IONOSPHERIC IMAGING WITH COMPRESSED SENSING

BY

BRIAN HARDING

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Associate Professor Jonathan J. Makela

ABSTRACT

Compressed sensing is a novel theory of sampling and reconstruction that has emerged in the past several years. It seeks to leverage the inherent sparsity of natural images to reduce the number of necessary measurements to a sub-Nyquist level.

We discuss how ideas from compressed sensing can benefit ionospheric imaging in two ways. Compressed sensing suggests signal reconstruction techniques that take advantage of sparsity, offering us new ways of interpreting data, especially for undersampled problems. One example is radar imaging. We explain how compressed sensing can be used for radar imaging and show results that suggest improved performance over existing techniques.

In addition to benefitting the way we use data, compressed sensing can improve how we gather data, allowing us to shift complexity from sensing to reconstruction. One example is airglow imaging, wherein we propose replacing CCD-based imagers with single-pixel, compressive imagers. This will reduce the cost of airglow imagers and allow access to spatial information at infrared wavelengths. We show preliminary simulation results suggesting this technique may be feasible for airglow imaging.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my adviser, Jonathan Makela, for his unspoken support and patience in allowing me to pursue my interests and for providing opportunities to cultivate these interests. I have truly enjoyed our work together and look forward to the coming years. I would also like to thank the staff at the Jicamarca Radio Observatory, where many of the ideas in this thesis were first generated. Thanks go specifically to Marco Milla and Koki Chau. Back at the University of Illinois, I must thank Nancy Morris for administrative support. To my fellow graduate students, thanks for all your help, academic and otherwise: Henry, Dan, Tim, Tom, Tony, Levent, and Kirk, to name a few. Of course, none of this would have been possible without the love and support from my family. Finally, thanks to Eileen for her love, encouragement, and understanding.

TABLE OF CONTENTS

CHAPTER 1

MOTIVATION

We have known since 1938 of the development of ionospheric irregularities and their effect on radio wave propagation [1]. The most spectacular of these irregularities are massive turbulent upwellings of plasma in the equatorial ionosphere that can reach heights of 1000 km or more, a phenomenon commonly known as "equatorial spread-F" [2, 3]. A more phenomenologically appropriate name is "convective ionospheric storm," but we will use the former to match previous literature [4]. In the intervening 75 years, we have gained an understanding of the seasonal variability in the occurrence of equatorial spread-F, but an explanation of its day-to-day variability remains elusive. Understanding the conditions that lead to its growth is a necessary first step to predicting the resulting communication outages, an overarching goal of the National Space Weather Program [5].

Equatorial spread-F comprises irregularities in electron density on spatial scales ranging from 100s of kilometers to fractions of a meter. Imaging the spatial and temporal dynamics on all scales is important in developing a holistic understanding of the physics of irregularity development and growth. The two most popular ionospheric imaging techniques are radar and optical airglow imaging. A backscatter radar imager is sensitive to small-scale irregularities on the order of half the radar wavelength because the dominant type of scattering is Bragg scatter. An airglow imager observes large, km-scale morphology of structures associated with equatorial spread-F, due to the fact that photon emission in the airglow layer is proportional to electron density, and the combination of nonzero pixel size, integration times, and lifetime of the emission results in a blurring of small-scale features. Both modalities are necessary for effectively monitoring irregularities; however, both have challenges. With radar imaging, we must overcome the underdetermined nature of the problem. With airglow imaging, typical CCD-based imagers are too expensive for global coverage and are only cost-effective at

visible wavelengths.

The emerging theory of compressed sensing can address both concerns. In this thesis, we show how compressed sensing can be used to improve imaging of equatorial spread-F, though the ideas presented here are easily generalized to other phenomena. In Chapter 2, we describe the problem of spatial electromagnetic imaging, of which radar and optical imaging are examples. We introduce the theory of compressed sensing and show how it can be used to solve such problems. In Chapter 3, we apply these ideas to radar imaging, and show results suggesting that compressed sensing can outperform existing methods on simulated and actual data. We use compressed sensing for airglow imaging in Chapter 4, which details the design of the single-pixel airglow imager and describes initial simulation results. We then conclude in Chapter 5.

CHAPTER 2

GENERAL SPATIAL ELECTROMAGNETIC IMAGING WITH COMPRESSED SENSING

Imaging is the process by a which a signal of space is estimated from measurements. This process comprises both the data gathering and the data processing. While these areas are often treated as the separate fields of instrumentation and image processing, a recent trend is to consider a more holistic approach. The idea is that deep knowledge of the sensor can inform better reconstruction algorithms, and conversely, a thorough understanding of the possibilities offered by signal processing can improve instrument design. In this vein, this chapter reviews both the hardware and software of imaging. The first section develops a general theory of electromagnetic imaging, oriented to the applications of optical and radio imaging. The second section introduces compressed sensing, an emerging signal processing theory that can be applied to solve electromagnetic imaging problems.

2.1 Spatial Electromagnetic Imaging

Students of electromagnetism learn as early as high school physics that light and radio are manifestations of the same underlying phenomenon, electromagnetic waves. However, in subsequent coursework the two fields diverge rapidly, developing independent language, notation, and conventions. This is likely due to the fact that the technologies in the two fields are substantially different; optical sensors can only measure amplitude, not phase, while radio sensors can measure both. This section is an attempt to bridge this gap in the context of imaging, providing a generalized framework for electromagnetic imaging that applies to both optics and radio.

2.1.1 Fundamentals

A traveling electromagnetic plane wave is characterized by an electric field of the following form:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E_0} \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \psi) \tag{2.1}$$

This describes a wave as a function of location in space, $\mathbf{r}$, and time, t, with an initial amplitude and direction given by $\mathbf{E_0}$, phase offset, ψ, oscillating with a radial frequency, ω, and traveling in the direction $\mathbf{k}$. To avoid having to deal with messy trigonometric identities, we leverage Euler's formula by performing all subsequent calculations with phasor notation. This phasor, $\tilde{\mathbf{E}}(\mathbf{r})$, is complex-valued, and the true time-domain signal can be recovered through:

$$\mathbf{E}(\mathbf{r}, t) = \Re\left\{\tilde{\mathbf{E}}(\mathbf{r})e^{j\omega t}\right\} \tag{2.2}$$

where $\Re$ denotes the real part. All additions and multiplications in the phasor domain are thus the identical operations in the time domain. However, we will have no need to revert to the time domain, so we will drop the tilde and simply write $\mathbf{E}(\mathbf{r})$. Our traveling wave phasor takes the form:

$$\mathbf{E}(\mathbf{r}) = Ae^{-j\mathbf{k}\cdot\mathbf{r}}\,\hat{\mathbf{E}}_\mathbf{0} \tag{2.3}$$

where $A = \|\mathbf{E_0}\|e^{j\psi}$ is a complex-valued amplitude and $\hat{\mathbf{E}}_\mathbf{0}$ is a unit vector in the direction of $\mathbf{E_0}$. For our purposes we need only consider transverse electromagnetic waves, for which $\hat{\mathbf{E}}_\mathbf{0} \perp \mathbf{k}$. Since we always know the relationship of $\hat{\mathbf{E}}_\mathbf{0}$ to $\mathbf{k}$, we can ignore the direction of $\mathbf{E}(r)$, simply writing:

$$E(\mathbf{r}) = Ae^{-j\mathbf{k}\cdot\mathbf{r}} \tag{2.4}$$

This simplified notation will allow us to have general results that apply to linearly, circular, and elliptically polarized waves without getting lost in notation.

2.1.2 Single Plane Wave Propagation

Consider the two-dimensional case of wave propagation in the xz-plane, as shown in Figure 2.1. In this work we will ignore the third dimension because

the generalization is trivial and the added complication would dilute the important concepts.
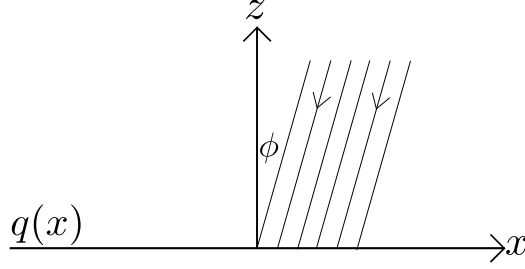

Figure 2.1: An electromagnetic wave impinging upon the $z = 0$ plane.

This wave is arriving at an angle ϕ with the z axis, so the wavevector is

$$\mathbf{k} = \begin{bmatrix} k_x \\ k_z \end{bmatrix} = \begin{bmatrix} -k\sin\phi \\ -k\cos\phi \end{bmatrix} \tag{2.5}$$

where $k = |\mathbf{k}| = \frac{2\pi}{\lambda}$ and λ is the wavelength. We can now write the phasor as:

$$E(x, z) = Ae^{jk(x\sin\phi + z\cos\phi)} \tag{2.6}$$

If we were to measure the electric field somewhere on the $z = 0$ plane, we would measure:

$$E(x, 0) = Ae^{jkx\sin\phi} \tag{2.7}$$

2.1.3 Plane Wave Continuum

We can think of $E(x, 0)$ as a spatial signal, a function of x. Imagine now that we have a multitude of plane waves arriving from a range of angles between $-\phi_0$ and ϕ_0. The plane wave from each angle has its own complex amplitude, a function we denote $A(\phi)$. The signal on the plane $z = 0$ is now the integration of all of these plane waves. This new signal is:

$$q(x) = \int_{-\phi_0}^{\phi_0} A(\phi)\, e^{jkx\sin\phi}\, d\phi \tag{2.8}$$

In many applications, we are interested in determining the angular distribution of incoming waves, $A(\phi)$, by making measurements of the spatial signal on a plane, $q(x)$. Equation (2.8) shows that these two functions have

a Fourier-transform-like relationship. This relationship is made more clear if one makes the small angle approximation:

$$\max|\sin\phi| = \sin\phi_0 << 1 \Rightarrow \sin\phi \approx \phi \qquad (2.9)$$

which yields an approximation for (2.8):

$$q(x) \approx \int_{-\phi_0}^{\phi_0} A(\phi)\, e^{jkx\phi}\, d\phi \qquad (2.10)$$

This is, precisely speaking, an inverse Fourier transform. If we take the inverse Fourier transform of A and evaluate it at kx, we get the function $q(x)$. Conversely, if we define a new function by rescaling q:

$$p(kx) = q(x) \qquad (2.11)$$

then we see that the functions p and A are Fourier transform pairs, which we denote with the following:

$$p \longleftrightarrow A \qquad (2.12)$$

2.1.4   Practical Issues

This is promising, since it gives us hope that by measuring p we can infer A. However, there are two serious obstacles that we must first overcome. First, to uniquely determine the function $A(\phi)$ we would need to know $p(x)$ for all x. This would require an uncountably infinite number of samples, but all we have in practice are finitely many measurements. Fortunately, we know that $A(\phi)$ is support-limited to within $\pm\phi_0$. In the language of signal processing, we can say that $p(kx)$ is bandlimited. The Shannon-Nyquist sampling theorem states that we can exactly reconstruct a bandlimited signal from its discrete samples, provided that the sampling rate is greater than twice the signal's bandwidth. Therefore, we can uniquely determine $A(\phi)$ as long as the sensors measuring $p(kx)$ are spaced by a distance less than the

Nyquist requirement:

$$f_{\text{sampling}} > 2B = 2\left(\frac{\phi_0}{2\pi}\right)$$

$$\frac{1}{\Delta(kx)} > \frac{\phi_0}{\pi} \qquad (2.13)$$

$$\Delta\left(\frac{x}{\lambda}\right) < \frac{1}{2\phi_0}$$

This Nyquist requirement relates the maximum sampling interval (measured in wavelengths) to the angular width of the signal. For example, consider a 50-MHz imaging radar with a transmitter beam width of 10°. The required antenna spacing is then about 17 meters. In the optical case ($\lambda = 400$ nm), the required spacing is 1 μm.

The second obstacle is that the Shannon-Nyquist theorem assumes we have a countably infinite number of samples spanning the entire x-axis. In practice, we only have finitely many sensors and thus can only sample out to a maximum distance x_0. One attractive solution to this problem is to continue using the Fourier-transform-based processing scheme, substituting zeros for those samples we are unable to obtain. Effectively, we are windowing $p(kx)$ with a rectangle function:

$$r(kx) = \begin{cases} 1 & \text{if } |kx| \leq kx_0 \\ 0 & \text{if } |kx| > kx_0 \end{cases} \qquad (2.14)$$

If we then attempt to recover $A(\phi)$ with a Fourier transform, we get a blurred version, as shown below:

$$p \cdot r \longleftrightarrow A * \mathscr{F}\{r\} = A * R \qquad (2.15)$$

where $\mathscr{F}$ denotes a Fourier transform, and we have used the Fourier transform property that multiplication in one domain is convolution in the other. The Fourier transform of a rectangle function is a sinc function, so after the correct scaling we can find the blurring function R:

$$R(\phi) = 2kx_0 \,\text{sinc}(kx_0\phi) \qquad (2.16)$$

7

where

$$\text{sinc}(x) \equiv \frac{\sin x}{x} \tag{2.17}$$

Roughly speaking, the width of the main lobe of the sinc function is $2\sqrt{3}$. This is found from the approximate value of x for which $\text{sinc}(x) = \frac{1}{2}$. Therefore, the width of the main lobe of $R(\phi)$ is:

$$\Delta\phi \approx \frac{2\sqrt{3}}{kx_0} = \frac{2\sqrt{3}}{2\pi\left(\frac{x_0}{\lambda}\right)} \approx \frac{1}{2\left(\frac{x_0}{\lambda}\right)} \tag{2.18}$$

Features of $A(\phi)$ on scales less than $\Delta\phi$ will be smeared out by the convolution with $R(\phi)$. The farther out we sample (higher x_0), the narrower $R(\phi)$ will be, and the more detail we can resolve in $A(\phi)$.

The preceding discussion inspires some questions. Since using the Fourier transform introduces a blur, is there some other inversion scheme with better resolution properties? Instead of assuming the non-measured samples to be zero, is there some clever way of extrapolating $p(kx)$ out past x_0, either implicitly or explicitly? What if we cannot, or prefer not to, have regular sampling? More seriously, what if we cannot sample at the Nyquist rate? All of these software-related questions find an answer in more advanced signal processing schemes such as compressed sensing, to be discussed in the second section of this chapter. For now, we will answer a hardware-related question. The procedure above relies on sampling $p(kx)$, a complex quantity. This is only possible with sensors that measure phase, but phase-sensitive sensors in the optical regime have not yet been realized. The mathematical problem of extracting the phase from an amplitude signal is known as phase retrieval, and despite decades of research, little headway has been made. Is there a way to perform the necessary Fourier transform without sampling and processing? The answer is yes, and the solution is of course to use a lens.

2.1.5 Analog Fourier Transform

Consider that instead of distributing sensors on the $z = 0$ plane, we place a lens, as shown in Figure 2.2. An ideal lens is simply a "phase screen," adding a position-dependent phase shift to the signal without affecting its amplitude. The ideal lens has a quadratic phase shift; the signal passing

through it is modulated by the following complex function [6]:

$$t(x) = e^{j\frac{kx^2}{2f}} \qquad (2.19)$$

where f is the focal length. Thus, the spatial signal just below the $z = 0$ plane is the product of (2.10) and (2.19):
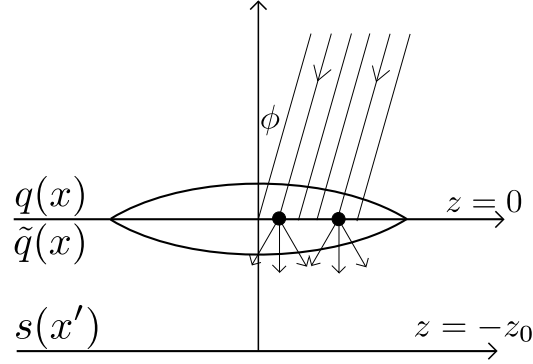
$$\tilde{q}(x) = q(x)t(x) \qquad (2.20)$$

Figure 2.2: An electromagnetic wave impinging upon a lens at the $z = 0$ plane, and propagating to the $z = -z_0$ plane.

We now propagate $\tilde{q}(x)$ to a new plane at $z = -z_0$. In keeping with the Huygens-Fresnel principle [6], we calculate this propagation by treating every point just below the $z = 0$ plane as the source of a new spherical wave. Consider just one ray emanating from just one point, x, on this plane to a point, x', on the $z = -z_0$ plane, as shown in Figure 2.3. This ray has initial amplitude $\tilde{q}(x)dx$ and propagates a distance d. After propagation, it encounters a phase decrease of kd. Its amplitude is also reduced by $4\pi d^2$. If we call the signal on the $z = -z_0$ plane $s(x')$, we can write:

$$s(x')dx = \frac{1}{4\pi d^2} \tilde{q}(x)e^{-jkd} dx \qquad (2.21)$$

The distance d is:

$$d = \sqrt{(x - x')^2 + z_0^2} \qquad (2.22)$$

which substituted into (2.21) gives:

$$s(x')dx = \frac{1}{4\pi \left((x - x')^2 + z_0^2\right)} \tilde{q}(x)e^{-jk\sqrt{(x-x')^2+z_0^2}} dx \qquad (2.23)$$
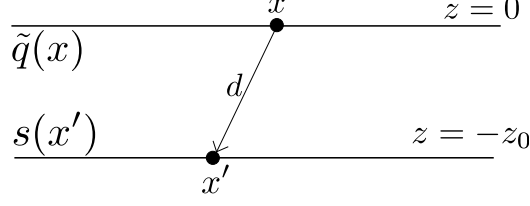
9

Figure 2.3: Using the Huygens-Fresnel principle to model electromagnetic propagation from one plane to another.

To simplify this expression, we use the Fraunhofer approximation [6], which basically assumes:

$$z_0^2 >> (x - x')^2 \qquad (2.24)$$

This assumption is valid when the planes are separated by a distance much greater than the size of the lens. Except to make this assumption, we will not explicitly assume a finite-sized lens, but will simply note that a finite lens produces a small blurring effect in an identical way to the finite sampling effect described above.

Using the Fraunhofer approximation, we can approximate the amplitude term and phase term in different ways, since the function $1/d$ is much less sensitive to small changes in d than e^{jkd}. The amplitude term can simply be approximated as:

$$\frac{1}{4\pi\left((x - x')^2 + z_0^2\right)} \approx \frac{1}{4\pi z_0^2} \qquad (2.25)$$

which is a constant amplitude term that is ignored in practice. For the phase term, we utilize the approximation $(1 + \delta)^n \approx 1 + n\delta$ for small δ.

$$
\begin{aligned}
e^{-jk\sqrt{(x-x')^2+z_0^2}} &= e^{-jkz_0\sqrt{1+\left(\frac{x-x'}{z_0}\right)^2}} \\
&\approx e^{-jkz_0\left(1+\frac{1}{2}\left(\frac{x-x'}{z_0}\right)^2\right)} \qquad (2.26) \\
&= e^{-jkz_0}\, e^{-jk\frac{x'^2}{2z_0}}\, e^{jkx\frac{x'}{z_0}}\, e^{-jk\frac{x^2}{2z_0}}
\end{aligned}
$$

Integrating both sides of (2.21) and substituting the approximations (2.25) and (2.26), we find an expression for the signal on the $z = -z_0$ plane:

$$s(x') = \frac{e^{-jkz_0}}{4\pi z_0^2}\, e^{-jk\frac{x'^2}{2z_0}} \int_{-\infty}^{\infty} \tilde{q}(x) e^{jkx\frac{x'}{z_0}}\, e^{-jk\frac{x^2}{2z_0}}\, dx \qquad (2.27)$$

We recognize that the integral is almost a Fourier transform, excepting the

quadratic phase term $e^{-jk\frac{x^2}{2z_0}}$. To remove this quadratic phase we use the definition of $\tilde{q}(x)$, Equation (2.20), to write

$$s(x') = \frac{e^{-jkz_0}}{4\pi z_0^2} e^{-jk\frac{x'^2}{2z_0}} \int_{-\infty}^{\infty} q(x)e^{jkx\frac{x'}{z_0}} t(x)e^{-jk\frac{x^2}{2z_0}} dx \qquad (2.28)$$

Not coincidentally, the ideal lens phase $t(x)$, given in Equation (2.19), exactly cancels this quadratic phase term if we choose $f = z_0$ to be the focal length of the lens, yielding:

$$s(x') = \frac{e^{-jkz_0}}{4\pi z_0^2} e^{-jk\frac{x'^2}{2z_0}} \int_{-\infty}^{\infty} q(x)e^{jkx\frac{x'}{z_0}} dx \qquad (2.29)$$

This Fourier transform cancels the Fourier transform used to create $q(x)$, giving us back the desired function $A(\phi)$. This last statement is made concrete by considering a new function u that is a rescaled and phase-adjusted version of s, defined such that:

$$u(-z_0 x') \frac{e^{-jkz_0}}{4\pi z_0^2} e^{-jk\frac{x'^2}{2z_0}} = s(x') \qquad (2.30)$$

With this substitution and the previously-defined substitution $p(kx) = q(x)$, Equation (2.29) can be succinctly written as a simple Fourier transform relationship between the signal p just above the plane $z = 0$ and the signal u on the plane $z = -z_0$:

$$p \longleftrightarrow u \qquad (2.31)$$

Coupled with the previous Fourier transform result, Equation (2.12),

$$p \longleftrightarrow A$$

we conclude that $u = A$. In other words, the signal on the image plane (ignoring constants and the small phase modulation $e^{-jk\frac{x'^2}{2z_0}}$) is a scaled version of the desired sky image $A(\phi)$. Therefore, to sample $A(\phi)$ we simply sample $u(-\frac{x}{z_0})$ where x is measured on the $z = -z_0$ plane. A popular way to perform this sampling is with a charge-coupled device (CCD), an array of sensors. While without the lens we sampled and performed the Fourier transform in software, here we let nature perform the Fourier transform, with the help of the lens to cancel the quadratic phase.

11

2.1.6 Generalized Sampling

In the previous sections we detailed the relationship between different electromagnetic imaging modalities and recognize that it is simply a matter of which domain is being sampled. A typical radio imager samples $q(x)$ directly, so it samples $A(\phi)$ in the Fourier domain. A typical optical imager uses a lens and samples $s(x')$ on the image plane, essentially sampling $A(\phi)$ directly. This is known as the canonical, or Dirac, domain. To generalize our notion of sampling, we model each sample as an inner product between our image $A(\phi)$ and a "test function" $h(\phi)$:

$$g_m = \langle A, h_m \rangle \equiv \int_{-\infty}^{\infty} A(\phi) h_m^*(\phi) \, d\phi \qquad (2.32)$$

where g_m is a sample, and the asterisk denotes complex conjugation. If we take the set of test functions to be complex exponentials, then we recognize that each measurement is a sample of the Fourier transform, which is the radio imaging case. If we take the test functions to be Dirac delta functions, we have the optical imaging case. Actually, since a photosensor cannot have infinitesimal area, but rather integrates the signal over a small range of values of x, the appropriate $\{h_m\}$ for CCD-based optical imaging consists of rectangle functions. However, more exotic test functions are certainly possible and can be beneficial, an example of which is the single-pixel camera discussed in Chapter 4. Equation (2.32) generalizes all forms of linear sampling.

Digital computing requires discrete problems, so much of the remainder of this thesis will be formulated in the discrete case. In order to discretize (2.32), we consider its associated Riemann sum approximation.

$$g_m = \sum_{i=-\infty}^{\infty} A(\phi_i) h_m^*(\phi_i) \Delta\phi \qquad (2.33)$$

Here, $\{\phi_i\}$ is a set of equally spaced independent variables separated by $\Delta\phi$. Since A has limited support, we only need to consider a finite number of values of i, a number we call N.

$$g_m = \sum_{i=0}^{N-1} A(\phi_i) h_m^*(\phi_i) \Delta\phi \qquad (2.34)$$

Suppose we have M such measurements, each with a different test function. By stacking these M measurements into a vector, we can rewrite (2.34) in matrix-vector notation:

$$\mathbf{g} = \mathbf{H}^*\mathbf{f} \tag{2.35}$$

where the columns of the matrix $\mathbf{H}$ are the discretized test functions, and we have relabeled the desired image vector $\mathbf{f}$ as the discretized version of $A(\phi)$. The asterisk denotes the Hermitian operator (conjugate transpose). The number of elements in $\mathbf{g}$ is M, the number of measurements, and the number of elements in $\mathbf{f}$ is N, the desired resolution. The matrix $\mathbf{H}$ is known as the sampling matrix. Equation (2.35) will serve as our primary generalized forward model for the remainder of this work.

2.1.7   Extension to Variance Imaging

Until now, for notational simplicity we have assumed that $A(\phi)$, the complex amplitude distribution, is the function we wish to estimate. However, in many applications, we do not care about the phase of $A(\phi)$, only the magnitude. For example, in optical imaging of a noncoherent source, the phase carries no useful information. To use a radar example, we often wish to know if a target is present in some direction and at some approximate range, but do not care whether this range is 5000 meters versus 5001. In these applications, the phase is a random variable while the magnitude is deterministic. In some applications, even the magnitude is a random variable. To capture all of these cases, we say that instead of estimating $A(\phi)$, we wish to estimate

$$f(\phi) = \mathbb{E}\big[|A(\phi)|^2\big] \tag{2.36}$$

where $\mathbb{E}[\cdot]$ is the expected value operator. This is the variance of the random variable $A(\phi)$ (assuming it has zero mean). Our previous measurements $\{g_m\}$ are related to $A(\phi)$, not its variance, so to get information on the second moment of A, we must find the second moment in our measurement space.

To do so, we calculate the correlation between two measurements:

$$
\begin{aligned}
\mathbb{E}[g_m g_k^*] &= \mathbb{E}\left[\left(\int_{-\infty}^{\infty} A(\phi) h_m^*(\phi)\, d\phi\right)\left(\int_{-\infty}^{\infty} A(\phi') h_k^*(\phi')\, d\phi'\right)^*\right] \\
&= \iint_{-\infty}^{\infty} \mathbb{E}[A(\phi) A^*(\phi')]\, h_m^*(\phi) h_k(\phi')\, d\phi\, d\phi'
\end{aligned}
\tag{2.37}
$$

The function $\mathbb{E}[A(\phi) A^*(\phi')]$ is the autocorrelation of $A(\phi)$. Assume that signals coming from different directions are uncorrelated. Then the autocorrelation function is:

$$
\mathbb{E}[A(\phi) A^*(\phi')] = f(\phi)\delta(\phi - \phi')
\tag{2.38}
$$

This assumption is justified, for example, in the case when $A(\phi)$ has a deterministic magnitude and a uniform random phase, ψ, that is spatially uncorrelated:

$$
\begin{aligned}
\mathbb{E}[A(\phi) A^*(\phi')] &= |A(\phi)|\, |A(\phi')|\, \mathbb{E}\left[e^{j(\psi_1(\phi) - \psi_2(\phi'))}\right] \\
&= |A(\phi)|\, |A(\phi')|\, \delta(\phi - \phi') \\
&= f(\phi)\delta(\phi - \phi')
\end{aligned}
\tag{2.39}
$$

This assumption is also justified if the magnitude is random, but is not shown here. In any case, substituting (2.38) into (2.37) yields a simplified expression:

$$
\begin{aligned}
\mathbb{E}[g_m g_k^*] &= \iint_{-\infty}^{\infty} f(\phi)\delta(\phi - \phi') h_m^*(\phi) h_k(\phi')\, d\phi\, d\phi' \\
&= \int_{-\infty}^{\infty} f(\phi) h_m^*(\phi) h_k(\phi)\, d\phi
\end{aligned}
\tag{2.40}
$$

Equation (2.40) is of the same form as (2.32): our "measurements" (actually cross-correlations of measurements) are generalized linear samples of the function we wish to estimate (the variance, $f(\phi)$), and the "test functions" are actually products of test functions. This means that Equation (2.35) is still valid as a forward model for variance imaging.

For the example of radio imaging, this result is particularly simple. In this case, as discussed above, the individual test functions are complex exponen-

tials, so (2.40) becomes:

$$\begin{aligned}
\mathbb{E}[g_m g_k^*] &= \int_{-\infty}^{\infty} f(\phi) e^{-jkx_m\phi} e^{jkx_k\phi} d\phi \\
&= \int_{-\infty}^{\infty} f(\phi) e^{-jk(x_k - x_m)\phi} d\phi
\end{aligned}$$

(2.41)

where x_m and x_k are the locations of the two sensors measuring g_m and g_k. Here we see that the new test function is also a complex exponential, maintaining the Fourier transform relationship between what we have (cross-correlations between sensors) and what we want (the distribution of signal energy across the sky). This is a well-known result in radio interferometry, radio astronomy [7], and more recently in radar imaging [8]. Cross-correlating the signal from two spatially separated antennas (the "visibility" function) yields a sample of the Fourier transform of the angular energy distribution (the "brightness" function).

It is worth pausing briefly to review the important distinction between correlation and sample correlation, since we have referred to the correlation as a measurement. Colloquially, the term "correlation" is often used to refer to the operation of taking two measurements and averaging their product over time or space. However, in the context used above, and in the strict sense, correlation is an *ensemble* average, a statistical concept. It is a fundamental property of a random signal; it cannot be measured. A temporal or spatial average can be used to estimate, not measure, the correlation, and only when the signal is ergodic. Any such temporal or spatial average is necessarily finite, leading to uncertainty in the estimate. As such, in a practical experiment, the left-hand side of Equation (2.40), $\mathbb{E}[g_m g_k^*]$, which is the "measurement," contains uncertainty. One must estimate this uncertainty in order to effectively use this data. This uncertainty is based on the properties of both the random signal, $A(\phi)$, and random sensor noise. It is tempting to lump statistical uncertainty and sensor noise together as "noise," but an understanding of the distinction is essential for proper use of one's data. A formal treatment and useful reference for these considerations for the case of an incoherent radar experiment is given by [9].

Finally, we note that this discussion about variance imaging relates primarily to the radio case, since optical sensors cannot measure phase and thus directly measure power, which is related to signal variance. However, as dis-

cussed above, Equation (2.35) encapsulates these complexities and continues to provide a valid discrete forward model for our problem.

With the forward model for spatial electromagnetic imaging developed in this section, we now turn to the latter half of our sampling and reconstruction problem.

2.2 Compressed Sensing

In the case of electromagnetic imaging, and in many other applications, our measurements are governed by the linear forward model (Equation (2.35)):

$$\mathbf{g} = \mathbf{H}^*\mathbf{f}$$

where $\mathbf{g}$ is a vector consisting of our measurements, $\mathbf{f}$ is the image vector we want, indexed by space, and $\mathbf{H}^*$ is the $M \times N$ linear operator that relates the signal of interest to our data. In the special case where $M \geq N$ and $\mathbf{H}^*$ is full rank (analogous to sampling at or above the Nyquist rate), we can reconstruct $\mathbf{f}$ exactly with a simple matrix inversion or least-squares solution. A more interesting and practical case is where $M < N$, a situation in which we refer to the problem as underdetermined. In this case, there are an infinite number of possible estimates, $\hat{\mathbf{f}}$, that agree with the data, $\mathbf{g}$, because $\mathbf{H}^*$ has a nonempty null space, $\mathcal{N}(\mathbf{H}^*)$. This is because, for any plausible solution $\hat{\mathbf{f}}$, we have:

$$
\begin{aligned}
\mathbf{g} &= \mathbf{H}^*\hat{\mathbf{f}} \\
&= \mathbf{H}^*\hat{\mathbf{f}} + \mathbf{0} \\
&= \mathbf{H}^*\hat{\mathbf{f}} + \mathbf{H}^*\mathbf{n} \quad \forall \mathbf{n} \in \mathcal{N}(\mathbf{H}^*) \\
&= \mathbf{H}^*(\hat{\mathbf{f}} + \mathbf{n}) \quad \forall \mathbf{n} \in \mathcal{N}(\mathbf{H}^*)
\end{aligned}
\tag{2.42}
$$

This means that, for any plausible solution $\hat{\mathbf{f}}$, an equally plausible solution is $\hat{\mathbf{f}} + \mathbf{n}$. How we are to choose among this infinitude of candidate solutions is the subject of regularization.

The basis of regularization is the contention that not all candidate solutions are equally plausible when we incorporate prior information. Typical regularization is to choose the solution that minimizes some cost function $c(\mathbf{f})$ while prescribing to the data. Formally, the procedure is to find the $\mathbf{f}$

in order to:

$$\begin{aligned} \text{minimize}: \quad & c(\mathbf{f}) \\ \text{subject to}: \quad & \mathbf{g} = \mathbf{H}^*\mathbf{f} \end{aligned} \qquad (2.43)$$

There are many possible choices for the cost function $c(\mathbf{f})$. For example, choosing

$$c(\mathbf{f}) = \|\mathbf{f}\|_2 \qquad (2.44)$$

yields Tikhonov, or 2-norm, regularization, which finds the candidate signal with the minimum energy. Choosing $c(\mathbf{f})$ to be the Shannon entropy:

$$c(\mathbf{f}) = -\sum_{i=0}^{N-1} f_i \ln f_i \qquad (2.45)$$

results in a form of regularization known as the Maximum Entropy Method [10].

Recently, a compelling new form of regularization has emerged from theoretical results in sampling and reconstruction theory known as "compressed sensing" or "compressive sampling" [11–15]. Briefly, the compressed sensing scheme is to regularize by the transform sparsity:

$$c(\mathbf{f}) = \|\mathbf{\Psi}^*\mathbf{f}\|_0 \qquad (2.46)$$

where $\|\cdot\|_0$ is the "0-norm," which counts the number of nonzero elements and $\mathbf{\Psi}$ is an $N \times N$ orthogonal matrix, whose columns define a basis in which $\mathbf{f}$ is believed to be sparse (or at least compressible, as we will see). The operator $\mathbf{\Psi}^*$ thus implements a transformation of $\mathbf{f}$ into a basis in which it is well-represented by a small number of coefficients.

In the following sections we introduce the basic theory of compressed sensing and its extensions, and discuss the reasons compressed sensing appears to be an improvement to current regularization schemes.

2.2.1 Basic Theory

We argue that compressed sensing offers a more compelling form of regularization in part due to surprising theoretical recovery results. Historically, the Shannon-Nyquist sampling theorem has guided engineers' thinking in instrument design and signal analysis. This theorem states that a signal is exactly

recoverable from its discrete samples provided its bandwidth is less than half the sample rate. Recent results indicate that exact signal recovery is possible under much looser constraints (e.g., [11, 12]). Instead of requiring the signal to be bandlimited, we require the signal to be sparse, meaning it consists mostly of zeros, with a small number of nonzero elements. Furthermore, we generalize our notion of sampling to include all linear samples instead of point samples (as discussed in Section 2.1.6). If the signal $\mathbf{f}$ is indeed sparse,

$$K = \|\mathbf{f}\|_0 \tag{2.47}$$

where $K << N$, then exact recovery of $\mathbf{f}$ by the measurements $\mathbf{g} = \mathbf{H}^*\mathbf{f}$ is possible by solving the following problem [11]:

$$\begin{aligned} \text{minimize}: \quad & \|\mathbf{f}\|_0 \\ \text{subject to}: \quad & \mathbf{g} = \mathbf{H}^*\mathbf{f} \end{aligned} \tag{2.48}$$

The solution to the above problem is equal to the original signal provided $\mathbf{H}^*$ satisfies the restricted isometry property (RIP) [11], which roughly requires that all sets of $2K$ columns of $\mathbf{H}^*$ are approximately orthogonal. This ensures that most of the signal energy is captured by the sensing matrix.

Of course, typical signals are not directly sparse. However, most natural signals are sparse when transformed into an appropriate basis. This assumption of transform sparsity is the basis of widely-used image compression algorithms such as JPEG, for example. To formalize this into our sensing framework, we may think of our desired signal not as $\mathbf{f}$ but rather as its transformation:

$$\mathbf{s} = \mathbf{\Psi}^*\mathbf{f} \tag{2.49}$$

where the columns of the $N \times N$ matrix $\mathbf{\Psi}$ define a basis in which $\mathbf{f}$ is sparsely representable. With this substitution, Problem (2.48) becomes:

$$\begin{aligned} \text{minimize}: \quad & \|\mathbf{s}\|_0 \\ \text{subject to}: \quad & \mathbf{g} = \mathbf{H}^*\mathbf{\Psi}\mathbf{s} \end{aligned} \tag{2.50}$$

Once the sparse transformed signal $\mathbf{s}$ is recovered by Problem (2.50), then the desired signal $\mathbf{f}$ can be recovered through $\mathbf{f} = \mathbf{\Psi}\mathbf{s}$. Note that Problem (2.48) is a special case of Problem (2.50) by considering $\mathbf{\Psi} = \mathbf{I}$, the identity

matrix, or Dirac basis.

Solving Problem (2.50) only succeeds if the matrix $\mathbf{H}^*\mathbf{\Psi}$ satisfies the RIP. The language of the RIP becomes slightly awkward and hard to interpret in this case. Since the validity of the RIP only depends on the relationship between the sensing matrix $\mathbf{H}$ and the sparsity matrix $\mathbf{\Psi}$, we will speak instead of the "incoherence" between these two bases, a property that determines the maximum K for which the RIP is satisfied [12]. Define the coherence of two bases as:

$$\mu \equiv \sqrt{N} \max_{i,j} |\langle h_i, \psi_j \rangle| \tag{2.51}$$

where $\langle h_i, \psi_j \rangle$ is the inner product between a column of $\mathbf{H}$ and a column of $\mathbf{\Psi}$. In this case, the RIP is satisfied and exact recovery is therefore possible if [11]:

$$K \leq C \frac{1}{\mu^2} \frac{M}{\log^4 N} \tag{2.52}$$

where C is a constant. The less coherent the measurement and sparsity bases are, the fewer measurements are needed for exact recovery.

The idea of incoherence is often described in terms of an uncertainty principle. It effectively measures how spread out basis vectors of $\mathbf{H}$ are when expressed in the $\mathbf{\Psi}$-basis, and vice versa. For example, the Fourier and Dirac bases are maximally incoherent, since a delta function in one domain has full support in the other. This is important because it guarantees that information from a sparse signal $\mathbf{s}$ will be spread out over all possible measurements; undersampling will not accidentally miss an important coefficient. To give a negative example, assume that we design a system in which the sensing and sparsity matrices are identical and therefore maximally coherent. Perhaps our signal is sparse in the Fourier domain, meaning it can be well-represented by a few sinusoids (but we do not know which ones), and our sensor measures a few of these sinusoids. In this case, there is no hope of exact recovery since we would have to be extraordinarily lucky to sample the important sinusoids. Most (if not all) of our measurements would be zero.

Research has revealed several rules of thumb for incoherence [16]. For example, most wavelet bases are reasonably incoherent with both the Dirac and Fourier bases (indicating the non-transitive nature of incoherence). Randomly selected basis vectors from noiselet domains are incoherent with wavelets. Finally, even completely random Bernoulli ± 1 or Gaussian matrices are, with

overwhelming probability, incoherent with almost all sparsity bases of practical interest.

Further intuition can be gained on the motivation for using compressed sensing as a regularizer by considering a modern digital camera. These devices make millions of measurements with a CCD, and then the signal is fed to an on-board image compression chip. This chip performs a basis transformation (a Discrete Cosine Transform for JPEG, or a wavelet transform for JPEG-2000), and then keeps only the most important coefficients, yielding a compressed version of the image that is often perceptually lossless. Most of the data obtained in the measurement step are thus immediately discarded. Compressed sensing offers a framework to remove this inefficiency, collecting a form of the signal that is already compressed

These preliminary compressed sensing results show that if we have a signal that is sparse in a known basis, and a sensing basis that is reasonably incoherent with the sparsity basis, we can get away with a severely undersampled problem. By solving Problem (2.50), we can recover the sparse signal with many fewer measurements than previously thought possible. This raises the question of how to solve Problem (2.50). The problem of ℓ_0-minimization is an NP Hard problem, which means that global optimization would require an exhaustive search of exponential complexity. Fortunately, as we will see in the next section, Problem (2.50) has the same global minimum as ℓ_1-minimization, provided the RIP is satisfied.

2.2.2 Leveraging ℓ_1-minimization

Perhaps one of the greatest breakthroughs in compressed sensing theory is the proof that ℓ_1-minimization is a "sparsity-inducing" regularizer. This has allowed compressed sensing to be elevated from a mere mathematical curiosity to a powerful practical framework. Optimization theory is rich with ℓ_1-minimization problems, which are solvable with a variety of techniques such as linear programming. These results show that the solution to Problem (2.50) is also the solution to the following problem:

$$
\begin{aligned}
\text{minimize} &: \quad \|\mathbf{s}\|_1 \\
\text{subject to} &: \quad \mathbf{g} = \mathbf{H}^*\boldsymbol{\Psi}\mathbf{s}
\end{aligned}
\tag{2.53}
$$

where the ℓ_1 norm is the sum-of-magnitudes:

$$\|\mathbf{s}\|_1 = \sum_{i=0}^{N-1} |s_i| \tag{2.54}$$

This equivalence is by no means obvious and is a deep mathematical result. In this work, we will simply provide a geometrical example to build intuition [15]. Consider the case where $M = 1$ and $N = 2$, and for simplicity, $\mathbf{\Psi} = \mathbf{I}$, meaning our signal is directly sparse. Thus, we have one measurement of the form:

$$g_0 = h_0^* f_0 + h_1^* f_1 \tag{2.55}$$

and two unknowns f_0 and f_1. The act of recovering the signal, $\mathbf{f}$, is identical to choosing a point in $\mathbb{R}^N$, which is $\mathbb{R}^2$ in our current example. The data constrain us to only consider points on the line described by (2.55). Any point on this line agrees with the data, so regularization decides which point on this line we choose.

Tikhonov regularization, as discussed above, chooses the point with minimum ℓ_2 norm. Geometrically, as shown in Figure 2.4, this can be thought of as starting with a small sphere (the sphere being the "ℓ_2 ball", or the locus of points with a given ℓ_2 norm), and expanding it until it is tangent to the line. The tangent point is the minimum ℓ_2-norm solution. Sparse solutions lie along axes, so the minimum ℓ_2-norm solution is almost never sparse.
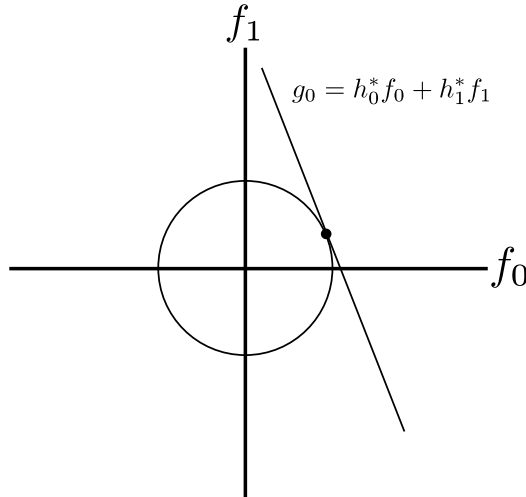

Figure 2.4: Regularizing with ℓ_2 minimization (Tikhonov regularization) does not produce sparse solutions.

On the other hand, consider ℓ_1-norm regularization. Here, we start with a small "ℓ_1 ball", and expand it until it touches the line, as shown in Figure 2.5. In $\mathbb{R}^2$, the ℓ_1 ball, the locus of all points with a given ℓ_1 norm, is a diamond. Here we can see that ℓ_1-minimization offers a sparse solution with high probability. Of course, low-dimensional intuition can only take us so far. In higher dimensions, the ℓ_1 ball becomes even more pointy, increasing the probability that the correct solution is recovered.
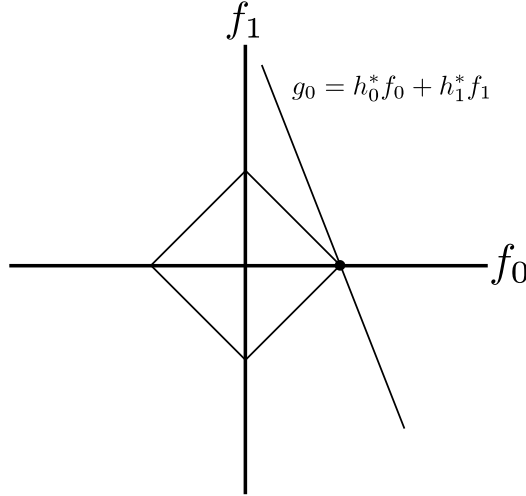

Figure 2.5: Regularizing with the ℓ_1 norm induces sparsity.

We have seen that ℓ_1-minimization is a substitute for ℓ_0-minimization, allowing for practical recovery of sparse signals. However, before compressed sensing can become truly practical, we must ensure that the problem is stable.

2.2.3 Statistical Estimation

The preceding discussions have all made two unrealistic assumptions: the data are noiseless, and the signal is actually sparse. In practice, noise is omnipresent, and signals, while possibly compressible, are never exactly sparse.

We first address the noise concern. Although our immediate intuition is that such a nonlinear problem would be extremely sensitive to noise, and that the surprising recovery results described above must only be true in the noiseless case, the truth is that this recovery scheme is indeed stable. However, one adjustment must be made. Instead of adhering exactly to the noisy data, we allow the algorithm to deviate from the data in search of

sparse solutions, provided that the deviation is within some estimate of the noise power, ϵ^2:

$$\begin{aligned} \text{minimize}: \quad & \|\mathbf{s}\|_1 \\ \text{subject to}: \quad & \|\mathbf{g} - \mathbf{H}^*\boldsymbol{\Psi}\mathbf{s}\|_2^2 \leq \epsilon^2 \end{aligned} \tag{2.56}$$

If this recovery scheme is used, then it can be proved that under some conditions related to the RIP, the worst case reconstruction error is bounded by a linear function of the noise level [13]:

$$\|\hat{\mathbf{s}} - \mathbf{s}\|_2 \leq C_1 \epsilon \tag{2.57}$$

where $\hat{\mathbf{s}}$ is the estimation of $\mathbf{s}$ and C_1 is a constant that in image processing applications is empirically around 10 [13]. Thus, sparse noisy recovery is stable.

The second concern is how this recovery scheme responds to non-sparse signals. Although many natural signals are compressible, meaning that they can be well-approximated by a small number of coefficients in some basis, they are not exactly sparse. Their lower-order coefficients are small but nonzero. Compression is necessarily lossy. Can we still use compressed sensing algorithms to reconstruct signals even if the signals are not sparse? The answer is a qualified yes. In these cases, we cannot hope to recover the true signal $\mathbf{s}$, but rather a compressed version of the signal, $\mathbf{s}_K$, which has all but the largest K coefficients set to zero. By solving Problem (2.56) in this case, we can be guaranteed [13]:

$$\|\hat{\mathbf{s}} - \mathbf{s}\|_2 \leq C_1 \epsilon + C_2 \frac{\|\mathbf{s} - \mathbf{s}_K\|_1}{\sqrt{K}} \tag{2.58}$$

Here we see that the reconstruction error is bounded by two error terms. The first is a function of the noise, as we have seen. The second is the inevitable error associated with approximating $\mathbf{s}$ as sparse. This term guarantees that the recovered solution is close to a compressed version of the actual signal. Again, C_2 is a constant that is empirically well-behaved [13]. Thus, our problem is indeed stable even when the data are noisy and the signal is not sparse, and since the solution is unique, the problem is also well-posed.

2.2.4 Compressed Sensing In Practice

In practice, inversions using compressed sensing depend heavily on the chosen sparsity basis, $\mathbf{\Psi}$. While a general-purpose basis may be discovered in the future, it does not appear that one basis will work for all problems. Nevertheless, certain bases have been found to be practical for many problems. For example, smoothly varying signals are compressible in a Fourier basis, piecewise constant signals are compressible with the Haar wavelets, and piecewise smooth signals are compressible with any number of alternate wavelet bases. Perhaps one of the most appealing properties of regularizing with compressed sensing is just this choice: allowing the user to choose the sparsity basis in order to tune the algorithm to certain signals of interest, or to bring out certain features in an image.

Algorithms to solve Problem (2.56) abound. This problem is known as Basis Pursuit Denoise, and algorithms for its solution are still emerging. Some are classical greedy algorithms, for example Matching Pursuit, Orthogonal Matching Pursuit (OMP), and Compressive Sampling Matching Pursuit (CoSaMP) [17]. These bear a striking resemblance to many ad hoc deconvolution methods such as the CLEAN algorithm [18]. Others solve smoothed versions of the ℓ_1 norm, for example Iterative Reweighted Least Squares [19]. Many others are completely novel approaches to the inversion problem; a few that we have found useful are SPGL1 [20], SpaRSA [21], and NESTA [22]. A major goal of current research in this direction is to find efficient methods for large scale problems.

A recent trend in compressed sensing is to use "overcomplete dictionaries" instead of an orthogonal sparsity basis. Essentially, this replaces the $N \times N$ matrix Ψ with a wider $cN \times N$ matrix, where c is an integer greater than one. This applies for signals which are sparse when expressed not in a single basis but as a combination of vectors from different bases. For example, some astronomical images consist of a smoothly varying background punctuated by stars, so the appropriate $\mathbf{\Psi}$ might be the concatenation of Fourier and Dirac bases. Another popular procedure for recovery of two-dimensional signals is to assume the signal is sparse in its Total Variation (TV) [12]:

$$TV(\mathbf{f}) = \sum_i |\nabla f_i| \qquad (2.59)$$

where ∇f_i is an approximation to the gradient of the underlying function f at location i. This is not strictly a basis but has found empirical success for signals that are piecewise constant (e.g., magnetic resonance imaging of the brain [23]).

The recent practical results and rich theoretical underpinning of compressed sensing make it a compelling choice for regularizing electromagnetic imaging problems. For the design of electromagnetic imagers, it is even more compelling. While current imaging architectures employ a "sample first, ask questions later" [24] approach to signal acquisition, compressed sensing is a more holistic approach to sampling. Furthermore, if a suitable measurement basis such as noiselets are chosen, the samples are universal (applying to almost any type of signal and sparsity basis), and future-proof (can be stored and decoded at a later date when processing algorithms have improved). Of course, the true value of compressed sensing for electromagnetic imaging must be judged not on theoretical grounds, but rather on the basis of simulation of real imaging scenarios, and application to real data. This will be the focus of the next two chapters.

CHAPTER 3

RADAR IMAGING WITH COMPRESSED SENSING

A conventional radar uses either a single antenna or multiple antennas connected together to simulate a large antenna. The radar can access range information by range-gating the signal returns, but it cannot access angular information in this simple configuration. Signals from different directions are summed together, weighted by the antenna beam pattern. Imaging requires access to this angular information, as discussed in Chapter 2. The only way to image in this case is to steer the antenna, either electronically or mechanically, or to keep the antenna stationary and allow the target to drift across the beam. In both cases, the spatial resolution is limited by the antenna beam width. Also, beam steering creates an ambiguity between temporal and spatial variation, as the true image may change as it is being observed.

A better solution is to use multiple antennas without connecting them together in a phased array, but rather recording the signals separately on each antenna. In this way, we obtain samples of the desired angular distribution, $f(\phi)$, in the Fourier domain. As we saw, this leads to an underdetermined problem in most practical cases. The optimal way to infer $f(\phi)$ from these incomplete measurements is a topic of debate in the field of radar imaging.

In this chapter we show how compressed sensing can be used to invert radar imaging data. We present simulation results in which compressed sensing outperforms existing methods and show some examples of its application to data collected at the Jicamarca Radio Observatory.

3.1 Radar Imaging Background

Before discussing the application of compressed sensing to invert radar imaging data, we first describe the history of ionospheric radar imaging and detail popular inversion methods.

3.1.1 History

Ionospheric radar imaging has its roots in early two-antenna interferometry experiments at the Jicamarca Radio Observatory in the 1970s and 80s [25,26]. These experiments were the first to make measurements from two antennas separately, and correlate the results to obtain angular information. As we know now, they were just measuring a single Fourier component of $f(\phi)$, but in these works they assumed a Gaussian shape for $f(\phi)$ and solved for the center location (which is related to the relative phase between the antennas), and the angular spread of the scattering region (known as the "coherence"). By monitoring the center location over time, they were able to deduce a drift velocity. This new method proved to be a rich source of information to study the horizontal dynamics of ionospheric irregularities.

While it is impressive what these researchers were able to achieve with one frequency component, it is clear that using more antennas would provide more angular information. However, it was not until 1991 that this technique was generalized, and the Fourier transform relationship between the measurements and the desired image was discovered in the context of radar imaging [27]. Interestingly, radio astronomers had known this result since at least the 1950s, and the formation of images in radio astronomy is extremely closely related to imaging with radar [7].

3.1.2 Basic Fourier Inversion

Although the 1991 work was a breakthrough in radar imaging, the method with which the data were inverted was naive, so the resolution of the resulting images was suboptimal. This method was based on the fact that the true image, $f(\phi)$, and the measurements, samples of $g(k\Delta x)$, are related with a Fourier transform, which we may write as

$$g(k\Delta x) = \int_{-\infty}^{\infty} f(\phi)e^{jk\Delta x\phi}\,\mathrm{d}\phi \qquad (3.1)$$

where k is the wavenumber and $g(k\Delta x)$ is obtained by cross-correlating the signals from two antennas separated by a distance Δx (see Section 2.1.7). This method then claimed that the best way to recover $f(\phi)$ is to attempt

to invert the Fourier transform as such:

$$f(\phi) = \int_{-\infty}^{\infty} g(k\Delta x) e^{-jk\Delta x \phi} \, \mathrm{d}k\Delta x \qquad (3.2)$$

Evaluating (3.2) requires measuring $g(k\Delta x)$ for all possible antenna separations. Since this is not practical, this early method attempts to mimic (3.2) using only the available measurements:

$$\hat{f}(\phi) = \sum_{m=0}^{M-1} g(k\Delta x_m) e^{-jk\Delta x \phi} \qquad (3.3)$$

where $\hat{f}$ is the estimate of f, ignoring constant multipliers. Note that this method implicitly assumes that the non-measured Fourier components are zero, a rather arbitrary choice not made in more advanced methods. A closer look at the ramifications of this assumption can provide some intuition on the resolution limits of this method.

Assuming that the non-measured values are zero is tantamount to assuming that the visibility function, $g(k\Delta x)$, is well-approximated by delta functions at its measured points:

$$\hat{g}(k\Delta x) = g(k\Delta x) \sum_{m=0}^{M-1} \delta(k\Delta x - k\Delta x_m)$$
$$\equiv g(k\Delta x)\, d(k\Delta x) \qquad (3.4)$$

where we define the function d as the delta function mask. Taking the Fourier transform of both sides and applying (3.3), we find:

$$\hat{f} = f * D \qquad (3.5)$$

where D is the Fourier transform of d, and the asterisk denotes convolution. In other words, the estimate of the image is the true image blurred by the function D. Evaluating D, we find:

$$D(\phi) = \sum_{m=0}^{M-1} e^{jk\Delta x_m \phi} \qquad (3.6)$$

Instead of indexing by the antenna separations, we will instead index by

the antenna locations directly, resulting in a double sum:

$$D(\phi) = \sum_i \sum_j e^{jk(x_i - x_j)\phi}$$

$$= \left(\sum_i e^{jkx_i\phi} \right) \left(\sum_i e^{jkx_i\phi} \right)^* \tag{3.7}$$

which we recognize as the squared magnitude of the receiving array beam pattern. The true image is convolved with the beam pattern, resulting in a resolution limit on the order of the beam width and ringing due to sidelobes. The sidelobes can be reduced by tapering the amplitudes elements in the array. For this reason, this method is often referred to in the literature as "spectral windowing," but, as we show above, it is no different than creating an image by sweeping the beam across the scene with a conventional single-receiver radar. Using this inversion method, the only advantage over a conventional radar is to separate temporal and spatial variation, but we do not overcome the spatial resolution limitations.

To fully realize the advantages of radar imaging, we must use more advanced techniques for inverting our data. We will describe in detail the two leading methods, Capon's method of imaging and the Maximum Entropy method.

3.1.3 Capon's Method

Capon's method of imaging is basically a minimum-variance beamformer. The original method was developed in the late 1960s to analyze seismic measurements [28], and it was applied to radar imaging in 1998 [29]. It can be viewed as a generalization of the beam steering approach. Instead of choosing a constant beam pattern in the form of a linear phased array, Capon's method optimizes the antenna weights in order to minimize interference, and it chooses different weights for different look directions.

Following the notation from Chapter 2, we denote the phasor measured at antenna i as $q(x_i)$. Stacking the measurements from all antennas yields the vector $\mathbf{q}$. The unknown antenna weights are denoted by $\mathbf{w}$. The output of the array is then

$$y = \mathbf{w}^* \mathbf{q} \tag{3.8}$$

where $\mathbf{w}^*$ denotes the conjugate transpose of $\mathbf{w}$.

We want y to be an optimal estimate of $A(\phi)$, the complex amplitude of the signal coming from the angle ϕ. One option is to choose the weights in the form of a linear phased array:

$$w_i = e^{-jkx_i\phi} \tag{3.9}$$

However, this is no different than the basic Fourier inversion described above. Signals from other directions will leak into the sidelobes of the beam pattern. We may think of the resulting output as being the sum of the desired signal with interference. To minimize the interference, we minimize the total output power while constraining the beam pattern to unity in the desired direction, ϕ. This constraint ensures that the desired signal, $A(\phi)$, passes to the output unmodified, while minimizing the interference power. Formally, Capon's method is:

$$\begin{aligned} \text{minimize:} \quad & \mathbb{E}\big[|y|^2\big] \\ \text{subject to:} \quad & \mathbf{e}^*\mathbf{w} = 1 \end{aligned} \tag{3.10}$$

where the vector $\mathbf{e}$ is:

$$\mathbf{e} = \begin{bmatrix} e^{jkx_0\phi} & e^{jkx_1\phi} & \cdots & e^{jkx_{I-1}\phi} \end{bmatrix}^{\mathrm{T}} \tag{3.11}$$

and I is the number of antennas.

Minimization under constraints is identical to unconstrained minimization of the following function with respect to the weight vector, $\mathbf{w}$, and the Lagrange multiplier, Λ:

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \Lambda) &= \mathbb{E}\big[|y|^2\big] + \Lambda(\mathbf{e}^*\mathbf{w} - 1) \\ &= \mathbb{E}[yy^*] + \Lambda(\mathbf{e}^*\mathbf{w} - 1) \\ &= \mathbb{E}\big[\mathbf{w}^*\mathbf{q}\mathbf{q}^*\mathbf{w}\big] + \Lambda(\mathbf{e}^*\mathbf{w} - 1) \\ &= \mathbf{w}^*\mathbb{E}\big[\mathbf{q}\mathbf{q}^*\big]\mathbf{w} + \Lambda(\mathbf{e}^*\mathbf{w} - 1) \\ &= \mathbf{w}^*\mathbf{R}\mathbf{w} + \Lambda(\mathbf{e}^*\mathbf{w} - 1) \end{aligned} \tag{3.12}$$

where we have defined the correlation matrix, $\mathbf{R} = \mathbb{E}\big[\mathbf{q}\mathbf{q}^*\big]$, and $\mathbf{q}^*$ denotes the conjugate transpose. Setting the partial derivatives of $\mathcal{L}$ equal to zero

results in a closed form for the weight vector:

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{e}}{\mathbf{e}^*\mathbf{R}^{-1}\mathbf{e}} \tag{3.13}$$

Recall from Section 2.1.7 that we are interested in imaging the variance of $A(\phi)$, so our estimate is:

$$\begin{aligned}
\hat{f}(\phi) &= \mathbb{E}\big[|y|^2\big] \\
&= \mathbf{w}^*\mathbf{R}\mathbf{w} \\
&= \frac{1}{\mathbf{e}^*\mathbf{R}^{-1}\mathbf{e}}
\end{aligned} \tag{3.14}$$

This provides a simple formula for estimating the power from a given direction based on the cross-correlations (which are the effective measurements, as described in Section 2.1.7). This procedure is repeated for many values of ϕ in order to generate an estimate of the function $f(\phi)$. Of all estimates of $f(\phi)$ that are linear functions of the received signals, Capon's method provides the minimum mean-square estimate. Another advantage of Capon's method is its relatively low computational complexity. For each pixel in the recovered image, we must only solve a linear system. Although Capon's method provides the optimal linear estimate, there is no reason we must restrict ourselves to linear processing schemes. Although nonlinear processing is generally more computationally complex, it allows access to more inversion methods with more theoretical and empirical justification. Both Maximum Entropy and compressed sensing use nonlinear inversions.

3.1.4 Maximum Entropy

The Maximum Entropy method (MaxEnt) is an attempt to adhere to the "first principle of data reduction" by Ables: "The result of any transformation imposed on the experimental data shall incorporate and be consistent with all relevant data and be maximally non-committal with regard to unavailable data." [30]. Many argue that the only way to conform to this principle and avoid making unfounded assumptions is to regularize by the entropy functional (e.g., [31–33]). In other words, of all the infinite candidate solutions that agree with the data, we should pick the solution with the

maximum entropy. The entropy is defined as:

$$S = -\sum_{i=0}^{N=1} f_i \log \frac{f_i}{F} \qquad (3.15)$$

where

$$F = \sum_{i=0}^{N-1} f_i \qquad (3.16)$$

and f_i denotes the ith element of $\mathbf{f}$, the discretized version of $f(\phi)$.

MaxEnt has a number of desirable properties. First, if we understand entropy as information content, then MaxEnt recovers an estimate with only as much information as allowed by the experimental data. To choose a solution with a higher entropy is to imply information not contained in the data. Second, entropy is a smoothness metric, in a certain sense. Images that are closer to the flat, uniform image have a higher entropy. Maximizing the entropy is equivalent to minimizing the relative entropy (or Kullback-Leibler divergence) between $\mathbf{f}$ and the uniform image. Of course, the entropy metric does not explicitly enforce spatial smoothness, since (3.15) is invariant under a relabeling of the indices. Finally, regularizing with entropy implicitly imposes a non-negativity constraint on $\mathbf{f}$, since negative images have an undefined entropy. In many practical applications, the desired image is known to be nonnegative. In the case of radar imaging, this is because the image represents the variance, which is related to scattered power.

Critics of MaxEnt claim that since there is nothing resembling "information" in the radio interferometry problem, the information-theoretical arguments that justify MaxEnt are not valid. They claim the practical success of MaxEnt arises because it keeps solutions close to the uniform image and enforces non-negativity, not because it has roots in information theory [34]. The apparent success of other similar regularization functions seems to support this claim [35].

Although the entropy expression seems somewhat arbitrary at first glance, it has an interpretation as a prior distribution. Many regularization schemes can be understood as Maximum a Posteriori (MAP) estimators, and MaxEnt is no exception, although this fact is rarely discussed in the literature. Imagine the proverbial monkey with a container of F radio photons, and an initially empty pixelated image, $\mathbf{f}$. The monkey takes a photon from the

32

bucket, chooses a pixel at random with uniform probability, and places the photon in this pixel. The monkey repeats this process until all photons have been distributed. Of all the possible images that agree with the data, we will choose the image that the monkey is most likely to create. The *a priori* probability of creating an image, $\mathbf{f}$, is proportional to the multiplicity of $\mathbf{f}$, the number of ways the monkey can create $\mathbf{f}$. By a simple counting argument, the multiplicity is given by the multinomial coefficient:

$$m(\mathbf{f}) = \frac{F!}{\prod_{i=0}^{N-1} f_i!} \tag{3.17}$$

We seek to maximize the prior probability, which we can do by maximizing the multiplicity. To make the computation easier, we will maximize the logarithm of the multiplicity, which leads to the same solution since the logarithm operator is convex. As we show below, the log-multiplicity is equal to the entropy.

$$
\begin{aligned}
\log m(\mathbf{f}) &= \log F! - \sum_{i=0}^{N-1} \log f_i! \\
&\approx F \log F - F - \sum_{i=0}^{N-1} (f_i \log f_i - f_i) \\
&= F \log F - \sum_{i=0}^{N-1} f_i \log f_i - F + \sum_{i=0}^{N-1} f_i \\
&= F \log F - \sum_{i=0}^{N-1} f_i \log f_i \\
&= \sum_{i=0}^{N-1} f_i \log F - \sum_{i=0}^{N-1} f_i \log f_i \\
&= -\sum_{i=0}^{N-1} f_i \log \frac{f_i}{F}
\end{aligned}
$$

We recognize the last line as the entropy. The second line used Stirling's approximation:

$$\log n! \approx n \log n - n \tag{3.18}$$

which is valid for large n. Note that this approximation is valid in our case, since the quantum nature of values of $\mathbf{f}$ are only a mathematical convenience.

We have the freedom to define the quanta size to be arbitrarily small, yielding arbitrarily large values of f_i. This effectively generalizes this procedure to our continuous case.

The MaxEnt procedure is often formulated as follows, in order to account for noisy measurements:

$$\text{maximize:} \quad -\sum_{i=0}^{N-1} f_i \log \frac{f_i}{F}$$
$$\text{subject to:} \quad \|\mathbf{g} - \mathbf{H}^*\mathbf{f}\|_2^2 \leq \epsilon^2 \tag{3.19}$$
$$F = \sum_{i=0}^{N-1} f_i$$

where $\mathbf{H}^*$ is the $M \times N$ linear operator implementing the forward model, $\mathbf{g}$ is the vector of measurements, and ϵ^2 is an estimate of the noise power (see Chapter 2). Of all images that could be described by the data corrupted by noise of power ϵ^2, we choose the one with the maximum entropy.

Instead of performing the constrained minimization in (3.19) directly, we can convert it to an unconstrained minimization of a function of Lagrange multipliers. The details are not shown here but can be found in the literature (e.g., [36, 37]). This results in the following set of equations:

$$f_i = \frac{Fe^{\sum_j h_{ij}^* \lambda_j}}{\sum_i e^{\sum_j h_{ij}^* \lambda_j}} \qquad \forall\, i \in [0, N-1]$$
$$\lambda_j = -2\Lambda \left(g_j - \sum_i h_{ij}^* f_i \right) \quad \forall\, j \in [0, M-1] \tag{3.20}$$
$$\Lambda^2 = \sum_j \frac{\lambda_j^2}{4\epsilon^2}$$

These $M + N + 1$ coupled equations can be reduced to M equations in the Lagrange multipliers $\{\lambda_j\}$. Performing a MaxEnt inversion is a matter of solving these M nonlinear equations. In the noiseless case, these equations have a unique solution, but we were unable to find an analogous result for finite ϵ^2. Empirically, applying Newton's method to solve these equations is found to yield a unique solution and converges for a wide range of initial guesses [35, 37].

MaxEnt inversions attain a higher resolution than basic Fourier inversions

with fewer sidelobe artifacts. An example of this can be seen in Figure 3.1, which shows actual data from observations of equatorial spread-F at the Jicamarca Radio Observatory. In the top pane, the width of main beam causes smearing artifacts in the reconstruction. In the bottom pane, we see that MaxEnt is able to resolve smaller-scale features. For example, a bifurcation is evident at 320 km, and vertically elongated structures on the bottomside layer between 250 and 275 km are seen to drift eastward and shear [35]. Subsequent work on this method has shown that further resolution improvements can be attained by accounting for correlated errors; since each measurement uses information from pairs of antennas, we should not expect our errors to be uncorrelated [36].
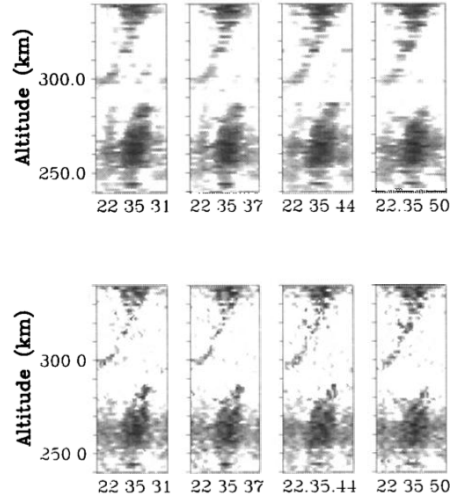

Figure 3.1: A comparison of the basic Fourier inversion (top) with MaxEnt (bottom) using observations of equatorial spread-F at the Jicamarca Radio Observatory. After [35]. Reproduced with permission of the American Geophysical Union.

An initial comparison of the Fourier, Capon, and MaxEnt inversions using a few representative image models and antenna layouts found that both Capon's method and MaxEnt are superior to the Fourier inversion. The results of Capon's method and MaxEnt were comparable, but MaxEnt was slightly superior, especially at low SNR [38]. A subsequent study reached similar conclusions [39]. As discussed in Chapter 2, radar imaging seems an ideal application for compressed sensing, yet to our knowledge there is no formal treatment in the literature. In the following section, we seek to bridge this gap by comparing compressed sensing to current methods in the context

of radar imaging.

3.2 Compressed Sensing Results

Radar imaging is usually an underdetermined problem. Each of the methods described above deals with this fact in different ways. The philosophy of compressed sensing is to recover the most parsimonious image consistent with the data, where parsimony is defined by a preference for a low-bit representation. In other words, we reconstruct a sparse image. It is our task to choose a domain in which the image will be sparse. This choice allows us to tailor the inversion to particular classes of signals or to bring out certain image features, a tool unavailable in the above methods. However, even with a simple sparsity domain such as wavelets, compressed sensing inversions can outperform existing methods, as we show in the following sections.

3.2.1 Simulation Details

We simulate a simple radar imaging experiment to test the compressed sensing method against Capon's method and MaxEnt. We used the antenna locations from the imaging configuration of the Jicamarca Radio Observatory near Lima, Peru. These are shown in Figure 3.2. For this experiment, we assume that ionospheric structure does not depend on magnetic latitude, and therefore we can reconstruct an image purely as a function of range and magnetic longitude. This assumes that scatterers are constant along magnetic field lines. This is a typical assumption made in ionospheric imaging that is theoretically valid for large, km-scale features, which tend to be elongated along the Earth's magnetic field [40]. For each range gate, we seek to recover the angular distribution in the magnetic east-west direction, $f(\phi)$. In this work we simulate this one-dimensional imaging problem. For the antenna locations, the relevant coordinate is the magnetic east coordinate.

In keeping with previous simulation studies, we start with a Gaussian as our true image, $f(\phi)$. We note that a real ionospheric image will not be as simple as a single Gaussian, nor is it the case that the results of this simulation generalize using superposition, since all of these techniques (except the Fourier inversion) are nonlinear in the sense that they adapt to the data.
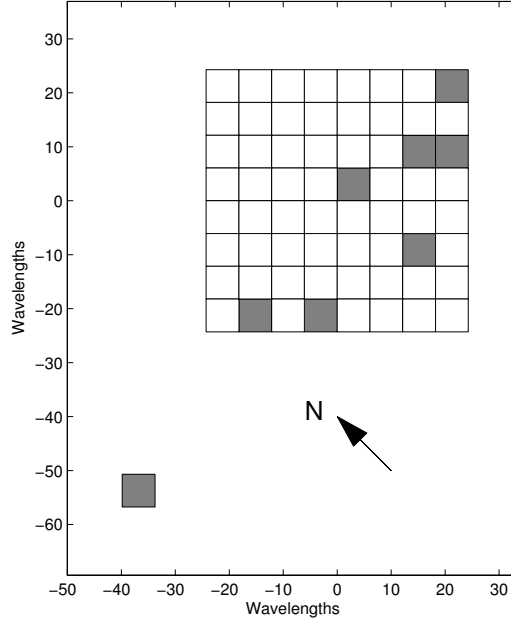
Figure 3.2: The antenna array at the Jicamarca Radio Observatory. The locations of the receiving antenna modules used in the simulation are shown in gray. Each module comprises 36 half-wave dipoles. The arrow indicates geomagnetic north, which is within 1° of geographic north.

Thus, we use a Gaussian as a proxy, and interpret our results conservatively. We use a Gaussian centered at $\phi = -3°$, with a width of 2°. We have tried other image models, such as multiple Gaussians, rectangles, and triangle functions, and we obtain qualitatively similar results.

The measurements, $g(k\Delta x)$, are Fourier coefficients of $f(\phi)$ specified by the antenna separations, as indicated by Equation (3.1). We add error to these measurements and reconstruct with all four methods. Realistically, this error arises from the finite time intervals used to estimate the signal correlations and is a function of the probability distribution of the signal and receiver noise (see Section 2.1.7). In the present simulation, we model this error as uncorrelated Gaussian noise for simplicity. This simplifying assumption was used in a previous simulation study [39]. A more detailed forward model [38] and a first-order error model for the inversion [36] can be found in the literature. We define the signal-to-noise ratio (SNR) as:

$$\text{SNR} \equiv \frac{\|\mathbf{g}_{\text{exact}}\|_2}{\|\mathbf{g} - \mathbf{g}_{\text{exact}}\|_2} \tag{3.21}$$

where $\mathbf{g}_{\text{exact}}$ is the exact measurement vector, comprising correlations without

error, and **g** includes the introduced error. We note that this definition of SNR does not correspond to the actual SNR at the radar receiver front end. Our definition includes the effect of signal variance and finite integration times. In any case, in this study we are not concerned with the absolute value of SNR, only with how the comparisons change as we change the uncertainty.

We reconstruct an image, $\hat{f}(\phi)$, onto the domain $\phi \in [-9°, 9°]$, which is large enough to include virtually all of the energy in $f(\phi)$. We reconstruct onto a grid with a coarser resolution than was used to simulate the measurements. The results of our inversions do not change qualitatively when we change the grid size. Both MaxEnt and compressed sensing require building a discrete forward linear operator as a matrix $\mathbf{H}^*$, such that our measurements are modeled linearly as

$$\mathbf{g} = \mathbf{H}^*\mathbf{f} \tag{3.22}$$

as described in Section 2.1.6. In order to reconstruct a real image, we must be careful with how we build $\mathbf{H}^*$, since the cross-correlations are in general complex. One option is to use the fact that the Fourier transform of a real function is even. This constraint creates two effective measurements per correlation, one each for the positive and negative spatial frequencies. This is identical to treating the antenna pair (i, j) as separate from the pair (j, i). However, we use a different but effectively identical option. We operate entirely with real numbers, treating the real and imaginary parts of $g(k\Delta x)$ as two separate measurements. This way, the matrix $\mathbf{H}^*$ is purely real and consists of sines and cosines instead of complex exponentials.

MaxEnt and compressed sensing both have a free parameter, ϵ, the estimate of the magnitude of the error. In this simulation, we assume this value is known exactly. In a real scenario, one can sample range and Doppler bins without signal to obtain an estimate of the receiver noise power and use this, along with the known integration time, to obtain an estimate of the error magnitude. Of course, this estimate of error magnitude is subject to error, but this is a complication we do not consider.

Obtaining solutions for the Fourier and Capon's method is straightforward. To obtain the MaxEnt solution, we solve the M nonlinear equations generated by (3.20) using MATLAB's `fsolve` function. We use a small, constant initial guess of $\lambda_j = 10^{-5}$, but we find that the solution is not sensitive to the initial

guess. As discussed in Section 2.2.3, the problem to solve for compressed sensing is (2.56):

$$\begin{aligned} \text{minimize} : \quad & \|\mathbf{s}\|_1 \\ \text{subject to} : \quad & \|\mathbf{g} - \mathbf{H}^*\boldsymbol{\Psi}\mathbf{s}\|_2^2 \leq \epsilon^2 \end{aligned} \tag{3.23}$$

where $\boldsymbol{\Psi}$ is the sparsity basis, and the image is subsequently recovered with $\hat{\mathbf{f}} = \boldsymbol{\Psi}\mathbf{s}$. This problem is often referred to as Basis Pursuit Denoise. To solve this problem, we use SPGL1 [20, 41].

For the sparsity basis, $\boldsymbol{\Psi}$, we use the Daubechies-4 wavelets. The motivation for using wavelets comes from the widespread success of the JPEG compression standard. The original JPEG uses a block-version of the Discrete Cosine Transform, which can be seen as an approximation to wavelets, and JPEG-2000 uses wavelets directly. Their success supports the claim that many natural images are inherently sparse; they have structure that is localized both in space and spatial frequency, and are thus compressible in the wavelet domain. Another attractive feature of wavelet transforms is that extremely fast implementations have been developed, of complexity $\mathcal{O}(N)$, which is even faster than a Fast Fourier Transform. To perform wavelet transforms we utilize MATLAB's Wavelet Toolbox.

The Daubechies-4 wavelets in particular were chosen because they were found useful in image processing applications [42]. However, this choice is somewhat arbitrary, as there exist many varieties of wavelets. The Daubechies-N wavelets are characterized by a maximum number of vanishing moments, meaning that they are optimal for polynomials up to order N. The higher N, the more vanishing moments, but the wider the support of the wavelet atom. Other wavelets have different design criteria and will give quantitatively different results, though the broad morphological features remain the same.

3.2.2   Simulation Results

In this section we show typical inversion results for high-, mid-, and low-SNR cases, and quantitatively compare the four inversion methods. A representative mid-SNR inversion is shown in Figure 3.3. The SNR is 15 dB. The forward model is calculated with 16,384 grid points, and the reconstruction grid is 256 points. Since the Fourier and Capon's methods do not return

meaningful absolute values, we have multiplied these reconstructions in order to make their maximum value equal the maximum value of the true image. The MaxEnt and compressed sensing inversions have not been modified. Qualitatively, it appears as if compressed sensing provides a reconstruction that is closer to the truth and has fewer artifacts.
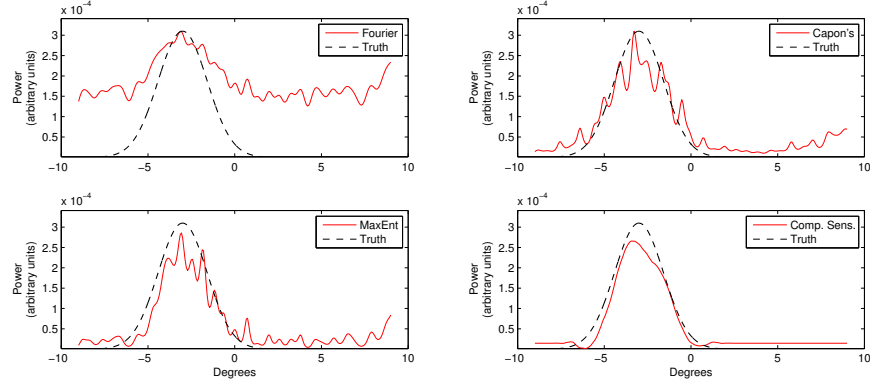

Figure 3.3: Inversions of simulated data using four methods with SNR = 15 dB. The black dotted line is the true image. The red line is the recovered image, using the basic Fourier inversion (top left), Capon's method (top right), MaxEnt (bottom left), and compressed sensing (bottom right).

At very high SNR, MaxEnt and compressed sensing both appear to converge to the true image. A representative inversion for a 25 dB SNR is shown in Figure 3.4. Capon's method does not converge to the true image, since side lobes in the beam pattern are still present, even though they have been arranged to minimize interference, and the main beam has nonzero width. In the Fourier method, we see a rather severe artifact near the right edge of the recovered image, possibly due to a large sidelobe or grating lobe.

At very low SNR, the signal is buried so deep in the noise that recovery is practically impossible with any method. With the definition of SNR used here, this occurs for SNR below about 5 dB. A representative inversion of data with a 5 dB SNR is shown in Figure 3.5. We can see that each method deals with this high-noise situation in different ways. With the Fourier inversion, white noise in the frequency domain looks like white noise in the space domain. Capon returns spikes at locations where it was unable to cancel the noise. Since the constraints imposed by the data are so loose in low-SNR cases, MaxEnt provides an image that is close to the globally maximum entropy image, which is constant. Compressed sensing provides the sparsest
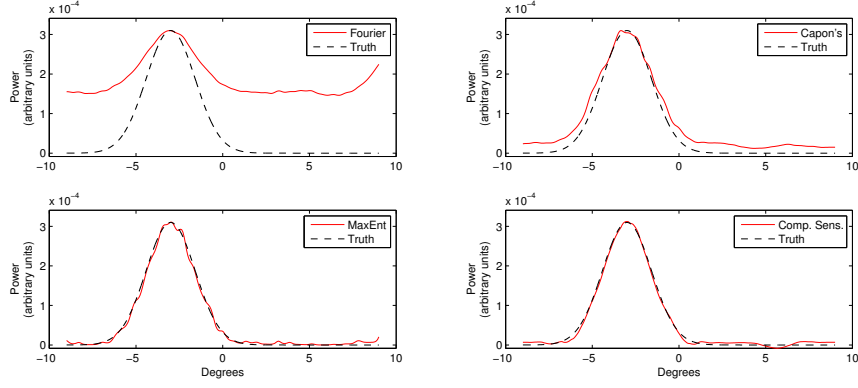
Figure 3.4: Inversions of simulated data using four methods with SNR = 25 dB. The black dotted line is the true image. The red line is the recovered image, using the basic Fourier inversion (top left), Capon's method (top right), MaxEnt (bottom left), and compressed sensing (bottom right).

image consistent with the data, which consists of one wavelet in this case, but is sometimes as low as zero if the data do not support even a single wavelet.
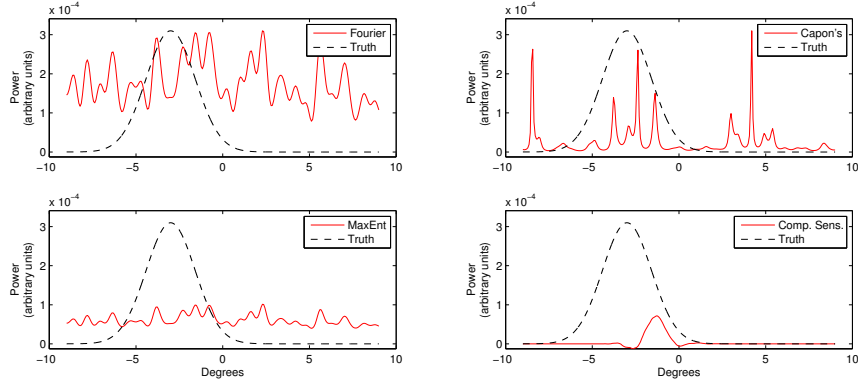

Figure 3.5: Inversions of simulated data using four methods with SNR = 5 dB. The black dotted line is the true image. The red line is the recovered image, using the basic Fourier inversion (top left), Capon's method (top right), MaxEnt (bottom left), and compressed sensing (bottom right).

In order to quantitatively compare these four methods, we use a correlation metric as a proxy for image quality, following previous work [38]. We define the correlation as:

$$\text{corr} \equiv \frac{\hat{\mathbf{f}}_{\mathbf{c}}^{\mathrm{T}} \mathbf{f}_{\mathbf{c}}}{\|\hat{\mathbf{f}}_{\mathbf{c}}\|_2 \|\mathbf{f}_{\mathbf{c}}\|_2} \tag{3.24}$$

where the subscript $\mathbf{c}$ denotes removal of the DC value. Removing the DC value is important so that this metric evaluates image structure, without in-

41

cluding the effect of a constant offset. In other words, a flat image should have a correlation of 0 with our Gaussian, and a perfect reconstruction should have a correlation of 1. The higher the correlation, the closer the reconstructed image matches the true image.

We run simulations for different values of SNR between 0 dB and 25 dB and measure the correlation for each of the four methods. For each SNR, we run 100 trials with randomized error realizations in order to minimize statistical fluctuations. The average correlation is calculated for each SNR and plotted in Figure 3.6. Compressed sensing outperforms all other methods for virtually all SNR, except for a small dip at 6 dB. Surprisingly, Capon's method is inferior for most SNR. Other image metrics than correlation could be used, such as mean-square error or other modified correlative measures. While quantitatively different, results from using some of these other methods are found to be qualitatively similar. These results match qualitatively the results in previous simulation studies, although other studies have shown better performance for Capon's method [38, 39].
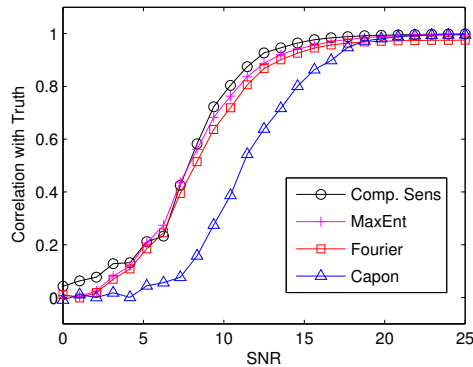

Figure 3.6: A statistical comparison of the four methods of inverting simulated radar imaging data. The average correlation is reported for each SNR for each method.

In comparing these four methods, one must also consider the tradeoff between the quality of the reconstruction and the running time. In our reconstructions, we find that Capon's method takes the same amount of time as the Fourier method. MaxEnt requires the most running time, at about 100 times that of the Fourier method. Compressed sensing inversions take about 30 times longer than Fourier inversions. We note that these values are not general. Since iterative methods are used in MaxEnt and compressed sens-

ing, running times can vary with SNR, image model, or even the particular error realization. Furthermore, if the number of grid points or number of measurements changes, these comparisons could change significantly.

Before moving on to real data, we must supply a word of caution. Nonlinear processing schemes can achieve higher resolution, but they can also produce spurious reconstructions. With linear reconstructions such as the Fourier method, the output is a linear function of the signal and noise. Thus, as the noise increases, the reconstruction degrades smoothly. This is not the case for nonlinear imaging, which often exhibits a "thresholding" behavior. In particular for compressed sensing, a small change in the noise level may lead to one more or fewer atom in the reconstruction, which can significantly change the interpretation of the image. Consider as an example the inversion in Figure 3.7, which uses 5 dB SNR. The compressed sensing inversion consists of a single wavelet in the wrong location. The Fourier inversion is noisy everywhere. Even though the MaxEnt inversion is nonlinear, the entropy constraint seems to inhibit this nonlinear thresholding effect. We see this pathology in about 5% of the low-SNR compressed sensing inversions. However, this effect is often not practically important, since low-SNR reconstructions are not given much weight in the formation of the full two-dimensional image.
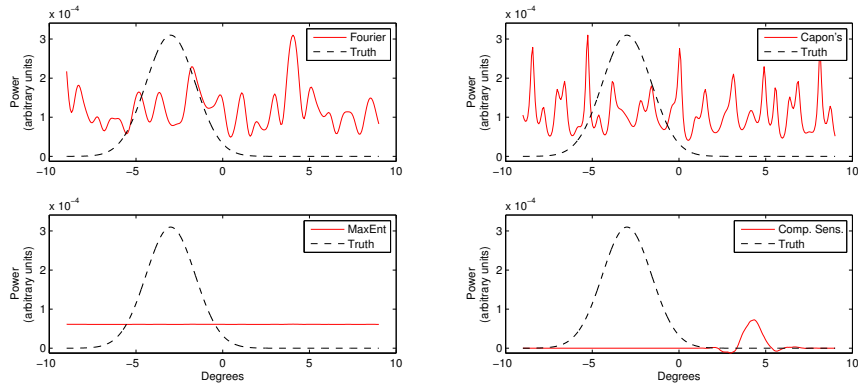

Figure 3.7: An example of the problem with nonlinear reconstructions, specifically compressed sensing. It can suggest signal structure where there is none. The SNR is 5 dB. The black dotted line is the true image. The red line is the recovered image, using the basic Fourier inversion (top left), Capon's method (top right), MaxEnt (bottom left), and compressed sensing (bottom right).

3.2.3 Jicamarca Radio Observatory Data

To further evaluate the performance of these four methods of inverting radar imaging data, we apply them to data from the Jicamarca Radio Observatory (JRO) near Lima, Peru (geographic: 11.95°S, 76.87°W). The magnetic dip angle at JRO is approximately 1°. The data were acquired at 22:18:30 LT on day 277 of 2011. An interval of 12 seconds was used to estimate the signal correlations. It is assumed that image features are stationary over such a short interval. We perform inversions for every range gate and Doppler bin, and add the results from different Doppler bins to obtain a grayscale two-dimensional image. In some applications, the Doppler bins are assigned colors, resulting in a color image, but we do not do so here. The results of different inversion methods are shown in Figure 3.8. As discussed in Chapter 1, these images represent backscattered power at 50 MHz, which is mostly determined by 3m-irregularities in electron density. These large turbulent plumes of plasma are indicative of a phenomenon known as equatorial spread-F. The genesis and growth of equatorial spread-F is of great interest in ionospheric research [2, 4].

The true image in this case is unknown, so it is impossible to accurately assess image quality. However, we believe that compressed sensing gives the best resolution and image clarity of the four methods. The MaxEnt method is more blurred and has a nonzero background level in some range gates. The Fourier and Capon images have severe artifacts that could be due to the large beam width and grating lobes. However, we suspect that this could also be due to a miscalibration with the phase measurements or antenna locations. MaxEnt and compressed sensing are affected differently because the data format for these methods was provided pre-calibrated. Additionally, MaxEnt and compressed sensing have more freedom to deviate from the data to correct any possible miscalibration.

3.2.4 Conclusion

Ionospheric radar imaging is an underdetermined problem, so it is necessary to incorporate prior information in order to recover a solution. In this chapter, we showed how compressed sensing can be used to invert these data. Simulations using a simple Gaussian image suggest that compressed sensing
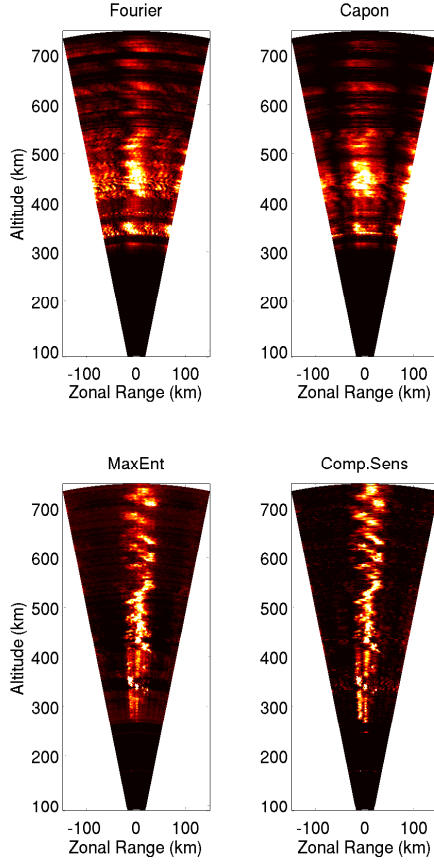
Figure 3.8: An image of equatorial spread-F recovered using four inversion methods. Data is from the Jicamarca Radio Observatory, year 2011, day 277, 22:18:30 LT.

can outperform MaxEnt, Capon's method, and the basic Fourier inversion, which are the popular current methods for radar imaging. This result appears to hold for most values of SNR. A more realistic forward and error model will be necessary to corroborate these results. The application to real data further suggests that higher resolution is possible with compressed sensing. One item for future work is to robustly impose a nonnegativity constraint, since scattered power cannot be negative. While there are some results regarding nonnegative compressed sensing, to our knowledge, there are no known algorithms for an arbitrary sparsity basis [43]. We also wish to investigate whether compressed sensing theory can aid in not only the analysis, but also the design of radar imaging experiments. The optimal antenna layout may

be the one yielding maximum incoherence with the sparsity basis.

Radar imaging is one example of how compressed sensing can be used for spatial electromagnetic imaging. In this case, our instrument was already built, and we use compressed sensing to improve our image reconstruction. We recognized our radar imager as a compressive sensor. However, a powerful impact of compressed sensing is to inform alternative instrument designs which optimize the data acquisition and the image reconstruction together, resulting in more efficiency in both steps. In the next chapter, we will discuss such an instrument and its application to optical imaging of airglow layers in the ionosphere.

CHAPTER 4

OPTICAL AIRGLOW IMAGING WITH COMPRESSED SENSING

In this chapter we review the concept of the single-pixel camera, an optical imaging instrument inspired by compressed sensing. This instrument can be applied to image irregularities and wave structure in naturally-occurring airglow emissions. A problem with conventional CCD-based imagers is that sensitive detectors are expensive and only viable at visible wavelengths. By replacing the CCD with a single photodiode, the single-pixel camera enables imaging for a larger range of wavelengths and at lower cost compared to conventional imagers. We present preliminary simulations of the single-pixel camera observing a typical airglow image, and suggest ideas for future work.

4.1   Airglow Imaging Background

The aurora is one well-known phenomenon resulting from natural optical emission in the ionosphere. It is a result of photon emission when atoms transition from an excited state to a ground state. These atoms are excited by collisions with solar wind particles precipitating down along magnetic field lines. However, ionospheric photon emission is not constrained to the polar regions. Indeed, at all latitudes there exist excited atoms and therefore optical emissions, referred to as airglow. Instead of particle precipitation, the source of airglow is ultraviolet radiation from the sun that ionizes neutral particles. As these ionized particles recombine over the course of the night, they create excited atoms that emit photons, resulting in airglow [44]. This airglow offers a natural probe for ionospheric measurements.

The interplay of neutral and plasma density profiles constrains particular emissions to certain altitude ranges. For example, the dissociative recombination of O_2^+ at 630.0 nm occurs almost entirely between 200 and 350 km. However, the majority of this emission occurs near 250 km, so a com-

mon approximation is to assume that the airglow layer is a thin shell [44]. Under this assumption, an image from a ground-based imager provides a two-dimensional map of airglow emission in the latitude-longitude plane. Of course, this assumption loses validity for low elevation angle observations, for which the thickness of the airglow layer leads to a blurring of the horizontal structure. If we are imaging the airglow layer during equatorial spread-F conditions, the airglow structures are known to map along magnetic field lines, and therefore, the latitudinal direction in the images corresponds to the altitudinal direction at the magnetic equator [45].

Thus, much like in Chapter 3, we obtain a two-dimensional altitude-longitude image. However, the physical interpretation of the image differs from Chapter 3, in which we described radar imaging of 3m-scale irregularities. In the airglow case, the emission is mostly, but not entirely, determined by electron density. Variations in electron density are associated with polarization electric fields. Large, km-scale features of this field map efficiently along magnetic field lines. Therefore, the image we obtain with optical imaging is of the large, km-scale features of the electron density.

Of course, airglow imaging is useful for more than just equatorial spread-F observations. It has a rich history of productive measurements in the both the mesosphere and thermosphere. These data provide a spatial context for studying the dynamics of waves and instabilities. In addition to equatorial spread-F, they have contributed significantly to our understanding of gravity wave propagation, for example [3, 46].

4.1.1   Existing Instrumentation

Before the advent of CCDs, airglow imaging was either performed with film or by spatially scanning a single photodetector. With film, the interpretation of absolute intensity was difficult, and with scanning, the separation of temporal and spatial structure was difficult. The use of CCDs bridged this gap and initiated a new era in airglow imaging with unprecedented spatial and temporal resolution [3].

The physical model for a typical CCD-based airglow imager is essentially captured by the discussion in Section 2.1.5 and Figure 2.2. Using a lens, the desired image, $f(\phi)$, is well-approximated by the signal on the image plane.

Sampling this signal corresponds to sampling $f(\phi)$. Using a CCD, this signal is directly sampled. To use the language developed in Section 2.1.6, the test functions are delta functions, which means that the measurement matrix, $\mathbf{H}$, is the identity matrix. This greatly simplifies signal reconstruction, since we can create an image directly from the measurements.

However, using a CCD has several disadvantages. The primary disadvantage is cost. To achieve the sensitivity needed for observations at 630.0 nm, for example, a CCD costs around \$25,000 (based off the Andor iKon-M 934 CCD). This cost is a hindrance to obtaining a global understanding of ionospheric dynamics, which requires many imagers deployed around the world. Also, while silicon has a reasonable quantum efficiency at visible wavelengths, the quantum efficiency falls off in the infrared band. To make infrared measurements, we would need to use other materials, which can make arrays such as CCDs prohibitively expensive. To make low-cost measurements of infrared airglow emissions, CCDs are not a viable option.

Another disadvantage of CCDs is evident when one considers the widespread success of image compression. In Figure 4.1, we show a rudimentary image compression example. In the left panel is a typical airglow image where the stars have been removed. To obtain the compressed version, we perform a wavelet transform using the Daubechies-4 wavelets, set to zero all but the largest 5% of the wavelet coefficients, and transform back. The resulting image is shown in the right panel. This compressed image retains most of the signal structure, including the correct absolute values, and even arguably provides some denoising. We are tempted to question the necessity of obtaining so many measurements of $f(\phi)$ when we could obtain a virtually identical image with many fewer pieces of data.

Compressed sensing, introduced in Chapter 2, offers a useful framework with which to address this concern. Indeed, using compressed sensing for optical imaging has inspired the so-called single-pixel camera. By replacing the CCD with a single photosensor, this single-pixel camera can reduce the cost of conventional airglow imagers and, because photosensors can be sensitive to the infrared, allow us access to spatial information at these wavelengths. In the remainder of this chapter we discuss this instrument and present some simulations pertaining to airglow imaging.
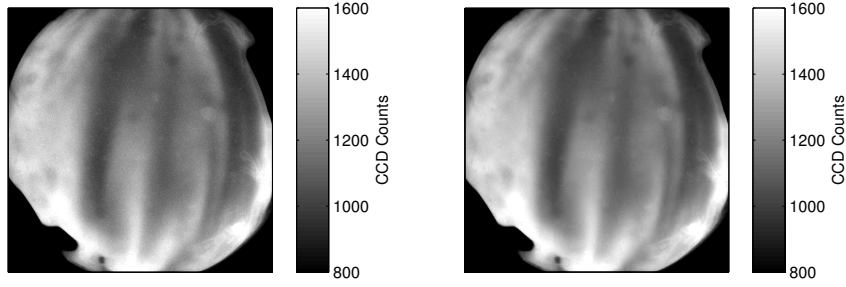
Figure 4.1: (Left) A typical airglow image showing the optical signature of Equatorial Spread F. (Right) The compressed version of the image, using the largest 5% of the wavelet coefficients.

4.2 Compressive Single-Pixel Airglow Imager

The intuition from Figure 4.1 suggests that if we were able to measure the important wavelets directly, then we could faithfully reconstruct the image with 20 times fewer measurements than a CCD. However, compressed sensing theory tells us that the optimal strategy is the exact opposite. We do not know ahead of time which wavelet coefficients will be important, so instead of measuring in the wavelet domain, we should measure in an incoherent domain. Although there exist many choices, one example used in practice is a domain consisting of test functions with random Bernoulli 0/1 entries, which is incoherent with wavelet domains [16].
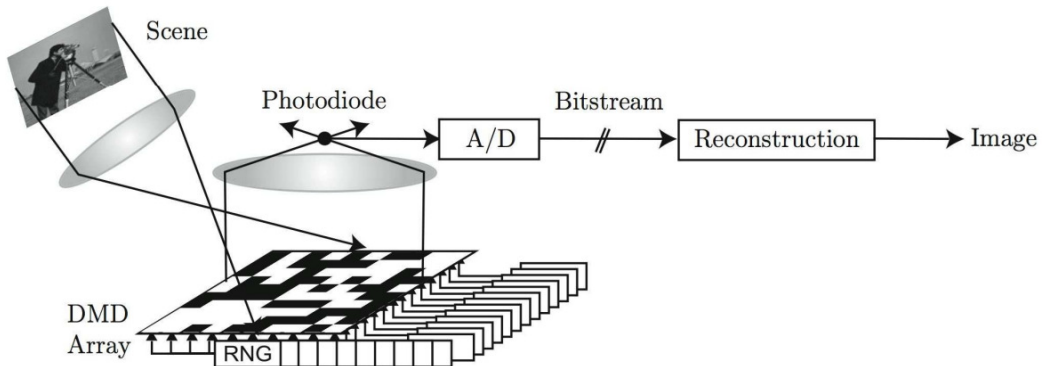

Figure 4.2: The signal path of the single-pixel camera. After [24]. Reproduced with permission of SPIE and the authors.

The single-pixel camera, originally developed in 2006, is one implementation of these ideas [24]. The architecture is shown in Figure 4.2. The initial portion of the signal path is identical to the CCD-based imager, but instead of a CCD, we place a digital micromirror device (DMD) in the image plane. A DMD is an array of programmable pivoting mirrors. The mirrors can be independently commanded to reflect light towards a photodiode, or away from it (represented by white and black squares in Figure 4.2). The DMD is the primary component of digital light projection (DLP) chips, which are an integral part of modern digital projectors and are thus widely available. The DMD pattern (the "mask") determines which pixels get measured and which ignored. The second lens focuses the reflected pixels into the photodiode, which records the intensity for this mask. After recording, the DMD is commanded to a new mask, and the process repeated. When enough measurements have been gathered, a nonlinear compressed sensing inversion is performed to estimate the image from these measurements.

Each measurement is simply the sum of intensities of certain regions of our image. These regions correspond to the locations of the reflective mirrors and can be thought of as the "pixels" of the image we wish to recover. If we index the N pixels in our image into a vector, $\mathbf{f}$, we can model a measurement as, for example:

$$g_m = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & \dots & 0 & 1 \end{bmatrix} \mathbf{f} \tag{4.1}$$

Stacking M of these measurements into a vector, $\mathbf{g}$, gives us our generalized forward model from Section 2.1.6, Equation (2.35):

$$\mathbf{g} = \mathbf{H}^*\mathbf{f} \tag{4.2}$$

where each column of the measurement matrix, $\mathbf{H}$, corresponds to a DMD mask.

The reconstruction is performed by solving the Basis Pursuit Denoise problem from Section 2.2.3 (Equation (2.56)):

$$\begin{aligned} \text{minimize}: \quad & \|\mathbf{s}\|_1 \\ \text{subject to}: \quad & \|\mathbf{g} - \mathbf{H}^*\mathbf{\Psi}\mathbf{s}\|_2^2 \leq \epsilon^2 \end{aligned} \tag{4.3}$$

where $\mathbf{\Psi}$ is the sparsity basis (e.g., wavelets), and $\mathbf{s}$ is the transformation of $\mathbf{f}$ into this basis, so $\mathbf{f} = \mathbf{\Psi}\mathbf{s}$.

An advantage of the single-pixel camera is the "sample first, ask questions later" philosophy [24]. Incoherent measurements are universal, which means that a string of measurements can be acquired without regard to how they will be used in the future. No knowledge of the signal details is necessary to design and run the instrument. Upon reconstruction, the important decisions are made: setting the acquisition time, choosing a sparsity basis, and choosing an algorithm with which to invert the data. This is in opposition to current imaging systems, where the important decisions are often made in the design phase.

It is assumed that the image $f(\phi)$ is constant over the time interval the measurements are being acquired. If $f(\phi)$ changes significantly from one mask to the next, then (4.2) would no longer provide a valid forward model since each measurement would correspond to a different $\mathbf{f}$. CCD-based imaging does not suffer from this problem since all measurements are taken simultaneously. However, the advantage of single-pixel imaging is that each measurement is a sum of pixels and therefore has a higher SNR, requiring a shorter exposure time per measurement. As we show below, as long as $M < N/2$, then the total exposure time for the single-pixel camera is actually shorter than that for a CCD.

Suppose we wish to ensure that the SNR of each measurement is equal to the SNR of one CCD pixel. We assume that Poisson noise dominates over read noise and dark current and that readout time is zero. Poisson noise has an SNR of:

$$\text{SNR}_{\text{Poisson}} = \frac{Pt}{\sqrt{Pt}} = \sqrt{Pt} \tag{4.4}$$

where P is the incident photon rate, and t is the exposure time per measurement. We denote the incident photon rate for a CCD pixel as P_0. For a single-pixel camera, if we assume that half the pixels in the mask are reflected into the photodiode, the photon rate is $\frac{N}{2}P_0$. Equating the SNRs, we find:

$$\text{SNR}_{\text{SPC}} = \text{SNR}_{\text{CCD}}$$

$$\sqrt{\frac{N}{2}P_0\, t_{\text{SPC}}} = \sqrt{P_0\, t_{\text{CCD}}} \tag{4.5}$$

$$t_{\text{SPC}} = \frac{2}{N}t_{\text{CCD}}$$

The total exposure time for the single-pixel camera is given by Mt_{SPC}, where

M is the number of measurements used in a single inversion. The total exposure time for the CCD is simply t_{CCD}. Thus, to ensure that the exposure time for the single-pixel camera is less than for the CCD, we must have:

$$M\, t_{\text{SPC}} < t_{\text{CCD}}$$
$$M < \frac{N}{2} \tag{4.6}$$

As we increase M, we have more measurements of our scene and thus reconstruct a more accurate image with a better spatial resolution. However, a larger M decreases our temporal resolution. As we show above, the "break-even point" for temporal resolution is at $M = N/2$. This is the point at which the exposure times for CCD and single-pixel camera are equal. The question is whether this many measurements can provide a suitable spatial resolution. To answer this question we must resort to simulations, presented in the next section.

Before moving on, we note that in practice this bound is probably improved, due to two effects. First, the nonlinear inversion used for the single-pixel camera tends to suppress noise for sparse images, so one must question whether setting the SNRs equal is the appropriate criterion. Second, a quantum efficiency of 100% was assumed. In reality, we can obtain higher quantum efficiencies with photodiodes than with CCDs. Both of these effects will reduce the necessary exposure time of the single-pixel camera compared to a CCD and thus improve upon the simple analysis above.

4.3 Simulations

As discussed in Chapter 2, it is difficult to theoretically prove the performance of compressed sensing techniques when applied to real problems. Thus, we must judge these techniques on the basis of simulation. For our simulations, we must construct a truth image, choose a measurement matrix, $\mathbf{H}$, add noise to the measurements, and invert the data to create an image.

4.3.1 Procedure

Although we could design an artificial truth image, we choose to create a truth image based off of data, in order to capture the complex details of realistic airglow images. We start with the raw 512 x 512-pixel image from an imager in Cajazeiras, Brazil, taken November 6, 2010 using a 630.0-nm filter. This image is shown in Figure 4.3. For an initial proof of concept, we remove the stars and increase the pixel size by a factor of 4 in each dimension. This downsampling is used to speed up the inversion and reduce noise in the image. The resulting 128 x 128-pixel image is shown in Figure 4.4, which will be our truth image.
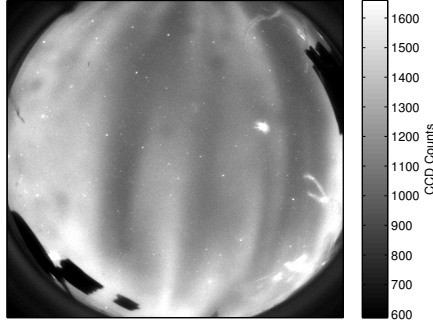

Figure 4.3: The raw image that is used to generate the truth image for the simulation (512 x 512 pixels).
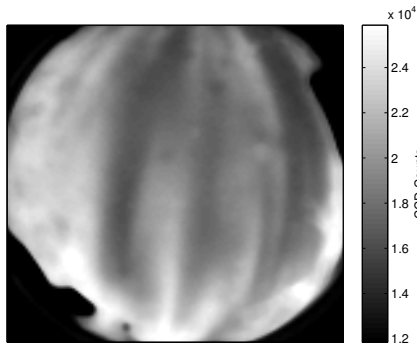

Figure 4.4: The truth image used in the simulation (128 x 128 pixels).

Our next choice is to design the pixel masks which compose the measurement matrix $\mathbf{H}$. Although many choices are possible, there are a few criteria that we would like $\mathbf{H}$ to satisfy. First, to ensure that each measurement sees

roughly the same incident power, we require that each mask contain $N/2$ ones and $N/2$ zeros. Second, we need reasonable incoherence with wavelets. Few results exist regarding the incoherence of arbitrary matrices, but since the wavelets are well-localized in space and frequency, we would do well to pick measurement atoms that are spread-out in both space and frequency. Third, it would be advantageous for each measurement to provide as much new information as possible. In the language of linear algebra, we wish the measurements to be orthogonal. Finally, because two-dimensional image processing often requires large N on the order of 10^5, storing the $N \times M$ matrix $\mathbf{H}$ in memory is often impossible. Thus, we require a matrix for which multiplication can be implemented as a fast transform.

For these reasons, we do not use completely random matrices as above. Instead, we choose to construct $\mathbf{H}$ by choosing random columns of a Hadamard matrix, which have all of the above properties, with the possible exception of incoherence. It has been reported that some groups have found Hadamard matrices useful for related problems [24]. A Hadamard matrix of size $N = 32$ is shown in Figure 4.5. An attractive feature of the Hadamard transform is its fast $\mathcal{O}(N \log N)$ implementation. The wavelet transform is even faster, $\mathcal{O}(N)$. Indeed, we find that even with its fast implementation, the Hadamard transform is our computational bottleneck, using over 90% of the computational resources of the inversion, so it is important to choose an $\mathbf{H}$ with an efficient implementation.
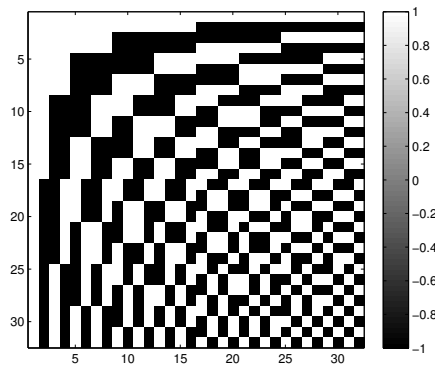
Figure 4.5: A 32×32 Hadamard matrix.

One problem with the Hadamard matrix is that it contains negative entries, which are not physically realizable with the proposed architecture. How-

ever, there exists a simple transformation that allows us to utilize the fast Hadamard transform. Simply put, we replace all instances of -1 with 0. Let us denote the $(-1, 1)$ Hadamard matrix $\tilde{\mathbf{H}}$ and the $(0, 1)$ Hadamard matrix $\mathbf{H}$. We acquire measurements, $\mathbf{g}$, using $\mathbf{H}$. These measurements are not orthogonal and they do not allow us to use the fast implementation. We can correct this by the simple transformation:

$$\tilde{g}_j = 2g_j - F \tag{4.7}$$

where g_j denotes the jth element of $\mathbf{g}$, and:.

$$F = \sum_{i=0}^{N-1} f_i \tag{4.8}$$

We show why this transformation is effective by expanding the right-hand side:

$$\tilde{g}_j = 2\langle \mathbf{h}_j\,,\,\mathbf{f}\rangle - \langle \mathbf{1}\,,\,\mathbf{f}\rangle$$
$$\tilde{g}_j = \langle 2\,\mathbf{h}_j - \mathbf{1}\,,\,\mathbf{f}\rangle \tag{4.9}$$

where the brackets denote an inner product and $\mathbf{1}$ is a vector of ones. We recognize that the vector $2\,\mathbf{h}_j - \mathbf{1}$ is simply the vector $\mathbf{h}_j$ where 0 has been replaced with -1. In other words, it is $\tilde{\mathbf{h}}_j$, so:

$$\tilde{g} = \langle \tilde{\mathbf{h}}_j\,,\,\mathbf{f}\rangle \tag{4.10}$$

which shows that our transformed measurements are related to the image through the $(1, -1)$ Hadamard matrix. By transforming our measurements with (4.7), it is as if we acquired them using $\tilde{\mathbf{H}}$, which has a fast transform. Recognizing that the vector $\mathbf{1}$ is simply the first column of $\mathbf{H}$, we can write this transformation succinctly in matrix notation as:

$$\tilde{\mathbf{g}} = \mathbf{Tg} \tag{4.11}$$

and

$$\tilde{\mathbf{H}} = \mathbf{TH} \tag{4.12}$$

where the transformation matrix $\mathbf{T}$ is:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 2 & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 2 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & 2 & 0 \\ -1 & 0 & 0 & 0 & \dots & 0 & 2 \end{bmatrix} \tag{4.13}$$

To summarize using matrix notation, the actual measurements use the $(0,1)$ Hadamard matrix and obey:

$$\mathbf{g} = \mathbf{H}\mathbf{f} \tag{4.14}$$

where, for simplicity, we used the Hadamard matrix property that $\mathbf{H}^* = \mathbf{H}$. To obtain the transformed measurements, we multiply both sides by $\mathbf{T}$:

$$\begin{aligned} \mathbf{T}\mathbf{g} &= \mathbf{T}\mathbf{H}\mathbf{f} \\ \tilde{\mathbf{g}} &= \tilde{\mathbf{H}}\mathbf{f} \end{aligned} \tag{4.15}$$

which gives us the formulation that we can use in compressed sensing algorithms.

We simulate the actual measurements, $\mathbf{g}$, and add noise. We assume that Poisson noise dominates over read noise and dark current. With photon counts on the order of 10^4, we can approximate the Poisson distribution as Gaussian. Therefore, to the measurement g_j is added a Gaussian random variable with mean zero and variance g_j. We assume that the noise in $\mathbf{g}$ is uncorrelated.

To recover the image, we transform to $\tilde{\mathbf{g}}$ and solve (4.3), the Basis Pursuit Denoise problem. For two-dimensional problems, it is important to use fast algorithms. The fastest we were able to find as of yet is Orthogonal Matching Pursuit (OMP) [47]. We assume the magnitude of the noise, ϵ, is known exactly. We use the 2-D Daubechies wavelets for the sparsity basis, $\mathbf{\Psi}$, but have found similar results using other wavelets.

4.3.2 Preliminary Results

The remaining design parameter is to choose the number of measurements, M. Here, we show results for different values of what we call the "sampling factor":

$$\alpha = \frac{M}{N} \tag{4.16}$$

If there were no noise and we chose $\alpha = 1$, we would have a square system of equations, and the reconstruction could be achieved with a simple matrix inversion. However, we are always limited by noise, and the sparsity of our image indicates that choosing such a large α is likely a waste of measurements. As we saw above, to a first approximation, choosing $\alpha = 0.5$ will match the temporal resolution of a CCD.
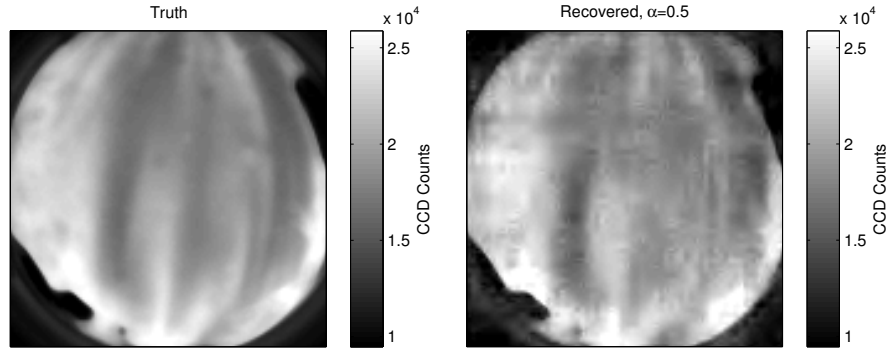

Figure 4.6: (Left) The original truth image. (Right) The recovered image using a sampling factor $\alpha = 0.5$.

Figure 4.6 shows an example reconstruction using $\alpha = 0.5$. The left panel shows the true image, and the right panel shows the recovered image. While the overall structure is discernible, we see small-scale artifacts, some of which are clearly due to single wavelets. We define the relative reconstruction error between the true image, $\mathbf{f}$, and the reconstructed image, $\hat{\mathbf{f}}$, as:

$$\text{relative error } (\%) = 100 \, \frac{\|\mathbf{f} - \hat{\mathbf{f}}\|_2}{\|\mathbf{f}\|_2} \tag{4.17}$$

The inversion shown in Figure 4.6 has a relative error of 3.1%. This can vary based on the particular noise realization. Despite the artifacts, it is promising that such a small error is possible with a problem that is both severely underdetermined and also noisy. Indeed, Poisson noise creates a relative error in the measurements, $\tilde{\mathbf{g}}$, of 20%. However, the level of error in this reconstruction is likely not low enough to support detailed airglow measurements such as drift velocity, although it may be good enough to study simple features such as bubble occurrence. A more detailed analysis of the usability of these images is let for future work.
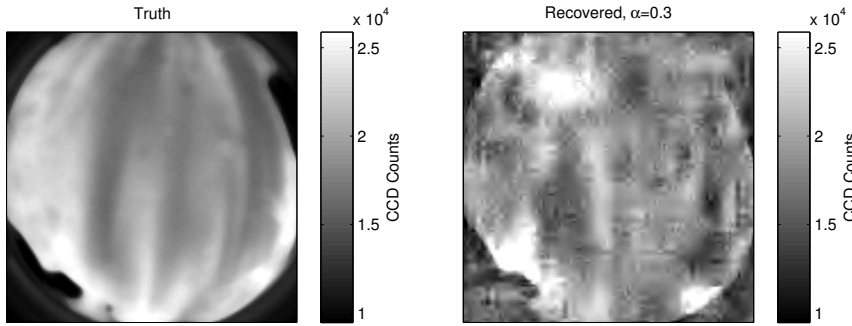

Figure 4.7: (Left) The original truth image. (Right) The recovered image using a sampling factor $\alpha = 0.3$.

Inversions with much smaller α typically fail to reconstruct important signal features. Shown in Figure 4.7 is such an example where $\alpha = 0.3$ and the relative error is 11.6%. The outline of the aperture is discernible, but there are virtually no distinguishable signal features. It appears that many of the initial wavelet atoms are being wasted in faithfully reconstructing the outline of the image, suggesting that an alternative, specialized inversion algorithm may be worth developing.

Inversions with larger α recover almost all of the signal features. Such an example is shown in Figure 4.8, for which $\alpha = 0.7$ and the relative reconstruction error is 1.1%. Although some small-scale artifacts are still evident, the reconstruction appears suitable for most of the typical airglow analyses such as bubble drift velocity, growth, and occurrence rate.

In order to better understand the relationship between α and the reconstruction error, we run many inversions for different values of α between 0.2
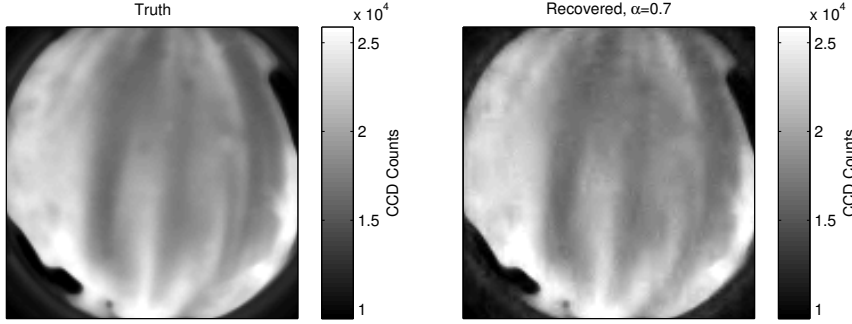
Figure 4.8: (Left) The original truth image. (Right) The recovered image using a sampling factor $\alpha = 0.7$.
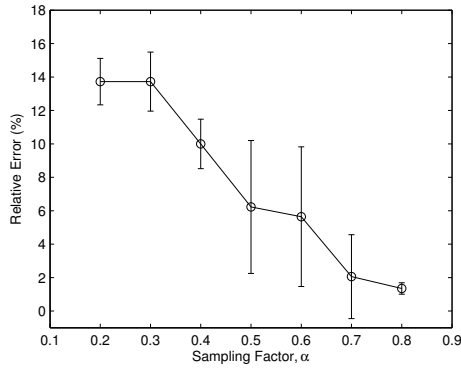

Figure 4.9: A statistical analysis of airglow inversions for different sampling factors. The circles represent the average reconstruction error, and the vertical bars indicate ± 1 standard deviation.

and 0.8. For each value, we run 10 simulations with random realizations of noise and randomly chosen columns of the Hadamard matrix, and we report the statistics of the relative reconstruction error. In Figure 4.9, we show the mean and standard deviation of the relative error for each value of α. The circle indicates the mean and the error bar shows plus and minus one standard deviation. For low α, the signal is not reconstructed well, and the relative error is consistently above 10%. At $\alpha = 0.4$, the error begins decreasing but is still around 10%. Sampling factors of 0.5 and 0.6 exhibit a wide range of relative errors. Closer examination reveals that there appears to be a bimodal distribution, with a cluster of low errors around $2 - 3\%$,

and a cluster around 10%. The cause of this clustering is uncertain but may indicate that randomly choosing columns of $\mathbf{H}$ is not optimal, and we may find that certain sets of columns work better than others. At $\alpha = 0.7$, all except one of the ten reconstructions are around 2% error. At $\alpha = 0.8$, the reconstructions are consistently around 1.5%.

4.4   Future Work

The inversions above show promise for eliminating the CCD in airglow imagers, which would reduce cost and allow imaging at infrared wavelengths. However, a number of simplifying assumptions were made to achieve these preliminary results. For one, we assumed no read noise or dark current. A more realistic forward model will have to include these effects for a realistic photodiode. Moreover, we must carefully account for losses associated with lenses and DMD reflections.

It was also assumed that the error magnitude, ϵ, was known exactly. In a realistic system, we will have to estimate this in a bootstrapped manner from the data, or use some *a priori* estimate. Another option is to treat ϵ as a general regularization parameter and use any number of ad hoc methods for regularization parameter selection such as Generalized Cross Validation or L-curve regularization, but these can be computationally expensive.

Another simplifying assumption was the removal of stars from the truth image. A real image has stars which can contribute to a large portion of the signal energy. In designing our algorithm, we will have to be careful to use the available signal energy to reconstruct airglow structure instead of wasting energy reconstructing minutiae of the star field. One possible solution would be to use a hybrid sparsity basis comprising wavelets and delta functions, to reconstruct airglow and stars, respectively. This hybrid basis is known as an overcomplete dictionary. This procedure could have the added benefit of improving the reliability of spatial calibration and star removal, since the star locations are effectively determined during the reconstruction. In general, the correct choice for the sparsity basis, $\boldsymbol{\Psi}$, is still unknown and finding general-purpose bases is a topic of current research. A nonlinear extension such as Total Variation (TV) minimization could be employed, if a robust and fast algorithm is found.

In addition to $\mathbf{\Psi}$, the optimal choice for $\mathbf{H}$ is unknown. Although attractive for certain reasons discussed above, the Hadamard basis may not be the best choice. It has been proven to be maximally incoherent with the Dirac basis [16], but the nature of its incoherence with wavelets is not known. By looking at Figure 4.5, we see that while no basis functions are localized in space, some appear to be localized in frequency, which may suggest some degree of coherence with the wavelets. Using a more incoherent measurement basis would allow for more accurate images to be reconstructed with fewer measurements. Ideally, we would like to find a measurement basis that is both incoherent with our sparsity basis and implementable with a fast transform. Additionally, the physical implementation of $\mathbf{H}$ could be changed. Instead of reflecting off of a DMD, we might use a MEMS-based shutter array in the image plane or aperture plane, a technique related to aperture coding. This may reduce reflective losses associated with the DMD mirrors.

The running time of these inversions is a problem. Even for the 4×4-pixel reduced resolution version shown here, an image that only takes 90 seconds to acquire takes 10-30 minutes to reconstruct on an Intel Core 2 Duo processor. This suggests a full-resolution inversion time of several hours. We would like to improve the running time by finding more efficient algorithms or developing specialized methods. One possibility is to break the image into blocks, acquiring and reconstructing each block separately in a manner similar to the original JPEG standard.

The measurement transformation $\mathbf{T}$ allows us to use the fast transform, but it also creates correlated errors. The implicit assumption in (4.3) is that errors are uncorrelated between measurements. This effect is probably small, but it should be accounted for.

Another assumption is that the image is stationary. A real airglow image has temporal structure. It will be necessary to model this structure and simulate our instrument in time in order to determine if the single-pixel camera is a viable option for motion capture of typical airglow images. This will create a lower limit on the temporal resolution and, consequently, an upper limit on the spatial resolution. The limits of and tradeoffs between temporal and spatial resolution will be important to estimate accurately before attempting to realize this device. A possible extension would be to treat the imaging problem as three-dimensional, with two spatial and one temporal dimension. This would allow us to leverage temporal as well as spatial sparsity,

thereby improving the quality of the reconstruction. However, the present simulations suggest that this idea is not computationally feasible with current technology.

CHAPTER 5

CONCLUSION

As we have shown, compressed sensing ideas are promising for ionospheric imaging applications. These ideas take advantage of the fact that many natural images are inherently sparse when expressed in certain bases, a fact that underlies the widespread success of image compression techniques such as JPEG. This inherent sparsity reduces the number of necessary measurements in an imaging context, and we reviewed surprising theoretical results pertaining to this recovery in Chapter 2. In this chapter, we also described the general electromagnetic imaging problem and how compressed sensing can be used to solve it.

Chapters 3 and 4 apply these ideas to radar imaging and airglow imaging. For radar imaging, compressed sensing offers us a better way to use our undersampled data, suggesting a sparsity-based reconstruction technique. We showed simulation results in which this technique returns better reconstructions than state-of-the-art methods. We also showed results of inverting data collected at the Jicamarca Radio Observatory during equatorial spread-F conditions. For airglow imaging, compressed sensing offers us a new way to design instrumentation to minimize the number of redundant measurements, effectively transferring complexity from expensive sensors to low-cost computation. The proposed instrument is the single-pixel camera, which directly gathers compressed data, combining the steps of sensing and compression. We showed preliminary simulation results that suggest the single-pixel camera may be feasible for airglow imaging, allowing us to develop low-cost sensors for much-needed global distribution and also image at infrared wavelengths. However, much work remains to be done before designing a prototype.

REFERENCES

[1] H. G. Booker and H. W. Wells, "Scattering of radio waves by the f-region of the ionosphere," *Terrestrial Magnetism and Atmospheric Electricity*, vol. 43, no. 3, pp. 249–256, 1938. [Online]. Available: http://dx.doi.org/10.1029/TE043i003p00249

[2] D. L. Hysell, "An overview and synthesis of plasma irregularities in equatorial spread F," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, no. 12, pp. 1037–1056, Aug. 2000. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S136468260000095X

[3] J. J. Makela, "A review of imaging low-latitude ionospheric irregularity processes," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 68, no. 13, pp. 1441–1458, Sep. 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S136468260600109X

[4] M. C. Kelley, J. J. Makela, O. de La Beaujardière, and J. Retterer, "Convective Ionospheric Storms: A Review," *Reviews of Geophysics*, vol. 49, no. 2, p. RG2003, June 2011. [Online]. Available: http://www.agu.org/pubs/crossref/2011/2010RG000340.shtml

[5] "The National Space Weather Program Strategic Plan," National Space Weather Program Council, Tech. Rep. June, 2010. [Online]. Available: http://www.ofcm.gov/nswp-sp/fcm-p30.htm

[6] R. Blahut, *Theory of Remote Image Formation*. Cambridge University Press, 2004. [Online]. Available: http://books.google.com/books?id=d8FMlHewYp0C

[7] A. Thompson, J. Moran, and J. George W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons, 2008. [Online]. Available: http://books.google.com/books?id=AwBN5bpuEU0C

[8] R. F. Woodman, "Coherent radar imaging: Signal processing and statistical properties," *Radio Science*, vol. 32, no. 6, p. 2373, 1997. [Online]. Available: http://www.agu.org/pubs/crossref/1997/97RS02017.shtml

[9] Z. Feng, E. Kudeki, R. Woodman, J. Chau, and M. Milla, "F region plasma density estimation at Jicamarca using the complex cross-correlation of orthogonal polarized backscatter fields," *Radio science*, vol. 39, no. 3, p. RS3015, 2004. [Online]. Available: http://www.agu.org/pubs/crossref/2004/2003RS002963.shtml

[10] E. Jaynes, "The Relation of Bayesian and Maximum Entropy Methods," in *Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, 1988, pp. 25–29.

[11] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, 2006. [Online]. Available: http://dialnet.unirioja.es/servlet/dcart?info=link\&codigo=3140640\&orden=237292

[12] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1614066

[13] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/cpa.20124/abstract

[14] E. Candès and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, no. March 2008, pp. 21–30, 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4472240

[15] J. Romberg, "Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, Mar. 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4472239

[16] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, June 2007. [Online]. Available: http://stacks.iop.org/0266-5611/23/i=3/a=008?key=crossref.bdb1d0e3390f5616a4395b50273198a9

[17] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1063520308000638

[18] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, pp. 417–426, 1974. [Online]. Available: http://adsabs.harvard.edu/full/1974A\%26AS...15..417H

[19] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, Jan. 2010. [Online]. Available: http://www.stormingmedia.us/ 01/0158/A015825.html

[20] E. van den Berg and M. P. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, Jan. 2009. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/080714488

[21] S. Wright, R. Nowak, and M. Figueiredo, "Sparse Reconstruction by Separable Approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, July 2009. [Online]. Available: http: //ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4799134

[22] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A Fast and Accurate First-Order Method for Sparse Recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, Jan. 2011. [Online]. Available: http://epubs.siam.org/doi/pdf/10.1137/090756855

[23] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs\ _all.jsp?arnumber=5490364 pp. 257–260.

[24] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Proc. of Computational Imaging IV at SPIE Electronic Imaging*, C. A. Bouman, E. L. Miller, and I. Pollak, Eds., Feb. 2006. [Online]. Available: http://proceedings.asmedigitalcollection.asme.org/ proceeding.aspx?articleid=728899 pp. 509–10.

[25] R. F. Woodman, "Inclination of the geomagnetic field measured by an incoherent scatter technique," *Journal of Geophysical Research*, vol. 76, no. 1, p. 178, 1971. [Online]. Available: http://www.agu.org/pubs/ crossref/1971/JA076i001p00178.shtml

[26] D. T. Farley, H. M. Ierkic, and B. G. Fejer, "Radar interferometry: A new technique for studying plasma turbulence in the ionosphere," *Journal of Geophysical Research*, vol. 86, no. A3, p. 1467, 1981. [Online]. Available: http://www.agu.org/pubs/crossref/1981/JA086iA03p01467. shtml

[27] E. Kudeki and F. Sürücü, "Radar interferometric imaging of field-aligned plasma irregularities in the equatorial electrojet," *Geophysical Research Letters*, vol. 18, no. 1, p. 41, 1991. [Online]. Available: http://www.agu.org/pubs/crossref/1991/90GL02603.shtml

[28] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1449208

[29] R. D. Palmer, S. Gopalam, T.-Y. Yu, and S. Fukao, "Coherent radar imaging using Capon's method," *Radio Science*, vol. 33, no. 6, p. 1585, Nov. 1998. [Online]. Available: http://www.agu.org/pubs/crossref/1998/98RS02200.shtml

[30] J. Ables, "Maximum entropy spectral analysis," *Astronomy and Astrophysics Supplement Series*, 1974. [Online]. Available: http://adsabs.harvard.edu/full/1974A\%26AS...15..383A

[31] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy." *Journal of the Optical Society of America*, vol. 62, no. 4, pp. 511–8, Apr. 1972. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/5019595

[32] S. F. Gull and G. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, 1978. [Online]. Available: http://www.nature.com/nature/journal/v272/n5655/abs/272686a0.html

[33] E. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939–952, 1982. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1456693

[34] T. J. Cornwell and K. Evans, "A simple maximum entropy deconvolution algorithm," *Astronomy and Astrophysics*, vol. 143, pp. 77–83, 1985. [Online]. Available: http://adsabs.harvard.edu/full/1985A\&A...143...77C

[35] D. L. Hysell, "Radar imaging of equatorial F region irregularities with maximum entropy interferometry," *Radio Science*, vol. 31, no. 6, p. 1567, 1996. [Online]. Available: http://www.agu.org/pubs/crossref/1996/96RS02334.shtml

[36] D. L. Hysell and J. L. Chau, "Optimal aperture synthesis radar imaging," *Radio Science*, vol. 41, no. 2, p. RS2003, 2006. [Online]. Available: http://www.agu.org/pubs/crossref/2006/2005RS003383.shtml

[37] R. Wilczek and S. Drapatz, "A high accuracy algorithm for maximum entropy image restoration in the case of small data sets," *Astronomy and Astrophysics*, vol. 142, pp. 9–12, 1985. [Online]. Available: http://adsabs.harvard.edu/full/1985A\&A...142....9W

[38] T.-Y. Yu, R. D. Palmer, and D. L. Hysell, "A simulation study of coherent radar imaging," *Radio Science*, vol. 35, no. 5, p. 1129, Sep. 2000. [Online]. Available: http://www.agu.org/pubs/crossref/2000/1999RS002236.shtml

[39] J. L. Chau and R. F. Woodman, "Three-dimensional coherent radar imaging at Jicamarca: comparison of different inversion techniques," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 63, no. 2-3, pp. 253–261, Jan. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364682600001425

[40] D. T. Farley, "A theory of electrostatic fields in the ionosphere at nonpolar geomagnetic latitudes," *Journal of Geophysical Research*, vol. 65, no. 3, p. 869, 1960. [Online]. Available: http://www.agu.org/pubs/crossref/1960/JZ065i003p00869.shtml

[41] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[42] M. Duarte, M. Davenport, D. Takhar, J. Laska, K. Kelly, and R. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4472247

[43] P. D. O'Grady and S. T. Rickard, "Compressive sampling of non-negative signals," in *2008 IEEE Workshop on Machine Learning for Signal Processing*. IEEE, Oct. 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4685468 pp. 133–138.

[44] J. J. Makela, "Midlatitude Ionospheric Studies Using the Global Positioning System and Airglow Cameras," Ph.D. dissertation, Cornell University, 2003.

[45] E. S. Miller, J. J. Makela, K. M. Groves, M. C. Kelley, and R. T. Tsunoda, "Coordinated study of coherent radar backscatter and optical airglow depletions in the central Pacific," *Journal of Geophysical Research*, vol. 115, no. A6, p. A06307, June 2010. [Online]. Available: http://www.agu.org/pubs/crossref/2010/2009JA014946.shtml

[46] M. Taylor, "A review of advances in imaging techniques for measuring short period gravity waves in the mesosphere and lower thermosphere," *Advances in Space Research*, vol. 19, no. 4, pp. 667–676, Jan. 1997. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0273117797001610

[47] S. Becker, "Orthogonal matching pursuit for sparse recovery," Dec. 2012, http://www.mathworks.com/matlabcentral/fileexchange/32402-cosamp-and-omp-for-sparse-recovery.