

THEORY AND APPLICATION OF SPARSITY FOR RADAR SENSING
OF IONOSPHERIC PLASMA

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
AERONAUTICS AND ASTRONAUTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ryan Volz
December 2014

© 2015 by Ryan Volz. All Rights Reserved.



This work is licensed under a Creative Commons Attribution 4.0 International License.
<http://creativecommons.org/licenses/by/4.0/>

This dissertation is online at: <http://purl.stanford.edu/kw265jg4383>

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sigrid Close, Principal Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Philip J. Erickson

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Mykel Kochenderfer

Approved for the Stanford University Committee on Graduate Studies

Patricia J. Gumpert, Vice Provost for Graduate Education

Abstract

The ionosphere plays an important role in the geospace system, perturbing and disrupting radio signals that pass through it and coupling the Earth with its space environment. Of particular note to spacecraft designers and operators are the small meteoroids that routinely impact spacecraft and are observable due to their passage through the ionosphere. Mechanical and electrical damage due to hypervelocity impacts with meteoroids pose a significant threat to spacecraft, and to fully characterize the threat it is necessary to observe the sporadic meteoroids through the meteor plasma that they create in the ionosphere. These observations, in addition to those of ionospheric plasmas that disrupt communications, can be made with high-power large-aperture radars located across the globe. Unfortunately, current radar techniques are not sufficient for elucidating the details of these plasmas due to inherent signal processing artifacts.

Sparsity has recently gained prominence in many scientific fields as a means of speeding, simplifying, or increasing the accuracy of information processing. Often this research falls under the moniker of *compressed sensing*, which exploits sparsity, global measurement, and nonlinear reconstruction to acquire data faster or with higher resolution. Mathematically, signals like images, sound waves, or radar returns can be represented with vectors. The vectors are paired with a dictionary describing atoms that, when combined according to the vector elements, can represent any given signal. Working with vectors gives a concrete definition of sparsity; it means that most of a vector's elements are zero. Natural phenomena tend to have structure, and that structure manifests itself as sparsity in the signal's vector representation with an appropriate dictionary. Knowing that natural signals are sparse allows one to constrain solutions to underdetermined measurements and uniquely find the signal's true representation.

In order to enable flexible high-resolution measurements of ionospheric plasma phenomena, a sparsity-based radar waveform inversion technique is formulated and found to eliminate processing artifacts caused by the standard matched filter approach. Taking direction from the theory of compressed sensing, sparsity of the radar target scene is employed as prior knowledge to successfully perform the inversion. The result is cleaner data that limits self-interference of range-spread targets and enables differentiation in crowded and variable environments. Though the approach has been applied to ionospheric radar, it is generally applicable and especially relevant for radar target scenes with multiple or distributed scatterers.

As a basis for the technique, a discrete radar model that captures signal sparsity in a delay-frequency dictionary is developed. This model is shown to have a strong connection to existing methods, resulting in an intuitive interpretation of the inversion technique as an iterative thresholding matched filter. An explicit formulation of the discrete model's representation of arbitrary distributed scatterers is derived, and it shows that sparsity is reasonably preserved in the discrete representation. Building on top of the model, waveform inversion is implemented using modern convex optimization techniques tailored for efficient computation and quick convergence. Finally, the real-world flexibility and effectiveness of the inversion technique is demonstrated by the elimination of filtering artifacts from meteor observations made with a variety of standard radar waveforms.

Acknowledgments

In terms of both scholarly and personal support, this dissertation would not have been possible without the help of many people. I would like to thank my thesis committee members, Professors Mykel Kochenderfer, Stephen Boyd, and Charbel Farhat, for providing many interesting discussions and taking the time to critique this work. I would also like to thank the researchers and staff of the Jicamarca Radio Observatory, particularly Dr. Jorge Chau and Karim Kuyeng, for hosting me and running my experiments. At Stanford, the Space Environment and Satellite Systems lab has been my home for most of my graduate studies, and I would like to thank its past and present members for making that home a good one. Thanks must also go to my classmates and friends in the Aero/Astro department, particularly my study group, Tom Coles, Brandon Fetroe, and Jason Forshaw, and my quals group, Brandon Morgan and Surajit Roy.

My first crash course in radar came on a two-week visit to the MIT Haystack Observatory to collect data and attend the AMISR workshop. I knew almost nothing about collecting and processing radar data at that point in time, but after those two weeks I knew enough to be dangerous. Credit and thanks go to the teachers at the workshop and the researchers at Haystack, most notably my co-advisor of sorts, Dr. Phil Erickson. On the first day of that visit, Phil sat me down in his office, gave me some Python code (with me never having used the language before), and let me go at it. He has answered my questions ever since, providing insight whenever I've been stuck on the thorny details of radar. I would like to thank Phil for being an outstanding teacher and mentor.

I was fortunate enough to be at the right place at the right time when I met my advisor, Professor Sigrid Close. Again, I knew nothing about radar or the ionosphere before that first meeting almost five years ago, and I had been beginning to wonder whether I would

be able to find a good fit for my interests at Stanford. Sigrid solved those problems by providing me with ALTAIR meteor data to analyze and giving me the freedom to explore. She was enthusiastic and supportive when I later came to her office with ambitious ideas about applying compressed sensing to radar, even if it meant that I never actually got around to analyzing that ALTAIR data. I would like to thank Sigrid for providing many work and travel opportunities and for guiding my scholarly explorations with the perfect blend of support and advice.

Finally, I would like to thank my friends and family for their encouragement, especially for not asking too many questions about when this thesis would finally be done. Most importantly, though, I want to thank my wife, Bekah, without whom none of this would have happened. It has been a long journey to get to this point, and I'm glad I could share the highs and lows with her.

*Ryan Volz
Stanford, California
December, 2014*

This research was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program, and by the National Science Foundation under Grant No. AGS-1056042.

Contents

Abstract	iv
Acknowledgments	vi
I Introduction	1
1 Introduction	2
1.1 The Ionosphere	3
1.1.1 Formation and Variability	4
1.1.2 Communications and Irregularities	5
1.1.3 System Science	7
1.1.4 Meteoroids and Meteors	7
1.2 Radar Measurements	9
1.2.1 Incoherent Scatter Radars	9
1.2.2 Examples of Ionospheric Radar Scatter	12
1.2.3 Challenges	14
1.3 Sparsity	16
1.3.1 Sparsity in Nature	16
1.3.2 Sparsity in Radar	18
1.3.3 Sparsity in Action	18
1.4 Contributions	19
1.5 Reader's Guide	21

II Radar Analysis	22
2 Radar Background	23
2.1 Pulse Design	24
2.1.1 Pulse Length	25
2.1.2 Pulse Waveforms	27
2.2 Matched Filtering	28
2.2.1 Autocorrelation and Range Sidelobes	30
2.2.2 Frequency Banks of Matched Filters	31
2.3 Measurement Ambiguity	32
2.3.1 Ambiguity Function	33
2.3.2 Sidelobes in Action	35
2.4 Current Techniques	38
2.4.1 Sidelobe Mitigation	38
2.4.2 Alternating Codes	38
2.4.3 Inverse Filters	39
2.4.4 Inversion	40
3 Radar Model	42
3.1 Defining the Model	42
3.1.1 Image Blur Analogy	42
3.1.2 Ambiguity Equation	44
3.1.3 The Radar Model	46
3.2 Derivation of Radar Model	48
3.2.1 Narrow-band Received Signal	48
3.2.2 Discretization	49
3.2.3 Simplification	50
3.2.4 Assumptions	52
3.3 Representation of Targets	52
3.3.1 Deterministic Reflectivity	52
3.3.2 Probabilistic Reflectivity	53
3.3.3 Reflectivity Coefficient Sparsity	54

III	Waveform Inversion	58
4	Sparse Solutions of Underdetermined Systems	59
4.1	Underdetermined Systems of Equation	60
4.1.1	Least Norm Solution	60
4.1.2	Sparsest Solution	61
4.2	Compressed Sensing	62
4.2.1	Conditions	63
4.2.2	Theoretical Bounds	64
4.2.3	Recovery Phase Transition	66
4.3	Convex Optimization	67
4.3.1	First-order Methods	68
4.3.2	Prox Operator Examples	69
4.3.3	Proximal Gradient	71
4.3.4	Improvements on Proximal Gradient	73
4.3.5	Related Prox Methods	74
4.3.6	Similar Convex Problems	79
5	Waveform Inversion Method	82
5.1	Related Work	83
5.1.1	Other Radar Applications	83
5.1.2	Theoretical Support	84
5.2	Solution Procedure	85
5.2.1	Regularization Parameter	85
5.2.2	Iterative Thresholding Matched Filter	87
5.2.3	Scaling and Noise	89
5.2.4	Solution Residual	90
5.3	Implementation	91
5.3.1	<code>radarmodel</code>	92
5.3.2	<code>prx</code>	93
5.3.3	Algorithm Comparison	94

6 Experimental Results	97
6.1 Jicamarca	98
6.1.1 Specifications	99
6.1.2 Array Segmentation	100
6.2 Setup	100
6.2.1 Waveforms	101
6.2.2 Pulse Parameters	102
6.2.3 Pulse Timing	102
6.2.4 Sampling	103
6.3 Procedure	103
6.4 Pseudorandom Pulse Results	103
6.4.1 Pulse Slices	104
6.4.2 Integrated RTI Plots	107
6.4.3 Frequency Slices	109
6.4.4 Sparse Solution and Residual	111
6.4.5 Detection Sensitivity	113
6.5 Waveform Comparison Results	113
6.5.1 Minimum Peak Sidelobe Code	113
6.5.2 Barker-13 Code	115
6.5.3 Uncoded Pulse	115
6.5.4 LFM Chirp	119
6.5.5 Inter-waveform Comparison	121
6.6 Undersampling Results	121
6.6.1 Discarding Half the Data	124
6.6.2 Recovering at Double Resolution	124
6.7 Recommendations	126
6.7.1 Waveform Selection	126
6.7.2 Frequency Analysis	129
6.7.3 Computation	130

IV Conclusion	131
7 Conclusion	132
7.1 Summary and Discussion of Contributions	133
7.1.1 Radar Model	133
7.1.2 Waveform Inversion Method	134
7.1.3 Experimental Results	136
7.2 Future Work	137
7.2.1 Incoherence and the Ambiguity Function	137
7.2.2 Super-resolution	137
7.2.3 Multi-pulse Analysis	138
7.2.4 Alternative Radar Models	139
7.2.5 Radar Self-calibration	139
7.2.6 Passive Radar	140
7.2.7 Near-universal Measurement Modes	140
V Appendix	142
A Reflectivity Coefficient Simplification	143

List of Tables

1.1	Incoherent scatter radar parameters	11
5.1	Convergence comparison of proximal optimization algorithms	95

List of Figures

1.1	Ionospheric plasma density during day and night	4
1.2	Scintillation of a radar calibration signal	6
1.3	Cumulative distribution of interplanetary bodies as a function of mass	8
1.4	Incoherent scatter radars spanning the globe	10
1.5	The equatorial ionosphere as measured by the Jicamarca ISR	13
1.6	Examples of meteor head and trail echoes	13
1.7	Meteor head echo oscillations	15
1.8	The structure of natural images	17
2.1	Transmitted and received signals for a short and long pulse	25
2.2	Waveform encoding examples	27
2.3	Autocorrelation of the Barker-13 code	29
2.4	Ambiguity function examples	34
2.5	Meteor head and trail measured with an uncoded pulse	36
2.6	Meteor head and trail measured with coded pulses	37
2.7	Comparison of matched and inverse filter ambiguity functions	40
3.1	Image blurring example	43
3.2	Reflectivity coefficients as a delay-frequency grid of point targets	48
3.3	Periodic sinc	53
3.4	Reflectivity coefficients for point targets	56
4.1	Phase transition for recovery of sparse solution	67
4.2	Soft thresholding function	70

6.1	The Jicamarca 50 MHz radar	98
6.2	Sequence of waveforms used in the Jicamarca experiment	101
6.3	Pulse slices comparing matched filter and waveform inversion	105
6.4	Pulse slice comparison for a meteor trail	106
6.5	RTI comparison of matched filter and waveform inversion	108
6.6	Frequency slices of waveform inversion	110
6.7	Decomposition of waveform inversion as sparse and residual solutions . .	112
6.8	Waveform inversion example for a different meteor	114
6.9	Waveform inversion with the minimum peak sidelobe code	116
6.10	Waveform inversion with the Barker-13 code	117
6.11	Waveform inversion with the uncoded pulse	118
6.12	Waveform inversion with the LFM chirp	120
6.13	Frequency-integrated waveform inversions of different waveforms	122
6.14	Zero-frequency waveform inversion slices of different waveforms	123
6.15	Waveform inversion with half the data	125
6.16	Waveform inversion at double the resolution	127
6.17	Waveform inversion at double the resolution, second example	128

Part I

Introduction

Chapter 1

Introduction

The fundamental mechanism of radar is to transmit a radio signal and listen for reflections to gain information about remote targets. Although the concept of radar was evident from the beginnings of electromagnetic theory, it was not until sufficient transmitter and receiver technology was developed around the time of World War II that practical radars came into use for remote detection of ships and aircraft [Skolnik, 2001]. Following the war, radar use branched out from military applications into scientific and commercial fields. Radar is still used with remote targets like aircraft and satellites for detection, tracking, and imaging, but it is more commonly encountered as a tool in weather or in highway speed limit enforcement [Melvin and Scheer, 2014]. Less known are the applications that drive the development of radar today, including autonomous landing systems for aircraft, driver-less cars, and terrain mapping (both surface and underground) of the Earth and other planets [Jol, 2009; Melvin and Scheer, 2012]. But of the many uses of radar, one is somewhat unique in its requirements: the study of ionospheric plasmas. These plasmas vary widely in size, density, and persistence, so their study has necessitated the development of transmission coding and signal processing techniques that are adaptable to a wide range of circumstances. Unfortunately, the current techniques produce inherent processing artifacts or are inflexible in their application, limiting our ability to learn more about some ionospheric plasmas [Virtanen, Lehtinen, and Vierinen, 2008].

The goal of this thesis is to develop new methods for radar data analysis that enable flexible, high-resolution measurements of ionospheric plasma phenomena. The key to achieving

this goal lies in leveraging the concept of sparsity; scenes typically observed by a radar have a simple representation relative to the space of possible signals, and this prior information can help to improve measurements and analysis. In the case of measurements, this thesis presents a waveform inversion technique for removing ambiguity and self-interference due to the transmitted signal. In the case of analysis, a clustering method is developed for automatically detecting and classifying related scattering events. For events with a single point scatterer, we simulate and analyze statistics of matched filter interpolation for estimating target range and range-rate with increased resolution. Though the ionosphere provides the motivation and application for this work, the research is primarily focused on techniques that are generally applicable to most radar scenarios, especially those with multiple or distributed scatterers.

The remainder of this chapter broadly covers the ionosphere, radar, and sparsity in order to motivate the research and set the stage for later developments. Section 1.1 gives background information on the ionosphere and describes its effects on radio communications, its role in studying the science of the geospace system, and its function in observing meteoroids that pose a threat to orbiting satellites. Section 1.2 describes the radars used to study the ionosphere, the types of scattering that are typically encountered, and the challenges they present. Section 1.3 details the function of sparsity in efficient information processing and its emerging role in many different fields of research. Finally, Section 1.4 outlines the contributions of this dissertation, and Section 1.5 provides a reader’s guide for the remaining chapters.

1.1 The Ionosphere

Surrounding the Earth is a region defined by ionization: the ionosphere. It is a portion of the upper atmosphere where charged particles, both electrons and ions, are present in addition to neutral air molecules. This region extends from approximately 60 km to 2000 km in altitude, a range which includes the low Earth orbits (LEO) inhabited by the majority of satellites. The ionized and neutral particles constitute a plasma, one that varies greatly over time and over the spatial extent of the ionosphere. Driving these variabilities are interactions with the lower atmosphere and the space environment, most notably radiation from the sun. As the

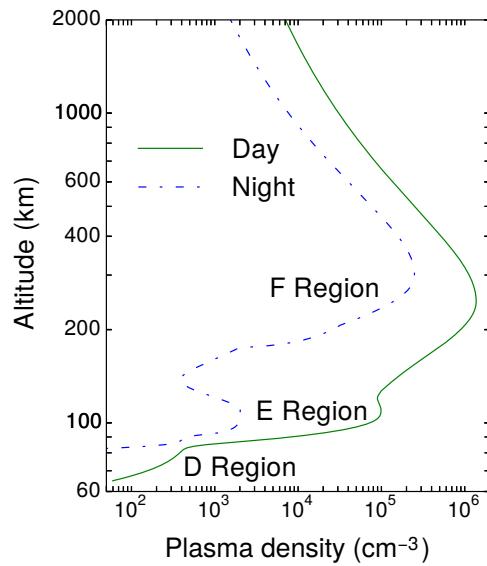


Figure 1.1 – Ionospheric plasma density during day and night. This plot shows predicted electron density at noon and midnight on December 1, 2014, according to the IRI-2012 model [Bilitza *et al.*, 2014]. The prediction is for the mid-latitude ionosphere above the Millstone Hill incoherent scatter radar.

interface between space and the Earth’s atmosphere, the ionosphere is difficult to study but critically important for understanding the geospace system [CEDAR Steering Committee, 2011].

1.1.1 Formation and Variability

Ionization of neutral air molecules by solar radiation is the primary cause for the existence of the ionosphere, and it is balanced by ion-electron recombination chemistry. Recombination limits the maximum ionized plasma density during the day and dissipates the plasma at night. Naturally, the temporal peak in plasma density happens near noon when the ionization rate is at its highest, but even then the plasma density is at most around one percent of the neutral background. Example plasma densities for the daytime and nighttime mid-latitude ionosphere are shown in Figure 1.1. Solar radiation ionizes air molecules down to altitudes of approximately 60 km, forming the ionosphere’s lower boundary. Its upper boundary at approximately 2000 km altitude is more tenuously defined; it is not a cutoff of ionization but rather one of sufficient density to exhibit bulk plasma properties. Between

these boundaries, the ionosphere is divided by altitude into different regions defined by variations in plasma density. Arranged by increasing altitude, these layers are called the D, E, and F regions. The different regions exist because the ionization rate varies with altitude: solar intensity decreases with decreasing altitude as more radiation is absorbed, while at the same time the neutral atmosphere increases in density and varies in composition with decreasing altitude due to gravity [Tascione, 2010; Kelley, 2009].

Though this simple model is useful to conceptualize the ionosphere, it is important to note that the ionosphere is part of a complex geospace system and therefore exhibits significant variability over time and space. Day/night variability has already been mentioned, but tidal, seasonal, and solar activity cycles are also significant drivers of the ionosphere. Weather is an important contributor on shorter timescales, both in terms of atmospheric weather from the dynamics of the neutral atmosphere and space weather from the flares and particle precipitation of solar storms. In terms of spatial variability, the changing orientation of the Earth's magnetic field lines with respect to latitude divides ionospheric behavior into three regions: equatorial, polar, and mid-latitude. In the equatorial region, the magnetic field lines are horizontal. This results in unique plasma dynamics that include the equatorial electrojet current sheet, turbulence-induced field-aligned irregularities, and equatorial bubbles [Kelley, 2009]. In the polar region, the magnetic field lines connect to the magnetosphere and solar wind and bring a supplemental source of energetic particles. The result is enhanced ionization and light displays known as aurora, among other phenomena [Kivelson and Russell, 1995]. The mid-latitude region lies between the other two and is defined through the absence of equatorial and polar effects. Boundary interactions are characteristic of this region, including the plasmasphere boundary layer of the sub-auroral zone [Carpenter and Lemaire, 2004] and notable disturbances like sudden stratospheric warming [Goncharenko *et al.*, 2013].

1.1.2 Communications and Irregularities

The ionosphere is felt most strongly in daily life through its effects on radio communications. Radio waves of lower frequency (less than about 13 MHz for a typical F-region peak plasma density [Tascione, 2010]) reflect off of the ionosphere. In practical terms,

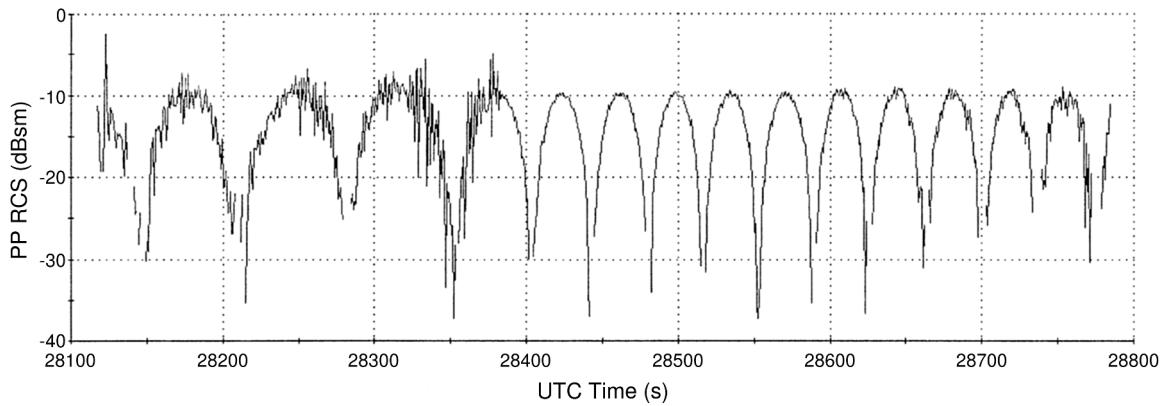


Figure 1.2 – Scintillation of a radar calibration signal. This plot shows the radar cross section (RCS) measured for the principal polarization (PP) from a calibration test with the ALTAIR radar at 160 MHz. The radar beam was swept back and forth across a fixed target, so the signal strength is expected to be consistent with each pass. Scintillation is evident on the first four and a half passes, while the rest are nearly scintillation-free.

AM radio broadcasts reflect (when they are not absorbed), while the higher-frequency FM radio signals pass through the ionosphere into space. However, even the higher microwave-frequency signals that are typically used for satellite communications are subject to attenuation, refraction, and delay [Kelley, 2009]. This is especially important to consider in the case of the global navigation satellite systems (GPS, GLONASS, Galileo); these depend on precise timing and require correction for ionospheric effects in order to achieve high positioning accuracy.

Ionospheric irregularities are even more important to characterize, due to their unpredictability and potential to completely disrupt typical satellite communications. Random variation of signal strength and phase, termed scintillation, occurs whenever there are abrupt changes in electron density along the signal path. These irregularities cause scattering and self-interference of a radio signal, a process analogous to looking through rough glass. The result is corrupted communication or even loss of signal. Figure 1.2 depicts the measured signal for a radar beam sweeping back and forth across a calibration target. The first four passes exhibit scintillation as strong variations in signal strength, while the remaining passes are nearly scintillation-free. Though categorized as an irregularity, scintillation is an everyday occurrence arising from typical variation of the ionosphere. Geomagnetic storms, in contrast, are truly irregular occurrences arising from solar activity. They disrupt the

ionosphere by increasing direct energy input and transport effects, thus triggering further irregularities and scintillation [Kelley, 2009].

1.1.3 System Science

Given how much the world depends on satellite communications and radio systems, studying the ionosphere in order to predict and mitigate disruptions is clearly important. In order to do that, though, it is necessary to examine the ionosphere's role in the overall geospace system. This system is formed of complex interactions between the Earth, with its atmosphere and magnetic field, and the space environment, with the dominating influence of the Sun. Scientific study of the geospace system in the past has been limited by the inability to investigate its complexity, eloquently described by the *CEDAR Steering Committee* [2011] as the “many interacting elements that, as a whole, exhibit properties not obvious from the properties of the individual parts.” Nevertheless, this so-called systems science is becoming more feasible every day with the advancement of computational and measurement techniques. Since the ionosphere is a key cog in the geospace system, it is important to develop new methods for studying the ionosphere that embrace complexity and enable improved understanding of the system as a whole. This requires measurements that are universal, encompassing large portions of the ionosphere, and it requires techniques that are flexible, measuring plasmas spanning a wide range of size, density, and persistence.

1.1.4 Meteoroids and Meteors

Of the many components of the space environment that interact with the ionosphere, meteoroids warrant special mention. *Meteoroid* is a term that encapsulates any small, naturally occurring solid body in space. Meteoroids range in size from 10 micrometers to approximately 1 meter in diameter; smaller particles are called dust, and larger bodies are called asteroids [Rubin and Grossman, 2010]. While an asteroid crossing paths with the Earth is a rare event of global disaster proportions, meteoroids routinely enter the Earth's atmosphere. Furthermore, meteoroids are more common as size decreases; roughly speaking, their number are inversely proportional to the mass of the meteoroid squared as shown in Figure 1.3. As a point of reference, on the order of 100 million meteoroids big enough to be

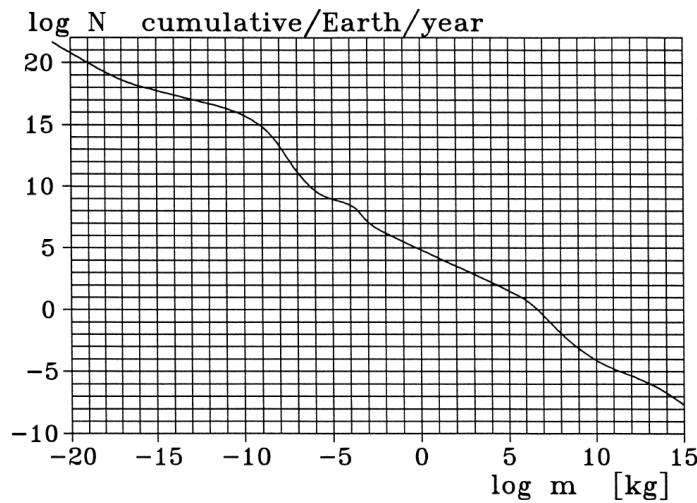


Figure 1.3 – Cumulative distribution of interplanetary bodies as a function of mass. This plot, reproduced from Cephecha *et al.* [1998], shows the cumulative count N of interplanetary bodies (including meteoroids and asteroids) with mass greater or equal to m that enter the Earth’s atmosphere per year. Both the count and mass are plotted as base-10 logarithms.

detected by high-power large aperture radars (0.1 micrograms or larger [Close *et al.*, 2007]) are estimated enter the Earth’s atmosphere every second on average [Cephecha *et al.*, 1998]; this is equivalent to approximately one meteoroid every second within a radar beam that is pointed upward with a one degree half-power beamwidth. If a meteoroid survives passage through the atmosphere and lands on the Earth, which is correspondingly more rare, the surviving body is called a *meteorite*.

The term *meteor* is not to be confused with meteoroid, although they are often incorrectly used interchangeably. A meteor denotes the plasma created by a meteoroid as it speeds through the atmosphere, typically forming between 80 and 120 km in altitude. Sometimes light is emitted when the meteoroid is large or fast enough, and this is also called a meteor or colloquially a shooting star. Meteoroids enter the atmosphere with such speed, typically between 11 and 72 km/s [Cephecha *et al.*, 1998], that they heat up, ablate their constituents, and cause some of the surrounding air to ionize. The resulting plasma is many times denser than the background ionospheric plasma, but it typically only retains that density for a fraction of a second before diffusing and recombining [Jones, 1995; Dyrud *et al.*, 2008]. The plasma dynamics spurred by meteors are significant to the overall ionospheric

system, while the ablated meteoroid elements play an important role in the chemistry of the upper atmosphere.

Aside from their role in driving ionospheric processes, meteoroids are also important because of the likelihood that they will impact satellites. The most prevalent meteoroids are small enough (< 1 microgram) that they can't do physical damage; however, meteoroids travel fast enough that even the smaller ones can cause electrical damage to satellites through the plasma that they create upon impact [Close *et al.*, 2010]. Lee *et al.* [2013] showed through ground experiments that this damage is possible and could explain nebulously attributed or otherwise unexplained anomalies and failures of a number of satellites. In order to better characterize the electrical threat to satellites, we need better methods of remotely measuring the meteoroid distributions of rate, size, speed, and composition; the best way to do that is by advancing radar techniques for observing meteors.

1.2 Radar Measurements

Due to its remoteness and size, spanning altitudes from 60 km to 2000 km, the ionosphere is difficult to study. Even the lowest reaches of the ionosphere are outside the range of aircraft, leaving balloons, rockets, and satellites as the only carriers for *in situ* measurement equipment. Not only are these methods expensive, but they also do not give enough data over time and space to perform more than localized studies. For large-scale observation of the ionosphere, scientists are left with methods of remote sensing, particularly radar.

1.2.1 Incoherent Scatter Radars

It takes a large radar to be sensitive enough to measure scattering from electrons in the background ionosphere; this type of signal is called incoherent scatter, and these large radars are incoherent scatter radars (ISRs) [Gordon, 1958]. There are only about a dozen ISRs in operation across the globe, with locations depicted in Figure 1.4a. Each radar covers a small slice of the ionosphere ranging from equatorial latitudes (e.g. Jicamarca outside Lima, Peru; Figure 1.4b) through mid-latitudes (e.g. Millstone Hill near Boston, MA; Figure 1.4d) to polar latitudes (e.g. the Poker Flat ISR in Alaska; Figure 1.4c). The main parameters of the

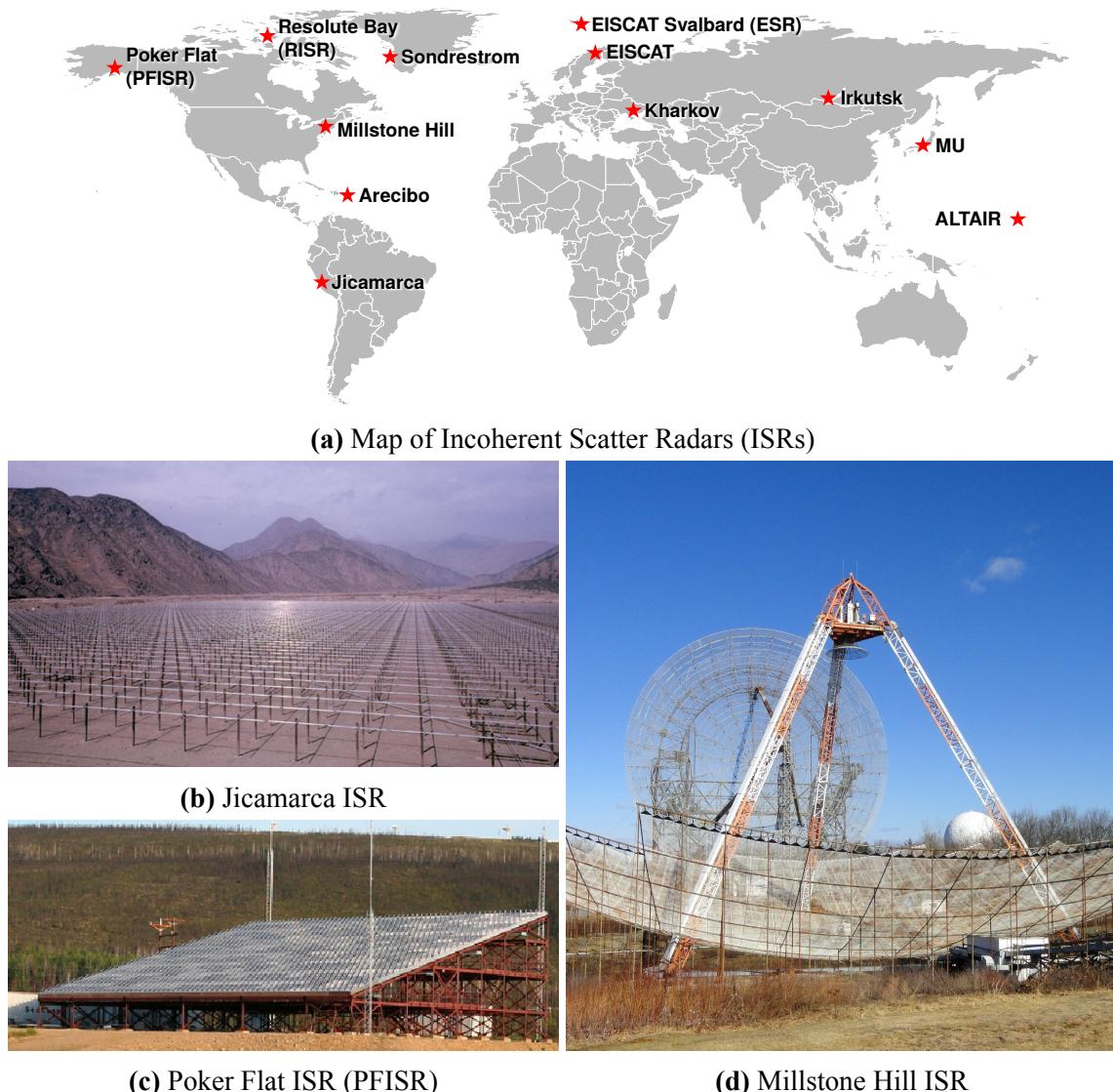


Figure 1.4 – Incoherent scatter radars spanning the globe. Locations of the current ISRs are shown in **(a)**. The Jicamarca ISR **(b)** is a phased array of dipole antennas located outside of Lima, Peru. The Poker Flat ISR **(c)** is a modern phased array located in Poker Flat, Alaska. The Millstone Hill ISR **(d)** is composed of one steerable dish antenna and one fixed zenith dish antenna and is located outside of Boston, Massachusetts.

Table 1.1 – Incoherent scatter radar parameters.

Radar	Frequency (MHz)	Bandwidth (MHz)	Peak Power (MW)	Gain (dBi)	Size (m)
ALTAIR ¹	160/422	7/18	6	34/42	46
Arecibo ²	430	1	2.5	62	305
EISCAT VHF ³	224	1	1.6	46	120 × 40
EISCAT UHF ³	931	1	2	48.1	32
ESR ³	500	1	1	42.5/45	32/42
Irkutsk ⁴	154	0.014	3.2	35	
Jicamarca ⁵	50	1	4.5	42.6	300 × 300
Kharkov ⁶	158	0.015	2	40	100
Millstone ⁷	440	1	2.5	42.5/45.5	46/68
MU ⁸	46.5	1.65	1	34	103
PFISR ⁹	450	1	2	43	30 × 30
RISR ¹⁰	442.5	1	2	43	30 × 30
Sondrestrom ¹¹	1290	1	3	50	32

[¹Tsunoda *et al.*, 1979; ¹Close, S. M. Hunt, *et al.*, 2002; ²Hagen, 2005; ³EISCAT Technical Specifications 2014; ⁴Potekhin *et al.*, 2008; ⁵Ochs, 1963; ⁵The Jicamarca Radio Observatory 2014; ⁶Burmaka, Taran, and Chernogor, 2006; ⁷Evans, 1965; ⁷MIT Haystack Observatory: ISR 2014; ⁸Fukao *et al.*, 1985; ⁹Chau *et al.*, 2009; ^{9,10}AMISR Specifications 2014; ¹¹Kelly *et al.*, 1995]

various ISRs are given in Table 1.1. Most of their characteristics are roughly similar, with megawatt-class transmitters, high gain, and large size. The baseband frequencies are all in the VHF and UHF bands, greater than 30 MHz and less than 3 GHz, which places them above the ionosphere’s peak plasma frequency and allows the electromagnetic waves to completely penetrate the plasma. The ISRs also come in one of two broad forms, with either focusing reflector dishes like Millstone Hill or distributed antenna arrays like Jicamarca and the Poker Flat ISR (PFISR). The dish radars are usually mechanically steerable which limits their view to one region of the sky within a short period of time, while the array radars are electronically steerable which allows the location of their electromagnetic sensitivity to be adjusted by changing the phase of the antenna connections. At older arrays like Jicamarca, this phasing is still done mechanically by hand, while newer arrays like PFISR are phased

by computer control and so can look at completely different portions of the sky from pulse to pulse [*Ochs*, 1963; *Chau et al.*, 2009].

1.2.2 Examples of Ionospheric Radar Scatter

In addition to the background ionospheric plasma, ISRs are also used to measure other plasma processes and irregularities which exhibit stronger signals. This thesis heavily utilizes data taken with the Jicamarca ISR, and one of the processes that is frequently encountered there is the equatorial electrojet. The electrojet arises from the unique geometry of the magnetic equator, where the magnetic field is parallel to the Earth's surface. This geometry, coupled with the existing east-west zonal current from E-region winds, creates a vertical electric field that greatly enhances the zonal current through $E \times B$ drift to form the electrojet [Kelley, 2009]. When the current of the electrojet is strong enough, it produces plasma density irregularities that coherently scatter a radar signal at many orders of magnitude greater than the strength of incoherent scatter. As shown in Figure 1.5, electrojet scatter is compact in range and resides near an altitude of 100 km. It also is relatively stationary, and so it produces no Doppler frequency shift to complicate the scattered signal. Though the electrojet is important in its own right, it is often seen as clutter when trying to measure other ionospheric processes.

Another ubiquitous class of scatterers in the equatorial ionosphere and globally are meteors. When viewed by a high-power large-aperture (HPLA) radar like the ISRs, meteors exhibit two types of scattering: head echoes and trail echoes. Both are shown in Figure 1.6. A head echo represents scattering from the dense ball of plasma traveling with the meteoroid [*Close et al.*, 2005]. In a radar signal intensity image plotted as range (or altitude) versus time (known as a range-time-intensity or RTI plot), head echoes appear as a narrow streak that descends in altitude with time. Because the meteoroid and head plasma are moving at high speed toward the radar as they descend in altitude, head echoes have a large positive Doppler frequency shift that complicates their measurement. Observing the head echo's speed, either by change in altitude or through the Doppler shift, allows one to determine the meteoroid's speed. If angular speed measurements are available through interferometry or monopulse, the full velocity and hence orbit of the meteoroid can be calculated. In addition,

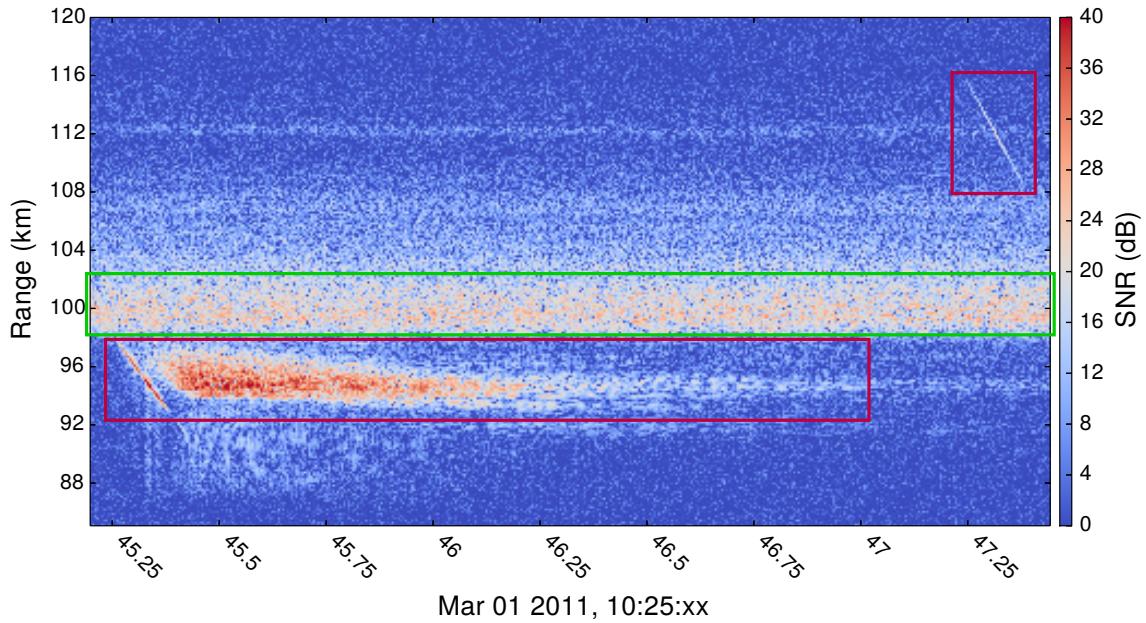


Figure 1.5 – The equatorial ionosphere as measured by the Jicamarca ISR. Scattering due to the **equatorial electrojet** is seen in a narrow band around 100 km range (altitude). The other significant scattering is due to two **meteors**.

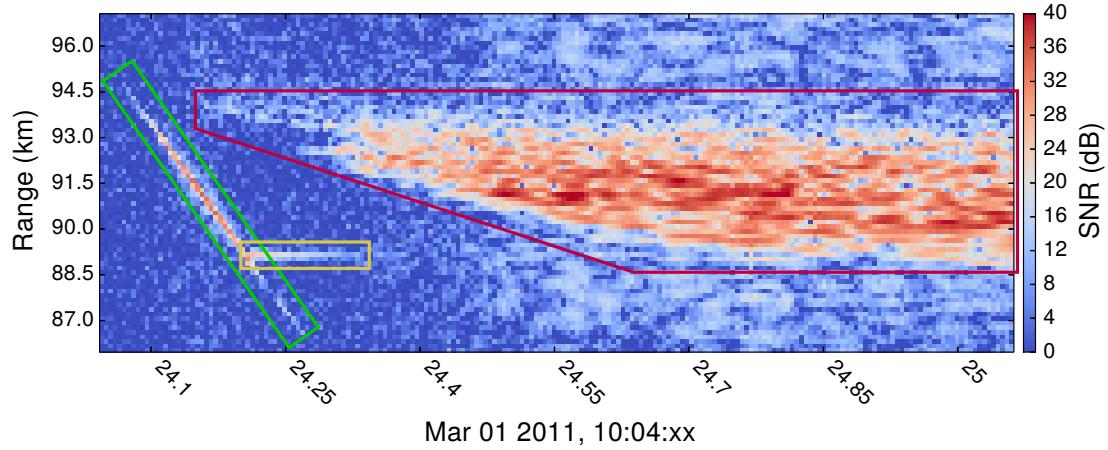


Figure 1.6 – Examples of meteor head and trail echoes. Two types of meteor scattering are observed from the same meteor by the Jicamarca ISR. The **head echo** is the narrow streak on the left, while the **non-specular trail echo** is the range-spread scattering on the right. The head echo also exhibits a **flare** with a sudden increase in signal strength and longer-lasting scattering at about 89 km range (altitude).

the signal strength and deceleration of the head echo can be used to determine the mass and density of the parent meteoroid [Close *et al.*, 2012].

Once the meteor plasma is created by the meteoroid, it diffuses quickly into a trail that is left behind in the meteoroid's wake. Scattering from the trail can then happen either specularly or non-specularly. For specular scattering, the line of trail plasma acts like a mirror to reflect the radar signal [Sugar, 1964]; with a monostatic radar system where the transmitting and receiving antenna are one in the same, this means that the meteor trail must be perpendicular to the radar pointing direction in order for it to be seen specularly. Non-specular scattering is only possible when the radar points near perpendicularly to the Earth's magnetic field; turbulence sets up field-aligned irregularities of the trail plasma within a few tenths of seconds, providing enhanced scattering in the direction perpendicular to the magnetic field until the irregularities dissipate after a few seconds [Dyrud *et al.*, 2002]. The example seen in Figure 1.6 is of a non-specular trail. Both types of trail echoes can be used to calculate meteoroid speed [Cervera, Elford, and Steel, 1997; Dyrud *et al.*, 2004] and observe upper atmospheric winds [Hocking, Fuller, and Vandepeer, 2001; Oppenheim *et al.*, 2009].

1.2.3 Challenges

Much of what we know about the ionosphere has been gleaned from radar data, but there are still processes that are difficult to analyze. Radar presents an inherent conflict between sensitivity and ambiguity; longer transmitter pulses are needed to put more energy on target, but they have a correspondingly poorer range discrimination capability. Pulse compression, encoding the waveform on transmission and decoding the returned signal on reception using matched filtering, is the standard way to settle the conflict. However, matched filtering still results in ambiguity in the form of self-clutter [Woodward, 1980]. Because of this conflict, it is difficult to make high quality measurements in situations where both sensitivity and high resolution are needed.

With meteors, for example, we lack good observations of two phenomena: fragmentation and flares. In the case of fragmentation, Mathews *et al.* [2010] show meteor head echoes with an oscillating signal intensity that they explain by modeling the signal as interference

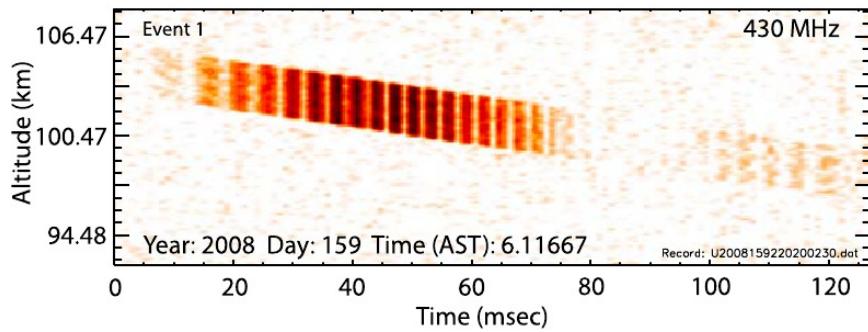


Figure 1.7 – Meteor head echo oscillations. This data from the Arecibo radar shows a meteor head echo with oscillating signal strength. The figure is reproduced from Mathews *et al.* [2010], and unfortunately no color scale for the signal strength was given. Mathews *et al.* [2010] theorize that this oscillation indicates interference between two distinct head echo scatterers produced by meteoroid fragmentation.

between two separate scattering centers, with the theory that such a signal represents the result of meteoroid fragmentation. An example of this oscillation is shown in Figure 1.7, but the fragmentation theory that could explain it has not been definitively proven. Flares and the related terminal events are a process sometimes seen with head echoes where the signal intensity suddenly increases and trail scattering is immediately visible at the location of the event. An example of this process can be seen in Figure 1.6, where a flare accompanies the meteor head echo. Currently no evidence has been observed that would explain why flares happen. Higher resolution data with reduced ambiguity is necessary to find answers to these mysteries, but such data cannot be produced with current techniques.

These challenges are somewhat unique to the applications found in the ionospheric radar community. Because of the wide variety of plasma processes and the corresponding demands on detail in space, time, and frequency, a wide variety of transmission waveforms are used. Some of these waveforms use deliberately-spread energy in frequency space to break the inherent relationship between sensitivity and resolution. Unfortunately, matched filter processing brings ambiguity and decoding artifacts, problems which cannot be ignored due to the high range of signal strength from one scatterer to the next. Current radar measurement techniques need to improve in four major areas to meet the challenges presented by the host of ionospheric plasma processes: 1) Differentiation of targets in crowded and variable environments, such as the ability to separate meteor and electrojet signals in the

equatorial ionosphere; 2) Elimination of self-interference of range-spread targets, which include non-specular meteor trails among others; 3) Higher resolution to observe small-scale processes, such as meteoroid fragmentation and meteor flares; and 4) Flexibility to achieve all of these goals at once.

1.3 Sparsity

The concept of sparsity is fundamental to how the brain processes information and manages to make sense of the deluge of sensory inputs that it receives. This principle, known as *sparse coding*, states that “information is represented [in the brain] by a relatively small number of simultaneously active neurons out of a large population” [*Olshausen and Field, 2004*]. Sparse coding makes the inherent structure of natural signals explicit, providing increased storage capacity and better energy efficiency in addition to simplifying higher-level reasoning. To achieve sparse coding, individual neurons respond to and represent just one particular aspect of a sensory input: an image is viewed through the retina, and a small number of neurons in the visual cortex respond to the image’s localized patterns and edges [*Olshausen and Field, 1997*]; a sound resonates at the individual frequencies of the inner ear hairs, and a sparse collection of neurons respond to the different frequency bands and corresponding time windows of the wave [*Lewicki, 2002*]. The responses of the individual neurons collectively form a dictionary from which a small number of elements are used to represent any one particular input. Though this neurological example represents just one case of the practical use of sparsity, the concept of finding simple descriptions to represent and reason with information is universally applicable. In a philosophical sense, sparsity is embodied by the principle of Occam’s Razor: the simplest explanation is best.

1.3.1 Sparsity in Nature

As evident with the success of sparse coding in the brain, many natural phenomena have sparse representations, and great use can be made of this fact to perform compression in the digital domain. Mathematically, signals like images, sound waves, or radar returns can be represented with vectors. In the case of images, this can be as simple as dividing the



Figure 1.8 – The structure of natural images. This digital image of alpacas at Jicamarca is 1280×960 pixels, and at 3 bytes per pixel it would require 3600 KiB of memory to store uncompressed. In its current JPEG compression format, it requires 420 KiB of memory instead. Structure in the image (patterns of color and continuity of edges) is used to reduce the required memory by a factor of about 8.5.

image into a fixed number of pixels and having each vector element give the color and intensity value of a single pixel. If the vector representation is sparse, with only a few of its elements nonzero (or even approximately sparse with most elements *close* to zero), then one can compress the signal by only storing the nonzero values (or values above a certain threshold).

But alas, the straightforward pixel representation of most images results in a vector that is dense with values, and this holds true for the straightforward representation of many other signals as well. As is the case in the brain, however, a sparse representation is often possible by choosing a dictionary of fundamental signals and composing any particular signal through combinations of the dictionary elements. If only a small number of the dictionary signals need to be used, then the vector giving the respective weights for a particular signal will be sparse. This is how images are processed in the visual cortex with a wavelet-like dictionary and how they are compressed in the JPEG format with a discrete cosine transform dictionary.

As a concrete example in the imaging case, consider Figure 1.8. Uncompressed, this

picture of alpacas at the Jicamarca radar site requires 3600 KiB of memory to store, 1280 × 960 pixels at 3 bytes per pixel. When compressed in the JPEG format through a discrete cosine transform and by discarding unimportant (small) coefficients to create an exactly sparse representation, this image requires 420 KiB of memory. That's about an 8.5 times savings, representative of the structure inherent in the image and captured in the chosen dictionary.

1.3.2 Sparsity in Radar

Radar signals are not exempt from the sparsity seen in nature. A radar transmits a signal and measures the intensity and phase of any reflections as a function of distance, with Doppler frequency shifts present in the reflection if a target is moving. Reflections from most radar scenes occupy only a small portion of that range-frequency space because the targets themselves are small in number and either spatially limited, have a compact Doppler spectrum from organized movement, or both. Most of the space being observed by the radar does not reflect any signal. Using a dictionary that decomposes signals on a joint range-frequency space captures the natural structure of radar targets and makes the sparsity exploitable.

1.3.3 Sparsity in Action

Whether influenced by research into brain physiology or not, the concept of sparsity has come to the fore in many fields of scientific research as a means of speeding, simplifying, or increasing the accuracy of information processing. Often this research falls under the moniker of *compressed sensing*, which exploits sparsity, global measurement, and nonlinear reconstruction to acquire data faster or with higher resolution. Conceptually, since images are often sparse, it does not make sense to measure each pixel individually (as with a regular camera) only to then throw most of that data away to compress it. Instead, compressed sensing takes the approach of directly measuring a compressed representation of the image in order to save time and memory.

Applications of compressed sensing are continually increasing in number. Notable examples that span a variety of fields include the following: the single pixel camera, which uses a filter to randomly mix components of an image and produce a series of mixed-image

single pixel readings from which the original image can be reconstructed [Duarte *et al.*, 2008]; compressed magnetic resonance imaging, which undersamples the spatial Fourier transform space of the image to speed acquisition and assumes image sparsity to achieve reconstruction [Lustig *et al.*, 2008]; and highly-parallel gene screening, which forms overlapping groups of genetic samples from a large set of donors, tests each group in bulk, and assumes rarity of the tracked gene pattern to determine which individual donors have it [Erlich *et al.*, 2010; Golan, Erlich, and Rosset, 2012]. Sparsity serves as a useful representation of prior knowledge in all of these and many more applications, so it is no wonder that many fields of research have found success in leveraging sparsity to make important advances.

1.4 Contributions

In developing a sparsity-based radar technique for flexible, high-resolution measurement of ionospheric plasmas, this dissertation makes several important contributions:

- *A new perspective for radar measurements is espoused in which the radar scene reflectivity is recovered by inverting a mathematical model.*

It is intuitively known that radar signals are sparse in a delay-frequency basis, but traditional approaches are not well-equipped to take advantage of that fact. A discrete radar model that captures signal sparsity is proposed using an image-blurring analogy, and this model is noted to be the adjoint of the standard pulse compression technique that uses a matched filter bank. Thanks to this connection between the two methods, an intuitive interpretation of the inversion technique as an iterative thresholding matched filter is presented.

- *A discrete radar model is rigorously derived from the continuous measurement equation for a general target scene.*

The derivation results in an explicit formulation for how the discrete model's reflectivity coefficients represent arbitrary distributed scatterers. Using this expression, it is shown that true target sparsity is reasonably preserved in the discrete representation provided that the number of frequency grid points is high.

- *A novel waveform inversion technique is formulated based on finding a sparse representation of the radar scene using the discrete model.*

Recovery of the true solution via l_1 -norm minimization is proposed based on the theory of compressed sensing. This theory is leveraged to provide theoretical guarantees for the recovery of the true solution provided that the radar pulse waveform produces sufficiently incoherent measurements and that the number of measurements is at least on the order of the solution's sparsity. Waveform inversion is shown to eliminate matched filter sidelobe artifacts and enable higher-resolution solutions.

- *The waveform inversion method is implemented using modern convex optimization techniques tailored for efficient computation.*

The focus is on algorithms based on the proximal operator because of their performance, ease of use, and ability to handle large problem dimensions. Since these algorithms are an active area of research, the implementation combines acceleration and adaptive step size enhancements from separate formulations to enable state-of-the-art convergence performance. All computer code is made freely available as open-source software to encourage further development, adoption, and collaboration.

- *The real-world flexibility and effectiveness of the inversion technique is demonstrated by the elimination of filtering artifacts from meteor observations made with a variety of standard radar waveforms.*

Successful inversion is shown to depend on the compactness of the peak of the waveform's ambiguity function, while quality of the solution is shown to depend on the fidelity of the transmitted waveform to its idealized model. Increased resolution of the inversion solution is demonstrated but not validated, although the results are promising for future applications.

1.5 Reader's Guide

This dissertation is divided into four parts that build on one another but are relatively self-contained: introduction, radar analysis, waveform inversion, and conclusion. The preceding introduction is intended to familiarize the reader with the primary topics of interest in order to motivate and connect the work as a whole. Part II concerns radar analysis and goes into more detail about how radar measurements are made: Chapter 2 describes current waveform coding and processing techniques and discusses their limitations; Chapter 3 details the derivation and analysis of a mathematical radar model, providing a different way of analyzing radar measurements. The next segment, Part III, introduces a waveform inversion method that combines the radar model with prior information of sparsity: Chapter 4 describes the theory of compressed sensing and related methods of convex optimization; Chapter 5 details the implementation of the waveform inversion technique, including algorithmic advances; and Chapter 6 provides experimental results using the Jicamarca radar to test waveform inversion and compare the effect of different waveforms. Finally, Part IV contains the concluding chapter, which discusses the benefits provided by the new techniques and highlights areas for future research. One appendix is included, Appendix A, which details the simplification of the reflectivity coefficient expression derived in Chapter 3 as part of the radar model.

Part II

Radar Analysis

Chapter 2

Radar Background

Radar come in a variety of forms, but for this work we are primarily concerned with pulsed monostatic radars. In this mode of operation, a radar transmits a pulse of electromagnetic energy and then switches to a listening mode to record the signal from any resulting reflections. The received power P_r backscattered from a target is described by the radar equation [Skolnik, 2001]:

$$P_r = \frac{P_t G}{4\pi r^2} \times \sigma \times \frac{A_e}{4\pi r^2}. \quad (2.1)$$

The transmitted power P_t is concentrated by the antenna with a gain of G , but it also decays with distance (or range) r according to $1/(4\pi r^2)$ as it spreads into space. The target then scatters a fraction of the power, the target's radar cross section σ , back toward the radar. The backscattered power once again decreases with range r before returning to the antenna and experiencing a gain dependent on the effective aperture area A_e . In this monostatic case with the transmitting and receiving antenna co-located, the aperture is related to the transmitting gain by $G = 4\pi A_e / \lambda^2$, where λ is the wavelength of the radio wave. Though it is important to remember how the received power relates to the radar's power and gain parameters (P_t , G , A_e) and the target's cross section and range (σ , r), we will mostly be able to ignore these effects by working only in terms of the received power as a function of range or, equivalently, time delay.

Once a pulse is sent and the received signal is recorded for as long as desired, the process is repeated to monitor how the radar targets change over time. The time from one pulse to the

next is called the *pulse period* and its inverse the *pulse repetition frequency* (PRF). The PRF is an important parameter in the design of a radar experiment since it specifies the amount of time during which the radar can listen for reflections. Hence, it also gives the maximum unambiguous distance at which targets can be measured. For ionospheric experiments, pulse periods on the order of one to ten milliseconds, at a minimum, are common. With a zenith-pointing radar, one millisecond corresponds to a maximum unambiguous range of 150 km, which falls in the E-region (the total distance traveled by a returning wave in 1 ms is $c[m/s] * 1e^{-3}[s] \approx 300[km]$ where $c \approx 3e8[m/s]$ is the speed of light, so the maximum range is half that).

Usually these pulsed monostatic radars operate at a fixed frequency, or multiple fixed frequencies in independent channels. Data is typically acquired by sampling the received waveform so that its zero frequency corresponds to the radar's baseband frequency. This frequency shift is accomplished either directly in the digital domain after high-frequency sampling or with analog filtering and mixing [Richards, 2005; Skolnik, 2008]. Either way, the end result is a digitally sampled signal of complex numbers encoding the magnitude and phase of the received signal, a vector that can be stored and further analyzed to extract information about the radar scene. We employ the following notation for representing the transmitted and received radar signals: the transmitted waveform is represented by the complex sequence $s[l]$ with L elements and indices $l = 0, \dots, L - 1$; the received waveform is represented by the complex sequence $y[m]$ with M elements and indices $m = 0, \dots, M - 1$. For consistency, we will always use l to index the transmitter sequence, m to index the received sequence, and the corresponding capital letters to denote the respective sequence lengths.

2.1 Pulse Design

While design parameters like the pulse period and system parameters like the transmitter power and antenna gain are important in designing a radar experiment, the pulse itself is critical. Imagine the pulse as a square wave centered at the radar's operating frequency as shown in Figure 2.1. In this figure, the delay space is divided into samples according to the sampling period, and the transmitted and received waveforms are shown along with

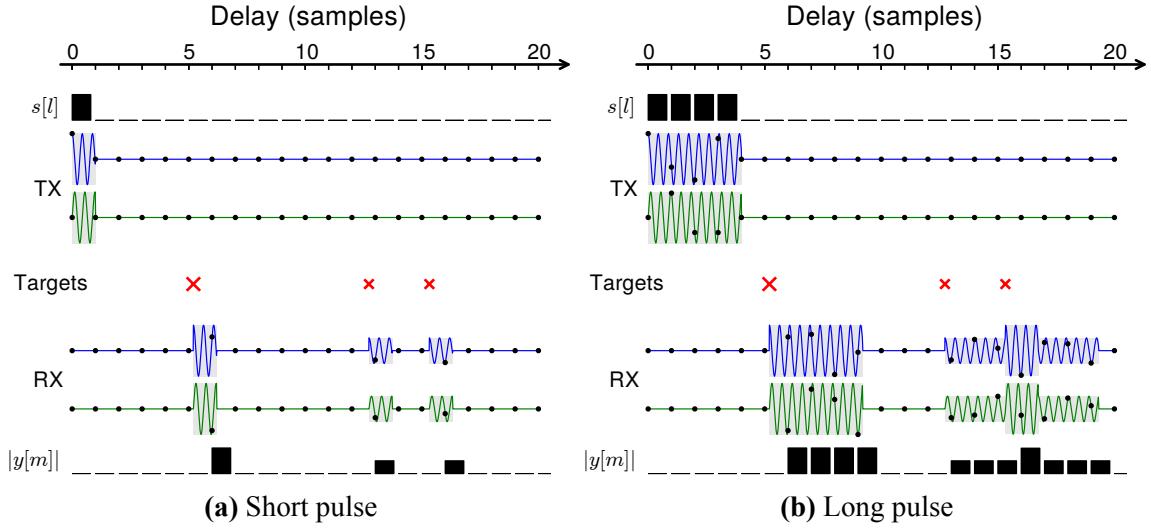


Figure 2.1 – Transmitted and received signals for a short and long pulse. In each case, transmitted (TX) and received (RX) waveforms are shown as a function of delay along with the corresponding sampled baseband signals $s[l]$ and $y[m]$. Each waveform has an in-phase (blue) and quadrature (green) component, representing the wave’s magnitude and phase when thought of as the real and imaginary parts of the complex signal. The received waveform encapsulates the delays and magnitude differences applied to the transmitted pulse by three point targets. Because of ambiguity created by the pulse length, the received signal for the short pulse is easier to interpret than for the long pulse, but the long pulse allows for increased total power through averaging over multiple samples.

their vector counterparts. The waveforms display a fixed frequency, the radar operating frequency, but are modulated on and off according to the pulse length. Although the ideal on/off of a square wave is never achieved in practice, it is a convenient way to conceptualize the waveforms.

2.1.1 Pulse Length

The simplest mode of operation matches the length of the transmitted pulse with the sampling rate of the received signal. In this way, each sample represents the sum of reflections over a particular range window with width corresponding to the pulse length. This short

pulse mode of operation is depicted in Figure 2.1a, where three point targets appear individually as three nonzero samples in the received signal. In general, range resolution is defined by the sampling rate, while range ambiguity, or the interference region for multiple scatterers, is defined by the pulse length. In the short pulse case, the range resolution matches the range ambiguity so that the ambiguity does not pose an interpretation problem.

The primary downside of a short pulse is that it limits the sensitivity of the radar in terms of energy on target. Since the peak power of a pulse is limited by the physical radar equipment, the only way to increase the total power is to increase the pulse length. With a pulse greater than the sampling period, one can add the returned signal over multiple samples within the pulse to achieve a higher sensitivity due to higher total power. However, a longer pulse also increases the range ambiguity of the measurement since the return from any single target will affect multiple receiver samples. This long-pulse mode of operation is depicted in Figure 2.1b. The example shows both the positives and negatives of a longer pulse: more samples for increased sensitivity with the first target, but ambiguity in the received signal between the latter two closely-spaced targets.

Two additional points are relevant when discussing pulse length: transmitter duty cycle and Fourier frequency resolution. In order to transmit large amounts of power instantaneously, transmitters often have a maximum duty cycle that describes the fraction of time that the transmitter is allowed to be on. A duty cycle is necessary so that there is enough off time to build up the power supply and allow physical components to cool. Because of these requirements, the pulse length cannot be increased indiscriminately and has a fixed limit with respect to the pulse period. The other consideration is Fourier frequency resolution. For persistent targets, pulse-to-pulse Fourier analysis is possible to measure how the target evolves in time, with the maximum observable frequency component given by half the pulse repetition frequency. For quickly-evolving or transient targets, the pulse-to-pulse method of Fourier analysis is not sufficient and intra-pulse Fourier analysis is necessary. In that case, the maximum observable frequency is half the sampling frequency and the resolution is the inverse of the pulse length. Thus, increasing the pulse length also has the benefit of improving the frequency resolution of intra-pulse Fourier analysis.

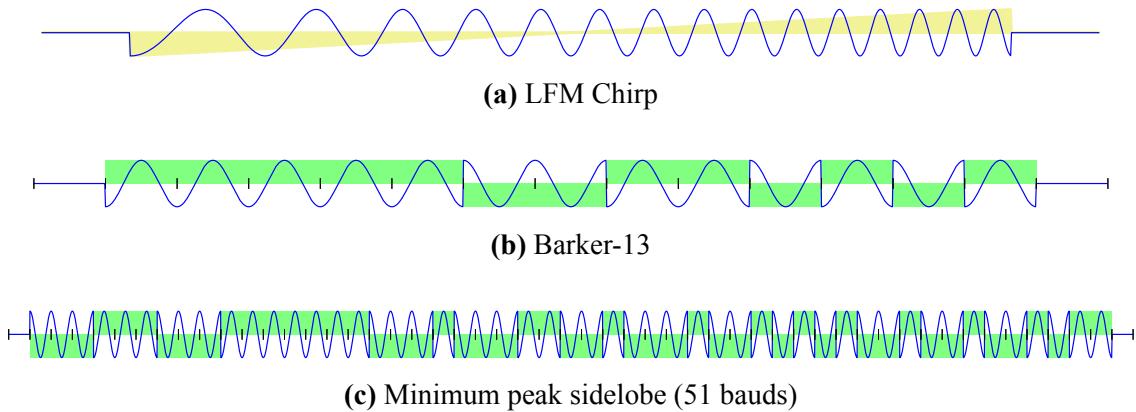


Figure 2.2 – Waveform encoding examples. These pulse waveforms depict three commonly used encodings: (a) an LFM chirp pulse with a linear frequency sweep from low to high, (b) a binary phase-coded pulse using the Barker-13 code, and (c) a binary phase-coded pulse using a minimum peak sidelobe code with 51 bauds. The yellow fill in the background of (a) indicates the pulse's frequency with respect to the baseband, while the green rectangles in (b) and (c) indicate the positive or negative phase of each baud.

2.1.2 Pulse Waveforms

One way to deal with the ambiguity problem is to code the pulse by modulating it with a designed waveform. Because of the way that coding allows for concentration of the pulse energy after decoding, this process is often called *pulse compression*. One simple and widespread method is linear frequency modulation (LFM), also known as a chirp waveform. A chirp pulse starts off at a low frequency and linearly increases in frequency over time to a high frequency, or vice-versa. An example is given in Figure 2.2a. Another common form of modulation is to divide the pulse into equal length segments called *bauds* and encode using the phase of each baud. Often the codes use only two opposite phases (e.g. 1 and -1) and the process is called binary phase coding, but the use of more phases is certainly possible. Two examples of commonly-used and well-behaved binary phase codes are shown in Figure 2.2. The Barker-13 code of Figure 2.2b has 13 bauds with phases described by the binary sequence

$$s_{B_{13}} = 1111100110101, \quad (2.2)$$

while the minimum peak sidelobe code of Figure 2.2c has 51 bauds with phases described by the binary sequence

$$s_{\text{MSL}_{51}} = 00011100011111100010001100100010010101001001001011. \quad (2.3)$$

2.2 Matched Filtering

When coded pulses are used, or even just long pulses, one needs a way to decode the received result and integrate the distributed power. The intuitive way of doing this, by starting with each sample in the received signal and operating over a window the length of the pulse to reverse the effects of the code and sum the result, is called a *matched filter*. Stated another way, the matched filter correlates every segment of the received signal with the transmitted pulse. Mathematically, the matched filter performs the correlation

$$u[p] = \sum_{m=0}^{M-1} s^*[m - p + (L - 1)]y[m] \quad (2.4)$$

or equivalently

$$u[p] = \sum_{l=0}^{L-1} s^*[l]y[p + l - (L - 1)], \quad (2.5)$$

where s^* signifies the complex conjugate of the code s and the matched filter result is denoted by $u[p]$ for $p = 0, \dots, P - 1$ with $P = M + L - 1$. These equations reference “out of bounds” indices of $y[m]$ and $s^*[l]$; for simplicity we take the convention that any out of bounds values are zero. The correlation process of the matched filter is shown in Figure 2.3 in the context of autocorrelation of the Barker-13 code, which corresponds to the ideal case when the received signal is due to perfect reflection from a point target. The benefits of the matched filter are its sensitivity and ease of computation: it produces the optimal signal-to-noise ratio (SNR) when decoding point targets and takes the form of a simple linear filter.

There is one degree of freedom when computing the matched filter: one can scale the

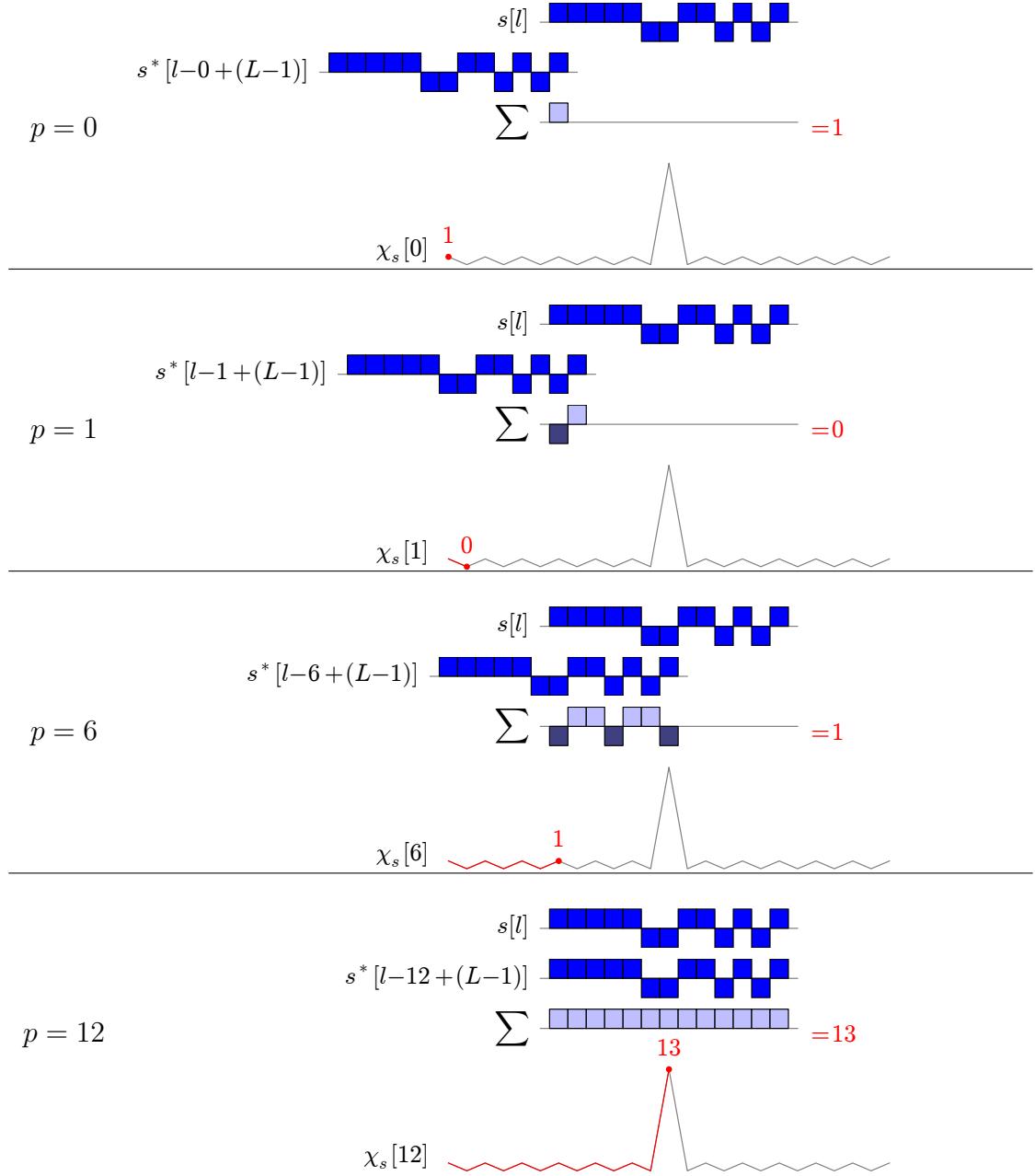


Figure 2.3 – Autocorrelation of the Barker-13 code. Each panel depicts the calculation of a single sample p for the autocorrelation of the Barker-13 code. The first two lines of each panel show the code sequence $s[l]$ and its shifted conjugate. The third line shows the resulting multiplication of the first two lines and sum total of the values. The fourth line gives a plot of the autocorrelation, highlighting the calculated sample.

code s by any amount and still have optimal SNR at the output. This is because increasing s by a scalar quantity not only increases the magnitude of the output when run against a matching sequence, but it also equally increases the magnitude of the output when run against non-matching random noise. Since the output is typically examined in terms of SNR, this arbitrary scaling makes no difference to the analysis. In a practical sense, however, it is often convenient to scale the code s so that it has an l_2 norm of one:

$$s'[l] = \frac{s[l]}{\sqrt{\sum_{k=0}^{L-1} |s[k]|^2}}. \quad (2.6)$$

When this is done, noise at the output of the matched filter has the same average power as noise at the input, and no recalculation of the noise power is necessary to compute SNR.

2.2.1 Autocorrelation and Range Sidelobes

Since targets essentially reflect the transmitted signal, just with phase and power differences and mixing from multiple or distributed targets, we can examine the effect of matched filtering by looking at a code's autocorrelation:

$$\chi_s[p] = \sum_{l=0}^{L-1} s[l] s^*[l - p + (L - 1)], \quad (2.7)$$

where $p = 0, \dots, P - 1$ with $P = 2L - 1$. As shown in Figure 2.3 for the Barker-13 code, the autocorrelation of any code has a peak in the middle, corresponding to when the received and transmitted codes are perfectly aligned. In the matched filter result, the peak occurs at the delay, or range, corresponding to the location of the reflecting target. Away from the peak, the autocorrelation is not zero but rather some pattern of varying response that depends on the code. When these off-peak responses are seen in the matched filter result, they are called *range sidelobes*.

Particular codes are selected for use because of their autocorrelation properties, where keeping sidelobe response to a minimum is desirable to minimize ambiguity and artifacts. The Barker-13 code is the longest in a class of Barker codes of differing lengths that have a maximum sidelobe level of one. It is possible to enforce similar minimum peak sidelobe

performance for codes of other (particularly longer) lengths; these are unsurprisingly called minimum peak sidelobe codes, and they have been found through simple brute force search [Lindner, 1975; Cohen, Fox, and Baden, 1990; Coxson, Hirschel, and Cohen, 2001; Coxson and Russo, 2005]. The 51-baud example given in Figure 2.2c is just one of the codes of different lengths with the minimum peak sidelobe property. The LFM chirp that we encountered earlier also has well-behaved range sidelobes, and in that case they decay from the peak according to a sinc function.

2.2.2 Frequency Banks of Matched Filters

Moving targets complicate the use of a matched filter to decode the received signal because the Doppler frequency shift causes a mismatch. If a target is moving toward (away from) the radar, the reflected signal is shifted up (down) in frequency. In order to correlate properly, the matched filter needs to be similarly frequency shifted. Of course, the true frequency shift due to target motion is not known ahead of time, so it is another unknown that must be determined along with the target's range. We cope with this problem by processing the received signal with a bank of filters, each with a different frequency shift. Computationally, it is convenient to have these filters be equally spaced in frequency up to the maximum detectable frequency shift, determined from the sample rate according to the Nyquist criterion. When that is done, the filter bank can be computed efficiently through the Discrete Fourier Transform (DFT). The equation describing the matched filter bank is similar to the single matched filter:

$$u[n, p] = \sum_{m=0}^{M-1} s^*[m - p + L - 1] e^{-2\pi i n(m-p+L-1)/N} y[m] \quad (2.8)$$

or equivalently

$$u[n, p] = \sum_{l=0}^{L-1} s^*[l] e^{-2\pi i n l / N} y[p + l - (L - 1)]. \quad (2.9)$$

The difference, of course, is that for each of the N frequencies indexed by n for $0, \dots, N-1$, the transmitted code is modulated at that frequency before correlation with the received signal.

Visualizing the results of a filter bank also poses a challenge. Since we will frequently have to deal with this type of visualization, a brief discussion is in order. Often we want to plot data from multiple pulses in order to see how the targets vary over time. With a filter bank, that means plotting signal strength as a function of three variables: range, frequency, and time. A natural representation is a video, with each frame being a plot of signal strength as a function of range versus frequency for a single pulse. Of course, that has the distinct disadvantage of not working well in print. A more suitable approach is to truncate one of the dimensions in some way. If all of the targets have a negligible or similar Doppler shift, then it is reasonable to just plot the data for that one frequency. This results in a standard range-time-intensity (RTI) plot with range as the vertical axis, pulse time as the horizontal axis, and signal intensity represented by color or shading. Figure 1.5 in the introduction is an example of this approach. However, when moving and stationary targets mix (as with the meteor head echo and trail shown in Figure 1.6), single frequency truncation is insufficient. In those cases and when an RTI-type plot is desired, the single frequency that gives the maximum response over all ranges is selected for each pulse individually. Most of the time this is sufficient to produce a well-decoded plot, but it still fails when multiple frequency-shifted targets appear in the same pulse. There are other approaches for visualizing filter bank outputs, but here we will only employ the peak-frequency-response approach for RTI plots.

2.3 Measurement Ambiguity

When analyzing the received radar signal for a single pulse, we are interested in how the transmitted signal was reflected as a function of delay (range) and frequency (range-rate). A frequency bank of matched filters produces a peak response at the delay and frequency of any target, but it also produces off-peak sidelobes as an artifact of the filtering process. The delay-frequency sidelobes can, depending on their magnitude, obscure weaker targets or bias the measurement through interference. We call these effects measurement ambiguity.

2.3.1 Ambiguity Function

As we saw in the single matched filter case, sidelobes are a consequence of the autocorrelation of a particular code. The same is true in the filter bank case, except we must consider the autocorrelation not only as a function of delay but as a function of frequency:

$$\begin{aligned} \chi_s[n, p; n', p'] = & \sum_{m=0}^{M-1} s[m - p' + L - 1] e^{2\pi i n'(m-p'+L-1)/N} \\ & \times s^*[m - p + L - 1] e^{-2\pi i n(m-p+L-1)/N}. \end{aligned} \quad (2.10)$$

It is necessary to explicitly write this function in terms of the two frequency indices n and n' and the two delay indices p and p' because of the nonlinear interaction between frequency and delay in the exponential term. As with before, $n, n' \in [0, N - 1]$, $p, p' \in [0, P - 1]$, and $M = P - (L - 1)$. The delay-frequency autocorrelation of a code is also known as its *ambiguity function*, although this term is sometimes used to denote the squared magnitude of the autocorrelation [Woodward, 1980]. The sidelobes of a code are completely described by the code's ambiguity function.

To make this concept more accessible, consider again the example codes encountered earlier. The ambiguity functions of the Barker-13, minimum sidelobe, and LFM chirp waveforms are shown in Figure 2.4. As a point of comparison, this figure also includes the ambiguity function of an uncoded pulse. All of the ambiguity functions have a peak centered at zero delay and zero frequency, or rather the reflecting target's delay and frequency. The uncoded pulse has a strong response for many samples surrounding the peak: along the zero-frequency axis, the response decays linearly on both sides of the peak, while along the zero-delay axis, the response decays with the periodic sinc function. In contrast, the peaks for the Barker-13 and minimum sidelobe codes are well-defined and isolated with the sidelobe response lower but spread out across the delay-frequency space. The LFM chirp is interesting because of its ridge-like response, with strong ambiguity along a line in delay-frequency space but minimal sidelobes elsewhere. Thus, the LFM chirp is popular for use with a single matched filter since it is insensitive to frequency shifts (aside from a corresponding shift in the apparent target range).

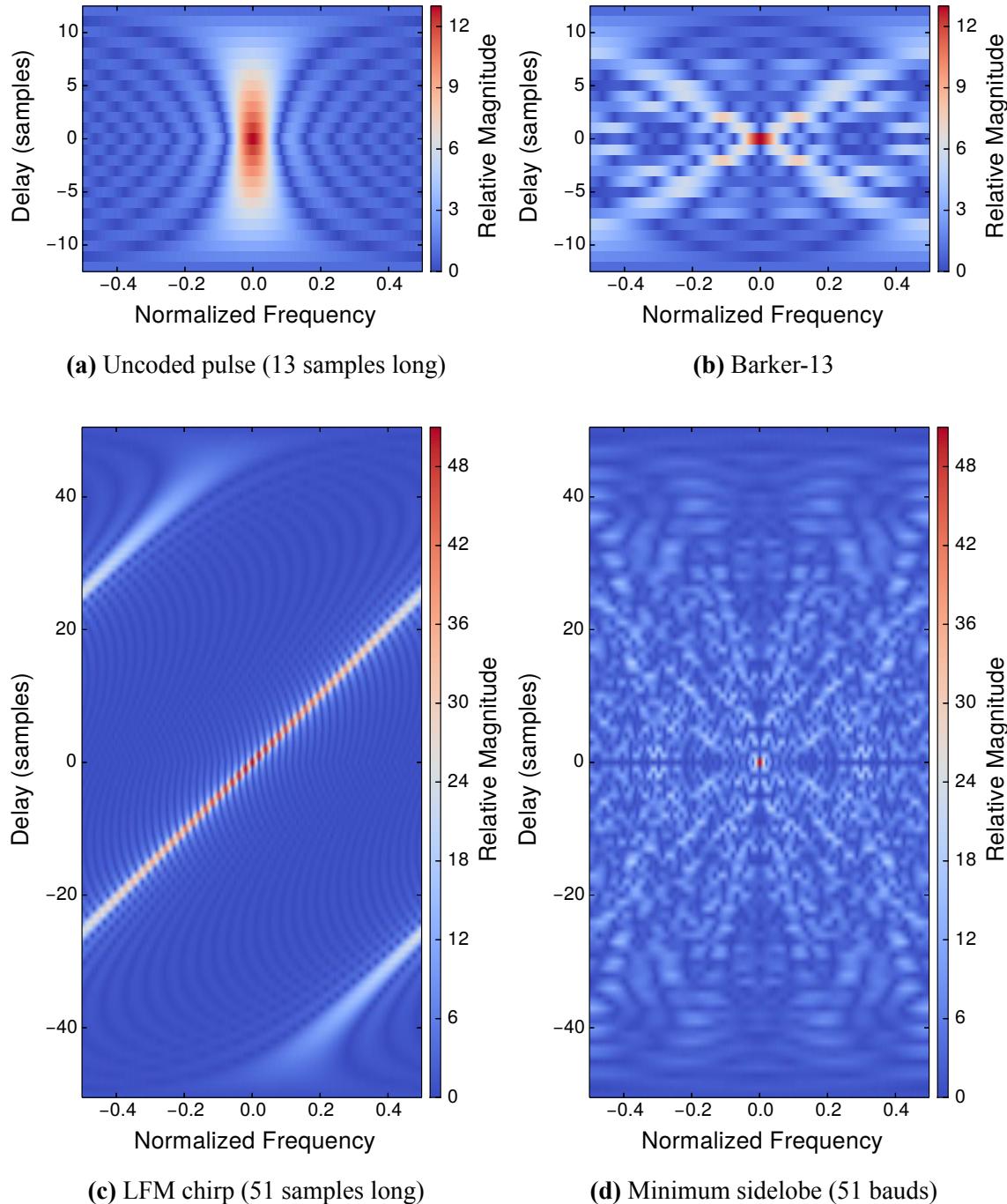


Figure 2.4 – Ambiguity function examples. These delay-frequency images show the ambiguity function magnitude with $N = 512$ frequencies for the (a) 13-sample uncoded pulse, (b) Barker-13 code, (c) 51-sample discretized LFM chirp with bandwidth equal to the sampling frequency, and (d) 51-baud minimum peak sidelobe code.

2.3.2 Sidelobes in Action

The differences in ambiguity between these codes have a large impact in practice. Figures 2.5 and 2.6 show a meteor head echo and trail measured separately using the 50 MHz Jicamarca radar with a $40\mu s$ uncoded pulse, a $13\mu s$ Barker-13 pulse, a $51\mu s$ minimum sidelobe pulse with 51 bauds, and a $51\mu s$ LFM chirp with 1 MHz frequency sweep. Since these are RTI plots, the matched filter results show the peak-frequency-response filter for each pulse. Because of that, the artifacts that crowd the plot are mostly the zero-frequency-axis range sidelobes for each code. Unsurprisingly, the uncoded pulse's sidelobes completely dominate the output when that code is used. The meteor trail in particular appears as a high SNR mess. The Barker-13 code produces much more manageable sidelobes where the head echo and primary regions of trail scatter are clearly visible. Unfortunately, the Barker-13 pulse is necessarily much shorter in length, and the reduced sensitivity is evident with lower overall SNR. The minimum sidelobe code has higher SNR while keeping the acceptable sidelobe performance of the Barker-13 code. However, the sidelobes are still clearly significant and can easily obscure weaker scatterers, especially when the scatterers are shifted in frequency and have to contend with the very significant off-axis sidelobes. Finally, the LFM chirp produces the cleanest-looking results, an outcome which is not surprising given its ambiguity function. It is important to note though that they come at the cost of range bias and loss of frequency information; those shortcomings are only mitigated in this instance by the use of the other codes to determine the proper decoding frequency.

The example meteor head echo and trail are particularly well-behaved because they occur separately and without any other targets interfering. Even though that is usually the case when only meteors are observed, sidelobes are the primary reason that many outstanding ionospheric science questions cannot be satisfactorily answered with current techniques. Any time multiple or distributed scatterers are observed, sidelobes obscure or bias the measurements. The prior examples of challenging meteor phenomena, flaring in Figure 1.6 and fragmentation in Figure 1.7, are just two important cases. Matched filtering is one way to decode coded radar measurements, and its limitations are clear. What is needed is a new decoding technique that keeps the benefits of the matched filter, sensitivity and efficient computation, but does not produce sidelobes. The problem in developing such a technique

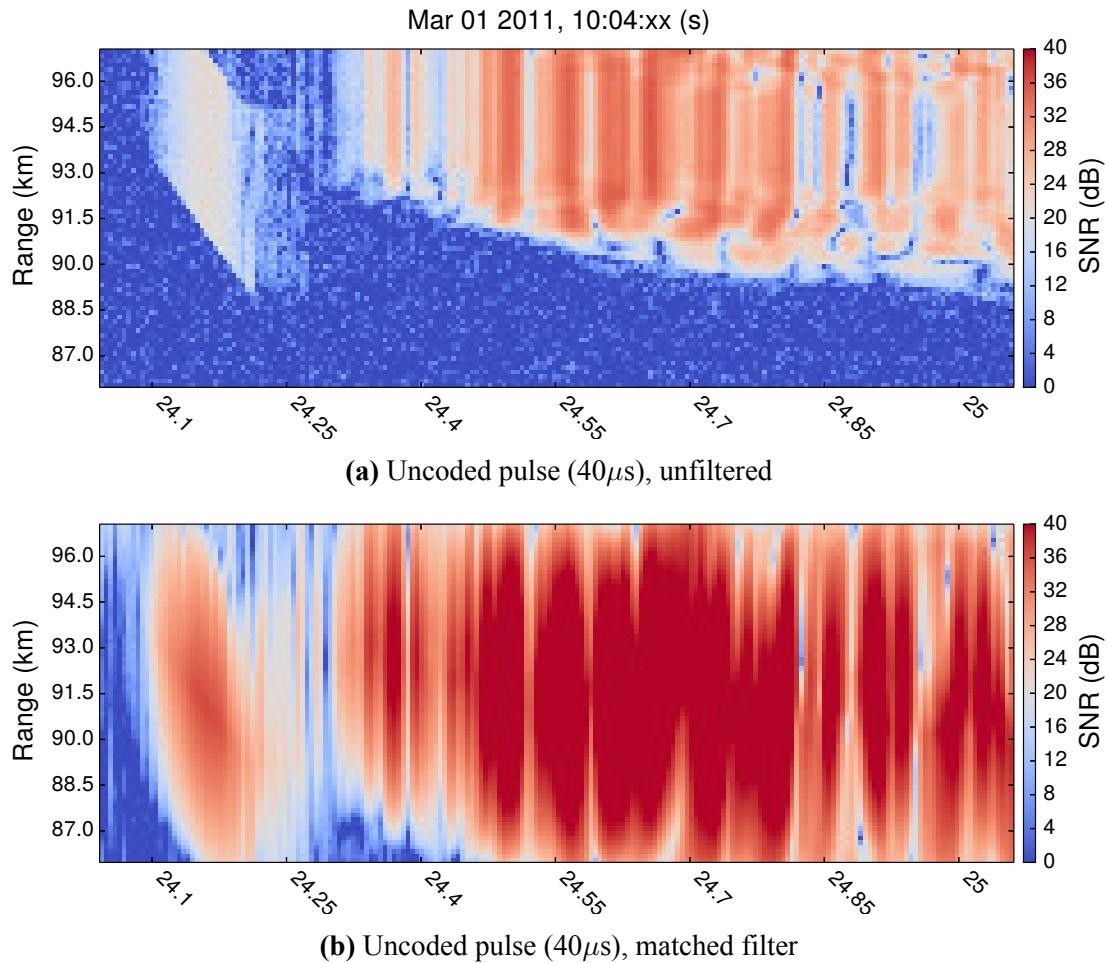


Figure 2.5 – Meteor head and trail measured with an uncoded pulse. These figures show the signal-to-noise ratio (SNR) as a function of range and time for one second of $1\mu\text{s}$ -sampled data taken with the 50 MHz Jicamarca radar. An uncoded pulse was used, with (a) showing the unfiltered result and (b) showing the peak-frequency-response matched filter result. The SNR is low before filtering, but the range sidelobes resulting from the filter completely obscure the meteor trail.

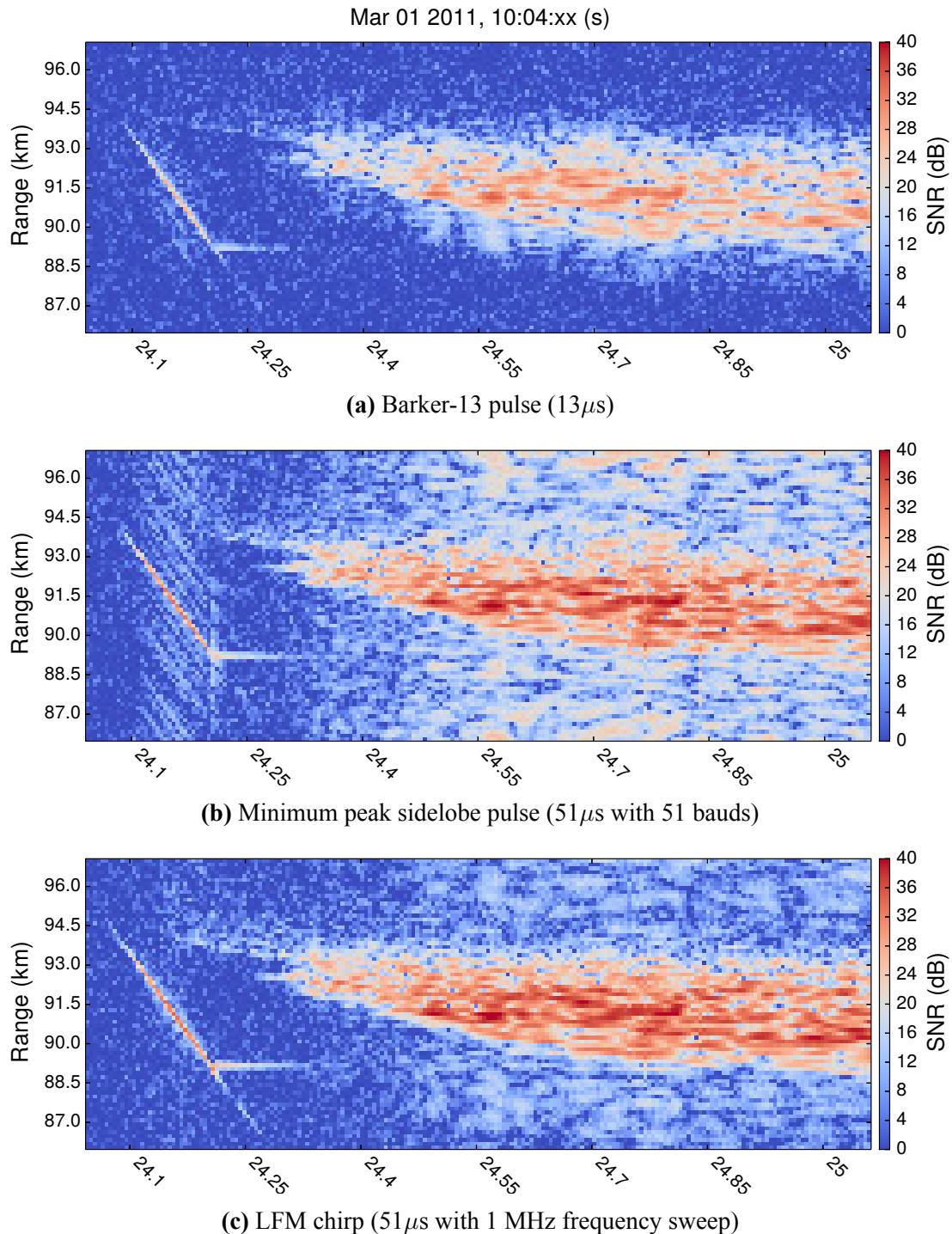


Figure 2.6 – Meteor head and trail measured with coded pulses. These figures show the peak-frequency-response matched filter results for different waveforms over one second of $1\mu\text{s}$ -sampled data taken with the 50 MHz Jicamarca radar. Range sidelobes from the codes create significant ambiguity and make interpretation difficult.

is that there are an infinite number of ways to construct a delay-frequency target scene that matches the measurements, and we're looking for the one that is both correct and sidelobe-free.

2.4 Current Techniques

2.4.1 Sidelobe Mitigation

Accounting for matched filter ambiguity is not a new problem, and there are many ways it has been addressed in the past. The most obvious method is the one we've already covered: sidelobe mitigation through the selection of well-behaved codes. Obviously waveforms like the Barker codes, minimum peak sidelobe codes, and LFM chirp have more manageable sidelobes than an uncoded pulse or even most other waveforms. This is by design, as those codes were formulated to have either the minimum peak range sidelobes or to be insensitive to frequency shifts. Their typical use in applications such as satellite tracking, where sensitivity is essential and point-like targets are the norm, demonstrates their effectiveness in particular circumstances. As we've seen, though, these codes are only effective when it is simple to ignore the sidelobes that do remain. It is especially hard to ignore those sidelobes when targets with multiple Doppler frequency shifts are present, since the off-zero-axis delay-frequency sidelobes are quite significant. This is a consequence of the codes being designed to minimize only the zero-frequency-axis range sidelobes. Codes employing a larger set of phases, such as quadriphase codes [Damtie *et al.*, 2008], show promise in improving performance over the widespread binary-phase codes [Virtanen *et al.*, 2013], but they still have many of the same fundamental sidelobe limitations.

2.4.2 Alternating Codes

Alternating codes are an approach to completely removing sidelobes by using a complementary set of codes over multiple pulses [Lehtinen and Häggström, 1987; Sulzer, 1993; M. Markkanen, Vierinen, and J. Markkanen, 2008]. When the multi-pulse data is summed together, sidelobes from the individual codes cancel out to make the set sidelobe-free. Since

incoherent scatter radars must already integrate over multiple pulses to get significant signal, alternating codes are widely used to measure incoherent scatter from the background ionosphere. As that application alludes to, alternating codes are an effective approach when target properties remain stationary over the time it takes to transmit the code set. Unfortunately, that is a constraint that limits flexibility and rules out many cases of interest.

2.4.3 Inverse Filters

It is also possible to leave matched filters behind altogether and use an alternative filtering or inversion technique. An inverse filter is one designed specifically to produce no range sidelobes at a point target's particular frequency shift. Theoretically, inverse filters are infinite in length, but practically, finite-length truncation can achieve nearly indistinguishable inversion behavior. Like the matched filter, the inverse filter is a function of the transmitted code and is performed by a simple linear filtering operation on the received radar data. Since they are not optimal-SNR matched filters, however, inverse filters result in a loss of sensitivity that depends on the particular code used. Recent work into so-called perfect codes [Lehtinen *et al.*, 2009] make inverse filters more attractive, since perfect codes are designed so that the inverse filter is also the matched filter. One reason that perfect codes are not more widespread is that they require amplitude modulation in addition to phase modulation for each baud, and this can be difficult to achieve in practice.

Another reason that perfect codes are not more widespread is that inverse filters still produce significant sidelobes when the target frequency does not match the decoding frequency. This is evident in Figure 2.7, which compares the ambiguity function for the Barker-13 code for both the matched filter and the inverse filter. On the global scale of the entire delay-frequency space, the two ambiguity functions look extremely similar. The Barker-13 code does not suffer much SNR loss due to the inverse filter, but most of the sidelobes remain as significant as with the matched filter. Only by examining the zero-frequency axis is it apparent that the inverse filter does indeed produce no range sidelobes when it matches the target's frequency. In the common case when the target Doppler shift is unknown or multiple frequencies must be decoded, the inverse filter provides no advantage.

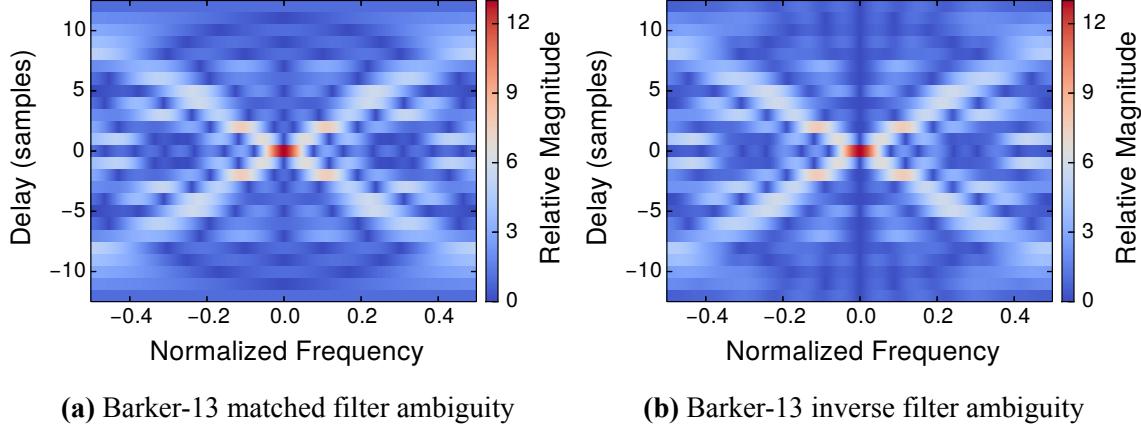


Figure 2.7 – Comparison of matched and inverse filter ambiguity functions. These delay-frequency images show the ambiguity function magnitude with $N = 512$ frequencies for the Barker-13 code. The **(a)** matched filter and **(b)** inverse filter ambiguities are nearly identical apart from the zero-frequency axis, indicating that the inverse filter is no more useful when decoding targets with unknown frequency shifts.

2.4.4 Inversion

Framing decoding as a standard inversion problem is the most general approach to handling coded waveforms, and this is the tactic that we will use in the subsequent chapters. First, however, we must note the existing methods that perform this type of waveform inversion. The common link in these approaches is the formulation of a forward model that translates the true radar scene into measurements. Once this model is known, it can be inverted using an optimization routine to recover the ambiguity-free radar scene from the received signal. The amplitude domain inversion method of *Vierinen, Lehtinen, and Virtanen* [2008] is similar to the one that we will develop later. The radar scene is decomposed over delay-frequency space just as with the matched filter, and a linear equation relates the delay-frequency scattering to the received signal. Just as we noted earlier, this model has an infinite number of solutions in the general case. However, *Vierinen, Lehtinen, and Virtanen* [2008] take the approach of limiting the analysis to targets with a narrow delay/range. The problem is then well-posed, and statistical inversion is achieved by maximizing the likelihood of the posterior distribution. Limiting the target to a known narrow range extent is akin to manually specifying that most of the delay space is free from significant scatterers.

It is a range sparsity assumption like the one that we will use, but their method requires the user to manually specify the target model whereas our approach is fully automated. That step in increased generality is significant for practical application.

Other existing inversion methods work with autocorrelation lag products of the received signal instead of the measurements directly. Lag profiles, the autocorrelation lags as a function of range, suffer from the same range ambiguity due to coding as the direct measurements. *Damtie, Lehtinen, and Nygrén* [2004] and *Virtanen et al.* [2008] describe how to invert the lag profiles by solving a linear maximum a posteriori (MAP) matrix equation. Another MAP method is applied for lag profile inversion by *Nikoukar et al.* [2008], with the difference being the use of regularization to stabilize the solution. These methods are hindered by the fact that they need either low SNR or the integration of many pulses in order to be applicable, otherwise the noise in the lag profile measurements is not Gaussian [*Lehtinen, Virtanen, and Vierinen*, 2008]. In practice, this constraint has limited their use to the decoding of the inherently low-SNR incoherent scatter measurements.

Chapter 3

Radar Model

With knowledge of the detrimental effects of delay-frequency sidelobes, we seek a radar measurement technique with the high sensitivity provided by coding but without the accompanying ambiguity wrought by matched filtering. In order to proceed, it is necessary to formalize the notion of sidelobe elimination. To do this, we develop a mathematical model for radar measurements and explore how this model relates to actual physical targets.

3.1 Defining the Model

3.1.1 Image Blur Analogy

Radar coding ambiguity is similar in effect to the blurring of images. Suppose there is a true image that we want to capture, $h[n, p]$. In capturing the image, the physical optics impose a point-spread function $\chi_s[n, p; n', p']$ that causes any point in the true image to be spread out in a determinable pattern in the measured image. The resulting blurred image could appear like in the example in Figure 3.1. Mathematically, this type of blurring takes the form of an inner product between the source image $h[n, p]$ and the point-spread function $\chi_s[n, p; n', p']$:

$$u[n', p'] = \langle h[n, p], \chi_s[n, p; n', p'] \rangle = \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} h[n, p] \chi_s^*[n, p; n', p'], \quad (3.1)$$

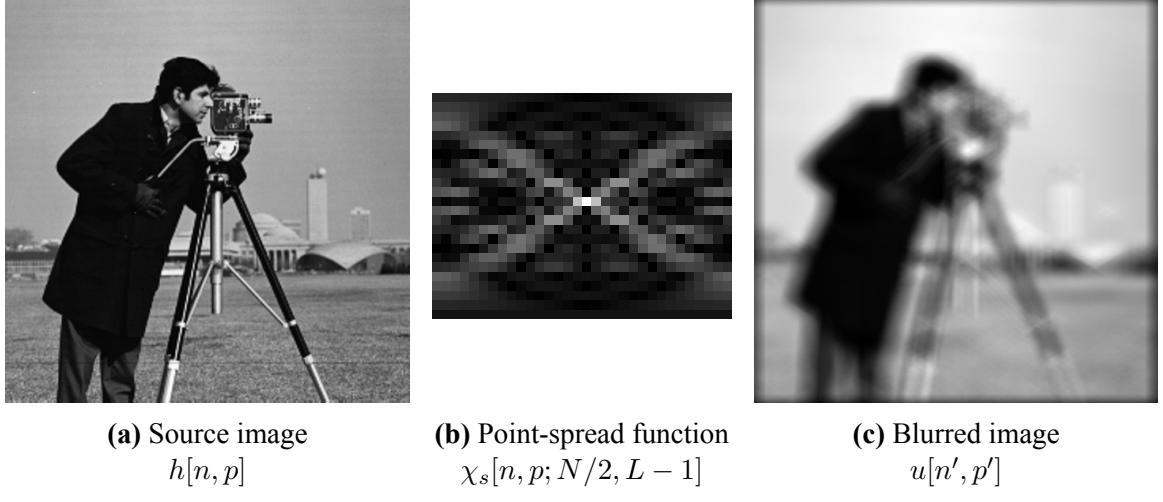


Figure 3.1 – Image blurring example. A blurred image (c) is formed by an inner product between a source image (a) and a point-spread function (b). The point-spread function describes how a block of neighboring pixels in the source image is weighted and combined to get a single pixel value in the blurred image. Because the point-spread function is shift invariant in this case, the inner product amounts to a two-dimensional convolution.

where $\langle \cdot, \cdot \rangle$ denotes an inner product (two-dimensional over the n and p indices in this case) and $u[n', p']$ is the resulting blurred image.

In terms of the radar analogy, the resulting image $u[n', p']$ is the output of matched filtering performed on the received signal, the point-spread function $\chi_s[n, p; n', p']$ is the ambiguity function, and the source image $h[n, p]$ is the reflectivity of the target scene. The matched filter bank result was originally defined in Section 2.2; it is calculated from the received signal $y[m]$ according to

$$u[n', p'] = \sum_{m=0}^{M-1} s^*[m - p' + L - 1] e^{-2\pi i n'(m - p' + L - 1)/N} y[m], \quad (3.2)$$

where $s[l]$ represents the transmitted waveform and $y[m]$ represents the received signal. The ambiguity function $\chi_s[n, p; n', p']$ for the code $s[l]$ was also defined earlier, in Section 2.3; it describes how matched filtering takes a point in delay-frequency space and spreads it into

sidelobes:

$$\begin{aligned}\chi_s[n, p; n', p'] = \sum_{m=0}^{M-1} s[m - p' + L - 1] e^{2\pi i n'(m-p'+L-1)/N} \\ \times s^*[m - p + L - 1] e^{-2\pi i n(m-p+L-1)/N}. \quad (3.3)\end{aligned}$$

The target reflectivity $h[n, p]$ is something we have not yet encountered; it is a matrix of complex numbers that represents the magnitude and phase of the reflected signal (including the power gains and losses described by the radar equation (2.1)) as indexed by delay p and frequency n . In this way it is very much like a complex-valued image of delay-frequency space. Substituting these representations into the blurring equation, (3.1), will allow us to define a mathematical radar model.

3.1.2 Ambiguity Equation

First, though, it is convenient to simplify these equations by assigning an operator name, A^* , to matched filtering. By defining A^* as

$$A^*(y)[n, p] = \frac{1}{\sqrt{N}} \sum_{m=0}^{M-1} s^*[m - p + L - 1] e^{-2\pi i n(m-p+L-1)/N} y[m], \quad (3.4)$$

equation (3.2) can be written compactly as

$$u[n', p'] = \sqrt{N} A^*(y[m]). \quad (3.5)$$

As the notation indicates, A^* is a function that takes a one-dimensional vector of length M and produces a two-dimensional matrix with indices n and p of lengths N and P . The matched filter operator also has an adjoint or conjugate transpose, A , defined by the inner product identity:

$$\langle A(x), y[m] \rangle = \langle x[n, p], A^*(y) \rangle. \quad (3.6)$$

From this, we can write A on its own as

$$A(x)[m] = \frac{1}{\sqrt{N}} \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[m - p + L - 1] e^{2\pi i n(m-p+L-1)/N} x[n, p]. \quad (3.7)$$

Again following the notation, A is a function that takes a two-dimensional matrix as its argument and produces a one-dimensional vector indexed by m . The scaling factor of $1/\sqrt{N}$ is included in the definitions for both A^* and A because it ensures that the compound operator $AA^*(y)$ is the identity scaled by $\|s\|_2^2$. Accordingly, it is convenient to define scaled reflectivity coefficients $x[n, p]$ as

$$x[n, p] = \sqrt{N} h[n, p] \Leftrightarrow h[n, p] = \frac{1}{\sqrt{N}} x[n, p]. \quad (3.8)$$

The measurement ambiguity equation can be formed by substituting the radar representations for the image blurring terms back into equation (3.1):

$$\begin{aligned} \sqrt{N} A^*(y[m]) &= \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} \frac{1}{\sqrt{N}} x[n, p] \sum_{m=0}^{M-1} s^*[m - p' + L - 1] e^{-2\pi i n'(m-p'+L-1)/N} \\ &\quad \times s[m - p + L - 1] e^{2\pi i n(m-p+L-1)/N} \\ &= \sum_{m=0}^{M-1} s^*[m - p' + L - 1] e^{-2\pi i n'(m-p'+L-1)/N} \\ &\quad \times \frac{1}{\sqrt{N}} \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[m - p + L - 1] e^{2\pi i n(m-p+L-1)/N} x[n, p]. \end{aligned} \quad (3.9)$$

Further simplification is possible by writing the right-hand side in terms of the A and A^* operators:

$$\begin{aligned} A^*(y[m]) &= \frac{1}{\sqrt{N}} \sum_{m=0}^{M-1} s^*[m - p' + L - 1] e^{-2\pi i n'(m-p'+L-1)/N} A(x[n, p]) \\ &= A^*(A(x[n, p])). \end{aligned} \quad (3.10)$$

We call this the ambiguity equation since it relates the ambiguous matched filter output to the

artifact-free reflectivity coefficients, and the composed operator A^*A completely describes the resulting delay-frequency sidelobes. Since it is easy to overlook, note that we now have the following definitions in terms of the operators A and A^* :

$$\text{Matched filter bank applied to } y: \sqrt{N} \cdot A^*(y) \quad (3.11)$$

$$\text{Ambiguity function centered at } (n_0, p_0): N \cdot A^*A(\delta_{n_0, p_0}), \quad (3.12)$$

where $\delta_{n_0, p_0}[n, p]$ is the discrete delta function that returns 1 at $[n_0, p_0]$ and 0 otherwise.

3.1.3 The Radar Model

The ambiguity equation is close to what we desire for a radar model, but one further simplification can be made. Since both sides of the equation have matched filtering applied to them, A^* can be removed to arrive at a measurement equation:

$$y[m] = A(x[n, p]). \quad (3.13)$$

The measured radar signal $y[m]$ is now just a linear function of the scaled delay-frequency target reflectivity $x[n, p]$. That linear function, A , is the radar model. Interestingly, the radar model is the adjoint, or conjugate transpose, of the matched filter bank operation. This means that while the matched filter correlates the received signal with the expected return from point targets with different delays and frequency shifts, the radar model simulates the received signal as the sum of returns from point targets with different delays and frequency shifts. Though that mirror relationship is just something of a curiosity for now, it will have importance later when it comes to solving for the target reflectivity.

The radar model equation is useful because it provides the formalism needed for eliminating delay-frequency sidelobes: one can use the measured signal and radar model to solve for the target reflectivity. As formulated, the target reflectivity is precisely the matched filter result but with sidelobes removed. Unfortunately, it is not a straightforward process to just solve for the target reflectivity. The radar model is not invertible because equation (3.13) is under-determined, with more unknowns in the target reflectivity than equations provided by the measurements; this is a natural consequence of allowing for unknown frequency shifts.

Since the system of equations is under-determined, there are an infinite number of solutions for the target reflectivity function that would satisfy the equation and match the measurements. The matched filter result is one of those solutions, but it has serious shortcomings. Finding the best solution for the target reflectivity is the subject of Part III, where it will be shown that sparsity of the radar scene is key to eliminating sidelobes.

Fully written out, the radar model is:

$$y[m] = \frac{1}{\sqrt{N}} \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[m - p + L - 1] e^{2\pi i n(m-p+L-1)/N} x[n, p] \quad (3.14)$$

or equivalently

$$y[m] = \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[m - p + L - 1] e^{2\pi i n(m-p+L-1)/N} h[n, p]. \quad (3.15)$$

In the theory of vector spaces, the model A is a tight frame, conceptually the equivalent of an overcomplete basis for delay-frequency space. In fact, the model is an instance of a Gabor frame with window given by the code s , also known as a short time Fourier transform [Mallat, 2008]. This representation can be interpreted using the notion of a grid of point targets, which is visualized in Figure 3.2. The target reflectivity $h[n, p]$ can be thought to represent point targets distributed uniformly over delay-frequency space. Each point target can reflect the transmitted radar waveform independently with a given magnitude and phase. The radar model describes how reflections from each of these point targets would manifest in the measured signal, and it sums the contributions of each to produce the measurement. Again, this is just the mirror operation to what the matched filter does in its role as adjoint, assuming point targets in delay-frequency space and matching against the modeled reflection.

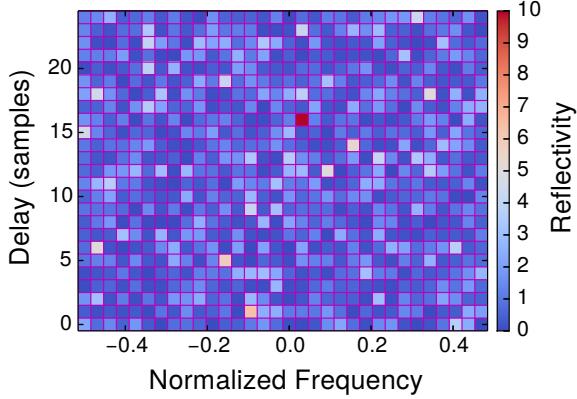


Figure 3.2 – Reflectivity coefficients as a delay-frequency grid of point targets. Each grid point represents a location in delay-frequency space indexed by delay p and frequency n . The color of each point corresponds to the amount of signal reflected at that particular delay and frequency shift, a value given by the magnitude of the target reflectivity coefficient $h[n, p]$. The radar scene is thus represented by a grid of independent point targets according to the radar model of equation (3.15).

3.2 Derivation of Radar Model

3.2.1 Narrow-band Received Signal

Though we now have a suitable radar model, it is instructive to derive it in another way that will allow us to relate the discrete reflectivity coefficients $h[n, p]$ to the complete reflectivity function. We begin with the narrow-band radar equation for (the complex envelope of) the baseband received signal from a fluctuating distributed target [Van Trees, 2001]:

$$y(t) = \int_{\rho_0}^{\rho_1} s(t - \rho) \tilde{a}\left(t - \frac{\rho}{2}, \rho\right) d\rho, \quad (3.16)$$

where $s(t)$ is the transmitted baseband modulation signal and $\tilde{a}(t, \rho)$ is the complex target reflectivity (magnitude and phase) as a function of time t and delay ρ . Note that this reflectivity function encompasses the power gained and lost according to the radar equation (2.1); it is not just the fraction of reflected power as described by the radar cross section σ . We have taken the integration interval to be $[\rho_0, \rho_1]$ to reflect the limited sampling window, and this carries with it the implicit assumption that $\tilde{a}(t - \rho/2, \rho) = 0$ outside that range. Whether one considers $\tilde{a}(t, \rho)$ to be random or deterministic does not matter at present. To

ease further derivation, we define a new function

$$\tilde{h}(t, \rho) = \tilde{a}\left(t + \frac{\rho}{2}, \rho\right) \quad (3.17)$$

so that the time variable represents the scattering for a signal *sent* at time t (which arrives at the target at $t + \rho/2$) rather than one reaching the target at time t . This results in

$$y(t) = \int_{\rho_0}^{\rho_1} s(t - \rho) \tilde{h}(t - \rho, \rho) d\rho. \quad (3.18)$$

3.2.2 Discretization

The next step is to begin discretizing the model. We assume that the received signal is sampled at a uniform rate τ_s starting at time t_0 and ending at t_1 (inclusive), so that $y(t)$ is represented by a complex discrete sequence

$$y[m] = y(m\tau_s + t_0) \quad m = 0, \dots, M - 1 \quad M = (t_1 - t_0)/\tau_s + 1. \quad (3.19)$$

In addition, we restrict our attention to discrete phase-modulated signals with L bauds and a baud length of τ_b , so that

$$s(t) = s[l] \quad \text{for } l\tau_b \leq t < (l+1)\tau_b, \quad l = 0, \dots, L - 1 \quad (3.20)$$

for a complex sequence $s[l] \in \mathbb{C}$. For ease of notation, let us also infinitely extend the sequence $s[l]$ by letting its value for non-existent indices be zero, $s[l] = 0$ for $l \neq 0, \dots, L - 1$. If we assume that the sampling time is an integer multiple of the baud length, then we have $\tau_s = R\tau_b$ where R is the under-sampling ratio. Note that if the reverse is true and over-sampling by an integer ratio is performed, we can simply duplicate the modulation sequence by the over-sampling ratio and let $R = 1$. Finally, the time interval $[t_0, t_1]$ can be chosen to correspond to the delay interval $[\rho_0, \rho_1]$ by requiring $t_0 = \rho_0 + L\tau_b$ and $t_1 = \rho_1$. This shrinks the sampling window to ensure that all returns from outside of the delay window have no contribution to the samples (each delay affects multiple later measurements because of the length of the code). Because of this relationship, we know that the delay

integration window can be divided into $(\rho_1 - \rho_0)/\tau_b = R(M-1) + L \equiv P$ equal segments of size τ_b . With these assumptions, we can break up the integral in equation (3.18) as follows:

$$\begin{aligned}
 y[m] &= y(m\tau_s + t_0) \\
 &= \int_{\rho_0}^{\rho_1} s(m\tau_s + t_0 - \rho) \tilde{h}(m\tau_s + t_0 - \rho, \rho) d\rho \\
 &= \sum_{p=0}^{RM+L-R-1} \int_{\rho_0+p\tau_b}^{\rho_0+(p+1)\tau_b} s(Rm\tau_b + L\tau_b + \rho_0 - \rho) \tilde{h}(Rm\tau_b + L\tau_b + \rho_0 - \rho, \rho) d\rho \\
 &= \sum_{p=0}^{P-1} s[Rm - p + L - 1] \int_{p\tau_b}^{(p+1)\tau_b} \tilde{h}(Rm\tau_b + L\tau_b - \rho, \rho + \rho_0) d\rho. \tag{3.21}
 \end{aligned}$$

3.2.3 Simplification

Now comes a key observation: $s(t)$ is nonzero only over $0 \leq t < L\tau_b$, and $\tilde{h}(t, \rho)$ is evaluated over the same values in equation (3.18). Thus, the result of the model will be the same if we replace $\tilde{h}(t, \rho)$ with its Fourier series representation on the interval $0 \leq t \leq N\tau_b$, where $N \geq L$ and $\nu_0 = 1/(N\tau_b)$ is the smallest frequency component. The Fourier series is given by

$$\tilde{h}(t, \rho) = \sum_{k=-\infty}^{\infty} h_k(\rho) e^{2\pi i k \nu_0 t} \quad \text{for } 0 \leq t \leq 1/\nu_0, \tag{3.22}$$

where

$$h_k(\rho) = \nu_0 \int_0^{1/\nu_0} \tilde{h}(t, \rho) e^{-2\pi i k \nu_0 t} dt \tag{3.23}$$

defines the Fourier coefficients. Substituting the Fourier series representation into equation (3.21) yields

$$\begin{aligned}
 y[m] &= \sum_{p=0}^{P-1} s[Rm - p + L - 1] \\
 &\times \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_k(\rho + \rho_0) e^{2\pi i k (Rm + L)/N} e^{-2\pi i k \nu_0 \rho} d\rho. \tag{3.24}
 \end{aligned}$$

This seems to have made the model more complicated for no benefit: there are now three discrete parameters and an infinite sum to boot. Notice, however, that $e^{2\pi ik(Rm+L)/N}$ as a function of k is periodic with period N . Re-parameterizing the infinite sum leads to

$$\begin{aligned} y[m] &= \sum_{p=0}^{P-1} s[Rm - p + L - 1] \\ &\quad \times \sum_{n=0}^{N-1} e^{2\pi in(Rm+L)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_{n+kN}(\rho + \rho_0) e^{-2\pi i(n+kN)\nu_0\rho} d\rho. \end{aligned} \quad (3.25)$$

Now the result of the integral is indexed by n and p where n parameterizes the frequency domain and p parameterizes the delay domain. However, it doesn't quite match the earlier radar model definition in equation (3.15). Factoring out the required terms produces

$$\begin{aligned} y[m] &= \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[Rm - p + L - 1] e^{2\pi in(Rm-p+L-1)/N} \\ &\quad \times e^{2\pi in(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_{n+kN}(\rho + \rho_0) e^{-2\pi i(n+kN)\nu_0\rho} d\rho. \end{aligned} \quad (3.26)$$

The form of this equation matches, and thus we define the discrete reflectivity coefficients as

$$h[n, p] = e^{2\pi in(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_{n+kN}(\rho + \rho_0) e^{-2\pi i(n+kN)\nu_0\rho} d\rho \quad (3.27)$$

and arrive once again at the discrete linear radar model:

$$y[m] = \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[Rm - p + L - 1] e^{2\pi in(Rm-p+L-1)/N} h[n, p]. \quad (3.28)$$

The only difference this time is the generalization to include undersampling of the transmitted waveform by a factor of R .

3.2.4 Assumptions

To sum up the derivation, recall the assumptions that underpin this model. We started with the narrow-band radar equation, which assumes that the bandwidth of the signal is much smaller than the baseband frequency so that the Doppler effect can be approximated only by frequency shifts and not time dilation. We also made the standard assumption that there is no scattering from outside the ranges parametrized by the model. In terms of the discretization, we assumed that the received signal is uniformly sampled at a fixed rate and that the transmitted waveform is piecewise constant over uniform intervals. We also required a uniformly-spaced frequency index with a number of frequency steps N greater than the number of bauds L in the transmitted waveform. All of these assumptions are congruous with standard radar practices and none are overly restrictive. From this we get an exact representation of the complete target reflectivity function as a discrete set of reflectivity coefficients $h[n, p]$.

3.3 Representation of Targets

3.3.1 Deterministic Reflectivity

The radar model allows us to analyze measurements in terms of delay-frequency reflectivity coefficients $h[n, p]$, but it is unclear as of yet how these coefficients embody actual radar target scenes. It is easy to think of the coefficients as representing a grid of point targets with different delays and frequency shifts, but this is highly unlikely to be the physical reality. Fortunately, the radar model derivation provides an exact relationship between the reflectivity coefficients and the complete reflectivity function, equation (3.27), that we can use to provide an intuitive interpretation. As shown in Appendix A, the reflectivity coefficient equation can be simplified to

$$h[n, p] = \int_{p\tau_b}^{(p+1)\tau_b} [h(f, \rho + \rho_0) e^{-2\pi i f(\rho - (p+1)\tau_b)} * b_{N, \tau_b}(f)](n\nu_0) d\rho, \quad (3.29)$$

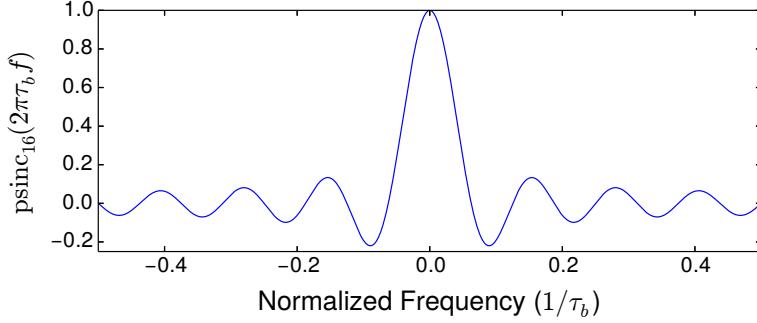


Figure 3.3 – Periodic sinc. The periodic sinc function, shown here for $N = 16$, results from taking the discrete-time Fourier transform of the N -sample wide rect function. It is periodic over the width of the displayed frequency window, $1/\tau_b$.

where the new term $b_{N,\tau_b}(f)$ is a frequency-blurring kernel involving the periodic sinc function. This kernel is defined as

$$b_{N,\tau_b}(f) = e^{-\pi i \tau_b f(N-1)} \text{psinc}_N(2\pi\tau_b f) \quad (3.30)$$

$$= \frac{1}{N} e^{-\pi i \tau_b f(N-1)} \frac{\sin(\pi N \tau_b f)}{\sin(\pi \tau_b f)}, \quad (3.31)$$

and the periodic sinc function $\text{psinc}_N(2\pi\tau_b f)$ is plotted in Figure 3.3 for reference. If the reflectivity function $h(f, \rho)$ is taken to be deterministic, the interpretation of equation (3.29) is straightforward: in the Doppler frequency variable n , the coefficients represent samples from the reflectivity function frequency spectrum $h(f, \rho)$ after it has been “smeared” by $b_{N,\tau_b}(f)$ through convolution; in the time delay variable p , the coefficients represent an integration of the reflectivity function over the corresponding delay window.

3.3.2 Probabilistic Reflectivity

Concerning the more general case of a probabilistic reflectivity function, we invoke the doubly-spread target model of *Van Trees* [2001]: assume that the reflectivity function $h(f, \rho)$ is a zero-mean complex Gaussian random variable with autocovariance given by

$$\mathbb{E}[h(f, \rho) h^*(f', \rho')] = S(f, \rho) \delta(f - f') \delta(\rho - \rho'), \quad (3.32)$$

where $E[\cdot]$ denotes expected value and $S(f, \rho)$ is the scattering function describing the returned *power* as a function of frequency and range. The process is zero-mean because the phase of $h(f, \rho)$ is assumed to follow a uniform distribution that is independent of the magnitude. In practical terms, the main difference between the deterministic model and the doubly-spread model is the phase of the returned signal: the former model produces phases that are a fixed function of range, while the latter model produces random phases with respect to range. Which of these is appropriate will depend upon the application; the deterministic model is best for simple targets with a known form, while the probabilistic model is best for targets that vary over time or present unpredictably complex scattering. Fortunately, both formulations fit equally well with the radar model. In terms of the reflectivity coefficients, the doubly-spread target model results in

$$E[h[n, p]h^*[n, p]] = \int_{p\tau_b}^{(p+1)\tau_b} [S(f, \rho + \rho_0) * B(f)](n\nu_0) d\rho \quad (3.33)$$

with

$$B(f) = \text{psinc}_N^2(2\pi\tau_b f) \quad (3.34)$$

$$= \frac{1}{N^2} \frac{\sin^2(\pi N\tau_b f)}{\sin^2(\pi\tau_b f)}. \quad (3.35)$$

In this case, the relationship between the coefficients and the actual quantity of interest $S(f, \rho)$ is even clearer: on average, the coefficients $|h[n, p]|^2$ give the total power returned by the target from ranges $p\tau_b \leq \rho - \rho_0 < (p+1)\tau_b$ and frequencies weighted by the squared periodic sinc $B(n\nu_0 - f)$. Both the deterministic and doubly-spread target models lead to similar interpretations for the reflectivity coefficients; the difference is in how the phases of the returned signals add up.

3.3.3 Reflectivity Coefficient Sparsity

With the deterministic and probabilistic interpretations of the reflectivity coefficients in hand, we can finally address a matter of vital importance: under what conditions are the

reflectivity coefficients sparse? Recall that radar scenes are typically sparse, with scatterers occupying only a small portion of delay-frequency space. Thus, the complete reflectivity function $h(f, \rho)$ can be assumed to have sparse support. The hope is that this sparsity translates into sparsity of the reflectivity coefficients. If it does, we can use that prior knowledge to find the sparsest solution for $h[n, p]$ that matches the measurements according to the radar model and have confidence that it is the true sidelobe-free solution.

To illustrate the reflectivity coefficient interpretations and analyze the coefficient sparsity for the sparsest possible target, consider the case of a point target with a reflectivity of A initially at range r and traveling toward the radar with a range rate v . The target's time delay and Doppler frequency shift are given by $\rho_t \approx \frac{2r}{c}$ and $f_t \approx \frac{2v}{c}f_0$ respectively, where c denotes the speed of light and f_0 is the baseband radar frequency. The appropriate target reflectivity function is

$$\tilde{h}(t, \rho) = Ae^{2\pi if_t t} \delta(\rho - \rho_t) e^{2\pi if_t \rho} e^{-2\pi if_0 \rho_t} \quad (3.36)$$

or equivalently

$$h(f, \rho) = A\delta(f - f_t)\delta(\rho - \rho_t)e^{2\pi if_t \rho}e^{-2\pi if_0 \rho_t}. \quad (3.37)$$

Plugging this into equation (3.29) shows how the discrete model represents a point target:

$$h[n, p] = \begin{cases} A b_{N, \tau_b}(n\nu_0 - f_t) e^{2\pi i f_t (\rho_0 + (p+1)\tau_b)} e^{-2\pi i f_0 \rho_t} & p = p_t \\ 0 & p \neq p_t, \end{cases} \quad (3.38)$$

where p_t gives the delay index which satisfies $p_t \tau_b \leq \rho_t - \rho_0 < (p_t + 1)\tau_b$. In order to look at sparsity, it will be easier to visualize the magnitude of the reflectivity coefficients. This is given by

$$\left| h[n, p_t] \right| = \frac{A}{N} \left| \frac{\sin(\pi N \tau_b (n\nu_0 - f_t))}{\sin(\pi \tau_b (n\nu_0 - f_t))} \right|. \quad (3.39)$$

Example point target reflectivity coefficients given by equation (3.39) are shown in Figure 3.4. Notice that if $n_t \nu_0 = f_t$ for some integer n_t , then all of the coefficients are zero except

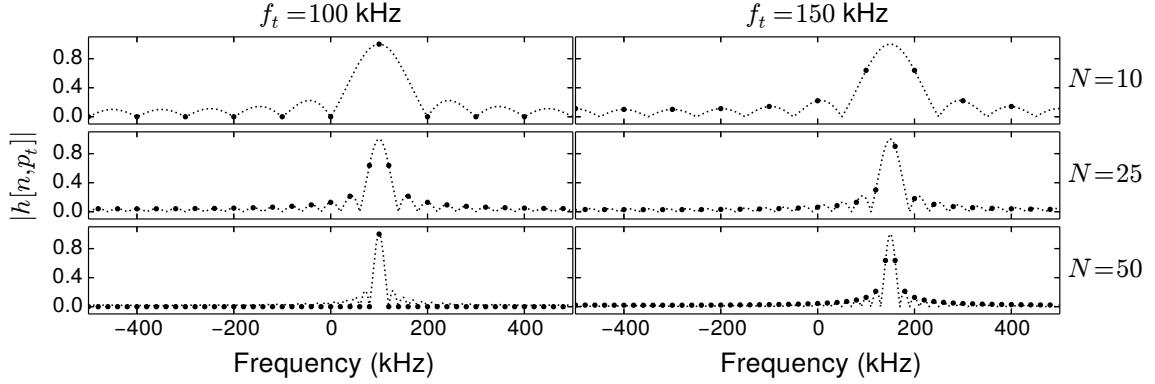


Figure 3.4 – Reflectivity coefficients for point targets. The reflectivity coefficient magnitudes are given by equation (3.39) with $A = 1$ and $\tau_b = 1 \mu\text{s}$ as a function of frequency for six point target cases. The left column shows the values for a point target with Doppler frequency shift of 100 kHz, while the right column shows values for a Doppler shift of 150 kHz. The first row shows the coefficient values when 10 frequencies are included in the discrete model, the second row shows values for 25 frequencies, and the third row shows values for 50 frequencies. The dotted line in each plot shows the underlying curve from which the coefficients are sampled.

for $h[n_t, p_t]$ and target sparsity translates directly into coefficient sparsity. However, in general all of the frequency coefficients corresponding to the correct range will be nonzero, and the question then becomes one of degree: are there few enough significant coefficients to say that they are sparse? In practical terms, “significant” means being above the noise level, since smaller values can be taken as zero and the reconstruction error will still be acceptable.

Figure 3.4 suggests one way of ensuring that the coefficient sparsity emulates the reflectivity function sparsity for point targets and by extension any general target: increasing N , the number of frequencies included in the discrete model. As N increases, the discretization effects become more localized and the values $h[n, p_t]$ for the point target look more like direct frequency samples from $h(f, \rho_t)$. This is a consequence of the frequency-smearing interpretation of the relationship between the reflectivity function and coefficients. Since the periodic sinc becomes more delta-like as N increases, the smearing it produces decreases. It is important to note that this is the same problem and solution encountered when relating the discrete Fourier transform of a sampled function to the complete function’s Fourier

transform. So we conclude that in general with N large enough, sparsity of $h(f, \rho)$ for any target translates directly to sparsity of $h[n, p]$.

Part III

Waveform Inversion

Chapter 4

Sparse Solutions of Underdetermined Systems

The method of least squares has been used for fitting models to data since Gauss invented the normal distribution and used it to predict the orbits of celestial bodies [Stigler; 1981]. With the law of large numbers ensuring that measurements often have Gaussian error, least squares serves its role well when no other prior information can be employed. This chapter is about the case when that isn't true, when we know something about the form of the solution itself: it is sparse. Images and audio are sparse; radar scenes are sparse; the brain uses sparsity to make sense of the world. Whenever something looks complex but is actually simple in the right form, that's sparsity. Across many applications, sparsity serves as a useful representation of prior knowledge.

The use of sparsity for inversion or fitting falls under a number of umbrella terms: compressed sensing, sparse approximation, low-rank decomposition, l_1 -regularization, and many similar variants. Each of these means something slightly different, but the machinery behind them is the same. In all cases, a simple or sparse representation is sought for some data, and the key to getting it is a nonlinear method. The solution methods are more diverse than even the number of applications for them, and the list continues to grow. For our purposes, we will focus on the theory of compressed sensing because it provides good guidance for what kind of approaches will work best. We will also focus on solution methods based on convex optimization and particularly the prox operator, since these are easy

to understand, flexible, have good theoretical underpinnings, and perform well in practice [Parikh and Boyd, 2013]. However, this choice is still somewhat arbitrary; the approach is the important part, and not so much the particular methods used.

4.1 Underdetermined Systems of Equation

When performing model inversion, a common challenge is solving an underdetermined system of equations. In this situation, one has a matrix/vector equation such as

$$y = Ax, \quad (4.1)$$

where y is a vector of length M representing measurements, A is a matrix that encompasses the mathematical model of the system, and x is a vector of length N giving unknown parameters. The equation is underdetermined because there are fewer measurements in y than unknowns in x , so there are an infinite number of x vectors that satisfy the equation. This situation can arise in many ways, but in general we can think of the underdetermined system as representing the decomposition of a signal y using an overcomplete dictionary. The columns of the matrix A , denoted a_n , are the atoms of the dictionary, and the entries of x are the coefficients corresponding to those atoms:

$$y = \sum_{n=0}^{N-1} x[n]a_n. \quad (4.2)$$

Since $M < N$, the dictionary atoms cannot form an orthogonal set, and there are many ways to form the decomposition [Mallat, 2008]. “Solving” the underdetermined system involves finding a preferable decomposition, one that at least has the qualities of the original measurement-generating solution when they are not the same.

4.1.1 Least Norm Solution

A common and linear solution method to an underdetermined system of equations is to take a least norm approach. Of the infinite number of solutions, it is simple to find the one with

the smallest l_2 -norm by performing the optimization

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \|x\|_2 \\ & \text{subject to} \quad y = Ax, \end{aligned} \tag{P}_2$$

where $\|x\|_2 = \sum_k |x[k]|^2$ denotes the l_2 -norm. When A is full rank (none of the measurements are redundant), AA^* is invertible and the solution to the least norm problem is

$$x_* = A^*(AA^*)^{-1}y. \tag{4.3}$$

For situations when the measurements are noisy, one can turn to the related l_2 -regularized least squares problem, also known as Tikhonov regularization:

$$\underset{x}{\text{minimize}} \quad \|y - Ax\|_2 + \lambda \|x\|_2, \tag{P}_{2\lambda}$$

where λ is the regularization parameter. This problem is solved by calculating

$$x_* = (A^*A + \lambda I)^{-1}A^*y, \tag{4.4}$$

where I is the identity matrix. As $\lambda \rightarrow 0$, the solution to l_2 -regularized least squares converges to the least norm solution. Since these solutions are the result of basic matrix operations and are therefore linear, they are often favored for their ease of computation [Boyd, 2008, Lecture 8].

4.1.2 Sparsest Solution

Motivated by the inherent simplicity of answers to many practical problems, another approach to selecting a particular solution of an underdetermined system is to choose the sparsest solution. This problem can be written using the l_0 pseudo-norm, which is just the count of the number of nonzero entries in a vector:

$$\|x\|_0 = \#\{k : x[k] \neq 0\}. \tag{4.5}$$

With this notation, the sparsest solution is found by the optimization problem

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_0 \\ & \text{subject to} && y = Ax. \end{aligned} \tag{P}_0$$

Unfortunately, this problem is NP-hard in general [Natarajan, 1995], requiring combinatorial optimization to solve. Essentially, one has to check each sparsity pattern (location of zeros/nonzeros) for a solution until the sparsest one is found.

Fortunately, the convex relaxation of this problem, which is easy to solve, provides the sparsest solution in most cases. Instead of minimizing the l_0 pseudo-norm, one minimizes the l_1 -norm:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 \\ & \text{subject to} && y = Ax, \end{aligned} \tag{P}_1$$

where $\|x\|_1 = \sum_k |x[k]|$ denotes the l_1 -norm. Then for the overwhelming majority of matrices A with columns normalized to one, there is a constant $\rho(M, N) > 0$ such that if x_0 is a solution to $y = Ax_0$ with at most ρM nonzeros, that solution is both the sparsest possible solution and the minimal l_1 -norm solution [Donoho, 2006b]. In other words, whenever there is a sufficiently sparse solution, l_1 minimization will find it. The proof by Donoho [2006b] does not provide a practically useful formula for the breakdown point ρ , but it does provide motivation for why an equivalence of this kind must exist. It is easy to see how the l_1 -norm promotes sparse solutions by equally penalizing both large and small nonzero values of x ; equivalence comes from the rarity of truly sparse x that actually solve the underdetermined system, causing l_1 -minimization to pick out the true solution due to lack of other possibilities.

4.2 Compressed Sensing

Compressed sensing [Donoho, 2006a; Candès and Wakin, 2008] expands on the theory for solving underdetermined systems of equations when the solution is known to be sparse. It provides guarantees for recovering the true solution to within noise bounds provided that the measurements are “incoherent” and that their number is on the order of the solution sparsity.

In essence, one only needs to measure at the information rate (sparsity) of the solution rather than fully sample the entire space of solutions. Not only that, but these optimal non-adaptive measurements, pre-determined and embodied in a measurement matrix, are known to perform nearly as well as the best adaptive measurements [Donoho, 2006a]. The key is sparsity, which acts as the prior information that enables narrowing in on the one true answer out of the infinite potential answers.

4.2.1 Conditions

The main requirement for invoking the theory of compressed sensing is that the sought signal is sparse, or at least well-approximated by a sparse representation. In terms of a specified dictionary, sparsity requires that the signal be well-represented by a relatively small number of coefficients corresponding to the dictionary atoms. A signal might not be sparse in its natural representation, but that does not mean that it is not sparse in some other representation. Images are generally not sparse when specified by individual pixel values, but they are almost always sparse in a Fourier or wavelet dictionary. The only instance where this is not true is for images with high entropy; they are random and have no structure [Mallat, 2008]. Even if a signal x is not exactly sparse, we say it has an S -sparse approximation if $\|x - x_S\|_2$ is small, where x_S is the vector of coefficients $x[n]$ with all but the largest S entries set to zero. As the success of lossy compression schemes demonstrate, many signals of interest satisfy at least this compressibility condition.

Measurement incoherence is the second condition for applying the theory of compressed sensing [Candès and Wakin, 2008]. Consider the simplest case of a 1-sparse signal x_1 measured as $y = Ax_1$, with the number of measurements M much less than the dimension of the signal N . At least one of the measurements must include a contribution from the nonzero entry of x_1 or else recovery from y is impossible. Since we don't know *a priori* which element is the nonzero one and because $M \ll N$, the only way to achieve a high probability of recovery is to have measurements that individually span a large portion of the unknown signal space. Each measurement must be global in some sense, embodying most of the unknown coefficients. This concept is known as measurement incoherence: none of the measurements are very concentrated (coherent) in the unknown space.

In the mathematical theory, incoherence can be guaranteed by a number of different but related conditions. Often it is theoretically convenient to allow the measurements to be random. For an orthogonal set of possible measurements given by the matrix U with scaling $U^*U = N \cdot I$, we can let the measurement matrix A be a random subset of M rows of U . The coherence $\mu(U)$ of the measurement set is then defined as

$$\mu(U) = \max_{k,n} |U[k, n]|^2, \quad (4.6)$$

and it varies in value from a minimum of 1 to a maximum of N [Candès and Romberg, 2007]. Thus, the measurements are incoherent under this framework when the set from which they are drawn has $\mu(U)$ close to 1. The coherence parameter can also be defined for a wider class of probabilistic measurement ensembles, including Gaussian and binary random measurements, as shown by Candès and Plan [2011].

Deterministic measurements can have their incoherence measured by the restricted isometry property (RIP) introduced by Candès and Tao [2006] and expanded upon in Candès, Romberg, and Tao [2006]. Let A_T with $T \subset \{0, \dots, N-1\}$ denote a subset of the columns of the measurement matrix A . Then the S -restricted isometry constant δ_S of A is defined as the smallest number such that

$$(1 - \delta_S)\|x_S\|_2^2 \leq \|A_T x_S\|_2^2 \leq (1 + \delta_S)\|x_S\|_2^2 \quad (4.7)$$

for all S -sparse vectors x_S and all subsets T with $|T| \leq S$. The restricted isometry property is satisfied and sufficient incoherence assured for a solution with sparsity S when δ_{2S} is small enough. This condition ensures that the measurements preserve the distance between any two S -sparse solutions, guaranteeing that any S -sparse answer is unique.

4.2.2 Theoretical Bounds

If the signal is compressible and incoherent sampling is performed, then essentially we know that each measurement contains a contribution from each of the S coefficients. Intuitively, we might then expect to be able to reconstruct the signal from about S measurements. This idea is made concrete with the incoherent sampling theorem presented below

for reconstruction via the l_1 -minimization problem (P_1). No matter the specific setting, the principle of this theorem remains the same: if the sampling is incoherent and we solve an appropriate convex optimization problem, we can reconstruct a signal to within noise and S -sparse approximation errors by using on the order of $S \log N$ measurements.

In the case where we know the coherence $\mu(U)$ of the measurements, the incoherent sampling theorem requires a number of measurements M that meets the condition $M \geq C \cdot \mu(U) \cdot S \cdot \log N$ for a sparsity level S and constant C . In the case where measurement incoherence is guaranteed in terms of the restricted isometry property, we require $\delta_{2S} < \sqrt{2} - 1$ for a specified sparsity level S . For random Gaussian or binary matrices, the RIP requirement on δ_{2S} is satisfied with overwhelming probability when $M \geq C \cdot S \cdot \log(N/S)$. In either case, the recovery bound has the same form. Denoting the result of the l_1 -minimization problem as x_* , the true solution as x , and the best S -sparse approximation of x as x_S , we have

$$\|x_* - x\|_2 \leq C_0 \|x - x_S\|_2, \quad (4.8)$$

where C_0 is a small known constant. The coherence parameter result is due to *Candès and Plan* [2011], while the RIP result is due to *Candès* [2008]. Both articles specify values for the constants C and C_0 , but they are believed to be sub-optimal in general, hence the imprecision. If x is exactly S -sparse, this says that l_1 -minimization finds the exact sparse solution. When x is not S -sparse, the l_1 solution is nearly as close to the true solution as the best S -sparse approximation. In other words, the reconstruction is almost as good as the best possible given by a so-called oracle; we could hardly do better if we knew the locations of the S most significant elements of x ahead of time.

The incoherent sampling theorem can also be extended to cases with measurement error. Suppose that the measurements are now given by

$$y = Ax + \eta, \quad (4.9)$$

where η is an unknown noise term. In this setting, we are interested in the solution to the

basis pursuit with denoising problem:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 \\ & \text{subject to} && \|y - Ax\|_2 \leq \epsilon, \end{aligned} \tag{P_{1\epsilon}}$$

where ϵ is a chosen noise bound. Then under the same restricted isometry condition $\delta_{2S} < \sqrt{2} - 1$ as before, the solution x_* to $(P_{1\epsilon})$ satisfies

$$\|x_* - x\|_2 \leq C_0 \|x - x_S\|_2 + C_1 \epsilon, \tag{4.10}$$

where C_0 and C_1 are small known constants with similar interpretations as before [Candès, 2008]. This remarkable result shows that l_1 -recovery is robust to noise; the error in the answer is still bounded only by the approximation error and the noise level. A similar result is given by *Candès and Plan* [2011] for the case when only the coherence $\mu(U)$ is known, although the noise term is handled a little differently.

4.2.3 Recovery Phase Transition

It may seem unlikely that adding noise to the problem does not dramatically worsen recovery of the true solution, but amazingly this has proven to be true in a growing list of cases. *Donoho, Maleki, and Montanari* [2011] completely characterize the phase transition from failed recovery to successful solution over the space of (δ, ρ) with $\delta = M/N$ and $\rho = S/M$ for measurement matrices A whose entries are i.i.d. Gaussian. For a given undersampling ratio δ , recovery is successful with overwhelming probability for ρ below the curve $\rho_{\text{MSE}}(\delta)$, and it fails with overwhelming probability otherwise. The phase transition curve is shown in Figure 4.1. For these results, successful recovery means that the measurement error is bounded for the worst-case noise and sparsity distributions, while failure means that the error is unbounded. The theoretical results are for the limit as M , N , and S tend to infinity with fixed (δ, ρ) , but simulations have shown that the transition holds for finite setups.

It must be emphasized that this phase transition represents the behavior of the *optimal* recovery algorithm when faced with the parameters of the problem. The key to the success of compressed sensing is that l_1 minimization *is* an optimal recovery algorithm; it achieves

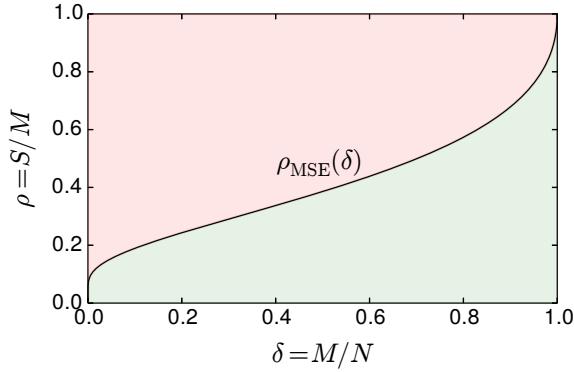


Figure 4.1 – Phase transition for recovery of sparse solution. For noisy measurements $y \in \mathbb{R}^M$ given by $y = Ax + \eta$, recovery of S -sparse $x \in \mathbb{R}^N$ by is successful (mean square error is bounded) when $\rho = S/M$ is below the curve $\rho_{\text{MSE}}(\delta)$ shown above with $\delta = M/N$. The result holds for many types of measurement matrices A and nonlinear recovery algorithms, particularly l_1 minimization.

this performance with appropriate choice of the noise parameter [Donoho and Tanner, 2009; Donoho, Maleki, and Montanari, 2011]. Furthermore, the optimal performance has been shown to extend empirically to other measurement ensembles [Donoho and Tanner, 2010; Monajemi et al., 2013] and theoretically to other optimization algorithms [Donoho, Johnstone, and Montanari, 2013]. Research in this area is ongoing and advancing rapidly, and it is expected that the recovery bounds and phase transition results apply to a wider class of measurement ensembles and recovery algorithms than have so far been proven. Along this line, Strohmer [2012] gives an overview of the open problems that remain to be solved.

4.3 Convex Optimization

Sparse solutions to an underdetermined system of equations can be sought with a bevy of different methods. Some of these are greedy algorithms like OMP (orthogonal matching pursuit) [Tropp and Gilbert, 2007], some are based on statistics like AMP (approximate message passing) [Donoho, Maleki, and Montanari, 2009], and most are variants that improve on a basic method in some way. Of particular note are the many methods that involve

a common but not universal theme: convex optimization, specifically involving minimization of the l_1 -norm. Our focus is on methods of this type because they are easy to understand and the theory is well-established. The base case is an optimization problem known as l_1 -regularized least squares:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1. \quad (\text{P}_{1\lambda})$$

This problem is equivalent to basis pursuit with denoising ($\text{P}_{1\epsilon}$) for a proper choice of the regularization parameter λ , but its form is more convenient for many algorithms.

4.3.1 First-order Methods

Before discussing l_1 -regularized least squares specifically, we restrict our attention to a particular class of optimization algorithms: first-order methods. For the types of problems that we are interested in, explicit representations of the measurement matrix A are not useful. The typical matrix is large and has additional structure that makes applying it via matrix multiplication inefficient. When A represents convolution, a Fourier transform, or other efficient transforms as is often the case, it is much faster to compute its result using the transform operation directly. First-order optimization algorithms are interesting because they do not require using an explicit matrix representation for A and A^* ; they only require the ability to evaluate the A and A^* operations, which we can do efficiently even for large problem dimensions.

The most natural first-order method for minimizing a differentiable convex function $G(x)$ is gradient descent. Starting at a point x^0 , gradient descent iteratively steps in the direction of the negative of the gradient $\mathbf{grad}_G(x) = \nabla G(x)$ until the minimum is reached.

Algorithm 1 Gradient Descent [Boyd, 2004]

given a starting point x^0

repeat

$x^{k+1} := x^k - \mu \mathbf{grad}_G(x^k)$ {step size μ chosen by line search}

until stopping criterion is satisfied

Gradient descent is guaranteed to converge to a minimum-value solution for smooth, convex $G(x)$ provided that the step size is chosen appropriately. This is typically done with a backtracking line search as detailed by Boyd [2004]. For the function $G(x) = \frac{1}{2}\|Ax - b\|_2^2$, the gradient is $\mathbf{grad}_G(x) = A^*(Ax - b)$. Thus in order to iteratively find the least squares solution to $Ax = b$ using gradient descent, we only need to be able to evaluate A and A^* .

Of course, the function we want to minimize to be able to solve the l_1 -regularized least squares problem is not differentiable, so gradient descent is not directly applicable. For non-smooth functions $F(x)$ like the l_1 -norm, an attractive alternative to the gradient is the *proximal operator*:

$$\mathbf{prox}_{\mu F}(v) = \arg \min_x \left(F(x) + \frac{1}{2\mu} \|x - v\|_2^2 \right), \quad (4.11)$$

where μ is a scaling factor that can be likened to a step size. The prox operator adds an l_2 -smoothing term to a non-smooth function and outputs the vector x which gives the minimum of the composite function. Using the prox operator as an iterative step in a similar manner to the gradient leads to an algorithm for minimizing non-smooth $F(x)$: the proximal point method.

Algorithm 2 Proximal Point [Parikh and Boyd, 2013]

```

given a starting point  $x^0$ 
repeat
     $x^{k+1} := \mathbf{prox}_{\mu F}(x^k)$  {scaling  $\mu$  chosen by line search}
until stopping criterion is satisfied

```

Note that as the iterates x^k approach the minimum x_* , the smoothing term goes to zero and the minimum of $F(x)$ is reached.

4.3.2 Prox Operator Examples

For many interesting functions $F(x)$, the prox operator is surprisingly easy to solve. The prox operator of the l_1 -norm will obviously be useful, and fortunately it has a closed form

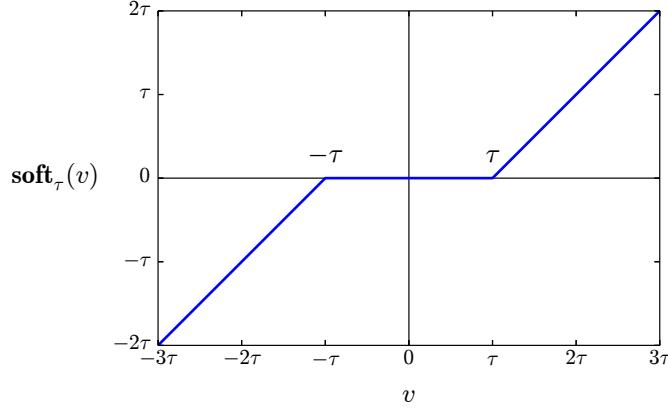


Figure 4.2 – Soft thresholding function. The real scalar soft thresholding function evaluates to zero when the magnitude of the input is below the threshold τ , and it returns a shrunken value of the input otherwise.

solution known as soft thresholding:

$$\mathbf{prox}_{\tau \|x\|_1}(v) = \mathbf{soft}_\tau(v)[k] = \begin{cases} \left(1 - \frac{\tau}{|v[k]|}\right)v[k], & |v[k]| > \tau \\ 0, & |v[k]| \leq \tau. \end{cases} \quad (4.12)$$

Soft thresholding operates element-wise on a vector v , setting all of the elements with magnitude less than the threshold τ to zero and shrinking all other elements toward zero by τ . A graphical depiction of scalar soft thresholding is given in Figure 4.2.

The l_2 -norm, though smooth, also has a closed form prox operator. For $F(x) = \|x\|_2$, the prox operator is the block soft thresholding operator,

$$\mathbf{prox}_{\tau \|x\|_2}(v) = \mathbf{bsoft}_\tau(v) = \begin{cases} \left(1 - \frac{\tau}{\|v\|_2}\right)v, & \|v\|_2 > \tau \\ 0, & \|v\|_2 \leq \tau. \end{cases} \quad (4.13)$$

Block soft thresholding operates on the entire vector v , setting it to zero if its norm is less than the threshold τ and shrinking its norm to $\|v\|_2 - \tau$ otherwise. The squared l_2 -norm $F(x) = \frac{1}{2}\|x\|_2^2$ has a different prox operator called the shrinkage function:

$$\mathbf{prox}_{\frac{\tau}{2}\|x\|_2^2}(v) = \mathbf{shrink}_\tau(v) = \frac{v}{1 + \tau}. \quad (4.14)$$

In contrast to soft thresholding, the shrinkage function shrinks all of the vectors' entries by a fixed fraction defined by the scale factor τ , penalizing the largest entries the most.

When the non-smooth function is the indicator $I_{\mathcal{C}}(x)$ for a closed nonempty convex set \mathcal{C} , which is defined as

$$I_{\mathcal{C}}(x) = \begin{cases} 0, & x \in \mathcal{C} \\ +\infty, & x \notin \mathcal{C}, \end{cases} \quad (4.15)$$

the prox operator acts as Euclidean projection onto \mathcal{C} :

$$\mathbf{prox}_{I_{\mathcal{C}}}(v) = \arg \min_x \left(I_{\mathcal{C}}(x) + \frac{1}{2} \|x - v\|_2^2 \right) = \arg \min_{x \in \mathcal{C}} \|x - v\|_2 = \Pi_{\mathcal{C}}(v). \quad (4.16)$$

When viewed in this way, prox operators can be seen as generalized projections. In the case of the l_1 -norm, soft thresholding “projects” onto a sparse space enforced by the l_1 -norm by setting small elements to zero. Practically, indicator functions are useful for enforcing constraints in optimization problems. By adding an appropriate indicator function to the objective, solutions can be required to lie in the corresponding convex set. For example, if a solution to an underdetermined system is known to have a particular sparsity pattern, the indicator function for the locations of the zeros can be added to a least squares objective to solve for the closest solution with that sparsity pattern.

4.3.3 Proximal Gradient

With the necessary operators in place, we finally return to the problem of l_1 -regularized least squares ($P_{1\lambda}$). This is a problem of the form

$$\underset{x}{\text{minimize}} \quad F(x) + G(Ax - b), \quad (4.17)$$

where $F(x)$ is (possibly non-differentiable) function with a known prox operator $\mathbf{prox}_F(v)$, A is a linear operator, b is a vector constant, and $G(z)$ is a differentiable function so that the gradient of the second term is $A^* \mathbf{grad}_G(Ax - b)$. With the split smooth and non-smooth objective, we can imagine a two-step iteration stage that first steps in the gradient direction of the smooth term and then evaluates the prox operator of the non-smooth term. This leads

to an optimization algorithm known as the proximal gradient method.

Algorithm 3 Proximal Gradient [Parikh and Boyd, 2013]

given a starting point x^0 , a step size $0 < \mu \leq 2/L$

repeat

$$z^{k+1} := x^k - \mu A^* \mathbf{grad}_G(Ax^k - b)$$

$$x^{k+1} := \mathbf{prox}_{\mu F}(z^{k+1})$$

until stopping criterion is satisfied

Stopping criteria for the algorithm vary, but we will choose one based on the residual of the optimality condition $0 \in \partial F(x_*) + \nabla_x G(Ax_* - b)$. This residual for the iterate x^{k+1} is given by

$$r^{k+1} = \frac{1}{\mu} (x^k - x^{k+1}) + A^* (\mathbf{grad}_G(Ax^{k+1} - b) - \mathbf{grad}_G(Ax^k - b)). \quad (4.18)$$

The algorithm is said to have converged if $\|r^{k+1}\|_2$ is small, less than a tolerance threshold. The basic proximal gradient method also uses a fixed step size $\mu > 0$, and to guarantee convergence it must satisfy $\mu \leq 2/L$ where L is the global Lipschitz constant of the gradient of $G(Ax - b)$. For all x, \tilde{x} in the domain of $\nabla_x G$, the gradient is globally Lipschitz continuous if

$$\|\nabla_x G(Ax - b) - \nabla_x G(A\tilde{x} - b)\|_2 \leq L\|x - \tilde{x}\|_2. \quad (4.19)$$

One interpretation of the proximal gradient method is as a majorization-minimization algorithm. At every step, the algorithm majorizes (upper bounds) the objective and then minimizes the majorization. In this case, the majorization involves taking a first-order approximation of the smooth portion of the objective and adding a trust region penalty:

$$G_\mu(x, \tilde{x}) = G(\tilde{x}) + \langle \nabla_x G(A\tilde{x} - b), x - \tilde{x} \rangle + \frac{1}{2\mu} \|x - \tilde{x}\|_2^2. \quad (4.20)$$

Then for $\mu \leq 1/L$, we have the majorization

$$F(x) + G(Ax - b) \leq F(x) + G(x^k) + \langle \nabla_x G(Ax^k - b), x - x^k \rangle + \frac{1}{2\mu} \|x - x^k\|_2^2 \quad (4.21)$$

about the point x^k . Minimizing the majorization is just a simple call to the prox operator of

F :

$$\begin{aligned} x^{k+1} &= \arg \min_x \left(F(x) + G(x^k) + \langle \nabla_x G(Ax^k - b), x - x^k \rangle + \frac{1}{2\mu} \|x - x^k\|_2^2 \right) \\ &= \arg \min_x \left(F(x) + \frac{1}{2\mu} \|x - (x^k - \mu A^* \mathbf{grad}_G(Ax^k - b))\|_2 \right) \\ &= \mathbf{prox}_{\mu F}(x^k - \mu A^* \mathbf{grad}_G(Ax^k - b)). \end{aligned} \quad (4.22)$$

This step is precisely the one taken in the proximal gradient method.

Proximal gradient applied to the l_1 -regularized least squares problem has a special name: iterative soft thresholding. If we divide the objective as $F(x) = \lambda \|x\|_1$ and $G(z) = \frac{1}{2}\|z\|_2^2$, the proximal operator $\mathbf{prox}_F(x) = \mathbf{soft}_\lambda(x)$ is soft thresholding, the gradient $\mathbf{grad}_G(z) = z$ is the identity, and the method takes the explicit form given in Algorithm 4.

Algorithm 4 Iterative Soft Thresholding

given a starting point x^0 , a step size $0 < \mu \leq 2/\|A\|_2^2$

repeat

$$z^{k+1} := x^k - \mu A^*(Ax^k - b)$$

$$x^{k+1} := \mathbf{soft}_{\mu\lambda}(z^{k+1})$$

until stopping criterion is satisfied

Clearly the iterative soft thresholding method is aptly named. With the form of the gradient known, we also know that its Lipschitz constant L is given by $\|A\|_2^2$, the squared induced l_2 -norm of the matrix A . If A is scaled so that its norm is unity, then the maximum fixed step size that guarantees convergence is 2.

4.3.4 Improvements on Proximal Gradient

The proximal gradient method can be improved both theoretically and in practice through the use of an acceleration term and a variable step size. Acceleration of first-order methods is based on the work of *Nesterov* [1983, 1988] for smooth objectives, which was extended by *Beck and Teboulle* [2009, 2010] to the proximal gradient method. *Nesterov* [2007] also

adapted acceleration to the case of a composite objective, but that particular approach requires multiple evaluations of the prox operator per step. Acceleration works by adding an appropriate ratio of the prior iterate to the current iterate before executing the gradient and prox steps, thus this simple modification improves convergence without significantly increasing per-step computation. More recently, *O'Donoghue and Candès* [2013] found that an adaptive restart scheme for the acceleration term improves convergence in practice even more significantly. Applied to the l_1 -regularized least squares problem, the accelerated method is known as FISTA, or the fast iterative soft thresholding algorithm [*Beck and Teboulle*, 2009].

A variable step size is the other modification that practically improves performance. Though convergence is guaranteed by bounding the steps by the global Lipschitz constant, such steps can be too conservative when the gradient's curvature is locally variable. When this is the case, even a standard backtracking line search for the step size is not sufficient since it can only decrease the step size. Both *Scheinberg, Goldfarb, and Bai* [2011] and *Becker, Candès, and Grant* [2011a], apparently independently, suggested an expansive backtracking line search scheme to accommodate increasing the step size if warranted. With this scheme, the step size is increased by a constant factor after every iteration; convergence is assured by checking the majorization condition at the next iterate and backtracking on the step size by a constant factor if necessary. Though backtracking results in repeated application of A , A^* , and the gradient and prox operators, the typical net result is fewer computations and faster convergence. With both the acceleration and variable step size enhancements, we get a proximal gradient method that significantly improves upon the performance of the basic method. The accelerated proximal gradient method, with adaptive restart and adaptive step size, is detailed as Algorithm 5.

4.3.5 Related Prox Methods

Naturally, there are other algorithms that solve the same (or a similar) split objective optimization problem and even make use of the prox operator. While proximal gradient requires part of the objective, $G(z)$, to be differentiable, these additional methods work in the more

Algorithm 5 Accelerated Proximal Gradient, with adaptive restart and adaptive step size

given starting point x^0
given initial step size $\mu^0 > 0$, step expansion factor $\alpha > 1$, backtracking factor $0 < \beta < 1$

given tolerance $\epsilon > 0$
 $x^{-1}, w^0 := x^0$
 $g^0 := A^* \mathbf{grad}_G(Aw^0 - b)$
 $t^0 := 1$
 $\mu^1 = \mu^0$
 $\gamma := 1$

repeat

- if** $\langle w^k - x^k, x^k - x^{k-1} \rangle > 0$ **then**
- $t^k := 1$
- end if**
- repeat**
- $t^{k+1} := \left(1 + \sqrt{1 + 4(t^k)^2\gamma}\right)/2$
- $\theta := (t^k - 1)/t^{k+1}$
- $w^{k+1} := x^k + \theta(x^k - x^{k+1})$
- if** $G(Ax^{k+1} - b) \leq G(Aw^{k+1} - b) + \langle x^{k+1} - x^k, g^{k+1} \rangle + \frac{1}{2\mu^{k+1}} \|x^{k+1} - w^{k+1}\|_2^2$ **then**
- break**
- else**
- $\mu^{k+1} := \beta\mu^{k+1}$
- $\gamma := \mu^k/\mu^{k+1}$
- end if**
- until**
- $r^{k+1} := (x^k - x^{k+1})/\mu^{k+1} + g^{k+1} - g^k$
- if** $\|r^{k+1}\|_2 \leq \epsilon$ **then**
- break**
- end if**
- $\mu^{k+1} := \alpha\mu^{k+1}$
- $\gamma = \mu^k/\mu^{k+1}$
- until**

Algorithm 6 Alternating Direction Method of Multipliers (ADMM)

given starting point x^0 with dual y^0 , penalty parameter $\rho > 0$
 $u^0 := \rho y^0$
 $z^0 := \text{prox}_{\rho G}(x^0 + u^0)$
repeat
 $x^{k+1} := \text{prox}_{\rho F}(z^k - u^k)$
 $z^{k+1} := \text{prox}_{\rho G}(x^{k+1} + u^k)$
 $r^{k+1} := x^{k+1} - z^{k+1}$
 $s^{k+1} := (z^k - z^{k+1})/\rho$
 $u^{k+1} := u^k + r^{k+1}$
until $\|r^{k+1}\|_2$ and $\|s^{k+1}\|_2$ are small

general case where only an easily-evaluated prox operator is required for each of the split-objective terms. Since this is a less-restrictive requirement, these methods are primarily of interest due to their broader applicability. Depending on the exact problem being solved, however, it is certainly possible for them to outperform even the accelerated proximal gradient method.

The alternating direction method of multipliers (ADMM) algorithm, also known as Douglas-Rachford splitting, solves the problem

$$\underset{x}{\text{minimize}} \quad F(x) + G(x) \quad (4.23)$$

where both F and G have prox operators given by $\text{prox}_F(x)$ and $\text{prox}_G(x)$. Notice that this definition does not include the affine transformation $Ax - b$ as with the proximal gradient method. The ADMM algorithm is similar to proximal gradient in that it splits the iteration into two steps, first applying $\text{prox}_F(x)$ and then applying $\text{prox}_G(x)$ as detailed in Algorithm 6. Amazingly, the algorithm converges for any positive value of the penalty parameter ρ . ADMM can vary in its exact form depending on which order the prox operators are applied or, equivalently, by how the objective is split and assigned to the two functions, so care must be taken to choose the optimal formulation for a given problem. A detailed treatment of ADMM, including per-iteration adjustment of the penalty parameter to help improve convergence, is given by *Boyd et al. [2011]*.

Linearized ADMM, also called the Split Inexact Uzawa method [*Esser, Zhang, and*

Chan, 2010], is a variant of ADMM that includes an affine transformation in the objective function with a matrix A and vector b :

$$\underset{x}{\text{minimize}} \quad F(x) + G(Ax - b), \quad (4.24)$$

where again F and G are not necessarily differentiable and have prox operators $\text{prox}_F(x)$ and $\text{prox}_G(z)$. It is not trivial to accommodate an affine transformation when evaluating the prox operator of most functions, so the explicit inclusion here is significant in practice and the reason linearized ADMM is often used over the original. Our version of linearized ADMM, given in Algorithm 7, is based on the one described by *Parikh and Boyd [2013]*. However, ours is unique because it includes not only a penalty adjustment scheme adapted from the ADMM method of *Boyd et al. [2011]* but also a novel expansive backtracking line search for the step size μ . In fact, the interpretation of the μ parameter as a step size is unique, and doing so allows us to adapt the majorization condition and line search scheme used in the accelerated proximal gradient algorithm (Algorithm 5) for use with linearized ADMM. We have not formally analyzed the convergence of linearized ADMM including these modifications or rigorously proved their benefits, but they do result in significant improvement in practice in a manner similar to the corresponding proximal gradient method schemes. As a point of reference, we have found that an expansion factor of $\alpha = 1.25$, a backtracking factor of $\beta = 0.5$, a maximum residual gap of $\gamma = 10$, and a penalty factor of $\rho = 1.5$ are the most effective parameters for a variety of problems.

The final algorithm of interest actually embodies a class of algorithms that include linearized ADMM as a special case: the primal dual hybrid gradient (PDHG) method. It solves the same problem as linearized ADMM, although that problem is sometimes written in the form

$$\min_x \sup_y F(x) + \langle y, Ax - b \rangle - G^*(y), \quad (4.25)$$

where G^* denotes the convex conjugate function of G . The conjugate function is defined as

$$G^*(y) = \sup_x \langle y, x \rangle - G(x), \quad (4.26)$$

and since $(G^*)^* = G$, the equivalence between the two minimization problems is clear.

Algorithm 7 Linearized ADMM

given starting point x^0 with dual y^0
given step size $\mu > 0$, step expansion factor $\alpha > 1$, backtracking factor $0 < \beta < 1$
given penalty parameter $\rho > 0$, maximum residual gap $\gamma > 1$, penalty factor $\tau > 1$
given penalty freezing index $k_{\rho_{\max}}$

$$u^0 := \rho y^0$$

$$u^{-1} := u^0$$

repeat

repeat

$x^{k+1} := \text{prox}_{\mu F}\left(x^k - \frac{\mu}{\rho} A^* (2u^k - u^{k-1})\right)$

if $\mu \|A(x^{k+1} - x^k)\|_2^2 \leq \rho \|x^{k+1} - x^k\|_2^2$ **then**

break

else

$\mu := \beta \mu$

end if

until

$z^{k+1} := \text{prox}_{\rho G}(Ax^{k+1} - b + u^k)$

$r^{k+1} := Ax^{k+1} - b - z^{k+1}$

$u^{k+1} := u^k + r^{k+1}$

$s^{k+1} := A^*(u^{k+1} - u^k)/\rho + A^*(u^{k-1} - u^k)/\rho + (x^k - x^{k+1})/\mu$

if $k < k_{\rho_{\max}}$ **then**

if $\|r^{k+1}\|_2 > \gamma \|s^{k+1}\|_2$ **then**

$\rho = \rho/\tau$

$u^{k+1} = u^{k+1}/\tau$

$u^k = u^k/\tau$

end if

if $\|s^{k+1}\|_2 > \gamma \|r^{k+1}\|_2$ **then**

$\rho = \rho\tau$

$u^{k+1} = u^{k+1}\tau$

$u^k = u^k\tau$

end if

end if

if $\|A(x^{k+1} - x^k)\|_2 \neq 0$ **then**

$\mu := \alpha \rho \frac{\|x^{k+1} - x^k\|_2^2}{\|A(x^{k+1} - x^k)\|_2^2}$

end if

until $\|r^{k+1}\|_2$ and $\|s^{k+1}\|_2$ are small

Algorithm 8 Primal dual hybrid gradient (PDHG)

given starting point x^0 with dual y^0
given primal step size $\gamma_p > 0$, dual step size $\gamma_d > 0$
given primal acceleration factor $0 \leq \theta_p^0 \leq 1$, dual acceleration factor $0 \leq \theta_d^0 \leq 1$
 $y^{-1} := y^0$
repeat
 $\bar{y}^k := y^k + \theta_d^k(y^k - y^{k-1})$
 $\hat{x}^{k+1} := x^k - \gamma_p A^*(\bar{y}^k)$
 $x^{k+1} := \text{prox}_{\gamma_p F}(\hat{x}^{k+1})$
 $\bar{x}^{k+1} := x^{k+1} + \theta_p^k(x^{k+1} - x^k)$
 $\hat{y}^{k+1} := y^k + \gamma_d(A\bar{x}^{k+1} - b)$
 $y^{k+1} := \text{prox}_{\gamma_d G^*}(\hat{y}^{k+1})$
 $p_{k+1} := (x^k - x^{k+1})/\gamma_p - A^*(y^k - y^{k+1}) - \theta_d^k A^*(y^k - y^{k-1})$
 $d_{k+1} := (y^k - y^{k+1})/\gamma_d - \theta_p^k A(x^k - x^{k+1})$
until $\|p^{k+1}\|_2$ and $\|d^{k+1}\|_2$ are small

Writing the objective in this alternative form makes it clear that the primal and dual objectives are near-mirror images, and it gives the algorithm a beautiful symmetry as shown in Algorithm 8. Equivalence to the basic linearized ADMM algorithm comes when the acceleration parameters are constant with $\theta_p = 0$ and $\theta_d = 1$. This relationship between the two methods, and the other forms of the PDHG method, are explored in detail by *Esser, Zhang, and Chan* [2010]. The generality of the PDHG algorithm is attractive, but unfortunately that generality is a negative when it comes to achieving performance on par with the other prox methods. Although we have not included adaptive parameter selection in our PDHG formulation, such improvements do exist; *Goldstein, Esser, and Baraniuk* [2013] detail one possible scheme.

4.3.6 Similar Convex Problems

With the additional prox-based algorithms, we can solve other l_1 -minimization problems that might be of interest. Depending on convergence performance in specific instances, some of the problems described in this section may even be preferable. The first problem is already a familiar one, and it is equivalent to l_1 -regularized least squares for an appropriate choice of the noise parameter ϵ : basis pursuit with denoising. Instead of a regularization

term, the noise constraint is explicit:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 \\ & \text{subject to} && \|Ax - b\|_2 \leq \epsilon. \end{aligned} \tag{P_{1\epsilon}}$$

In this form, the meaning of the noise constraint is easier to understand and to adapt to particular use cases. Unfortunately, we lose the ability to use proximal gradient methods: BPDN can be solved using linearized ADMM or PDHG by letting F represent the l_1 -norm and G represent the indicator for the l_2 -ball of radius ϵ .

The Dantzig selector problem replaces the l_2 noise constraint with an l_∞ noise constraint:

$$\begin{aligned} & \underset{x}{\text{minimize}} && \|x\|_1 \\ & \text{subject to} && \|A^*(Ax - b)\|_\infty \leq \delta, \end{aligned} \tag{P_{1\delta}}$$

where $\|v\|_\infty = \max_k |v[k]|$. Like the prior estimators, the Dantzig selector is suitable for finding a sparse solution in accordance with the theory of compressed sensing. This problem was originally proposed by *Candès and Tao* [2007], and they did so for two reasons that both relate to the application of A^* to the residual term $Ax - b$ in the noise constraint. The first reason is that it makes the noise constraint invariant to orthonormal transformations applied to the residual. The second reason is that it ensures that no elements of the solution's residual remain highly correlated with the model dictionary. With the l_2 noise constraint, it is possible for significant model-correlated components of the residual to remain if they are small in magnitude. Although these two properties are desirable, the form of the Dantzig selector requires an extra application of A and A^* per iteration when compared to the other problems, and that cost can be significant. The Dantzig selector can be solved using linearized ADMM or PDHG by letting F represent the l_1 -norm and having G represent the indicator for the l_∞ -ball of radius δ .

Tibshirani's original LASSO (least absolute shrinkage and selection operator) estimator [Tibshirani, 1996] flips the optimization problems we've seen thus far on their heads, instead

minimizing the squared residual and constraining the l_1 -norm:

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \frac{1}{2} \|Ax - b\|_2^2 \\ & \text{subject to} \quad \|x\|_1 \leq \tau. \end{aligned} \tag{P_{2\tau}}$$

This problem is not to be confused with other problems commonly called the LASSO, whereby either l_1 -regularized least squares or basis pursuit is meant. The LASSO is particularly useful when solved repeatedly for different values of the l_1 -norm constraint τ . By doing so in sequence from low to high τ , one can observe a family of solutions with different levels of sparsity while taking advantage of the efficiency gained by reusing the prior solution to speed convergence of the subsequent optimization. Statistical selection techniques can then be used to pick the most appropriate solution out of the family, perhaps based on a sudden change in the form of the solutions. The LASSO can be solved using proximal gradient, linearized ADMM, or PDHG by letting F represent the indicator of the l_1 -ball of radius τ and having G represent half the squared l_2 -norm.

Finally, the zero-constrained least squares problem is useful when the sparsity pattern of the solution is known:

$$\begin{aligned} & \underset{x}{\text{minimize}} \quad \frac{1}{2} \|Ax - b\|_2^2 \\ & \text{subject to} \quad x \in \mathcal{Z}, \end{aligned} \tag{P_{2z}}$$

where \mathcal{Z} represents the set of vectors with a particular zero pattern. This situation can arise when one seeks to de-bias the solution of another sparse estimator. Because the l_1 -norm tends to shrink all solutions, it produces estimates that are biased toward zero. To remove this bias, one can take the zero locations discovered by the minimum l_1 -norm solution and use them to find an unbiased sparse solution via zero-constrained least squares. This problem can be solved using proximal gradient, linearized ADMM, or PDHG by letting F represent the indicator of the zero set \mathcal{Z} and having G represent half the squared l_2 -norm.

Chapter 5

Waveform Inversion Method

In the quest to eliminate matched filter sidelobes from decoded radar data, we have developed a model that represents the radar scene as discrete reflectivity coefficients $h[n, p]$ in a delay-frequency space:

$$y[m] = \sum_{p=0}^{P-1} \sum_{n=0}^{N-1} s[Rm - p + L - 1] e^{2\pi i n(Rm - p + L - 1)/N} h[n, p], \quad (5.1)$$

written compactly as

$$y = \sqrt{N} A h \quad \Leftrightarrow \quad y = Ax. \quad (5.2)$$

Solving for the reflectivity coefficients from the measured signal y provides a sidelobe-free and potentially higher-resolution ($R > 1$) radar scene. The primary barrier is that the model represents an underdetermined system of equations with infinite solutions.

We have already encountered the least-norm solution: the matched filter result. Recall that for the radar model, the composed operator AA^* is the identity scaled by the l_2 norm of s , which we can choose to be 1 without loss of generality. Thus, the least-norm solution is given by equation (4.3) as

$$x_{l_2} = A^* y, \quad (5.3)$$

which is just the (scaled) matched filter bank applied to the measured signal. Just as the

least-norm solution to a general underdetermined system may not always be the most appropriate, in this case we already know that the simple linear method provides an inadequate solution.

Fortunately, one of the reasons we sought a delay-frequency representation of the radar scene is that we expect such a representation to be sparse. With the delay-frequency radar model, measurement incoherence is completely determined by the choice of transmission waveform s , although it remains to be seen which values for s will result in minimal coherence. Provided that the measurements described by the radar model are sufficiently incoherent, the theory of compressed sensing says that we can recover the true sparse signal using convex optimization algorithms like the proximal gradient method. In this chapter, we bring these ideas to bear on the radar problem and detail how to implement them to perform waveform inversion.

5.1 Related Work

5.1.1 Other Radar Applications

Sparse approximation and the radar problem are a natural match, and researchers have proposed applying compressed sensing to radar almost since the theory was first developed [*Baraniuk and Steeghs, 2007*]. Some applications are concerned only with acquiring the returned radar signal at lower cost or with a higher sampling rate. In this case, the signal is undersampled with respect to the Nyquist rate and compressed sensing is performed in the truest sense of the term. Theory specific to this problem, with the sparse signal represented in an overcomplete and thus coherent dictionary, is examined by *Candès et al. [2011]*. On the practical side, *Yoo et al. [2012]* demonstrate wideband compressed sensing of radar pulse parameters, with the recovery done completely in hardware. The potential for compressed sensing to reduce radar hardware complexity and cost is also noted by *Baraniuk and Steeghs [2007]* and *Ender [2010]*. This research is interesting because of its potential to completely replace existing receivers with ones that are both cheaper and higher-bandwidth, but it is almost completely orthogonal to the problem of coded waveform ambiguity.

The role of sparsity in radar signal processing and the relationship between compressed

sensing techniques and established processing methods is discussed by *Potter et al.* [2010], but their focus is on imaging with synthetic aperture radar. *Harding and Milla* [2013] simulate a related technique, compressed sensing interferometric imaging, and apply it to observe equatorial spread-F using the Jicamarca ISR. Their results compare favorably to the existing Capon’s method and maximum entropy techniques. Radar imaging applications also exhibit waveform ambiguity and would benefit in theory from the elimination of sidelobes, but these and other methods safely ignore the sidelobe problem by either working with stationary targets and using a chirp waveform ([*Potter et al.*, 2010]) or using short uncoded pulses ([*Harding and Milla*, 2013]). Because of this, the existing imaging techniques are more akin to compressed sensing tomography than to waveform inversion.

5.1.2 Theoretical Support

Two works focusing on the time-frequency representation problem do have direct relevance to waveform inversion. *Bajwa, Sayeed, and Nowak* [2008] propose using a representation similar to our radar model for learning the parameters of communication channels, while *Herman and Strohmer* [2009] explore the use of compressed sensing for increased resolution of radar target detection using nearly the same model. Both are complementary to our work because they focus on waveform selection and theoretical conditions for sparse recovery rather than the practical details of applying the technique to real data. Consequently, it is useful to know that their measurement matrices have the restricted isometry property when the transmitted waveform is given by a random binary code [*Bajwa, Sayeed, and Nowak*, 2008] or the Alltop sequence [*Herman and Strohmer*, 2009], the latter of which is a type of discrete quadratic chirp.

A number of authors address the compressed sensing problem for general convolutional sensing matrices, again either with random codes [*Tropp et al.*, 2006; *Bajwa et al.*, 2007; *Romberg*, 2009; *Pfander and Rauhut*, 2010] or with the Alltop sequence [*Pfander and Rauhut*, 2010]. The specifics vary, but all prove measurement incoherence and show that sparse recovery is guaranteed with a number of measurements on the order of the solution sparsity. These results establish specific forms of the general incoherent sampling theorem from Chapter 4. The differences between these models and ours are minor but notable, so

the recovery guarantees are not directly applicable. All of the existing literature is based on a dense measurement matrix, which can only be produced by an exceptionally long pulse relative to the delay window size. As a result, the typical measurement incoherence of our model is reduced, and we can expect to require either more measurements or higher sparsity for success. Nevertheless, we are justified theoretically in seeking sparse recovery with measurements of this form.

5.2 Solution Procedure

With incoherent measurements of the form $y = Ax$ and an assumed sparse x , we are primed to solve an l_1 -minimization problem to recover x . In reality, radar data has noise, so the measurements are more accurately approximated by

$$y = Ax + \eta, \quad (5.4)$$

where η represents a vector of i.i.d. zero-mean complex Gaussian noise. To ensure that A is properly scaled with $\|A\|_2 = 1$ and that any noise has the same average power after applying the adjoint (matched filter) A^* , we normalize the code s so that $\|s\|_2 = 1$. Of the sparse recovery problems that include measurement noise, we will focus on l_1 -regularized least squares because of its simplicity and generally quick convergence. In the notation of our model, this problem solves the following:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1. \quad (P_{1\lambda})$$

Choosing the regularization parameter λ is crucial since it balances the importance of finding a sparse solution with the importance of minimizing deviation from the measurements.

5.2.1 Regularization Parameter

The solution is expected to deviate from the measurements due to noise, and it is through quantifying the noise that we can choose an appropriate λ . It is common practice to estimate

the noise variance $\hat{\sigma}^2$ by averaging the power of presumed noise-only samples y_η :

$$\hat{\sigma}^2 = \frac{1}{K} \sum_{k=0}^{K-1} |y_\eta[k]|^2, \quad (5.5)$$

where K is the number of noise samples. The noise measurement is usually acquired separately from the data measurement (or perhaps each is a different segment of a continuous stream of samples), and it typically corresponds to large delays/ranges for which no target backscatter is assumed. It can also be helpful to stabilize the noise estimate by averaging it across multiple pulses, provided one can assume that the noise power is fixed over some period of time. Certainly it makes sense to use a single noise estimate when analyzing a slice of data that encompasses multiple pulses, if only so that slight changes in the estimate do not cause apparent changes in signal-to-noise ratio (SNR) that might be interpreted instead as signal variance. No matter how the noise power estimate $\hat{\sigma}^2$ is made, accuracy is important since the estimate is used to determine the regularization parameter.

To set the regularization parameter λ , we use a property of the solution x_* to $(P_{1\lambda})$:

$$\|A^*(Ax_* - y)\|_\infty \leq \lambda, \quad (5.6)$$

In words, every entry of the result of applying A^* to the residual must have a magnitude less than λ . This fact is evident from the objective's first-order optimality condition from subdifferential calculus:

$$0 \in \partial \left(\frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1 \right) \Big|_{x=x_*}, \quad (5.7)$$

which evaluates to

$$(A^*(Ax_* - y))_i \in \begin{cases} \{\lambda x_{*i}/|x_{*i}|\}, & |x_{*i}| > 0 \\ \{\lambda : |\lambda| \leq 1\}, & x_{*i} = 0 \end{cases} \quad (5.8)$$

for every individual entry x_{*i} of x_* . One possible choice for λ is just below the critical value $\lambda_{\text{crit}} = \|A^*y\|_\infty$ at which $x = 0$ ceases to be the optimal solution. The choice of

$\lambda = 0.99\lambda_{\text{crit}}$ or $\lambda = 0.95\lambda_{\text{crit}}$ provides a maximally sparse nonzero solution, but it can allow the residual $y - Ax_*$ to have a higher variance than would be expected with a given noise level.

Since we have an estimate for the noise power, we can do better. As was shown with the derivation of the radar model, A^* scaled by \sqrt{N} is equivalent to matched filtering. Therefore, we can think of $\sqrt{N}A^*(Ax_* - y)$ as the matched filter applied to the residual. In order for x_* to be an acceptable solution, we restrict the filtered residual so that its entries have a magnitude less than the noise level. Otherwise, we would consider that element to represent scattered signal from a target, which should be represented in the solution x_* instead of in the residual. Since $\sqrt{N}\lambda$ is an upper bound on the filtered residual from equation (5.6), λ specifies the detection threshold. For a given false alarm rate for detection, we can set the detection threshold at a fixed multiple C_{FAR} of the noise power. In other words, we want to set a solution constraint of

$$\left\| \sqrt{N}A^*(Ax_* - y) \right\|_\infty^2 \leq N\lambda^2 = C_{\text{FAR}}\hat{\sigma}^2, \quad (5.9)$$

which implies that we want

$$\lambda = \sqrt{C_{\text{FAR}}} \frac{\hat{\sigma}}{\sqrt{N}}. \quad (5.10)$$

If the noise is truly zero-mean complex Gaussian, then $C_{\text{FAR}} = 4$ corresponds to a false alarm rate of about 2%. Targets with an SNR less than 4 (after the sidelobes from higher-signal targets are removed) won't be included in the solution, and about 2% of the noise will be included in the solution. This is a reasonable trade-off between detection and model complexity, so our default value for λ is $2\hat{\sigma}/\sqrt{N}$.

5.2.2 Iterative Thresholding Matched Filter

One method for recovering the sparse solution for the reflectivity coefficients is to use the iterative soft thresholding algorithm derived from the proximal gradient method. Examining the steps of this algorithm leads to an intuitive interpretation, in radar terminology, of the nonlinear procedure for arriving at a sparse solution. We describe this interpretation as an *iterative thresholding matched filter*.

Algorithm 4 Iterative Soft Thresholding

given a starting point x^0 , a step size $0 < \mu \leq 2/\|A\|_2^2$
repeat

$$z^{k+1} := x^k - \mu A^*(Ax^k - y)$$

$$x^{k+1} := \mathbf{soft}_{\mu\lambda}(z^{k+1})$$

until stopping criterion is satisfied

Consider the algorithm's first step for an initial guess of $x^0 = 0$ and step size $\mu = 1$:

$$z^1 := A^*y. \quad (5.11)$$

In radar terms, this first step is to apply the matched filter to the received signal. The next step produces a first “guess” of the solution by thresholding the matched filter result:

$$x^1 := \mathbf{soft}_\lambda(z^1) = \mathbf{soft}_\lambda(A^*y). \quad (5.12)$$

Values in the matched filter result that are below the detection threshold set by λ are set to zero, and the remaining values are shrunk toward zero. In other words, the first guess is simply everything that the matched filter would classify as a detected signal, which can include sidelobes.

Looping to the next iteration, the third step is:

$$z^2 := x^1 + A^*(y - Ax^1). \quad (5.13)$$

The interpretation is that the expected received signal based on the first guess is calculated and compared to the actual measurement. The difference is matched filtered to see what newly-revealed components of the signal were left out the first time, and that result is added to the original guess. As before, the fourth step thresholds the matched filter result to get an improved guess:

$$x^2 := \mathbf{soft}_\lambda(z^2) = \mathbf{soft}_\lambda(A^*(x^1 + A^*(y - Ax^1))). \quad (5.14)$$

Further iterations proceed in the same manner, applying matched filtering and thresholding to steadily improve the solution estimate until convergence.

5.2.3 Scaling and Noise

It is convenient to work in terms of x so that the model operations A and A^* are well-scaled, but $h = x/\sqrt{N}$ is what we actually want given that the matched filter bank has the scaling $\sqrt{N}A^*y$. The reflectivity coefficients h are indeed the matched filter result with modeled sidelobes and the filtered residual removed:

$$\begin{aligned}\sqrt{N}A^*y &= \sqrt{N}A^*(y - Ax + Ax) \\ &= NA^*Ah + \sqrt{N}A^*(y - Ax) \\ \underbrace{\sqrt{N}A^*y}_{\text{matched filter}} &= h + \underbrace{(NA^*A - I)h}_{\text{sidelobes}} + \underbrace{\sqrt{N}A^*(y - Ax)}_{\text{matched filtered residual}}.\end{aligned}\quad (5.15)$$

Recall that the operation $N \cdot A^*A$ adds matched filter ambiguity to its input (equation 3.12), hence subtracting the identity I and applying it to h leaves only the sidelobes generated from h . The residual mostly represents measurement noise, but a significant part of it is the bias that is left after l_1 -minimization overly-shrinks the solution toward zero.

Solution bias can be a problem if it is not handled properly. One method to acquire an unbiased estimate is to find the least-squares solution with the sparsity pattern that matches x_* . This can be solved efficiently using any of the prox methods detailed in Chapter 4. With reasonable false-alarm rates, however, the sparsity pattern of x_* will include noise coefficients, and this de-biasing step can cause those coefficients to take on unreasonable values in an effort to match all of the noise. Fortunately for this application, the de-biasing step is generally unnecessary since in most cases we can work with a modified solution that includes the residual term. If all we want to do is remove sidelobes from the matched filter solution, then we can subtract only that term from equation (5.15) and define

$$\hat{h}_* = h_* + \sqrt{N}A^*(y - Ax_*), \quad (5.16)$$

or equivalently

$$\hat{h}_* = \frac{x_*}{\sqrt{N}} + \sqrt{N}A^*(y - Ax_*). \quad (5.17)$$

This modified reflectivity estimate \hat{h}_* nicely encapsulates the entire solution. The first term is the sparse reflectivity, embodying the signal that is well-represented by the radar model. The second term is matched filtering applied to the residual, embodying everything that is not captured by the radar model. If the sparse solution leaves out significant components that fall below the noise threshold, those signals will be visible in the residual term. Instead of calling everything in the residual “noise” and discarding it, we incorporate it into the solution and let it be judged on its own merits.

5.2.4 Solution Residual

Care must be taken when working with the residual for purposes other than comparison to the full delay-frequency matched filter result. Since waveform inversion eliminates side-lobes, there is a new viable option for representing the solution as a range-time-intensity (RTI) plot: totaling the reflected power across all frequencies and displaying it as a function of range and time. In other words, we sum the square magnitude of the reflectivity over the frequency index and plot that instead of a single-frequency slice. Doing this with the estimate \hat{h}_* produces incorrect results since the noise is not independent across frequency. The residual values $\zeta = y - Ax_*$ are independent, but the matched filtered residual across delay-frequency space is clearly not. Summing across the frequency index of $A^*\zeta$ is acceptable, and it produces the zero-frequency matched filter result of correlating ζ with the transmitted waveform s^* . The problem arises from the added scaling factor of \sqrt{N} in the residual term of \hat{h}_* , which magnifies the noise by that factor after frequency summing. Therefore, when calculating a frequency-integrated RTI plot, we drop the scaling factor on the residual term and proceed as normal.

It must be noted that adding the residual to the sparse solution in the manner of \hat{h}_* does not produce a solution that matches the measurements exactly. A nearly-sparse exact

solution x_0 of $y = Ax$ is given by adding the unscaled residual to x_* :

$$x_0 = x_* + A^*(y - Ax_*). \quad (5.18)$$

This solution is nearly-sparse because it has few nonzero values with magnitude greater than λ , but many nonzero values with magnitude less than λ . Applying the forward model produces

$$\begin{aligned} Ax_0 &= Ax_* + AA^*(y - Ax_*) \\ &= Ax_* + y - Ax_* \\ &= y \end{aligned} \quad (5.19)$$

since AA^* is the identity. Though the scaled solution $h_0 = x_0/\sqrt{N}$ is an unbiased estimate of the true reflectivity coefficients, it is not particularly useful. One negative is that it is not exactly sparse, but the difference is small enough to be of almost no importance. The main problem with h_0 is that it spreads the residual out over delay-frequency space in a manner that masks its true representation in the measurement space. The noise coefficients $x_\eta = A^*(y - Ax_*)$ that get added to x_* are simply the least-norm estimate for $y - Ax_* = Ax_\eta$, which has no physical meaning. In order to understand where the model fails in approximating the true solution, it is preferable to work with the estimate \hat{h}_* since it can be compared directly to the matched filter result.

5.3 Implementation

Performing waveform inversion in practice requires the ability to calculate the radar model and solve a sparse approximation problem using it. Obviously there was no existing computer code that implemented our specific radar model, so that had to be written from scratch. As for sparse approximation, many packages are freely available that solve the various forms of that problem. However, many of them do not accommodate specifying A and A^* as functional operators rather than explicit matrices, so that leaves far fewer software packages to choose from. The TFOCS package described by *Becker, Candès, and Grant* [2011a]

and available at <http://cvxr.com/tfoocs/> is the only one that implements the algorithms in Chapter 4 and includes the performance enhancements that speed convergence. The initial exploration of sparse approximation using the radar model was done using TFOCS, and it was well-suited to that purpose.

Unfortunately, two obstacles were encountered. The first is that it was difficult to tune the algorithm parameters to achieve reliably good performance. The second is that TFOCS is written using MATLAB, which became a problem when we decided to implement the radar model in Python for ease of reuse. In the end, it was necessary to also code all of the convex optimization algorithms from scratch. To prevent this from happening to other researchers in the future, the radar model and proximal operator software packages have been made available as BSD-licensed open source Python code [*Python Software Foundation, 2014; Oliphant, 2007*], which can be found on GitHub at <http://github.com/ryanvolz>. The radar model package was given the creative name of `radarmodel`, while the proximal optimization routines are packaged together as `prx`. For the purposes of reproducibility, all of the code for producing the results in this thesis is also available at the same GitHub site. Special recognition must go to the developers of `matplotlib` [*Hunter, 2007*] and `IPython` [*Pérez and Granger, 2007*]; their plotting and interactive programming tools have made this work possible.

5.3.1 `radarmodel`

The `radarmodel` package implements the forward operator A and adjoint operator A^* , which are necessary to compute solutions using the radar model of Chapter 3. Users first specify the transmission code s , the undersampling ratio R , the length of the measurement vector M , and the number of frequency steps N for a particular problem formulation. Then, the program creates tailored functions for the forward and adjoint operators which can be used to quickly compute those operations repeatedly. As with most scientific computing in Python, `radarmodel` makes extensive use of the `numpy` package [*Walt, Colbert, and Varoquaux, 2011*] for fast array computations.

Since the primary goal in writing special-purpose functions to compute the radar model is to make them as fast and efficient as possible, great care was taken in implementing

and benchmarking the operators. Multiple formulations that deconstruct the operations into different convolution, Fourier transform, and array arithmetic steps were tested, and `radarmodel` can select the best formulation for a particular set of problem parameters on the fly. The different formulations all make use of the FFT operator through the `pyffw` package [Gomersall, 2014], and some of them use the compiled-language add-ons provided by `cython` or `numba`. The `cython` package [Behnel *et al.*, 2011] allows one to take a restricted form of the Python language, add static declarations, and compile it into C code. The `numba` package [Oliphant and The Numba Contributors, 2014] allows one to decorate customized `numpy` routines and compile them just-in-time (JIT) by dynamically invoking the `llvm` compiler. Both packages provide optimized performance for intensive calculations without having to leave Python, greatly easing development and debugging. The result is a radar model implementation that is one or two orders of magnitude faster and less memory-intensive than a naive or matrix implementation.

5.3.2 prx

The `prx` package implements the proximal gradient, accelerated proximal gradient, ADMM, linearized ADMM, and PDHG methods for solving a number of l_1 -minimization problems. It also provides functions for computing the prox operators that are necessary to solve those problems. As with `radarmodel`, `prx` makes extensive use of the `numpy` package [Walt, Colbert, and Varoquaux, 2011]. Each optimization algorithm requires the user to specify the objective functions F and G , the linear operators A and A^* , the data vector b , and algorithm parameters for step size and convergence tolerance. The optimization algorithms can be used to solve any problem with the correct split-objective form, provided that F and G can be specified. These objective functions can be composed from a selection of elemental functions with known gradients and/or prox operators, or entire standard problems like l_1 -regularized least squares can be solved using specialized functions. The package is designed to be easily-extensible so that if a new objective function is required, it can be implemented in a standard form and used with the optimization routines just like the standard functions. This design has proven effective in testing a variety of algorithms and optimization problems, including ones not originally envisioned.

In contrast to the `radarmodel` package, the key to efficient computation with `prx` lies in optimizing the individual algorithms and making them easy to use. All of the enhancements mentioned in Chapter 4, and a few additional ones, are included so that the algorithms converge as quickly as possible. Moreover, all of the required tuning parameters can be specified if required, but the default values work well enough that tuning won't be necessary in most cases. Since the package is written completely in Python, it is easy to examine and modify the code to figure out what is happening behind the scenes, making the implementation far more useful than a computational black box. Special care was taken to provide user-visible documentation that describes each of the optimization routines, gradient and proximal operators, and standard problem formulations in detail, with references included if more information is required.

5.3.3 Algorithm Comparison

With a variety of proximal optimization algorithms available to solve the l_1 -regularized least squares problem, it was necessary to test each algorithm to determine which converges fastest in general when applied with the radar model. Computation time for all of these algorithms is dominated by the evaluation of the forward and adjoint operators of the radar model, so the primary interest is in minimizing the number of iterations required for convergence. Table 5.1 lists the iteration count until convergence for some of the algorithms detailed in Chapter 4 and implemented in the `prx` package. All of the algorithms were applied to the same instance of an l_1 -regularized least squares problem using the radar model provided by the `radarmodel` package. The required single pulse of radar data came from the Jicamarca ISR using a minimum peak sidelobe waveform, and it contained scattering from a meteor head echo. Convergence was determined using the same absolute and relative tolerance parameters for all of the algorithms. Regardless of these details, the results are generally descriptive of our experience using these algorithms on actual radar data.

Table 5.1a compares the proximal gradient method with its improvements, each version adding a feature to the last. The unmodified proximal gradient algorithm is somewhat slow to converge, while the accelerated scheme is about a third faster. Adding adaptive restart of the acceleration term brings a significant improvement, dropping the number of steps down

Table 5.1 – Convergence comparison of proximal optimization algorithms. The tables list the number of steps until convergence (and backtracking steps, if applicable) for the same instance of l_1 -regularized least squares.

(a) Proximal gradient variations		(b) Three different methods	
Method	Iterations	Method	Iterations
Proximal gradient	4463	Accelerated proximal gradient	51 (36)
+ Acceleration	2915	Linearized ADMM (adaptive)	79 (96)
+ Adaptive restart	399	PDHG (fixed step)	4466
+ Adaptive step	51 (36)		
– Acceleration	58 (27)		

an order of magnitude. Adding the adaptive step size scheme drops the number of steps by another order of magnitude, converging in less than a hundred total steps (complete steps and partial backtracking steps). Finally, keeping the adaptive step size but dropping the acceleration scheme just barely increases the number of iterations over the accelerated version, with even fewer backtracks. Because of this vast improvement from step size adjustment, it seems that the radar model coupled with l_1 -regularized least squares is particularly prone to large variations in the optimal step size. This would appear to be an exceptional requirement, since as far as we are aware, an expanding adaptive step size is included in only one of the available proximal gradient implementations (TFOCS), and it is not well-publicized in the literature. Also of note is how the acceleration scheme only becomes really effective once adaptive restart is allowed, but even that improvement is inconsequential compared to having an adaptive step size.

The accelerated and fully adaptive proximal gradient method is compared to linearized ADMM and PDHG in Table 5.1b (plain ADMM is not applicable to this l_1 -regularized least squares problems because of the affine transformation). Given that both linearized ADMM and PDHG can handle more general problems, it is not surprising that they both take longer to converge than the optimized proximal gradient method. The linearized ADMM method uses an adaptive step size scheme of our own devising that is similar to that for the proximal gradient method. Considering the importance of the adaptive step for the latter method, it is no wonder that linearized ADMM converges much faster than the fixed step PDHG. The

PDHG algorithm used here is the basic one; it can be improved with adaptive parameter selection, but that implementation was deemed unnecessary for now given the relative success of the proximal gradient method. It is interesting, and probably not a coincidence, that both of the fixed step algorithms (PDHG and basic proximal gradient) converge in almost the same number of iterations. This fact not only shows how closely-related all of these algorithms are, but it also emphasizes just how step-size dependent this problem is.

Limited performance testing was also done for the basis pursuit with denoising and Dantzig selector problems, with both using the linearized ADMM algorithm. Approximately equivalent sparse answers can be achieved with appropriate parameter selection, but both algorithms require more iterations to converge than l_1 -regularized least squares. Roughly speaking, basis pursuit with denoising takes about twice as many iterations, while the Dantzig selector takes about ten times as many iterations. Little effort has been put into investigating optimal algorithm parameters for these problems, so it is likely that performance can be improved. Nevertheless, it is unlikely that either will be able to converge faster than l_1 -regularized least squares using the accelerated proximal gradient method. Given the success of that combination, we feel justified in continuing to use it almost exclusively.

Chapter 6

Experimental Results

The real value in using sparsity for waveform inversion manifests through its application. Backed by an efficient implementation of the radar model operators and the quick convergence of tuned sparse recovery algorithms, we are primed to demonstrate the waveform inversion method on real radar data. Though the eventual goal in applying this technique is to enable highly-flexible and artifact-free measurements, the goals of this experiment are to (1) show that waveform inversion produces results that are consistent with our expectations, (2) directly compare the effectiveness of different waveforms, and (3) collect high-quality meteor data.

We expected that inversion would eliminate delay-frequency sidelobes and strictly improve upon matched filtering, and the experiment is successful in demonstrating that. We know that the pulse waveform directly affects measurement incoherence in terms of compressed sensing, and the experiment shows that successful inversion is dependent on the compactness of the peak of the waveform’s ambiguity function. Regarding meteors, the hope was that eliminating sidelobes and increasing resolution through waveform inversion would allow us to address the open questions concerning flares and micrometeoroid fragmentation [Mathews *et al.*, 2010]. Further experimentation will be required to answer those questions, but the experiment proves that the technique is still useful for eliminating interference between meteor head echoes and stationary targets like non-specular trails and the equatorial electrojet. Without sidelobe interference, each of those phenomena can be detected more reliably and studied independently.



Figure 6.1 – The Jicamarca 50 MHz radar. The author is pictured, pre-experiment, in front of the Jicamarca Radio Observatory’s main radar, a large phased array operating at 50 MHz and consisting of 18,432 crossed half-wave dipoles. Located outside of Lima, Peru, the radar’s primary use is to study the equatorial ionosphere from its vantage point near the magnetic equator.

From February 28 through March 2, 2011, we collected approximately 48 hours of nearly-continuous data observing the equatorial ionosphere using the Jicamarca 50 MHz radar. That amounted to over two terabytes of receiver samples. Even with fast code and good algorithms, it is not reasonable at this time to apply the waveform inversion technique to anything more than tiny segments of that data. We have chosen to focus on a handful of examples collected from the Jicamarca data set in order to analyze them in detail and infer the method’s effectiveness in general. With the knowledge gained from this experience, we are able make recommendations to aid future applications of the waveform inversion technique.

6.1 Jicamarca

The Jicamarca 50 MHz radar is a phased array composed of 18,432 crossed half-wave dipoles in a 96 by 96 square grid (Figures 1.4b and 6.1), covering an area of almost 85,000 square meters. It is located inland from Lima, Peru, at a geographic latitude and longitude

of 11.95° south, 76.87° west, placing it near the magnetic equator [Ochs, 1963]. Accordingly, the main use of Jicamarca is the study of the equatorial ionosphere, and monitoring its background plasma through incoherent scatter is one primary objective. Since the radar is able to point perpendicular to the Earth’s magnetic field (or nearly so), it enjoys a number of rare capabilities. Ultra-precise measurement of ionospheric electric fields is one benefit of that configuration, as is the ability to view field aligned irregularities, which are frequently fueled by meteor ionization and radar-visible as non-specular meteor trails [Woodman and Hagfors, 1969; Chapin and Kudeki, 1994]. The study of these and other unique transient events, including spread-F and the equatorial electrojet, is another focus of Jicamarca’s scientific program. Thanks to the radar’s low VHF frequency relative to other high-power large-aperture (HPLA) radars, more meteor head echoes are detectable with the Jicamarca array than with almost all of those other HPLAs given the inverse scaling of frequency with received signal strength, which scales approximately as $1/f^2$. Head echoes, trails, and electrojet are near-constant features of the daytime ionosphere at Jicamarca. This crowded and challenging environment is a great proving ground for new measurement techniques, and it provides a surfeit of opportunities for testing our waveform inversion method.

6.1.1 Specifications

The observatory is currently equipped with three transmitters that each have a peak power output of about 1.5 MW, for a total possible transmitter power of 4.5 MW. The transmitter system supports a variety of pulsed waveforms, including binary-phase codes and the LFM chirp. These pulses can be transmitted with a maximum duty cycle of 6%, a maximum bandwidth of about 1 MHz (shortest pulse of 1 microsecond), and a maximum length of 2 milliseconds. Four phase-coherent receivers can be used to sample the acquired signal with a maximum bandwidth of 1 MHz. Each receiver downmixes the signal to baseband frequency and outputs complex samples representing in-phase and quadrature components [The Jicamarca Radio Observatory 2014].

6.1.2 Array Segmentation

The four receiver channels can measure signals from distinct subsets of the whole array. One typical configuration involves sampling the four quadrants of the array independently. Comparison of signal phase between these quadrants through interferometry allows one to determine the signal's direction of arrival. The channels can also be used to sample independent polarizations or more complex interferometry setups with multiple measurement baselines. The dipoles are actually organized into 64 separate, square modules; the modules can be sampled in any combination, but they can also be phased independently to form different beam patterns. Tuned appropriately, the full array has a one-way half power beamwidth (full-width at half maximum) of 1.1° and can be electronically steered within about 7° of zenith. Currently, individual module phasing must be carried out manually by adjusting cable lengths, but a computer-controlled phasing system is being developed to enable pulse-to-pulse beam pointing [*The Jicamarca Radio Observatory* 2014].

6.2 Setup

The suitability of different radar waveforms for sparse recovery is an open question. We expect from theoretical results that random or pseudorandom waveforms will produce suitably incoherent measurements, but most of the standard waveforms also merit investigation. Because they are designed to minimize the effects of ambiguity, codes with ideal sidelobe properties are likely to also be useful for inversion. Our experiment using the Jicamarca radar was designed to test the effectiveness of a number of these common waveforms by comparing them side-by-side while scattering from the same targets. Switching the code on a pulse-to-pulse basis allows us to judge waveforms against each other under the same circumstances. As long as we eventually repeat the same waveform after a short time relative to the persistence of the target, we can build up equivalent views of a radar scene using a number of different waveforms.

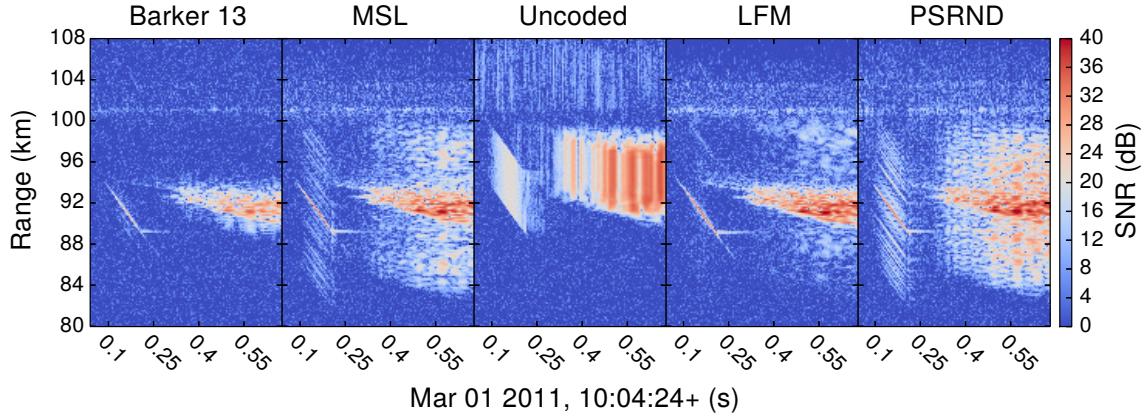


Figure 6.2 – Sequence of waveforms used in the Jicamarca experiment. The images show example meteor data collected with the Jicamarca radar using five different pulse waveforms. The radar transmits just one of the five waveforms with each pulse, stepping through the sequence of five in the order shown before repeating.

6.2.1 Waveforms

For this experiment, we selected five different waveforms and stepped through them sequentially before repeating the sequence. The five waveforms are depicted using a meteor head and trail in Figure 6.2; the waveforms are, in order, a Barker-13 code (B_{13}), a minimum peak sidelobe code (MSL_{51}), an uncoded pulse, a linear frequency modulation (LFM) chirp, and a pseudorandom code ($PSRND_{51}$). The first four were chosen because they are widely used for collecting meteor head echo and trail data [Wannberg, Pellinen-Wannberg, and Westman, 1996; Erickson *et al.*, 2001; Mathews *et al.*, 2008; Close, Oppenheim, S. Hunt, *et al.*, 2002]. The pseudorandom code was chosen because it was expected to have good waveform inversion performance based on compressed sensing theory [Bajwa, Sayeed, and Nowak, 2008; Pfander and Rauhut, 2010]. The three binary phase codes are described by the following sequences:

$$s_{B_{13}} = 1111100110101$$

$$s_{MSL_{51}} = 00011100011111100010001100100010010101001001001011$$

$$s_{PSRND_{51}} = 011111000010000001001100001011000001001011010011100.$$

6.2.2 Pulse Parameters

With the exception of the uncoded pulse, a bandwidth of 1 MHz (the maximum at Jicamarca) was fixed as the primary design point. A high bandwidth equates to better pulse compression and smaller regions of ambiguity. For the phase-coded pulses, that requirement specifies a baud length of $1 \mu\text{s}$. For the LFM chirp, it specifies a total frequency sweep of 1 MHz, centered around the radar frequency. The Barker-13 code is the longest binary code with a maximum sidelobe level of 1 when the peak is normalized to the code length, but it is still short, with a total pulse length of $13 \mu\text{s}$, compared to the other pulses. We wanted a longer minimum peak sidelobe code in order to test the effects of pulse length, so the code of length 51 was chosen since it is the longest that has a peak sidelobe level of 3 or less. Hence, the MSL_{51} pulse has a length of $51 \mu\text{s}$. The lengths of the LFM and pseudorandom pulses were chosen to match the minimum sidelobe pulse. The pseudorandom code was indeed produced by a random number generator; we drew a few binary codes of length 51 until we found one with a somewhat-uniform autocorrelation. Finally, the uncoded pulse bears almost no relation to the others. It was included as a baseline so that any unusual results from the coded pulses could be compared to the simplest data possible. A length of $40 \mu\text{s}$ was arbitrarily chosen to be long enough for reasonable sensitivity but short enough for reasonable ambiguity, and it is close to the value used by *Mathews et al.* [2003] at Arecibo.

6.2.3 Pulse Timing

Meteors are typically observed at altitudes ranging from 80 to 120 kilometers, although they are sometimes observed as low as 70 kilometers and as high as 140 kilometers. With that as our primary range window of interest, we chose an inter-pulse period of 1.02 ms, with the corresponding maximum unambiguous range of 153 km, to minimize the time required to repeat the waveform sequence. It is advantageous to repeat each waveform as quickly as possible to ensure that even transient events would be sampled sufficiently by each type of pulse. This is also the reason that the waveform sequence was limited to five pulses. With that inter-pulse period, the longest pulses in our pulse sequence nearly achieve the maximum duty cycle of Jicamarca's transmitter.

6.2.4 Sampling

High resolution is an obvious goal for meteor measurements so that accurate range and range rate estimates are possible. Waveform inversion is also dependent on having enough measurements relative to the sparsity level to ensure success. The highest sampling rate is desirable, and at Jicamarca that means a sampling period of $1 \mu\text{s}$. With the selected waveforms consuming 1 MHz of bandwidth and switching phases at a period of $1 \mu\text{s}$, no other choice would have made sense.

6.3 Procedure

The waveform inversion procedure specified in Section 5.2 was applied to second-long windows of the Jicamarca data. To ensure proper normalization for inversion, the radar model was defined with the code sequence s normalized to a power of 1. To ensure that the number of frequency steps exceeded the length of the longest code sequences, the number of frequency steps N was set at 128. Specifying a power of 2 ensures maximally-efficient Fourier transforms, and the higher value ensures just enough frequency resolution to differentiate moving head echoes from stationary targets. Except where noted, the model was defined with no undersampling, $R = 1$. In all cases, we used the accelerated proximal gradient method with adaptive restart and adaptive step size to iteratively solve for the sparse reflectivity from the measurements. Each pulse was analyzed individually, with no shared information other than the uniformly-specified regularization parameter λ , which was set based on the estimated noise level. Using a recent laptop computer to perform the computations, the solution for each one-second segment of data took anywhere from 15 minutes to 1 hour to calculate. For a single pulse, the solution converged in approximately 1 second, with denser solutions taking correspondingly longer.

6.4 Pseudorandom Pulse Results

Given the theoretical backing, it was expected that the pseudorandom waveform would be suitable for inversion. Fortunately, that is the case. Many of the waveform inversion

features of the pseudorandom code are shared by the other waveforms. Therefore, we can summarize many of the basic features of the technique by first limiting our attention to the pseudorandom pulses.

6.4.1 Pulse Slices

Figures 6.3 and 6.4 compare waveform inversion and matched filter bank results over a series of individual pseudorandom pulses. Each slice represents data from a single pulse and shows part of the delay-frequency decomposition of the radar signal, giving SNR as a function of range and Doppler range rate. Figure 6.3 shows the time evolution of a meteor head echo that ends in a flare. The head echo is first visible as a single point with negative range rate coming toward the radar, and over time it descends in range until flaring and exiting the radar beam. Figure 6.4 depicts the initial appearance and later fluctuation in the non-specular trail that follows the head echo. The scattering has a Doppler range rate of zero and fades in and out over the ranges containing the meteor plasma, according to the scattering characteristics of the evolving field-aligned irregularities.

The range rate axis has particularly poor resolution for these and all of the other results. The meteor head echo, traveling at approximately 40 km/s toward the radar, is only two range rate/frequency steps away from zero out of the 128 total grid points. This is a result of Jicamarca's relatively low baseband frequency of 50 MHz. At that frequency, even fast moving targets can barely generate a significant frequency shift relative to the available pulse lengths. Better frequency/range rate resolution is only possible by increasing the length of time covered by the frequency analysis, and that means analyzing the same target over multiple pulses. As such, the frequency resolution cannot be improved without the development of a multi-pulse waveform inversion technique.

Though it makes no practical difference for a single point target like the meteor head echo, the first thing to notice from the results is just how significant the delay-frequency sidelobes are in the matched filter solution. Waveform inversion effectively eliminates those sidelobes, leaving only the head echo points and noise, the latter of which has been added back to the sparse solution from the residual. It is also important to note the second row of head echo images, in which the solitary target appears to occupy multiple points spanning

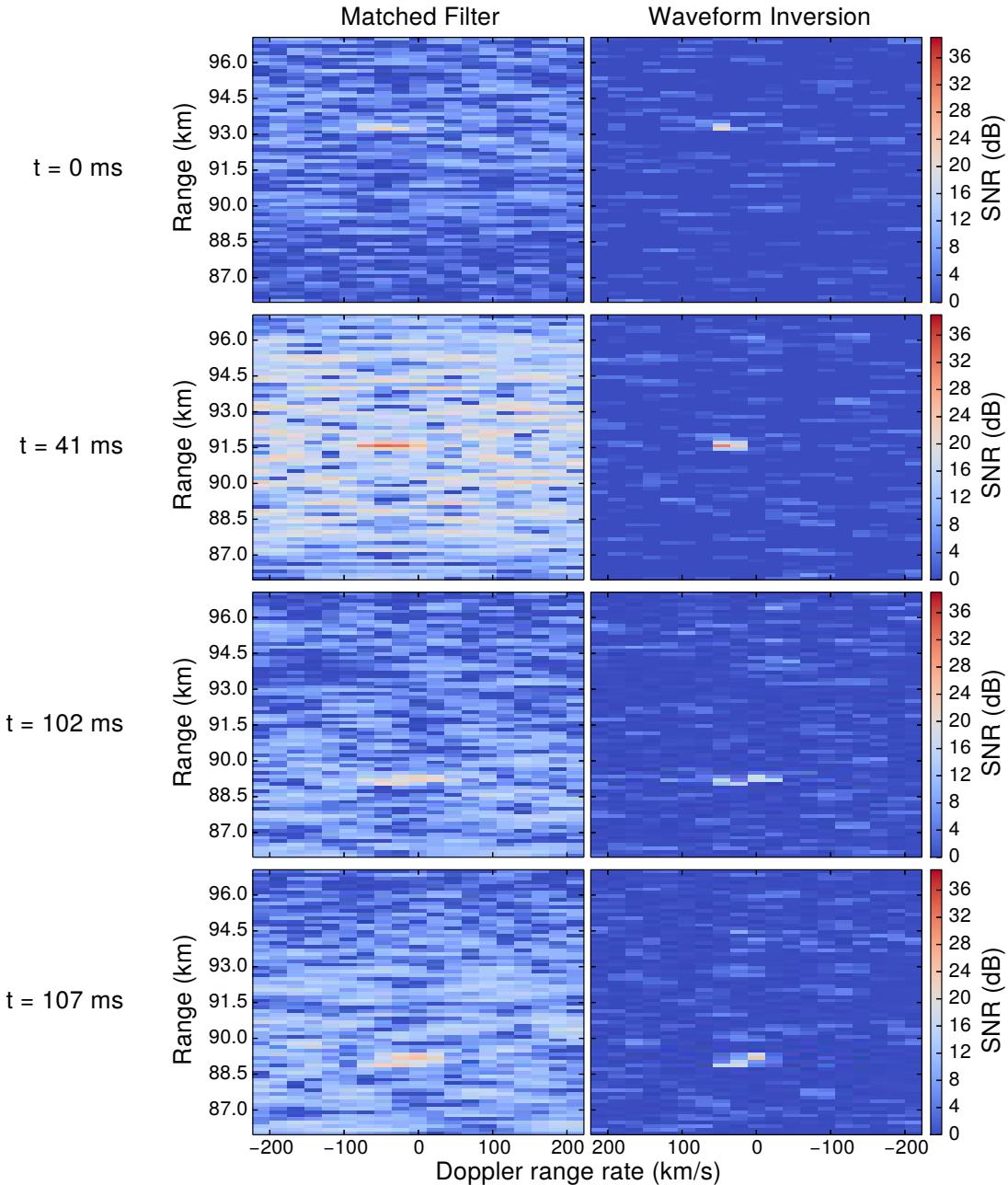


Figure 6.3 – Pulse slices comparing matched filter and waveform inversion. The images in each row depict meteor head echo scattering as a function of range (delay) and Doppler range rate (frequency) for a single pseudorandom pulse. Corresponding pulse times, relative to the first pulse, are shown at left. The differentiated signal in the waveform inversion result (right column), somewhat obscured in the matched filter result (left column), shows the head echo descending in range and flaring.

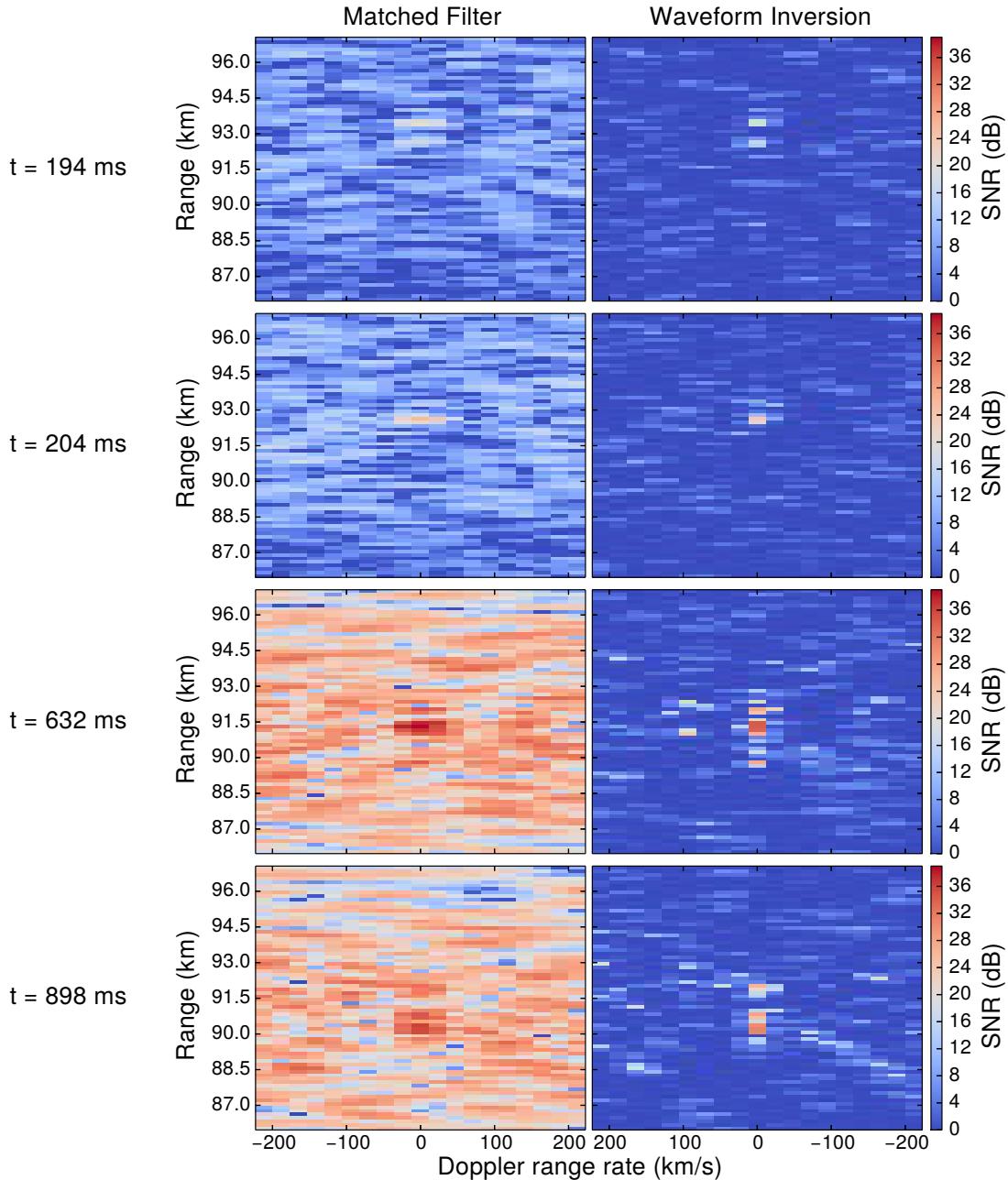


Figure 6.4 – Pulse slice comparison for a meteor trail. The images in each row depict non-specular trail scattering as a function of range (delay) and Doppler range rate (frequency) for a single pseudorandom pulse. Corresponding pulse times, relative to the time of the first pulse shown in Figure 6.3, are shown at left. Waveform inversion (right column) removes the sidelobes obscuring the matched filter result (left column) to make the fluctuating scattering at zero-range rate clearly visible.

two range rate bins. This is as expected for a point target, due to the blurring experienced as a result of discretization (see Section 3.3.3). The true range rate lies between those two grid points, and the discrete model copes by spreading the signal over both. The flaring seen in the last two rows of head echo images is notable because sidelobe removal allows us a clearer view of the signal, which can be seen spreading to zero range rate while the head echo continues to descend. Given the poor frequency resolution, it is impossible to draw any conclusions from this result, but it is a good illustration of a potential future application.

The non-specular meteor trail is most notable for how the results go from completely cluttered in the matched filter case to relatively clean in the waveform inversion case, all thanks to the elimination of sidelobes. The most significant scattering is still noticeable with the matched filter, but it is also impossible to tell if the rest represents actual scattering or just delay-frequency sidelobes. Waveform inversion answers that question by clearly separating true scattering from sidelobe. Nevertheless, the waveform inversion solution is not perfect. It includes some significant signal at impossibly high frequencies/range rates, which we are inclined to identify as reconstruction artifacts. The best guess is that the artifacts represent the result of the transmitted signal not exactly matching the ideal, infinite-bandwidth code used in the model. As a result, the sparse solution is forced to include whatever coefficients add together to make up the difference between the actual and expected pulse shape. Imperfect matching is sufficient for good pulse compression, but the difference is much harder to account for in the sidelobes.

6.4.2 Integrated RTI Plots

The same meteor head echo, flare, and trail is shown in Figure 6.5 as a range-time-intensity (RTI) plot. The matched filter result in part (a) is for the peak-frequency-response filter for each individual pulse, selected from the entire frequency bank of 128 filters. The waveform inversion result in part (b) is for the total signal summed across all frequency coefficients, which is sensible only because no delay-frequency sidelobes are present. The sparse solution's residual was matched filtered and added to the sparse coefficients to form the waveform inversion result as described in Section 5.2.

As with the time slices of this data, waveform inversion is notable for removing the

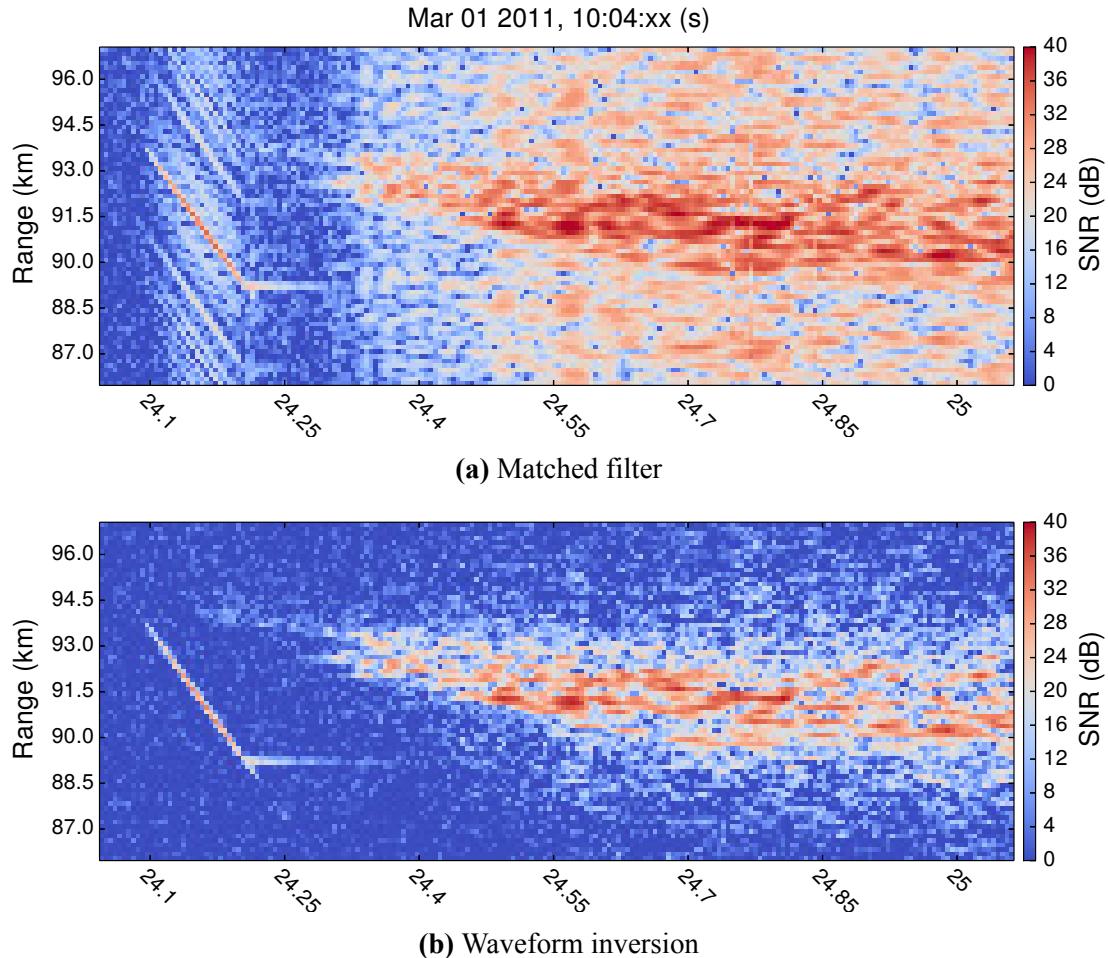


Figure 6.5 – RTI comparison of matched filter and waveform inversion. These images show the same pseudorandom-coded meteor head, flare, and trail from Figures 6.3 and 6.4. The matched filter result is for the peak-frequency-response of each pulse, while the waveform inversion result is the integrated power over all frequency coefficients. The elimination of delay-frequency sidelobes with waveform inversion reveals previously-obsured trail scattering and makes the true signal easier to identify.

range sidelobes that clutter the matched filter result while retaining the high SNR of the true signal. There can be no doubt here that the waveform inversion method is beneficial. The pseudorandom code's autocorrelation function is clearly visible in the matched filter image, surrounding the descending head echo streak. With waveform inversion, those range sidelobes are completely eliminated to reveal previously-obscured trail scattering. Much of the clutter is also removed for the non-specular trail, although some artifacts remain. As previously discussed, these artifacts are likely the result of mismatch between the actual pulse waveform and its idealized representation in the radar model. Even with these artifacts, the result is much improved.

6.4.3 Frequency Slices

One way to minimize the effect of recovery mismatch artifacts is to analyze the individual frequency slices of the waveform inversion result. Frequency decomposition also makes it easy to separate and identify different types of signals. It is generally not possible to perform this type of analysis with a matched filter, since sidelobes of one target often obscure the features of another target even when separated in frequency. To demonstrate, Figure 6.6 shows the first three positive frequency shifts, starting at zero, of the previous waveform inversion result.

As nearly stationary targets, meteor trails exhibit only a small Doppler shift that is on the order of 100 m/s and is associated with the background neutral winds. Because this potential shift is very small compared to the waveform inversion frequency resolution, the flare and non-specular trail are clearly visible in the zero-frequency result. In contrast to the frequency-integrated RTI plot, the trail as shown at zero frequency is free from clear mismatch artifacts. The artifacts are spread randomly throughout the entire frequency space, so it is reasonable that they would not be very noticeable in any one frequency slice. The meteor head echo is visible in the image showing the first positive frequency shift, 7.8 kHz or -23.4 km/s in Doppler range rate. Although the true meteoroid velocity is higher, frequency spreading from discretization causes the signal to appear in this neighboring slice. A faint background of the trail is also visible in this image, along with a few points of higher SNR. The ghost trail is due to the frequency ambiguity of the code, and it appears as a result

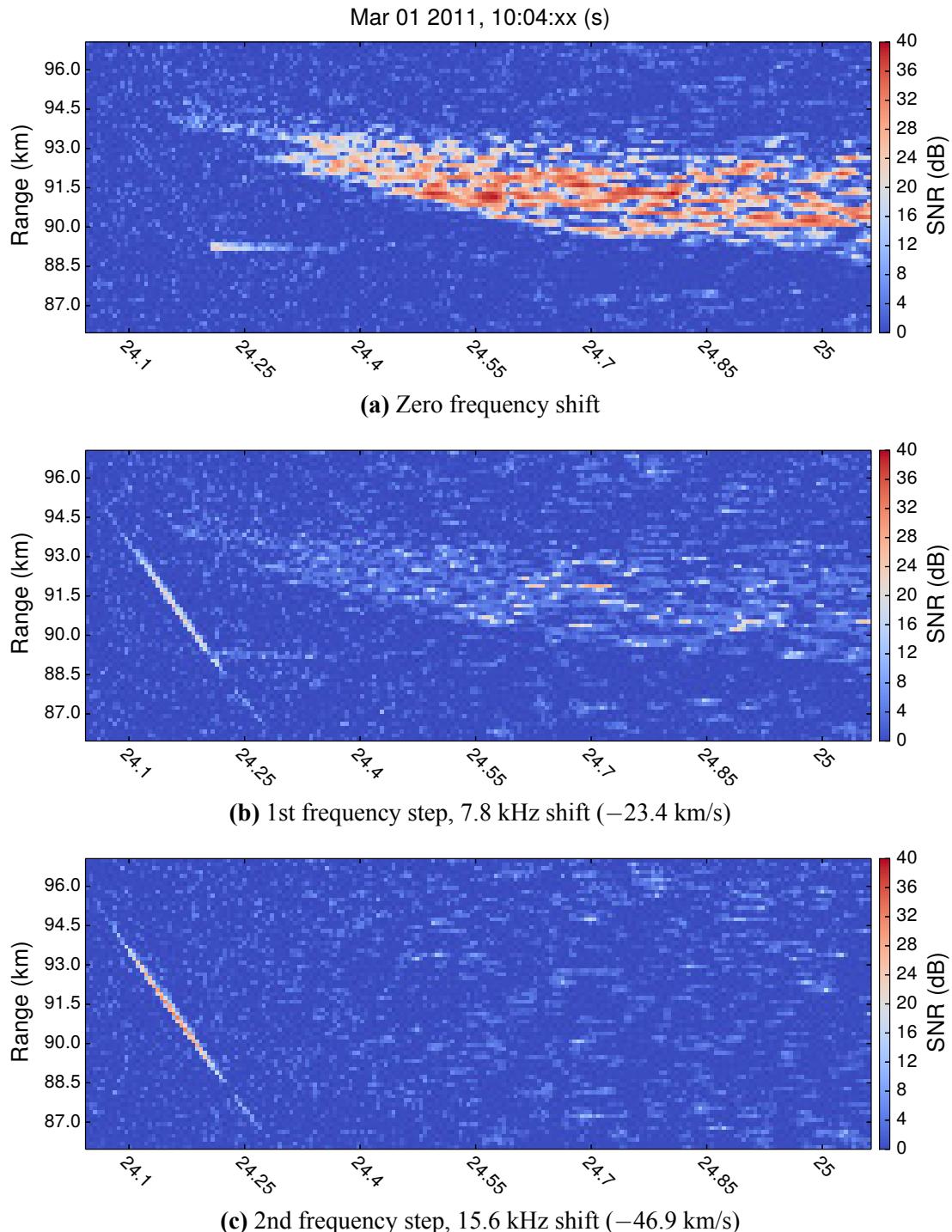


Figure 6.6 – Frequency slices of waveform inversion. These images show the same meteor head, flare, and trail from Figure 6.5 decomposed over the first three frequency steps of the waveform inversion result. The stationary flare and trail appear in the zero-frequency slice, while the moving head echo appears at positive frequency shifts.

of applying the matched filter to the sparse solution's residual. This step is necessary to de-bias the sparse result and restore the signal that was shrunk toward zero. The second frequency step, 15.6 kHz or -46.9 km/s, is closest to the Doppler range rate of the head echo, and so its signal is strongest in the third plot. For the first time, we can also see that the head echo continues after disappearing for a few pulses; this is most likely due to the meteor exiting the main radar beam and reappearing in an antenna sidelobe. It is impossible to identify this feature in the matched filter or even the frequency-integrated result, but frequency decomposition allows us to find it. The absence of trail signal at the second frequency step is also reassuring, lending support to the theory that the signal seen in the first step is due to discretization spreading.

6.4.4 Sparse Solution and Residual

Next we investigate the separate sparse and measurement residual terms that form the waveform inversion result. The composition of the zero-frequency slice of the previous meteor example is shown in Figure 6.7. Figure 6.7a repeats the complete waveform inversion image, while Figures 6.7b and 6.7c show the sparse and residual components.

As expected, the sparse solution contains all of the significant coefficients that represent the flare and trail scattering. It is truly sparse, with most of the shown zero-frequency coefficients and the vast majority of the unshown nonzero-frequency coefficients equal to zero. The sparse solution is excellent for identifying targets and examining the high-SNR components, but it does not provide all of the needed information. Since the sparse solution by itself is biased toward zero, the residual contains not only noise but also remaining elements of the signal that fall below the detection threshold. The additional signal that is necessary to de-bias the sparse solution is clearly visible for the flare and trail. This low-SNR element also includes crucial details like the initial formation of the trail and the continuation of the flare. Since it is matched filtered, the residual also contains sidelobes for that signal, but the sidelobes are necessarily so far below the noise level that they are inconsequential. When combined together into the waveform inversion result, all of the relevant features of the signal are visible for easy comparison and analysis. Nevertheless, it is still important to remember how those elements were formed.

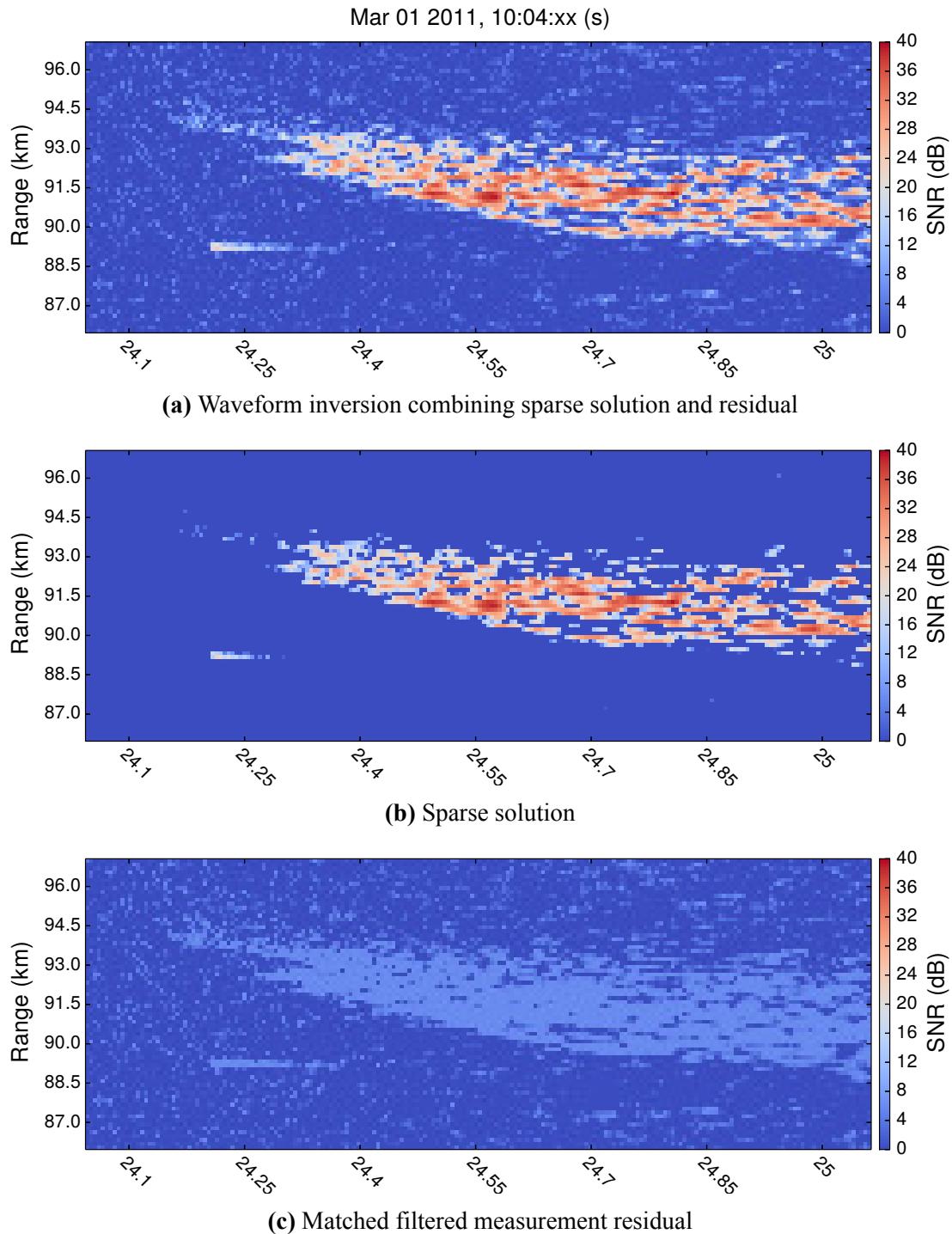


Figure 6.7 – Decomposition of waveform inversion as sparse and residual solutions. The sparse solution is found by minimal l_1 fitting to the radar model, while the residual term is formed by matched filtering the remaining measurement error. Both contain important features that are added together to compose the waveform inversion solution.

6.4.5 Detection Sensitivity

In Figure 6.8, we show RTI images for a different meteor in order to lend more support to our waveform inversion observations. The matched filter result for the pseudorandom code is shown, followed by the corresponding waveform inversion result. A third image showing the matched filter result for the LFM chirp is also given in order to compare detection of the entire head echo. The overall conclusion taken from this data is the same as that from the previous example: waveform inversion is very successful at eliminating sidelobes and making obscured features visible, but it can also include artifacts from waveform mismatch. Comparison with the LFM chirp, which features the best matched filter results due to its unique delay-frequency ambiguity, reveals something new: signals with lower SNR are not always recoverable with a given waveform. The head echo is longer and better-defined with the LFM chirp. Waveform inversion makes more of the head echo evident than what is visible with the comparable matched filter, but we have clear evidence that it is possible to do better. The loss of signal on the head echo could just be due to actual- and ideal-code mismatch, but it is at least something to note for analyzing other results. Overall, waveform inversion with the pseudorandom code gives results that are clearly superior to the LFM chirp.

6.5 Waveform Comparison Results

In this section we compare the inversion performance of the various waveforms used in the Jicamarca experiment. Our results show that the success of the technique depends primarily upon the waveform's ambiguity function, but the quality of the recovery is more sensitive to pulse length and the fidelity of the transmitted waveform to its idealized representation.

6.5.1 Minimum Peak Sidelobe Code

The minimum peak sidelobe code enables waveform inversion with as much success as the pseudorandom code. Their identical length means that they should be able to recover signals with the same sensitivity, but it is reassuring to know that the minimum sidelobe code is suitably incoherent. The original meteor example returns in Figure 6.9 with a comparison

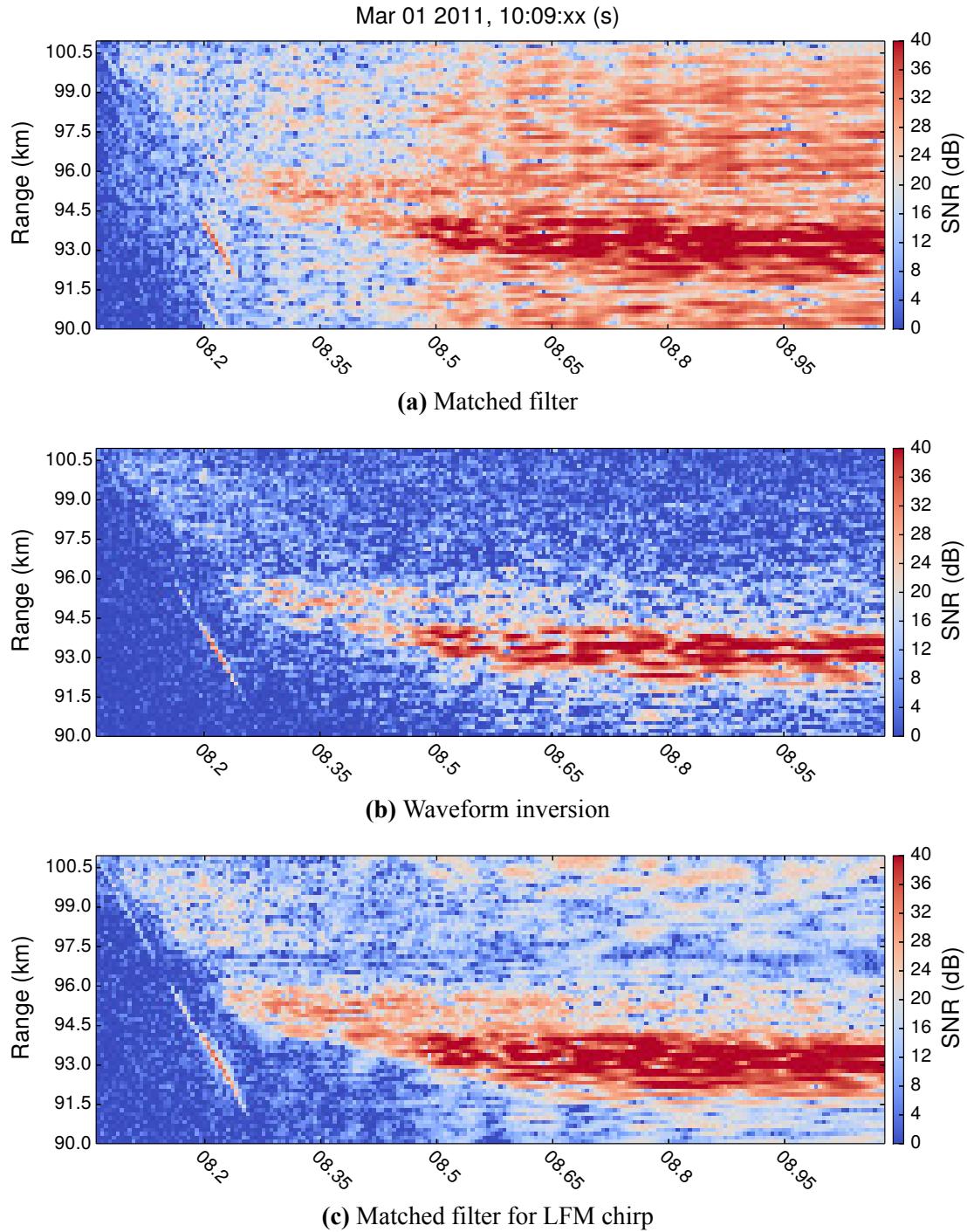


Figure 6.8 – Waveform inversion example for a different meteor. The method is successful even with head echo and non-specular trail signals overlapping in the same pulse. This comparison includes the matched filter result for the neighboring LFM chirp pulses, which does a better job of differentiating the head echo.

between the matched filter, frequency-integrated waveform inversion, and zero-frequency waveform inversion results. Aside from a little leakage of the head echo into the zero-frequency result, the minimum sidelobe results exhibit the same features as the pseudorandom results. Given the better range sidelobe properties of the optimal minimum sidelobe code, one might hope that it might somehow be able to do better than pseudorandom code. While this may still be the case for scenes that push the sparsity limit, the difference between the codes is inconsequential in this case. This fact is not entirely surprising either, since the complete set of delay-frequency sidelobes for the minimum sidelobe code are not necessarily any more suitable for waveform inversion than those of the pseudorandom code.

6.5.2 Barker-13 Code

The Barker-13 code also has good sidelobe properties, so we would expect it to find success as well. Given its shorter length, however, it is less sensitive and more likely to have signals lost in the noise. Figure 6.10 provides the now-familiar comparison images between matched filtering and waveform inversion. Since the observed meteor has a high SNR, waveform inversion is successful with the Barker-13 code in this case. The example also shows the slice for the second frequency shift of the waveform inversion solution. This plot best demonstrates the effects of another downside of a shorter pulse: wider frequency spread in the matched filter ambiguity function. Because the matched filter is applied to the sparse solution’s residual, the wider frequency spread results in the zero-frequency residual ghost trail being visible two frequency steps away. If shorter pulses must be used, the Barker-13 code is still a suitable one for waveform inversion. However, given the benefits of using a longer code and the success we have demonstrated in eliminating sidelobes, there no longer seems to be a reason to restrict oneself to shorter codes when the necessary computing resources are available to do better.

6.5.3 Uncoded Pulse

The uncoded pulse has a terrible ambiguity function, and we might expect that it would be correspondingly ill-suited to waveform inversion. This is not the case, as shown in Figure 6.11. Waveform inversion successfully eliminates the sidelobes of the uncoded pulse and

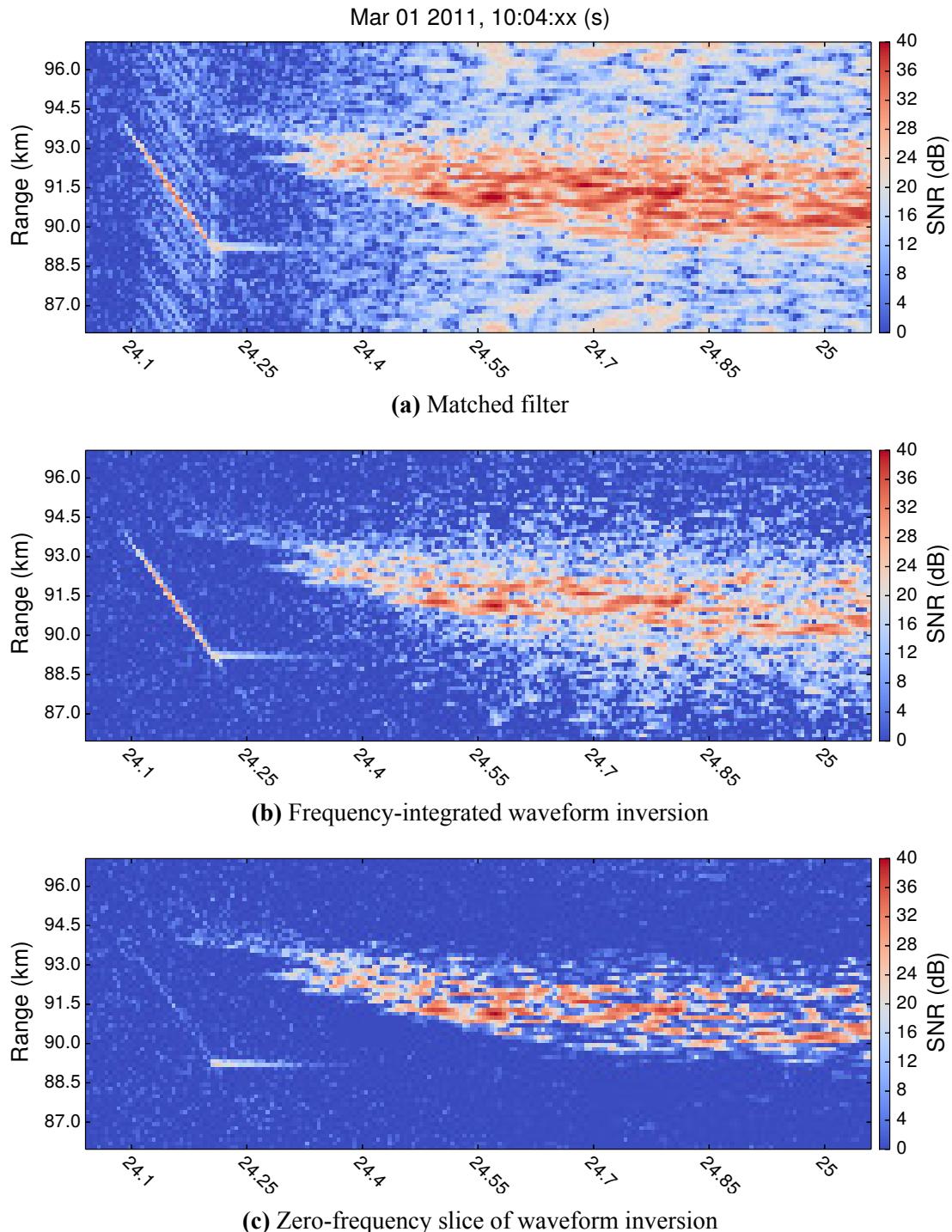


Figure 6.9 – Waveform inversion with the minimum peak sidelobe code. With its good sidelobe properties, the minimum sidelobe code also demonstrates successful waveform inversion. Sidelobes are eliminated to provided a cleaner image, although the frequency-integrated result still suffers from mismatch artifacts.

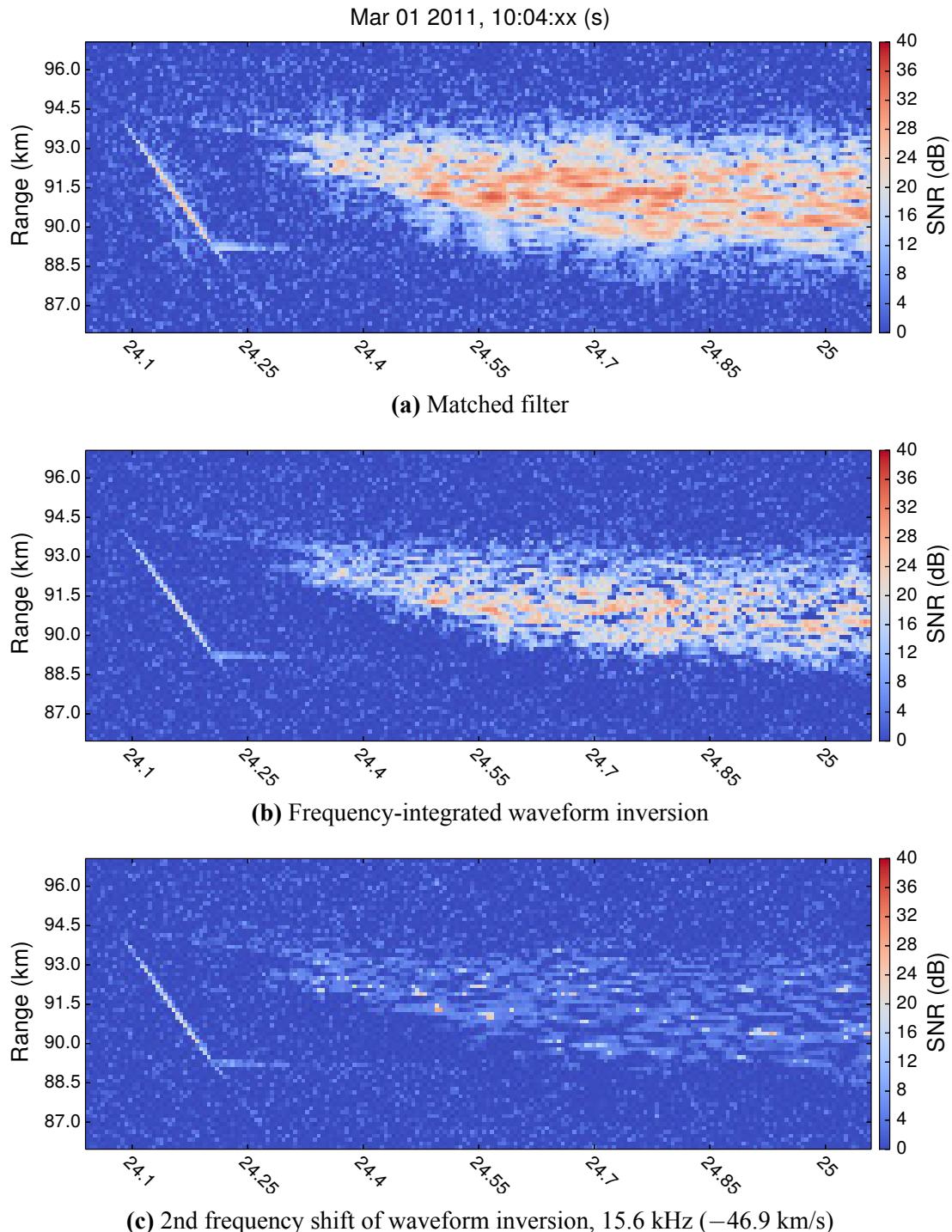


Figure 6.10 – Waveform inversion with the Barker-13 code. The technique is moderately successful with the Barker-13 code. It effectively eliminates sidelobes, but as a result of the pulse's short length, a ghost trail from the zero-frequency matched filtered residual spreads into the 2nd frequency step.

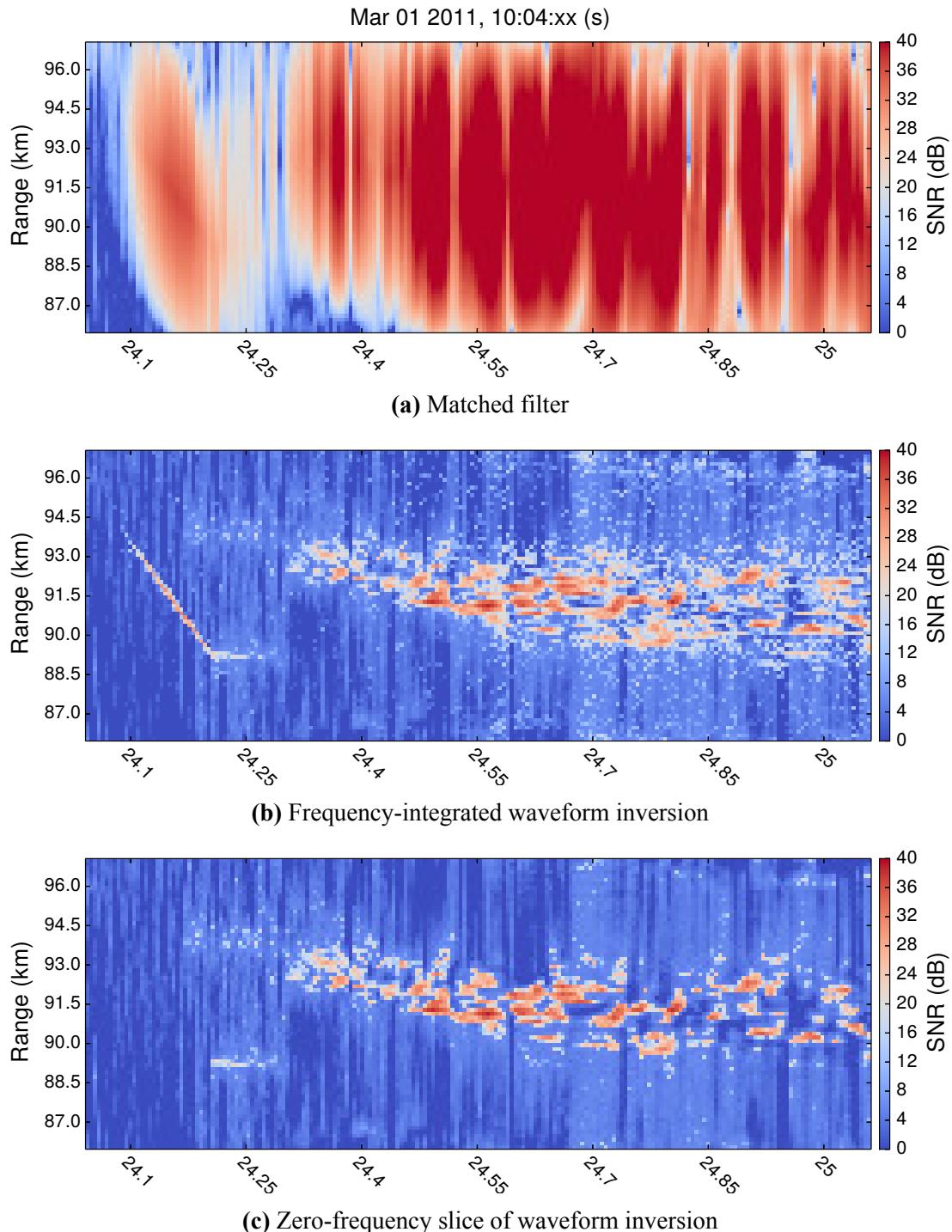


Figure 6.11 – Waveform inversion with the uncoded pulse. Despite sidelobes that completely obscure the matched filter result, waveform inversion is successful at recovering the meteor scatter. Mismatch artifacts are also minimized in comparison to the coded pulses, producing a cleaner frequency-integrated result.

recovers a solution that is similar to the other waveforms. There are both negatives and positives with the uncoded pulse. On the one hand, the zero-frequency result shows that it misses some of the lower-SNR details that the other waveforms reveal in the meteor trail, and the matched filtered residual still produces prominent ghost sidelobes. On the other hand, the mismatch between the ideal waveform and its idealized model is reduced for the uncoded pulse without the complexities of coding, and the frequency-integrated result is relatively free of artifacts. It is still expected that an uncoded pulse will produce measurements that are more coherent than most codes, causing the waveform inversion to struggle in recovering denser solutions. However, the uncoded pulse has taught us that the more important consideration is waveform fidelity. Therefore, uncoded or lower-bandwidth pulses may be used if needed to ensure fidelity.

6.5.4 LFM Chirp

The LFM chirp has a very special ambiguity function; it is compact along a line in delay-frequency space, but within that line it is extremely ambiguous. As it turns out, that ambiguity is too much to overcome, and the LFM chirp is the only waveform that we tested that completely fails waveform inversion. The evidence is given in Figure 6.12. Without a clear delay-frequency peak produced by the matched filter bank, the sparse approximation algorithm can't pinpoint scatterers, and recovery is impossible. In the parlance of compressed sensing, the measurements are too coherent in the delay-frequency dictionary to allow for undersampling. To make matters worse, it must be noted that the radar model doesn't really accommodate the LFM chirp since it requires a discrete code with constant bauds. To perform waveform inversion (or apply digital matched filtering), the continuous frequency sweep has to be discretized. These problems should not be read as an indictment of all chirp waveforms, however. Truly discrete chirps require no special treatment, and the ridge ambiguity is only a property of the linear chirp. Judging from the ambiguity function, we expect that a quadratic or other type of chirp (e.g. an Alltop sequence) would have properties sufficient for inversion.

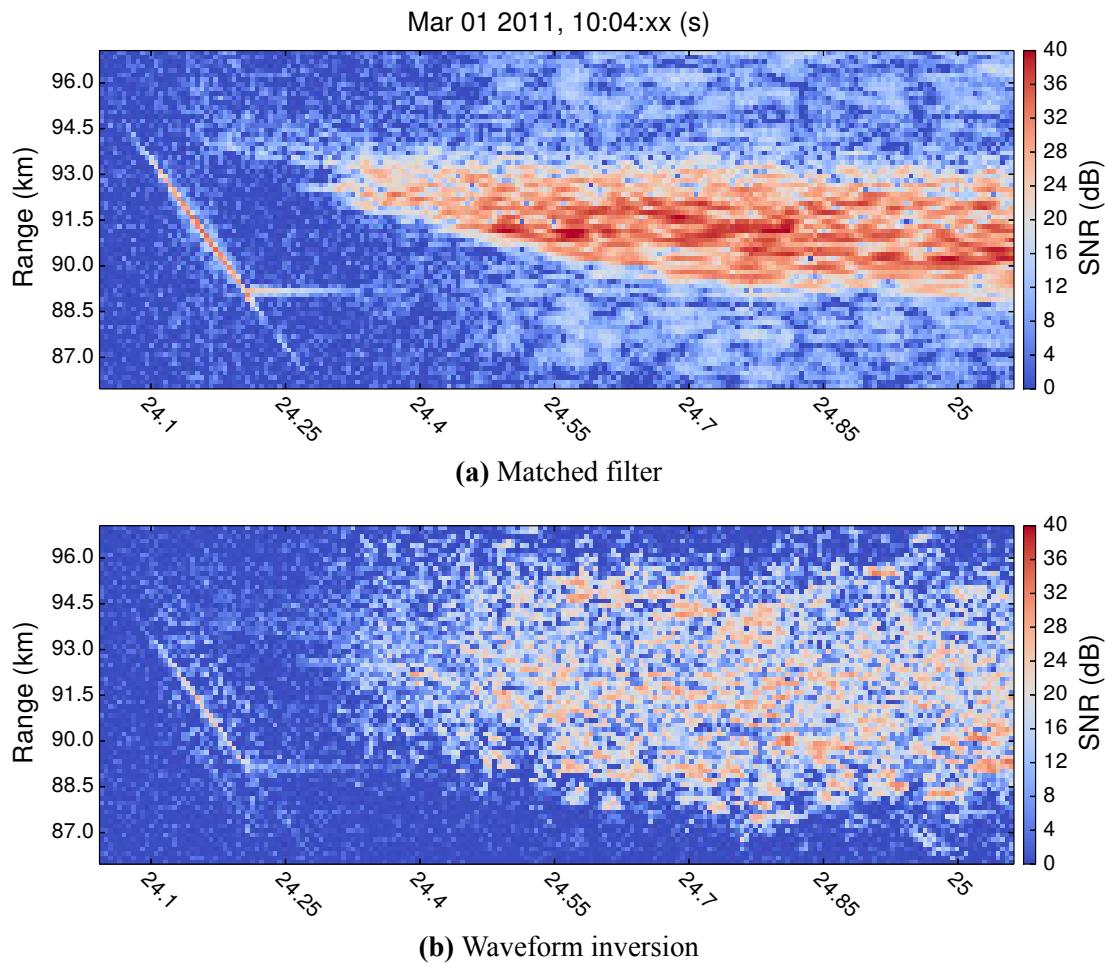


Figure 6.12 – Waveform inversion with the LFM chirp. The ridge of delay-frequency ambiguity, a defining property of the LFM chirp, makes waveform inversion impossible. The head echo gains more sidelobes but is still visible, while the trail scattering is represented by completely non-physical reflectivities.

6.5.5 Inter-waveform Comparison

The relative performance of each of the waveforms can already be inferred by comparing the plots of each of their results, but displaying them together will make the comparison easier. We return to the second meteor example and present the waveform inversion results for the pseudorandom code, minimum sidelobe code, and uncoded pulse. These are shown in Figure 6.13 as frequency-integrated RTI plots and in Figure 6.14 as zero-frequency slices. The comparison of these results is our basis for claiming that the artifacts present in the solution, and most noticeable in the frequency-integrated plots, are based on waveform mismatch. Each of the waveforms produces an image that agrees for the location and strength of the significant scattering. Sensitivity to low SNR scattering is more variable, but by examining the frequency slices it becomes clear which of the lower-SNR signals are true scattering and which are artifacts spread across all frequency steps. Since the form of these artifacts changes with each of the codes, they are a code-dependent property. Since the artifacts are minimized with the uncoded pulse, we conclude that they are a result of mismatch between the actual waveform and its idealized model.

The zero-frequency slices are more illuminating when it comes to revealing the recovery properties of the waveforms themselves and not that of these particular imperfect measurements. The slices clearly show that the coded pulses are strictly better than the uncoded pulse in accurately recovering the meteor trail. Significant features present in the coded pulse results are absent with the uncoded pulse, and it seems more likely that it is a failing of the latter rather than a coincidence of the former. Stripping away most of the waveform artifacts, as with these zero-frequency inversion results for the coded pulses, gives the clearest image of trail scattering that we can produce. Their degree of similarity is amazing, and it must be representative of the fact that we are getting close to the true solution.

6.6 Undersampling Results

One unexplored property of the waveform inversion method is the ability to set the model undersampling factor R to a value greater than 1, which would achieve higher resolution results. The method has proven useful for sidelobe elimination, but sparsity could allow

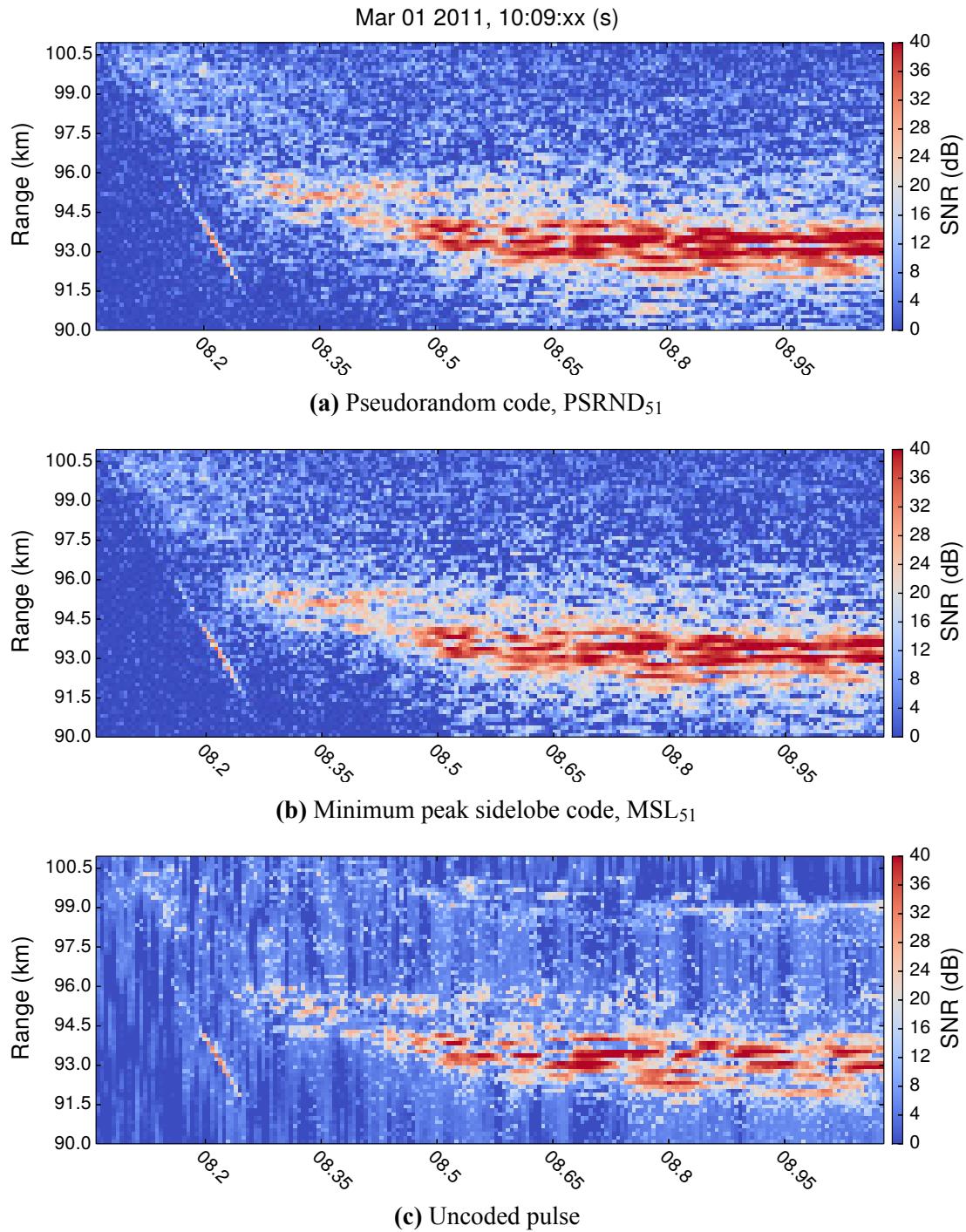


Figure 6.13 – Frequency-integrated waveform inversions of different waveforms. These images compare results for the second meteor example. They reveal that the significant signals are common to all of the waveforms, but the lower-SNR artifacts differ. This is evidence that the artifacts are related to transmitted waveform mismatch.

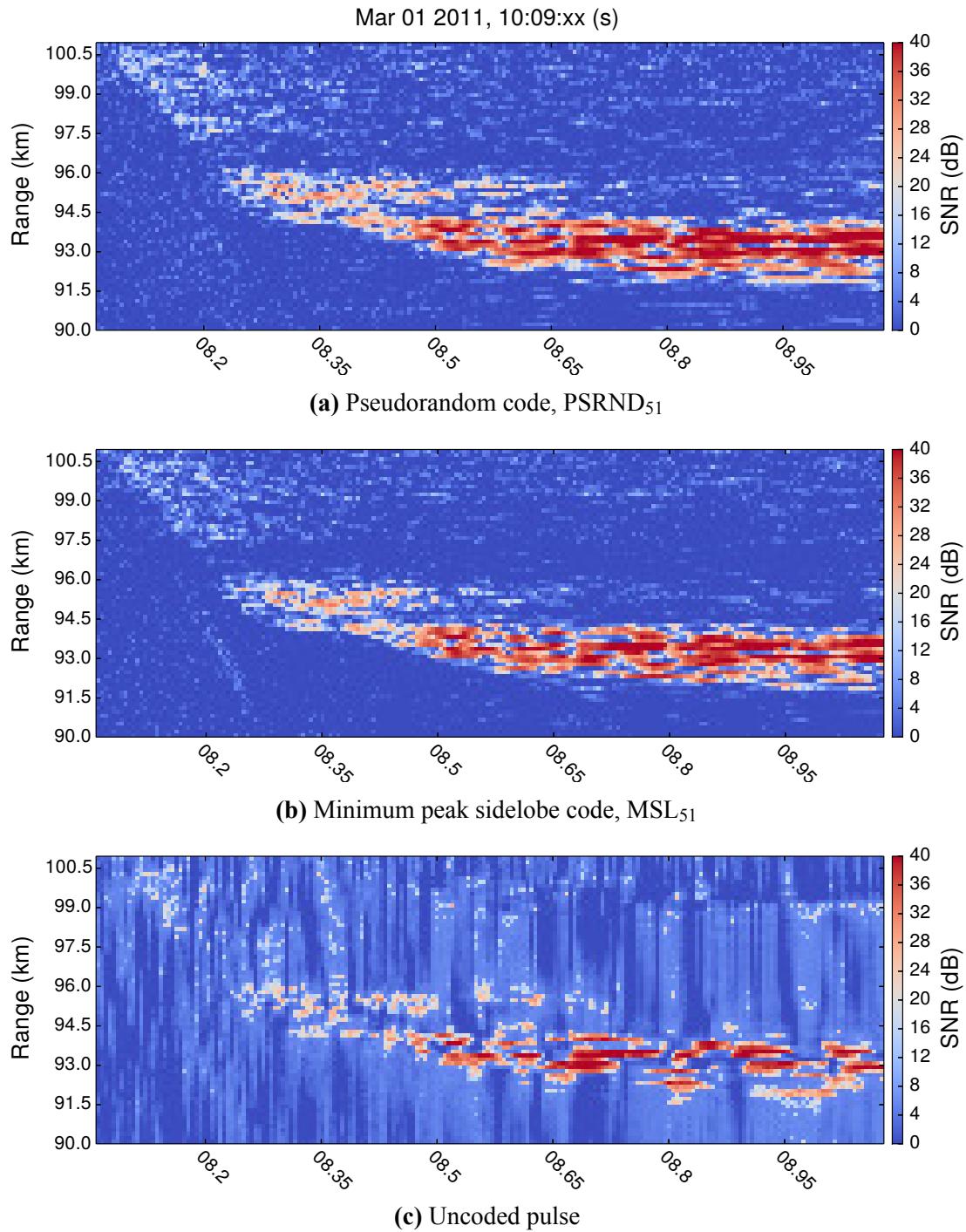


Figure 6.14 – Zero-frequency waveform inversion slices of different waveforms. The coded pulses produce remarkably similar results when mismatch artifacts are not considered, while the uncoded pulse lags behind somewhat. The consistency is evidence that the results are close to the true sparse solution.

us to accomplish more by truly taking advantage of the compression aspect of compressed sensing. Theoretically, the key to recovery at increased resolution is still measurement incoherence, and the best way to likely ensure that is to increase the bandwidth of the transmitted waveform. As we have seen with the uncoded pulse, however, that is not a necessary condition for success.

6.6.1 Discarding Half the Data

In lieu of being able to validate the increased resolution with $R > 1$ using the matched filter, as we have done with the previous results, we can deliberately undersample the existing data and use our full knowledge as a basis of comparison. In Figure 6.15, we compare waveform inversion with half of the data samples (keeping every other one) to waveform inversion with the full data set. Since the best waveform inversion results have been the single-frequency slices with the pseudorandom pulse, that is what we have compared between the half-sampled and fully-sampled data. Though the frequency-integrated image is rife with mismatch artifacts, overall they are not much worse than what was frequently encountered with the fully-sampled data. The zero-frequency slice tells the full story, though, and the results are mixed. With half the samples, we expect the result to have half the SNR, but much more of the signal is missing. The features with the highest SNR are still present for the most part, but the surrounding scattering with lower SNR is eliminated with half the data. However, the head echo is still recovered quite well, matching our expectations from the full data set exactly. It seems most likely that half the data is not enough to recover a signal with the number of nonzero coefficients required by the trail. Instead of the full solution, what we get is the closest sparse approximation with the maximum number of coefficients that can be recovered with the given number of measurements.

6.6.2 Recovering at Double Resolution

Even with mixed success at recovering consistent results from half-sampled data, we proceed to attempt recovery at double the resolution of the fully-sampled data. At a minimum, we hope that the head echo resolution can be increased successfully. The transmitted waveforms were sampled at their bandwidth, so doubling the model sampling rate nominally

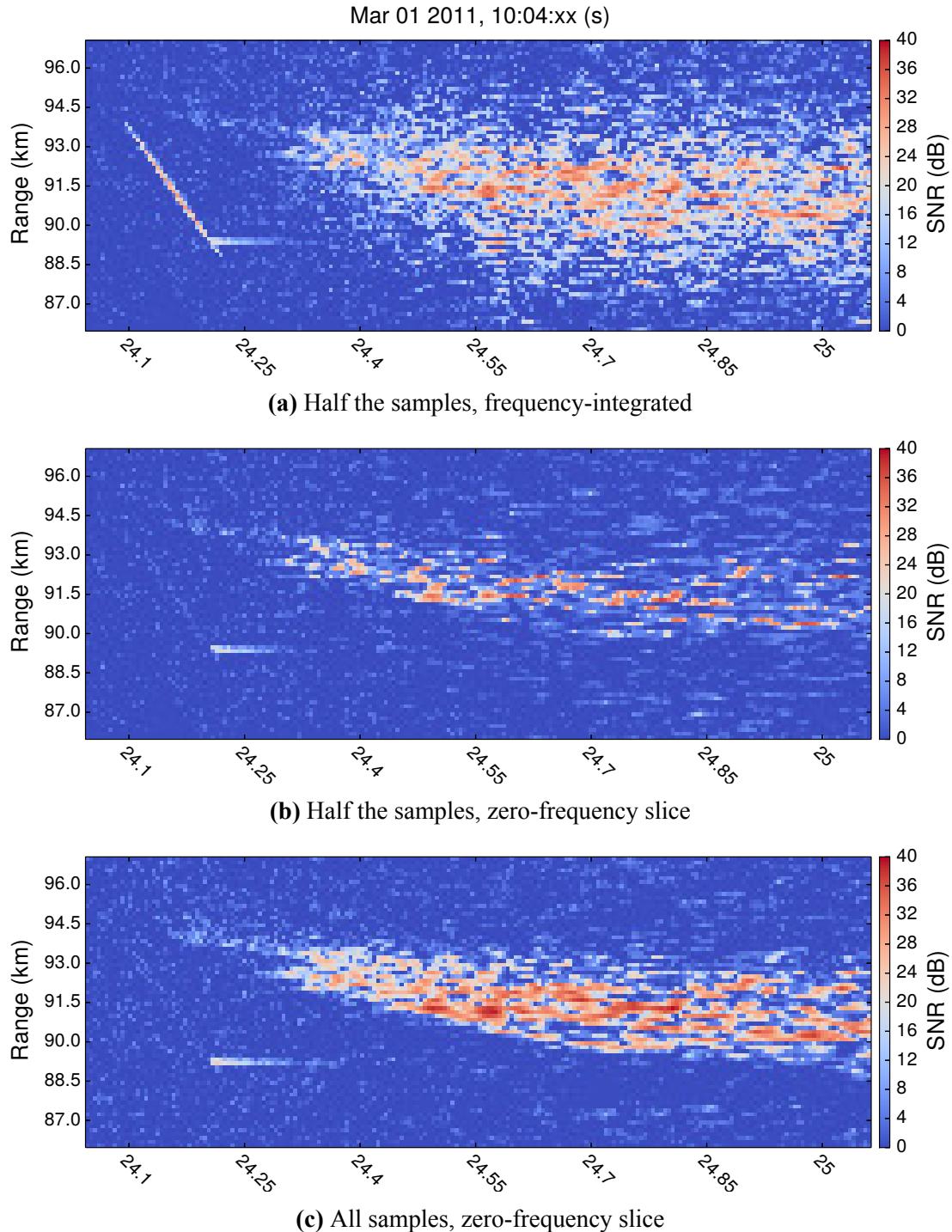


Figure 6.15 – Waveform inversion with half the data. Waveform inversion was performed after discarding half of the pseudorandom pulse samples, and the results are compared to fully-sampled inversion. Head echo recovery is congruent with the previous results, but the trail is too sparse, indicating that the undersampling is too severe.

makes the waveforms ambiguous over the width of two samples. This is not necessarily an insurmountable problem, as recovery with the fully ambiguous uncoded pulse shows. A consistent theme has been that the actual waveforms don't exactly match the idealized waveforms, and in reality they do vary within the "constant" bauds of each code. To perform this double-resolution recovery, we have attempted to use the non-ideality to our advantage. Rather than doubling the length of the ideal codes by repeating each of the bauds, we simulated the effects of the transmitter's limited bandwidth by upsampling and lowpass filtering the ideal codes. The result is a double-resolution waveform that should vary in a manner similar to the actual waveform.

Figure 6.16 shows the double-resolution recovery results for the first meteor example using the pseudorandom waveform data and the upsampled, lowpass-filtered pseudorandom code. Figure 6.17 shows the same for the second meteor example. Whether the difference is a better relative sparsity or the fact that the lowpass-filtered code better represents the true waveform, these double resolution results *look* better than the half-sampled results. In both examples, the mismatch artifacts appear to be noticeably reduced in the frequency-integrated image, and the double-resolution image has all of the same significant features as the normal-resolution image. In fact, it is easy to imagine that the original results represent a blurred version of the "true" increased-resolution results. Do the localized striations of the double-resolution trails give a better representation of the plasma and scattering physics? Without further support and given the half-sampled recovery results, it is impossible to declare victory and claim that we have successfully doubled the range resolution of our measurements. Regardless, the method looks promising and definitely merits further research.

6.7 Recommendations

6.7.1 Waveform Selection

The waveform inversion results have demonstrated the method's usefulness, eliminating sidelobes in these examples and in our additional informal testing. It works with a variety of codes, even with uncoded pulses, but the best results were achieved with the longest

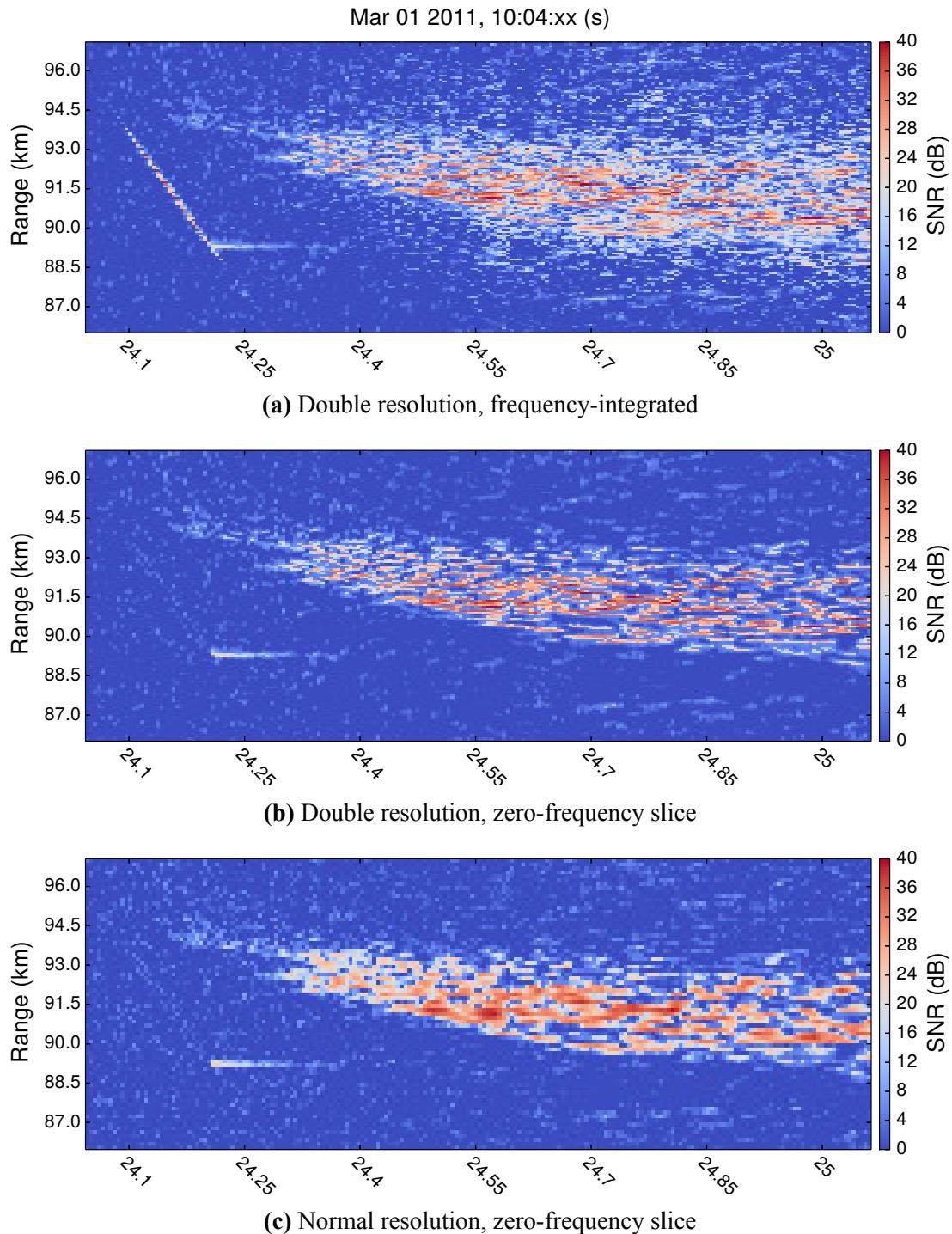


Figure 6.16 – Waveform inversion at double the resolution. To recover a double-resolution solution for the first meteor example, we upsampled and lowpass-filtered the pseudorandom code to simulate the effects of the actual transmitter's limited bandwidth. The results are promising, with all features represented consistently.

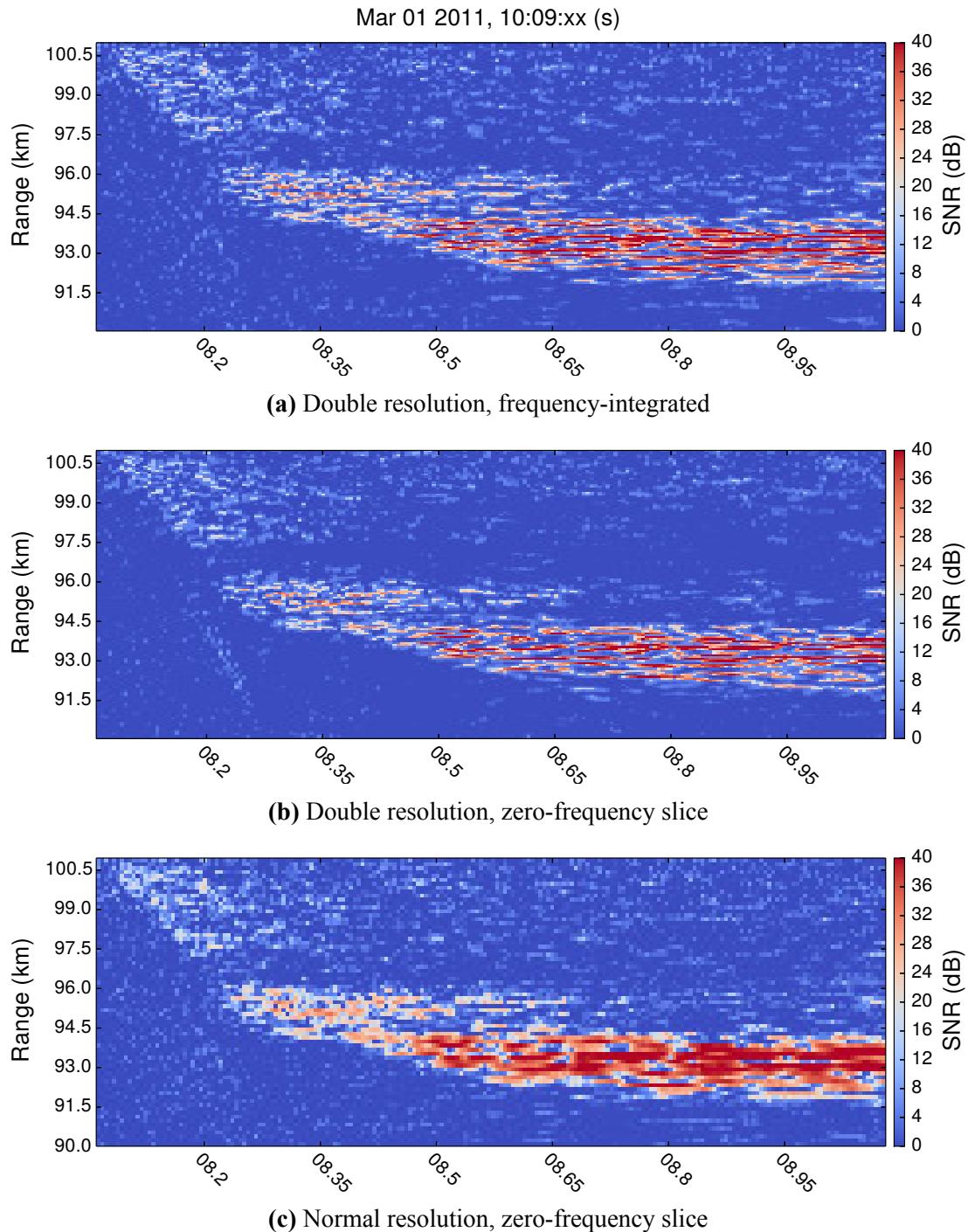


Figure 6.17 – Waveform inversion at double the resolution, second example. Doubling the resolution of the second meteor example also produces promising results. One can easily imagine that the normal-resolution image is a blurred version of the double-resolution image. However, it is impossible to determine whether that is truly the case.

codes. It is impossible to replace the sensitivity provided by a long code, and with the effect of sidelobes minimized, there is no reason to not use the longest possible pulse length. The other consideration is the code itself, and on that point we have less evidence. Removing pulse length and the clearly-suboptimal uncoded and LFM chirp ambiguities from the equation, the remaining three codes seemed to find similar success. Based on our knowledge of the solution procedure as iterative matched filtering and thresholding steps, we can make the educated guess that codes with lower peak sidelobes across the entire delay-frequency space will perform better. There is obviously less work for the recovery algorithm to do when the matched filter already does a good job. Overall, the recommendation is to use long codes with minimal delay-frequency sidelobes, which are not necessarily the minimum peak range sidelobe codes.

The other waveform consideration is fidelity with the actual transmission. The most significant artifacts in the inversion result are likely an effect of mismatch between the transmitted waveform and its idealized representation in the model. The relative success of the uncoded pulse, at least with respect to artifacts, points to this conclusion. Codes should be selected knowing the bandwidth constraints and performance characteristics of the transmitter in order to ensure that the outgoing waveform is as desired. There will always be non-idealities, however, so it is recommended that one take the additional step of sampling the transmitted waveform. With transmitter samples, it is possible to use code values that most closely match the actual waveform, and thus mismatch artifacts should be minimized.

6.7.2 Frequency Analysis

Another problem we encountered is the poor frequency resolution that results from using a relatively low-frequency radar like Jicamarca. Future work could involve developing a multi-pulse inversion method, or alternatively, use a radar that operates at a higher frequency. Another possibility for increasing frequency resolution is to increase the number of frequency steps N included in the radar model. Doing so can help to improve relative sparsity as discussed in Section 3.3.3, but there are limits. Larger N means more unknown variables, which means proportionally longer calculations. Larger N also means making

the measurements more coherent in terms of the solution space, so there is a point at which increasing N would result in the failure to find the true solution. For these reasons, we recommend setting N as low as possible to achieve the desired frequency resolution, optimally to the minimum value of the length L of the waveform code. When this is done with sufficiently long codes, the true solution is likely to still be sparse enough to permit recovery.

The ability to decompose the waveform inversion solution into different frequency slices is a real strength of the analysis method. Though a frequency-integrated plot gives a good overview of the recovered scene, the individual frequency slices can be used to identify particular targets and better-observe their details. Even when targets that scatter at different frequency shifts are present at the same ranges, waveform inversion can separate them for individual analysis. This ability is most useful for crowded radar scenes like those encountered in the equatorial ionosphere at Jicamarca. Making full use of this capability when designing waveform inversion experiments is highly recommended.

6.7.3 Computation

One final difficulty with the waveform inversion method is the time it takes to compute the recovery solution. Instead of applying the matched filter once per pulse, inversion requires, at minimum, applying it and the forward model on the order of a hundred times. Even with an efficient implementation and quick convergence, that takes a significant amount of time. Fortunately, the method as it stands now is eminently parallelizable; each pulse is processed individually, and there is no reason that they cannot be processed at the same time instead of in sequence. For larger applications, this would certainly be recommended.

Part IV

Conclusion

Chapter 7

Conclusion

With the goal of enabling flexible, high-resolution measurements of ionospheric plasma phenomena, this thesis draws upon the subjects of radar analysis, sparse approximation, and convex optimization to develop and apply a novel waveform inversion method. In Part I, we introduced these subjects and described the challenges encountered with ionospheric radar measurements that arise from the inherent conflict between sensitivity and ambiguity. Current techniques struggle to meet the following goals: 1) Differentiation of targets in crowded and variable environments, which is necessary to separate meteor and electrojet signals in the equatorial ionosphere; 2) Elimination of self-interference of range- or frequency-spread targets, including non-specular meteor trails; 3) High range and frequency resolution for observing small-scale processes, such as meteoroid fragmentation and meteor flares; and 4) Flexibility to achieve these goals for a variety of waveform and measurement requirements arising from experimental and hardware constraints. Waveform inversion can achieve these goals. In Part II, we reviewed the traditional radar matched filtering technique and the ambiguity that it introduces, and we proposed a mathematical model to describe radar measurements in terms of ambiguity-free reflectivity coefficients. In Part III, we addressed the complete problem of waveform inversion, from the theory of compressed sensing via convex optimization to the implementation of efficient algorithms and the application to actual radar data. In this chapter, we conclude with a discussion of contributions and a sketch of potential further research.

7.1 Summary and Discussion of Contributions

7.1.1 Radar Model

A new perspective for radar measurements is espoused in which the radar scene reflectivity is recovered by inverting a mathematical model.

Radar measurements in their basic form, a sequence of values sampled from the received signal, are not amenable to information processing when pulses longer than the sampling period are used. Our radar model addresses this by representing the measurements as a function of delay-frequency reflectivity coefficients. The use of a delay-frequency dictionary is useful because the coefficients are typically sparse. A similar delay-frequency analysis is also possible with a bank of frequency-shifted matched filters, but that representation is blurred with respect to the true reflectivities due to waveform ambiguity. We first introduced the radar model in Chapter 3 with exactly this imaging analogy; the ambiguity function acts as a point-spread function when matched filtering is applied to the measurements, blurring the true radar scene's reflectivity. The forward radar model is the adjoint of the matched filter bank operation, and as a consequence, matched filtering gives the least-norm solution to the radar model measurement equation. Instead, knowing that a sparse solution is most likely to be correct, we posed the following waveform inversion problem: invert the radar model with sparsity as a constraint to recover the sidelobe-free radar scene.

A discrete radar model is rigorously derived from the continuous measurement equation for a general target scene.

We followed the informal radar model presentation with a rigorous derivation to determine precisely how general radar scenes are represented. Starting from the continuous narrow-band radar measurement equation, the discrete radar model was re-derived through discretization and Fourier analysis. Notably, this derivation only assumes what is typical for matched filter analysis: a narrow-band signal, no scattering from outside the delay window of interest, uniform sampling of the received signal, a transmission waveform represented by a code sequence describing constant bauds, and a number of frequency steps greater than the length of the code. The resulting radar model is discrete, with a fixed number

of reflectivity coefficients corresponding to a delay-frequency grid. Point targets falling on the grid points are naturally represented, but off-grid point targets and distributed targets are also accommodated. The derivation shows that arbitrary targets are written exactly, though not uniquely, in terms of the gridded coefficients. With respect to delay, the coefficients represent the integrated reflectivity over the corresponding delay window; with respect to frequency, the coefficients represent samples from the complete reflectivity function after it has been convolved with a periodic sinc function. Because of this known relationship, we can infer that sparsity of the full radar scene, as represented by the continuous reflectivity function, is reasonably preserved as sparsity of the discrete reflectivity coefficients. Thus, the waveform inversion procedure is justified.

7.1.2 Waveform Inversion Method

A novel waveform inversion technique is formulated based on finding a sparse representation of the radar scene using the discrete model.

Sparse solutions to underdetermined systems of equations is the domain of the theory of compressed sensing, which states that minimizing the l_1 -norm when constrained by enough incoherent measurements guarantees recovery of the true sparse solution. In Chapter 5, a review of theoretical results indicated that recovery of the reflectivity coefficients using the radar model is likely to be successful provided that a suitable waveform is used. All that is required are global, incoherent measurements, which are ensured by random binary-phase sequences, the Alltop quadratic chirp sequences, and likely many more waveforms. With successful recovery, waveform inversion meets all of the goals that existing methods fail to reach. Elimination of delay-frequency sidelobes, and their ambiguity, is inherent to the method. That, coupled with the simultaneous decoding over the signal's full frequency spectrum provided by waveform inversion, enables target differentiation for multiple or distributed scatterers in crowded and variable environments. Given a sufficient level of sparsity, inversion can also provide solutions with a higher resolution than that specified by the sampling rate, thus enabling better observations of small-scale processes. Since recovery is likely with many different waveforms, the method is also flexible enough to be used

when there is little or no freedom in choosing the waveform or dictating the experimental setup.

In practical terms, the waveform inversion solution procedure is straightforward. Our approach, detailed in Chapter 5, is to solve an l_1 -regularized least squares problem using a first-order convex optimization method that involves the soft-thresholding proximal operator. Algorithms like this have an intuitive interpretation in radar terms as an iterative thresholding matched filter: each step involves applying the matched filter to detect under-represented components and thresholding the new representation to induce sparsity. The optimization problem's regularization parameter λ is chosen to ensure a constant false-alarm rate for detection based on an independently-estimated noise level. In order to compare the inversion to matched filtering and de-bias the result, the solution is scaled and added to its matched filtered measurement residual. Adding the residual term also has the benefit of including any signal that falls below the detection threshold. As a result, the regularization parameter (detection threshold) serves as the threshold for sidelobe elimination: sidelobes are removed for signals with magnitude above the threshold and left intact for signals with magnitude below the threshold. Larger thresholds result in faster convergence, so there is a direct trade-off between speed and sidelobe elimination.

The waveform inversion method is implemented using modern convex optimization techniques tailored for efficient computation.

Our waveform inversion method was implemented as a pair of Python packages, `radar-model` and `prx`, that provide the radar model operators and the proximal convex optimization algorithms. The forward and adjoint operations of the radar model are applied repeatedly in iterating to the solution, so they must be as efficient as possible. Our implementation uses the FFT (fast Fourier transform) algorithm and multiple formulations of each operator to minimize computation time for arbitrary model dimensions. To ensure quick convergence to the solution, the convex optimization algorithms that we implemented include adaptive acceleration and step size schemes, including one of our own devising in the case of the linearized ADMM (alternating direction method of multipliers) algorithm.

We tested multiple first-order convex optimization algorithms that use the proximal operator to show that the accelerated proximal gradient method applied to l_1 -regularized least

squares converges the most quickly, in general, for the radar model. The most important point concerning these implementations, however, is that they are flexible and easy to understand. Algorithm development for problems of this type is advancing rapidly, and it is reasonable to expect faster, better algorithms to emerge before long, if they do not exist already. The `radarmodel` package can be used with any future algorithm that permits direct computation of the model operators, while the proximal methods found in `prx` are simple enough to easily accommodate improvements or, if completely obsoleted, serve as an intuitive explanation of the inversion process.

7.1.3 Experimental Results

The real-world flexibility and effectiveness of the inversion technique is demonstrated by the elimination of filtering artifacts from meteor observations made with a variety of standard radar waveforms.

Compressed sensing theory indicates that the waveform inversion technique is well-posed, but the more pertinent question concerns its practical application to analyzing ionospheric radar data. As proof of its effectiveness, we performed waveform inversion on meteor measurements collected using the Jicamarca 50 MHz radar. The results were dissected pulse by pulse, analyzed as individual frequency slices, and integrated over all frequencies to show that matched filter sidelobes are effectively eliminated with minimal remaining artifacts. Five common waveforms were compared in the experiment by transmitting them in a repeating sequence, one per pulse, so that they could be observed scattering from the same targets. Based on this sample, in which inversion was a moderate success using an uncoded pulse and a complete failure using a linear frequency-modulated pulse, we concluded that successful inversion depends on the compactness of the peak of the waveform's ambiguity function. Comparing the inversions to each other revealed that waveform-dependent artifacts remaining in the solutions were likely due to mismatch between the transmitted waveform and its idealized model. Additional inversions were attempted with undersampled data, producing mixed results that are promising for increased-resolution recovery but require further testing to verify. For future applications, we concluded that long, coded

waveforms with good fidelity should be the goal, in order to achieve the best sensitivity and delay-frequency resolution with minimal mismatch artifacts.

7.2 Future Work

In this dissertation, we have built a strong foundation for the use of sparsity to achieve waveform inversion of radar measurements. Based on demonstrated and theoretical potential, this foundation warrants further development so that its promise is realized. There are many avenues for future research based on this work. Some involve specific applications of the waveform inversion technique, while others would strengthen the theory or develop additional methods. The following topics are illustrative of all of these possibilities.

7.2.1 Incoherence and the Ambiguity Function

Our experimental results have hinted at a strong connection between the radar ambiguity function and the compressed sensing concept of measurement incoherence. Indeed, they both describe the degree to which individual measurements are related. With a little effort, a direct relationship giving the incoherence in terms of the ambiguity function should be easy to establish. One consequence would be the ability to prove that certain waveforms are optimal for compressed sensing. The more interesting aspect of this concept, however, comes from the fact that both topics are individually well-studied. What known results concerning waveform ambiguity can be applied to incoherence, and vice-versa? Perhaps codes that are well-known to the radar community would be very useful in other fields of compressed sensing, and maybe the reverse is true as well. It is certainly possible to continue to apply the waveform inversion technique without further exploring this relationship, but we believe applications would benefit from more theoretical background in this area.

7.2.2 Super-resolution

The experimental results concerning double resolution reconstruction are enticing in their supposed quality, but we have not produced enough evidence as of yet to verify them. Even beyond doubling the resolution, sparse recovery has the potential to increase resolution right

up to the undersampling limit prescribed by the recovery phase transition. A proper verification of these capabilities would involve deliberately discarding data when performing the reconstruction, much as we did for our half-sample results. Where we believe our results failed, and where future experiments must succeed, is in controlling for the effect of transmitted waveform fidelity. Incorporating the effects of lowpass filtering seemed to improve our double-resolution results compared to the half-sample results, and it makes sense that correctly matching the transmitted waveform with the model would be more important when working with less data. If super-resolution can be demonstrated and perfected, waveform inversion would be unique in reliably and automatically providing that capability.

7.2.3 Multi-pulse Analysis

Another avenue of research encouraged by the experimental results is the ability to combine data from multiple pulses to improve recovery. Provided that the radar scene maintains the same delay-frequency reflectivity over the length of multiple pulses, the combined data can be used to significantly increase the recovered frequency resolution. Even if some of the target properties change from pulse to pulse, the data is still likely to be correlated in some way. In a sense, this makes the scene even more sparse; instead of recovering each pulse individually, we can use correlation from pulse to pulse as a type of time sparsity to recover the reflectivities for multiple pulses using fewer measurements. Many formulations of this type of recovery are possible. One simple method would be to weight the individual coefficients in the l_1 -norm penalty by the inverse of the previous pulse's solution, placing a higher penalty on coefficients expected to be zero. A statistical approach could involve a Kalman or particle filter to model the state of the radar scene and include measurement uncertainty; each pulse's sparse estimate could then be recovered while incorporating and updating the estimated statistical distribution. While the benefits of multi-pulse analysis are enticing, implementations will have to be careful to not significantly increase the dimensions of the model and thereby make recovery practically impossible.

7.2.4 Alternative Radar Models

The grid-of-point-targets radar model is successful and attractive because of its simplicity and direct relationship to established techniques, but there are aspects of it that can be improved. One troublesome feature is the blurring of the radar scene's frequency spectrum in the reflectivity coefficients by the periodic sinc function. Though it doesn't completely eliminate sparsity, it does increase the number of non-zero coefficients. Another troublesome aspect of the current radar model is its frequency coherence. In order to increase frequency resolution, it is necessary to increase the coherence of the measurements. The radar model is functionally equivalent to time-frequency analysis using the short-time Fourier transform, and their shortcomings are the same. Windowing is often used in Fourier analysis to minimize frequency spreading from discretization, so a windowed radar model could be an improvement. Wavelets are often a better method for analyzing signals that are compact in time and frequency, so a wavelet-based radar model would provide a better representation of the expected sparsity. Other techniques from the field of time-varying signal processing could surely be adapted as well. The key to implementing and actually using these enhancements in practice would be to ensure that they do not significantly increase the already-restrictive computational complexity of the model.

7.2.5 Radar Self-calibration

Waveform fidelity is an important issue for minimizing mismatch errors in the recovered solution, and the best way to know the transmitted waveform is to measure it directly. When this is not possible, however, a type of radar self-calibration could suffice. Consider the typical situation, in which we use an assumed ideal waveform to recover a sparse solution. The result includes the signal that we want, but it can also have mismatch artifacts that manifest as seemingly-random non-signal components in the sparse solution. In other words, we know that the true solution should actually be even more sparse. Ideally, we would be able to adjust the modeled transmitted waveform so that the recovered solution reaches the true level of sparsity and is free of artifacts. Self-calibration would do this by flipping the waveform inversion problem on its head, instead using the recovered reflectivity to solve for the transmitted waveform from the measurements. Perhaps, instead, it simultaneously

recovers both the transmitted waveform and the reflectivity coefficients. Either way, we would apply knowledge about the form of transmitted waveform, along with knowledge that the reflectivity coefficients should be sparse, to constrain the solution. Since it is completely ambiguous whether a specific effect is attributed to the transmitted waveform or the reflectivity coefficients, applying the appropriate constraints would be the difficult part. If it could be made to work reliably, however, the quality of the solution would be unsurpassed.

7.2.6 Passive Radar

Passive radar is the practice of intercepting existing radio transmissions to make radar measurements [Lind *et al.*, 2013]. One notable existing transmission is digital television, and with their large coverage and wide bandwidth, digital television signals make a very attractive target for passive radar. Since we obviously cannot control the waveform of the intercepted signal, it is not possible to use codes with good sidelobe properties to make the measurements. Consequently, delay-frequency sidelobes can pose a problem to passive radar accuracy, and waveform inversion provides an obvious improvement. The principal problem with direct application of the technique is that passive radar already requires significant processing power when using just a matched filter. Adding hundreds of iterations to perform waveform inversion would make the computation time unmanageable. Though the application to passive radar would appear to be straightforward, computational efficiency is the problem that would have to be solved to enable success.

7.2.7 Near-universal Measurement Modes

One difficulty with current ionospheric measurements is that the experiment has to be tailored to the specific phenomena being measured. The varied scattering processes cover a very wide range of both spatial and temporal coherence as a function of altitude, and these properties impose constraints on the pulse length and inter-pulse period needed to ensure unambiguous sampling. The requirements are often at odds for different processes, preventing simultaneous measurement with the same radar. For instance, ionospheric D-region scattering is relatively compact in altitude extent but has a long correlation time, requiring long sequences of short pulses to measure its autocorrelation function. Conversely, F-region

scattering has a wide altitude extent but a short correlation time, which requires oversampling a long pulse to measure the autocorrelation well. Delay-frequency sidelobes prevent measuring the D-region with a long pulse, so these requirements are at odds. Waveform inversion, however, eliminates sidelobes and makes simultaneous measurement possible. In fact, in order to properly sample any autocorrelation function over a wide range of altitudes under a transmitter duty cycle constraint, it is necessary to use aperiodic waveform sequences [*Virtanen, Vierinen, and Lehtinen, 2009*]. Matched filtering is wholly inadequate for decoding such sequences, but waveform inversion can decode them and enable a near-universal measurement mode for the entire ionosphere.

Part V

Appendix

Appendix A

Simplification of the Reflectivity Coefficient Equation

The radar target scene is represented in the discrete radar model by reflectivity coefficients $h[n, p]$, defined in equation (3.27) as a result of the radar model derivation as

$$h[n, p] = e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_{n+kN}(\rho + \rho_0) e^{-2\pi i(n+kN)\nu_0\rho} d\rho. \quad (\text{A.1})$$

This equation, with its infinite-sum Fourier series expansion of the reflectivity function, is not very illuminating when it comes to finding an intuitive interpretation of the reflectivity coefficients. Fortunately, such an interpretation, given in Section 3.3, is possible after some simplification. This chapter documents the steps for finding a simpler representation of the reflectivity coefficients.

We start by recalling the definition of the Fourier series coefficients and writing them in terms of a windowed Fourier transform:

$$h_k(\rho) = \nu_0 \int_0^{1/\nu_0} \tilde{h}(t, \rho) e^{-2\pi i k \nu_0 t} dt \quad (\text{A.2})$$

$$= \nu_0 \mathcal{F}_t \left[\tilde{h}(t, \rho) \Pi_{1/\nu_0} \left(t - \frac{1}{2\nu_0} \right) \right] (k\nu_0) \quad (\text{A.3})$$

$$= h_w(k\nu_0, \rho), \quad (\text{A.4})$$

where \mathcal{F}_t represents the Fourier transform with respect to t and $\Pi_T(t)$ is the rect function defined as

$$\Pi_T(t) = \begin{cases} 1 & |t| < T/2 \\ 0 & |t| \geq T/2. \end{cases} \quad (\text{A.5})$$

As is clear from this definition, we can think of the Fourier series coefficients as samples from the scaled and windowed Fourier transform of \tilde{h} . Substituting (A.4) into (A.1) and using convolution and Dirac comb shorthand notation yields:

$$\begin{aligned} h[n, p] &= e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} h_w((n+kN)\nu_0, \rho + \rho_0) e^{-2\pi i(n+kN)\nu_0\rho} d\rho \\ &= e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} h_w(f, \rho + \rho_0) e^{-2\pi i f \rho} \delta((n+kN)\nu_0 - f) df d\rho \\ &= e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \sum_{k=-\infty}^{\infty} [(h_w(f, \rho + \rho_0) e^{-2\pi i f \rho}) * \delta_{-kN\nu_0}(f)](n\nu_0) d\rho \\ &= e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} [(h_w(f, \rho + \rho_0) e^{-2\pi i f \rho}) * \text{III}_{N\nu_0}(f)](n\nu_0) d\rho \end{aligned} \quad (\text{A.6})$$

where $*$ is the convolution operator and $\text{III}_\alpha(f) \equiv \sum_{k=-\infty}^{\infty} \delta(f - k\alpha)$ is the Dirac comb with spacing α . Using the rect function notation from (A.3), the function in brackets in equation (A.6) can be further simplified by taking the inverse Fourier transform, combining terms, and taking the Fourier transform to arrive at an equivalent representation:

$$\begin{aligned} &[(h_w(f, \rho + \rho_0) e^{-2\pi i f \rho}) * \text{III}_{N\nu_0}(f)](s) \\ &= \mathcal{F}_u \mathcal{F}_s^{-1} [(h_w(f, \rho + \rho_0) e^{-2\pi i f \rho}) * \text{III}_{N\nu_0}(f)](s) \\ &= \mathcal{F}_u \left[\mathcal{F}_f^{-1} (h_w(f, \rho + \rho_0) e^{-2\pi i f \rho}) \cdot \left(\frac{1}{N\nu_0} \text{III}_{1/(N\nu_0)}(u) \right) \right] \\ &= \mathcal{F}_u \left[\mathcal{F}_f^{-1} (h_w(f, \rho + \rho_0))(u - \rho) \cdot \left(\frac{1}{N\nu_0} \text{III}_{\tau_b}(u) \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \mathcal{F}_u \left[\tilde{h}(u - \rho, \rho + \rho_0) \Pi_{1/\nu_0} (u - \rho - 1/(2\nu_0)) \nu_0 \frac{1}{N\nu_0} \text{III}_{\tau_b}(u) \right] \\
&= \mathcal{F}_u \left[\tilde{h}(u - \rho, \rho + \rho_0) \frac{1}{N} \Pi_{N\tau_b} (u - \rho - N\tau_b/2) \text{III}_{\tau_b}(u) \right] \\
&= \mathcal{F}_u \left[(\tilde{h}(u - \rho, \rho + \rho_0)) \cdot \left(\frac{1}{N} \sum_{k=\lceil \rho/\tau_b \rceil}^{\lceil \rho/\tau_b \rceil + N - 1} \delta(u - k\tau_b) \right) \right] \\
&= \left[(h(f, \rho + \rho_0) e^{-2\pi i f \rho}) * \left(\frac{1}{N} \sum_{k=\lceil \rho/\tau_b \rceil}^{\lceil \rho/\tau_b \rceil + N - 1} e^{-2\pi i f k \tau_b} \right) \right](s), \tag{A.7}
\end{aligned}$$

with $\lceil \rho/\tau_b \rceil = \text{ceil}(\rho/\tau_b)$ and $h(f, \rho) = \mathcal{F}_t \tilde{h}(t, \rho)$. Substituting this result back into equation (A.6) yields the simplification

$$\begin{aligned}
h[n, p] &= e^{2\pi i n(p+1)/N} \int_{p\tau_b}^{(p+1)\tau_b} \left[h(f, \rho + \rho_0) e^{-2\pi i f \rho} * \frac{1}{N} \sum_{k=\lceil \rho/\tau_b \rceil}^{\lceil \rho/\tau_b \rceil + N - 1} e^{-2\pi i f k \tau_b} \right] (n\nu_0) d\rho \\
&= e^{2\pi i n(p+1)\nu_0\tau_b} \int_{p\tau_b}^{(p+1)\tau_b} \left[h(f, \rho + \rho_0) e^{-2\pi i f \rho} * \frac{1}{N} \sum_{k=p+1}^{p+N} e^{-2\pi i f k \tau_b} \right] (n\nu_0) d\rho,
\end{aligned}$$

where we use the fact that $\lceil \rho/\tau_b \rceil = p + 1$ on the integration interval $[p\tau_b, (p + 1)\tau_b]$. The next step is to combine exponential terms,

$$h[n, p] = \int_{p\tau_b}^{(p+1)\tau_b} \left[h(f, \rho + \rho_0) e^{-2\pi i f \rho} * \frac{1}{N} e^{-2\pi i (f - n\nu_0)(p+1)\tau_b} \sum_{k=0}^{N-1} e^{-2\pi i f k \tau_b} \right] (n\nu_0) d\rho,$$

and then recognize that since the combined exponential term is evaluated at $(f - n\nu_0)$, it can be moved to the other side of the convolution operator and combined with the exponential on that side:

$$h[n, p] = \int_{p\tau_b}^{(p+1)\tau_b} \left[h(f, \rho + \rho_0) e^{-2\pi i f(\rho - (p+1)\tau_b)} * \frac{1}{N} \sum_{k=0}^{N-1} e^{-2\pi i f k \tau_b} \right] (n\nu_0) d\rho. \tag{A.8}$$

This equation is much simpler than the original, but there is one more simplification

that can be performed. The final step makes use of the periodic sinc, sometimes known as the wrapped sinc or Dirichlet function, defined as:

$$\text{psinc}_N(\omega) = \frac{1}{N} \frac{\sin(\omega N/2)}{\sin(\omega/2)}. \quad (\text{A.9})$$

From the finite geometric sum formula and the complex exponential representation of sine, the sum in (A.8) can be written as a frequency-blurring kernel $b_{N,\tau_b}(f)$:

$$\begin{aligned} b_{N,\tau_b}(f) &\equiv \frac{1}{N} \sum_{k=0}^{N-1} e^{-2\pi i f k \tau_b} \\ &= \frac{1}{N} \frac{1 - e^{-2\pi i N \tau_b f}}{1 - e^{-2\pi i \tau_b f}} \\ &= \frac{1}{N} e^{-\pi i \tau_b f (N-1)} \frac{\sin(\pi N \tau_b f)}{\sin(\pi \tau_b f)} \end{aligned} \quad (\text{A.10})$$

$$= e^{-\pi i \tau_b f (N-1)} \text{psinc}_N(2\pi \tau_b f). \quad (\text{A.11})$$

This allows us to finally write the discrete reflectivity coefficients $h[n, p]$ as:

$$h[n, p] = \int_{p\tau_b}^{(p+1)\tau_b} [h(f, \rho + \rho_0) e^{-2\pi i f(\rho - (p+1)\tau_b)} * b_{N,\tau_b}(f)](n\nu_0) d\rho. \quad (\text{A.12})$$

Further discussion of this representation can be found in Section 3.3 of Chapter 3.

Bibliography

- AMISR Specifications*, (2014), Online, <http://amisr.com/amisr/about/amisr-overview/amisr-specifications/>.
- Bajwa, W. U., A. M. Sayeed, and R. Nowak, (2008), Learning sparse doubly-selective channels. In *2008 46th annual allerton conference on communication, control, and computing*, 575–582, IEEE, September, doi:10.1109/ALLERTON.2008.4797610.
- Bajwa, W. U., J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, (2007), Toeplitz-Structured Compressed Sensing Matrices. In *IEEE/SP 14th workshop on statistical signal processing, 2007. SSP '07*, 294–298, IEEE, August, doi:10.1109/SSP.2007.4301266.
- Baraniuk, R., and P. Steeghs, (2007), Compressive Radar Imaging. In *2007 IEEE radar conference*, 128–133, IEEE, April, doi:10.1109/RADAR.2007.374203.
- Beck, A., and M. Teboulle, (2009), A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, *SIAM Journal on Imaging Sciences*, 2 (1): 183–202, doi:10.1137/080716542.
- . (2010), Gradient-Based Algorithms with Applications to Signal Recovery Problems. In *Convex optimization in signal processing and communications*, edited by D. P. Palomar and Y. C. Eldar, 33–88, Cambridge University Press, January.
- Becker, S. R., E. J. Candès, and M. C. Grant, (2011a), Templates for convex cone problems with applications to sparse signal recovery, *Mathematical Programming Computation*, 3 (3): 165–218, doi:10.1007/mpc.v3i3.58.

- Becker, S. R., E. J. Candès, and M. C. Grant, (2011b), *TFOCS: Templates for First-Order Conic Solvers*, version 1.0c, available at <http://cvxr.com/tfocs/>.
- Behnel, S., R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, (2011), Cython: The Best of Both Worlds, *Computing in Science Engineering*, 13 (2): 31–39, doi:10.1109/MCSE.2010.118.
- Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. Reinisch, (2014), The International Reference Ionosphere 2012 – a model of international collaboration, *Journal of Space Weather and Space Climate*, 4:A07, doi:10.1051/swsc/2014004.
- Boyd, S. P., (2004), *Convex optimization*, Cambridge, UK ; New York: Cambridge University Press.
- . (2008), *Lecture Notes for EE263*, Stanford University.
- Boyd, S. P., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, (2011), Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Found. Trends Mach. Learn.* 3 (1): 1–122, doi:10.1561/2200000016.
- Burmaka, V. P., V. I. Taran, and L. F. Chernogor, (2006), Wave-like processes in the ionosphere under quiet and disturbed conditions. 1. Kharkov incoherent scatter radar observations, *Geomagnetism and Aeronomy*, 46 (2): 183–198, doi:10.1134/S0016793206020071.
- Candès, E. J., (2008), The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathematique*, 346 (9-10): 589–592, doi:10.1016/j.crma.2008.03.014.
- Candès, E. J., Y. C. Eldar, D. Needell, and P. Randall, (2011), Compressed sensing with coherent and redundant dictionaries, *Applied and Computational Harmonic Analysis*, 31 (1): 59–73, doi:10.1016/j.acha.2010.10.002.

- Candès, E. J., and Y. Plan, (2011), A Probabilistic and RIPless Theory of Compressed Sensing, *IEEE Transactions on Information Theory*, 57 (11): 7235–7254, doi:10.1109/TIT.2011.2161794.
- Candès, E. J., and J. K. Romberg, (2007), Sparsity and incoherence in compressive sampling, *Inverse Problems*, 23 (3): 969–985, doi:10.1088/0266-5611/23/3/008.
- Candès, E. J., J. K. Romberg, and T. Tao, (2006), Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, 59 (8): 1207–1223, doi:10.1002/cpa.20124.
- Candès, E. J., and T. Tao, (2006), Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?, *IEEE Transactions on Information Theory*, 52 (12): 5406–5425, doi:10.1109/TIT.2006.885507.
- . (2007), The Dantzig selector: Statistical estimation when p is much larger than n , *The Annals of Statistics*, 35 (6): 2313–2351, doi:10.1214/009053606000001523.
- Candès, E. J., and M. Wakin, (2008), An Introduction To Compressive Sampling, *IEEE Signal Processing Magazine*, 25 (2): 21–30, doi:10.1109/MSP.2007.914731.
- Carpenter, D. L., and J. Lemaire, (2004), The Plasmasphere Boundary Layer, *Ann. Geophys.* 22 (12): 4291–4298, doi:10.5194/angeo-22-4291-2004.
- CEDAR Steering Committee, (2011), *CEDAR: The New Dimension : Strategic Vision for the National Science Foundation Program on Coupling, Energetics and Dynamics of Atmospheric Regions*, National Science Foundation.
- Ceplecha, Z., J. Borovička, W. G. Elford, D. O. ReVelle, R. L. Hawkes, V. Porubčan, and M. Šimek, (1998), Meteor Phenomena and Bodies, *Space Science Reviews*, 84 (3-4): 327–471, doi:10.1023/A:1005069928850.
- Cervera, M. A., W. G. Elford, and D. I. Steel, (1997), A new method for the measurement of meteor speeds: The pre-to phase technique, *Radio Science*, 32 (2): 805–816, doi:10.1029/96RS03638.

- Chapin, E., and E. Kudeki, (1994), Plasma wave excitation on meteor trails in the equatorial electrojet, *Geophysical Research Letters*, 21 (22): 2433–2436, doi:10.1029/94GL01705.
- Chau, J. L., F. R. Galindo, C. J. Heinselman, and M. J. Nicolls, (2009), Meteor-head echo observations using an antenna compression approach with the 450 MHz Poker Flat Incoherent Scatter Radar, *Journal of Atmospheric and Solar-Terrestrial Physics*, Advances in high latitude upper atmospheric science with the Poker Flat Incoherent Scatter Radar (PFISR), 71 (6–7): 636–643, doi:10.1016/j.jastp.2008.08.007.
- Close, S., P. Brown, M. Campbell-Brown, M. M. Oppenheim, and P. Colestock, (2007), Meteor head echo radar data: Mass–velocity selection effects, *Icarus*, 186 (2): 547–556, doi:10.1016/j.icarus.2006.09.007.
- Close, S., P. Colestock, L. Cox, M. Kelley, and N. Lee, (2010), Electromagnetic pulses generated by meteoroid impacts on spacecraft, *Journal of Geophysical Research*, 115:A12328, doi:10.1029/2010JA015921.
- Close, S., S. M. Hunt, F. M. McKeen, and M. J. Minardi, (2002), Characterization of Leonid meteor head echo data collected using the VHF-UHF Advanced Research Projects Agency Long-Range Tracking and Instrumentation Radar (ALTAIR), *Radio Science*, 37 (1): 9–1, doi:10.1029/2000RS002602.
- Close, S., M. M. Oppenheim, D. Durand, and L. P. Dyrud, (2005), A new method for determining meteoroid mass from head echo data, *Journal of Geophysical Research: Space Physics*, 110 (A9): A09308, doi:10.1029/2004JA010950.
- Close, S., M. M. Oppenheim, S. Hunt, and L. Dyrud, (2002), Scattering characteristics of high-resolution meteor head echoes detected at multiple frequencies, *Journal of Geophysical Research: Space Physics*, 107 (A10): 1295, doi:10.1029/2002JA009253.
- Close, S., R. Volz, R. Loveland, A. Macdonell, P. Colestock, I. Linscott, and M. M. Oppenheim, (2012), Determining meteoroid bulk densities using a plasma scattering model with high-power large-aperture radar data, *Icarus*, 221 (1): 300–309, doi:10.1016/j.icarus.2012.07.033.

- Cohen, M., M. Fox, and J. Baden, (1990), Minimum peak sidelobe pulse compression codes. In *Radar conference, 1990., record of the IEEE 1990 international*, 633–638, May, doi:10.1109/RADAR.1990.201102.
- Coxson, G., and J. Russo, (2005), Efficient exhaustive search for optimal-peak-sidelobe binary codes, *IEEE Transactions on Aerospace and Electronic Systems*, 41 (1): 302–308, doi:10.1109/TAES.2005.1413763.
- Coxson, G., A. Hirschel, and M. Cohen, (2001), New results on minimum-PSL binary codes. In *Proceedings of the 2001 IEEE radar conference, 2001*, 153–156, doi:10.1109/NRC.2001.922969.
- Damtie, B., M. S. Lehtinen, and T. Nygrén, (2004), Decoding of Barker-coded incoherent scatter measurements by means of mathematical inversion, *Ann. Geophys.* 22 (1): 3–13, doi:10.5194/angeo-22-3-2004.
- Damtie, B., M. S. Lehtinen, M. Orispää, and J. Vierinen, (2008), Mismatched Filtering of Aperiodic Quadriphase Codes, *IEEE Transactions on Information Theory*, 54 (4): 1742–1749, doi:10.1109/TIT.2008.917655.
- Donoho, D. L., (2006a), Compressed sensing, *IEEE Transactions on Information Theory*, 52 (4): 1289–1306, doi:10.1109/TIT.2006.871582.
- Donoho, D. L., (2006b), For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution, *Communications on Pure and Applied Mathematics*, 59 (6): 797–829, doi:10.1002/cpa.20132.
- Donoho, D. L., I. Johnstone, and A. Montanari, (2013), Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising, *IEEE Transactions on Information Theory*, 59 (6): 3396–3433, doi:10.1109/TIT.2013.2239356.
- Donoho, D. L., A. Maleki, and A. Montanari, (2011), The Noise-Sensitivity Phase Transition in Compressed Sensing, *IEEE Transactions on Information Theory*, 57 (10): 6920–6941, doi:10.1109/TIT.2011.2165823.

- Donoho, D. L., A. Maleki, and A. Montanari, (2009), Message-passing algorithms for compressed sensing, *Proceedings of the National Academy of Sciences*, 106 (45): 18914–18919, doi:10.1073/pnas.0909892106.
- Donoho, D. L., and J. Tanner, (2009), Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367 (1906): 4273–4293, doi:10.1098/rsta.2009.0152.
- . (2010), Precise Undersampling Theorems, *Proceedings of the IEEE*, 98 (6): 913–924, doi:10.1109/JPROC.2010.2045630.
- Duarte, M., M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, (2008), Single-Pixel Imaging via Compressive Sampling, *IEEE Signal Processing Magazine*, 25 (2): 83–91, doi:10.1109/MSP.2007.914730.
- Dyrud, L. P., K. Denney, S. Close, M. Oppenheim, J. Chau, and L. Ray, (2004), Meteor velocity determination with plasma physics, *Atmospheric Chemistry and Physics*, 4 (3): 817–824, doi:10.5194/acp-4-817-2004.
- Dyrud, L. P., M. M. Oppenheim, S. Close, and S. Hunt, (2002), Interpretation of non-specular radar meteor trails, *Geophysical Research Letters*, 29 (21): 2012, doi:10.1029/2002GL015953.
- Dyrud, L. P., D. Wilson, S. Boerve, J. Trulsen, H. Pecseli, S. Close, C. Chen, and Y. Lee, (2008), Plasma and Electromagnetic Simulations of Meteor Head Echo Radar Reflections, *Earth, Moon, and Planets*, 102 (1-4): 383–394, doi:10.1007/s11038-007-9189-8.
- EISCAT Technical Specifications*, (2014), Online, <https://www.eiscat.se/about/specifications>.
- Ender, J. H., (2010), On compressive sensing applied to radar, *Signal Processing*, 90 (5): 1402–1414, doi:10.1016/j.sigpro.2009.11.009.

- Erickson, P. J., F. D. Lind, S. M. Wendelken, and M. A. Faubert, (2001), Meteor head echo observations using the Millstone Hill UHF incoherent scatter radar system, 495:457–463, November.
- Erlich, Y., A. Gordon, M. Brand, G. Hannon, and P. Mitra, (2010), Compressed Genotyping, *IEEE Transactions on Information Theory*, 56 (2): 706–723, doi:10.1109/TIT.2009.2037043.
- Esser, E., X. Zhang, and T. F. Chan, (2010), A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science, *SIAM Journal on Imaging Sciences*, 3 (4): 1015–1046, doi:10.1137/09076934X.
- Evans, J. V., (1965), Radio-echo studies of meteors at 68-centimeter wavelength, *Journal of Geophysical Research*, 70 (21): 5395–5416, doi:10.1029/JZ070i021p05395.
- Fukao, S., T. Sato, T. Tsuda, S. Kato, K. Wakasugi, and T. Makihira, (1985), The MU radar with an active phased array system: 1. Antenna and power amplifiers, *Radio Science*, 20 (6): 1155–1168, doi:10.1029/RS020i006p01155.
- Golan, D., Y. Erlich, and S. Rosset, (2012), Weighted pooling—practical and cost-effective techniques for pooled high-throughput sequencing, *Bioinformatics*, 28 (12): i197–i206, doi:10.1093/bioinformatics/bts208.
- Goldstein, T., E. Esser, and R. Baraniuk, (2013), Adaptive Primal-Dual Hybrid Gradient Methods for Saddle-Point Problems, pre-print, May, arXiv: 1305.0546 [math.NA].
- Gomersall, H., (2014), *pyFFTW*, version 0.10.0, available at <http://hgomersall.github.io/pyFFTW/>.
- Goncharenko, L. P., V. W. Hsu, C. G. M. Brum, S.-R. Zhang, and J. T. Fenzke, (2013), Wave signatures in the midlatitude ionosphere during a sudden stratospheric warming of January 2010, *Journal of Geophysical Research: Space Physics*, 118 (1): 472–487, doi:10.1029/2012JA018251.

- Gordon, W., (1958), Incoherent Scattering of Radio Waves by Free Electrons with Applications to Space Exploration by Radar, *Proceedings of the IRE*, 46 (11): 1824–1829, doi:10.1109/JRPROC.1958.286852.
- Hagen, J., (2005), *Arecibo 430 MHz Radar System: Operation and Maintenance Manual*, Manual, Arecibo Observatory, May.
- Harding, B. J., and M. Milla, (2013), Radar imaging with compressed sensing, *Radio Science*, 48 (5): 528–588, doi:10.1002/rds.20063.
- Herman, M. A., and T. Strohmer, (2009), High-Resolution Radar via Compressed Sensing, *IEEE Transactions on Signal Processing*, 57 (6): 2275–2284, doi:10.1109/TSP.2009.2014277.
- Hocking, W. K., B. Fuller, and B. Vandepeer, (2001), Real-time determination of meteor-related parameters utilizing modern digital technology, *Journal of Atmospheric and Solar-Terrestrial Physics*, Radar applications for atmosphere and ionosphere research - PIERS 1999, 63 (2–3): 155–169, doi:10.1016/S1364-6826(00)00138-3.
- Hunter, J., (2007), Matplotlib: A 2D Graphics Environment, *Computing in Science Engineering*, 9 (3): 90–95, doi:10.1109/MCSE.2007.55.
- Jol, H. M., (2009), *Ground penetrating radar theory and applications*, Amsterdam, Netherlands; Boston: Elsevier Science.
- Jones, W., (1995), Theory of the initial radius of meteor trains, *Monthly Notices of the Royal Astronomical Society*, 275:812–818.
- Kelley, M. C., (2009), *The Earth's Ionosphere: Plasma Physics & Electrodynamics*, 2nd ed., Academic Press, June.
- Kelly, J. D., C. J. Heinselman, J. F. Vickrey, and R. R. Vondrak, (1995), The sondrestrom radar and accompanying ground-based instrumentation, *Space Science Reviews*, 71 (1-4): 797–813, doi:10.1007/BF00751351.

- Kivelson, M. G., and C. T. Russell, eds., (1995), *Introduction to space physics*, Cambridge ; New York: Cambridge University Press.
- Lee, N., S. Close, A. Goel, D. Lauben, I. Linscott, T. Johnson, D. Strauss, S. Bugiel, A. Mocker, and R. Srama, (2013), Theory and experiments characterizing hypervelocity impact plasmas on biased spacecraft materials, *Physics of Plasmas (1994-present)*, 20 (3): 032901, doi:10.1063/1.4794331.
- Lehtinen, M. S., I. I. Virtanen, and J. Vierinen, (2008), Fast comparison of IS radar code sequences for lag profile inversion, *Ann. Geophys.* 26 (8): 2291–2301, doi:10.5194/angeo-26-2291-2008.
- Lehtinen, M. S., B. Damtie, P. Piironen, and M. Orispää, (2009), Perfect and almost perfect pulse compression codes for range spread radar targets, *Inverse Problems and Imaging*, 3 (3): 465–486, doi:10.3934/ipi.2009.3.465.
- Lehtinen, M. S., and I. Häggström, (1987), A new modulation principle for incoherent scatter measurements, *Radio Science*, 22 (4): 625–634, doi:10.1029/RS022i004p00625.
- Lewicki, M. S., (2002), Efficient coding of natural sounds, *Nature Neuroscience*, 5 (4): 356–363, doi:10.1038/nn831.
- Lind, F. D., P. J. Erickson, A. J. Coster, J. C. Foster, J. R. Marchese, Z. Berkowitz, and J. D. Sahr, (2013), Intercepted signals for ionospheric science, *Radio Science*, 48 (3): 248–264, doi:10.1002/rds.20034.
- Lindner, J., (1975), Binary sequences up to length 40 with best possible autocorrelation function, *Electronics Letters*, 11 (21): 507–507, doi:10.1049/el:19750391.
- Lustig, M., D. Donoho, J. Santos, and J. Pauly, (2008), Compressed Sensing MRI, *IEEE Signal Processing Magazine*, 25 (2): 72–82, doi:10.1109/MSP.2007.914728.
- Mallat, S., (2008), *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, December.

- Markkanen, M., J. Vierinen, and J. Markkanen, (2008), Polyphase alternating codes, *Ann. Geophys.* 26 (8): 2237–2243, doi:10.5194/angeo-26-2237-2008.
- Mathews, J. D., S. J. Briczinski, A. Malhotra, and J. Cross, (2010), Extensive meteoroid fragmentation in V/UHF radar meteor observations at Arecibo Observatory, *Geophysical Research Letters*, 37 (4): L04103, doi:10.1029/2009GL041967.
- Mathews, J. D., S. J. Briczinski, D. D. Meisel, and C. J. Heinselman, (2008), Radio and Meteor Science Outcomes From Comparisons of Meteor Radar Observations at AMISR Poker Flat, Sondrestrom, and Arecibo, *Earth, Moon, and Planets*, 102 (1-4): 365–372, doi:10.1007/s11038-007-9168-0.
- Mathews, J. D., J. Doherty, C. .-.H. Wen, S. J. Briczinski, D. Janches, and D. D. Meisel, (2003), An update on UHF radar meteor observations and associated signal processing techniques at Arecibo Observatory, *Journal of Atmospheric and Solar-Terrestrial Physics*, 65 (10): 1139–1149, doi:10.1016/j.jastp.2003.07.009.
- Melvin, W. L., and J. A. Scheer, eds., (2012), *Principles of Modern Radar: Advanced Techniques*, vol. 2, Raleigh, NC: SciTech Publishing.
- , eds., (2014), *Principles of Modern Radar: Radar Applications*, vol. 3, Raleigh, NC: SciTech Publishing.
- MIT Haystack Observatory: ISR*, (2014), Online, <http://www.haystack.mit.edu/atm/mho/instruments/isr/>.
- Monajemi, H., S. Jafarpour, M. Gavish, D. L. Donoho, S. Ambikasaran, S. Bacallado, D. Bharadia, et al., (2013), Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices, *Proceedings of the National Academy of Sciences*, 110 (4): 1181–1186, doi:10.1073/pnas.1219540110.
- Natarajan, B. K., (1995), Sparse Approximate Solutions to Linear Systems, *SIAM Journal on Computing*, 24 (2): 227–234, doi:10.1137/S0097539792240406.

- Nesterov, Y., (1983), A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In *Soviet mathematics doklady*, 27:372–376.
- . (1988), On an approach to the construction of optimal methods of minimization of smooth convex functions, *Ekonomika i Mateaticheskie Metody*, 24:509–517.
- . (2007), *Gradient methods for minimizing composite objective function*, CORE Discussion Paper 2007076, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), September.
- Nikoukar, R., F. Kamalabadi, E. Kudeki, and M. Sulzer, (2008), An efficient near-optimal approach to incoherent scatter radar parameter estimation, *Radio Science*, 43:RS5007, doi:10.1029/2007RS003724.
- Ochs, G., (1963), The large 50 mc/s dipole array at Jicamarca radio observatory. In *Antennas and propagation society international symposium, 1963*, 1:237–241, July, doi:10.1109/APS.1963.1148638.
- O'Donoghue, B., and E. J. Candès, (2013), Adaptive Restart for Accelerated Gradient Schemes, *Foundations of Computational Mathematics*:1–18, doi:10.1007/s10208-013-9150-3.
- Oliphant, T. E., (2007), Python for Scientific Computing, *Computing in Science Engineering*, 9 (3): 10–20, doi:10.1109/MCSE.2007.58.
- Oliphant, T. E., and The Numba Contributors, (2014), *Numba*, version 0.11.1, available at <http://numba.pydata.org/>.
- Olshausen, B. A., and D. J. Field, (1997), Sparse coding with an overcomplete basis set: A strategy employed by V1?, *Vision Research*, 37 (23): 3311–3325, doi:10.1016/S0042-6989(97)00169-7.
- Olshausen, B. A., and D. J. Field, (2004), Sparse coding of sensory inputs, *Current Opinion in Neurobiology*, 14 (4): 481–487, doi:10.1016/j.conb.2004.07.007.

- Oppenheim, M. M., G. Sugar, N. O. Slowey, E. Bass, J. L. Chau, and S. Close, (2009), Remote sensing lower thermosphere wind profiles using non-specular meteor echoes, *Geophysical Research Letters*, 36 (9): L09817, doi:10.1029/2009GL037353.
- Parikh, N., and S. P. Boyd, (2013), Proximal Algorithms, *Foundations and Trends in Optimization*, 1 (3): 123–231.
- Pérez, F., and B. Granger, (2007), IPython: A System for Interactive Scientific Computing, *Computing in Science Engineering*, 9 (3): 21–29, doi:10.1109/MCSE.2007.53.
- Pfander, G. E., and H. Rauhut, (2010), Sparsity in Time-Frequency Representations, *Journal of Fourier Analysis and Applications*, 16 (2): 233–260, doi:10.1007/s00041-009-9086-9.
- Potekhin, A. P., A. V. Medvedev, A. V. Zavorin, D. S. Kushnarev, V. P. Lebedev, and B. G. Shpynev, (2008), Development of diagnostic capabilities of the Irkutsk incoherent scattering radar, *Cosmic Research*, 46 (4): 347–353, doi:10.1134/S0010952508040102.
- Potter, L. C., E. Ertin, J. T. Parker, and M. Cetin, (2010), Sparsity and Compressed Sensing in Radar Imaging, *Proceedings of the IEEE*, 98 (6): 1006–1020, doi:10.1109/JPROC.2009.2037526.
- Python Software Foundation, (2014), *Python Language Reference*, version 2.7, available at <http://www.python.org>.
- Richards, M. A., (2005), *Fundamentals of Radar Signal Processing*, 1st ed., McGraw Hill Professional, June.
- Romberg, J. K., (2009), Compressive Sensing by Random Convolution, *SIAM Journal on Imaging Sciences*, 2 (4): 1098–1128, doi:10.1137/08072975X.
- Rubin, A. E., and J. N. Grossman, (2010), Meteorite and meteoroid: New comprehensive definitions, *Meteoritics & Planetary Science*, 45 (1): 114–122, doi:10.1111/j.1945-5100.2009.01009.x.

- Scheinberg, K., D. Goldfarb, and X. Bai, (2011), *Fast First-Order Methods for Composite Convex Optimization with Backtracking*, pre-print, April.
- Skolnik, M. I., (2001), *Introduction to Radar Systems*, 3rd ed., McGraw Hill.
- . (2008), *Radar Handbook*, 3rd ed., McGraw-Hill Education.
- Stigler, S. M., (1981), Gauss and the Invention of Least Squares, *The Annals of Statistics*, 9 (3): 465–474, doi:10.1214/aos/1176345451.
- Strohmer, T., (2012), Measure What Should be Measured: Progress and Challenges in Compressive Sensing, *IEEE Signal Processing Letters*, 19 (12): 887–893, doi:10.1109/LSP.2012.2224518.
- Sugar, G. R., (1964), Radio propagation by reflection from meteor trails, *Proceedings of the IEEE*, 52 (2): 116–136, doi:10.1109/PROC.1964.2801.
- Sulzer, M. P., (1993), A new type of alternating code for incoherent scatter measurements, *Radio Science*, 28 (6): 995–1001, doi:10.1029/93RS01918.
- Tascione, T. F., (2010), *Introduction to the Space Environment*, 2nd ed., Malabar, Fla.: Krieger Publishing Company.
- The Jicamarca Radio Observatory*, (2014), Online, <http://jicamarca.ece.cornell.edu/overview.html>.
- Tibshirani, R., (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1): 267–288.
- Tropp, J. A., and A. C. Gilbert, (2007), Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, *IEEE Transactions on Information Theory*, 53 (12): 4655–4666, doi:10.1109/TIT.2007.909108.

- Tropp, J. A., M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk, (2006), Random Filters for Compressive Sampling and Reconstruction. In *2006 IEEE international conference on acoustics, speech and signal processing, 2006. ICASSP 2006 proceedings*, 3:III–III, IEEE, May, doi:10.1109/ICASSP.2006.1660793.
- Tsunoda, R. T., M. J. Baron, J. Owen, and D. M. Towle, (1979), Altair: an incoherent scatter radar for equatorial spread F studies, *Radio Science*, 14 (6): 1111–1119, doi:10.1029/RS014i006p01111.
- Van Trees, H. L., (2001), *Detection, Estimation, and Modulation Theory, Part III*, New York, USA: John Wiley & Sons, Inc., September.
- Vierinen, J., M. S. Lehtinen, and I. I. Virtanen, (2008), Amplitude domain analysis of strong range and Doppler spread radar echos, *Ann. Geophys.* 26 (8): 2419–2426, doi:10.5194/angeo-26-2419-2008.
- Virtanen, I. I., M. S. Lehtinen, T. Nygrén, M. Orispää, and J. Vierinen, (2008), Lag profile inversion method for EISCAT data analysis, *Ann. Geophys.* 26 (3): 571–581, doi:10.5194/angeo-26-571-2008.
- Virtanen, I. I., M. S. Lehtinen, and J. Vierinen, (2008), Towards multi-purpose IS radar experiments, *Ann. Geophys.* 26 (8): 2281–2289, doi:10.5194/angeo-26-2281-2008.
- Virtanen, I. I., J. Vierinen, and M. S. Lehtinen, (2009), Phase-coded pulse aperiodic transmitter coding, *Ann. Geophys.* 27 (7): 2799–2811, doi:10.5194/angeo-27-2799-2009.
- Virtanen, I. I., F. D. Lind, L. Roininen, P. J. Erickson, W. C. Rideout, M. Orispää, J. Vierinen, and M. S. Lehtinen, (2013), Polyphase-coded incoherent scatter measurements at Millstone Hill, *Radio Science*, 48:519–526, doi:10.1002/rds.20058.
- Walt, S. van der, S. Colbert, and G. Varoquaux, (2011), The NumPy Array: A Structure for Efficient Numerical Computation, *Computing in Science Engineering*, 13 (2): 22–30, doi:10.1109/MCSE.2011.37.

- Wannberg, G., A. Pellinen-Wannberg, and A. Westman, (1996), An ambiguity-function-based method for analysis of Doppler decompressed radar signals applied to EISCAT measurements of oblique UHF-VHF meteor echoes, *Radio Science*, 31 (3): 497–518, doi:10.1029/96RS00088.
- Woodman, R. F., and T. Hagfors, (1969), Methods for the measurement of vertical ionospheric motions near the magnetic equator by incoherent scattering, *Journal of Geophysical Research*, 74 (5): 1205–1212, doi:10.1029/JA074i005p01205.
- Woodward, P. M., (1980), *Probability and information theory, with applications to radar*, Dedham, Mass.: Artech House.
- Yoo, J., C. Turnes, E. Nakamura, C. Le, S. Becker, E. Sovero, M. Wakin, et al., (2012), A Compressed Sensing Parameter Extraction Platform for Radar Pulse Signal Acquisition, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2 (3): 626–638, doi:10.1109/JETCAS.2012.2214634.