider a circular cylinder — the point of reflection is on the surface closest to the radar and changes as the aspect angle changes. The location can be parameterized as a function of θ around the center of the cylinder (x_c, y_c) using the radius of the cylinder R_0. The sinusoid equation changes in the following manner:

$$\rho(\theta) = (x_c - R_0 \cos\theta)\cos\theta + (y_c - R_0 \sin\theta)\sin\theta = x_c \cos\theta + y_c \sin\theta - R_0. \quad (2.17)$$

A generalization for any migration that can be parameterized around a center is:

$$\rho(\theta) = x_c \cos\theta + y_c \sin\theta - R(\theta). \quad (2.18)$$

In discussing stationary scattering centers, the spatial location (x_p, y_p) and the scattering center p are synonymous. However, care must be taken when discussing migratory scattering centers — some invariant location $(\bar{x}_p, \bar{y}_p)$ is needed to discuss the function $s(\bar{x}_p, \bar{y}_p, \theta)$ for example. This invariant spatial location $(\bar{x}_p, \bar{y}_p)$ is taken to be the location the scattering center appears at when $\theta = 0$. When $\theta = 0$, $\bar{x} = x_c - R(0)$ and $\bar{y} = y_c$. This leads to the following expression for phase history with migratory point scatterers:

$$r\left(f, \theta\right) = \sum_{p=1}^{N_s} s\left(\bar{x}_p, \bar{y}_p, \theta\right) \exp\left\{-j\frac{4\pi f}{c}\left(\left(\bar{x}_p + R_p(0)\right)\cos\theta + \bar{y}_p \sin\theta - R_p(\theta)\right)\right\}. \quad (2.19)$$

2.2 Inverse Problems and Sparse Signal Representation

The field of inverse problems concerns itself with the study of operators that are fairly simply defined in the forward direction, but difficult to invert. Meanwhile, through sparse signal representation, an attempt is made to decompose complicated signals into a few signals from a collection in a parsimonious way. In general, in both of these fields, signals may be infinite and continuous, but for the purposes of this thesis, dealing with finite duration, discrete-time signals is sufficient. Also, operators and ways to combine signals from a collection is restricted to linear operations, also discrete and finite. Even under these restrictions, inverse problems and sparse signal representation are rich areas. Some of that richness is recounted in this section.

2.2.1 Ill-Posed Inverse Problems and Regularization

In inverse problems, one has observations, possibly noisy, and attempts to discover the cause of those observations. In mathematical notation, $\mathbf{r} = \mathbf{\Phi}(\mathbf{a}) + \mathbf{n}$ is to be solved, where $\mathbf{a}$ is the unknown, $\mathbf{r}$ is the observed data, $\mathbf{\Phi}$ is the observation process, and $\mathbf{n}$ is noise. Now, under the restriction that $\mathbf{\Phi}$ is a linear operator, $\mathbf{a} \in \mathbb{C}^M$, and $\mathbf{n}, \mathbf{r} \in \mathbb{C}^N$, the problem is:

$$\mathbf{r} = \mathbf{\Phi}\mathbf{a} + \mathbf{n}, \mathbf{\Phi} \in \mathbb{C}^{N \times M}. \tag{2.20}$$

Inverse problems are ill-posed if no solution $\mathbf{a}$ exists, if the solution is not unique, or if the solution does not depend continuously on the data [28]. The condition number of a matrix κ indicates how difficult it is to invert: the identity matrix has $\kappa = 1$ and singular matrices have infinite κ. Inverse problems with ill-conditioned matrices, i.e. those with large condition number, have a noise amplification effect. Small example operators $\mathbf{\Phi}$ are displayed in Table 2.1, illustrating ill-posedness and ill-conditioning. A well-posed example is also shown for reference.

Type	Exemplar	κ	Notes
well-posed	$\begin{bmatrix} 0.6 & 0 \\ 0 & 0.4 \end{bmatrix}$	1.5	$\mathbf{\Phi}^{-1} = \begin{bmatrix} 1.67 & 0 \\ 0 & 1.5 \end{bmatrix}$
no solution	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	∞	$\mathbf{r} = \begin{bmatrix} 1 \\ 0.01 \end{bmatrix}$ has no corresponding $\mathbf{a}$
solution not unique	$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$	1.73	$\mathbf{r} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ has infinite solutions: $\mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{a} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \ldots$
ill-conditioned	$\begin{bmatrix} 0.01 & 1 \\ 0 & 0.01 \end{bmatrix}$	10^4	for $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{\Phi}^{-1}(\mathbf{\Phi}\mathbf{a}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, but $\mathbf{\Phi}^{-1}\left(\mathbf{\Phi}\mathbf{a} + \begin{bmatrix} 0.001 \\ 0.001 \end{bmatrix}\right) = \begin{bmatrix} -8.9 \\ 1.1 \end{bmatrix}$

Table 2.1: Ill-posed and ill-conditioned problems.

If $M = N$ and $\mathbf{\Phi}$ is nonsingular, then the matrix inverse $\mathbf{\Phi}^{-1}$ exists and may be used to obtain a solution $\mathbf{a} = \mathbf{\Phi}^{-1}\mathbf{r}$. If $\mathbf{\Phi}$ does not have an inverse, then generalized inverses, such as $\mathbf{\Phi}^\dagger = (\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H$, which satisfy $\mathbf{\Phi}^\dagger\mathbf{r} = \mathbf{a}$ and exist for all matrices, can be used [40, 44]. The vector $\mathbf{\Phi}^\dagger\mathbf{r}$ is the minimum norm least-squares solution.

Although the generalized inverse specifies a unique estimate for $\mathbf{a}$ from $\mathbf{\Phi}$ and a noisy version of $\mathbf{r}$, this estimate may not be meaningful or may be inappropriate for the application at hand, e.g. a different choice of solution from infinite possibilities may be better for a certain application. Regularization — the inclusion of prior knowledge to stabilize the solution in the presence of noise and to give meaningful and reasonable estimates — is used to deal with issues of ill-posedness when estimating $\mathbf{a}$ from $\mathbf{r}$. A regularization approach to solving inverse problems constructs a cost function of the form:

$$J(\mathbf{a}) = J_1(\mathbf{a}) + \alpha J_2(\mathbf{a}), \tag{2.21}$$

where J_1 is a data fidelity term, often $\|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2$; J_2 is a regularization term; and α is a regularization parameter, which trades off data fidelity and regularization. The $\mathbf{a}$ that minimizes the cost function is the desired solution to the inverse problem. When $\alpha = 0$, the solution reduces back to least-squares.

In one approach to regularization developed by Tikhonov, the regularization term is quadratic, i.e. $J_2(\mathbf{a}) = \|\mathbf{a}\|_2^2$ [53]. This method and other quadratic regularization methods are tractable and have closed form solutions. The solutions resulting from quadratic regularization are 'smooth,' i.e. the elements of $\mathbf{a}$ vary smoothly. As different applications have different goals and different prior information, quadratic regularization may or may not be the preferred method of solution. A regularization term giving preference to sparse solutions is described in Sec. 2.2.3. Ill-posed inverse problems abound in the field of mathematical physics, including problems dealing with the propagation of electromagnetic waves [51].

2.2.2 Sparse Signal Representation

The main idea behind sparse signal representation is that given a class of signals of dimension N, if one takes a collection of $M > N$ signals that span the class, it should be possible to represent any signal as a linear combination of a *few* members of the collection. In other words, any given signal should be representable using only a sparse subset of the full collection. The collection is more than a basis, and is termed an overcomplete basis or dictionary. The elements are known as basis vectors or atoms. An in-depth coverage of sparse signal representation may be found in [19]; a similar, but condensed exposition is given in this section.

From linear algebra, it is known that the signal representation problem with overcomplete basis has infinite solutions. Let us define a precise notion of sparsity before proceeding on to a discussion of obtaining the optimally sparse solution. The ℓ_p-norm, $\|\cdot\|_p$, is defined as follows:

$$\|\mathbf{a}\|_p = \sqrt[p]{\sum_{i=1}^{M} |(a)_i|^p}, \tag{2.22}$$

where $(a)_i$ indicates the i^{th} entry of $\mathbf{a}$. The ℓ_p-norm is a true norm for $p \geq 1$, but fails to satisfy the triangle inequality for $0 \leq p < 1$, becoming a quasi-norm. When $p = 0$, the quasi-norm counts the number of non-zero elements in a vector. An optimally sparse solution is a solution which has the smallest ℓ_0-norm among all solutions. In other words, the sparsest coefficient vector $\mathbf{a}^*$ is optimal for the following problem:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{a}\|_0 \\ \text{subject to} \quad & \mathbf{r} = \mathbf{\Phi}\mathbf{a}, \end{aligned} \tag{2.23}$$

where $\mathbf{r}$ is the signal that is to be represented and the columns of $\mathbf{\Phi}$ are the basis vectors. This problem is difficult to solve directly; consequently, a number of indirect techniques have been developed to obtain solutions. The two main approaches go by the names 'matching pursuit' and 'basis pursuit,' with alternative more descriptive names 'orthogonal greedy algorithm' and 'convex relaxation.'

Matching pursuit approaches the problem in an iterative fashion. In the first iteration, the algorithm selects the single basis vector that minimizes the ℓ_2 error if included in an expansion and then subtracts away its contribution from the measurement, leaving a residual vector. The same process is continued in subsequent iterations, taking the basis vector that minimizes the error with the residual into the expansion. When stopped after i iterations, the expansion contains i basis vectors, which is a sparse representation if i is small. The iteration may be continued until the residual is zero.

In basis pursuit, the problem (2.23) is relaxed by replacing the ℓ_0-norm with the ℓ_1-norm, giving the following problem:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{a}\|_1 \\ \text{subject to} \quad & \mathbf{r} = \mathbf{\Phi}\mathbf{a}. \end{aligned} \tag{2.24}$$

This problem has a piecewise linear convex objective function and is straightforwardly cast as a linear programming problem, efficiently solved by a number of optimization techniques. Under certain conditions on the overcomplete basis $\mathbf{\Phi}$, it can be shown that the original problem and the relaxed problem have the same solution [18, 36].

When the measurements $\mathbf{r}$ are noisy, the signal representation problem turns into a signal approximation problem with essentially the same form. An optimally sparse solution is a solution to the following problem:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{a}\|_0 \\ \text{subject to} \quad & \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2 \leq \delta, \end{aligned} \tag{2.25}$$

where δ is a small noise allowance. Instead of satisfying $\mathbf{r} = \mathbf{\Phi}\mathbf{a}$ exactly, the solution coefficient vector $\mathbf{a}$ is allowed to satisfy the relationship approximately. The matching pursuit algorithm does not change in the presence of noise, but may be terminated when the residual vector has ℓ_2-norm less than δ. Basis pursuit relaxation from ℓ_0 to ℓ_1 with the approximation constraint can be cast as a convex quadratic program and solved efficiently.

Matching pursuit and basis pursuit are effective methodologies for sparse signal representation and sparse approximate signal representation usually applied to real-valued data. A different approach to sparsity, employing the ℓ_p-norm, with $0 < p < 1$, is presented in the next section.

2.2.3 Sparsifying Regularization

This section pulls together pieces from the previous two sections into a common framework, yielding a sparsity formulation distinct from matching pursuit and basis pursuit. If the sparse signal representation problem is approached as an ill-posed inverse problem requiring regularization, then the regularization term should favor sparse solutions. As discussed in Sec. 2.2.2, the ℓ_p-norm is a useful tool for considering sparsity, suggesting a regularization term of the form $J_2 = \|\mathbf{a}\|_p^p$ and regularization cost function:

$$J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi a}\|_2^2 + \alpha \|\mathbf{a}\|_p^p. \tag{2.26}$$

The above cost function may also be recognized as a Lagrangian relaxation to the optimization problem given below.

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{a}\|_p^p \\ \text{subject to} \quad & \|\mathbf{r} - \mathbf{\Phi a}\|_2^2 \leq \delta'. \end{aligned} \tag{2.27}$$

When $p \neq 2$, J_2 is not quadratic and not in the Tikhonov form. When $p = 1$, the regularization cost function (2.26) corresponds to basis pursuit. Ideally, p should be zero for sparsity; however, the relaxed case $0 < p < 1$ is of interest in the remainder of the thesis. With $p < 1$, the cost (2.26) is not of the Tikhonov form, and thus there is no closed form solution for its minimum. Additionally, the cost function is not convex. However, the optimization can be performed with a quasi-Newton method detailed in [9] and discussed in Chapter 4. Chapter 3 makes clear how sparsifying regularization fits with wide-angle SAR imaging. The next section departs from inverse problems and sparse signal representation in examining methods of anisotropy characterization.

2.3 Prior Work on Anisotropy Characterization

The problem of detecting, estimating, and modeling aspect-dependent scattering behavior has received attention lately. This section briefly surveys a few proposed methods for anisotropy characterization. SAR imaging has many closely aligned avenues of research, but the survey is restricted in scope to the anisotropy problem. Anisotropy characterization methods may be broadly categorized into those that operate in the phase history domain, employing parameterizations for angle-dependent scattering, and those that operate in the image domain. Both classes are put forward along with an overall discussion on their positive and negative aspects.

2.3.1 Parametric Methods

As seen in Sec. 2.1, specifically equations (2.16) and (2.19), the phase history domain observation model with anisotropy is composed of $s(\theta)$ functions for each scattering center. A general approach is to posit a parametric model for $s(\theta)$ and estimate the model parameters for each scattering center. Often, scattering center spatial locations (x_p, y_p) are also estimated within the parameter estimation. Electromagnetic theory provides motivation for a number of parametric functions including parameterizations of frequency dependence, some of which are given below.

A model obtained from the geometrical theory of diffraction is:

$$s\left(x_p, y_p, f, \theta\right) = A_p \left(\frac{jf}{f_c}\right)^{\alpha_p} \exp\left\{\beta_p \theta\right\} \tag{2.28}$$

where f_c is the center frequency of the radar pulse, and A_p, α_p, and β_p are the parameters to estimate [47]. The α_p parameters are restricted to be half-integers and relate to the number of bounces a radar pulse takes before coming back to the antenna. Two other parametric models also motivated by the geometrical theory of diffraction [26] include, for localized scattering mechanisms such as corners and edges of manmade objects:

$$s\left(x_p, y_p, f, \theta\right) = A_p \exp\left\{-2\pi f \gamma_p \sin\theta\right\} \tag{2.29}$$

and for distributed scattering mechanisms, such as flat plates and cylinders:

$$s\left(x_p, y_p, f, \theta\right) = A_p \left(\frac{jf}{f_c}\right)^{\alpha_p} \operatorname{sinc}\left(\frac{2\pi f}{c} L_p \sin\left(\theta - \varphi_p\right)\right). \tag{2.30}$$

The parameters to be estimated are A_p, α_p, γ_p, L_p, and φ_p. Iterative algorithms for maximum likelihood parameter estimation for these three models exist [14].

A different parametric model for scattering behavior uses an expansion of $s(x, y, \theta)$ into spherical harmonic functions with θ as their argument [17]. The coefficients for the spherical harmonic basis are obtained by taking inner products. In [54], the scattering function is parameterized with Gaussians. The model is:

$$s\left(x_p, y_p, \theta\right) = A_p \exp\left\{-\frac{\left(\theta - \varphi_p\right)^2}{2\sigma_p^2}\right\} \tag{2.31}$$

and the parameters A_p, φ_p, and σ_p are found with a given efficient numerical algorithm.

In addition to parameterizing anisotropy, the aspect dependence of spatial location (migratory scattering) may also be parameterized. The work in [25] parameterizes anisotropy with polynomials and also range profile with polynomials:

$$\rho_p\left(\theta\right) = a_p + b_p\theta + \frac{c_p}{2}\theta^2. \tag{2.32}$$

The stationary point scatterer expression for range profile can be recovered by setting $a_p = x_p$, $b_p = y_p$, and $c_p = -x_p$, yielding $\rho_p(\theta) = x_p \cos(\theta) + y_p \sin(\theta)$ using Maclaurin series approximations for sine and cosine.

Parametric methods in the phase history domain yield well-defined estimation problems based on the physics of canonical scattering or other prior knowledge. The next section is concerned with methods that incorporate very little prior information.

2.3.2 Image-Domain Methods

Image-domain methods use a multiaperture approach for characterizing anisotropy. The radar measurements are divided into smaller pieces, with cuts along the θ axis. An image, known as a subaperture image, is then formed from each segment of the measurements. It should be noted that the subaperture images have poorer cross-range resolution than an image formed from the full aperture would. Also, subapertures are of fixed angular extent; consequently, any subaperture analysis is limited in its ability to characterize anisotropy extent.

In [23], the division of the aperture is a partition into equally-sized subapertures and the goal is the detection of anisotropy. After subaperture images are formed through conventional imaging and registered, changes among corresponding pixels are detected. In essence, the procedure is like looking at a flipbook and picking out little flashes. The detection is posed as a binary hypothesis testing problem for each location and the likelihood functions are estimated using a trained hidden Markov model. The result can be summarized as a Boolean image that is true at pixel (x_p, y_p) if $s(x_p, y_p, \theta)$ is anisotropic. By dividing the aperture into non-overlapping subapertures, anisotropic behavior that straddles a subaperture boundary may be missed. Overlapped subapertures are not used in [23] due to the increase in computation. Similar methods are described in [50].

The method described above can be visualized as passing the phase history through a filterbank with non-overlapping rectangular windows. A different procedure for forming subapertures, given in [41], is to use a filterbank with overlapping Hamming windows. Further processing used in [41] involves subaperture image formation using point-enhanced imaging [10]. Anisotropy is characterized by considering corresponding pixels in each subaperture image — the subaperture image with the maximum pixel value corresponds to the direction of anisotropy. Other approaches use filters different from Hamming and rectangular windows in the filterbank that are matched to expected responses [2]. Other subaperture approaches are detailed in [13, 34, 22].

The above methods consider anisotropy of fixed angular extent only. A method of detecting anisotropic scattering [32] that is in the same spirit as the previous methods additionally uses multiple levels of subapertures. Within each level, the subapertures have the same angular extent and are overlapping, but the extent in each level is half that of the previous level. The layers of subapertures form a tree-structured subaperture pyramid, with the full aperture as the root. Each of the subapertures can be considered as a hypothesis. For each pixel (x_p, y_p) independently, the tree is searched for the best hypothesis to explain the data, considering two levels at a time, starting at the root. The search is continued if a child subaperture is more likely to explain the data than the parent. The likelihoods are obtained using a derived statistic analogous to the total energy contained in the subaperture.

In general, image-domain multiaperture methods operate on images already formed by conventional or enhanced image formation techniques, proceeding based on the simple fact that partial data is enough to still form reasonable images. The next section discusses aspects of this class of methods side-by-side with parametric methods.

2.3.3 Discussion

The parametric methods and image-domain methods described above have been shown to work well in different situations. Notably, the parametric models incorporate much prior information about expected scattering behavior. Also, the estimated parameters have physical significance, e.g. the parameter L_p corresponds to the physical length of the scattering mechanism. The image-domain methods are robust, easy to reason about conceptually, and can be applied to already formed images. They can be used to characterize anisotropy to various levels of detail, from simple detection — isotropic vs. anisotropic — to more complicated questions such as the angular extent or center angle of anisotropy.

The parametric model formulation of the anisotropy characterization problem is of course predicated on the correct modeling of natural phenomena. The above parametric models often do not hold in wide-angle imaging scenarios [41]. This fact is especially evident in equation (2.32), the range profile parameterization, in which the basic point scatterer model cannot be recovered in wide-angle scenarios. Model misspecification error is considered in [14], but with the limitation that the unknown true model is of a known parametric family. However, moving from narrow-angle SAR to wide-angle SAR is not simply a transition of the parametric model within a family of models. Within the image-domain methods, the subaperture pyramid framework moves towards allowing a continuum of aspect angle extents, but is still limited to full-, half-, quarter-, $\cdots$-, apertures. All of the image-domain methods use subapertures which inherently reduce resolution. Also, in most of the techniques, the characterization of anisotropy in different spatial locations (different pixels) is done independently.

A formulation using an overcomplete basis novel to the anisotropy characterization problem is presented in Chapter 3. While still taking advantage of prior information, this method is flexible enough to admit solutions that are not from a prespecified parametric family. It jointly treats spatial locations and suffers no reduction in cross-range resolution.

Chapter 3

Overcomplete Basis Formulation for Anisotropy Characterization and Image Formation

> Interpretation of inaccurate, insufficient, and inconsistent data: Many problems in physical science involve the estimation of a number of unknown parameters which bear a linear or quasi-linear relationship to a set of experimental data. The data may be contaminated by random errors, insufficient to determine the unknowns, redundant, or all of the above.
>
> — *D.D. Jackson (1972), in Geophysical Journal of the Royal Astronomical Society*

This chapter describes a formulation of the anisotropy characterization problem different from the subaperture and parametric formulations discussed in Sec. 2.3. The problem is approached through the sparsifying regularization framework described in Sec. 2.2.3 by constructing an overcomplete basis and appropriately using the phase history measurements. The idea is to expand the scattering function $s(\theta)$ at each spatial location as a superposition of basis vectors and then determine coefficients for those vectors. The first part of the chapter leaves the overcomplete basis fully general; the chapter concludes with a consideration of specific basis choices.

3.1 Anisotropy Characterization Inverse Problem

In this section, the observation models for stationary and migratory point scattering with anisotropy, equations (2.16) and (2.19), are expanded into overcomplete bases. The expansions are expressed in both summation notation and in matrix-vector form.

3.1.1 Stationary Scattering Centers

The starting point for the overcomplete expansion is the observation model for stationary scattering centers, derived in Sec. 2.1, which is reproduced below.

$$r\left(f,\theta\right)=\sum_{p=1}^{P}s\left(x_p,y_p,\theta\right)\exp\left\{-j\frac{4\pi f}{c}\left(x_p\cos\theta+y_p\sin\theta\right)\right\}. \tag{3.1}$$

For a single spatial location p, the aspect-dependent scattering is expanded as follows:

$$s\left(x_p,y_p,\theta\right)=\sum_{m=1}^{M}a_{p,m}b_m\left(\theta\right), \tag{3.2}$$

yielding the following overall $M\cdot P$ basis expansion:

$$r\left(f,\theta\right)=\sum_{p=1}^{P}\sum_{m=1}^{M}a_{p,m}b_m\left(\theta\right)\exp\left\{-j\frac{4\pi f}{c}\left(x_p\cos\theta+y_p\sin\theta\right)\right\}. \tag{3.3}$$

Taking the phase history measurements at K discrete frequencies and N discrete aspect angles, let us define length N vectors $\mathbf{r}_k=r(f_k,\theta)$, $\mathbf{b}_m=b_m(\theta)$, and $\boldsymbol{\varepsilon}_{k,p}=\exp\left\{-j\frac{4\pi f_k}{c}\left(x_p\cos\theta+y_p\sin\theta\right)\right\}$. Then, taking $\boldsymbol{\phi}_{k,p,m}=\mathbf{b}_m\bullet\boldsymbol{\varepsilon}_{k,p}$, where $\mathbf{M}_1\bullet\mathbf{M}_2$ is the elementwise multiplication of matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, the basis expansion may be simply expressed as:

$$\mathbf{r}_k=\sum_{p=1}^{P}\sum_{m=1}^{M}a_{p,m}\boldsymbol{\phi}_{k,p,m},\ k=1,\ldots,K. \tag{3.4}$$

The inverse problem is to determine the $M\cdot P$ complex-valued coefficients $a_{p,m}$ that satisfy or approximately satisfy the linear equations (3.4). The number of basis vectors M is chosen such that $M > N$, making the basis overcomplete.

Now, moving to matrix-vector equations, the collection of all phase history measurements can be stacked as the following tall $N \cdot K$-vector $\mathbf{r}$.

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_K \end{bmatrix}. \tag{3.5}$$

The set of all basis vectors at a particular frequency f_k and spatial location (x_p, y_p) can be collected into a matrix $\mathbf{\Phi}_{k,p}$.

$$\mathbf{\Phi}_{k,p} = \begin{bmatrix} \boldsymbol{\phi}_{k,p,1} & \boldsymbol{\phi}_{k,p,2} & \cdots & \boldsymbol{\phi}_{k,p,M} \end{bmatrix}. \tag{3.6}$$

In the same manner, the $\mathbf{b}_m$ vectors can be concatenated into a $\mathbf{B}$ matrix.

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M \end{bmatrix}. \tag{3.7}$$

These two matrices are related by the following expression:

$$\mathbf{\Phi}_{k,p} = \mathbf{B} \bullet \left(\boldsymbol{\varepsilon}_{k,p} \mathbf{1}_M^T \right), \tag{3.8}$$

where $\mathbf{1}_M$ is the M-vector of all ones. The factor $\mathbf{B}$ is subject to design in the anisotropy characterization procedure, but $\boldsymbol{\varepsilon}_{k,p}$ is fundamental to the SAR phase history measurements. The choice of $\mathbf{B}$ is discussed in the second half of this chapter.

Putting together all frequencies and spatial locations, the overall overcomplete basis $\mathbf{\Phi}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{1,1} & \mathbf{\Phi}_{1,2} & \cdots & \mathbf{\Phi}_{1,P} \\ \mathbf{\Phi}_{2,1} & \mathbf{\Phi}_{2,2} & \cdots & \mathbf{\Phi}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Phi}_{K,1} & \mathbf{\Phi}_{K,2} & \cdots & \mathbf{\Phi}_{K,P} \end{bmatrix}. \tag{3.9}$$

Defining the length $M \cdot P$ vector of coefficients as:

$$\mathbf{a} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,M} & a_{2,1} & \cdots & a_{P,M} \end{bmatrix}^T , \tag{3.10}$$

the statement $\mathbf{r} = \mathbf{\Phi a}$ in matrix-vector form is completely equivalent to the summation form (3.4). In this form, it is readily apparent that there are $N{\cdot}K$ linear equations with $M{\cdot}P$ unknowns and $M > N$. Often in the choice of $\mathbf{B}$, $M \gg N$, so regardless of P and K, $M{\cdot}P > N{\cdot}K$ and the system is an underdetermined set of linear equations. Solving the inverse problem with sparsifying regularization, mentioned in Sec. 2.2.3, is discussed in greater depth in Chapter 4.

By forming this $\mathbf{\Phi}$ matrix, the spatial locations are treated jointly within one system of equations. The first M elements of $\mathbf{a}$ depend on the first spatial location $p = 1$, elements $M + 1$ to $2M$ of $\mathbf{a}$ depend on the second position $p = 2$, and so on. Thus, by setting up the problem in this manner, it is possible to decompose the phase history data into contributions from different point scatterers and in the process, characterize amplitude and anisotropy for each one. As a notational convenience, $\mathbf{a}_p$ is the M-vector of coefficients corresponding to position p and:

$$\mathbf{\Phi}_p = \begin{bmatrix} \mathbf{\Phi}_{1,p} \\ \mathbf{\Phi}_{2,p} \\ \vdots \\ \mathbf{\Phi}_{K,p} \end{bmatrix} ,$$

the subset of basis vectors corresponding to position p. There is no requirement that all spatial locations under consideration contain a scatterer. If there is no scatterer at a particular spatial location p, then all of the elements of $\mathbf{a}_p$ should come out to be zero. It is thus possible to use a grid of pixels as the set of spatial locations.

3.1.2 Migratory Scattering Centers

With migratory scattering, there is an additional degree of freedom over stationary scattering. The overcomplete basis expansion of aspect-dependent scattering is the same as above, giving:

$$r\left(f,\theta\right)=\sum_{p=1}^{P}\sum_{m=1}^{M}a_{p,m}b_{m}\left(\theta\right)\exp\left\{-j\frac{4\pi f}{c}\left(\left(\bar{x}_{p}+R_{p}(0)\right)\cos\theta+\bar{y}_{p}\sin\theta-R_{p}\left(\theta\right)\right)\right\}. \tag{3.11}$$

The observation model is further expanded in migration. Restricting to circular migration (see equation (2.17)), the observations are related to the overcomplete basis and coefficients by:

$$r\left(f,\theta\right)=\sum_{p=1}^{P}\sum_{l=1}^{L}\sum_{m=1}^{M}a_{p,l,m}b_{m}\left(\theta\right)\exp\left\{-j\frac{4\pi f}{c}\left(\left(\bar{x}_{p}+R_{p,l}\right)\cos\theta+\bar{y}_{p}\sin\theta-R_{p,l}\right)\right\}. \tag{3.12}$$

Again defining vectors $\boldsymbol{\varepsilon}_{k,p,l}=\exp\left\{-j\frac{4\pi f_k}{c}\left(\left(\bar{x}_{p}+R_{p,l}\right)\cos\theta+\bar{y}_{p}\sin\theta-R_{p,l}\right)\right\}$ and $\boldsymbol{\phi}_{k,p,l,m}=\mathbf{b}_{m}\bullet\boldsymbol{\varepsilon}_{k,p,l}$, of length N, the basis expansion is:

$$\mathbf{r}_{k}=\sum_{p=1}^{P}\sum_{l=1}^{L}\sum_{m=1}^{M}a_{p,l,m}\boldsymbol{\phi}_{k,p,l,m},\ k=1,\ldots,K. \tag{3.13}$$

Thus, with circular migratory scattering there are $L{\cdot}M{\cdot}P$ complex-valued coefficients. Each basis vector corresponds to a different invariant spatial location, different radius of migration, and different anisotropy, i.e. $\mathbf{b}_{m}$.

The conversion from summation notation to matrix-vector notation follows in the same manner as for stationary scattering. The main extensions are highlighted below:

$$\mathbf{\Phi}_{k,p,l}=\begin{bmatrix}\boldsymbol{\phi}_{k,p,l,1} & \boldsymbol{\phi}_{k,p,l,2} & \cdots & \boldsymbol{\phi}_{k,p,l,M}\end{bmatrix}, \tag{3.14}$$

$$\mathbf{\Phi}=\begin{bmatrix}\mathbf{\Phi}_{1,1,1} & \mathbf{\Phi}_{1,1,2} & \cdots & \mathbf{\Phi}_{1,1,L} & \mathbf{\Phi}_{1,2,1} & \cdots & \mathbf{\Phi}_{1,P,L}\\ \mathbf{\Phi}_{2,1,1} & \mathbf{\Phi}_{2,1,2} & \cdots & \mathbf{\Phi}_{2,1,L} & \mathbf{\Phi}_{2,2,1} & \cdots & \mathbf{\Phi}_{2,P,L}\\ \vdots & \vdots & \cdots & \cdots & \cdots & \ddots & \vdots\\ \mathbf{\Phi}_{K,1,1} & \mathbf{\Phi}_{K,1,2} & \cdots & \mathbf{\Phi}_{K,1,L} & \mathbf{\Phi}_{K,2,1} & \cdots & \mathbf{\Phi}_{K,P,L}\end{bmatrix}, \tag{3.15}$$

and

$$\mathbf{a} = \begin{bmatrix} a_{1,1,1} & a_{1,1,2} & \cdots & a_{1,1,M} & a_{1,2,1} & \cdots & a_{1,L,M} & a_{2,1,1} & \cdots & a_{P,L,M} \end{bmatrix}^T. \qquad (3.16)$$

The overcomplete bases and inverse problems for the anisotropy characterization and migratory scattering problems have now been stated. The choice of vectors in $\mathbf{B}$ is covered in the next section, whereas solving the inverse problem is covered in the next chapter.

3.2 Choice of Basis Vectors

The set of vectors $\mathbf{B}$ for the overcomplete basis $\mathbf{\Phi}$ is to be chosen such that its cardinality is much greater than the dimension of θ and linear combinations of very few basis vectors accurately represent plausible angle-dependent scattering behaviors. The choice of $\mathbf{B}$ is a way to incorporate prior information about angle-dependent scattering.

3.2.1 Contiguous Anisotropy

As seen in Sec. 2.3, methods employing subaperture analysis and parametric models expect to find contiguous intervals in θ for which there is non-zero scattering. Fig. 3-1 shows an example of the aspect-dependent scattering function of a realistic point scattering center — the anisotropy is contiguous. Basis vectors used in this thesis are chosen to similarly represent contiguous segments of anisotropy. The $\mathbf{b}_m$ are chosen to be pulses with all possible angular extents and all possible starting angles, in other words all widths and shifts. For example, if $N = 8$ and the pulse shape is rectangular, then $\mathbf{b}_1$, the isotropic vector is $[1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]^T$, $\mathbf{b}_2 = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 0]^T$, $\mathbf{b}_3 = [0\ 1\ 1\ 1\ 1\ 1\ 1\ 1]^T$, and the final pulse with the finest anisotropy $\mathbf{b}_M = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 1]^T$. The $\mathbf{b}_m$ have unit maximum amplitude; solving the inverse problem gives the complex amplitude coefficients. The full set $\mathbf{B}$ for $N = 8$ is illustrated in Fig. 3-2. The dots represent entries that have a non-zero value

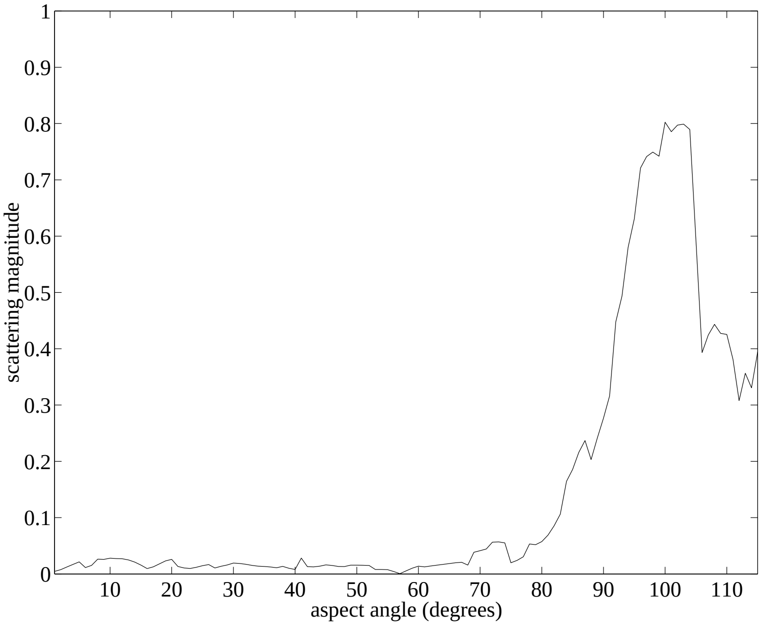


Figure 3-1: Contiguous anisotropy: large scattering response over contiguous intervals of aspect angle and nearly zero response elsewhere.

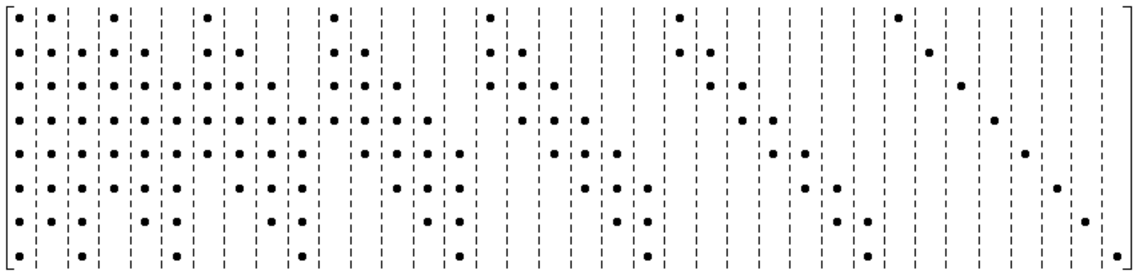


Figure 3-2: Illustration of matrix **B** for $N = 8$.

and spaces without dots represent zero-valued elements.

For this choice, there is a simple relationship between M, the number of vectors in **B**, and N, the number of angle samples. M is the N^{th} triangular number:

$$M = \binom{N+1}{2} = \frac{1}{2}N^2 + \frac{1}{2}N. \tag{3.17}$$

Various pulse shapes may be enlisted in the overcomplete basis, e.g. triangle, raised triangle, and raised cosine. These shapes are discussed further in Sec. 3.2.3. The next section describes a structure for the basis that comes in handy in developing methods of solution for the inverse problem.

3.2.2 Graph-Structured Interpretation of Basis

The overcomplete basis described in the previous section has a nice, intuitive graph-structured interpretation. The vectors $\mathbf{b}_m$ can be arranged as nodes in a weakly connected directed acyclic graph. The graph is given in Fig. 3-3 for $N = 8$, with nodes labeled to the left with their corresponding **b** vectors when the pulse shape is rectangular. A topological sort for this graph is exactly the ordering from left to right of the vectors in **B**, illustrated in Fig. 3-2.

The root node is the isotropic basis vector; traversing down the graph corresponds to decreasing angular extent of anisotropy. Any such traversal from coarse to fine anisotropy is analogous to the path a shot or chip takes when dropped into a pegboard apparatus in which the pegs are arranged in a quincunx pattern. Such apparatus,

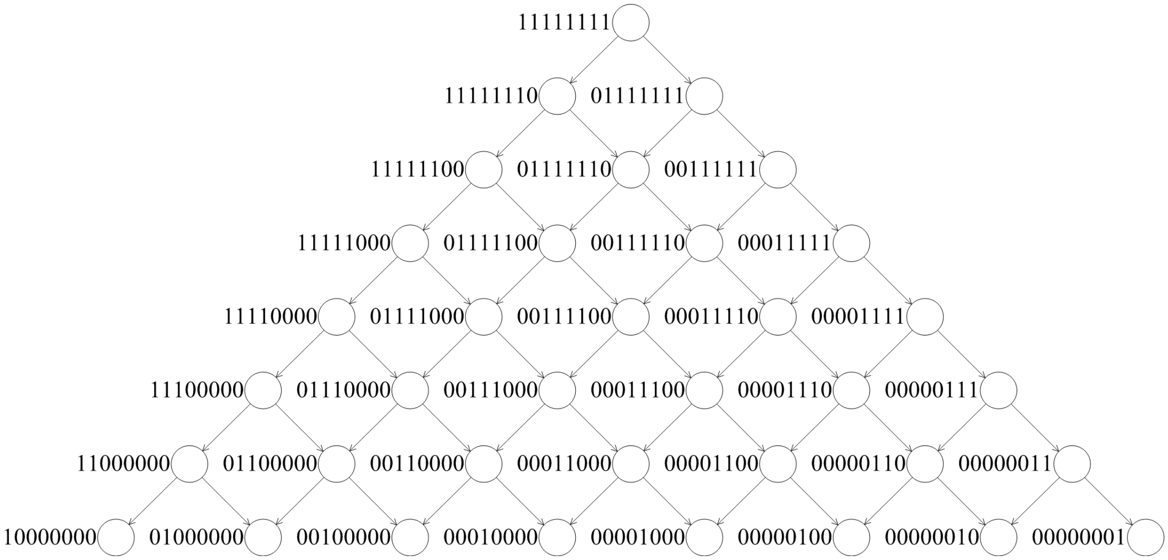


Figure 3-3: Directed acyclic graph representation of **B**.

(a) (b)

Figure 3-4: Quincunx peg apparatus (a) Galton board[1], and (b) plinko board[2].

going by names such as the Galton board and plinko board have been used to study statistics [24] and in games of chance [46], see Fig. 3-4. Consequently, in this thesis a graph of this form with N rows is referred to as an N-level *plinko graph*.

The graph-structured representation need not have directed edges. Edges are directed in this section for pedagogic reasons only; plinko graphs are undirected in the remainder of the thesis.

---

[1] image from page 63, Fig. 7 of Galton [24]

[2] image from Mobliss website (http://www.mobliss.com/_img/game/detail/plinko_150x200.gif)

3.2.3 Mutual Coherence

Mutual coherence of an overcomplete basis, $\mu(\mathbf{\Phi})$, is an important quantity in convergence and stability conditions of sparse signal representation algorithms such as matching pursuit and basis pursuit [19]. Mutual coherence is defined as:

$$\mu\left(\mathbf{\Phi}\right) = \max_{i \neq j} |\boldsymbol{\phi}_i^H \boldsymbol{\phi}_j|, \tag{3.18}$$

where the $\boldsymbol{\phi}$ vectors are normalized to unit ℓ_2 magnitude. Equivalently, it is the largest non-diagonal entry of the Gram matrix $\mathbf{G} = \mathbf{\Phi}^H\mathbf{\Phi}$ in magnitude. By definition, μ is between zero and one. Mutual coherence indicates how similar the two most similar basis vectors in the overcomplete basis are. Orthonormal bases have zero mutual coherence. Overcomplete bases with small μ are referred to as incoherent, whereas overcomplete bases with large μ are referred to as coherent.

Mutual coherence gives an indication of how difficult or ill-posed the overcomplete basis is. Sparse signal representation algorithms employing ℓ_1 relaxations can be shown to always find solutions that are exactly the optimally sparse solution, i.e. with minimum ℓ_0-norm, when the overcomplete basis is sufficiently incoherent [36, Chapter 7], [19]. Essentially, mutual incoherence enables the construction of a sufficient condition for ℓ_1-ℓ_0 equivalence, hence guaranteeing the solution of a combinatorial optimization problem by tractable optimization algorithms. In this sense, mutual coherence is an indicator of how demanding a particular overcomplete basis is for a sparse signal representation algorithm.

When considering the mutual coherence of the overcomplete basis for anisotropy characterization described in this chapter, intuitively, vectors corresponding to the same spatial location p and radius of migration l are the most similar. Consequently, the two basis vectors which provide the maximum inner product to the mutual coherence calculation originate in the same p and l. Thus, it is $\mathbf{B}$ and not the $\boldsymbol{\varepsilon}_{k,p,l}$ vectors which determine $\mu(\Phi)$.

This begs the question, how does pulse shape in $\mathbf{B}$ affect mutual coherence? Are there easier and more difficult pulse shapes? The mutual coherence of different pulse shapes is now compared.

The mutual coherence of an overcomplete basis $\mathbf{B}$ with rectangular pulses or boxcar window shapes is easily determined analytically as a function of the number of angle samples N. The vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ have the maximum inner product. The two vectors are both non-zero in $N-1$ entries and have normalization factors $\frac{1}{\sqrt{N}}$ and $\frac{1}{\sqrt{N-1}}$, resulting in:

$$\mu\left(\mathbf{B}_N^{\text{boxcar}}\right) = \sqrt{\frac{N-1}{N}}. \tag{3.19}$$

The mutual coherence for triangle window shapes constructed using the `triang` function in Matlab is more complicated, but can be worked out to be:

$$\mu\left(\mathbf{B}_N^{\text{triang}}\right) = \begin{cases} \frac{N^3-N}{\sqrt{N^6+N^4-2N^2}}, & N = 2,4,6,\ldots \\ \frac{N^3-4N+3}{\sqrt{N^6-2N^4-6N^3+N^2+6N}}, & N = 3,5,7,\ldots \end{cases}. \tag{3.20}$$

Even and odd N have different expressions because the triangle window is defined differently for the two cases. As N goes to infinity, μ goes to one for both the boxcar and triangle window, but the convergence to unity is slower for the boxcar window, implying that both types of overcomplete bases are difficult, but the triangle is more difficult.

The two above mutual coherence expressions are plotted in Fig. 3-5 along with the mutual coherences for overcomplete bases employing several other pulse shapes. These other shapes include the Hamming window, as defined by the Matlab function `hamming`; a house shape, given by `(triang(N)+1)/2`; and a v shape, given by `circshift(triang(N),ceil(N/2))`. It can be observed that the more a pulse shape tapers to zero at its edges, the more coherent the basis using that pulse shape. Thus the v-shaped pulse is the least coherent (it grows towards its edges); unfortunately, this type of shape is not very plausible in explaining anisotropy.

The choice of pulse shape should foremost be representative of expected aspect-dependent scattering. However, solving the sparse signal representation problem becomes more difficult with peaky, tapering shapes. These two factors may be in contention and may require a compromise. The next section defines a new mutual

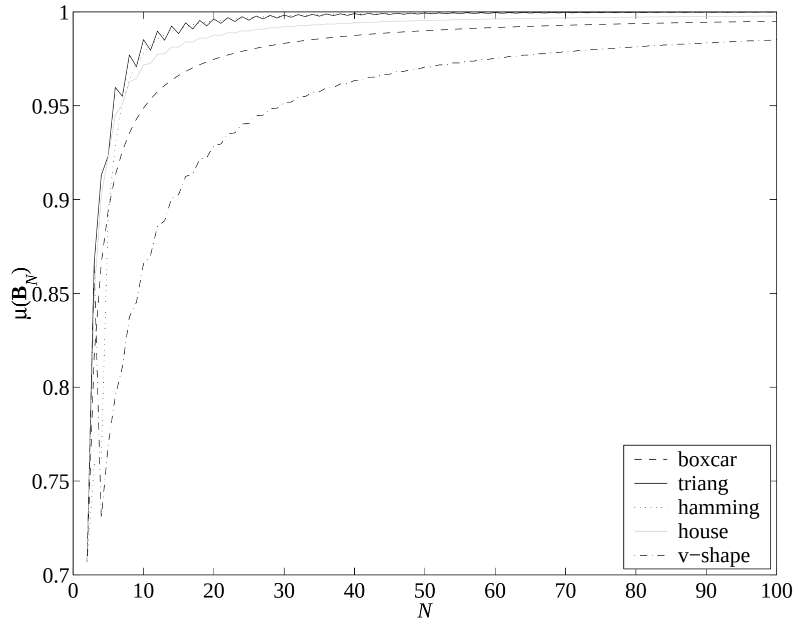


Figure 3-5: Mutual coherence as a function of N using different pulse shapes.

coherence-like quantity specific to the overcomplete basis for anisotropy characterization to further investigate how different parameters affect the difficulty of the problem.

3.2.4 Cross-Coherence

As discussed above, mutual coherence does not depend on the SAR $\boldsymbol{\varepsilon}_{k,p,l}$ vectors. Here a new notion, cross-coherence, is defined that does depend on the $\boldsymbol{\varepsilon}_{k,p,l}$. The maximum inner product is taken over all pairs of basis vectors such that they do not share common p and l. In terms of the Gram matrix, instead of disregarding just the diagonal as done in the mutual coherence calculation, a block diagonal portion is disregarded. Specifically,

$$\nu\left(\boldsymbol{\Phi}\right) = \max_{\substack{\phi_{p',l',m'} \notin \boldsymbol{\Phi}_{p,l} \\ \phi_{p,l,m} \notin \boldsymbol{\Phi}_{p',l'}}} |\boldsymbol{\phi}_{p,l,m}^{H} \boldsymbol{\phi}_{p',l',m'}|. \tag{3.21}$$

Cross-coherence may be used to investigate trends in the difficulty of the anisotropy characterization sparse signal representation problem under various SAR parameters. The effects of the number of angle samples N, the number of frequency samples K, the aperture width, and the frequency bandwidth are investigated.

When the two closest positions are (x_A, y_A) and (x_B, y_B), then the functional form for cross-coherence, $\nu(\mathbf{\Phi})$, is simply:

$$\nu\left(\mathbf{\Phi}\right) = \frac{1}{NK}\left|\sum_{n=1}^{N}\sum_{k=1}^{K} e^{-j\frac{4\pi}{c}f_k\left((x_B-x_A)\cos\theta_n+(y_B-y_A)\sin\theta_n\right)}\right|. \tag{3.22}$$

However, it is difficult to intuit the effects of the various parameters from the above expression. In order to develop this intuition, first let the N aspect angles be equally spaced from $-\hat{\theta}$ to $+\hat{\theta}$ and the K frequencies be equally spaced from $f_c - \frac{B}{2}$ to $f_c + \frac{B}{2}$. For $N = 1$, equation (3.22) reduces to:

$$\nu\left(K, B\right) = \begin{cases} \frac{2}{K}\sum_{k=1}^{\frac{K}{2}} \cos\left(\frac{4\pi}{c}\frac{2k-1}{K-1}\frac{B}{2}\left(x_B - x_A\right)\right), & K \text{ even} \\ \frac{1}{K} + \frac{2}{K}\sum_{k=1}^{\frac{K-1}{2}} \cos\left(\frac{4\pi}{c}\frac{2k}{K-1}\frac{B}{2}\left(x_B - x_A\right)\right), & K \text{ odd} \end{cases}, \tag{3.23}$$

as a function of K and B. For ranges of parameters often encountered, $\frac{B|x_B - x_A|}{c} < \frac{1}{4}$, so ν gets values from the first quarter cycle of cosine waves; hence ν is positive and monotonically decreasing with increasing bandwidth B. As K increases, cosines with longer periods contribute and thus ν increases. These trends are illustrated in Fig. 3-6 with $x_B - x_A = 0.001$. As bandwidth is increased, diversity in the measurements increases, thus reducing the cross-coherence and making the problem easier. Because frequency dependence is not modeled in the overcomplete basis, measurements at additional frequencies provide redundant information; redundancy makes the inverse problem more difficult, hence the increase in ν with K.

A clean simplification $\nu(N, \hat{\theta})$ similar to equation (3.23) does not fall out by setting $K = 1$ in general; however, $\nu(N, \hat{\theta})|_{N=2}$ is $\cos(\frac{4\pi}{c} f_c (y_B - y_A) \sin\hat{\theta})$. The widest of wide-angle apertures have angular extent less than 180°, so $\hat{\theta} < \frac{\pi}{2}$, and $\sin(\hat{\theta})$ is a monotonically increasing function. Thus, as $\hat{\theta}$ increases, $\sin(\hat{\theta})$ increases and conse-

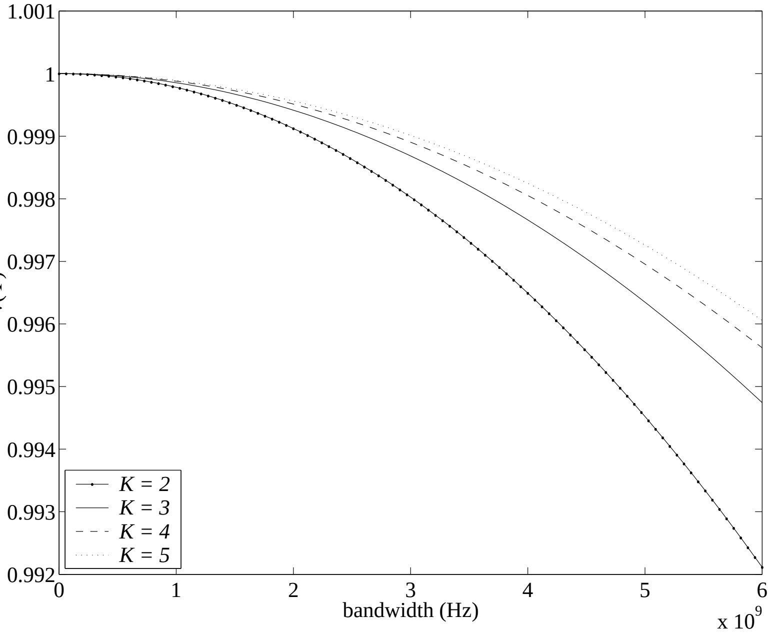


Figure 3-6: Cross-coherence as a function of bandwidth for different numbers of frequencies.

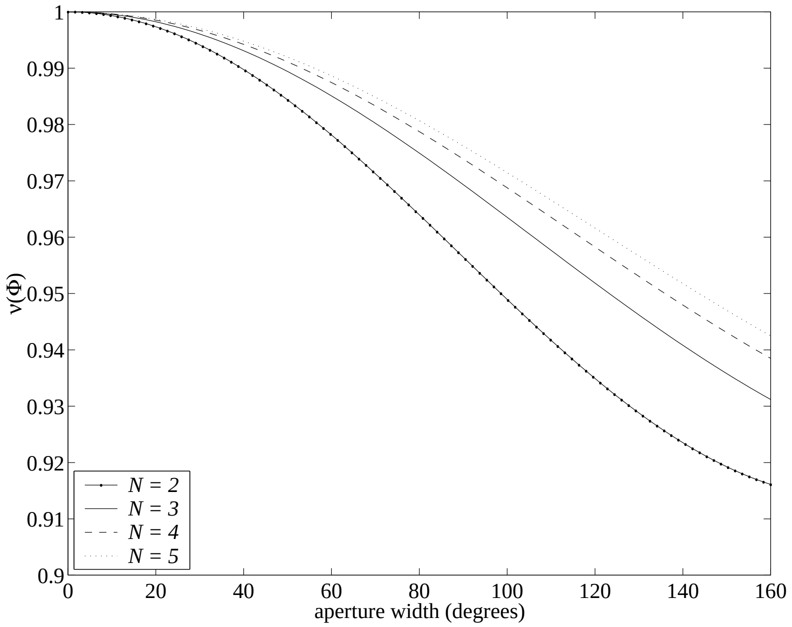


Figure 3-7: Cross-coherence as a function of aperture width for different numbers of aspect angles.

quently ν decreases. As N increases, ν increases, mirroring the trends for frequency. The reasoning for the aspect-angle trends also mirrors the reasoning for the frequency trends. Cross-coherence as a function of aperture width, i.e. $2\hat{\theta}$, for different values of N is shown in Fig. 3-7. For this plot, $x_B - x_A = 0.001$, $y_B - y_A = 0.001$, and $f_c = 10^{10}$.

Using phase history measurements corresponding to just two frequencies bracketing the widest possible bandwidth is a good strategy because small K and large B minimize ν. This reasoning would seem to suggest that not making use of all available data is advantageous; however, throwing away data, no matter how redundant, is not good for robustness in the presence of noise. It may be better to use measurements at several frequencies, but not necessarily all available phase history measurements. The tradeoff of noise robustness and difficulty of the sparse signal representation problem is investigated further in Chapter 5. By a similar logic, it would seem to follow that N should be kept small as well. However, the goal of this work is anisotropy characterization; not making use of all data in aspect angle limits the quality of characterized anisotropy. Thus, measurements at all angles should be used.

Employing cross-coherence as an indicator for the difficulty of the anisotropy characterization sparse signal representation problem, this section has shown in conjunction with the previous section that the problem as posed is difficult. As expected, diversity in the data makes the problem easier and an interesting tradeoff concerning redundancy in the data appears. The overcomplete basis defined in this chapter is fairly coherent, nevertheless viable methods of solution exist and are discussed in the next chapter.

Chapter 4

Optimization Methods

Structures are the weapons of the mathematician.

— *Nicolas Bourbaki*[1]

The thesis has heretofore developed a precise formulation of anisotropy characterization and image formation as a sparse signal representation inverse problem with a specific overcomplete basis and a specific methodology for solution — minimization of a regularization cost function. The contents of this chapter deal with actually carrying out the minimization. The techniques in this chapter make varying use of the structure of the overcomplete basis developed in Chapter 3.

4.1 Quasi-Newton Method

The cost function at hand in the minimization problem, $J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p$, $p < 1$, is certainly non-linear and non-quadratic; it may be shown that $J(\mathbf{a})$ is non-convex as well. To say that non-convex optimization is challenging would be an understatement. Further complicating matters, $\mathbf{\Phi}$, $\mathbf{a}$, and $\mathbf{r}$ are complex-valued. This section describes a quasi-Newton numerical optimization technique applicable to $J(\mathbf{a})$ developed in [9].

---

[1] This quotation is often attributed to the pseudonymous Nicolas Bourbaki, but the original source is quite elusive.

4.1.1 Iterative Algorithm

In order to arrive at the quasi-Newton algorithm, the cost function is approximated by a function differentiable everywhere, which is then differentiated and set to zero. The derivation of the algorithm is sketched below; a more thorough derivation may be found in [9].

The cost function being minimized is reproduced below in equation (4.1), with the $\|\cdot\|_p^p$ term expressed explicitly in summation form.

$$J(\mathbf{a}) = \|\mathbf{r} - \boldsymbol{\Phi}\mathbf{a}\|_2^2 + \alpha \sum_{i=1}^{M} |(a)_i|^p. \tag{4.1}$$

The $\|\cdot\|_p^p$ function forms a cusp at the origin; consequently $J(\mathbf{a})$ is not differentiable at $\mathbf{a} = \mathbf{0}$. A simple approximation $J_\epsilon(\mathbf{a})$, defined below, differentiable at $\mathbf{0}$, is used proceeding forward.

$$J_\epsilon(\mathbf{a}) = (\mathbf{r} - \boldsymbol{\Phi}\mathbf{a})^H(\mathbf{r} - \boldsymbol{\Phi}\mathbf{a}) + \alpha \sum_{i=1}^{M} \left((a)_i^2 + \epsilon\right)^{p/2}. \tag{4.2}$$

Clearly, $J_\epsilon(\mathbf{a}) = J(\mathbf{a})$ as ϵ goes to zero. Also, for small ϵ, J_ϵ and J are discernibly different only locally near the origin.

The next step towards reaching the iterative algorithm is determining the gradient of J_ϵ. The gradient of the quadratic term is:

$$\nabla \left((\mathbf{r} - \boldsymbol{\Phi}\mathbf{a})^H(\mathbf{r} - \boldsymbol{\Phi}\mathbf{a})\right) = -2\boldsymbol{\Phi}^H(\mathbf{r} - \boldsymbol{\Phi}\mathbf{a}). \tag{4.3}$$

The partial derivative of the sparsifying term with respect to entry $(a)_i$ is:

$$\frac{\partial}{\partial (a)_i}\left(\sum_{i=1}^{M} \left((a)_i^2 + \epsilon\right)^{p/2}\right) = \frac{p}{2}\left((a)_i^2 + \epsilon\right)^{p/2-1}(2(a)_i). \tag{4.4}$$

Then, after canceling twos and noting that the partial derivative with respect to variable $(a)_i$ depends only on $(a)_i$, the gradient may be formed as follows with a diagonal matrix:

$$\nabla\left(\sum_{i=1}^{M}\left((a)_i^2+\epsilon\right)^{p/2}\right)=p\begin{bmatrix}\left((a)_1^2+\epsilon\right)^{p/2-1} & 0 & \cdots & 0\\ 0 & \left((a)_2^2+\epsilon\right)^{p/2-1} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \left((a)_M^2+\epsilon\right)^{p/2-1}\end{bmatrix}\mathbf{a}. \tag{4.5}$$

Defining the matrix in equation 4.5 as $\mathbf{\Lambda}(\mathbf{a})$ and collecting the two gradient expressions leads to:

$$\begin{aligned}\nabla J_\epsilon(\mathbf{a}) &= -2\mathbf{\Phi}^H(\mathbf{r}-\mathbf{\Phi}\mathbf{a})+\alpha p\mathbf{\Lambda}(\mathbf{a})\mathbf{a}\\ &= -2\mathbf{\Phi}^H\mathbf{r}+\left(2\mathbf{\Phi}^H\mathbf{\Phi}+\alpha p\mathbf{\Lambda}(\mathbf{a})\right)\mathbf{a}.\end{aligned} \tag{4.6}$$

Once again, the notation is simplified by defining a matrix, $\mathbf{H}(\mathbf{a})=2\mathbf{\Phi}^H\mathbf{\Phi}+\alpha p\mathbf{\Lambda}(\mathbf{a})$. Setting ∇J_ϵ equal to zero leads to:

$$\begin{aligned}-2\mathbf{\Phi}^H\mathbf{r}+\mathbf{H}(\mathbf{a})\mathbf{a}&=\mathbf{0}\\ \mathbf{H}(\mathbf{a})\mathbf{a}&=2\mathbf{\Phi}^H\mathbf{r}.\end{aligned} \tag{4.7}$$

If $\mathbf{H}$ were not a function of $\mathbf{a}$, then there would be a closed form solution for the optimal $\mathbf{a}$, but alas, that is not the case.

Instead, an iterative approach is taken. The following iteration is used to converge to an $\mathbf{a}$ that minimizes the cost function.

$$\mathbf{a}^{(n+1)}=\mathbf{H}^{-1}\left(\mathbf{a}^{(n)}\right)2\mathbf{\Phi}^H\mathbf{r} \tag{4.8}$$

The iteration is terminated when the percent change between successive iterates falls below a small constant.

It is shown in [9] that this quasi-Newton method is also an extension of half-quadratic optimization and inherits the convergence properties of half-quadratic optimization. The technique is robust and is initialized with $\mathbf{a}^{(1)}=\mathbf{\Phi}^H\mathbf{r}$. The algorithm works well in practice; examples are given in Sec. 4.1.2. Next, a few aspects of the iterative update that may be taken advantage of under special structure in $\mathbf{H}$ are noted.

In the algorithm, $\mathbf{\Phi}^H\mathbf{\Phi}$ and $\mathbf{\Phi}^H\mathbf{r}$ are calculated once, whereas $\mathbf{\Lambda}$ is recalculated at every iteration. The set of linear equations (4.7) are resolved on every iteration as well. Efficient methods for solving large systems of equations are often specialized to systems with a special structure, such as triangular matrices and banded matrices [56]. Systems with diagonally dominant matrices, i.e. matrices that have the property that for each row, the sum of the non-diagonal magnitudes is smaller than the diagonal entry magnitude, have iterative algorithms for solution such as the Gauss-Jacobi and Gauss-Seidel methods. Iterative methods may also be used to solve systems approximately by early termination. Fortunately, the $\mathbf{H}(\mathbf{a})$ produced by the anisotropy characterization overcomplete basis is diagonally dominant.

In the examples in this thesis, the conjugate gradient method, another iterative method for solving systems of linear equations is used. The conjugate gradient technique repeatedly computes the forward linear operation $\mathbf{H}$. With stationary scattering centers, if the spatial locations (x_p, y_p) are equally spaced, then the matrix $\mathbf{\Phi}^H\mathbf{\Phi}$ is a block convolution matrix. Hence, computations may be sped up through the use of convolution rather than matrix multiplication.

4.1.2 Examples

This section contains three examples: an example with stationary scattering centers using synthetic data, an example with stationary scattering centers using XPatch data, and an example with migratory scattering centers using synthetic data. Together these examples illustrate that the overcomplete basis formulation is quite effective for joint image formation and anisotropy characterization.

The first example has $N = 16$ equally-spaced angles over $[-55°, +55°]$. There are $K = 3$ frequencies 7.047 GHz, 7.059 GHz, and 7.070 GHz, the speed of propagation $c = 2.998 \times 10^8$ m/s, and there are $P = 25$ spatial locations on a square lattice going from zero to four meters in both x and y. Of the 25 positions, five contain point

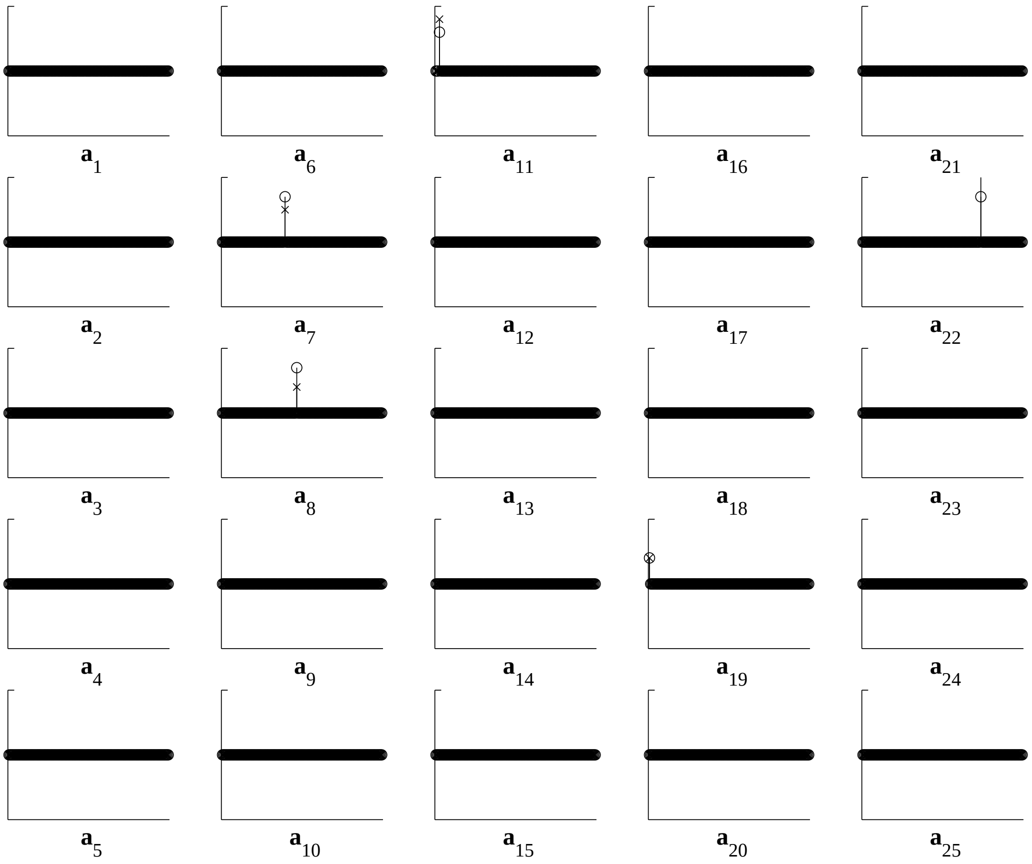


Figure 4-1: Complex amplitudes of true coefficients $\mathbf{a}$ with real part $\circ$ and imaginary part $\times$.

scatterers, with a variety of rectangular anisotropy extents and complex amplitudes — Fig. 4-1 shows the true coefficient vector $\mathbf{a}$. In the figure, $\mathbf{a}$ is partitioned into subvectors $\mathbf{a}_p$, with each $\mathbf{a}_p$ arranged according to its corresponding spatial location as in an image. With $N = 16$, there are $M = 136$ coefficients per position. The 136 coefficients are displayed as a stem plot, in a left to right ordering similar to Fig. 3-2, with the $\circ$ being the real part of the complex amplitude and the $\times$ being the imaginary part of the complex amplitude.

First, $\mathbf{r} = \mathbf{\Phi a}$ is solved using the pseudo-inverse to obtain the least-squares, minimum norm solution $\hat{\mathbf{a}}^{\mathrm{LS}}$ as a baseline for comparison. In Fig. 4-2, $\hat{\mathbf{a}}^{\mathrm{LS}}$ is displayed in the same manner as $\mathbf{a}$ in Fig. 4-1. This solution puts energy into coefficients of all $\hat{\mathbf{a}}_p^{\mathrm{LS}}$, giving non-zero scattering to all spatial locations. The coefficients are mapped

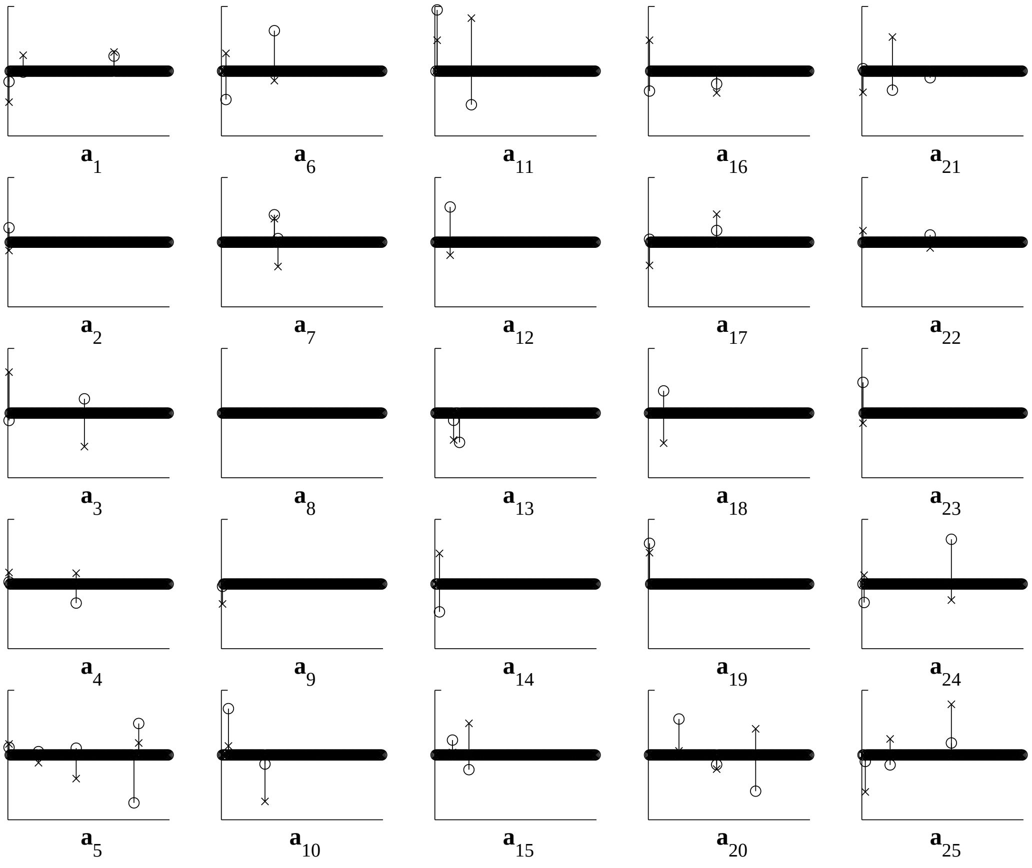


Figure 4-2: Complex amplitudes of coefficients in least-squares solution $\hat{\mathbf{a}}^{\mathrm{LS}}$ with real part $\circ$ and imaginary part $\times$.

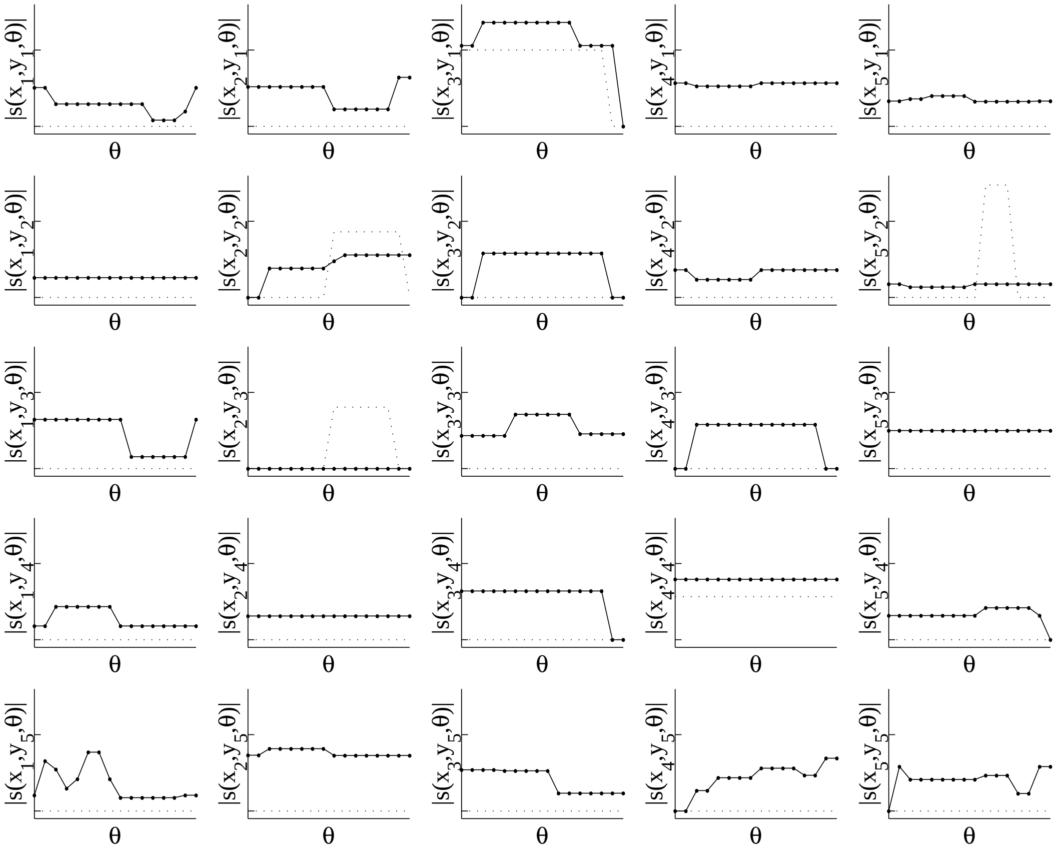


Figure 4-3: Least-squares solution scattering magnitude overlaid on true scattering magnitude.

to their corresponding scattering functions in Fig. 4-3. Once again, the scattering functions for individual scattering centers are arranged as in an image based on the spatial location of the scattering centers. The solid line with markers is the least-squares solution and the dotted line is the true scattering. As there is no additive noise, the product of $\hat{\mathbf{a}}^{\mathrm{LS}}$ and $\boldsymbol{\Phi}$ equals $\mathbf{r}$, but in this solution, the contributions are not assigned to individual spatial locations in a way that accurately reflects the underlying truth.

In contrast, the use of sparsifying regularization gives a solution to the inverse problem that explains the observations consistent with the true scattering function. Solution coefficients $\hat{\mathbf{a}}^{\mathrm{SR}}$, shown in Fig. 4-4, are obtained with the regularization term being the $\ell_{0.1}$-norm and the regularization parameter $\alpha = 3$. The scattering function

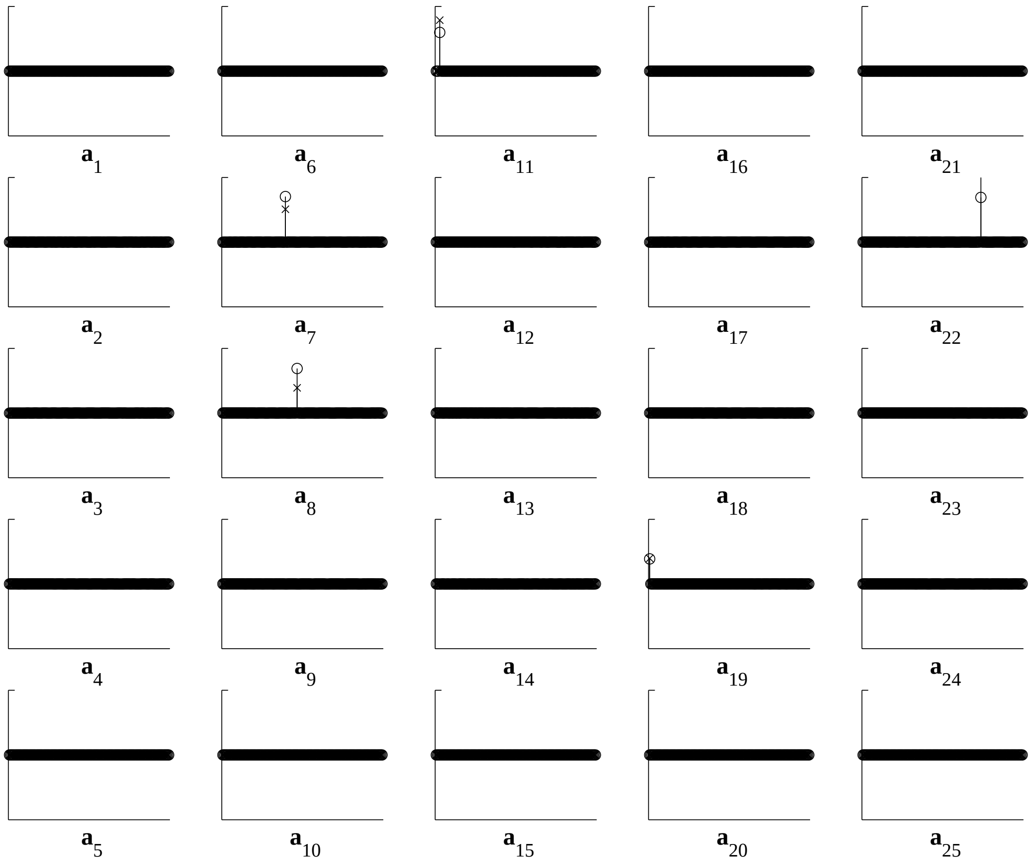


Figure 4-4: Complex amplitudes of coefficients in sparsifying regularization solution $\hat{\mathbf{a}}^{\mathrm{SR}}$ with real part $\circ$ and imaginary part $\times$.

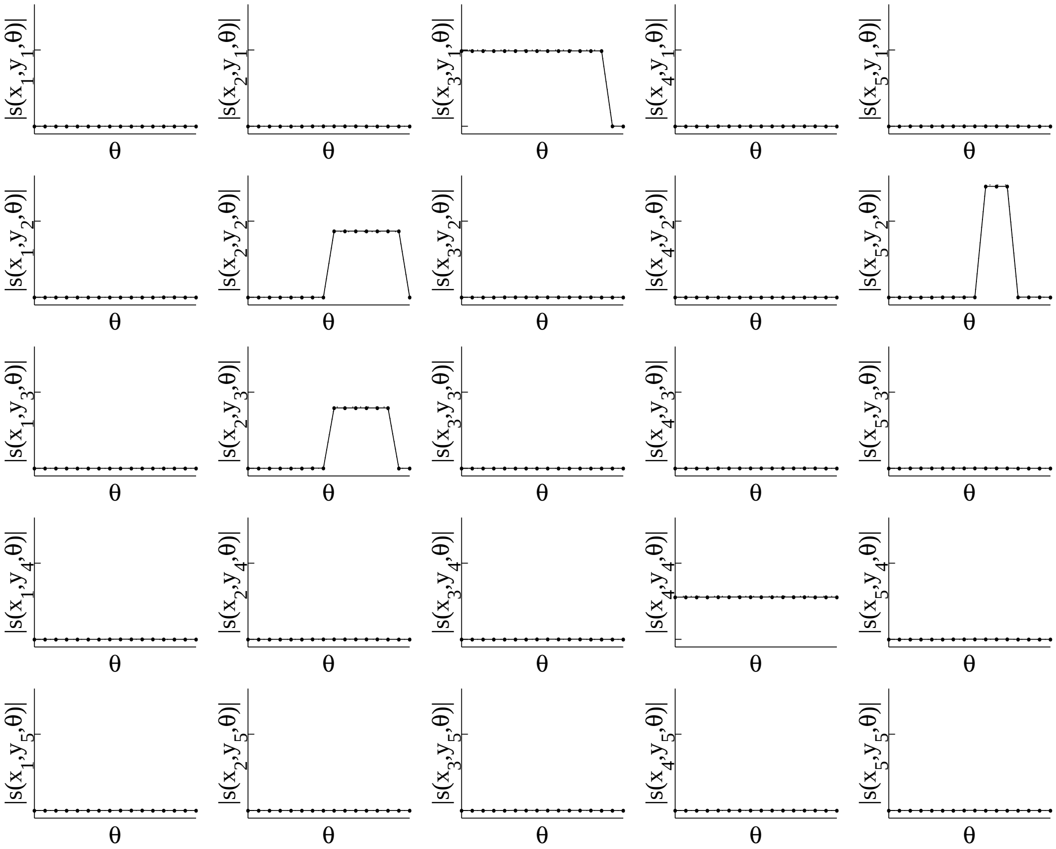


Figure 4-5: Sparsifying regularization solution scattering magnitude overlaid on true scattering magnitude.

corresponding to these coefficients, shown in Fig. 4-5, illustrates that this methodology of an overcomplete basis with sparsifying regularization is an effective method for characterizing anisotropy of point scatterers from phase history measurements. It also illustrates the fact that all spatial locations need not contain point scattering centers.

In the previous example the scatterer anisotropies exactly match the basis vectors, but in realistic imaging scenarios, that is not necessarily the case. In this second example, there are four spatial locations; two of the spatial locations have no scatterers and the scattering centers at the other two spatial locations exhibit realistic, i.e. from XPatch predictions, aspect-dependent scattering behavior. The four positions are at $(0,0)$, $(0,\frac{1}{2})$, $(\frac{1}{2},0)$, and $(\frac{1}{2},\frac{1}{2})$ meters, measurements are at $K=3$ frequencies 9 GHz,

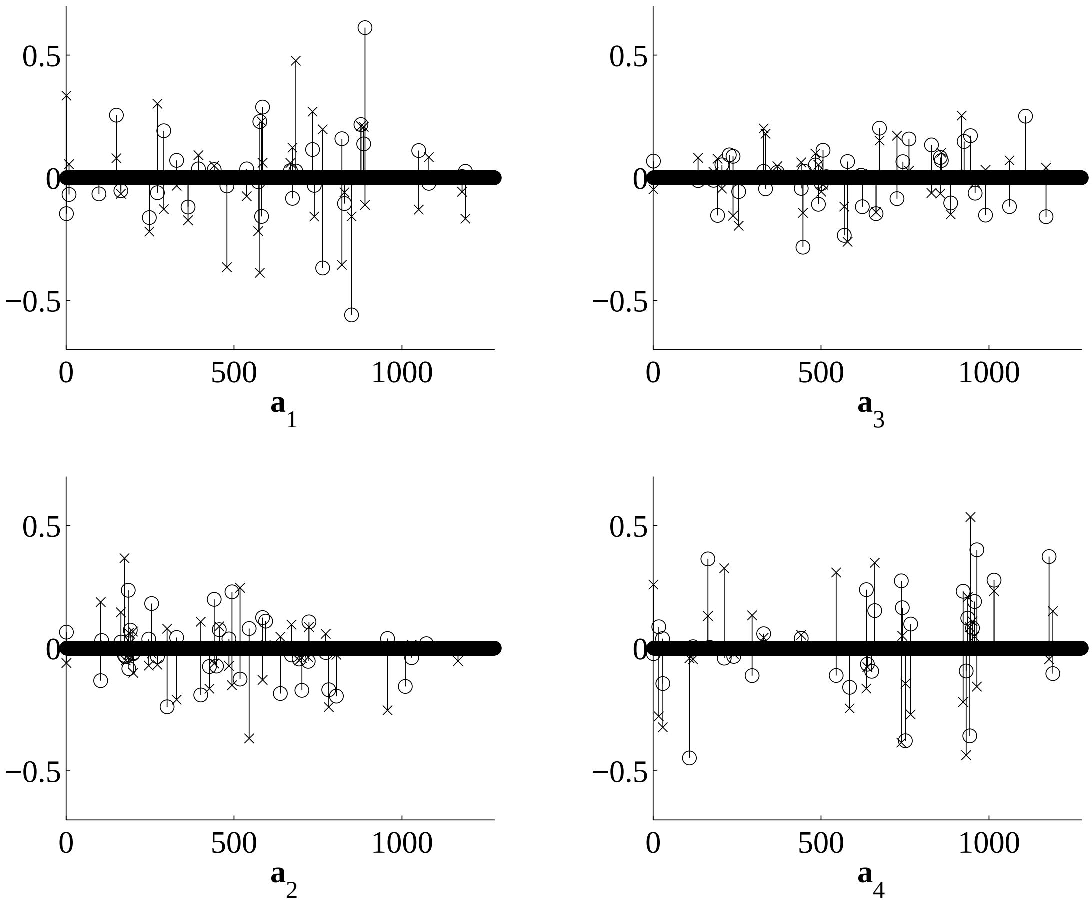


Figure 4-6: Complex amplitudes of coefficients in least-squares solution $\hat{\mathbf{a}}^{\mathrm{LS}}$ with real part $\circ$ and imaginary part $\times$.

9.016 GHz, and 9.032 GHz over the $N = 50$ angle samples equally spaced over a 98° aperture. The least-squares solution and a solution using the quasi-Newton method with the ℓ_p-norm having $p = 0.1$ and the regularization parameter $\alpha = 1$ are shown in Fig. 4-6 through Fig. 4-9.

As expected, $\hat{\mathbf{a}}^{\mathrm{SR}}$ is much sparser due to the $\ell_{0.1}$ regularization term. As in the first example, the coefficients corresponding to spatial locations without scattering centers are all nearly zero in the sparse solution, whereas the least-squares solution has many large-valued coefficients. Inspecting the scattering functions, the sparse solution is more accurate in its representation of the underlying truth than the least-squares solution, giving an indication that basis vectors of contiguous anisotropy are fairly good at sparsely representing realistic aspect-dependence.

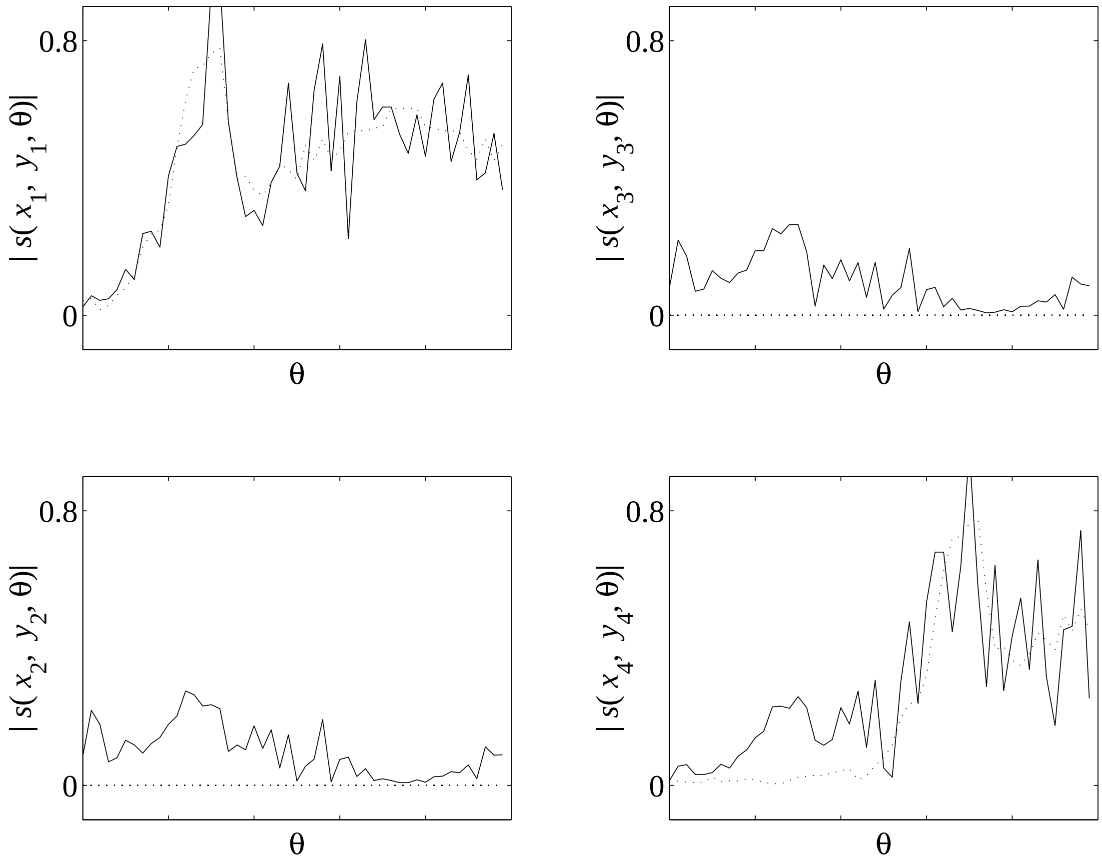


Figure 4-7: Least-squares solution scattering magnitude overlaid on true scattering magnitude.

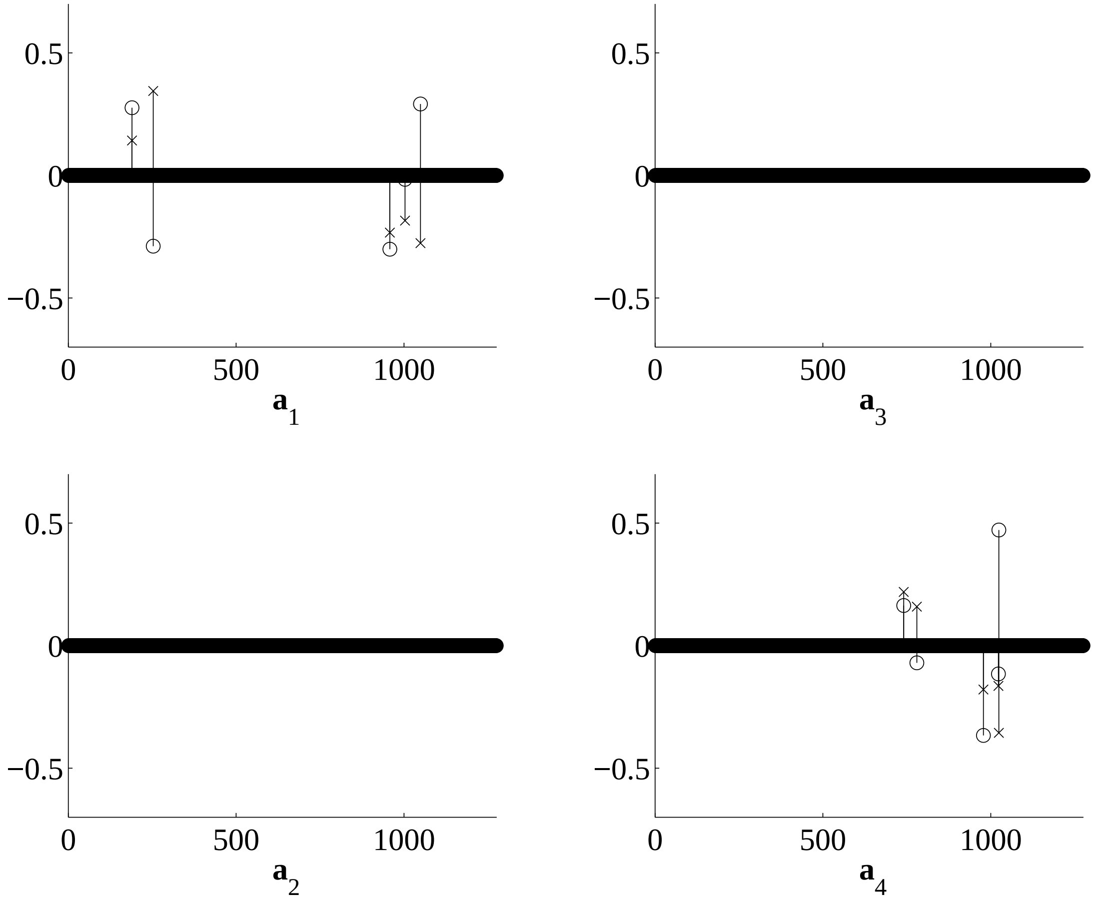


Figure 4-8: Complex amplitudes of coefficients in sparsifying regularization solution $\hat{\mathbf{a}}^{\mathrm{SR}}$ with real part $\circ$ and imaginary part $\times$.

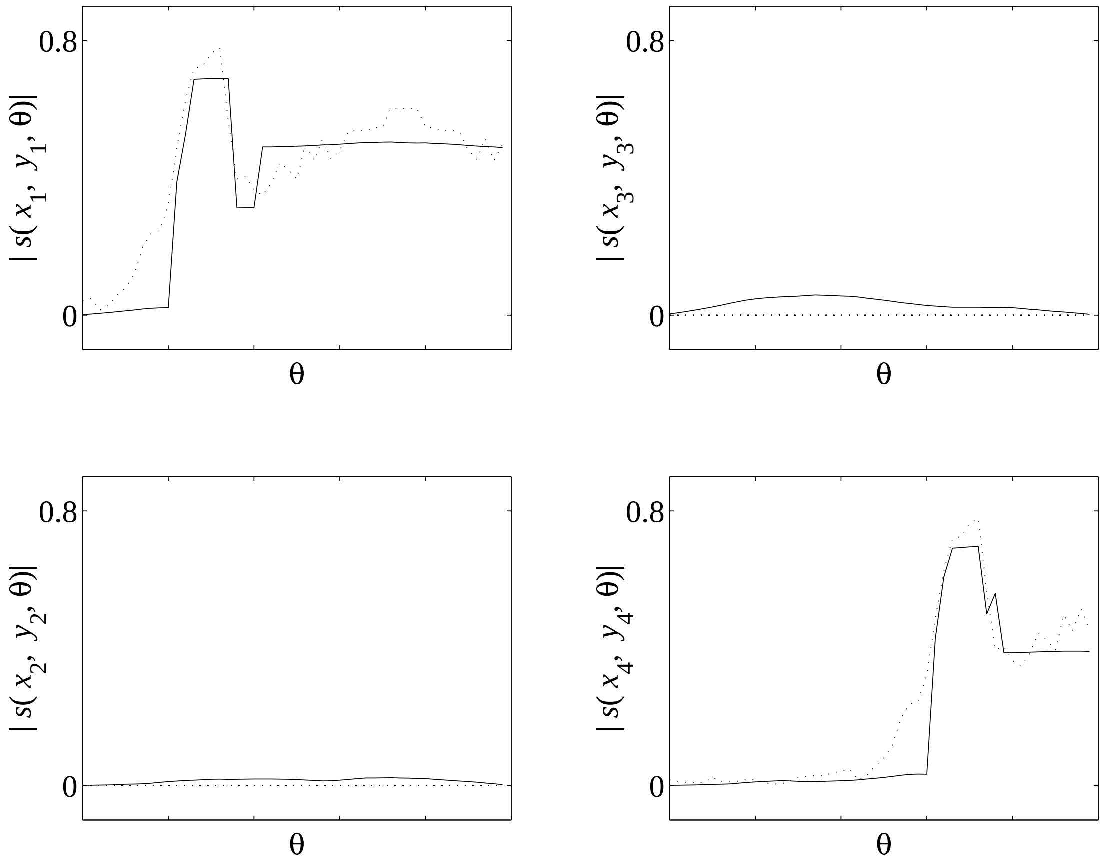


Figure 4-9: Sparsifying regularization solution scattering magnitude overlaid on true scattering magnitude.

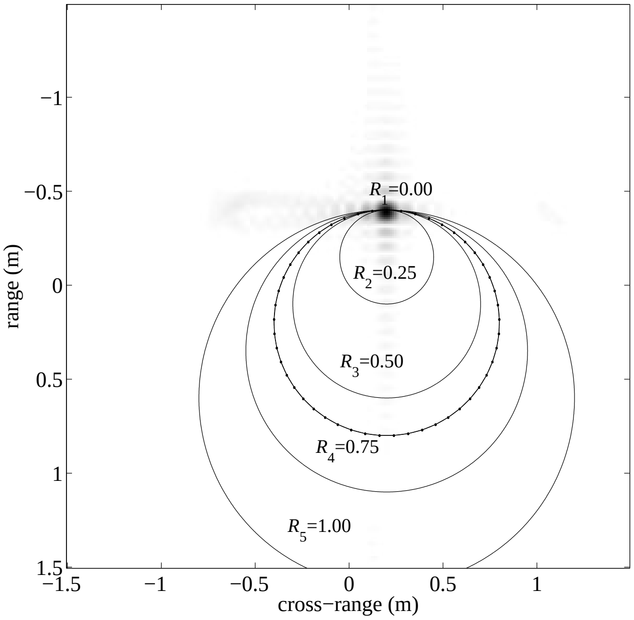


Figure 4-10: Illustration of basis vectors of different radii of migration along with true radius of migration, the circle with dots, overlaid on conventionally formed image.

The previous two examples deal with stationary scattering centers, whereas the third example contains a migratory scattering center. There is just one scattering center in the scene, i.e. $P = 1$, with $N = 15$ angle samples over a 14° aperture. The scatterer has a certain rectangular anisotropy and has circular migration with radius 0.6 meters. The overcomplete basis has $L = 5$ radii, with the R_l being 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and 1 meters. These different R_l are illustrated in Fig. 4-10 along with the true radius of migration overlaid on an image of the scene formed by conventional processing.

The inverse problem is solved with $K = 5$ frequencies 9.00 GHz, 9.49 GHz, 9.98 GHz, 1.05 GHz, and 1.10 GHz, by the quasi-Newton method. The coefficient magnitudes of the solution are shown in Fig. 4-11; two of the radii, $R_3 = \frac{1}{2}$ and $R_4 = \frac{3}{4}$, have non-zero coefficients corresponding to the true rectangular anisotropy. The true radius, 0.6, falls between $\frac{1}{2}$ and $\frac{3}{4}$, so the solution follows expected behavior. Putting different discrete values for radii into the overcomplete basis is somewhat limiting; a finer discretization of radius would have given more precise results. The next section

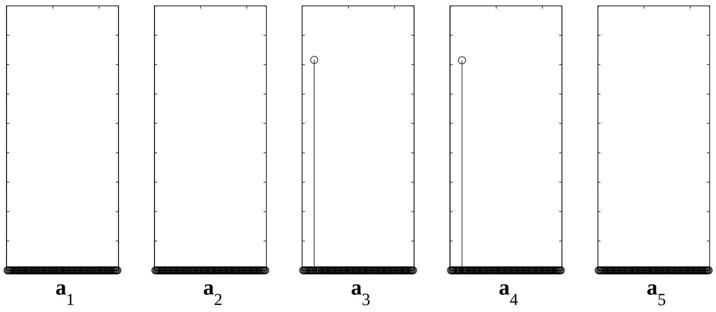


Figure 4-11: Magnitudes of coefficients in sparsifying regularization solution of migratory scattering example.

discusses a different approach to characterizing the radius of migration in circular migratory scattering centers that does not discretize radius.

4.2 Regression Over Radius

The overcomplete basis expansion with circularly migratory scattering centers:

$$\sum_{p=1}^{P}\sum_{l=1}^{L}\sum_{m=1}^{M} a_{p,l,m} b_m(\theta) \exp\left\{-j\frac{4\pi f}{c}\left((\bar{x}_p + R_{p,l})\cos\theta + \bar{y}_p \sin\theta - R_{p,l}\right)\right\}$$

requires the possible radii of migration to come from a discrete set of L values per scattering center. As seen in the previous section, this approach is quite effective, but causes the overcomplete basis to have $\mathcal{O}(LN^2P)$ basis vectors. Additionally, it is desirable to allow any value for the radius rather than just the chosen $R_{p,l}$. In this section, a method is developed that uses a $\mathbf{\Phi}$ matrix with $\mathcal{O}(N^2P)$ columns and converges to a radius that does not come from a discrete set, but may be any value.

4.2.1 Nonlinear Least-Squares

As described in Sec. 2.1.4 and Sec. 3.1.2, there are invariant spatial locations $(\bar{x}_p, \bar{y}_p)$ for each migratory scattering center p, but R_p is not known. In the technique considered in this section, the R_p are initially zero and then increased until settling onto hopefully the correct values. In a sense, the method is like blowing up P balloons simultaneously. The details are given below.

Starting with the P scattering centers and invariant positions $(\bar{x}_p, \bar{y}_p)$, the following vectors as functions of R_p are defined:

$$\boldsymbol{\varepsilon}_{k,p}(R_p) = \exp\left\{-j\frac{4\pi f_k}{c}\left((\bar{x}_p + R_p)\cos\theta + \bar{y}_p \sin\theta - R_p\right)\right\},$$

and $\boldsymbol{\phi}_{k,p,m}(R_p) = \mathbf{b}_m \bullet \boldsymbol{\varepsilon}_{k,p}(R_p)$. Concatenating the $\boldsymbol{\phi}_{k,p,m}(R_p)$ yields a matrix $\boldsymbol{\Phi}(\mathbf{R})$, where $\mathbf{R} = \begin{bmatrix} R_1 & R_2 & \cdots & R_P \end{bmatrix}^T$. For fixed anisotropy characterization coefficients $\mathbf{a}$, as $\mathbf{R}$ changes, aspect-dependent magnitudes for individual scattering centers remain the same, but phases do not. Phase affects how the contributions from the various scattering centers combine. The approach is to optimize over the radius vector $\mathbf{R}$ to minimize the data fit error $\|\mathbf{r} - \boldsymbol{\Phi}(\mathbf{R})\mathbf{a}\|_2^2$, but since anisotropy is also to be characterized with sparsity, the $\mathbf{a}$ in the previous expression is a sparse solution. The optimization problem is thus:

$$\min_{\mathbf{R}} \left\| \mathbf{r} - \boldsymbol{\Phi}(\mathbf{R}) \arg\min_{\mathbf{a}} \left\{ \|\mathbf{r} - \boldsymbol{\Phi}(\mathbf{R})\mathbf{a}\|_2^2 + \alpha \|\mathbf{a}\|_p^p \right\} \right\|_2^2. \tag{4.9}$$

To solve this problem, the outer minimization is performed using an iterative nonlinear least-squares algorithm trust-region reflective Newton to converge to an optimal $\mathbf{R}$ [15]. On every iteration, the update of $\mathbf{R}$ requires the evaluation of $\mathbf{r} - \boldsymbol{\Phi}(\mathbf{R})\mathbf{a}$, which requires the inner minimization to be carried out, using the quasi-Newton method of Sec. 4.1. The procedure is initialized with $\mathbf{R} = \mathbf{0}$, exactly the problem for stationary scattering centers, giving a first $\mathbf{a}$. On subsequent iterations, the inner sparse signal representation minimization is carried out with fixed $\boldsymbol{\Phi}(\mathbf{R})$, $\mathbf{R} \neq \mathbf{0}$. The outer minimization is terminated when the change in $\mathbf{R}$ falls below a small constant. Constraints such as $\mathbf{R} \geq \mathbf{0}$ or $R_p \leq R_p^{\max}$ may also be imposed on the outer minimization if prior information is available. Often, the coefficient vector $\mathbf{a}$ will retain its sparsity structure throughout the trust-region reflective Newton iterations because even for $\mathbf{R} = \mathbf{0}$, characterized anisotropy may be close to correct, or at least have the correct support. Examples of using this approach are given in the next section.

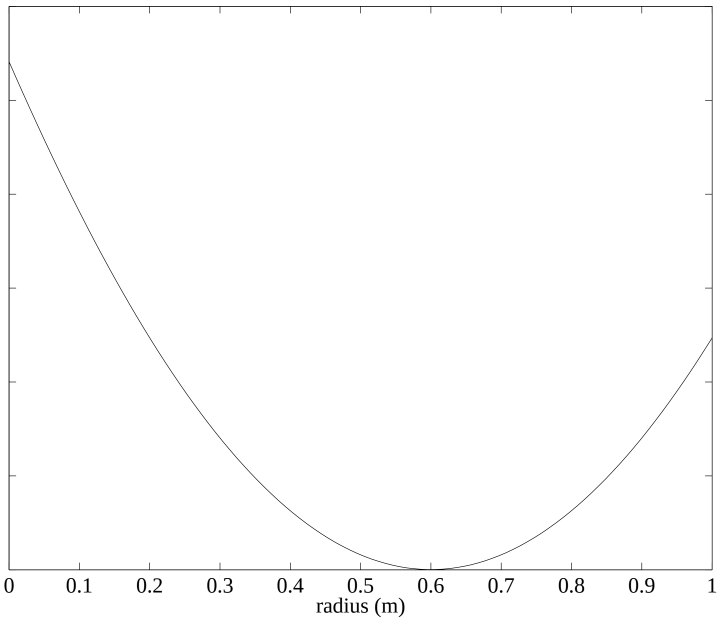


Figure 4-12: Cost function for minimization over radius in synthetic example.

4.2.2 Examples

The first example in this section is the same problem as the last example of Sec. 4.1.2, with a single scatterer with true radius 0.6. The cost function of the outer minimization for this problem is plotted in Fig. 4-12. The cost function is one-dimensional because there is one scattering center. Calculation of the cost function requires solving the inner anisotropy characterization problem for fixed radius. The minimum is located at the true radius of 0.6 meters. The numerical optimization technique returns a radius of 0.5999995 meters along with the correct anisotropy characterization. From this example, it seems that optimizing over radius is a promising technique and better suited to the task of migration characterization than the inclusion of basis vectors corresponding to different radii in the overcomplete basis. The next example is more complicated with more scattering centers in the scene.

In the second example, there are six scattering centers. Three of the scatterers are stationary point scatterers and the other three have non-zero radius, as shown

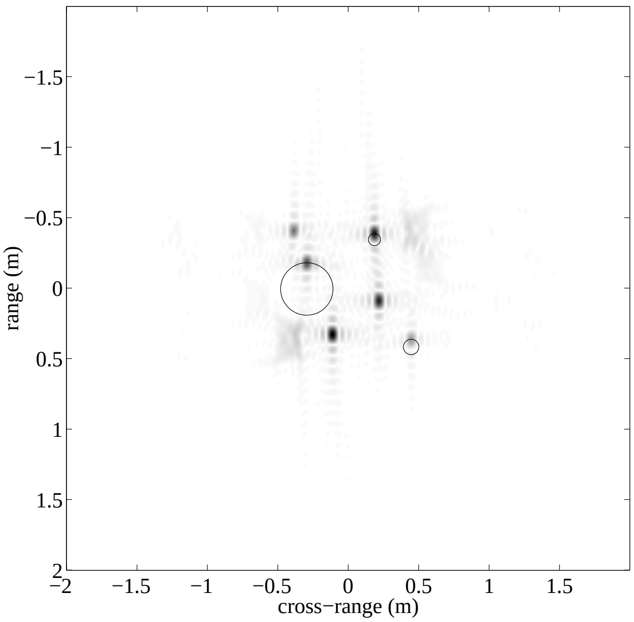


Figure 4-13: True migratory scattering radii overlaid on conventionally formed image.

in Fig. 4-13. $P = 6$ spatial locations are extracted using the method described in Sec. 5.2. In this case, the optimization over radius is of a six-dimensional cost function. There are $N = 19$ angle samples over 18° degrees and $K = 3$ frequencies 9 GHz, 10 GHz, and 11 GHz. The solution that is obtained is fairly good, i.e. the characterized anisotropy has the right support for all six scatterers and the characterized radii are fairly close to the truth, as seen in Fig. 4-14. The method of optimizing over radius works well in practice even with multiple scattering centers.

The final example of this section uses XPatch data of a tophat scatterer with the same angles and frequencies as the first example. The angle-dependent scattering, shown in Fig. 4-15, is peaky; hence raised triangle basis vectors are used. The invariant spatial location $(\bar{x}, \bar{y})$ is obtained from a conventionally formed image. Similarly to Fig. 4-12, the cost function of the outer minimization can be calculated for this example and is shown in Fig. 4-16. It should be reiterated that obtaining the cost function requires solution of the anisotropy characterization problem for every data

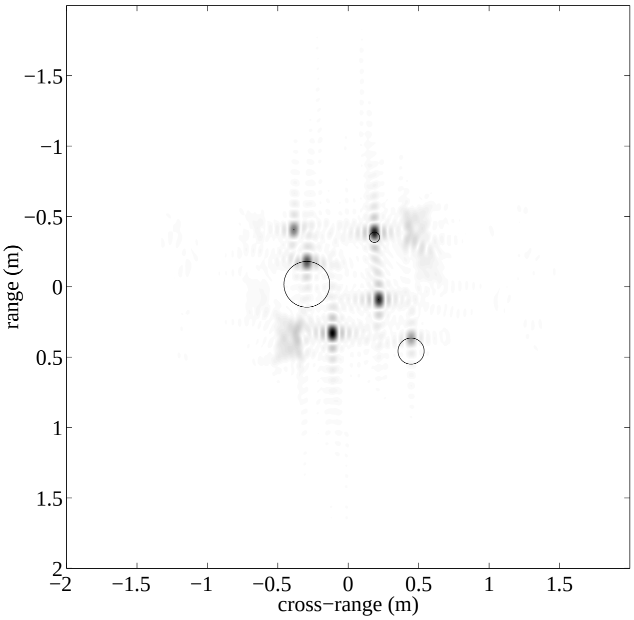


Figure 4-14: Solution migratory scattering radii overlaid on conventionally formed image.

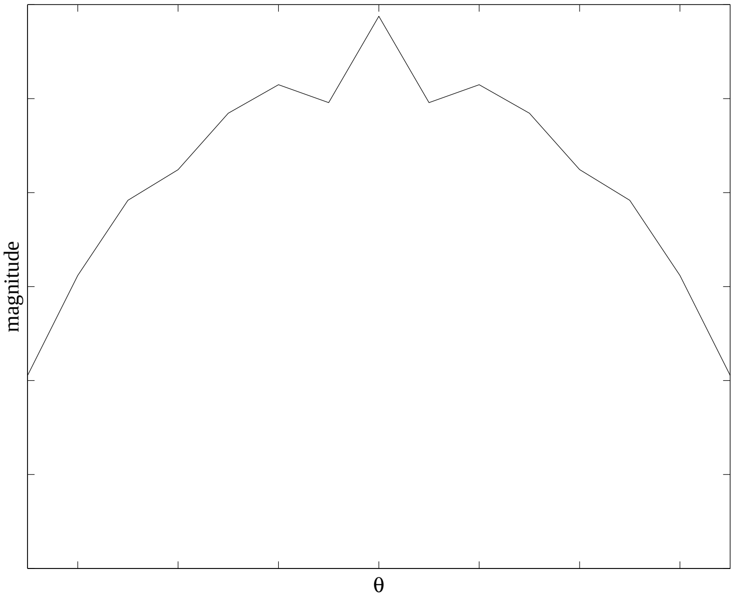


Figure 4-15: True aspect-dependent scattering function of tophat scatterer.

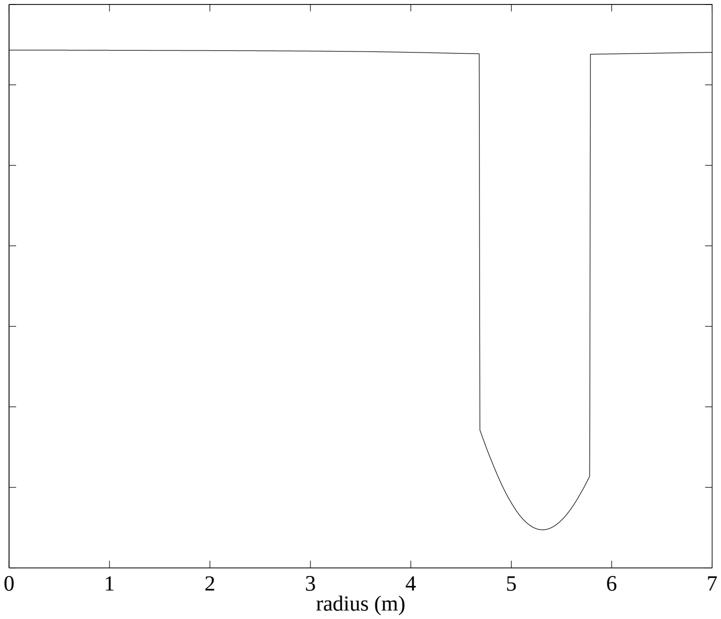


Figure 4-16: Cost function for minimization over radius in XPatch example.

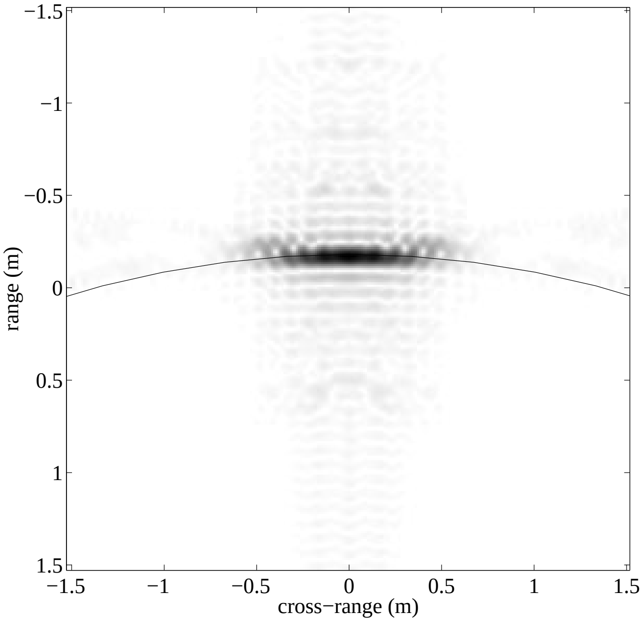


Figure 4-17: Image showing radius of minimum cost in XPatch example overlaid on conventionally formed image.

point on the curve. For the cost function plotted in the figure, the regularization parameter $\alpha = 55$. The cost in this example resembles that of the synthetic example only near the minimum and is nearly flat, but sloping inward, elsewhere; the almost flat portions are a consequence of regularization.

Initialized with $R = 0$, nonlinear least-squares optimization is successful at finding the minimum, the best radius. The solution is 5.314 meters; for reference, the circular migration corresponding to a radius of 5.314 meters is displayed in Fig. 4-17 and matches well with the conventionally formed image also shown in the figure.

In the examples considered in this section, the method of optimization over radius has been shown to perform well with realistic data as well as with several scatterers in a scene, with solutions giving information about both anisotropy and migration. Although not explored further in this thesis, a method using a coarse discretization in radius through the overcomplete basis may be used to provide an initialization for the outer minimization. The approach is extended in Chapter 6 to consider migration that is not just circular, but of other shapes as well.

4.3 Greedy Graph-Structured Algorithm

In theory, there is no restriction on the size of the problem that the quasi-Newton method can be applied to. However, the number of columns of $\mathbf{\Phi}$, which is $\mathcal{O}(N^2P)$ for the choice of overcomplete basis in Sec. 3.2, is restrictive in terms of memory as well as computation in realistic imaging scenarios with hundreds of angle samples and spatial locations. (In this section, only stationary point scattering is considered, but the methodology does not exclude migratory point scattering.) This section develops a suite of greedy algorithms with reduced memory requirements taking advantage of the graph structure described in Sec. 3.2.2.

4.3.1 Hill-Climbing Search

The idea to have in mind during the exposition of the algorithm is of a small subgraph, given the name *guiding graph*, iteratively moving through an N-level plinko graph (recall Sec. 3.2.2), searching for a parsimonious representation. The specifics of the guiding graph, the search strategy, and search steps are presented below.

In the N-level plinko graph, the nodes represent the basis vectors in the overcomplete basis. The basis is designed such that a few or often just one basis vector per position p' is sufficient to represent the aspect-dependent scattering function $s(x_{p'}, y_{p'}, \theta)$. Thus, the sparse signal representation problem may be reformulated as a search for a node or a few nodes in the plinko graph. In addition to finding nodes, complex amplitudes must also be determined. In general, there are $P > 1$ positions in the problem, and consequently P coexisting plinko graphs. Thus, to solve the problem, there is not just one search to be done, but P simultaneous searches. To be most effective, these searches should not be performed independently, but rather should interact and influence each other.

Plinko graphs can be converted to trees through node repetition, as shown in Fig. 4-18 for a 3-level plinko graph, transforming the problem to search on a tree.

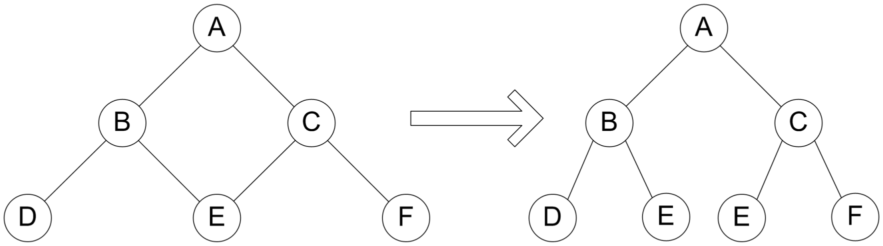


Figure 4-18: Conversion of plinko graph to tree.

Search on a tree is a well-studied problem with well-studied algorithms, with the only requirement that it should be obvious when the goal node is reached. Depth-first search and breadth-first search are two blind search algorithms that incorporate no prior information to guide the search. In heuristic search algorithms, however, some notion of closeness to the goal is available during the search process, allowing the search to proceed along paths that are likely to lead to the goal. Heuristic search algorithms have reduced average-case running time than blind search algorithms.

One example of a heuristic search algorithm is hill-climbing search, also known as guided depth-first search. In the depth-first algorithm, one path is followed from root to leaf in a predetermined way, such as: "always proceed to the left-most child." If the goal is not found along the way and the bottom of the tree is reached, then there is back-tracking also in a predetermined way. In hill-climbing, the search still proceeds from root to leaf along one path, but this path is not predetermined. Instead of always proceeding to the left-most child, the algorithm will "proceed to the most promising child based on a heuristic." Selection of back-tracking paths also involves use of the heuristic. Both depth-first search and hill-climbing search always find the goal if it exists in the tree.

As described thus far, hill-climbing search finds a single node — a single basis vector — per plinko graph. However, the algorithm proposed here is able to find a small subset of basis vectors by employing a guiding graph. The guiding graph is chosen to be a G-level plinko graph, $G \ll N$, with its root at the current node of the search. The search procedure moves the guiding graph around within the full plinko graph by moving the root node in the hill-climbing manner. Thus it is as if the guiding graph root node leads the search and the rest of the guiding graph is along for the ride; an alternative interpretation is as follows. Guiding graphs themselves may be taken as nodes in a search graph. An example of the same search with the two different interpretations for $N = 8$, $G = 4$ is shown in Fig. 4-19. With the alternate interpretation, the search is once again a search for a single node. The first interpretation is used in the remainder of this thesis.

The plan going forward is to use simultaneous hill-climbing searches on P plinko graphs to arrive at a solution for the anisotropy characterization sparse signal representation problem. Hill-climbing search requires heuristics and stopping conditions, which are discussed in the next section.

4.3.2 Heuristics and Stopping Conditions

In standard search problems on trees, nodes are labeled and the goal of the search is specified with a label, e.g. "find node K." Also, the goal remains fixed. Thus the stopping criterion for the search is simply whether the current node's label matches the goal of the search. Also, there is often a notion of intrinsic distance between nodes that leads to simple search heuristics.

When the sparse signal representation problem for a single spatial location is reformulated as a search on an N-level plinko graph, stopping criteria and heuristics are not obvious. One clear desideratum is that calculation of both should be much less memory and computation intensive than solving the full problem through the quasi-Newton method. Heretofore, the guiding graph concept has been used to allow a subset of basis vectors rather than a single basis vector to be used in the representation. However, the main role of the guiding graph is to guide the search by providing search heuristics and stopping conditions, as seen next.

Intuition about the problem suggests that if basis vectors corresponding to true scattering behavior are not included in the guiding graph when the inverse problem is solved in a sparsity enforcing manner, then the resulting solution coefficient vector $\mathbf{a}$ will have a non-zero coefficient for the basis vector most 'similar' to the truth. In terms of the N-level plinko graph, intuition suggests that if the truth is far down in

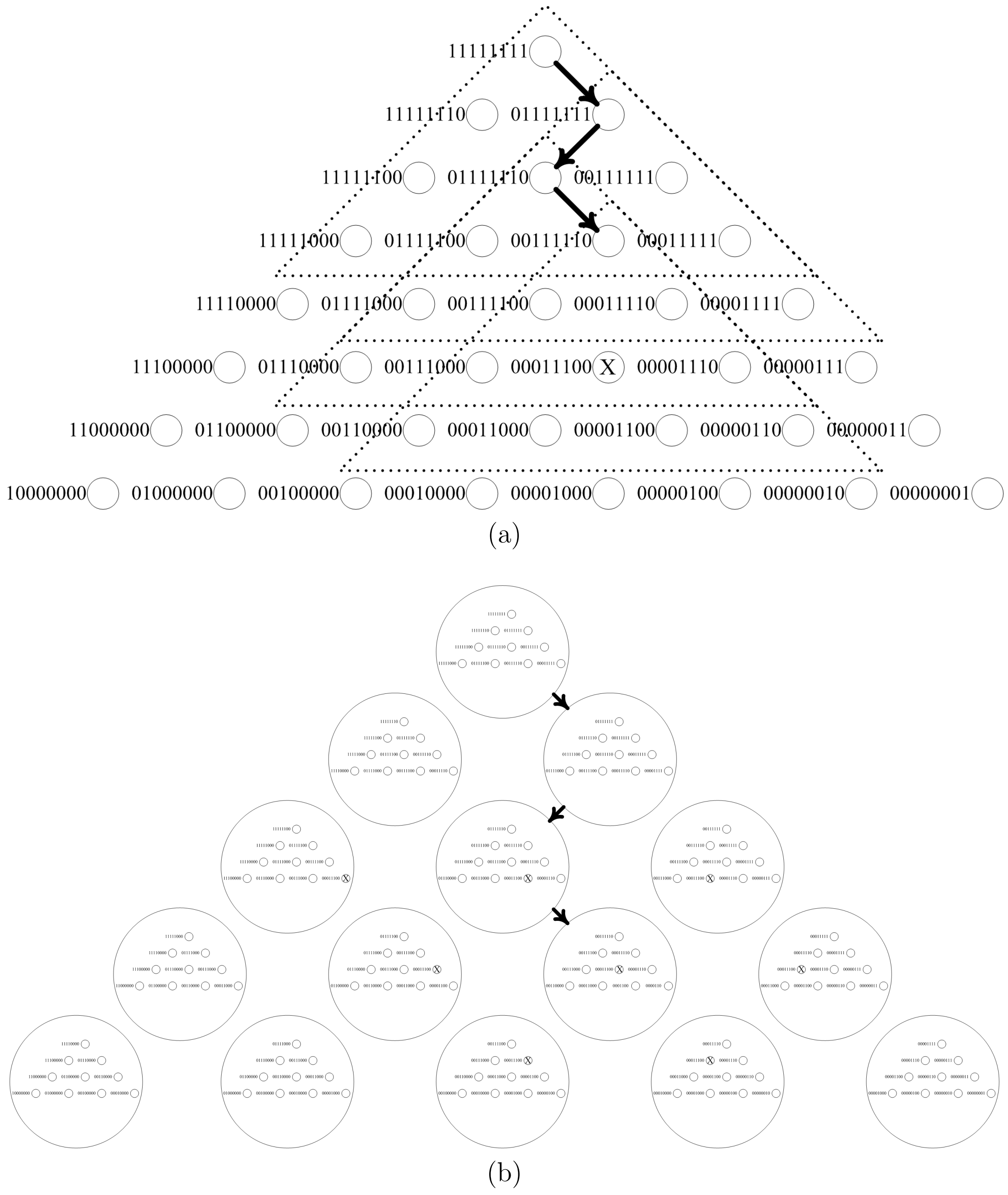


Figure 4-19: The (a) primary and (b) alternate interpretations of search with a guiding graph, where nodes marked with X indicate true anisotropy.

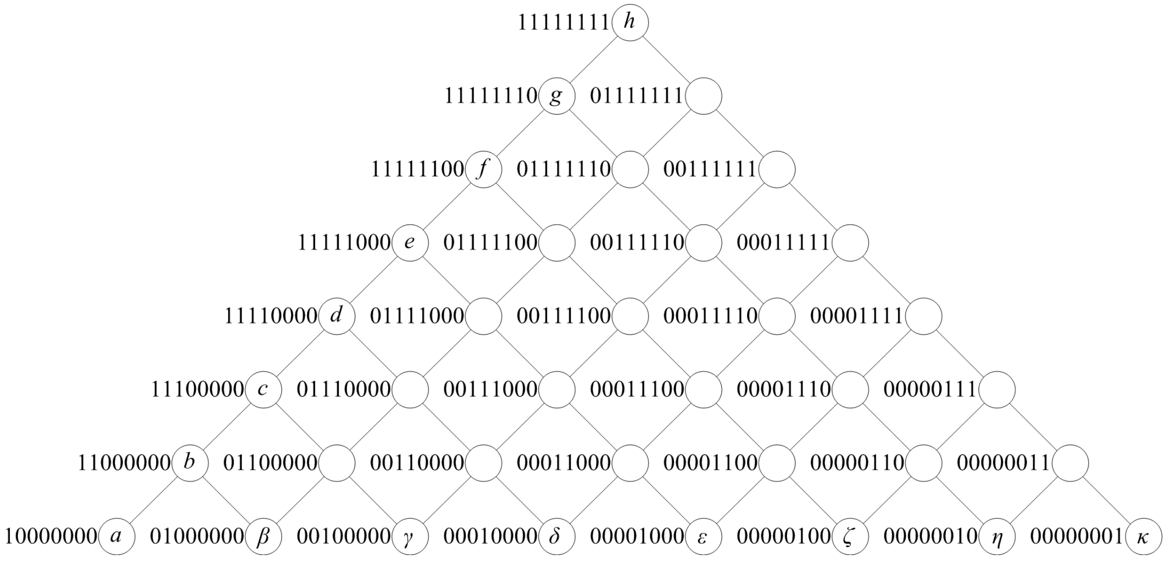


Figure 4-20: An 8-level guiding graph.

the plinko graph, but the inverse problem is solved with only basis vectors from a guiding graph near the top of the N-level plinko graph, then coefficients in the first $G-1$ levels will be zero and coefficients in level G may be non-zero. In the same vein, if the guiding graph is rooted below the truth, then the root coefficient may be non-zero and the coefficients in levels two through G will be zero. Again, intuition suggests that if the guiding graph is such that it contains the truth, then coefficients explaining the truth will be non-zero and the rest of the coefficients zero.

Before arriving at the search heuristics and stopping conditions, the above intuition is first confirmed through experimentation for $N = 400$, $G = 8$, and implicitly $P = 1$. The 400 angle samples are over the interval $[-55°, +55°]$, the number of frequencies $K = 3$ with values 7.047 GHz, 7.059 GHz, and 7.070 GHz, and the regularization parameter $\alpha = 150$. The 8-level guiding graph contains 36 nodes, as shown in Fig. 4-20. In the first experiment, with results in Fig. 4-21, the guiding graph is fixed with root at the left-most node of level 200 in the plinko graph. The true scattering behavior is varied from isotropic, to anisotropic with medium angular extent, to anisotropic with just one angle sample non-zero. In terms of the 400-level plinko graph, the true coefficient is varied, starting at the root node, through all nodes along the left edge of the graph, to the left-most node of level 400, as diagrammed in the left

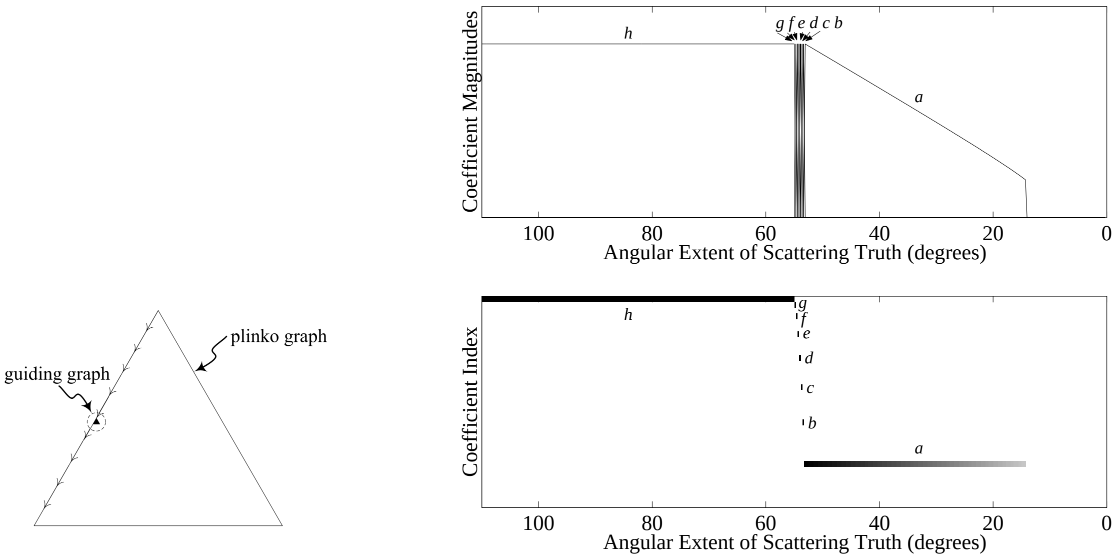


Figure 4-21: Coefficient magnitudes in 8-level guiding graph as true scattering behavior is varied from isotropic to highly anisotropic. The guiding graph is fixed with top node having angular extent 55.3° and eighth row nodes having angular extent 53.1°.

portion of Fig. 4-21. The large triangle is the 400-level plinko graph, the tiny filled triangle is the fixed guiding graph, and the arrows along the left edge indicate the variation of the true node. In the two plots, the angular extent of the true scattering behavior is plotted on the horizontal axis. In the top plot, the coefficient magnitudes for all 36 coefficients are plotted on the vertical axis, whereas in the bottom plot, coefficient magnitudes are indicated by shading (white is zero) and each horizontal strip is for each of the 36 different coefficients. The coefficient values are obtained by solving the inverse problem with sparsifying regularization. Most coefficients are zero for all true scattering behaviors in this experiment. Lines on the plots are labeled in correspondence with node labels in Fig. 4-20. The figure shows that in agreement with intuition, in the regime where the guiding graph is below the true coefficient, the root node (node h) is non-zero. In the regime where the guiding graph covers the true coefficient, the correct node is non-zero (nodes b-g). Also in agreement with intuition, when the guiding graph is above the true coefficient, the node in the last

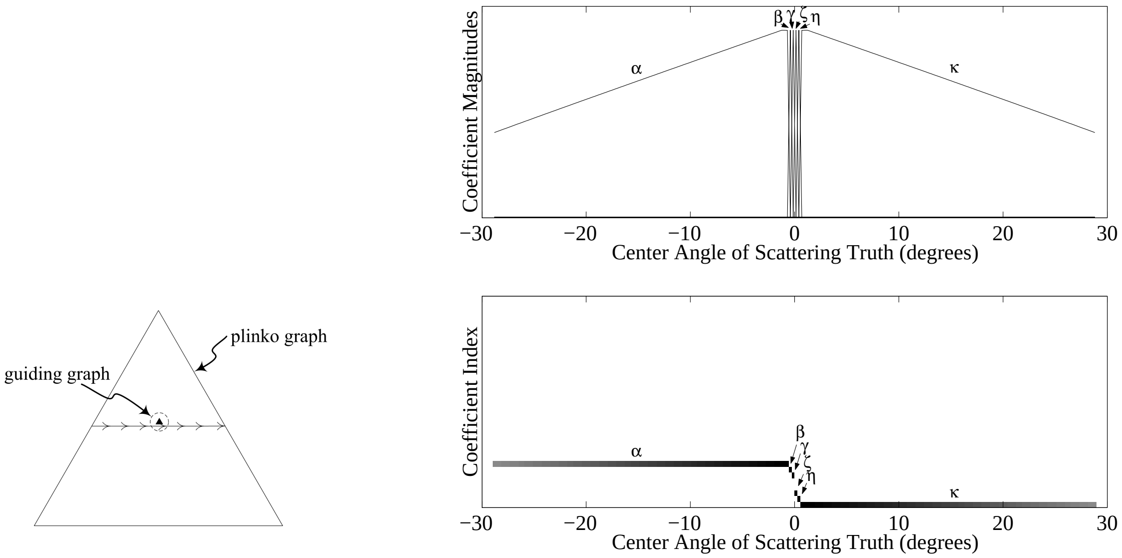


Figure 4-22: Coefficient magnitudes in 8-level guiding graph as center angle of true scattering behavior is varied. The guiding graph is fixed, covering center angles $[-1.0°, +1.0°]$.

level closest to the truth (node a) is non-zero and others are zero. It should be noted that the influence of the finest angular extent scattering, when the true coefficient has angular extent less than about 13°, does not reach up to make any guiding graph nodes non-zero (a consequence of regularization).

The experiment yielding the results of Fig. 4-22 has the same setup, but the guiding graph is fixed with root at the center node of level 200 instead of the left-most node. The true node is varied from left to right across the plinko graph at level 210, three levels below the bottom of the guiding graph, effectively changing the center angle of the anisotropy, but leaving the extent constant. This figure is organized in the same manner as Fig. 4-21, but the horizontal axis features the center angle rather than angular extent. From these results, first it is apparent that only nodes in the last level of the guiding graph are non-zero, reconfirming results from the previous experiment. Second, it can be seen that when the truth is to the left of the guiding graph, the left-most node of the G^{th} level (node α) is non-zero. Similarly, when the truth is to the right, the right node (node κ) is non-zero; when the truth is underneath the 8-level graph, nodes along the last level (nodes β-η) are non-zero.

Intuition along with these experimental validations suggests simple stopping conditions and heuristics. Two similar stopping criteria are apparent: first, stopping when all of the nodes in level G (nodes α-κ) are zero during the search, or second, stopping when in addition to all of the nodes in level G (nodes α-κ) being zero, the top node of the guiding graph (node h) is also zero. The second stopping condition is more amenable to back-tracking because it ends the search when the truth is neither above nor below the current guiding graph location. These two stopping criteria may terminate a search prematurely when the true anisotropy is of very small angular extent due to regularization effects as seen in the first experiment, in which anisotropy of narrow angular extent resulted in all coefficients becoming zero.

A heuristic for the search is also apparent based on the coefficient values of the G nodes in level G (nodes α-κ). Due to the structure of the plinko graph, each node has two children, so the heuristic is used to determine whether to proceed to the left child or the right child. Based on the second experiment, one reasonable idea is to take the weighted average of the coefficient magnitudes of the bottom level — the search can then be guided towards the side the average indicates to be stronger. The basis used in calculating the given heuristic and stopping criterion has $\mathcal{O}(1)$ columns for each spatial location and $\mathcal{O}(P)$ columns for P spatial locations, providing savings in terms of memory.

In summary, the greedy algorithm based on the plinko graph and hill-climbing search is as follows. For the case of a single spatial location, $P = 1$, the inverse problem $\mathbf{r} = \mathbf{\Phi}^{(i)}\hat{\mathbf{a}}^{(i)}$ is solved for each iteration i of the search, where $\mathbf{\Phi}^{(i)}$ contains basis vectors belonging to the current guiding graph. Then, $\hat{\mathbf{a}}^{(i)}$ is tested for the stopping condition. If the search is to continue, the heuristic is calculated to determine which one of two choices $\mathbf{\Phi}^{(i+1)}$ will be, i.e. where the guiding graph will proceed. In the depth-first approach, if a leaf of the N-level plinko graph is reached without the stopping condition being met, then there is back-tracking; the search back-tracks and again starts searching downwards starting at the highest guiding graph in the plinko graph for which $\hat{\mathbf{a}}$ had a non-zero top node. The initial set of basis vectors $\mathbf{\Phi}^{(1)}$ is the set with the largest angular extent located in the top G levels of the N-level plinko graph.

For the general case of multiple spatial locations, P searches are performed simultaneously, but not independently. As in the single position case, $\mathbf{r} = \mathbf{\Phi}^{(i)}\hat{\mathbf{a}}^{(i)}$ is still solved on each iteration, but now individual submatrices $\mathbf{\Phi}_p^{(i)}$ evolve based on their corresponding $\hat{\mathbf{a}}_p^{(i)}$. For example, the first spatial location's coefficients $\hat{\mathbf{a}}_1^{(i)}$ may satisfy the stopping condition, in which case $\mathbf{\Phi}_1^{(i+1)} = \mathbf{\Phi}_1^{(i)}$. The second spatial location's coefficients may indicate through the heuristic that the search should proceed to the left child, so $\mathbf{\Phi}_2^{(i)}$ is updated accordingly. The third spatial location's coefficients may guide the search on $\mathbf{\Phi}_3^{(i)}$ to the right child, and so on. The overall search terminates when all of the $\hat{\mathbf{a}}_p^{(i)}$ satisfy the stopping criterion. The P searches are coupled because the inverse problem is solved jointly for all spatial locations on every iteration; when there are multiple spatial locations, contributions from different positions interact.

The experimental confirmations leading to the heuristics and stopping criteria in this section were performed using rectangular pulse shapes. Harkening back to Sec. 3.2.3, overcomplete bases with the rectangular shape have smaller mutual coherence than other shapes as a result of not tapering at the edge of the contiguous segment of anisotropy. A related phenomenon presents itself when solving the inverse problem with only a subset of basis vectors. When the experimental validation of Fig. 4-22 is repeated with the house-shaped pulse, the same intuition follows, as seen in Fig. 4-23. However, with the triangle pulse shape the results are quite different, as shown in Fig. 4-24. Coefficients corresponding to interior bottom row nodes are never non-zero, coefficients corresponding to nodes α and κ are only non-zero over a limited range of center angles, and coefficients corresponding to nodes not in the bottom row are non-zero for a certain range of center angles. This dashes any hope of using search heuristics and stopping criteria based on guiding graphs for the triangle and other tapered pulse shapes.

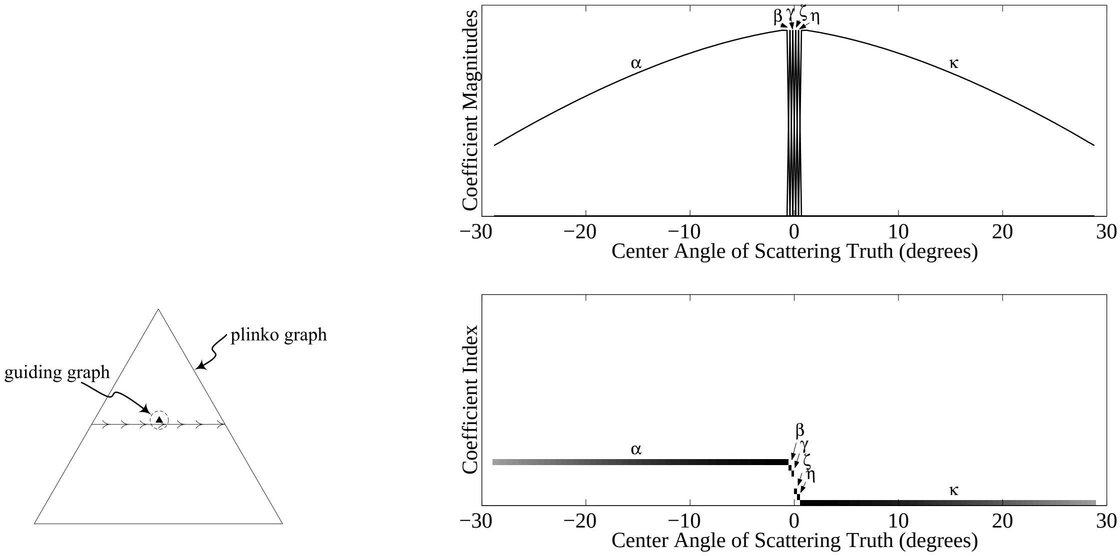


Figure 4-23: Coefficient magnitudes in 8-level guiding graph as center angle of true scattering behavior is varied, with raised triangle pulse shape.

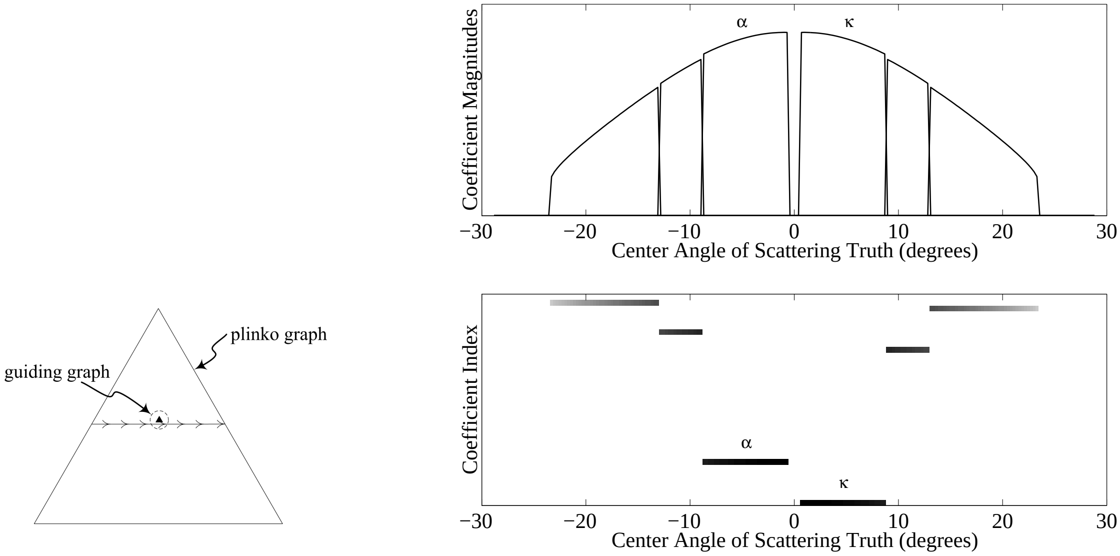


Figure 4-24: Coefficient magnitudes in 8-level guiding graph as center angle of true scattering behavior is varied, with triangle pulse shape. Beyond center angles of ±9°, coefficients corresponding to nodes not in the bottom row are non-zero.

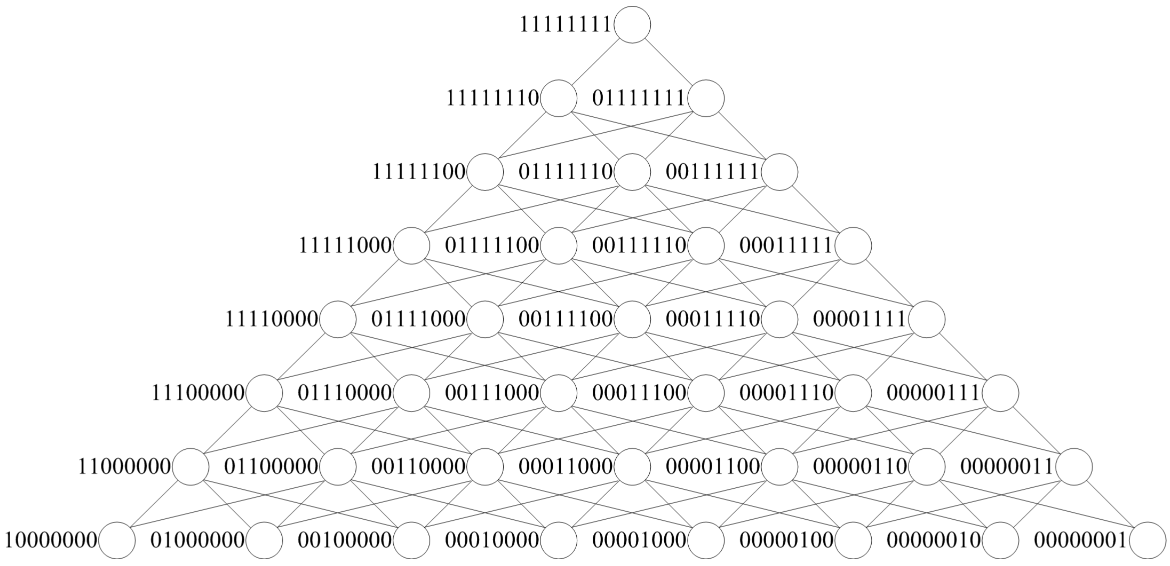


Figure 4-25: Modified plinko graph with additional edges.

4.3.3 Variations on the Theme

The greedy graph-structured algorithm described thus far uses the full hill-climbing search including back-tracking, taking steps of single levels per iteration based on a heuristic employing guiding graphs taking the form of G-level plinko graphs. A number of variations to the basic algorithm may be made that either make the method less greedy or further reduce memory or computation.

Hill-climbing search always finds the goal node because of back-tracking. In a first variation, the search is limited so that back-tracking is not allowed. Consequently, the success of the search is not guaranteed. However, it is observed in examples (see Sec. 4.3.4) that searches terminate on the first pass down plinko graphs before reaching leaves, in which case, the algorithm remains unchanged. Without back-tracking, the search takes $\mathcal{O}(N)$ iterations, whereas with back-tracking there are $\mathcal{O}(N^2)$ iterations.

The plinko graph is structured such that in a greedy hill-climbing search without back-tracking, one wrong step eliminates many nearby nodes and paths because each node has only two children. The graph can be modified to increase the number of children per node and consequently not disallow as many nodes and paths per search step. An example of a modified 8-level plinko graph is given in Fig. 4-25. A modified heuristic to go along with this modified graph is to use the G coefficients in level G of the guiding graph as before, but instead of determining whether the weighted average of the coefficient magnitudes is in the left half or the right half, determining which quarter the weighted average falls into. If the left-most quadrant is chosen, then the search proceeds to the node in the next level that is two to the left of the current node. If the middle left quadrant is chosen, then the next node is one to the left in the next level, and so on. In a way, the modified plinko graph allows the search to center itself above the true node. With these additional edges, the method is less greedy with no additional cost, since calculating this modified heuristic is no more costly than calculating the original heuristic. An example in Sec. 4.3.4 compares search paths and solutions of hill-climbing search without back-tracking with modified and unmodified plinko graphs.

The guiding graph need not be a G-level plinko graph; for example, the graph may be thinned and include the top node, nodes in level G, and nodes in a few intermediate levels rather than all intermediate levels, further reducing the number of columns in $\mathbf{\Phi}$. These basis vectors are sufficient for calculating the heuristic and stopping condition. Some examples in Sec. 4.3.4 make use of thinned guiding graphs. Also, searches may take larger steps than moving guiding graphs just down one level in the plinko graph per iteration.

The algorithm allows for contributions from more than one basis vector per spatial location in the final solution, but those basis vectors must be within the span of a guiding graph. The guiding graph may be enlarged to allow for contributions from disparate basis vectors at additional expense, the extreme being to take the guiding graph as the full plinko graph. The search strategy discussed so far uses a single guiding graph to traverse the plinko graph. Although not explored in this thesis, it may be possible to use heuristic search strategies from the literature other than hill-climbing, including strategies that allow multiple candidates in the search and multiple goal nodes, possibly through the use of multiple guiding graphs per spatial location.

The greedy graph-structured algorithm reduces the number of basis vectors per spatial location from $\mathcal{O}(N^2)$ to $\mathcal{O}(1)$, but does nothing to reduce the number of spatial locations P. A further variation to the hill-climbing search without back-tracking may be introduced that reduces the average-case dependence of the number of total basis vectors on P. It is observed that, despite interactions among contributions from different spatial locations, once the search on a particular spatial location stops it does not restart in general, but may occasionally restart after a few iterations. Fixing the contribution from a spatial location after its coefficients have been found is thus natural to consider. In the algorithm, this implies that once the stopping criterion is met at position p, the observation data $\mathbf{r}$ is updated to be $\mathbf{r}' = \mathbf{r} - \mathbf{\Phi}_p \mathbf{a}_p$, and $\mathbf{\Phi}_p$ is removed from matrix $\mathbf{\Phi}$, thereby reducing the number of columns in $\mathbf{\Phi}$. In practice, the removal is performed some iterations after the stopping criterion is met and maintained to allow for a possible restart.

One way to further reduce the number of spatial locations P is through a preprocessing step that first identifies positions of interest with large scattering magnitude and only characterizes anisotropy for those positions. The justification for this preprocessing is that locations with little scattering magnitude are likely background pixels and characterizing their anisotropy is not imperative for applications. Such preprocessing is discussed in Sec. 5.2, but is used in a couple of examples in Sec. 4.3.4.

Many other variations to the greedy graph-structured algorithm are possible that either reduce complexity or improve performance. Algorithms that use one variation or use a few variations together can be used to solve the inverse problem. Depending on the size of the problem and the requirements of the application, one algorithm can be selected from this suite of possible algorithms. In the next section, examples of varied problem size are given, solved with different variations of the greedy algorithm.

4.3.4 Examples

In this section, several examples are given that show the workings of the greedy graph-structured algorithm, culminating in a large, complicated example using the backhoe dataset. The first example in this section solves the same problem as in the first example of Sec. 4.1.2.

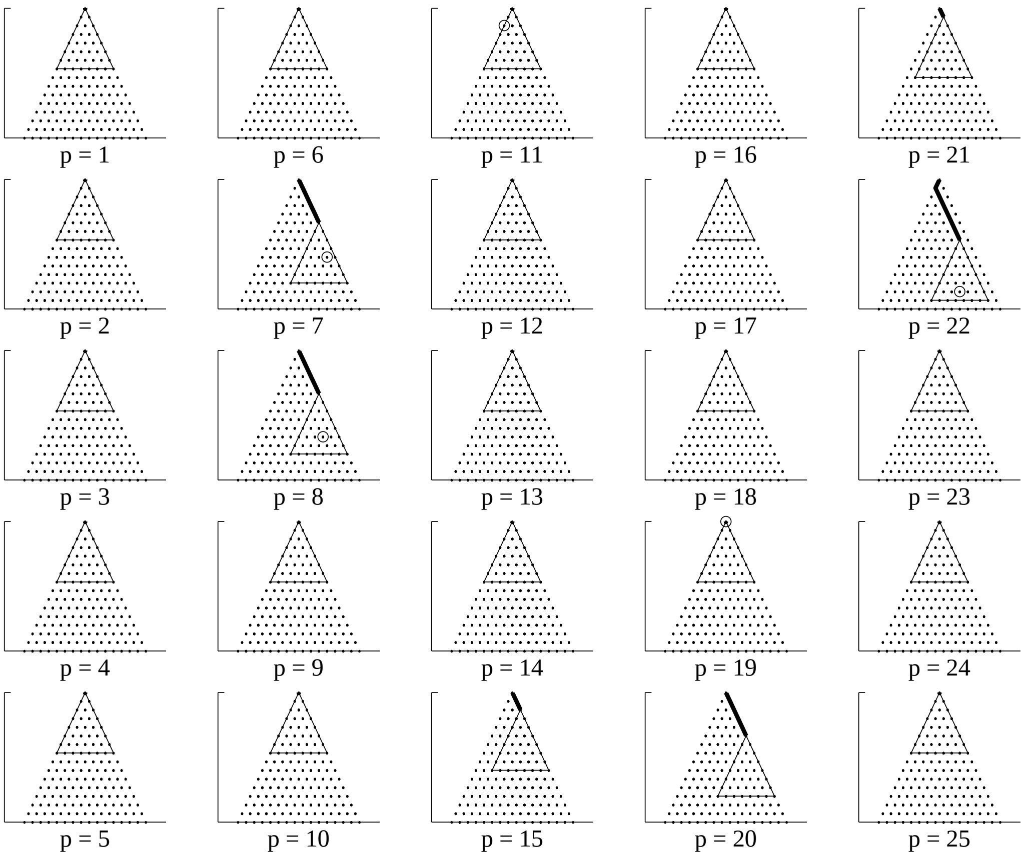

p = 1 p = 6 p = 11 p = 16 p = 21
p = 2 p = 7 p = 12 p = 17 p = 22
p = 3 p = 8 p = 13 p = 18 p = 23
p = 4 p = 9 p = 14 p = 19 p = 24
p = 5 p = 10 p = 15 p = 20 p = 25

Figure 4-26: Search paths.

Illustrating that the greedy algorithm described above may lead to solutions comparable in quality to solutions from solving the full inverse problem at once, the setup of the scene is exactly the same as the example in Sec. 4.1.2 with 25 locations. The heuristic and stopping criterion are calculated with an 8-level guiding graph, and the regularization parameter α is 4; for the stopping criterion, any coefficient magnitude less than 0.01 is considered zero. Fig. 4-26 shows the search paths for all 25 spatial locations as a dark solid line overlaid on the full 16-level plinko graph, with dots indicating nodes. For the 5 spatial locations with scatterers, the node corresponding to the true scattering behavior is encircled. The triangle shows the coverage of the final 8-level guiding graph for each spatial location. The search is performed without back-tracking. Fig. 4-27 shows the final scattering function determined by the algorithm. It is apparent in comparison with Fig. 4-5 that the greedy algorithm finds a

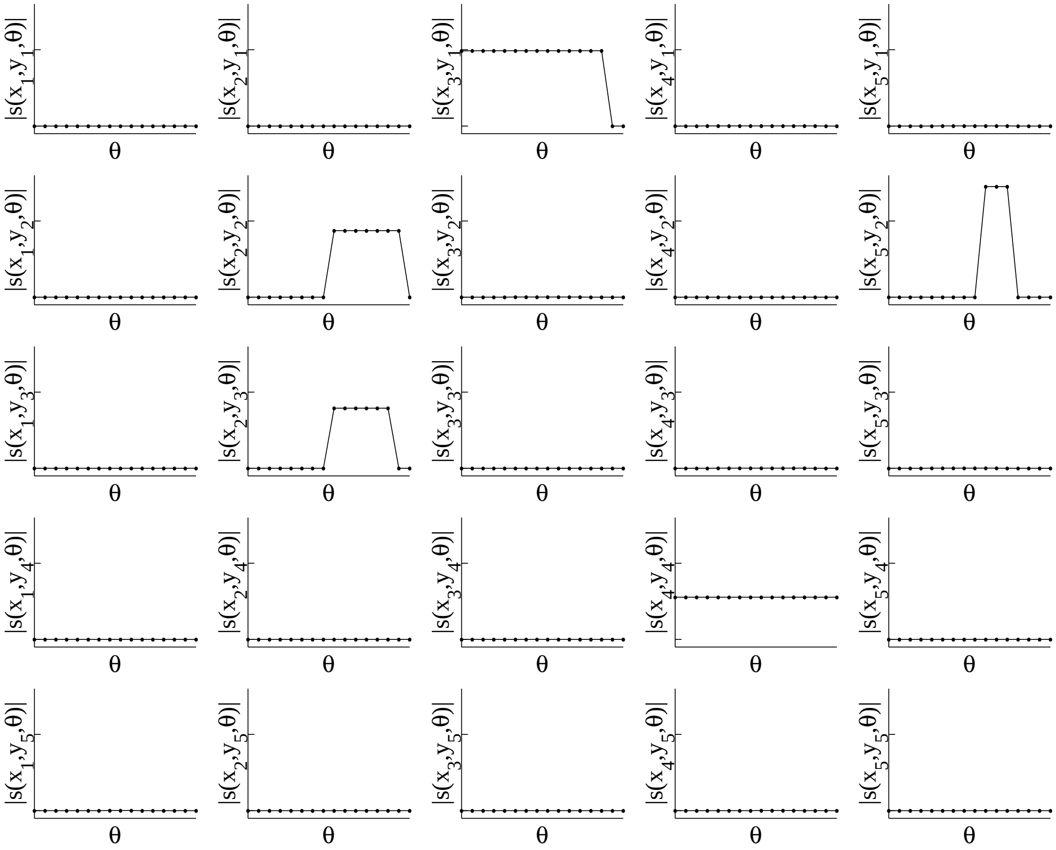


Figure 4-27: Hill-climbing search without back-tracking solution scattering magnitude.

solution that corresponds well to the underlying truth. Looking at the search paths, despite not containing point scatterers, positions 15, 20, and 21 iterate nonetheless, but in the end correctly give all zero coefficients. Also, the searches on positions 7 and 8 do not terminate at the earliest possible opportunity, but still give correct coefficients. These effects, although not adverse to the solution in this case, are a result of the interaction between different spatial locations. In addition to memory savings, the running time for the greedy algorithm is one fourth the running time of the quasi-Newton method on the full inverse problem for this example.

The problem size in the second synthetic example is larger than the first example. The range of angles is still $[-55°, +55°]$, but now there are $N = 160$ angles. The 3 frequencies and 25 spatial locations are as before and the anisotropy of the point scatterers is similar, as seen in Fig. 4-28. Memory constraints do not allow the full inverse problem to be solved in this case — the full $\mathbf{\Phi}$ matrix has $322,000$ columns. Two variations of the algorithm — with and without the additional step of removing spatial locations from the measurement vector and overcomplete basis after they have stopped iterating — are employed and compared. For both, the search uses an 8-level guiding graph; for the search with removal, the removal of found spatial locations is done after 8 iterations of the guiding graph remaining in the same place. The two search paths are given in Fig. 4-29, where the dotted line is the path for the algorithm without removal and the solid line is the path for the algorithm with removal. The large triangle outlines the full 160-level plinko graph. The resulting scattering function solutions are plotted in Fig. 4-28 along with the true scattering function. The three lines in the plots — the truth and the two solutions — are indistinguishable due to the similarity and quality of the results; in this and many other examples, the variation of removing found spatial locations to reduce average-case complexity does not cause the solution to suffer.

The third example is an example in which use of the modified plinko graph (see Fig. 4-25) improves the performance of the greedy algorithm. In this example, $N = 800$ over 110° and $K = 3$ with frequencies 9.9 GHz, 10.0 GHz, and 10.1 GHz. There are two scatterers with fairly close spatial locations, at $(0.05, 0.02)$ and $(0.02, 0.07)$,

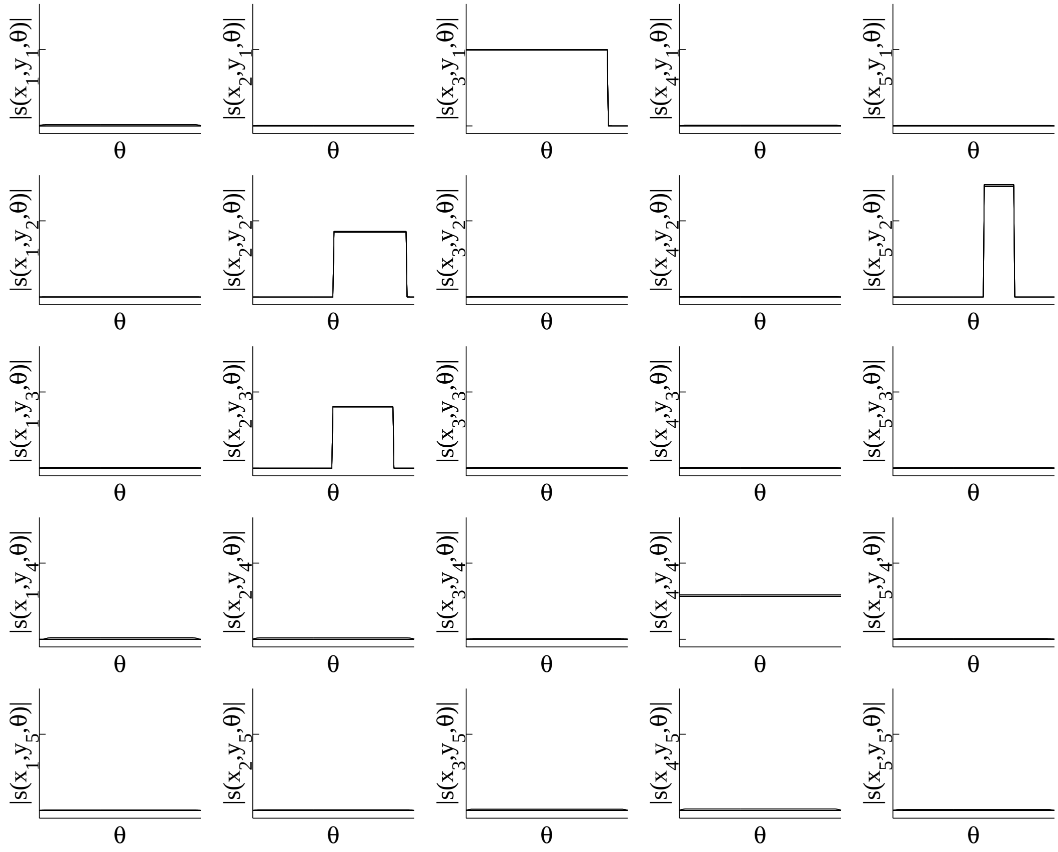


Figure 4-28: Scattering function magnitudes of truth, solution from hill-climbing search without back-tracking, and solution from hill-climbing search without back-tracking with removal of found spatial locations. The three lines are nearly indistinguishable.

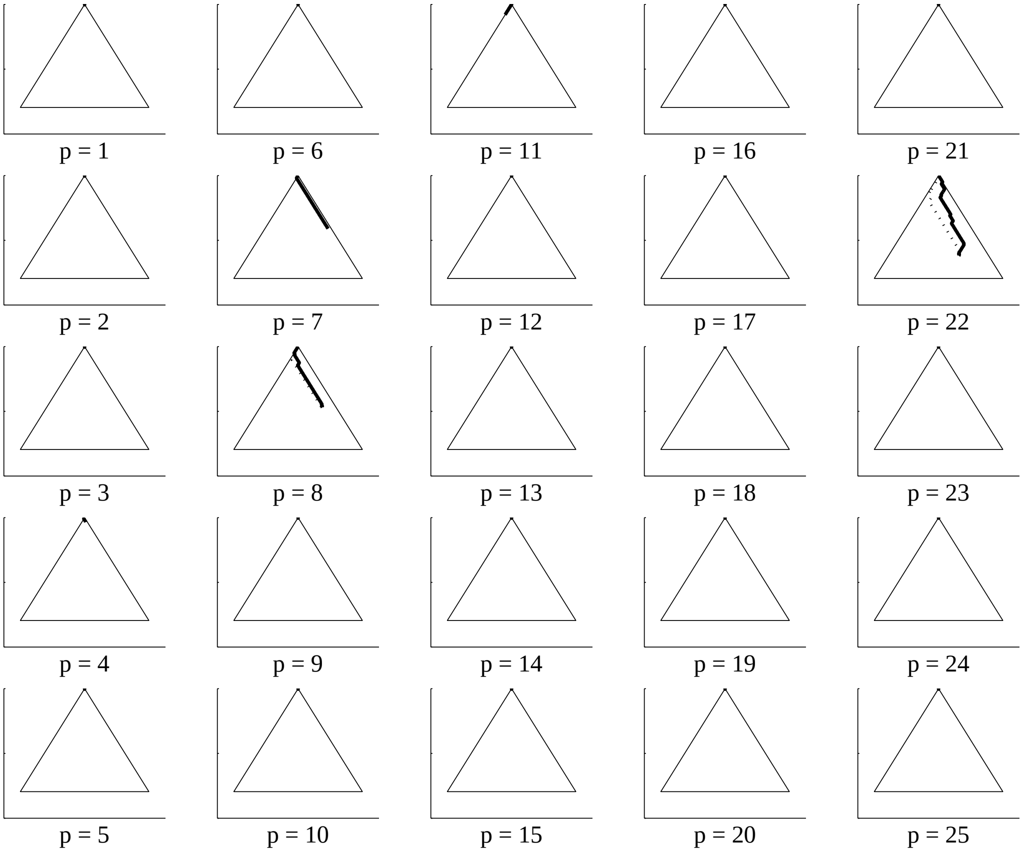
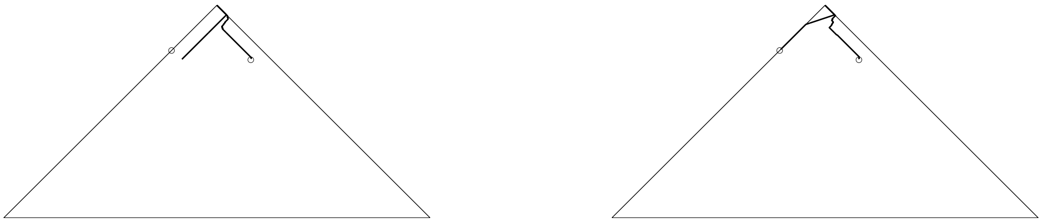


Figure 4-29: Search paths, where dotted line is path of algorithm without removal of found locations and solid line is path of algorithm with removal.



(a) (b)

Figure 4-30: Search paths with (a) unmodified plinko graph and (b) modified plinko graph. True anisotropy is indicated by $\circ$.

thus influencing each other during the early part of the greedy search process. Fig. 4-30 shows the search paths taken by hill-climbing search without back-tracking using the standard plinko graph and the modified plinko graph with additional edges, with 8-level guiding graphs. The searches on both spatial locations are shown on the same graph — the paths are coincident during the first few iterations — along with true synthetic rectangular anisotropies, indicated by $\circ$. Without the additional edges of the modified plinko graph, a single wrong step eliminates many possible guiding graph placements, but with the additional edges, more placements remain in play. In this example, a few wrong steps at the beginning of the search cause the characterized anisotropy of one of the spatial locations to be inaccurate. The additional edges reduce algorithm greediness without any increase in cost and in examples such as this, improve performance.

The fourth example makes use of XPatch data and uses an image formed by conventional processing to obtain spatial locations of interest. The conventional image, shown in Fig. 4-31, is used to extract positions with large magnitude scattering response. In this case, six locations (pixels) are selected, as shown in Fig. 4-32 (this image is zoomed in). The data has $N = 120$ angles equally spaced in an aperture of 119°. In forming the conventional image, 128 frequencies from 9 GHz to 11 GHz are

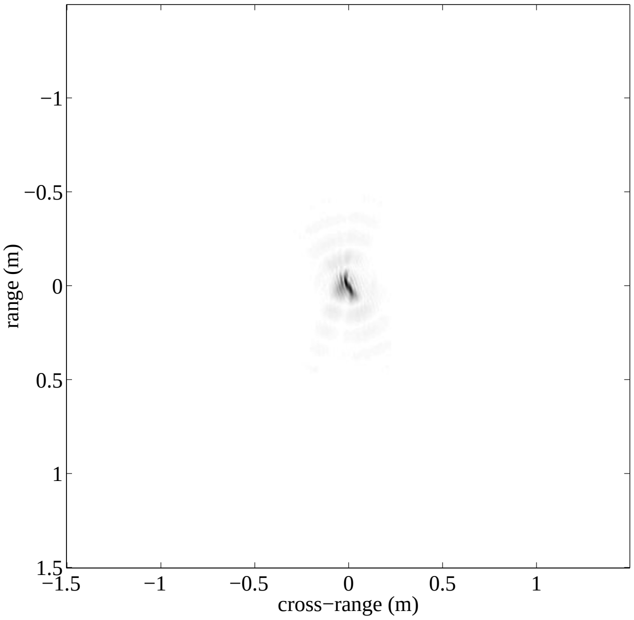


Figure 4-31: Conventionally formed image of example with XPatch data. Shading is on a linear scale.

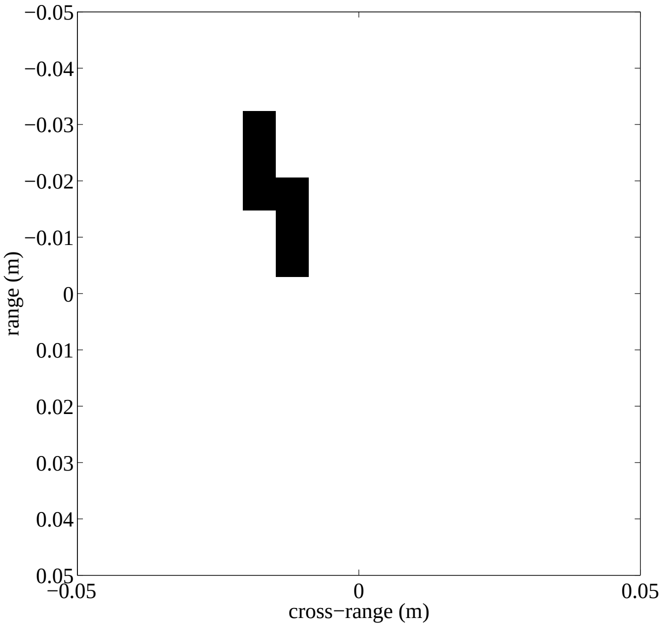
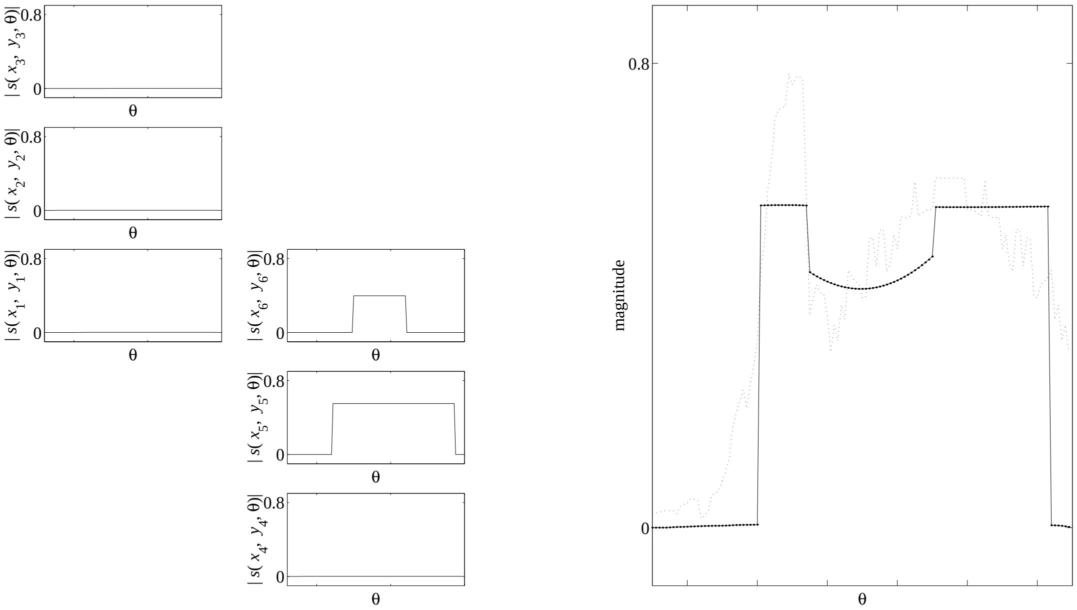


Figure 4-32: Spatial locations extracted from conventionally formed image.

used, but in anisotropy characterization, only data from the first $K = 3$ frequencies 9 GHz, 9.016 GHz, and 9.032 GHz is employed. The true aspect-dependent scattering is shown as the dotted line in Fig. 4-33b.

Hill-climbing search without back-tracking, with removal of found spatial locations, with the unmodified plinko graph, and 8-level guiding graphs is applied to the $P = 6$ spatial locations. The resulting scattering function magnitudes are shown in Fig.4-33a, arranged as in the image Fig. 4-32. Two spatial locations have non-zero responses that conspire to match the true scattering parsimoniously and well, as seen in Fig. 4-33b. With many closely spaced pixels, it is possible for the method to assign contributions to different positions such that the overall summed phase history is matched to the measurements when those contributions are not within the span of a single guiding graph. One way to reduce this effect is to cluster pixels in the conventional image that correspond to a single scattering center and only take one spatial location from the cluster as done in the next example and discussed in



(a) (b)

Figure 4-33: Scattering function magnitude solutions for (a) individual spatial locations and (b) all spatial locations combined as phase history — the marked line is the solution and the dotted line is the true scattering function magnitude.

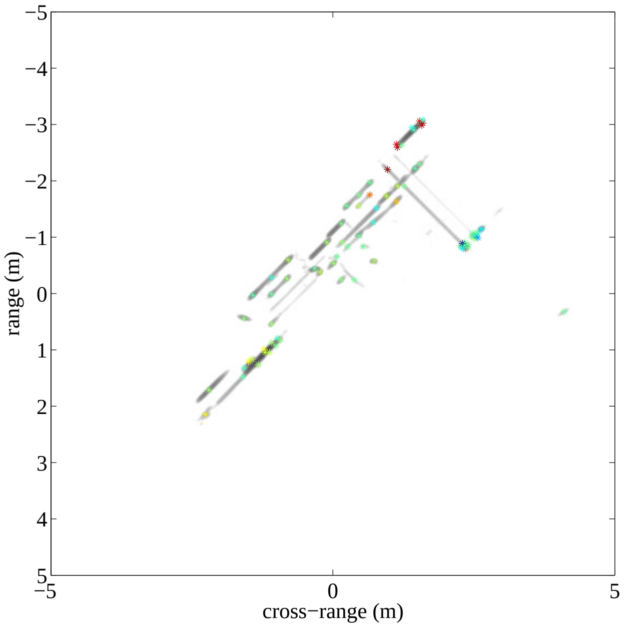


Figure 4-34: Backhoe spatial locations color-coded with anisotropy center of solution.

Sec. 5.2. Another way is to include additional terms in the sparsifying regularization cost function that explicitly favor sparsity in the image or range profile domains, as discussed in Sec. 6.1.

The final example, with a dataset from the Backhoe Data Dome [4], uses hill-climbing search without back-tracking and a thinned 16-level guiding graph. The data has $N = 1541$ angle samples over the range $[-10°, 100°]$. $P = 75$ spatial locations are extracted from a composite image of conventionally-formed subaperture images [12] and then the greedy graph-structured algorithm is applied to the data with $K = 3$ and frequencies 7.047 GHz, 9.994 GHz, and 12.953 GHz. Each of the 75 spatial locations, shown in Fig. 4-34, comes from a different cluster of pixels. The center angle of the characterized anisotropy in the solution is indicated in Fig. 4-34 as a color, where dark red corresponds to the center angle closest to $-10°$ and blue corresponds to the center angle closest to $100°$, with the colors cycling red to green to blue. The characterized anisotropies of a few scattering centers are shown in Fig. 4-

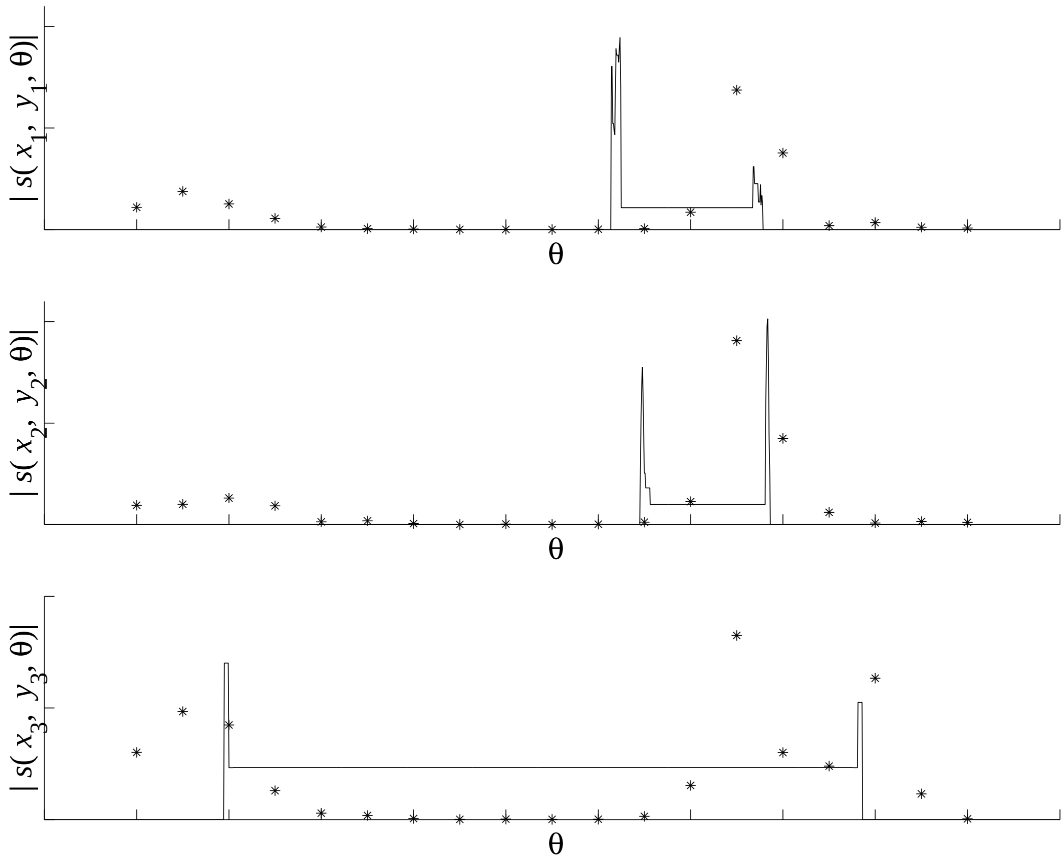


Figure 4-35: Sample scattering function magnitudes from backhoe example with subaperture pixel values also plotted as asterisks.

35. The solid line indicates the solution from the overcomplete basis formulation and graph-structured algorithm, whereas the asterisks show subaperture pixel values from conventional imaging with overlapping 20° subapertures [12]. The vertical axes for the line and the set of asterisks are scaled differently to allow comparison. For the first two scattering centers, the subaperture pixel values indicate contiguous extents of anisotropy and the overcomplete basis formulation also detects strong responses at those angles. However, the type of solution that may be produced is more detailed in θ, as subaperture analysis only gives one value per subaperture whereas the overcomplete basis formulation gives values for all angle samples spaced small fractions of degrees apart. The results indicate that anisotropy persistence is not matched to the 20° subaperture width because the solutions have thinner anisotropy. In the

third scattering center, the subaperture pixel values indicate two disjoint segments of anisotropy. However, the greedy algorithm may only use basis vectors that lie within a guiding graph to explain the anisotropy. Nevertheless, the algorithm does the best it can to produce two peaks via a positive-valued guiding graph root coefficient and negative-valued leaf coefficient. Multiple candidate search is an approach that would allow for better performance in such instances.

Chapter 5

Practical Issues

> Of beings, those that are endowed with life are superior. Of living beings, those that are endowed with intelligence are superior. Of intelligent creatures, men are superior. Of men, the twice-born are superior. Of the twice-born, students of the Veda are superior. Of students of the Veda, those of cultured understanding are superior. Of cultured men, practical persons are superior.
>
> — *Drupada, in Mahābhārata*[1]

In this chapter, a few lingering topics that are of practical importance are discussed. In particular, parameter selection in the sparsifying regularization cost function, preprocessing techniques to identify spatial locations of interest (used in examples in Chapter 4 and Chapter 6), and performance with noise are covered.

5.1 Parameter Selection

The sparsifying regularization cost function $J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p$ has two parameters, α and p, that must be specified. The regularization parameter α is a positive-valued number that trades off data fidelity for sparsity, whereas the parameter p of the ℓ_p-norm is between zero and one, and controls how close the relaxation is to the ℓ_0-norm. The effects of these parameters are looked at in this section.

In all of the examples in this thesis, the regularization parameter α is chosen manually to give good results. As a general trend it is true that as α increases,

---

[1] Translation from The Mahabharata of Krishna-Dwaipayana Vyasa, by Kisari Mohan Ganguli.

the number of non-zero coefficients decreases. However, this trend is not always monotonic; it is possible for the number of non-zero coefficients to actually increase as α increases. When α is too large, then the coefficients in resulting solutions are all nearly zero. The set of non-zero coefficients in solutions remains the same for long stretches as α is varied. However, there are sudden jumps where the set of non-zero coefficients discontinuously changes but then stays the same for a range of α afterwards.

A recently developed method to solve the ℓ_1-relaxation basis pursuit is able to efficiently trace out solution paths for the full range of α, allowing one to automatically select the regularization parameter if the desired number of non-zero coefficients is known a priori [37]. In general, automatic regularization parameter selection is an open research problem.

Although the trend that as α increases, sparsity increases, is not strict, the subjective choices made in this thesis have been guided by it. Enough sparsity is required when using the quasi-Newton method such that the contributions of different spatial locations are separated correctly in the inverse problem and there is not an overfitting of the data, but not so much so that complicated realistic scattering functions cannot be explained. When using the graph-structured algorithm, it is important to have sparse solutions so that the heuristics and stopping criteria can work as envisioned; thus, it is better to sacrifice some of the freedom to explain complicated scattering functions and get more sparsity.

The parameter p defining the ℓ_p-relaxation is set to $p = 0.1$ throughout the examples of the thesis. To remain faithful to the ℓ_0-norm, p should be as small as possible. The effect of p is seen most clearly without the complications of multiple spatial locations. If the measurement vector $\mathbf{r}$ is such that it equals column m of the overcomplete basis $\mathbf{\Phi}$, then as p approaches zero, in the solution $\mathbf{a}$, $(a)_m$ goes to one and the other coefficients go to zero. As p is relaxed, the value of $(a)_m$ falls off, but not too quickly. However as p reaches one, the falloff is significant. An example is shown in Fig. 5-1, with the line marked noiseless, for the case with rectangular pulse shape, $N = 8$, $K = 1$, $M = 36$, $m = 11$, spatial location $(0, 0)$, and α fixed

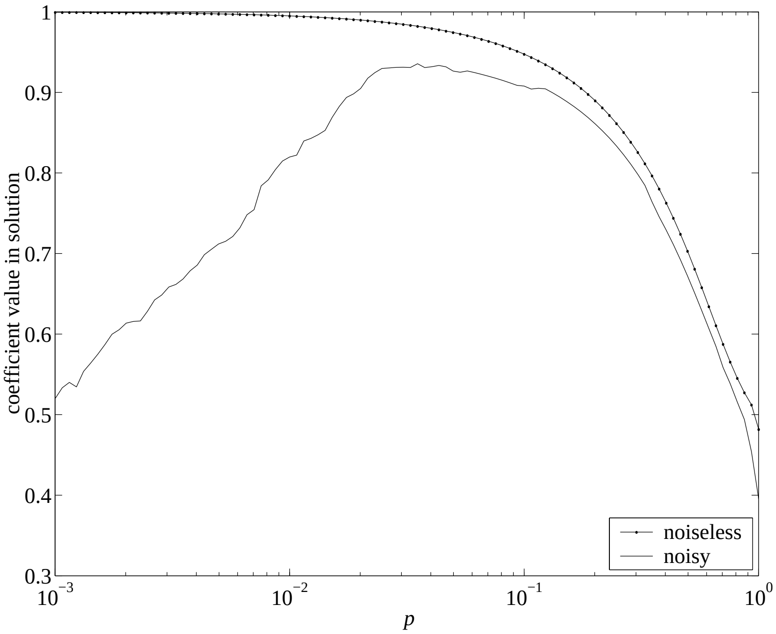


Figure 5-1: Solution as a function of p.

at 4. In the plot, p varies from 0.001 to 1 on a logarithmic scale. The decrease in coefficient $(a)_{m=11}$ is noticeable, becoming objectionable as p passes 0.1 or 0.2. Thus it is sensible to take p fairly small.

However, when p is very small, some issues come up. Consider these three exponentiations: $(10^{-100})^{0.1} = 10^{-10}$, $(10^{-100})^{0.01} = 0.1$, and $(10^{-100})^{0.001} = 0.794$. Any quantity, even a tiny value that is zero for all intents and purposes, goes to nearly unity when raised to the p power for p very small. If p is very small, then the solution is sensitive to even a small amount of noise added on to the measurements $\mathbf{r}$. With the same setup as above (the rectangular pulses have unit amplitude), the inverse problem is solved with zero mean white Gaussian noise with variance $\frac{1}{25}$ added on to the measurement. The average value of the coefficient $(a)_{m=11}$ from 100 trials is shown in Fig. 5-1 as the line marked noisy. When p is less than approximately 0.04, the solution is not consistent with the noiseless solution and truth. In light of the effects discussed when p is large and when p is very small, $p = 0.1$ is a reasonable value to use.

5.2 Preprocessing to Obtain Spatial Locations of Interest

The overcomplete basis $\mathbf{\Phi}$ has $\mathcal{O}(LN^2P)$ columns. Sec. 4.2 discusses a method to reduce the dependence on L by optimization over radius and Sec. 4.3 discusses methods to reduce the dependence on N through guiding graphs. The number of spatial locations P is reduced on average by removing positions that have stopped iterating in the graph-structured algorithm. However, it is better from a complexity perspective to reduce P before even coming to anisotropy characterization and forming $\mathbf{\Phi}$. This section discusses preprocessing to identify spatial locations of interest.

Forming images through conventional techniques such as the polar format algorithm is fairly quick in practice. Hence, conventional images provide a good starting point for preprocessing. Characterizing the anisotropy of locations with small scattering magnitude, which are likely background pixels, is not critical in many applications. Preprocessing determines spatial locations with large scattering magnitude; anisotropy is only characterized for those positions. There are various approaches to preprocessing; the approach used in examples here is the method of [33].

The method is a watershed algorithm that starts with the conventionally formed magnitude image, which can be treated as an elevation map. Large magnitude scattering corresponds to spire-like maxima in elevation. Water is filled so that it covers everything and then drained so that the maximum peak is exposed. Water continues to be drained, growing a region around the maximum. This cluster of pixels is taken as a group of spatial locations of interest. In the next step of the method, the cluster of pixels is subtracted away from the image and water is again filled. This process is repeated to get the set of spatial locations. As an alternative to taking the full cluster of pixels as spatial locations, just the peak of each cluster may be taken as the spatial location of interest.

5.3 Performance with Measurement Noise

The measurements in the phase history domain, $\mathbf{r}$, are never received unimpaired. A common model for measurement noise is additive white zero-mean circularly symmetric complex Gaussian noise [9]. The regularization cost function formulation handles additive Gaussian noise unchanged. This section briefly investigates the effect of noise, focusing on a tradeoff mentioned in Sec. 3.2.4, which informally is that the inverse problem is harder to solve but resistance to noise is better with an increased number of frequency measurements for a fixed bandwidth. A more comprehensive analysis of performance under noise corruption may be performed, but is not the focus here.

Let an example with a specific setup be considered; the conclusions generalize however. The scene has two stationary point scatterers at $(0.01, 0.01)$ and $(0.01, -0.01)$. Anisotropy characterization is performed using the quasi-Newton method with these $P = 2$ spatial locations. There are $N = 9$ angles over a 16° aperture, resulting in $M = 45$. The true anisotropies of the scatterers correspond to coefficients $m_1 = 13$ and $m_2 = 17$ respectively. The frequency band is from 9.9 GHz to 10.1 GHz with the number of frequencies K varying. In the regularization cost, $\alpha = 10$ and $p = 0.1$. Complex Gaussian noise with varying variance is added to the measurements.

The sum of the magnitudes of coefficients $(a)_{m_1}$ and $(a)_{m_2}$ in the solution are plotted in Fig. 5-2 as a function of K. Each curve is for a different noise variance and is the average from 500 trials. Solutions with the largest $|(a)_{m_1}| + |(a)_{m_2}|$ are the closest to the truth. The key feature to note in the curves is that the maximum is at a small value of K and that as noise variance increases, the peak of the curve shifts to higher values of K. The best solutions occur neither at the smallest value of K nor the largest value. Thus, it is advantageous not to use all frequency measurements. Not using all frequencies reduces the cross-coherence of the overcomplete basis, making the inverse problem less demanding, and has the added benefit of reducing memory requirements. Because measurements at additional frequencies give redundant information, it is to be expected that the peak of the curve moves to the right as noise increases — more

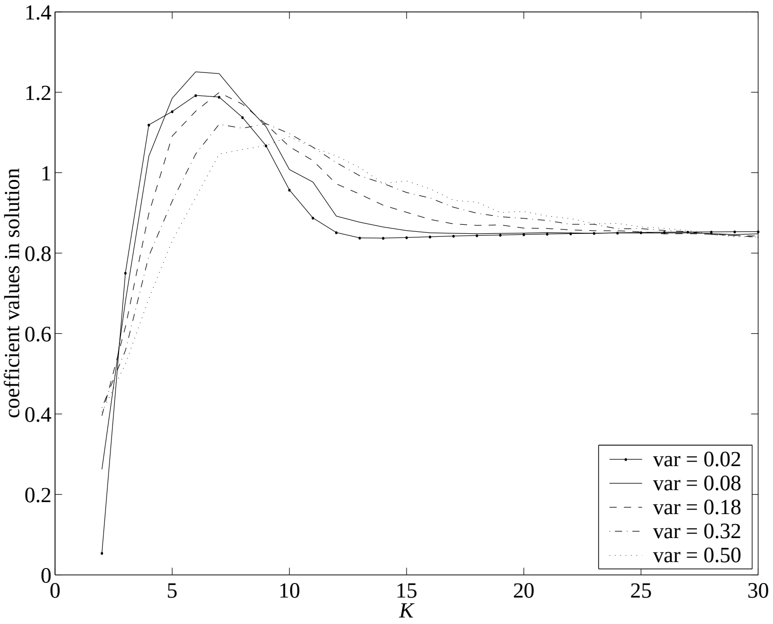


Figure 5-2: Solution as a function of K with additive Gaussian noise of different variances.

measurements, i.e. larger K, increase robustness to noise. Overall, the greater the noise power, the more frequencies should be used; however using too many frequencies is not of benefit either. Different analysis is required to give a well-justified statement about the overall robustness to noise of the anisotropy characterization procedures using the specific overcomplete basis.

Chapter 6

Objects and Image Formation

> A chariot is not asserted to be other than its parts, nor non-other. It also does not possess them. It is not in the parts, nor are the parts in it. It is not the mere collection of its parts, nor is it their shape. The self and the aggregates are similar.
>
> — *Chandrakirti (c. 650), in Madhyamakavatara*[1]

The model of point scattering is a model of the observation process, but scattering centers do not occur in isolation, they arise from real-world objects such as vehicles on the ground. The anisotropy characterization formulation for stationary scattering centers considered up to now has not considered interrelationships, especially geometric interrelationships in the x-y domain, among points or pixels as parts of objects. The work on migratory scattering centers builds up a notion of objects, but has thus far been limited in scope to circular objects. This chapter looks at extensions concerned with these two object-level issues in SAR imaging.

6.1 Object-Level Regularization

Different domains in which SAR measurements may be viewed were discussed in Chapter 2, including the phase history domain, the range profile domain, and the image domain. This section makes extensive use of the range profile domain in inserting

---

[1] Translation adapted from Emptiness Yoga, by Jeffrey Hopkins.

additional regularization terms related to spatial geometry into the regularization cost function. The range profile domain is closely related to the normal parameter space of the Hough transform, which has traditionally been used to detect parameterized shapes such as lines and circles in binary images.

6.1.1 Range Profile, Radon Transform, Hough Transform

The continuous range profile expression (2.3), reproduced here:

$$\hat{R}(\rho,\theta) = \iint\limits_{x^2+y^2\leq L^2} s\left(x,y\right)\delta\left(\rho - x\cos\theta - y\sin\theta\right)dx\,dy,$$

exactly expresses the Radon transform [49] of $s(x,y)$, $\mathscr{R}\{s(x,y)\}$, aside from the limits of integration. Taking $s(x,y)$ to be zero outside of the ground patch,

$$\mathscr{R}\{s(x,y)\} = \iint\limits_{-\infty}^{\infty} s\left(x,y\right)\delta\left(\rho - x\cos\theta - y\sin\theta\right)dx\,dy = \hat{R}(\rho,\theta). \tag{6.1}$$

The Radon transform is defined for all ρ and all $\theta \in [0,\pi]$. If $s(x,y)$ is real-valued then the Radon transform is also real-valued, but if $s(x,y)$ is complex-valued then $\mathscr{R}\{s(x,y)\}$ is complex-valued. Wide-angle SAR measurements cover a significant part of, but not the complete range of angles θ, and $\hat{R}(\rho,\theta)$ is in the form of samples, not a continuous field. Also, by using K-point one-dimensional DFTs of phase history measurements to obtain range profiles, data is not available for all ρ.

The Hough transform is not strictly a transform, but a method in image analysis for detecting parameterized curves in binary images [29]. The original formulation of the method was to detect straight lines parameterized by slope and intercept, but was improved and generalized in [21]. Still looking at the problem of detecting straight lines, the parameterization of [21], which is called the normal parameterization, uses the angle of a line's normal θ and its algebraic distance ρ from the origin of the image. With x and y as coordinates in the image plane, the equation for a line is $x\cos\theta + y\sin\theta = \rho$. The parameter space, the ρ-θ plane, and the image space, the

x-y plane, are related by the following properties: a point in image space corresponds to a sinusoid in parameter space and a set of points lying on the same line in image space corresponds to a set of sinusoids that intersect at a common point in parameter space. Also, a point in parameter space corresponds to a line in image space and a set of points lying on the same sinusoidal curve in parameter space correspond to a set of lines that intersect at a common point in image space. The Hough transform method of detecting straight lines makes use of these properties.

Let the binary image be such that the background is made up of zero-valued pixels and lines of one-valued pixels. Parameter space is gridded into ρ-θ cells and each one-valued pixel 'votes' for all cells along the sinusoid corresponding to that pixel. If many one-valued pixels are along a common straight line, then their corresponding sinusoids will intersect in one parameter space cell. With parameter space cells acting as accumulators of votes from image domain pixels, a cell with a high count indicates a line in image space. The approach has been extended with different parameters looking for different parameterized curves.

The normal parameter space of the Hough transform is a discretization of the Radon transform domain [55] and consequently corresponds to the discrete samples of the range profile domain. The Hough transform parameter space properties lend a geometric interpretation to the set of range profiles that is used in the next section.

6.1.2 Glint

Pixels may be treated as scattering centers, but this ignores the fact that a single point scatterer may correspond to a spatially distributed scattering mechanism. For one type of scattering, the migratory model is able to assign a single scatterer to a distributed area in image space, but cannot do so for spatially-distributed but stationary scattering. One important type of scattering, referred to as flash or glint that comes from long, flat metal plates has very thin anisotropy and corresponds to a line segment in image space oriented at the same angle as the center angle of the anisotropy. A parsimonious representation ought to explain scattering with a single scatterer rather than a collection of scatterers along a line. The goal of this section is to extend the regularization cost function (2.26) to favor sparsity along lines in addition to favoring sparsity among basis vectors, making use of the Hough transform properties discussed in the previous section.

In [1], a Hough space sparsifying regularization approach is employed to enhance and detect straight lines in positive real-valued images by imposing sparsity when taking the image data to the ρ-θ plane. Parameter space cells with small counts are suppressed and cells with large counts are enhanced; thus, non-line features are suppressed and line features are enhanced in image space, making the line detection problem painless. The goals in this section are very different from those in [1] and consequently, the regularization terms are of a different flavor as well: the Hough transform conception of accumulators to detect lines is turned on its head.

The idea here is to have sparsity in each cell of the ρ-θ plane rather than having sparsity among cells. As points on a line in the image domain transform to sinusoids coincident at a point in the range profile domain, sparsity among scatterers in individual ρ-θ cells achieves the goal of sparsity among points on a line. This qualitative description is translated into mathematical terms in the remainder of the section.

The regularization cost $J(\mathbf{a})$ is a function of the coefficient vector $\mathbf{a}$; consequently, in order to work with range profiles, the coefficients must be mapped to that domain first. P separate range profile planes, coming from each of the P scatterers, are required to achieve sparsity among the scatterers in ρ-θ cells. As discussed in Sec. 2.1.2, the range profile domain and the phase history domain are a single one-dimensional discrete Fourier transform away from each other. Also, the overcomplete basis $\mathbf{\Phi}$ is exactly the mapping from coefficients to the phase history domain. However, taking the coefficients through the overcomplete basis inherently sums the contributions of each spatial location coherently, which is undesirable when seeking to keep data from the P scatterers separate. Hence, in mapping from coefficients to a set of P range profile planes, a block diagonal matrix $\tilde{\mathbf{\Phi}}$ with $\mathbf{\Phi}_p$ matrices on the diagonal is used in conjunction with a matrix $\mathbf{F}$, which is like a DFT matrix. The values are exactly those that would appear in a $K \times K$ DFT matrix, but rearranged to fill an $N \cdot K$ by $N \cdot K$ area and replicated P times. An illustration is given in Fig. 6-1.

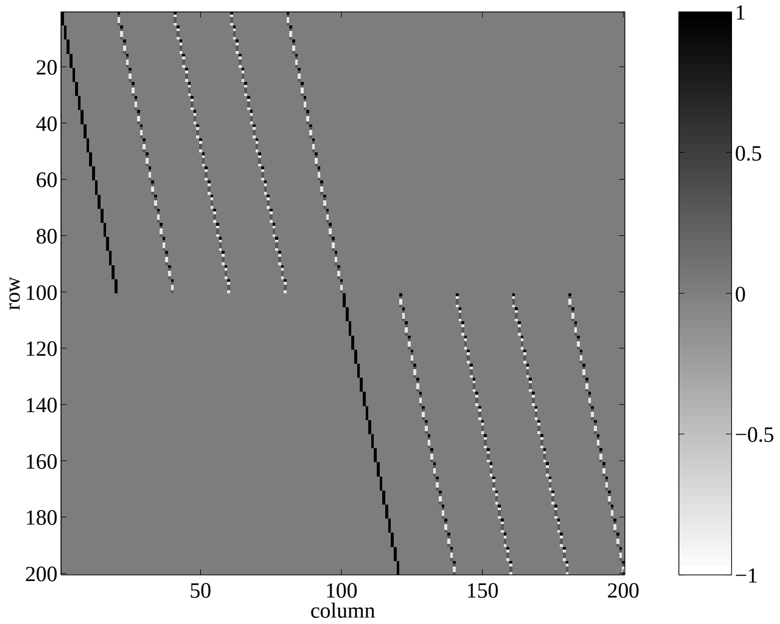


Figure 6-1: An illustration of the DFT-like matrix $\mathbf{F}$ for $N = 20$, $K = 5$, and $P = 2$. The real part of the values in the matrix are shown.

Additionally, to select data from a cell ($\rho = \rho_k, \theta = \theta_n$) in the range profile domain, a matrix $\mathbf{S}_{k,n}$ with P rows and $N \cdot K \cdot P$ columns composed of mostly zeroes and P ones is used. Specifically, $\mathbf{S}_{k,n}$ is defined as follows with entries indexed by row $i = 1, \ldots, P$, and column $j = 1, \ldots, N \cdot K \cdot P$:

$$(S_{k,n})_{i,j} = \begin{cases} 1, & j = (k-1)N + n + (i-1)NK \\ 0, & \text{otherwise} \end{cases} . \tag{6.2}$$

Thus, a length P vector of values for an individual range profile cell (ρ_k, θ_n) is obtained by the multiplication $\mathbf{L}_{k,n}\mathbf{a}$, where $\mathbf{L}_{k,n} = \mathbf{S}_{k,n}\mathbf{F}\tilde{\mathbf{\Phi}}$, and has P rows and $M \cdot P$ columns. The $\mathbf{L}_{k,n}$ matrices need not be calculated through matrix multiplication; the $\mathbf{F}\tilde{\mathbf{\Phi}}$ product may be calculated analytically in a straightforward manner based on the discrete Fourier transform of the pulse shape in the $\mathbf{b}_m$ vectors along with the phase term $\boldsymbol{\varepsilon}_{k,p}$, and the operation $\mathbf{S}_{k,n}$ simply involves extracting out the correct elements from the Fourier transform result.

It follows that for sparsity among scatterers in cell (ρ_k, θ_n), a regularization term of the form $\||\mathbf{L}_{k,n}\mathbf{a}|\|_p^p$, $p < 1$ is used. Then, continuing to maintain sparsity among basis vectors, the overall regularization cost function including sparsity in all range profile cells is:

$$J_{\text{line}}(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha_0 \|\mathbf{a}\|_{p_0}^{p_0} + \sum_{k=1}^{K}\sum_{n=1}^{N} \alpha_{k,n}\||\mathbf{L}_{k,n}\mathbf{a}|\|_{p_{k,n}}^{p_{k,n}}. \tag{6.3}$$

Taking the regularization parameters for all cells to be the same and taking the p of the ℓ_p-norm to be the same throughout, the cost function simplifies to:

$$J_{\text{line}}(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha_0 \|\mathbf{a}\|_p^p + \alpha_1 \sum_{k=1}^{K}\sum_{n=1}^{N} \||\mathbf{L}_{k,n}\mathbf{a}|\|_p^p. \tag{6.4}$$

This extended cost function $J_{\text{line}}(\mathbf{a})$ may be minimized using the quasi-Newton method described in Sec. 4.1, but with $\mathbf{H}(\mathbf{a})$ composed of additional terms. The $\mathbf{H}(\mathbf{a})$ matrix in this extension, $\mathbf{H}_{\text{line}}(\mathbf{a})$, is:

$$\mathbf{H}_{\text{line}}(\mathbf{a}) = 2\mathbf{\Phi}^H\mathbf{\Phi} + \alpha_0 p\mathbf{\Lambda}(\mathbf{a}) + \alpha_1 p\mathbf{\Lambda}_{\text{line}}(\mathbf{a}), \tag{6.5}$$

where,

$$\mathbf{\Lambda}_{\text{line}}(\mathbf{a}) = \sum_{k=1}^{K}\sum_{n=1}^{N} \mathbf{L}_{k,n}^H \mathbf{D}_{k,n}(\mathbf{a})\mathbf{L}_{k,n}$$

and

$$\mathbf{D}_{k,n}(\mathbf{a}) = \begin{bmatrix} \left(\left|\sum_{j=1}^{MP} (L_{k,n})_{1,j}\,(a)_j\right|^2 + \epsilon\right)^{p/2-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\left|\sum_{j=1}^{MP} (L_{k,n})_{P,j}\,(a)_j\right|^2 + \epsilon\right)^{p/2-1} \end{bmatrix}.$$

Accounting for the glint phenomenon, the extended formulation incorporating sparsity along lines in the spatial domain has now been stated. An example comparing the anisotropy characterization and image formation of glint scattering with and without the additional regularization terms is presented in the next section. Glint and sparsity among points on a line is just one imaging scenario, but an important one; other extensions to the regularization cost function for other scattering phenomena and objects may be developed, either based on properties of the Hough normal parameter space or other parameter spaces and domains.

6.1.3 Example

The example presented in this section uses XPatch data of glint type anisotropy and shows how the extended cost function with both sparsifying terms, the original one and the new one, leads to a parsimonious representation, whereas a cost with either of the sparsifying terms alone with the data fidelity term does not. The scene contains a single scatterer located at $(0, 0)$ with aspect-dependent scattering as shown in Fig. 6-2. There are $N = 20$ angles over a 19° aperture centered around zero degrees. There

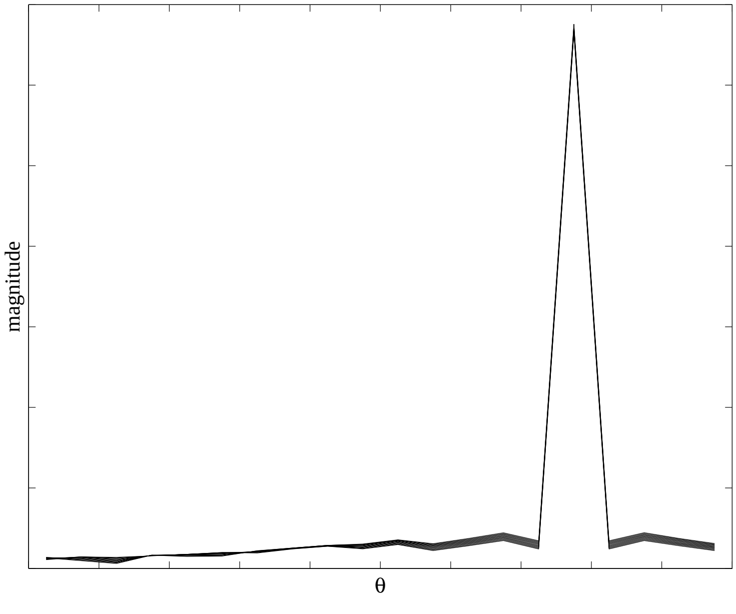


Figure 6-2: True scattering magnitude of glint anisotropy from XPatch data, with lines for ten different frequency measurements.

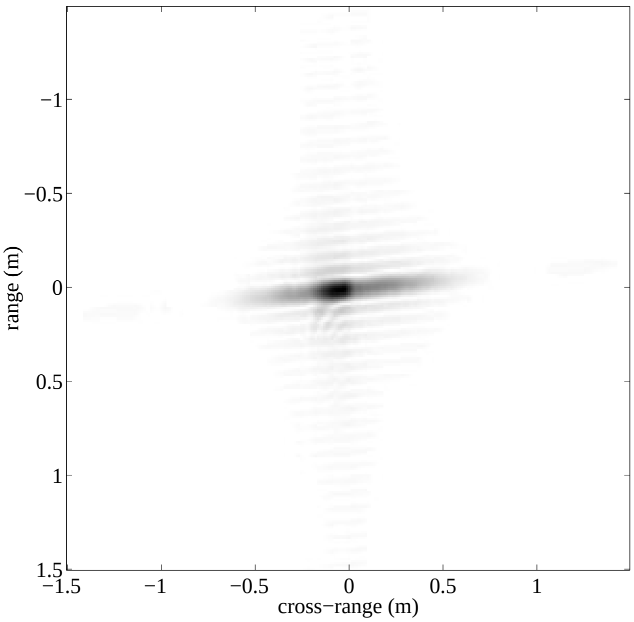


Figure 6-3: Conventionally formed image of glint anisotropy.

is a spike in scattering response at 5.5°, which is the flash or glint. The figure shows the magnitude of the scattering at ten different frequencies in this XPatch data — since there is almost no frequency dependence, the lines are nearly indistinguishable.

In a conventionally formed image using data with a bandwidth of 2 GHz, Fig. 6-3, the glint shows up as a spread out line segment oriented at 5.5°. From this image, $P = 24$ pixels are chosen as spatial locations for joint anisotropy characterization and image formation. The spatial locations range from $-\frac{9}{90}$ m to $-\frac{1}{90}$ m in the x direction and from $-\frac{1}{90}$ m to $\frac{2}{90}$ m in the y direction, with a uniform pixel spacing of $\frac{1}{90}$ m in both directions.

Then, with $K = 10$ frequencies in the range 9.00 GHz to 9.14 GHz, the anisotropy is characterized with three different pairs of values for the regularization parameters α_0 and α_1. The first set of regularization parameters is $\alpha_0 = 30$ and $\alpha_1 = 0$, i.e. without the extension to the cost function discussed in the previous section. The magnitudes of the coefficients for the twenty-four spatial locations are plotted in Fig. 6-4, ar-

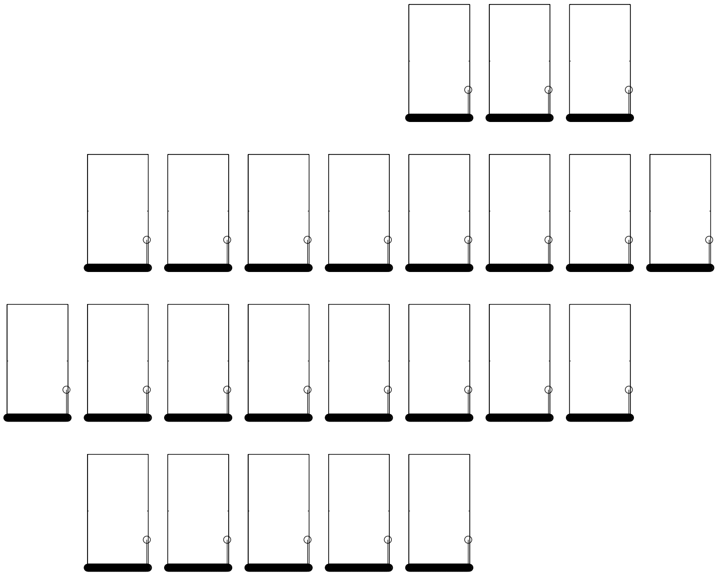


Figure 6-4: Solution coefficients in glint anisotropy example, $\alpha_0 = 30$, $\alpha_1 = 0$.

ranged as in an image, and the scattering function magnitudes for each of the spatial locations are given in Fig. 6-5, also arranged as in an image. The anisotropy has been characterized correctly, but split up and assigned to all of the spatial locations. This solution is parsimonious in basis vectors per spatial location, but is not parsimonious in the number of spatial locations used.

The second set of regularization parameters is $\alpha_0 = 0$ and $\alpha_1 = 20$: just sparsity among spatial locations along a line. As seen in Fig. 6-6, the solution in this case has non-zero coefficients at just one spatial location. This spatial location is the closest among all $P = 24$ spatial locations to $(0, 0)$, the true location of the scatterer. However, there are many coefficients with large values, not just one as in the previous case. The coefficients and corresponding basis vectors are such that they add to match the true anisotropy well, as seen in Fig. 6-7, but the representation is not parsimonious in terms of basis vectors per spatial location.

The third set of parameters is chosen such that both sparsifying terms in the regularization cost function are significant. With $\alpha_0 = 30$, $\alpha_1 = 20$, the solution

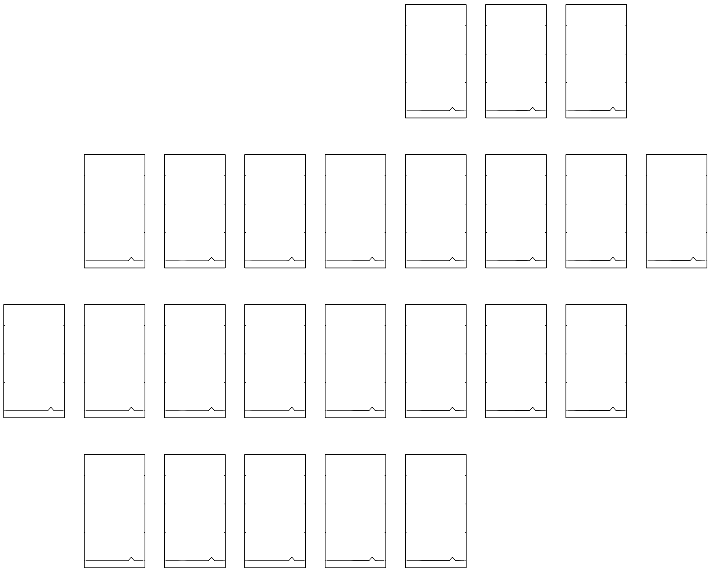


Figure 6-5: Solution scattering magnitudes in glint anisotropy example, $\alpha_0 = 30$, $\alpha_1 = 0$.

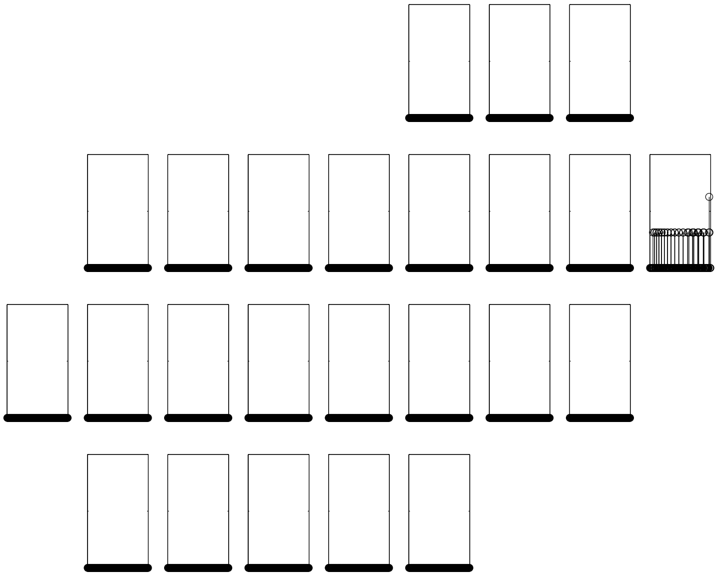


Figure 6-6: Solution coefficients in glint anisotropy example, $\alpha_0 = 0$, $\alpha_1 = 20$.

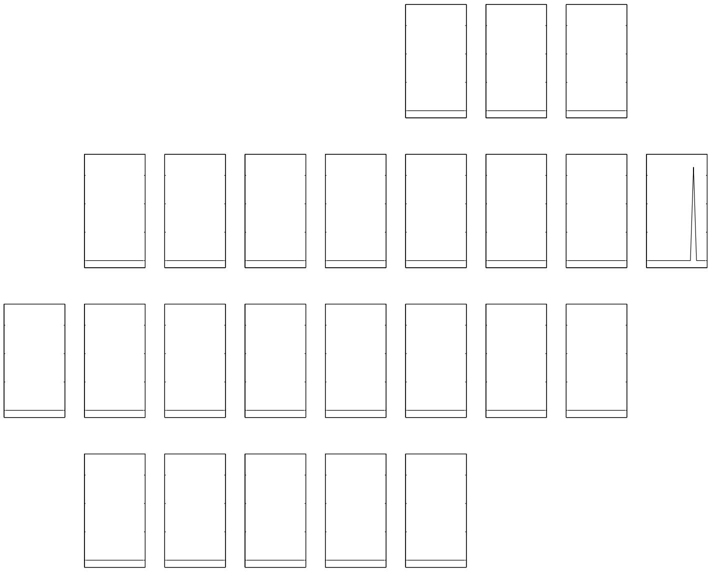


Figure 6-7: Solution scattering magnitudes in glint anisotropy example, $\alpha_0 = 0$, $\alpha_1 = 20$.

coefficient vector has only one non-zero coefficient seen in Fig. 6-8. The coefficient corresponds to a basis vector with a single non-zero angle sample, shown in Fig. 6-9, and is thus parsimonious in both spatial locations and basis vectors.

The original sparsifying regularization cost function has the effect of favoring solutions with sparsity among spatial locations, seen in examples of Chapter 4, because the vector **a** has coefficients associated with all spatial locations. The additional regularization terms of this section also favor sparsity among spatial locations because spatial locations along a line are general spatial locations as well. However, the distinguishing characteristic of the additional regularization terms is that the favored sparsity is specially adapted for the object-level idea that individual point-scattering centers affect linear regions in space.

Through the example it has been seen that both types of sparsity — sparsity among basis vectors and sparsity among spatial locations along a line — are necessary in the regularization in order to recover a solution that represents the scattering

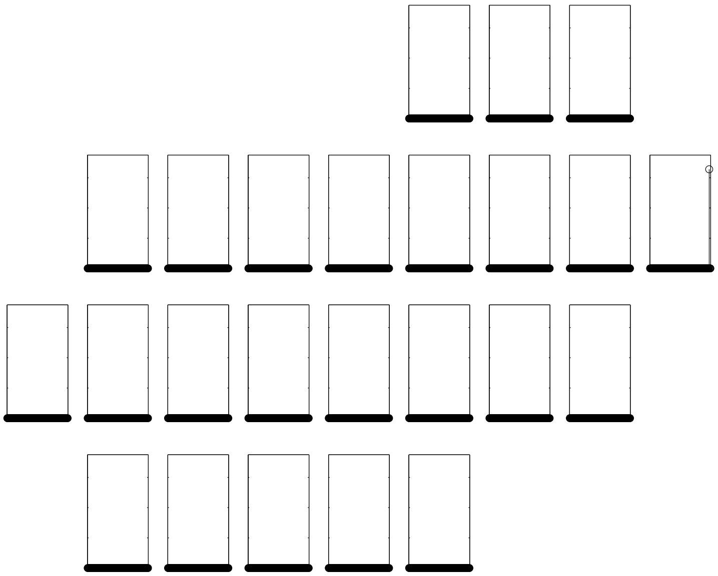


Figure 6-8: Solution coefficients in glint anisotropy example, $\alpha_0 = 30$, $\alpha_1 = 20$.

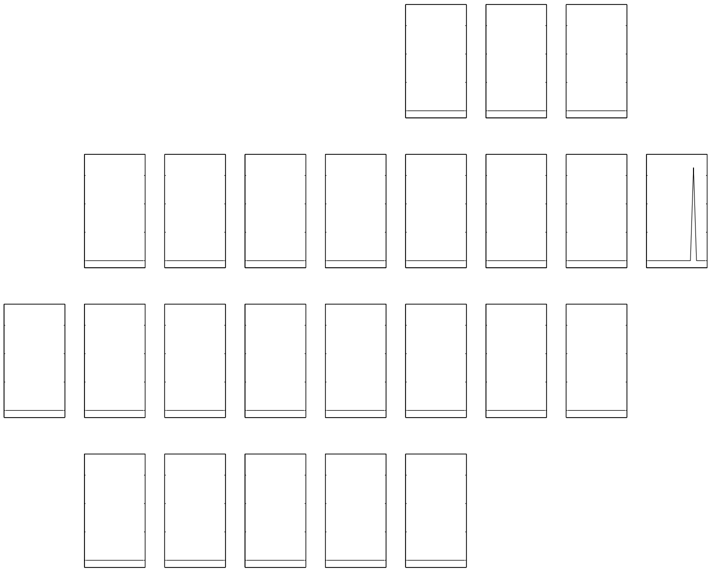


Figure 6-9: Solution scattering magnitudes in glint anisotropy example, $\alpha_0 = 30$, $\alpha_1 = 20$.

as coming from a single point and with very thin anisotropy explained by a single basis vector. With this representation, spatial properties about the object being imaged, such as orientation and physical extent, may be inferred; thin anisotropy corresponds to objects of large physical extent and wider anisotropy to objects with smaller physical extent. Also, the center angle of anisotropy indicates orientation in the spatial domain. Points have more meaning than just pixels with aspect-dependent amplitudes.

6.2 Object Shape of Migratory Scattering Centers

The machinery developed in the thesis thus far to handle migratory scattering has restricted the migration to be in a circle. Circular migration comes about from cylinders and tophats, which are often encountered in practice. However, if the machinery were more general purpose, then it would be possible to characterize objects of various other shapes as well. In this section, a methodology is suggested based on the observation that smooth convex shapes can be locally approximated by circles.

6.2.1 Circular Migration in Subapertures

Convex shapes in image space may be parameterized around a center (x_c, y_c) by a radius function $R(\theta)$. When a scattering center migrates along such a shape, the range profile is no longer a sinusoid, but the function $R(\theta)$ wrapped around a sinusoid, as discussed in Sec. 2.1.4 and given in equation (2.18), reproduced here:

$$\rho(\theta) = x_c \cos\theta + y_c \sin\theta - R(\theta).$$

When a convex shape is viewed over only a small angular sector, then it looks very much like a circle (also over that same small angular sector). Also, small non-sinusoidal segments may be approximated by small sinusoidal segments. These small angular sectors or segments are in fact subapertures in the context of SAR data.

The work in this thesis has stayed away from subaperture analysis due to the associated reduction in cross-range resolution. Here, however, a method is developed that uses data from subapertures to locally fit circles to more complicated shapes. This problem of determining migration shape is different from that of estimating the shape of convex sets in images from tomographic projections, tackled in [48, 27], because although migratory scattering centers may move along a convex-shaped path, scattering only occurs from a single point on the surface of the object per observation angle θ, in contrast to tomographic applications in medical imaging, for example, in which objects are filled with mass and line integrals through the entire object contribute to the measurements.

The methodology to determine the shape of objects producing migratory scattering is then to divide the phase history measurements into subapertures, optimize over radius as described in Sec. 4.2 to find the best circle for each subaperture, and then take all of the circles from the subapertures together to obtain a unified picture of object shape. This approach is fairly crude and there is much room for improvement in the formulation; nonetheless it may be used effectively, as shown in the next section through an example.

6.2.2 Example

The example in this section considers migratory scattering along a non-circular shape, illustrating that the method of finding best circular segments over subapertures is promising. The shape of migration is an ellipse with eccentricity 0.7 and semimajor axis of length 0.4 meters, centered at the point $(0.1, 0.1)$ and rotated by 20°. The true ellipse is shown in Fig. 6-10 with the solid line. The background in the figure is the subaperture image formed with an 18° subaperture with center angle 0°. The dotted line circle is the best explanation of migration using a circle, determined using the regression method of Sec. 4.2 over that subaperture.

Overall, data is available over a 108° wide-angle aperture; the scattering in this synthetic example is isotropic across the entire aperture. Nineteen overlapping 18° subapertures with center angles ranging from -45° to 45° at 5° increments are used. The conventional image formed from the full 108° aperture is shown in Fig. 6-11.

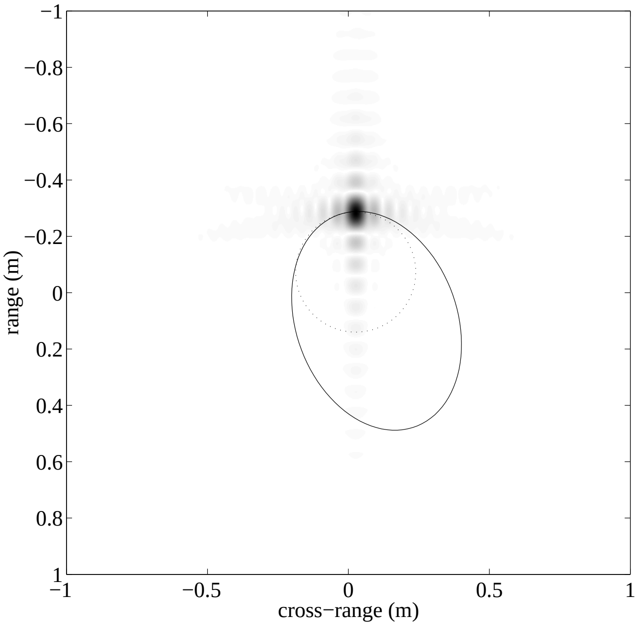


Figure 6-10: True elliptical migration as solid line overlaid on subaperture image along with best circular explanation over the subaperture as a dotted line.

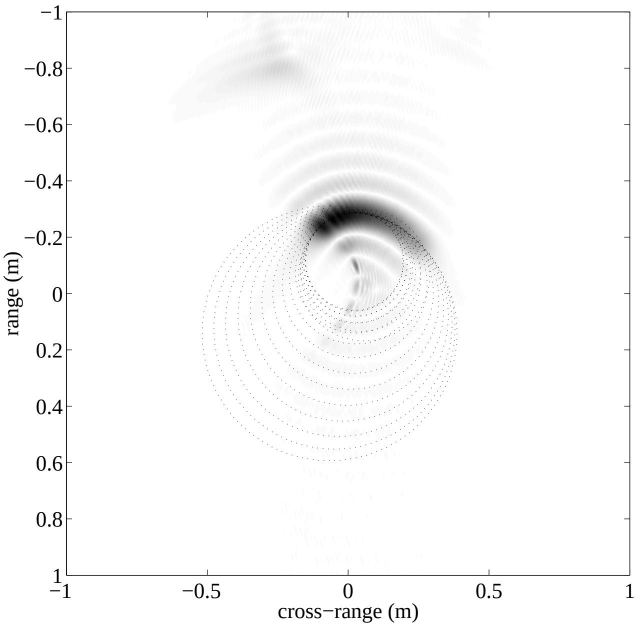


Figure 6-11: Solution subaperture circles overlaid on conventionally formed wide-angle image.

Overlaid on the image are the 19 circle solutions of optimizations over radius for each of the subapertures. The 19 optimizations are not coupled in any manner, but further extensions of the methodology may include some way of optimizing jointly.

Each of the circle solutions is only valid over its corresponding subaperture. Hence, it is only sensible to view circular segments, not entire circles. The circular segment solutions are shown in Fig. 6-12, overlaid on the same 108° aperture image of Fig. 6-11, but enlarged. Clearly, the segments taken together provide a good approximation to the true elliptical migration in the spatial domain. The segments also well-approximate the true migration in the range profile domain, as seen in Fig. 6-13. The top plot shows the ideal range profile with elliptical migration and the bottom plot shows the overlapping range profiles segments of the circle solutions.

This chapter, building upon stationary point scattering and circular migratory scattering discussed in earlier chapters, has moved in a direction away from a pixel representation towards more of an object-type representation. If the subaperture

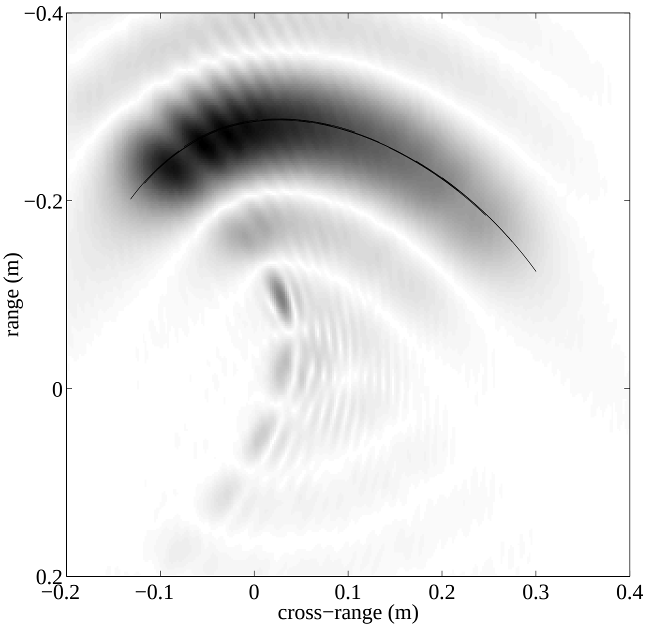


Figure 6-12: Solution circular segments overlaid on conventionally formed wide-angle image.

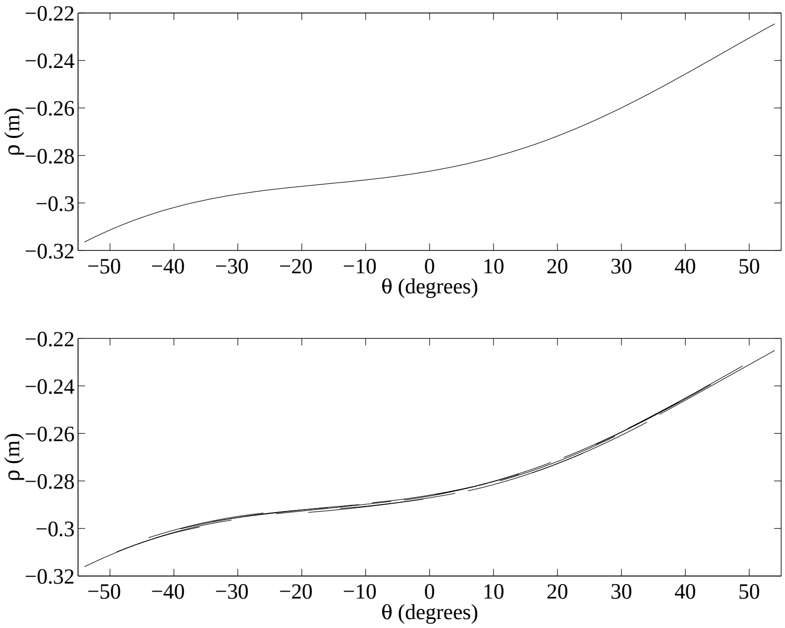


Figure 6-13: Range profile domain view of circular segments locally fitting elliptical migration.

segments were to be stitched together, then the result would be a unified object-level parsimonious representation akin to circular migratory scattering and glint scattering representations. This object-level conception may be taken even further through the use of basis vectors that are not necessarily localized in space; an elaboration is provided in Sec. 7.2.2.

Chapter 7

Conclusion

> I just don't think you have time. I've never been a great believer in reflection. I don't think it's a productive enterprise. I think that when you're done, then you look back and it'll be great to look back at a lot of things, a whole number of things. That'll actually be fun to do that, to look back and think about each team, the different situations.
>
> — *Jim Boeheim (2005), in The Post-Standard*

The first part of this concluding chapter wraps up the work in the thesis with a brief summary. However, the material cannot really be packaged nicely and neatly; many issues and directions remain sticking out like hooks. In the second part of the chapter, suggestions for future work are suggested.

7.1 Brief Summary

If one were to describe the aims of the research in this thesis generally, it would come down to taking problems and issues associated with wide-angle SAR and turning them into opportunities to squeeze out as much information and understanding from the data as possible. The issue in wide-angle SAR imaging considered primarily in the thesis is aspect-dependent scattering or anisotropy. Anisotropy may be viewed as a nuisance to conventional two-dimensional image formation techniques such as the polar format algorithm, but if anisotropy can be characterized within the image formation process, then the additional information can be useful as a feature for automatic target recognition and as an indicator about spatial domain properties of the objects being imaged. Similarly, the secondary wide-angle issue investigated in the thesis, migratory scattering, may also be viewed as either a problem or opportunity. Migration characterization gives information about object shape directly.

In SAR imaging, it is determination of the complex-valued scattering function of the ground patch s that is of interest. (Of course, the real goal is trying to learn as much as possible of what is out there, what it is doing, etc.) With anisotropy, the scattering function takes on an aspect angle dependence and is thus $s(x, y, \theta)$. By joint image formation and anisotropy characterization, it is meant that $s(x, y, \theta)$ is to be recovered from the phase history measurements. The point-scattering model of radar allows the problem to be one of determining scattering over a discrete set rather than determining a continuous field.

The approach taken in this thesis makes use of an overcomplete basis expansion and sparsifying regularization to obtain a solution for $s(x, y, \theta)$. This formulation of anisotropy characterization is new. The problem is then to solve for coefficients in the basis expansion. Overcomplete bases give rise to an infinite subspace of solution coefficient vectors; hence some way of selecting a favorite solution is required. Overcomplete bases are often designed, including in this thesis, such that basis vectors parsimoniously describe the class of signals that will be encountered, i.e. linear combinations of very few basis vectors are needed to well-represent signals. In this light, sparse solutions, i.e. coefficient vectors with few non-zero entries, are to be favored. Finding optimally sparse solutions is a combinatorial optimization problem, however.

The overcomplete basis is designed with a basis vector for every possible width of angular persistence of anisotropy and every possible shift of that contiguous anisotropy across the aperture. This choice is motivated by the physics of scattering, which indicate that objects have strong scattering responses over contiguous segments of aspect angle, to allow for parsimony in the representation of anisotropy. In the thesis, it is recognized that this overcomplete set has a structure to it that may be represented by a graph.

Reverting to the issue of finding sparse solutions, a relaxation employing the ℓ_p-

norm with $p < 1$ in a regularization cost function is used as an objective to minimize; coefficient vectors that minimize the cost function are sparse. A previously developed, non-combinatorial, quasi-Newton algorithm is applied to perform the minimization, yielding good results on examples with SAR data. A novel suite of algorithms are developed in the thesis with reduced complexity at the expense of greediness based on the graph structure of the overcomplete basis. The algorithms, best explained as heuristic search on the graph also achieve fine results on SAR examples, even on complicated scenes such as the backhoe public release dataset [4].

The overcomplete basis formulation solves for the anisotropy of all spatial locations within one system of equations, taking interactions among scattering centers into account. Also, the formulation is more flexible than parametric methods, but nevertheless does incorporate prior information about anisotropy through the choice of basis vectors. Data from the full wide-angle aperture is used throughout and the solutions have more detail in aspect angle than subaperture methods.

The ideas and terms point scattering center, spatial location, and pixel are fairly interchangeable unless more meaning is injected into the idea of a scattering center. With scatterers that appear to move or migrate as a function of the viewing direction, each scattering center corresponds to an area in space, not just a single spatial location or pixel. The overcomplete basis formulation is extended in a straightforward manner to incorporate migratory scattering centers with circular migration, through one additional dimension, radius, in the collection of basis vectors. These new basis vectors with migration radii provide more of an object-level description.

As an alternative to including basis vectors with different radii of migration in the overcomplete basis, an optimization methodology is developed in the thesis that both optimizes radius and characterizes anisotropy. It is a nonlinear least-squares regression technique over radii that performs the sparsifying cost function minimization as part of function evaluations. Through the use of subapertures and local approximation of non-circular migration by circular migration, the beginnings of an approach are laid to lend object-level descriptions to scattering centers with any general convex migration shape.

Object-level descriptions for stationary scattering centers are also come upon through the use of Hough transform-inspired regularization terms for favoring solutions that assign scattering contributions from objects distributed in space to single points. Object-level descriptions of stationary scattering centers can be further extended as actual basis vectors in the overcomplete basis as described in the next section as a direction for future research. The next section also discusses other avenues of research that have been illuminated by the inquiry into anisotropy characterization via sparse signal representation in this thesis.

7.2 Suggestions For Future Work

There are many different research directions that branch from the core work presented in this thesis, ranging from the development of automatic parameter selection methods to extending the formulation to three-dimensional imaging where the scattering function with angle-dependent scattering takes the form $s(x, y, z, \theta, \varphi)$. Frequency-dependent scattering may also be characterized. Effective techniques for the visualization of $s(x, y, \theta)$ solutions may be investigated, considering movies, lenticular images [35], and various ways to use color and shading.

The use and performance of characterized anisotropy as a feature in functioning automatic target recognition systems has not been examined. An extensive investigation into the effects of basis vector pulse shape not matching true anisotropy is worth pursuing as well. Also, there are many imaging application domains other than SAR, such as sonar, ultrasound, digital holography, and laser imaging that share the characteristics of coherency and complex-valued signals [11]; there are bound to be effects closely related to anisotropic or migratory scattering with these modalities, lending an opportunity to apply the formulation of this thesis. In the remainder of this section, two research directions — analysis and extension of the graph-structured algorithm, and further object-level processing — are commented upon.

7.2.1 Graph-Structured Algorithm Analysis and Extensions

The novel graph-structured algorithm developed in Sec. 4.3 is based on fairly simple heuristics and stopping criteria based on solutions with sparsifying regularization. Less computationally intensive methods, e.g. matching pursuit, could be considered to calculate the heuristics and stopping conditions as well. In the thesis, the algorithm is described and motivated intuitively and validated empirically, but no detailed analysis on its properties is performed. A next logical step for research is to analyze the algorithm first for the single spatial location case and then for the multiple spatial location case.

The minimum argument of the regularization cost function, $J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p$, has a statistical interpretation as the maximum a posteriori estimate for $\mathbf{a}$ when it is taken as a random vector a with prior probability density function $f_{\mathsf{a}}(\mathbf{a}) \propto e^{-\alpha\|\mathbf{a}\|_p^p}$ and the measurements $\mathbf{r}$ are corrupted by additive white circularly symmetric complex Gaussian noise [9]. Additionally, if each choice of left or right is independent and uniformly distributed for a chip falling down the plinko graph, then the distribution at the bottom level, level N, approaches a Gaussian distribution as N approaches infinity [24]. Putting these two facts together with perhaps some additional machinery, it may be possible to give a probabilistic guarantee on the convergence of the hill-climbing search method without back-tracking, at least for the single spatial location case. For the multiple spatial location case, showing something similar may be more difficult, but not out of the question.

The suite of graph-structured algorithms of this thesis can be related to other methods in science and engineering, e.g. multigrid methods [20]. One-way multigrid is similar to hill-climbing search without back-tracking on a plinko graph. The nested iteration algorithm or full multigrid, operating in a coarse-to-fine manner with corrections included, is in the same spirit as hill-climbing with back-tracking on a plinko graph. Connections such as this may lead to innovations in the graph-structured algorithm.

Considering the alternate interpretation of the search procedure, shown in Fig. 4-

19b, the idea of taking guiding graphs themselves as nodes is closely related to the idea of junction trees, although it is not exactly the same [16]. With a junction tree or something similar as a starting point along with the probabilistic interpretation discussed earlier in this section, more complicated procedures beyond the simple hill-climbing search, including multiscale or multiresolution methods, may be envisioned. Also, if a multiple candidate search procedure that allows contributions from disparate guiding graphs in the solution coefficient vector were to be developed, it would improve the generality and reduce the greediness of the graph-structured algorithms.

Sec. 4.3.3, variations on the theme, is by no means comprehensive; any number of slight modifications or tweaks are possible. However, principled approaches such as search on trees, multigrid, and inference algorithms on graphical models using junction trees, may have a more radical effect on the theory of and extensions to the graph-structured algorithm than minor modifications.

7.2.2 Object-Level Processing

The work presented in Chapter 6 approaches a general framework for objects with anisotropic scattering, an associated region in space, and possibly migration, but may be taken even further. The approach of Sec. 6.2 fits sinusoidal segments to general curves in the range profile domain. However, a more general approach may be to fit splines or other similar functions to the entire curve. In a general sense, this type of approach was followed in [48]. Another possibility is to use ideas from snakes and curve evolution [43]. The method of Sec. 4.2 is like blowing up circular balloons starting from points; the use of an outward curve evolution flow that could converge to non-circular shapes is a possible approach. It is interesting that the same migration is a sinuous curve in the range profile domain and a closed convex curve in the image domain, both concentrated into lines, not regions.

The stationary and circularly migratory basis vectors used in the thesis may be viewed as curves in a three-dimensional space of x, y, and θ, following visualizations in [5]. Stationary basis vectors are straight lines and migratory basis vectors are spirals, as shown in Fig. 7-1. Thinking in this space, glint anisotropy can also be

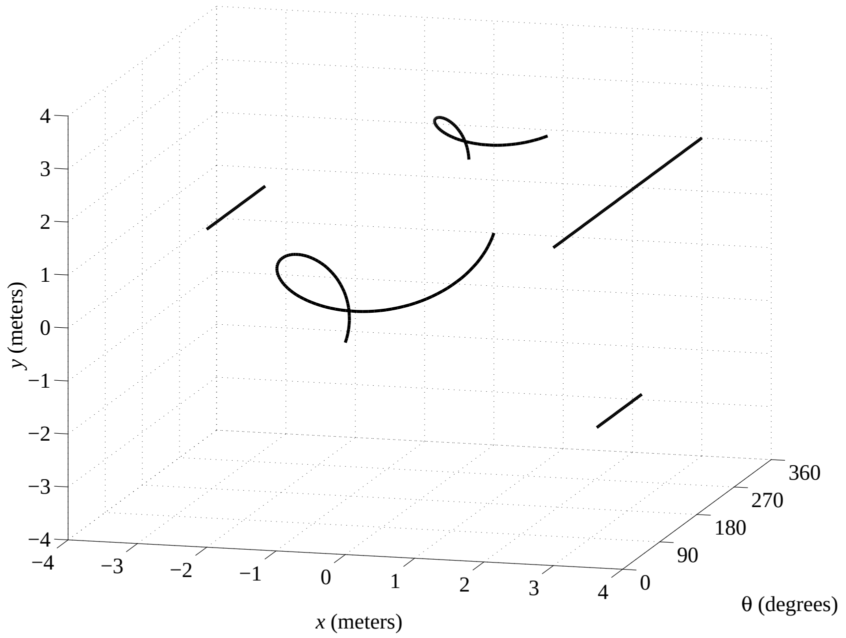


Figure 7-1: Stationary and migratory basis vectors in x-y-θ space.

represented by a line, but one that lies in a plane parallel to the x-y plane, as seen in Fig. 7-2. (Stationary basis vectors lie in planes perpendicular to the x-y plane.) Also, non-circular migration maps to generalized spirals. Continuing to think in this three-dimensional space, the overcomplete collection of basis vectors should ideally contain line segments and spiral-like curves of all different types and orientations. Basis vectors of this type would couple spatial extent and anisotropy in a physically meaningful way.

A very different approach to determine the size and shape of objects from radar measurements, the Bojarski-Lewis inverse scattering method from the electromagnetics literature, is ill-posed unless data is available at all frequencies and aspect angles [6, 45]. The method recovers a binary function $g(x, y, z)$ in 3-D that is one inside and zero outside of the object. This function g is the inverse Fourier transform of a function G which is related to the scattered signal from the object, $r(x', y', z')$, with

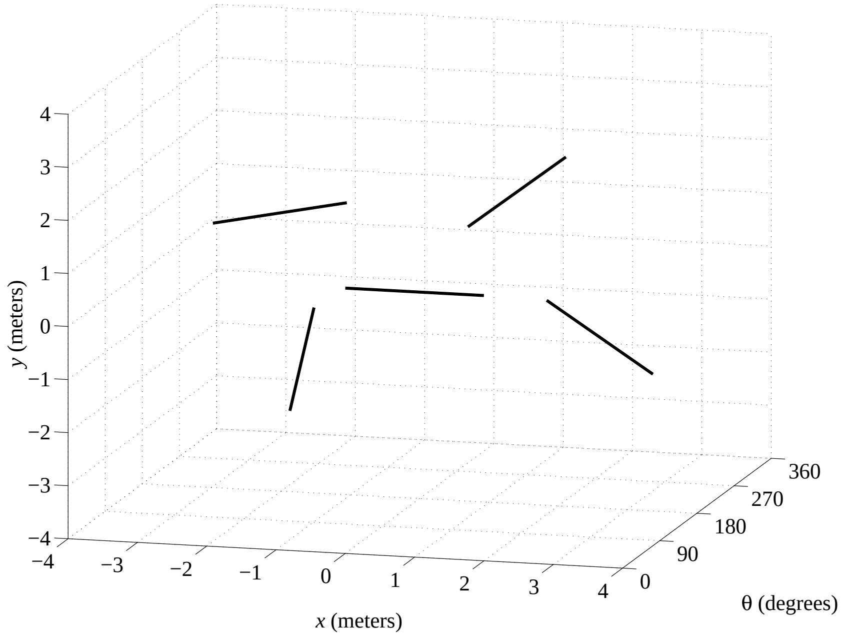


Figure 7-2: Glint anisotropy in x-y-θ space.

a radar pulse at frequency $2\pi f = \frac{c}{2}\sqrt{x'^2 + y'^2 + z'^2}$, by the following expression:

$$G(x', y', z') = \frac{4\sqrt{\pi}\Re\{r(x', y', z')\}}{x'^2 + y'^2 + z'^2}, \tag{7.1}$$

where x', y', and z' are spatial frequencies. The exact relationship between $r(x', y', z')$ and SAR phase history measurements needs to be investigated, but if done, the physics-based approach may provide insight for developing new object-level regularization terms which would both take care of the ill-posedness of the Bojarski-Lewis method and bring the method into a common frame as the overcomplete basis formulation for image formation and anisotropy characterization.

Moving even one layer of abstraction higher, object-type scattering centers are interrelated as parts of more complicated objects and the eventual goal in SAR imaging is detecting and recognizing these larger complicated objects, e.g. vehicles. Recent work in the computer vision community has looked at learning hierarchical models of scenes, objects, and parts, but this work relies on crude low-level interest operators for features such as image patches, corners, edges, etc. that are hoped to be invariant under many variables [52]. It is here that scattering centers may be employed as physically and semantically meaningful descriptors for use in higher level tasks such as object classification. This may eventually lead to a vertically integrated system starting at the lowest level of physics, moving through signal processing, statistical inference, and going all the way to understanding and awareness.

Bibliography

[1] Nitin Aggarwal and William Clement Karl. Line detection in images through regularized Hough transform. *IEEE Transactions on Image Processing*, 15(3):582–591, March 2006.

[2] Mark R. Allen and Lawrence E. Hoff. Wide-angle wideband SAR matched filter image formation for enhanced detection performance. In Dominick A. Giglio, editor, *SPIE Symposium, Algorithms for Synthetic Aperture Radar Imagery*, Orlando, Florida, April 1994.

[3] Amir Averbuch, Ronald R. Coifman, David L. Donoho, Michael Elad, and Moshe Israeli. Fast and accurate polar Fourier transform. *Applied and Computational Harmonic Analysis*, in press, corrected proof, available online, December 27, 2005.

[4] Backhoe data dome and Visual-D challenge problem. Available at Air Force Research Laboratory Sensor Data Management System (https://www.sdms.afrl.af.mil/main.php), 2004.

[5] Rajan Bhalla, Ann Marie Raynal, Hao Ling, John Moore, and Vincent J. Velten. Angular description for 3D scattering centers. In Edmund G. Zelnio and Frederick D. Garber, editors, *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XIII*, Orlando, Florida, April 2006.

[6] Norbert N. Bojarski. Three-dimensional electromagnetic short pulse inverse scattering. Technical report, Syracuse University Research Corporation, Syracuse, New York, February 1967.

[7] Ronald N. Bracewell. Numerical transforms. *Science,* New Series, 248(4956):697–704, May 11, 1990.

[8] Walter G. Carrara, Ron S. Goodman, and Ronald M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, Norwood, Massachusetts, 1995.

[9] Müjdat Çetin. *Feature-enhanced synthetic aperture radar imaging.* PhD thesis, Boston University, Boston, Massachusetts, 2001.

[10] Müjdat Çetin and William Clement Karl. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. *IEEE Transactions on Image Processing*, 10(4):623–631, April 2001.

[11] Müjdat Çetin, William Clement Karl, and Alan S. Willsky. Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging. *Optical Engineering*, 45(1):017003, January 2006.

[12] Müjdat Çetin and Randolph L. Moses. SAR imaging from partial-aperture data with frequency-band omissions. In Edmund G. Zelnio and Frederick D. Garber, editors, *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XII*, Orlando, Florida, March 2005.

[13] Ronald D. Chaney, Alan S. Willsky, and Leslie M. Novak. Coherent aspect-dependent SAR image formation. In Dominick A. Giglio, editor, *SPIE Algorithms for Synthetic Aperture Radar Imagery*, Orlando, Florida, April 1994.

[14] Da-Ming Chiang. *Parametric signal processing techniques for model mismatch and mixed parameter estimation.* PhD thesis, The Ohio State University, Columbus, 1996.

[15] Thomas F. Coleman and Yuying Li. A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4):1040–1058, 1996.

[16] Robert Cowell. Introduction to inference for Bayesian networks. In Michael I. Jordan, editor, *Learning in Graphical Models*, pages 9–26. The MIT Press, Cambridge, Massachusetts, 1999.

[17] Bradley S. Denney and Rui J. P. de Figueiredo. Scattering-based tomography for HRR and SAR prediction. *Multidimensional Systems and Signal Processing*, 14(1-3):207–222, January 2003.

[18] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, March 4 2003.

[19] David L. Donoho, Michael Elad, and Vladimir N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, January 2006.

[20] Craig C. Douglas. Multigrid methods in science and engineering. *IEEE Computational Science and Engineering Magazine*, 3(4):55–68, Winter 1996.

[21] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, January 1972.

[22] Laurent Ferro-Famil, Andreas Reigber, Eric Pottier, and Wolfgang-Martin Boerner. Scene characterization using subaperture polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10):2264–2276, October 2003.

[23] Layne R. Flake, Stanley Carlton Ahalt, and Ashok K. Krishnamurthy. Detecting anisotropic scattering with hidden Markov models. *IEE Proceedings - Radar, Sonar & Navigation*, 144(2):81–86, April 1997.

[24] Francis Galton. *Natural Inheritance*. Macmillan and Company, London, United Kingdom, 1889.

[25] Anna Genell and David Lösaus. On the Cramer Rao bound in determining scattering center parameters using high resolution radar. Master's thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2001.

[26] Michael J. Gerry, Lee C. Potter, Inder J. Gupta, and Andria van der Merwe. A parametric model for synthetic aperture radar measurements. *IEEE Transactions on Antennas and Propagation*, 47(7):1179–1188, July 1999.

[27] Alexander Goldenshluger and Vladimir Spokoiny. Recovering convex edges of an image from noisy tomographic data. *IEEE Transactions on Information Theory*, 52(4):1322–1334, April 2006.

[28] Jacques-Salomon Hadamard. Sur les probl'emes aux d'eriv'ees partielles et leur. *Princeton University Bulletin*, 13:49–52, 1902.

[29] Paul V.C. Hough. Method and means for recognizing complex patterns. U.S. Patent 3,069,654, December 1962.

[30] Charles V. J. Jakowatz, Jr., Daniel E. Wahl, Paul H. Eichel, Dennis C. Ghiglia, and Paul A. Thompson. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach.* Kluwer Academic Publishers, Norwell, Massachusetts, 1996.

[31] Joseph B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America*, 52(2):116–130, February 1962.

[32] Andrew J. Kim, John W. Fisher, III, and Alan S. Willsky. Detection and analysis of anisotropic scattering in SAR data. *Multidimensional Systems and Signal Processing*, 14(1-3):49–82, January 2003.

[33] Michael A. Koets and Randolph L. Moses. Feature extraction using attributed scattering center models on SAR imagery. In Edmund G. Zelnio, editor, *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery VI*, Orlando, Florida, April 1999.

[34] Balaji Krishnapuram, Jeffrey Sichina, and Lawrence Carin. Physics-based detection of targets in SAR imagery using support vector machines. *IEEE Sensors Journal*, 3(2):147–157, April 2003.

[35] Gabriel Lippmann. La photographie intégrale. *Comptes rendus de l'Académie des Sciences*, 146:446–451, 1908.

[36] Dmitry M. Malioutov. A sparse signal reconstruction perspective for source localization with sensor arrays. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2003.

[37] Dmitry M. Malioutov, Müjdat Çetin, and Alan S. Willsky. Homotopy continuation for sparse signal representation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 733–736, Philadelphia, Pennsylvania, March 2005.

[38] Guglielmo Marconi. Radio telegraphy. *Proceedings of the IRE*, 10(8):215–238, August 1922.

[39] James Clerk Maxwell. A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, 155:459–512, 1865.

[40] Eliakim H. Moore. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26:394–395, 1920.

[41] Randolph L. Moses, Lee C. Potter, and Müjdat Çetin. Wide angle SAR imaging. In Edmund G. Zelnio and Frederick D. Garber, editors, *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XI*, Orlando, Florida, April 2004.

[42] David C. Munson, Jr., James Dennis O'Brien, and W. Kenneth Jenkins. A tomographic formulation of spotlight-mode synthetic aperture radar. *Proceedings of the IEEE*, 71(8):917–925, August 1983.

[43] Nikos Paragios, Yunmei Chen, and Olivier Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer, 2005.

[44] Roger A. Penrose. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51:406–413, 1955.

[45] William L. Perry. On the Bojarski-Lewis inverse scattering method. *IEEE Transactions on Antennas and Propagation*, 22(6):826–829, November 1974.

[46] Plinko. In *The Price is Right*. CBS Broadcasting, Los Angeles, California, 1972-present.

[47] Lee C. Potter and Randolph L. Moses. Attributed scattering centers for SAR ATR. *IEEE Transactions on Image Processing*, 6(1):79–91, January 1997.

[48] Jerry L. Prince and Alan S. Willsky. Hierarchical reconstruction using geometry and sinogram restoration. *IEEE Transactions on Image Processing*, 2(3):401–416, July 1993.

[49] Johann Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Berichte Sächsische Akademie der Wissenschaften, Leipzig, Mathematisch-Physikalische Klasse*, 69:262–277, 1917.

[50] Paul Runkle, Lam H. Nguyen, James H. McClellan, and Lawrence Carin. Multi-aspect target detection for SAR imagery using hidden Markov models. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1):46–55, January 2001.

[51] Tapan K. Sarkar, Donald D. Weiner, and Vijay K. Jain. Some mathematical considerations in dealing with the inverse problem. *IEEE Transactions on Antennas and Propagation*, 29(2):373–379, March 1981.

[52] Erik B. Sudderth, Antonio Torralba, William T. Freeman, and Alan S. Willsky. Learning hierarchical models of scenes, objects, and parts. In *Proceedings of the Tenth IEEE International Conference on Computer Vision*, volume 2, pages 1331–1338, Beijing, China, October 2005.

[53] Andrei N. Tikhonov. O regulyarizatsii nekorrektno postavlennykh zadach. *Doklady Akademii nauk SSSR*, 151:501–504, 1963.

[54] Luiz C. Trintinalia, Rajan Bhalla, and Hao Ling. Scattering center parameterization of wide-angle backscattered data using adaptive Gaussian representation. *IEEE Transactions on Antennas and Propagation*, 45(11):1664–1668, November 1997.

[55] Michael van Ginkel, Cris L. Luengo Hendriks, and van Vliet Lucas J. A short introduction to the Radon and Hough transforms and how they relate to each other. Technical Report QI-2004-01, Delft University of Technology, Quantitative Imaging Group, Delft, Netherlands, 2004.

[56] Charles F. Van Loan. *Introduction to Scientific Computing.* Prentice Hall, Upper Saddle River, New Jersey, 2000.