

**COMPRESSED SENSING IN RADAR WITH STRUCTURED
INTERFERENCE**

A Thesis
Presented to
The Academic Faculty

by

Peter Benjamin Tuuk

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2017

Copyright © 2017 by Peter Benjamin Tuuk

COMPRESSED SENSING IN RADAR WITH STRUCTURED INTERFERENCE

Approved by:

James McClellan, Advisor
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Douglas Williams
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Mark Davenport
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Aaron Lanterman
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Yao Xie
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Date Approved: 1 June 2017

To my son, Caleb,

whose wide-eyed wonder as he meets the world

reveals our common existence for the miracle it is

ACKNOWLEDGEMENTS

I want to first thank my advisor, Professor McClellan. He has given me always-helpful advice on approaches to the various topics I have considered and provided pointers to prior work from a wide range of research areas. As I have sought to synthesize work from a number of fields, his broad perspective has helped me considerably. And I would like to thank him for his patience as I worked at my own pace.

I would like to thank my co-workers at GTRI who have helped me formulate the motivating problem, grounded my work in the realm of the applicable, and have had to endure inconveniences small and large when I have taken time to focus on my studies. In particular I would like to thank Greg Showman for his careful reading of my work and frequent advice, Bill Melvin who sought and provided data for testing my techniques and who has made certain that our lab maintains a commitment to scholarship, and George Brown whose frequent question – “*Have you defended yet?*” – more than once served as a helpful prod. And I owe gratitude to GTRI, the organization, for supporting my studies.

To my class- and lab-mates, thank you, especially to Chris Turnes, Aurèle Balavoine, and Kyle Krueger who provided the existence proof that it can be done.

And I must thank my family more than anyone else. To my parents, thank you for simultaneously providing unconditional love and instilling a sense of expectation. And to my wife, Alexandra, thank you; we have come some ways together and you have walked beside me the whole way.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
I INTRODUCTION	1
1.1 History and Motivation	1
1.2 Organization	3
II LITERATURE SURVEY	4
2.1 Radar Use Case and Design	4
2.2 Signal Model	7
2.3 Phenomenology of Correlated Interference	11
2.3.1 Clutter	11
2.3.2 RFI and Jamming	13
2.3.3 System Model Error	13
2.4 Signal Processing Treatment of Clutter	14
2.4.1 Early Techniques	14
2.4.2 Space-Time Adaptive Processing	15
2.4.3 Reduced Rank STAP	17
2.4.4 Robust Covariance Estimators	18
2.5 Compressed Sensing	19
2.5.1 Formulation and Assumptions	19
2.5.2 CS in Radar	21
2.5.3 Matrix Completion	23
2.6 Recent Work	25
2.7 Summary	27
III APPLYING INTERFERENCE COVARIANCE IN COMPRESSED SENSING ESTIMATION	28
3.1 Introduction	28

3.2	Method	29
3.2.1	ASPEN	29
3.2.2	Solution Methods	30
3.2.3	Proposed Extension	31
3.3	Results	32
3.3.1	Scoring Metrics	32
3.3.2	Signal Quality Definitions	33
3.3.3	Simulation Description	34
3.3.4	CS in Noise	34
3.3.5	CS and Covariance-Aware CS in Clutter	37
3.4	Conclusion	38
3.4.1	Review	38
3.4.2	Future Work	43
IV	ESTIMATING INTERFERENCE COVARIANCE FROM COMPRESSED MEASUREMENTS	45
4.1	Problem Formulation	45
4.1.1	Sensing Model	45
4.1.2	Interference	45
4.1.3	Sampling Compression	46
4.2	Prior Work	46
4.2.1	Space-Time Adaptive Processing	46
4.2.2	Diagonal Loading	48
4.2.3	Reduced Rank STAP	48
4.2.4	CS and Clutter	48
4.3	Proposed Technique	49
4.3.1	Concept	49
4.3.2	Singular Value Thresholding (SVT) Algorithm	49
4.4	Results	55
4.4.1	Evaluation Criteria	55
4.4.2	Synthetic Low-Rank Covariance Matrices	55
4.4.3	High-Fidelity Simulated Clutter	58

4.4.4	Flight Test Data	64
4.5	Conclusion	70
V	APPLYING RANDOM MATRIX THEORY IN ADAPTIVE RADAR	74
5.1	Problem Formulation	74
5.2	Random Matrix Theory	77
5.2.1	RMT Overview	77
5.2.2	Marčenko-Pastur	78
5.2.3	Tracy-Widom	78
5.2.4	Non-Null Stieltjes Transform Distribution	81
5.2.5	Characterizing Detectability	86
5.3	Application to STAP Parameter Selection	92
5.3.1	Problem Statement	92
5.3.2	Results	93
5.3.3	Summary	100
5.4	“Inverting” Sample Eigenvalues	100
5.4.1	Problem Statement	101
5.4.2	Applicability to Interference Estimation	103
5.5	Conclusion	104
VI	CONCLUSION	105
6.1	Contribution	105
6.2	Summary	106
6.3	Future Work	107
	APPENDIX A — IMPLEMENTATION OF SPECTRAL DISTRIBUTION CALCULATION	109
	REFERENCES	113
	VITA	120

LIST OF TABLES

1	Parameters used for synthetic structured interference experiments	56
2	Parameters for MCARM data collection experiments	66
3	The ensembles of classical random matrix theory. The commonly used Gaussian ensembles (Gaussian orthogonal ensemble, Gaussian unitary ensemble, and Gaussian symplectic ensemble) have been identified, showing their membership in this broader structure.	77

LIST OF FIGURES

1	Historical examples of airborne radar	2
2	A simplified block diagram of a single subarray of a large airborne phased array radar. The signal path includes a low-noise amplifier, analog filtering, phase shifting, RF combining, mixing to IF, and sampling. The digital samples from this and all other subarrays feed into the digital signal and data processor for pulse compression, beamforming, detection, association, tracking, prediction, and other functions.	6
3	The sampled distribution of the observed clutter in one realization, along with the underlying gamma distribution which is defined by a shape parameter of 10/3.	13
4	The adjoint estimate $\hat{\mathbf{x}}_{adj}$, reshaped into the range-angle-Doppler cube, and projected along each of the three dimensions. The true target location is $r = 290$ m, $\theta = -30^\circ$ and $v = 1$ m/s. The marker indicates the true target location in each view.	16
5	The STAP estimate $\hat{\mathbf{x}}_{stap}$, reshaped into the range-angle-Doppler cube, and projected along each of the three dimensions. The true target location is $r = 290$ m, $\theta = -30^\circ$ and $v = 1$ m/s. The marker indicates the true target location in each view.	16
6	Detection performance of compressed sensing recovery in white noise. The performance degrades with subsampling rate; each additional octave of undersampling results raises the noise floor by a factor of two, or 3 dB.	34
7	Detection performance as a function of SNR, varying quantization. Both CS and adjoint techniques are robust to eight bit quantization.	35
8	Detection performance as a function of SNR, varying P_{FA} . CS and adjoint techniques show similar changes in performance.	36
9	Detection performance as a function of SNR, varying number of point targets. By placing additional targets in the search volume the probability of detection decreases comparably in the compressed sensing and the matched filter techniques.	36
10	The CA CS method subsamples the data just as the standard CS method does, however it takes into account the covariance matrix that describes the interference structure. By doing so, it improves the probability of detection over the CS case as well as beyond the fully sampled, matched filter case that does not use the covariance information. These results are shown with a probability of false alarm of 0.005.	39

11	A comparison of various solution methods applied to a representative sample problem with input SNR of 0 dB, and input SCR of -20 dB. These plots show the relative amplitude of all the bins in the estimate produced by the identified technique. The red circle in each indicates the amplitude of the bin closest to the true target location. These results show that only the adaptive techniques, STAP and CA CS, correctly assign the highest estimated amplitude to the bin closest the target.	40
12	A related evaluation criterion, the detection quantile Q_d measures the average ranking of the true-target bin among all the bins in the estimate. A Q_d of zero is perfect. The fully-sampled STAP estimate performs better than all other techniques. Also, the $20\times$ undersampled CA CS estimate consistently achieves performance near that of the fully-sampled STAP and better than of the fully-sampled adjoint. The $40\times$ undersampled CA CS estimate is inferior to the fully-sampled adjoint, in contrast to the performance as measured by the P_d in Figure 10.	41
13	A comparison of reconstruction error (E_{norm}) performance frontiers shows that the sparsity-favoring techniques achieve better reconstruction error of this sparse signal.	42
14	A comparison of detection (P_D) performance frontiers shows that all these techniques achieve approximately equal detection performance.	42
15	Iteratively modified compressed sample covariance eigenvalues (shown in blue). As Algorithm 2 modifies the eigenvalues they eventually fall below the noise-only prediction line, producing the rank estimate. In this case the estimate is accurate, as the true interference rank is 14.	53
16	Performance of SVT rank estimation technique over a parametric evaluation space for different undersampling factors. For these cases, the interference rank structure is structure 1 (simple step function), the number of channels is 8, number of pulses is 128, and number of fast-time samples is 512.	54
17	The step size update function which uses a measure of consistency between the calculated residual and the previous residual. If the steps are in the same direction, the step size is increased to speed convergence. If the direction between steps changes significantly, the step size is reduced to improve stability.	54
18	Synthetic structured interference eigenvalue decay functions. These functions are shown with width parameter of 10. For other values, of this parameter, the spectral structure is stretched or compressed proportionally.	57
19	The spectral decay structures of the various covariance estimates of the synthetic structured interference. The SVT estimate more closely matches the truth than does the compressed SMI which uses the same data as input.	58

20	Example iteration results for the SVT algorithm operating on a synthetic structured interference. (a) Estimate of the interference rank increases as the SVT iteration progresses, which confirms that the resultant estimate of the covariance matrix improves. Improvement shown for three metrics: (b) direct matrix estimation error, (c) signal-to-interference ratio for a point target embedded in the interference (using the estimate as a filter), and (d) optimization residual.	59
21	Example target recovery results for the various covariance estimates on synthetic structured interference. After making a set of estimates of the covariance matrix, a number of targets are embedded in the interference, each of the resultant adaptive filters applied, and the signal-to-interference ratio recorded. The statistics of one such run are shown in a box and whisker plot. The box contains the 25%-75% quantiles of the results while the whiskers contain the 5%-95% quantiles.	60
22	Average probability of detection as a function of data under-sampling factor for cases of synthetic structured interference. As the under-sampling factor increases and less data is available for the two compressed estimates, the accuracy of those estimates degrades. Notably, the SVT estimate maintains much better performance than compressed SMI as the USF increases. . . .	60
23	Average probability of detection as a function of data under-sampling factor for cases with favorable interference structure. The SVT estimate based on the compressed data performs better than the diagonally loaded estimate based on the full data and nearly as well as the true covariance matrix. The interference rank is 5 and the structure is simple (structure ID = 1).	61
24	Average probability of detection as a function of CNR for cases of synthetic structured interference. At low clutter-to-noise ratios, noise is the dominant interference source. But as clutter becomes dominant, the structure of the interference can be estimated and used to develop a usable adaptive filter. Total signal-to-interference ratio is constant.	61
25	Average probability of detection as a function of number of range samples for cases of synthetic structured interference. As the number of range samples (snapshots) increases the two compressed estimates, which use compressed data (compressed snapshots), both improve in estimation accuracy.	62
26	Average probability of detection as a function of interference structure for cases of synthetic structured interference. The performance of various estimators is tested with four different clutter structures, having the four eigen-spectra illustrated in Figure 18. The performance of the estimators varies as a function of the spectra, with lower rank clutter being easier to filter out and higher rank clutter being more difficult.	62
27	Average probability of detection as a function of interference rank for cases of synthetic structured interference. As the clutter becomes higher rank it is more difficult to filter out.	63

28	The true eigenspectrum of the ASPEN-generated clutter. It exhibits concentration of spectral structure but is not strictly low-rank.	64
29	Example iteration results for the SVT algorithm operating on high-fidelity simulated clutter data. (a) Estimate of the interference rank increases as the SVT iteration progresses, which confirms that the resultant estimate of the covariance matrix improves. Improvement shown for three metrics: (b) direct matrix estimation error, (c) signal-to-interference ratio for a point target embedded in the interference (using the estimate as a filter), and (d) optimization residual.	65
30	The spectral decay structures of the various covariance estimates for high-fidelity simulated clutter. The SVT estimate more closely matches the truth than does the compressed SMI which uses the same data as input.	66
31	Angle-Doppler map results from each of the tested estimators using high-fidelity simulated clutter data with embedded target signal. The location of the true target is indicated with the square and signal-to-interference-plus-noise indicated in the subfigure titles.	67
32	Angle-Doppler map results from each of the tested estimators using artificially-rank-reduced simulated clutter data with embedded target signal. The location of the true target is indicated with the square and signal-to-interference-plus-noise indicated in the subfigure titles.	68
33	MCARM data record RD050575 averaged angle-Doppler map	69
34	MCARM data record RD050575 averaged angle-range map	69
35	MCARM data record RD050575 averaged range-Doppler map	70
36	Angle-Doppler maps for range bin 363 in MCARM data record RD050575. The undersampling factor is 2 for the compressed cases. The location of a presumed target is indicated in each plot with a square. This target location is used to provide the indicated signal-to-interference ratio.	71
37	The eigenspectra of the tested estimates of the interference covariance for MCARM data record RD050575. The undersampling factor is 2 for the compressed cases.	71
38	The performance of the tested estimators for each MCARM log file. Certain files feature more prominent targets in the observation space and targets vary in their location relative to the clutter ridge. The results shown use an undersampling factor of 2.	72
39	The performance of the tested estimators on the MCARM dataset as a function of undersampling factor. Results are averaged over all MCARM log files. In all undersampled cases the SVT estimate improves on the diagonally-loaded compressed SMI estimate.	72
40	A set of sample eigenvalues is shown for a noise-only distribution, with many spread substantially from the true distribution. The parameters are $m = 1024$, $n = 512$, $k = 0$, $\nu = 1$	75

41	A set of sample eigenvalues is shown for a noise and clutter distribution, with many spread substantially from the true distribution. Clutter rank $k = 77$, clutter structure ID is 2, degrees of freedom $m = 1024$, number of observations $n = 512$, noise variance $\nu = 1$	76
42	Projection of eigenvectors into the clutter subspace. Some smaller eigenvalues still correspond to eigenvectors with significant clutter component. As in Figure 41, clutter rank $k = 77$, clutter structure ID is 2, degrees of freedom $m = 1024$, number of observations $n = 512$, noise variance $\nu = 1$	76
44	The Tracy-Widom function gives the distribution of the largest eigenvalue.	80
45	The sample eigenvalues verify two different RMT-derived predictions. The Marčenko-Pastur prediction is shown in red, which closely matches the sample values. And the Tracy-Widom 95% confidence test for the maximum noise-only eigenvalue is shown as a dashed line, which closely matches the maximum sample value.	81
46	Plot of $x(m)$ which is used to calculate the support of the sample eigenvalue distribution. The red markers indicate the positive range over which $x(m)$ is non-increasing and represents the support of the sample eigenvalue distribution. Here $H(t) = \frac{7}{8}\delta(t - 1) + \frac{1}{24}\delta(t - 2.25) + \frac{1}{24}\delta(t - 6) + \frac{1}{24}\delta(t - 21)$ and has half as many samples as the dimension of the distribution.	83
47	The number of Newton iterations to converge to a solution of (63). This number is typically small, since the solution to the previous two iterations are used to linearly predict a starting guess. Here this example uses synthetic interference structure 4 and CNR = 13 dB and has half as many samples as the dimension of the distribution.	84
48	The Stieltjes transform-predicted sample eigenvalue distribution. This calculation predicts three groups of eigenvalues.	85
49	The Stieltjes transform-predicted sample eigenvalue distribution (magenta line) shown with a sample (blue circles). The calculation based on Stieltjes transform accurately predicts the eigenvalues of the sample.	85
50	The spectral density of noise, clutter, and the sum (noise+clutter). The sum is not a linear combination of the components.	86
51	The cumulative spectral density of noise, clutter, and the sum (noise+clutter). The sum is not a linear combination of the components.	87
52	Identifiability as a function of eigenvalue magnitude. Large eigenvalues can be identified as belonging to the Noise + Clutter distribution but smaller ones could have come from either the Noise-only or Noise+Clutter distributions. The black dashed line shows the Marčenko-Pastur maximum eigenvalue, so any eigenvalue above that magnitude is unlikely to be from the Noise-only distribution.	88

53	The fraction of dimensions identified as signal-bearing as a function of the decision threshold for both the Noise-only and Noise + Clutter distributions. For reasonable thresholds (> 0.5) fewer dimensions than the true number of clutter dimensions are detectable.	88
54	Identifiability as a function of sampling ratio. A higher sampling ratio indicates more samples per estimated dimension.	89
55	Identifiability as a function of clutter-to-noise ratio. Higher CNR separates the clutter eigenvalues from those of the noise and allows them to be detected an isolated more easily.	90
56	Performance of the adaptive filter as a function of the number of training samples. When the number of samples is close to the sample size the performance of SMI decreases. Rank reduction maintains good performance through this space. Parameters include interference structure ID of 4 and a CNR of 8 dB.	91
57	The predicted sample eigenvalue densities for several different training sample sizes. When the number of samples is close to the sample size the probability of small eigenvalues increases. These small eigenvalues inhibit stable matrix inversion. Parameters include interference structure ID of 4 and a CNR of 8 dB.	91
58	Covariance matrix error relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity. . . .	96
59	Precision matrix error relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large errors have been cropped for clarity.	97
60	Mean SINR loss relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity. . . .	98
61	Signal-to-interference ratio relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity.	99
62	The two components that form the distance metric used as the minimization objective. The horizontal and vertical distances are indicated on the CDF plots.	102
63	Distance between the observed sample eigenvalues and the predicted sample eigenvalues as a function of clutter rank and CNR.	102

64 The optimization objective as a function of iteration. The objective is minimized effectively when the assumed parameterization matches the true structure (red case). However, when the model does not match the true structure the algorithm does not converge well (blue case). The objective here is the area between the predicted CDF and the observed eigenvalue CDF. 103

SUMMARY

Ground clutter has challenged designers of airborne radar since it was first developed in the 1940's. Since that time, pulse-Doppler processing and space-time adaptive processing (STAP) have provided significant gains in clutter mitigation. This research examines the mitigation of ground clutter in compressed sensing (CS) radars, which use randomized sampling schemes. The work shows three results: (1) that clutter can be mitigated in these types of systems using the interference covariance matrix, (2) that the interference covariance can be estimated from compressed measurements using low-rank approximation, and (3) that results from random matrix theory can be used to improve processing for both CS radars and in the standard STAP context.

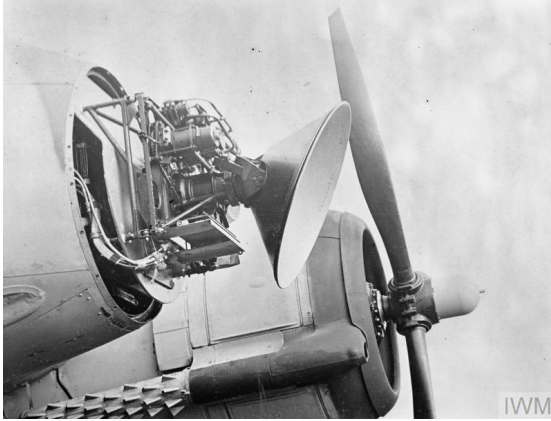
CHAPTER I

INTRODUCTION

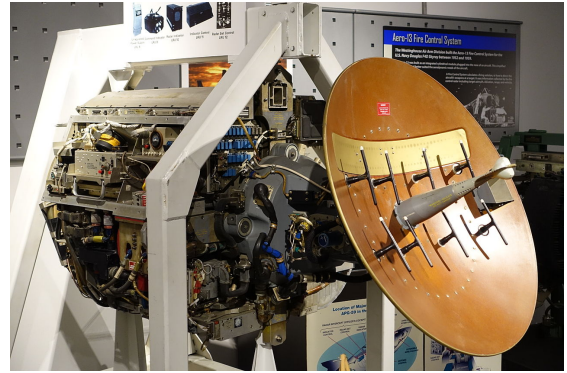
1.1 History and Motivation

The problem of clutter has posed challenges to radar designers and engineers since the early days of the sensing modality itself. Indeed, as early as the Second World War attention was paid to mitigating false detections from terrain [1]. One such early radar is shown in Figure 1a. Generally the problem of clutter was mitigated by constructing the observation geometry so that targets were above the radar and sensed against the background of sky. Early techniques included a simple notch filter at the transmitted frequency that removed returns with zero Doppler shift. This is effective in some cases for stationary radar systems but does not compensate for the effects of platform motion that frequency shifts the clutter returns. A subsequent development, pioneered in the 1950's, is the displaced phase center technique that phase shifts the returns from a series of pulses to align them, allowing for effective cancellation [2]. The next major development in airborne radar was a revolution. Pulse-Doppler radar, enabled by high accuracy of timing circuitry and early digital memory, coherently processed a set of pulses for clutter-rejection and other purposes. These techniques allowed effective airborne early warning (AWACS) development and look-down, shoot-down modes for fighter aircraft [3]. One such early pulse-Doppler radar, employed on the US F-4 Phantom is shown in Figure 1b.

In the late 1970s and 1980s, the concept of space-time adaptive processing (STAP) was introduced [4]. STAP takes advantage of improvements in digital signal processing to extend the clutter cancellation to the two-dimensional domain. It does so by introducing a spatial or array channel dimension. These additional degrees of freedom allow improved cancellation in the joint domain and extend work on clutter cancellation to that of other structured interference sources. In the years since, STAP theory and practice have improved with the introduction of more array channels, prior-knowledge-aided processing, and introduction of



(a) Early airborne radar: S-Band Mk. VIII radar on a Bristol Beaufighter (Imperial War Museum, Wikimedia Commons)



(b) Pulse Doppler Radar: X-Band AWG-10 radar from F-4 Phantom (National Electronics Museum, Wikimedia Commons)

Figure 1: Historical examples of airborne radar

computationally expensive matrix decompositions.

As STAP and other advanced radar processing become more mature technologies, there is a desire to employ it on smaller platforms. But cost, size, weight, power, and other considerations make large, high bandwidth array antennas infeasible in these settings. Compressed sensing offers the hope that lower sampling requirements and data volumes could simplify data acquisition requirements and allow advanced techniques on lower-end platforms. In some contexts the computational costs of compressed sensing reconstruction are prohibitive today. But in other contexts the signal acquisition problem is intractable under traditional Nyquist-rate sampling. As these techniques, approaches, and technologies mature, the need to consider additional sources of interference beyond noise becomes more pressing.

It is at this point that this work picks up the thread. This thesis examines the topic of compressed sensing in radar with a focus on mitigating structured interference, such as clutter, in the compressed sensing context. To do so, prior work from adaptive filtering, low-rank matrix approximation, and random matrix theory is exploited. Results of this thesis show that the covariance of the interference can be incorporated into the compressed sensing estimation process to improve performance. In addition, if the interference has low rank, statistics of the covariance can be reliably estimated from highly compressed measurements. To do so, results from random matrix theory on the expected structure of

sample covariance matrices are applied successfully to the adaptive radar filtering problem in both the compressed and uncompressed cases.

1.2 Organization

First, Chapter 2 introduces the topics under consideration including a detailed description of the clutter cancellation problem and the foundational results of compressed sensing. Chapter 3 shows how the clutter covariance matrix can be incorporated into a compressed sensing framework to improve estimation and detectability of targets. In Chapter 4 it is shown that results from low-rank matrix estimation can be effectively leveraged to estimate the statistics of the structured interference from compressed measurements which then enables the cancellation of interference and detection of targets. This work leads naturally into topics of Random Matrix Theory. Chapter 5 shows how results from that branch of mathematical analysis can be leveraged to improve the estimation of interference statistics in both the compressed and uncompressed cases. Finally, Chapter 6 provides concluding remarks to tie together previous work before offering several topics for further study.

CHAPTER II

LITERATURE SURVEY

2.1 Radar Use Case and Design

Radar systems are often tasked with characterizing objects within some field of regard. The radar performs this task by transmitting radio frequency electromagnetic energy into the surrounding medium. This energy propagates through the medium and impinges on objects in that environment. The objects reflect some portion of the energy back to the radar where it is processed to determine characteristics of the environment.

To more precisely describe this process, define the transmitted waveform as $w(t)$. This waveform propagates through the medium at the speed of light, c . Thus the range of the target is proportional to the time between the transmission of the waveform and the reception of the reflected echo. For a given range reflectivity profile $x(r)$ the received signal is the convolution of that profile with the transmitted waveform: $y(t) = w(t) * x(ct)$. Since we want to access the full set of tools from linear algebra, we express this and other processes as linear operators that operate on the discretized range reflectivity profile: $\mathbf{y} = \mathbf{S}\mathbf{x}$.

The most basic use of a radar system is to calculate range to a target by measuring the time between transmission of a pulse and the time the reflection from the target is received. Another fundamental measurement that may be made with a radar receiver is to calculate target velocity by measuring the Doppler shift of the reflected pulse. For multi-pulse radar the Doppler shift is calculated over multiple pulses to increase the observation time, improving the Doppler resolution and minimum detectable target velocity. Multi-channel digital receivers may estimate the angle of arrival of a reflected target signal using the differential time, or phase delay, between measurements at the sampled phase centers. These three sampling dimensions correspond to three dimensions in the target space:

1. Receiver channel: Elements of the antenna array are separated into some number of channels. The received energy collected by the elements of a channel is coherently

combined and sampled. This sampling dimension is used to determine the direction of arrival of signals from targets.

2. Slow Time: The radar transmits a series of pulses, samples the returns from each, and processes these samples coherently. This set of pulses constitutes a coherent processing interval (CPI) and the pulses are transmitted with some frequency, the pulse repetition frequency (PRF). This sampling dimension is used to determine the range rate of targets.
3. Fast Time: The analog-to-digital converter samples the incoming radio-frequency signal at a rate determined by the bandwidth of the transmitted waveform. This sampling dimension is used to determine the range of targets.

This three-dimensional conception of the received signal is known as the data cube [5].

The received-signal processing chain is built of a number subsystems: antenna, analog signal processing, digital signal processing, and data processing. These subsystems sequentially refine this input signal. This includes using a set of matched filters to generate an estimate of the true range profile. These matched filters can be expressed as the conjugate transpose of the sensing model: $\hat{\mathbf{x}} = \mathbf{S}^H \mathbf{y}$. The matched filter in the slow-time and phase center dimensions amount to Fourier transforms and the matched filter in the fast-time dimension is a convolution, which can be performed in the Fourier domain as well. The upstream portion of this processing is shown in Figure 2 for a notional radar subarray.

Waveform design is a rich field of research and various types of waveforms have been designed to produce different target response and matched-filter characteristics [6, 7, 8]. Wider bandwidth waveforms produce higher range resolution. The range resolution of a waveform with bandwidth β hertz is

$$\Delta_R = \frac{c}{2\beta}.$$

Commonly used wideband waveforms include linear frequency-modulated (LFM) chirp waveforms that can be easily synthesized and then processed using a frequency shifting system known as stretch processing in which the measured frequency is extracted to produce

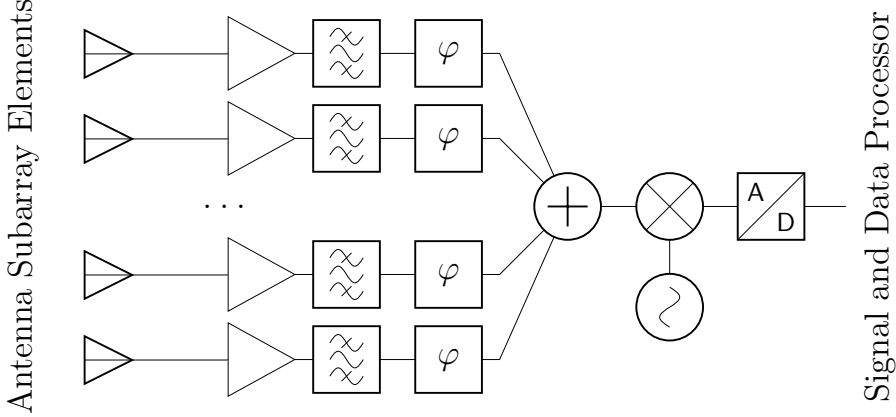


Figure 2: A simplified block diagram of a single subarray of a large airborne phased array radar. The signal path includes a low-noise amplifier, analog filtering, phase shifting, RF combining, mixing to IF, and sampling. The digital samples from this and all other subarrays feed into the digital signal and data processor for pulse compression, beamforming, detection, association, tracking, prediction, and other functions.

the measured range profile. Another class of commonly used waveforms are phase coded which can be further subdivided into bi-phase and multi-phase codes. Phase coded waveforms derive their bandwidth from the narrow phase chips which make up the waveform. Commonly used codes include Barker codes and maximal length sequences both of which minimize the waveform correlation sidelobes.

The radar system is subject to interference from various sources that inhibits accurate detection of targets. The most commonly treated type of interference is thermal noise that accumulates along the analog processing chain and at the analog-to-digital converter (ADC). This noise can be expressed in a linear model as $\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{n}$. To describe the power of the noise interference relative to the power of the desired signal, the signal quality metric known as signal-to-noise ratio (SNR) is used.

The signal to noise ratio can be calculated using the equation

$$\text{SNR} = \frac{P_t G_A^2 \sigma \lambda^2 G_{ibw} G_{dop}}{(4\pi)^3 k T B R^4 L N_f} \quad (1)$$

from [9], where P_t is transmitted signal power, G_A is antenna gain, σ is target radar cross section (RCS), λ is the wavelength of carrier, G_{ibw} is pulse compression gain, G_{dop} is Doppler processing gain, k is Boltzman's constant, T is system temperature, B is receiver bandwidth,

R is target range, L is other system losses, and N_f is receiver noise figure. This equation makes clear that the SNR is dependent on the geometry of the problem, as well as the physical characteristics of the antenna and receiver. On the processing side, gains may be made by using longer pulses so that the pulse compression gain is increased, or by coherently integrating more pulses so that the Doppler processing gain is increased. Both of these fundamentally mean spending more energy on the target to make the coherent signal stand out from the incoherent noise.

Finally, a detector is used to generate a number of detections and the locations of those detections [10, 11]. These detections may then be used to generate tracks on the targets over time, or for other real-time uses. The detector is applied as a threshold; entries that exceed it are identified as targets and others are not. Setting this threshold is a statistical exercise in maximizing the probability of detecting true targets (probability of detection, P_D) subject to some specified probability of falsely detecting non-targets (probability of false alarm, P_{FA}). By either calculating or estimating the statistical properties of the interference, the threshold can be set and by applying that threshold to the statistics of the target signal the resultant probability of detection can be estimated.

If the statistics of the interference change from one time instance to the next they may be adaptively estimated using the constant false alarm rate (CFAR) estimator and threshold. This uses the observations of the bins surrounding the cell under test to estimate the local interference statistics and to adaptively set the detection threshold.

2.2 Signal Model

A radar receives a post-pulse-compression data cube of dimensions $N_c \times N_s \times N_f$, for the channel, slow time, fast time sampled dimensions. The expected sampled impulse response of a point target at any location in the observation extent can be expressed as a function of the transmitted waveforms, sensing geometry, timeline, and pulse compression processing. These responses can be combined to form a sensing matrix \mathbf{S} which is a transform from the discretized target space (angle, radial velocity, and range of dimension $N_a \times N_v \times N_r$) to

the sampled data cube. Together, the discretized sensing model can be described as:

$$\mathbf{y} = \mathbf{S}\mathbf{x} \quad (2)$$

where \mathbf{y} contains the samples in time and space, and \mathbf{x} is an indicator vector giving radar cross section of scatterers at each location in the observation extent – meaning it contains mostly zeros because many gridpoints contain no target. The three dimensions underlying \mathbf{y} and \mathbf{x} have been vectorized.

In the real world targets are not confined to a discrete set of locations, nor is clutter fully resolved by the sampling rate in either range or cross-range. Nevertheless, we use this simpler model for analysis and computation, cognizant of its limits. When we simulate and test algorithms we do so with data generated by higher resolution and higher fidelity models that include off-grid targets, unresolved returns, and other realistic nonlinearities.

Each column of \mathbf{S} is a steering vector to a bin in the range-angle-Doppler cube. Much of the work in STAP and other adaptive radar literature uses the term “steering vector” to mean a strictly angle-Doppler dictionary element. Some work has been done regarding the estimation and use of the 3-D adaptive filter, for instance in [12]. Also, radars are becoming more fully digitized as analog-to-digital converters continue to move closer to the individual antenna elements; pulse compression is increasingly performed digitally [13]. Thus we mean the steering vector to represent the three-dimension dictionary element. This basis can be constructed from the bases that describe the response along each individual dimension:

$$\mathbf{S} = \mathbf{S}_r \otimes \mathbf{S}_a \otimes \mathbf{S}_d. \quad (3)$$

This model describes simple propagation phenomena. Specify the target location by placing the i -th point target at range r_i from the antenna and angle θ_i from the array boresight with range rate v_i . Define the range rate as the derivative of the target’s range, $v_i = \frac{d}{dt}r_i$. In the case of the simulation v_i is selected from a uniform random distribution over the sampled Doppler frequency space. This target is illuminated by a series of n_s identical waveforms with carrier frequency f_0 , i.e., each waveform $p(t) = e^{2\pi j f_0 t} e^{2\pi i \phi(t)}$ with pulse repetition interval T_s . The waveform $p(t)$ has some bandwidth β , whether by swept

frequency chirp, phase code sequence, or some other modulation function. The illumination experienced at the i -th target is then

$$e_i(t) = \alpha_i \sum_{q=0}^{n_s-1} p(t - qT_s - (r_i + v_i qT_s)/c) \quad (4)$$

for some scalar α_i .

A moving target imparts a Doppler frequency shift on the waveform commensurate with its radial velocity (positive shift for decreasing range), and some of this energy is reflected back to the antenna array to be received. The array consists of n_e individual array elements uniformly separated by a distance d . We neglect the element pattern of any array element, and instead model them as isotropic receivers. Each of these array elements makes n_f uniformly-spaced fast time samples on in-phase and quadrature channels, i.e., these samples are points in the complex plane. The signal reflected from target i received at element k is

$$f_{i,k}(t) = \nu \sum_{q=0}^{n_s-1} p\left(t - qT_s - \frac{2(r_i + qv_i T_s) - kd \sin \theta_i}{c}\right) e^{2\pi j \frac{v_i}{f_0 c} t} \quad (5)$$

for some scalar ν . In a digital receiver $f_{i,k}(t)$ is sampled in fast time every $T_f = 1/\beta$ sec.

The observable Doppler frequencies, dictated by the Nyquist sampling rate, are

$$-\frac{PRF}{2} \leq f_d \leq \frac{PRF}{2}. \quad (6)$$

In the range basis, \mathbf{S}_r , each column is a shifted copy of the transmitted waveform, w , with leading and trailing zeros. The first column is the response from a target at the minimum range in the range window and in each subsequent column the waveform is shifted down by one entry.

$$\mathbf{S}_r = \begin{bmatrix} w(1) & 0 & 0 & \cdots & 0 \\ w(2) & w(1) & 0 & \cdots & 0 \\ w(3) & w(2) & w(1) & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & w(n_r) \end{bmatrix} \quad (7)$$

The angle and Doppler bases, \mathbf{S}_a and \mathbf{S}_d , are both frequency bases. The Doppler frequency is evaluated across pulses in the coherent processing interval. The angle basis is the

relative phase delay introduced at the elements of the antenna array as an incoming planar wavefront reaches each element sequentially. The time delay between receivers becomes a simple phase shift for narrow-band signals. For wideband signals, the spatial phase pattern becomes frequency-dependent. To coherently process signals in a wideband setting, other approaches can be used to minimize the losses due to phase mismatch. These include true time delay units or corrections applied in the digital domain to counteract the known phase errors introduced. For radar systems made up of subarrays, true phase combining can be used at the subarray level if the subarray is small enough and/or the bandwidth is small enough to adequately control phase migration.

This linear model can be represented explicitly as a matrix. For more details about the following construction of the forward model see [14] and [15]. Each column of the matrix \mathbf{S} is the return from a target at the corresponding range-angle-Doppler position in space. The size of the matrix grows rapidly as the dimension of the sample space and search space increase. For a system with 8 channels, 128 pulses and 512 range samples the matrix \mathbf{S} has 2.8×10^{11} elements. Therefore, the most efficient way to implement this model is not by explicitly storing the matrix but by performing discrete Fourier transforms along the angle and velocity dimensions and a convolution in the range dimension. For computational gains, these can be implemented using the Fast Fourier Transform (FFT). By this means we reduce storage requirements and processing time considerably.

This work assumes a single-dimension uniform linear array, but all these results are generalizable to a planar array. In this work we neglect that fourth dimension in the interest of clarity and computational tractability. Furthermore, any polarization effects are neglected in this model. For systems that record dual polarization simultaneously or sequentially, an added dimension could be used to represent that variable.

In short, the i -th column of \mathbf{S} is the return one expects to receive from a target in the i -th voxel of the range-angle-Doppler cube. Of course targets move through continuous space, and so any discretization will necessarily be only an approximation. Finer and finer discretization can reduce the associated errors, but at the cost of increasing the correlation between columns of the matrix. With any processing there are diminishing returns as the

discretization becomes finer than the resolution of the sensor. Furthermore, as will become clear in Section 2.5, this increased correlation is particularly detrimental to compressed sensing and sparse recovery processing techniques.

Circular Gaussian noise $\mathbf{n} \in \mathbb{C}^n$ also enters the measurements with a variance set by the noise level of the receivers and other elements in the processing chain. As described above in the radar range equation, these noise sources can be summarized by the system noise figure or temperature. Quantization noise, though deterministically related to the sampled signal, can be treated as Gaussian for reasonable resolution analog-to-digital converters. Thus the measurement can be decomposed into signal and interference components:

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{n}. \quad (8)$$

2.3 Phenomenology of Correlated Interference

The phenomenology of white noise interference is the simplest and most commonly treated interference type. In some applications it is the dominant interference source. This is particularly true for targets at long ranges and for radars that detect targets against an empty background. But other sources of interference exist. Some of these sources of interference are deterministic and thus exhibit structure according to the mechanism by which they are caused. Mitigating these sources of interference requires more than using longer waveforms or more pulses. Such techniques increase the target energy but also increase the correlated interference energy. This interference is said to be correlated because its structure can be described statistically by its expected autocorrelation. Correlated interference takes on a number of forms. The three most common sources are distributed ground clutter, in-band radio frequency interference (RFI) including adversarial jamming, and differences between the true and the modeled sensing system.

2.3.1 Clutter

Clutter returns come from energy that is reflected back to the radar by objects other than the intended targets. Thus, the definition of clutter is application-specific. For a system designed to detect motor vehicles in forested terrain, the trees and ground are the clutter.

However, for the purpose of performing geographic land use surveys, the ground and trees are targets. Clutter and other structured interference degrade performance in this and other applications. This work considers the case where the targets are moving objects, including especially motor vehicles, and the clutter consists of terrain, foliage, and buildings.

To include this in the sensing model, a clutter vector \mathbf{c} is introduced, which contains the geometry- and terrain-dependent clutter reflectivity at the grid locations.

$$\mathbf{y} = \mathbf{S}(\mathbf{x} + \mathbf{c}) + \mathbf{n}. \quad (9)$$

The clutter is illuminated by the same waveform as the targets themselves and thus signal processing that is used to improve signal-to-noise ratio of the targets will also cause the clutter to be accentuated as well. So more energetic waveforms or more integration will not be useful in improving the signal-to-clutter ratio (SCR).

Clutter can result from land or sea surface reflections and the statistics of the interference depend strongly on the particularities of the terrain being surveilled. Much research has focused on describing the expected returns from clutter as a function of terrain, radar band, radar resolution, and other parameters. The simplest model is to assume a Gaussian distribution; for low-bandwidth radars this assumption is frequently sufficient. A common distribution for clutter amplitude is gamma function [16, 17]. But upon more detailed analysis some data shows significant skewness and kurtosis that do not match either the normal or gamma distributions. Having examined K , log-normal, Weibull, and Rayleigh distributions, work in [18] showed that the Weibull distribution best described the data collected in a flight test over open farmland in Saskatchewan. In [19] research led to a compound distribution composed of two separate gamma distributions that describe the modulation and speckle observed in high-resolution radar.

So with all this in mind, we have elected to use a a constant gamma-distributed random model variable as in [16], with a shape factor of 10/3 in our simulation. This model approximates a terrain with relatively open land and the absence of man-made scatterers. Figure 3 shows the amplitude of clutter observed in a realization of this clutter model. Other models may be more appropriate at very low grazing angles and for other types of terrain [19], [18].

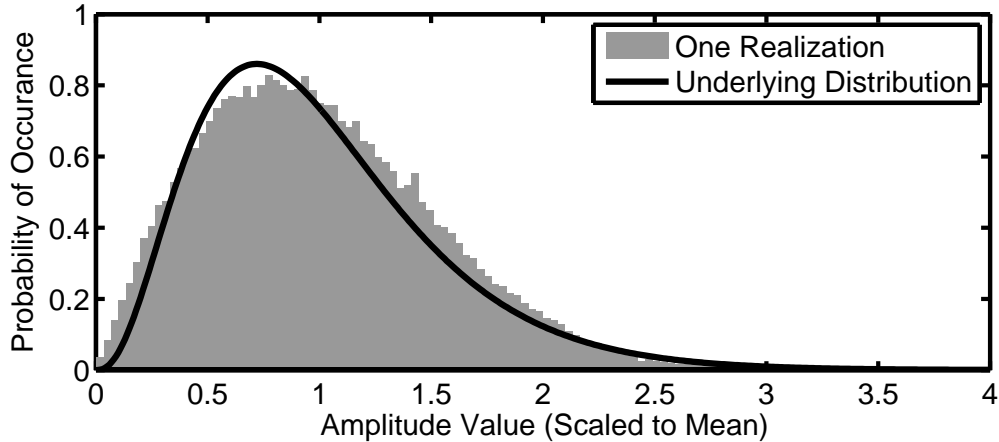


Figure 3: The sampled distribution of the observed clutter in one realization, along with the underlying gamma distribution which is defined by a shape parameter of $10/3$.

2.3.2 RFI and Jamming

In-band radio-frequency interference, to include adversarial jamming, is not necessarily correlated with the transmitted waveform (though in the case of more advanced jammers it may be). Thus it enters the linear sensing equation in the same way as the system noise:

$$\mathbf{y} = \mathbf{S}(\mathbf{x} + \mathbf{c}) + \mathbf{j} + \mathbf{n}. \quad (10)$$

Once processed using matched filtering or other techniques this energy may be spread throughout the target estimation space which is then observed as a raised noise floor, or it may be focused into a smaller number of bins and appear to be false targets. The statistical structure of the interference may change from one time instance to the next making estimation difficult.

2.3.3 System Model Error

Other sources of interference result in system model error. In these cases, the true measurement system is not equal to the estimated system. Common causes of model error include mismatches between channels in the analog signal processing, non-idealities in components like antenna phase shifters, or sensor position and orientation estimation error. These types

of errors can be expressed as

$$\mathbf{y} = (\mathbf{S}_{est} + \mathbf{S}_{err})(\mathbf{x} + \mathbf{c}) + \mathbf{j} + \mathbf{n}. \quad (11)$$

These types of model error have been subject to much attention by radar designers in an attempt to minimize hardware errors and biases. Whether these errors significantly affect the ability of the system to identify targets depends on the magnitude of the errors themselves [20]. Several processing approaches have been proposed to simultaneously calibrate an estimated sensing matrix while providing a target state estimate [21].

2.4 Signal Processing Treatment of Clutter

Processing only using matched filtering applies a non-adaptive technique to a target scene with statistics that exhibit inter-bin correlation structure which can then be used to improve target detection performance. Several approaches to treating the correlated clutter problem are described in this section.

2.4.1 Early Techniques

The two-pulse canceler is a filtering technique in which the returns from two consecutive pulses are subtracted. If the clutter is stationary from pulse to pulse, it will be nulled by this filtering. If targets exhibit radial motion their return will not be subject to as much nulling. The null produced by this technique can be quite broad, and might hide slow-moving targets or targets moving nearly perpendicular to the radar's radial vector. In addition, this filter introduces nulls at Doppler frequencies equal to integer multiples of the PRF which correspond to blind velocities [22]. The canceler technique can be employed non-coherently with reduced performance if radar system phase tolerances are not sufficient to support pulse-to-pulse coherent processing. A number of improvements can be made to this basic filtering concept. Differencing filters can be cascaded to improve filter performance at the null locations. Additionally, the PRF can be staggered within a CPI to resolve blind velocities and null the clutter at zero frequency while passing moving targets.

2.4.2 Space-Time Adaptive Processing

The principal technique for reducing the effects of structured interference in a multi-channel radar system is adaptive filtering. This technique estimates the interference statistics to produce a filter to be applied to the received signals. In the multi-channel pulse-Doppler radar case under examination a two-dimensional filter, known as space-time adaptive processing (STAP), may be employed. STAP uses the three-dimensional data cube to estimate the interference structure over the joint angle-velocity space using the range samples as training data [4, 23, 24, 25]. Especially in downward looking airborne radar, the clutter may interfere with the target in either the Doppler or angle dimension but by processing them jointly the target often falls outside the clutter support.

The Wiener filter defines the optimal filter for maximizing signal-to-noise ratio in Gaussian interference. The measurements defined above with noise, clutter, and other interference $\mathbf{y} = \mathbf{S}(\mathbf{x} + \mathbf{c}) + \mathbf{j} + \mathbf{n}$ can be broken into signal and interference portions: $\mathbf{y} = \mathbf{y}_s + \mathbf{y}_i = (\mathbf{S}\mathbf{x}) + (\mathbf{S}\mathbf{c} + \mathbf{j} + \mathbf{n})$. The covariance matrix can then be formed to describe the interference statistics: $\mathbf{R} = E(\mathbf{y}_i \mathbf{y}_i^H)$. This covariance matrix can be used to produce the filter that whitens the interference and maximizes target detectability: $\mathbf{W}_{Wiener} = \mathbf{R}^{-1}\mathbf{S}$ and the target estimate $\hat{\mathbf{x}}_{Wiener} = \mathbf{W}_{Wiener}\mathbf{y}$.

In the most basic expression, STAP techniques estimate the interference covariance matrix for each range bin $\mathbf{R} = E[\mathbf{y}_i \mathbf{y}_i^H]$ from a set of training data that usually includes nearby range bins. This sample covariance matrix is then inverted to generate a whitening filter: $\mathbf{W} = \mathbf{R}^{-1}\mathbf{S}$. Computing \mathbf{R}^{-1} directly through sample matrix inversion (SMI) can be computationally expensive. Other methods exploit the low-rank structure of the clutter to estimate \mathbf{R}^{-1} from an eigen-decomposition of the sampled data. The sample covariance matrix is a consistent estimator of the true covariance matrix, but only when the number of training samples is large. In cases with limited training this estimate is biased, as will be shown in Chapter 5, on random matrix theory. In the case of sufficient training data, the constructed filter maximizes signal-to-interference-plus-noise ratio (SINR) when applied to the measured data:

$$\hat{\mathbf{x}}_{stap} = \kappa \mathbf{W}^H \mathbf{y} = \kappa \mathbf{S}^H \mathbf{R}^{-1} \mathbf{y}. \quad (12)$$

where the scaling factor $\kappa = (\mathbf{S}^H \mathbf{R}^{-1} \mathbf{S})^{-1}$. Here SINR is defined as $E \left(\frac{\|\mathbf{W}^H \mathbf{y}_s\|_2}{\|\mathbf{W}^H \mathbf{y}_i\|_2} \right)$. This can be compared to the adjoint, or matched filter, estimate which does not make use of the interference covariance information:

$$\hat{\mathbf{x}}_{adj} = \mathbf{S}^H \mathbf{y}. \quad (13)$$

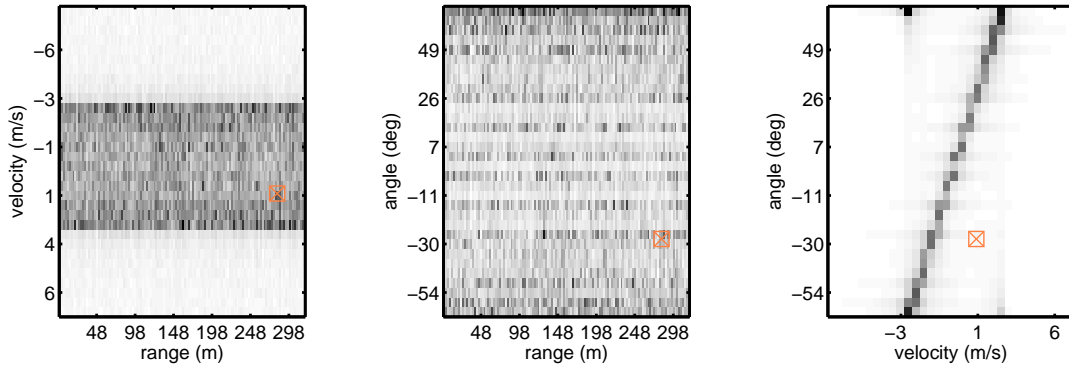


Figure 4: The adjoint estimate $\hat{\mathbf{x}}_{adj}$, reshaped into the range-angle-Doppler cube, and projected along each of the three dimensions. The true target location is $r = 290$ m, $\theta = -30^\circ$ and $v = 1$ m/s. The marker indicates the true target location in each view.

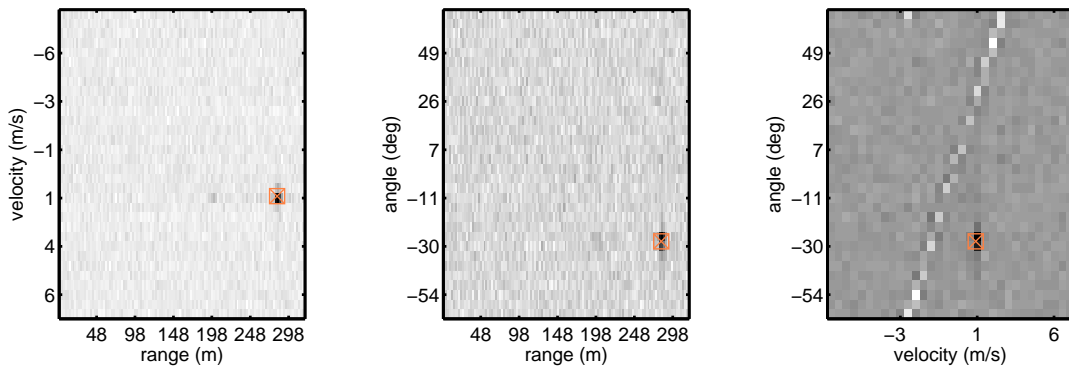


Figure 5: The STAP estimate $\hat{\mathbf{x}}_{stap}$, reshaped into the range-angle-Doppler cube, and projected along each of the three dimensions. The true target location is $r = 290$ m, $\theta = -30^\circ$ and $v = 1$ m/s. The marker indicates the true target location in each view.

Figures 4 and 5 illustrate the ability of the STAP filter to reduce the contribution of clutter in the estimate while maintaining target detectability in uncluttered regions. The line of clutter appears as a slice in the third frame of the figures due to the interaction of the sensor platform motion with the stationary ground clutter. In the matched-filter-only

estimate that clutter line is such a strong ridge in the angle-Doppler plane that it renders the target invisible. In the STAP estimate, the clutter has been nulled by the filter and the target can be easily located in all three projections. This is a high signal-to-clutter-ratio (SCR) example to enhance the visibility of the target relative to the clutter. This same framework will be used to generate experimental results in the latter portion of this chapter, although with different parameters for signal, clutter, and noise power.

Acquiring sufficient training data is a constant challenge for adaptive algorithms. The statistics of the interference must be estimated from data in adjacent or nearby range bins. The Reed-Mallet-Brennan rule [26] states that a stable estimate can be made from a set of training data twice as large as the number of degrees of freedom in the covariance. Especially in cases of inhomogeneous clutter this requirement can be difficult to meet. This rule of thumb provides a reasonable goal, but it does not show how the expected estimates change as a function of samples used. Random matrix theory, which will be introduced in Chapter 5 provides such information. Furthermore, the topic of data sufficiency is complicated by the introduction of compressed sampling as will be introduced in Section 2.5.2.

2.4.3 Reduced Rank STAP

The accuracy of the estimate of the interference statistics decreases with the number of degrees of freedom in the estimate (for a fixed training data set). To improve the stability of the estimates at the expense of resolution, reduced rank STAP is utilized. These techniques work well because the interference information can be described using relatively few basis vectors. In [12], it is shown that signal dependent rank reduction can improve performance. The cross-spectral metric for designing the rank reducer incorporates information on the desired signal steering vectors rather than only the interference statistics to improve performance in cases of limited training data. In [27], it is shown that reduced-rank STAP techniques offer better performance than full-rank estimation under constraints on training time and data. In [28], it is shown how one may select a dimensionality reduction basis efficiently. In addition to a performance improvement given limited training data, these

reduced-rank techniques also offer reduced computational burden relative to a full-rank algorithm. The optimal rank may be set by processing timeline constraints, but even in the unconstrained case learning the optimal number of dimensions to use for estimating the covariance is not straightforward. This topic will be treated in Section 5.3 by applying results from random matrix theory.

2.4.4 Robust Covariance Estimators

Regularizing the estimate of the covariance matrix to account for small sample sets or outlier observations has been studied and several approaches exist. The utility of these approaches depends somewhat on the intended application space. The concept of M -estimation of multi-variate distributions from observations found its first elaboration in Huber [29]. Subsequently Maronna extended this work to include the scatter matrix estimate [30]. These results generalize the covariance estimation problem beyond the multi-variate normal case to elliptical distributions which can cover heavy-tailed phenomenologies that drive the need for robust estimators. Maronna defines the M -estimator as a matrix \mathbf{V} and vector \mathbf{t} satisfying

$$\begin{aligned} n^{-1} \sum_{i=1}^n u_1 \left(\sqrt{(\mathbf{x}_i - \mathbf{t})^H \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{t})} \right) (\mathbf{x}_i - \mathbf{t}) &= \mathbf{0} \\ n^{-1} \sum_{i=1}^n u_2 \left((\mathbf{x}_i - \mathbf{t})^H \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{t}) \right) (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^H &= \mathbf{V} \end{aligned} \tag{14}$$

for functions $u_1(\cdot)$ and $u_2(\cdot)$ nonnegative, nonincreasing, and continuous. The existence and uniqueness of the solution to these equations has been proved. Furthermore, this estimate converges to the true mean and covariance of the distribution as $n \rightarrow \infty$.

Tyler [31] proposes as an estimator of the scatter matrix the solution to

$$\frac{m}{n} \sum_{i=1}^n \frac{(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^H}{(\mathbf{x}_i - \mathbf{t})^H \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{t})} = \mathbf{V}. \tag{15}$$

This provides the same objective as Maronna in the case $u_2(s) = m/s$. This estimator is argued to be the “most robust” estimator for the scatter matrix of an elliptical population, which is true if robustness is defined in a minimax sense. Tyler’s approach is only valid when the number of samples exceeds the dimensionality of the variables. This limitation is remedied with the modified fixed-point iterations provided in [32], and further improved

upon in [33], which provides optimal shrinkage given a specified rank estimate. We will return to these robust estimators in Section 5.3 when comparing performance of the Tyler estimator to a reduced-rank and diagonally loaded estimate.

Another approach, known as rank-constrained maximum likelihood uses a prior estimate of the clutter rank based on the sensing geometry to estimate the covariance [34]. A disadvantage to this approach is its reliance on an estimate of the clutter rank, which can vary substantially as a function of ground cover, wind speed, or characteristics of the radar processing. Additionally, other types of interference can cause changes to the spectral structure which would not be captured in the prior estimate.

2.5 *Compressed Sensing*

2.5.1 Formulation and Assumptions

Compressed sensing is a recently developed theory by which some vectors may be exactly or stably recovered from an underdetermined system of linear equations. If the problem to be solved obeys certain conditions an exact or approximate solution can be found with a rigorously bounded degree of certainty [35, 36, 37]. Compressed sensing operates under two assumptions [38]:

1. Sparsity: The signal to be reconstructed has few nonzero elements, or can be represented in some basis by few nonzero elements.
2. Incoherence: The measurement model is incoherent with the sparsifying basis.

These may be expressed as linear operations:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{z}, \quad \mathbf{\Phi} \in \mathbb{C}^{m \times n}. \quad (16)$$

In this formulation \mathbf{z} is the unknown vector in the large n -dimensional space. The number of observations made is m , each of which is an inner product with a known measurement vector (a row of $\mathbf{\Phi}$).

This may be generalized to allow unknown vectors that can be sparsified in some other basis, $\mathbf{\Psi}$:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{x}, \quad \mathbf{\Psi} \in \mathbb{C}^{n \times n} \quad (17)$$

The vector \mathbf{x} may be recovered efficiently by solving a convex optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{x} \quad (18)$$

If there is any error (noise or otherwise) in the observations a relaxation may be introduced:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{x}\|_2^2 < \tau \quad (19)$$

or

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (20)$$

These formulations can be shown to yield equivalent results for appropriate choices of τ and λ . This approach is variously known as LASSO, basis pursuit denoising, or ℓ_1 -regularized least squares. In all these cases the 1-norm is used to promote sparsity while maintaining convexity. Recall that the ℓ_p -norm of some vector \mathbf{x} is defined as $(\sum_i |x_i|^p)^{\frac{1}{p}}$.

The required degree of incoherence can be described by the restricted isometry property. If the combined operator $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$ preserves the ℓ_2 norm of any k -sparse vector, CS will succeed with high-probability. Define $\epsilon_k(\mathbf{A})$ as the minimum ϵ such that

$$1 - \epsilon \leq \frac{\|\mathbf{A}\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \leq 1 + \epsilon \quad (21)$$

for all k -sparse $\mathbf{z} \in \mathbb{C}^n$. The result assumes the columns of \mathbf{A} have been normalized to unit norm. Smaller ϵ_k implies reconstruction is more likely to succeed. Finding $\epsilon_k(\mathbf{A})$ requires evaluating all $\binom{n}{k}$ possible support sets of \mathbf{z} .

Adherence to this rule is difficult to ascertain except by an expensive exhaustive search. A measure that supports weaker claims while being easier to calculate is the mutual coherence. For some matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ normalized such that each column has unit norm, define

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{n} \max_{i,j} |\langle \phi^i, \psi_j \rangle| \in [1, \sqrt{n}] \quad (22)$$

where ϕ^i is the i -th row of $\mathbf{\Phi}$ and ψ_j is the j -th column of $\mathbf{\Psi}$. If μ is the maximum correlation between elements of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ then smaller μ implies reconstruction is easier and more likely to succeed.

For certain types of sparsifying bases there exist deterministic measurement operators that obey the incoherence rule (22). For other measurement operators such a deterministic solution cannot be shown. It can be proved, however, that measurement operators drawn from certain distributions of random variables are overwhelmingly likely to be incoherent with the sparsifying bases [38].

Others have identified the compressed matched filter as an effective way to perform classification and detection tasks with lower computational burden than the convex optimization approach [39]. This technique generates another estimate of \mathbf{x} :

$$\hat{\mathbf{x}} = \Psi^H \Phi^H \mathbf{y}. \quad (23)$$

Although $\hat{\mathbf{x}}$ is not a sparse estimate, for some cases this result may provide nearly the same utility as one that is. This is especially true if the sidelobes introduced either by the sensing model or by the compression operation itself are well-behaved. For true random sampling these sidelobes can manifest mostly as an elevation to the noise floor with broad, even distribution through the estimated vector. But for other sampling approaches with more structure the sidelobes can negatively affect the utility of the compressed matched filter.

2.5.2 CS in Radar

Applications of compressed sensing are numerous including geophysical sensing, medical imaging, and genomic testing. Even within the field of radar sensing these ideas have diverse utility and have generated widespread interest. Topics include multiple-input, multiple output (MIMO) [40, 41, 42], synthetic aperture radar (SAR) [43], detection [44], subsurface imaging [45], and Doppler processing [46]. Many of these applications come about because of the relatively slow progress in analog to digital converter (ADC) performance compared to the rapid increases in general computation performance [47]. Some working prototype radar systems that gather and process compressed measurements have been built [48].

To take advantage of this body of work the conditions on sparsity and sampling incoherence must be met. Properties of the radar sensing problem presented in (9), including the linear formulation and sparse target vector, make application of compressed sensing

techniques a natural step. Sparsity is evident in the formulation of (9). Sparsity is natural in any surveillance application where the density of targets is very low. In a search space that may be discretized into thousands or millions of bins, targets are rare and form a sparse support to be estimated. In many realistic applications this will hold true. For instance, in an airspace monitoring radar most points in the observable sky do not contain an airplane but those that do compose the support set of the nonzero elements of a sparse vector. Other applications like synthetic aperture radar (SAR) in which images are formed of a land area, do not exhibit native sparsity and may require representation in some other sparsifying basis. Because of the large extent of the observed clutter, downward looking MTI radar of the sort we consider herein does not possess a native sparsity.

Coherence depends on the structure of the sampling process but can be guaranteed with high probability by sampling the inner product of the signal with a (pseudo-) random process. To describe the contributions to the area of interest build on the syntax introduced in (9) by adding a compression operator \mathbf{C} that undersamples the incident signals. If \mathbf{y} is the set of Nyquist-sampled measurements then it represents all the (bandlimited) electromagnetic information passing over the measurement aperture during the period of observation (a single coherent processing interval). The compressed measurements are modeled as another vector

$$\mathbf{z} = \mathbf{C}\mathbf{y} \tag{24}$$

where $\mathbf{z} \in \mathbb{C}^m$ and $\mathbf{C} \in \mathbb{C}^{m \times n}$. As indicated above, these compressive measurements must be incoherent with the sparsifying basis \mathbf{S} . The undersampling factor (USF) is defined as the ratio $u = n/m$ where a higher USF indicates that the measurements have been subject to a greater degree of compression.

A simple estimate of the target vector from the compressed measurements \mathbf{z} can be computed by performing a compressed adjoint as in (23) which applies the matched filter to the available measured data.

$$\hat{\mathbf{x}}_{\text{cadj}} = \mathbf{S}^H \mathbf{C}^H \mathbf{z}. \tag{25}$$

This method has low computational cost but does not necessarily yield a sparse solution.

A compressed sensing estimate of the target vector can be computed by solving a convex linear program such as an ℓ_1 -regularized least-squares

$$\hat{\mathbf{x}}_{cs} = \arg \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{C}\mathbf{S}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (26)$$

where the first term of the minimization objective is the Euclidean norm of the residual that enforces fidelity to the measured data, and the second term is the ℓ_1 norm of the estimate that promotes sparsity in the solution. The parameter λ enables a trade off between these competing priorities.

Much attention has been paid to how the random sampling at the core of compressed sensing can be realized in hardware. The exact content of the compression matrix \mathbf{C} will depend on the measurement process it describes. In pulse-Doppler radar this process may introduce incoherence in fast-time by mixing incoming signals with pseudo-random modulation sequences before low-pass filtering and sampling slowly [49], it may introduce incoherence in slow-time by staggering the pulse repetition interval [50], it may introduce incoherence in the spatial domain using a random measurement array [51, 52], or a coprime thinned array [53].

These ideas will not be treated in this work. Given that the designer enjoys some freedom in the design of the measurement process that \mathbf{C} describes, it could be adapted in coordination with the waveform and sensing geometries described by \mathbf{S} and the interference described by \mathbf{R} to maximize target detection. We simply construct \mathbf{C} as a random matrix filled with independent and identically distributed (i.i.d.) Rademacher random variables that take on the values ± 1 with equal likelihood. An implemented radar system would likely not use a measurement process that is well-described by this matrix, but it is an analytically useful matrix that provides an upper limit on performance.

2.5.3 Matrix Completion

Related to compressed sensing recovery of sparse vectors is an area of work around the recovery of low-rank matrices from fewer samples than matrix elements. This area of work is known as matrix completion. In the same way that the ℓ_1 norm offers a convex relaxation

of a direct sparsity (ℓ_0 “norm”) objective, the nuclear norm (denoted as $\|\cdot\|_*$), which is the sum of the matrix singular values, offers a convex relaxation of the $\text{rank}(\cdot)$ objective.

Let some unknown matrix $\mathbf{X} \in \mathbb{C}^{n \times n}$ exist with rank r . Some number m observations, $\mathbf{M}_{i,j}$, of this matrix are available only at the support set $(i, j) \in \Omega$. In [54] it is shown that low-rank matrix recovery from few measurements is not too ill-posed and is convex. The convex optimization problem

$$\min \|\mathbf{X}\|_* \text{ s. t. } \mathbf{X}_{i,j} = \mathbf{M}_{i,j}, (i, j) \in \Omega \quad (27)$$

will recover the matrix \mathbf{X} exactly with high probability if $m \geq Cn^{1.2}r \log n$, for a specified constant C .

In [55], the topic of matrix completion is surveyed and it is shown that $n \times n$ matrices of rank r can be recovered from m noise-corrupted direct samples via nuclear norm minimization with high probability if $m \geq Cnr \log^2 n$ with an error on the order of the noise level. In [56], this idea is expanded to show that matrices can be recovered from expansion coefficients with respect to a known matrix basis as long as that basis is not coherent with the matrix being recovered. This is directly analogous to the required incoherence between the sparsifying basis and the sensing basis in compressed sensing theory. The work in [57] provides information theoretic lower bounds on the number of samples needed to recover certain types of low-rank matrices. One result is that $m = Cnr \log n$ is the lower limit on the number of samples needed to recover a random $n \times n$ matrix with rank r . In [58], an algorithm for efficient matrix completion is provided; this approach uses iterative singular value thresholding and projection back onto the observation set. This algorithm is able to recover large matrices in low run-time relative to interior-point methods. Starting from \mathbf{M} , the observations, and $\mathcal{P}_\Omega(\cdot)$, the projection onto the observation domain, the iteration involves repeated application of

$$\begin{aligned} \mathbf{X} &= \text{shrink}(\mathbf{Y}, \tau) \\ \mathbf{Y} &= \mathbf{Y} + \delta \mathcal{P}_\Omega(\mathbf{M} - \mathbf{X}) \end{aligned} \quad (28)$$

returning \mathbf{X} . The $\text{shrink}(\cdot)$ operation is soft-thresholding of the singular values of its argument. When the SVD of matrix $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, where $\mathbf{S} = \text{diag}(\mathbf{s}) = \text{diag}([s_1, \dots, s_n]^T)$,

$$\begin{aligned} \text{shrink}(\mathbf{A}; \tau) &:= \mathbf{U} \text{diag}(\text{soft}(\mathbf{s}; \tau)) \mathbf{V}^H, \\ \text{soft}(\mathbf{s}; \tau) &:= [\text{soft}(s_i; \tau), i \in 1 \dots n]^T, \\ \text{soft}(s_i; \tau) &:= \frac{s_i}{|s_i|} \max(0, |s_i| - \tau). \end{aligned} \tag{29}$$

The iteration in Equation (28) solves

$$\begin{aligned} \min \tau \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X}\|_F^2 \\ \text{s. t. } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}) \end{aligned} \tag{30}$$

This iteration approaches the direct nuclear norm objective for large values of τ , and smaller τ improves the stability of the solution. This technique solves for large 1000×1000 matrices in several seconds on a personal computer. This work is the launching pad for the proposed low-rank covariance estimation technique developed in this thesis. Also, we note that approaches from the low-rank matrix approximation body of theory have been applied to the moving target detection problem, for example in [59].

2.6 Recent Work

This section outlines some prior and ongoing work related to the topic of reducing clutter in the compressed sensing radar context.

The work in [60] treats the topic of estimating simultaneously sparse and low-rank matrices from rank-one measurements. This measurement model consists of a series of sketches of the underlying matrix $y[i] = \mathbf{a}_i^H \mathbf{\Sigma} \mathbf{a}_i$. In [61] this same “sketching” measurement model is considered and shows specifically that the number of measurements required for stable estimation scales linearly with the rank of the matrix and the sparsity of the matrix and with the logarithm of the number of rows. Though this model differs from that which we will develop in this thesis, the results provide a basis for confidence that by exploiting the low-rank nature of the covariance an improvement in performance can be expected.

In [62], it is shown that the low-rank assumption can be used to improve the estimation accuracy of the covariance matrix using a standard STAP benchmark dataset in the case

of limited training data. Their proposed algorithm out-performs other lower computational complexity algorithms by using a dictionary learning approach.

In [63] and [64], a CS-STAP technique is developed in which a small amount of training data, in some cases one snapshot, can be used to estimate the covariance statistics and to build the whitening filter. This technique assumes direct sparsity in the interference covariance matrix. The validity of this assumption depends on the type of interference being described. For certain types of electromagnetic interference, or for certain types of man-made clutter, a concentration in this domain can be very pronounced. For other types of natural ground cover the spectrum can be more distributed. This distribution is especially pronounced for foliage being blown in the wind.

In [65] an approach is proposed to estimate the sample covariance matrix (that would be obtained from the uncompressed data) from a set of compressed measurements. The goal of estimating the sample covariance matrix is slightly different than that of estimating the underlying interference statistics, but does isolate the two stages of sampling limits: first, a sample covariance matrix deviates from the true covariance matrix because it is based on a limited number of realizations, and second, the compressed sample covariance matrix differs from the full sample covariance matrix because of the dimensionality reduction. The nature of the second limit depends on the manner in which the samples are compressed and the estimator used to generate the sample covariance matrix.

Additional work from a sparse-estimation perspective has taken place recently. One paper that presents an ℓ_1 -regularized MTI is given by Yang, et. al., [66]. This technique assumes sparsity in the rank of the interference subspace relative to the number of system degrees of freedom.

In [46] a filter is developed that assumes sparsity of the target vector in the angle-Doppler domain to solve a regularized optimization problem over the portion of the plane that is judged to be outside the clutter ridge. For the purposes of estimating the statistics of clutter for use in adaptive filtering, compressing the samples exacerbates challenges related to paucity of training data.

2.7 Summary

Despite years of research, radar sensing continues to be an active field of inquiry and development for engineers, scientists and mathematicians. New applications, new platform requirements, and new technologies have allowed fresh ideas to take root. One of those fresh ideas is the application of compressed sensing theory to radar sensing. Much attention has been paid to the design and structure of a compressed radar and to the detection of objects under modest noise interference. Less attention has focused on mitigating other types of interference like clutter or RFI in a compressed sensing radar application. Some recent work exists here and more mature related topics exist, especially reduced-rank STAP. This dissertation seeks to provide a path forward to estimate the covariance structure from the compressed measurements and to use that covariance estimate to improve detection in a convex optimization framework.

CHAPTER III

APPLYING INTERFERENCE COVARIANCE IN COMPRESSED SENSING ESTIMATION

This chapter describes the topic of how the covariance matrix can be incorporated into the compressed sensing solution framework. In the optimization framework that balances the ℓ_2 -norm of the measurement residual and the ℓ_1 -norm of the estimate, the two norm can be modified to form an elliptical norm that uses the covariance matrix to appropriately shape the optimality surface. This approach is tested on simulated clutter data to measure target detection performance.

3.1 Introduction

Main-beam clutter may be safely neglected in some radar applications, but in many airborne and surveillance applications strong ground returns swamp target energy and must be attenuated. While a more energetic waveform can bring targets out of noise interference it illuminates target and clutter alike leaving the signal-to-clutter ratio unchanged. Ground returns can be tens of decibels stronger than those of the target making detection of targets difficult. But clutter interference exhibits structure that depends on the radar platform velocity, radar boresight, and ground geometry. This structure can be described by an interference covariance matrix that represents the relationship between interference in the various measurement cells. The interference covariance matrix can be used to build a filter that minimizes clutter energy while preserving target energy.

Compressed sensing techniques have shown that certain signals can be sampled at rates less than the frequency extent of the signal. In ground moving target indicator (GMTI) radar systems sampling occurs in three dimensions: intra-pulse (fast time), pulse-to-pulse (slow time), and across the antenna aperture (spatial). By reducing the sampling requirements in these dimensions compressed sensing offers the potential to reduce the cost and complexity of such systems.

But little prior work exists on techniques for incorporating the covariance information into a compressed sensing problem solution. This work addresses that shortcoming by describing such a technique and showing that it improves the probability of detection of targets in the midst of strong clutter returns.

Additionally much of the literature of compressed sensing has focused on the reconstruction of vectors. In the radar community, however, much more attention is paid to detection performance metrics like probability of detection and probability of false alarm. We align our work along these detection metrics so as to compare the proposed solution techniques to those already in use and to make the performance metrics more valuable to practitioners.

3.2 Method

3.2.1 ASPEN

To test against realistic data we simulate measurements in the Adaptive Sensor Prototyping ENvironment (ASPENTM) tool developed at Georgia Tech Research Institute’s Sensors and Electromagnetic Applications Laboratory. ASPEN is a high fidelity clutter modeling simulation that supports flexible sensor definitions. We model a 32-element uniform linear array, 32-pulse coherent processing interval, and 128-sample range window, yielding a data cube of $2^{17} = 131,072$ samples. The ASPEN outputs consist of data (\mathbf{y}) measured at the Nyquist sampling rate, true target locations (from which we derive \mathbf{x}), single range bin interference covariance matrices (\mathbf{R}_1) as well as simulation parameters like the transmitted waveform and relevant geometry. We generate the all-range interference covariance matrix (\mathbf{R}) by assuming that the interference covariance statistics are stationary in range; this block-diagonal matrix can be expressed as the Kronecker product $\mathbf{R} = \mathbf{I} \otimes \mathbf{R}_1$.

Much work in the STAP literature focuses on the estimation of the covariance matrix from finite training data. We use the clairvoyant interference matrix to isolate the utility of using the interference information in the CS solution framework from the accuracy by which that matrix can be estimated. The conclusion to this work identifies covariance matrix estimation from compressed measurements as a key area for future work.

The ASPEN measurement model is nonlinear, breaking the assumption in (9). Targets

can exist off grid locations and nearby targets interfere nonlinearly. We use the forward model described in (9) as an approximation for processing. As in any real estimation problem, the true measurement model differs somewhat from the model used for processing. Though this linear model does not exactly match the more realistic simulation it provides theoretical and computational tractability while maintaining fidelity to the reference and serving the purpose of estimating target parameters.

3.2.2 Solution Methods

Various estimates of the target scene can then be computed from the measured data. We compute the matched filter (13), least-squares estimate (using conjugate gradient method), STAP estimate (12), or a compressed sensing estimate using the Templates for First-Order Conic Solvers (TFOCS) algorithm [67] and implementation [68].

The TFOCS tool set implements a number of state-of-the-art solution techniques for large-scale convex problems and provides a flexible framework for the solution of alternative problem formulations. It solves problems of the form

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathcal{A}(\mathbf{x}) + \mathbf{b} \in \mathcal{K}. \end{aligned} \tag{31}$$

In the objective, $\mathbf{x} \in \mathbb{R}^n$ and f is convex but not necessarily smooth. In the constraint, $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear operator, $\mathbf{b} \in \mathbb{R}^m$ is interference, and $\mathcal{K} \subseteq \mathbb{R}^m$ is a closed, convex cone. This framework accommodates the set of common compressed sensing optimization problems. TFOCS operates by:

1. Casting the problem to be solved in the form (31)
2. Formulating a dual problem that is more computationally efficient
3. Smoothing the dual problem to ensure differentiability over \mathcal{K} and thus speed convergence
4. Applying optimal first-order methods to the smoothed problem

For the first-order solution algorithm we chose that of Auslender and Teboulle for its simplicity and convergence rate [69].

Solving the ℓ_1 -regularized least-squares problem (26) requires a value be selected for λ . Any $\lambda \geq \lambda_{\max} = 2 \|\hat{\mathbf{x}}_{\text{cadj}}\|_{\infty}$ will cause $\hat{\mathbf{x}}$ to return as the zero vector. We select $\lambda = 0.005\lambda_{\max}$ by experimentation to produce consistent results and good detection performance.

3.2.3 Proposed Extension

None of the mentioned compressed sensing solution methods, as described, are equipped to handle structured interference. If \mathbf{R} is the covariance matrix of the fully sampled interference, then the covariance matrix of the interference in the compressed domain can be expressed as $\mathbf{R}_c = \mathbf{C}\mathbf{R}\mathbf{C}^H$. This use of the dimensionally-reduced covariance matrix is a non-data-adaptive reduced-dimension STAP formulation. The literature of reduced-dimension STAP provides a natural way to obtain an estimate of the scene from the compressed measurements while including the covariance information. Define

$$\hat{\mathbf{x}}_{\text{cstap}} = \mathbf{S}^H \mathbf{C}^H \mathbf{R}_c^{-1} \mathbf{z}. \quad (32)$$

This is analogous to the $\hat{\mathbf{x}}_{\text{stap}}$ solution defined in (12) in that the matched filter is post-multiplied by the covariance matrix inverse to whiten the interference. It is identical to the STAP solution in the case that \mathbf{C} is the identity matrix (i.e., the fully-sampled case). To compute this estimate one incurs the cost of inverting the covariance matrix (or parametrically estimating the inverse), but gains a good deal of clutter suppression as we will show in our results. However, this technique does not necessarily favor sparse solutions.

In addition to this technique, we propose a covariance-aware CS (CA CS) that accounts for structured interference in a compressed sensing framework. As in STAP methods, this is accomplished via the interference covariance matrix inverse. The original ℓ_1 regularized least squares problem of (26) can be modified as follows:

$$\hat{\mathbf{x}}_{\text{cacs}} = \arg \min_{\mathbf{x}} (\mathbf{z} - \mathbf{C}\mathbf{S}\mathbf{x})^H \mathbf{R}_c^{-1} (\mathbf{z} - \mathbf{C}\mathbf{S}\mathbf{x}) + \gamma \|\mathbf{x}\|_1. \quad (33)$$

This generalization of (26) is identical (for some value of γ) in the clutter free case. In that case \mathbf{R}_c is a scaled identity matrix. Here γ is set relative to the entries in \mathbf{R}_c^{-1} .

Specifically $\gamma = \lambda \|\mathbf{R}_c^{-1}\|_F$ using the Frobenius norm where $\|\mathbf{Q}\|_F = \sqrt{\text{Tr}(\mathbf{Q}\mathbf{Q}^H)}$. This relationship is used so as to provide similar performance as in (26) for a selected value of λ .

The first term in the objective function penalizes deviations from the measurements in interference-free regions, but less so in interference regions. This term is akin to a Mahalanobis distance. The second term penalizes large entries in the solution. Thus the entries in the interfered region are unconstrained by the first term and they are allowed to be driven to zero by the second term. The first term maintains fidelity to the measurements in the clear areas while the second term promotes sparsity. Here the γ parameter serves the same purpose as λ in (26): balancing the weight of the sparsity promoting ℓ_1 norm against the fidelity-preserving ℓ_2 norm.

3.3 Results

3.3.1 Scoring Metrics

We are interested in the ability of a radar system to identify locations of targets in a field of noise and structured interference. Therefore, we present results in terms of probability of false alarm (P_{fa}) and probability of detection (P_d) rather than the more standard (in the CS literature at least) Euclidean error norm. Additionally, we use a non-standard performance measure, detection quantile Q_d , to evaluate the consistency of performance.

Consider t targets placed arbitrarily in the observation area. The scene description vector $\mathbf{x} \in \mathbb{C}^n$ contains mostly zeros, but at those entries that correspond to the grid-points closest to the true target locations entries are nonzero values. Call these locations $\mathcal{P} = \{p_1 \dots p_t\}$. By some technique $\hat{\mathbf{x}}$ is produced which is an estimate of \mathbf{x} . For a given probability of false alarm P_{fa} a detection threshold D_{th} can be computed: let $\hat{\mathbf{x}}_s$ contain all the elements of $|\hat{\mathbf{x}}|$ sorted in increasing order, then $D_{th} = \hat{\mathbf{x}}_s(\lceil P_{fa}n \rceil)$. Finally, P_d is the fraction of elements of the estimate at the true target locations that have magnitude greater than or equal to the threshold: $P_d = \text{frac}(|\hat{\mathbf{x}}(\mathcal{P})| \geq D_{th})$.

Detection quantile is, then, the P_{fa} required to achieve a perfect P_d . Define $Q_d = \text{frac}(|\hat{\mathbf{x}}| > \min(|\hat{\mathbf{x}}(\mathcal{P})|))$. If $Q_d = 0$ it implies that the true target position was the highest-amplitude bin in the estimate.

Other possible scoring metrics include the false discovery proportion and non-discovery proportion described in [70] and the earth mover’s distance [71].

3.3.2 Signal Quality Definitions

Detection results are only interesting when presented in the context of a known signal quality. Definitions of signal-to-noise ratio (SNR), signal-to-clutter ratio (SCR), and signal-to-interference ratio (SIR) are often slippery and vary with usage. Furthermore, these values can be measured at either the input or the output of the signal processing chain. In this work, we define these functions as described in this section. Again let \mathcal{P} refer to the t true target locations in \mathbf{x} and \mathcal{A} refer to all locations in \mathbf{x} . Also, consider measurements collected as $\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{c}) + \mathbf{n}$, and $\hat{\mathbf{x}}$ produced by some technique to estimate \mathbf{x} .

Define input SNR as:

$$\frac{\max(|\mathbf{Ax}|)^2}{\text{var}(\mathbf{n})}. \quad (34)$$

Define input SCR as:

$$\frac{\max(|\mathbf{Ax}|)^2}{\text{var}(\mathbf{Ac})}. \quad (35)$$

Define input SIR as:

$$\frac{\max(|\mathbf{Ax}|)^2}{\text{var}(\mathbf{Ac} + \mathbf{n})}. \quad (36)$$

On the input side, where we use the $\max(\cdot)$ function we do so to represent the modulus of the complex waveform that is received. If the waveform returns from more than one target overlap (before pulse compression), the modulus of the overlapping segment may constructively or destructively interfere based on the relative phases of the targets.

On the output side of the receiver, pulse compression or some other estimation technique has been performed and the power for each individual target can be measured independently. In the event of multiple true targets, we take the arithmetic mean of the corresponding SIR values. Define output SIR as:

$$\frac{\text{mean}(\hat{\mathbf{x}}(\mathcal{P})^2)}{\text{var}(\hat{\mathbf{x}}(\mathcal{A} \setminus \mathcal{P}))}. \quad (37)$$

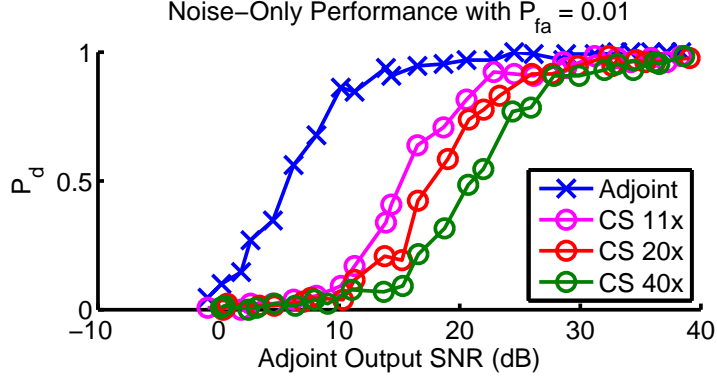


Figure 6: Detection performance of compressed sensing recovery in white noise. The performance degrades with subsampling rate; each additional octave of undersampling results raises the noise floor by a factor of two, or 3 dB.

3.3.3 Simulation Description

In the results shown here we vary the signal quality, either SNR or SCR, while holding all other parameters related to the measurement model and target model constant. We then run a large number of simulation trials in which the target location is drawn from a uniform random distribution over the observation space and the noise and clutter realizations are drawn from their respective distributions and scaled to produce the desired SNR and SCR. At each SNR point we calculate an average probability of detection P_d over all the trials. By observing the average P_d as a function of SNR or SCR one can compare the various solution methods on an equal basis.

As described above, we model a 32-element uniform linear array, 32-pulse coherent processing interval, and 128-sample range window, yielding a data cube of 131,072 samples. The waveform modeled is a swept linear-frequency modulated pulse with a bandwidth of 60 MHz and a carrier frequency of 9.7 GHz. The platform is modeled at an altitude of 10 km, and the array is oriented perpendicular to the platform motion vector with a ground grazing angle of 5.3° .

3.3.4 CS in Noise

To test the effectiveness of compressed sensing techniques in the presence of clutter we first develop results demonstrating the performance of these methods in noise only. We are

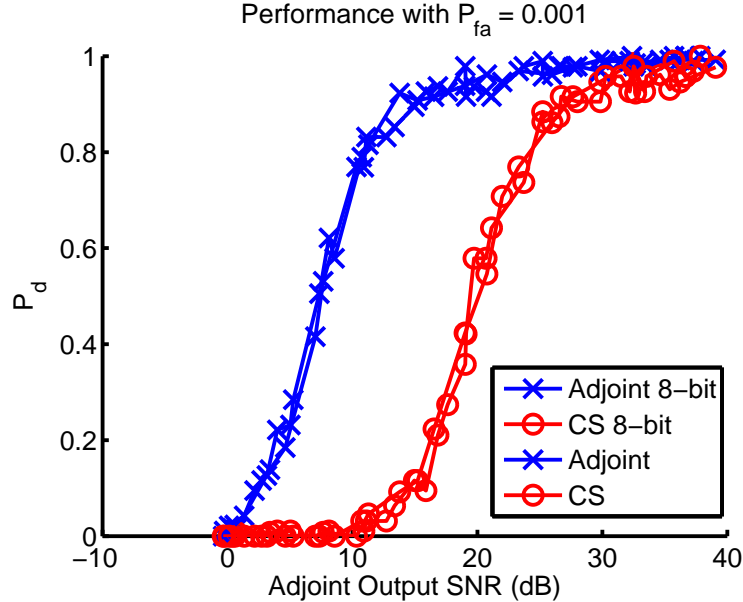


Figure 7: Detection performance as a function of SNR, varying quantization. Both CS and adjoint techniques are robust to eight bit quantization.

primarily interested in the ability to detect targets based on the compressed measurements.

These results are shown in Figure 6. Notably, this plot shows that a reduction in sampling rate (by the compressive sampler) results in a commensurate increase in noise floor as shown in [72]. This is evident in the 13 dB rightward shift from the adjoint solution to the $20\times$ undersampled CS solution. Accordingly the other solutions are shifted right by 3 dB for every additional octave ($2\times$) of undersampling. So it appears that for applications in which signal power competes with noise power, compressed sensing meets fundamental limitations. Sampling more slowly causes the noise to additively fold over the sampled spectral region competing with fixed signal energy.

To test the robustness of these results we modify the problem slightly in three different ways: by quantizing the returns at the point of measurement, by varying the probability of false alarm threshold, and by adding more targets to be detected. In each of these cases we maintained a $20\times$ USF. In Figure 7 we show that the results of these experiments are virtually unchanged by quantizing the measured returns to 8 bits. In Figure 8 we show that though performance is decreased upon demanding a higher P_{fa} , the impact on the CS estimate is not greater than the impact on a standard adjoint estimator. In Figure 9 we

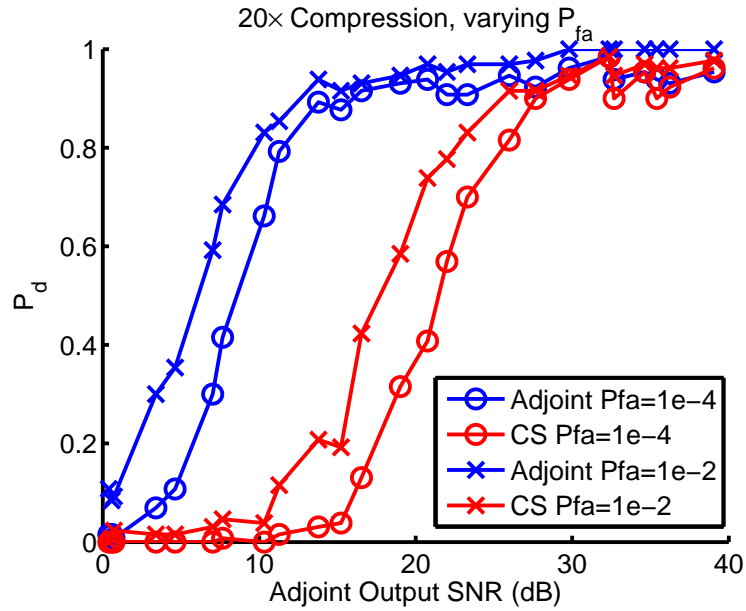


Figure 8: Detection performance as a function of SNR, varying P_{FA} . CS and adjoint techniques show similar changes in performance.

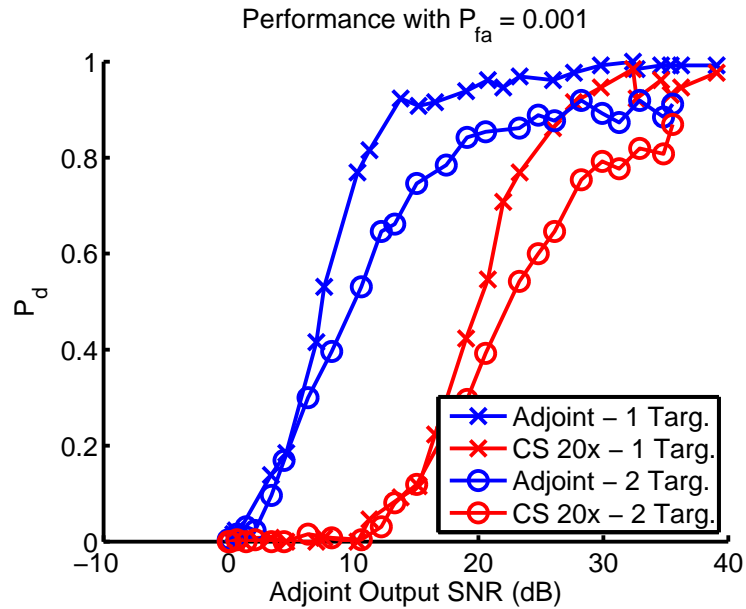


Figure 9: Detection performance as a function of SNR, varying number of point targets. By placing additional targets in the search volume the probability of detection decreases comparably in the compressed sensing and the matched filter techniques.

show that target detectability is impacted by the addition of more targets to the scene but, again, the impact is no worse than that observed in the adjoint case.

3.3.5 CS and Covariance-Aware CS in Clutter

Clutter presents a different challenge. Clutter returns are correlated with the transmitted waveform in the same way the signal returns are. The sparsity-favoring objective of the compressed sensing solution is as likely to identify strong clutter returns as it is the true target returns.

Figure 10 compares the performance of various estimation techniques on problems with clutter as the dominant interference. In this plot the shapes indicate different levels of undersampling: squares indicate full sampling, circles indicate $20\times$ undersampling, and crosses $40\times$ undersampling. The various colors indicate the estimation techniques used: red for full-sampled STAP (12), blue for CA CS (33), magenta for compressed STAP (32), black for the classic matched filter (13), green for a standard compressed sensing formulation (26), and cyan for the compressed adjoint (25).

Here CA CS shows success beyond that of the CS estimates. By using the covariance information CA CS can even surpass the performance of the fully-sampled (but covariance-ignorant) matched filter. Still, the fully sampled adaptive STAP filter remains the gold standard. Of particular interest is the $20\times$ undersampled CA CS solution that achieves nearly the same detection performance as the STAP solution. Also notable is the fact that the performance of the classically-formulated estimation methods (compressed STAP and compressed adjoint) is equal to that of the CS-formulated ones (CA CS and CS).

This performance of CA CS relative to the non-adaptive CS and relative to the fully sampled adjoint can be observed in Figure 11. This figure shows that the estimates that use the covariance information are able to correctly separate the target from the surrounding noise and clutter while the non-adaptive techniques are less-able to do so.

Aggregating this type of result over many random trials produces the results shown in Figure 12. An interesting comparison can be drawn between Figures 10 and 12. When

judging by P_d , as in Figure 10, the $40\times$ undersampled CA CS solution outperforms the fully-sampled adjoint. However when judging by Q_d , as defined in Section 3.3.1 and shown in Figure 12, the fully-sampled adjoint shows superior performance to the $40\times$ undersampled CA CS estimate. This shows that the adjoint solution often assigns a high amplitude to the correct bin, but doesn't as often put it among the *highest* bins to be counted as a detection. This, then, indicates the level of consistency of the various algorithms apart from their detection threshold behavior. The $20\times$ undersampled CA CS outperforms both of these estimates as measured by both P_d and Q_d .

However, when compared to the compressed STAP (CSTAP) methods, there is no advantage to using CA CS. It is also evident that the compressed adjoint performs as well as the CS solution in the interference limited case. This result hold over both $20\times$ and $40\times$ undersampling factors as well as for both the covariance-aware (CA CS vs. CSTAP) and covariance-unaware (CS vs. Compressed Adjoint) cases. Furthermore, this performance equivalence holds over a range of tested probability of false alarm settings.

This equivalent detection performance between CSTAP and CA CS was notable enough to bear further examination. To do so, we plotted the performance frontiers at which the estimate drops below some specified threshold. These frontiers are show in Figure 14 and 13. Notably, the detection performance of all the estimators is comparable. In contrast, the sparsity-favoring ℓ_1 -regularized least squares solution gives a better reconstruction error. Note, also, that the simple addition of the soft-thresholding operator reduces the reconstruction error significantly, and a more aggressively applied threshold could reduce it further.

3.4 Conclusion

3.4.1 Review

Structured interference, including clutter, presents a different challenge to estimation techniques than does white noise. This structure can be exploited to improve performance. While estimation of sparse vectors in the presence of white noise has been well-studied in

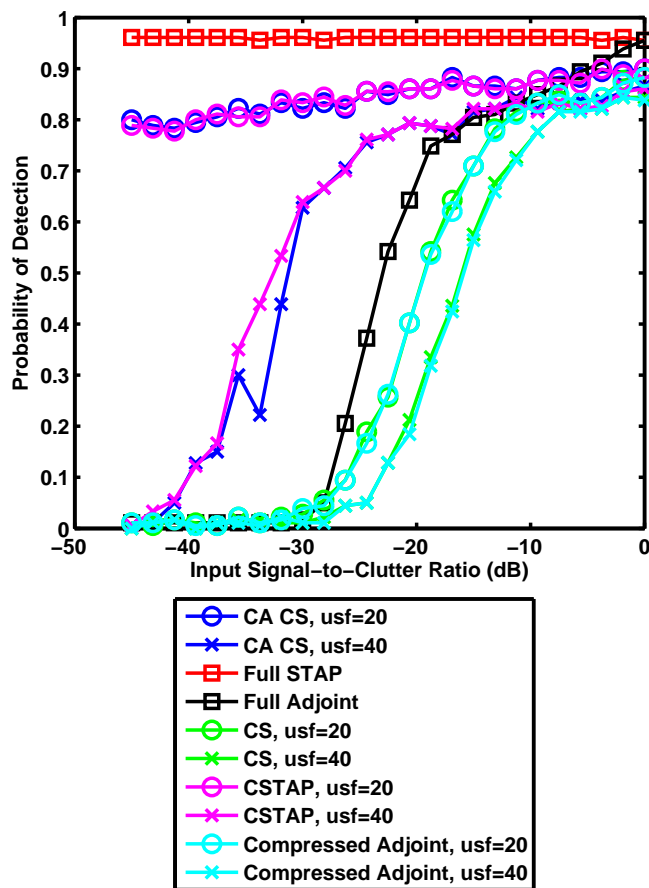


Figure 10: The CA CS method subsamples the data just as the standard CS method does, however it takes into account the covariance matrix that describes the interference structure. By doing so, it improves the probability of detection over the CS case as well as beyond the fully sampled, matched filter case that does not use the covariance information. These results are shown with a probability of false alarm of 0.005.

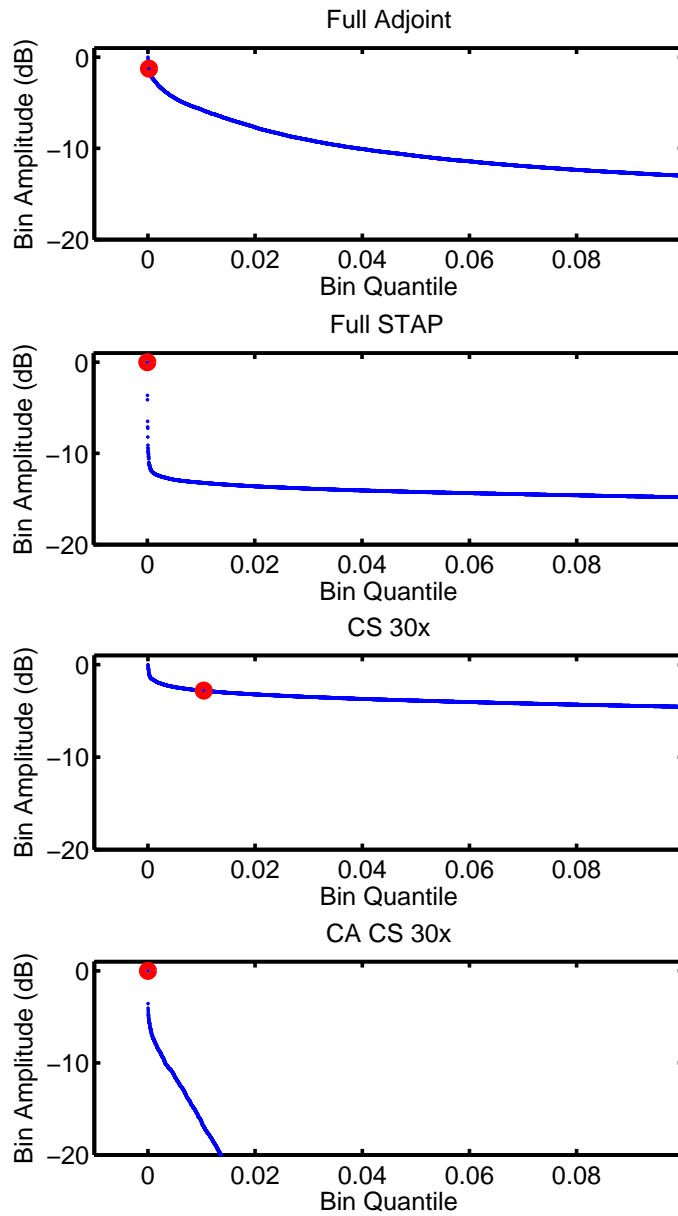


Figure 11: A comparison of various solution methods applied to a representative sample problem with input SNR of 0 dB, and input SCR of -20 dB. These plots show the relative amplitude of all the bins in the estimate produced by the identified technique. The red circle in each indicates the amplitude of the bin closest to the true target location. These results show that only the adaptive techniques, STAP and CA CS, correctly assign the highest estimated amplitude to the bin closest the target.

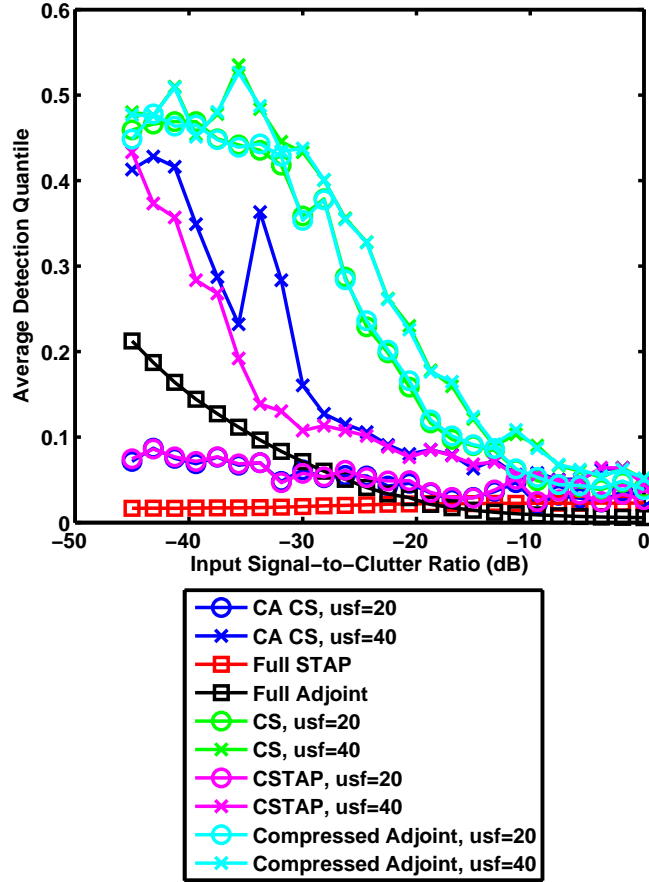


Figure 12: A related evaluation criterion, the detection quantile Q_d measures the average ranking of the true-target bin among all the bins in the estimate. A Q_d of zero is perfect. The fully-sampled STAP estimate performs better than all other techniques. Also, the $20\times$ undersampled CA CS estimate consistently achieves performance near that of the fully-sampled STAP and better than of the fully-sampled adjoint. The $40\times$ undersampled CA CS estimate is inferior to the fully-sampled adjoint, in contrast to the performance as measured by the P_d in Figure 10.

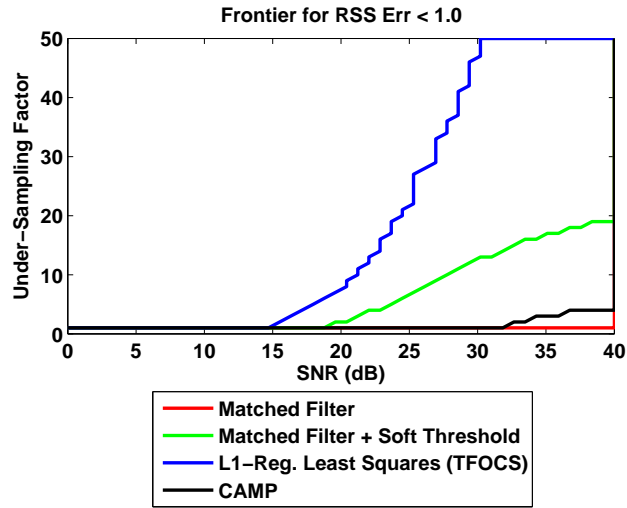


Figure 13: A comparison of reconstruction error (E_{norm}) performance frontiers shows that the sparsity-favoring techniques achieve better reconstruction error of this sparse signal.

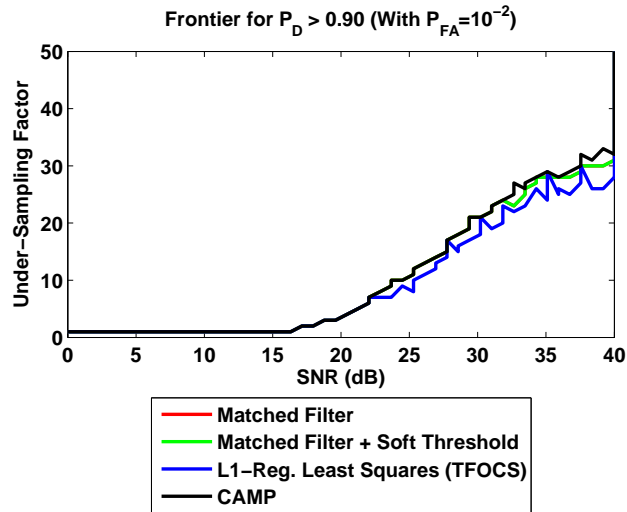


Figure 14: A comparison of detection (P_D) performance frontiers shows that all these techniques achieve approximately equal detection performance.

compressed sensing literature, estimation in structured noise has not. We show that compressed sensing techniques can be robust to structured interference and propose a method for exploiting the interference structure to improve detection statistics. However justifying the use of CS and CA CS techniques, from a probability of detection perspective, is difficult given that the compressed adjoint and compressed STAP methods perform just as well.

3.4.2 Future Work

A number of points remain to be investigated. Future work should proceed along many dimensions:

- Work remains within the CS community to clearly identify domains of noise power, clutter power, and undersampling factor, and scene sparsity over which compressed sensing methods offer better detection performance than traditional matched filter techniques on compressed data.
- The results presented here test the performance of the CA CS algorithm in the presence of clutter; by simulating jammer returns these techniques could be tested in the presence of another common structured interference type.
- The accuracy with which the interference covariance matrix can be estimated from compressed measurements remains to be explored.
- Theoretical bounds that describe the number of measurements required for a given sparsity level and interference structure would guide theorists and designers alike.
- The CA CS approach modified the ℓ_2 norm observation fidelity term; other approaches could modify the ℓ_1 sparsity-favoring term to further shape the favored solution space, but this will be left as future work as the correct modification to the norm is not obvious.
- We cited work that could be applicable to producing compressed measurements in the three sampling dimensions of this problem. But to be convincing to practitioners, these techniques must be paired with a more detailed description of the hardware

required to collect the types of measurements required in a CS receiver as well as the size, weight, and power gains that could be made by designing to a CS specification.

- Finally, the proliferation of airborne research platforms, like unmanned aerial vehicles (UAVs), offers the opportunity to gather data from flight and process using the proposed techniques.

There remains much progress to be made but the work presented here shows that CS radar systems can perform even in the presence of clutter.

CHAPTER IV

ESTIMATING INTERFERENCE COVARIANCE FROM COMPRESSED MEASUREMENTS

4.1 *Problem Formulation*

4.1.1 Sensing Model

A radar receives a post-pulse-compression data cube of dimensions $N_c \times N_s \times N_f$, for the channel, slow time, fast time sampled dimensions. The expected sampled impulse response of a point target at any location in the observation extent can be expressed as a function of the transmitted waveforms, sensing geometry, timeline, and pulse compression processing. These responses can be combined to form a sensing matrix \mathbf{S} which is a transform from the discretized target space (angle, radial velocity, and range of dimension $N_a \times N_v \times N_r$) to the sampled data cube. Together, the discretized sensing model can be described as:

$$\mathbf{y} = \mathbf{S}\mathbf{x} \tag{38}$$

where \mathbf{x} is a vector giving radar cross section of scatterers at each location in the observation extent, and \mathbf{y} contains the samples in time and space. The three dimensions of \mathbf{y} and \mathbf{x} have been vectorized.

The signal processing includes using a set of matched filters to generate an estimate of the true target profile. These angle-Doppler matched filters can be expressed as the conjugate transpose of the sensing model:

$$\hat{\mathbf{x}} = \mathbf{S}^H \mathbf{y}. \tag{39}$$

4.1.2 Interference

The receivers that make up the sensor are subject to thermal noise, quantization noise, and other Gaussian-like error sources that corrupt the measurements. The model of the signal and noise produces a familiar measurement structure

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{n}. \tag{40}$$

Because the noise is uncorrelated with the sensing model, the matched filter in (39) improves the ability to extract the signal from the noise.

Clutter components of the received signal are modeled as

$$\mathbf{y} = \mathbf{S}(\mathbf{x} + \mathbf{c}) + \mathbf{n}. \quad (41)$$

Clutter can result from land or sea surface features and the statistics of the interference depend strongly on the particularities of the terrain being surveilled. Much research has focused on describing the expected returns from clutter as a function of terrain, radar band, radar resolution, and other parameters [16, 17, 18, 19]. If the terrain exhibits significant variability as a function of range, the available training data for estimating local interference statistics is reduced. Finally, we may reorganize the received data \mathbf{y} as $\mathbf{Y} = [\mathbf{y}_1 | \dots | \mathbf{y}_{N_f}] \in \mathbb{C}^{N_c N_s \times N_r}$ for subsequent processing.

4.1.3 Sampling Compression

To introduce the compression and incoherence to the problem under consideration, as introduced in Section 3.1, we make use of a linear compression operator

$$\mathcal{C}(\cdot) : \mathbb{C}^{N_c N_s \times N_f} \rightarrow \mathbb{C}^{N_c N_s N_f / u}, \quad (42)$$

with adjoint $\mathcal{C}^*(\cdot)$, which undersamples the incident signals by a factor of u :

$$\mathbf{z} = \mathcal{C}(\mathbf{Y}). \quad (43)$$

The problem at hand is then to estimate the interference covariance matrix from the compressed measurements, \mathbf{z} , so as to optimally filter the interference and detect targets despite their relatively low return energy. The structure of the clutter and jamming, imposed by the sensing geometry provides a basis for separating the targets from interference. This estimate could then be used in a compressed estimator, as examined in [15].

4.2 Prior Work

4.2.1 Space-Time Adaptive Processing

Let the measurement data \mathbf{y} be reorganized as $\mathbf{Y} = [\mathbf{y}_1 | \dots | \mathbf{y}_{N_f}] \in \mathbb{C}^{N_{cs} \times N_r}$ where $N_{cs} = N_c N_s$. Putting aside any target energy, each of these vectors is a sample from a complex

normal distribution $\mathbf{y}_i \sim \mathcal{CN}(\mathbf{0}_{N_{cs}}, \mathbf{R}_y)$. This interference distribution has both clutter and noise components: the noise contributes a diagonal component to the covariance matrix, while the clutter contributes a component with low- or approximately-low-rank structure as a result of the sensing geometry and timeline. Thus

$$\mathbf{R}_y = \mathbf{R}_n + \mathbf{R}_c = \nu^2 \mathbf{I}_{N_{cs}} + \mathbf{V}\mathbf{D}\mathbf{V}^H \quad (44)$$

where ν^2 is the measurement noise variance, $\mathbf{V} \in \mathbb{C}^{N_{cs} \times k}$ with structure determined by the characteristics of the observed clutter, and \mathbf{D} is diagonal with entries $\sigma_1^2, \dots, \sigma_k^2$. Here we model the clutter as having rank $k \leq N_{cs}$, though the degree to which this low-rank model holds true will be examined in greater detail in this chapter. Clutter is stronger than noise (necessitating adaptive filtering), though this may be quite variable. The sample covariance matrix used in STAP is then defined as:

$$\hat{\mathbf{R}}_y = N_f^{-1} \mathbf{Y}\mathbf{Y}^H. \quad (45)$$

As more samples from this distribution are collected, $N_f/N_{cs} \rightarrow \infty$, and this estimate will converge to the true covariance \mathbf{R}_y . This estimate of the interference statistics can be inverted to form an SINR-maximizing linear filter [24] which leads to an estimate of the target scene for range bin i as

$$\hat{\mathbf{x}}_i = \kappa \mathbf{S}^H \hat{\mathbf{R}}^{-1} \mathbf{y}_i. \quad (46)$$

where the scaling factor $\kappa = (\mathbf{S}^H \hat{\mathbf{R}}^{-1} \mathbf{S})^{-1}$.

The quantity of training data available to estimate the covariance, as in (45), is often limited. The statistics of the interference must be estimated from data in adjacent or nearby range bins to avoid using non-representative statistics in non-homogeneous environments. The Reed-Mallet-Brennan rule states that a stable estimate can be made from a set of training data with twice as many snapshots as degrees of freedom in the covariance [26]. Various approaches that have been applied in attempts to maximize performance with limited data are discussed next.

4.2.2 Diagonal Loading

Diagonal loading is a commonly used technique to improve sample covariance accuracy and stability [73]. It adds a diagonal (usually a scaled identity) component to regularize the covariance matrix estimate before inversion. After introducing the diagonal loading term, the sample covariance matrix estimate becomes:

$$\hat{\mathbf{R}}_{full} = N_f^{-1} \mathbf{Y} \mathbf{Y}^H + \delta \mathbf{I}_{N_{cs}} \quad (47)$$

which is identified as the “full” estimate because it uses the full (Nyquist-rate) set of data and also draws a distinction with other estimates to be introduced. Diagonal loading has the effect of reducing the matrix condition number if the matrix is badly conditioned. It also places a limit on the filter null depth. The optimal loading level depends on the type of covariance estimate, as well as the high sensitivity to clutter-to-noise ratio and even the temporal sampling rate. Optimal selection is largely based on heuristic approaches and empirical results. For this work we use a diagonal loading factor of 10^{-6} which will appropriately regularize the inverse without degrading performance by overwhelming the observed sample structure.

4.2.3 Reduced Rank STAP

As described in Section 2.4.3, reduced-rank STAP offers improved convergence of estimates of the interference covariance matrix in cases with few training snapshots available. This is accomplished by reducing the number of effective degrees of freedom by applying assumed prior knowledge that the clutter has a compact eigen-spectrum. Improvement in convergence by reducing the number of degrees of freedom comes at the cost of some resolution of the clutter null.

4.2.4 CS and Clutter

The central question of this work is how well samples produced from a compression operation \mathcal{C} , as in (43), can be used to form an estimate of the covariance matrix. The simplest approach is compressed sample matrix inversion (SMI) in which the adjoint of the compression operator is used to bring the samples back to the full ambient signal space before

forming the covariance estimate:

$$\hat{\mathbf{R}}_{comp} = N_f^{-1} \mathcal{C}^*(\mathbf{z}) \mathcal{C}^*(\mathbf{z})^H + \delta \mathbf{I}_{N_{cs}}. \quad (48)$$

As in (47) a diagonal loading term is introduced to regularize the matrix inverse. This technique will be compared to the proposed approach described next in Section 4.3.

4.3 Proposed Technique

This section introduces a singular value thresholding technique, which will subsequently be compared to the covariance estimators described up to this point. To review, the previously-identified covariance estimates are:

- $\hat{\mathbf{R}}_{none} = \mathbf{I}_{N_{cs}}$
- $\hat{\mathbf{R}}_{full} = \mathbf{Y} \mathbf{Y}^H + \delta \mathbf{I}_{N_{cs}}$
- $\hat{\mathbf{R}}_{comp} = \mathcal{C}^*(\mathbf{Z}) \mathcal{C}^*(\mathbf{Z})^H + \delta \mathbf{I}_{N_{cs}}$

Like the compressed sample matrix covariance matrix, $\hat{\mathbf{R}}_{comp}$, the proposed technique will use compressed data to generate an interference covariance estimate.

4.3.1 Concept

The structure of ground clutter in airborne radar data provides the basis for using of adaptive filtering. This structure includes concentration of the singular values of the clutter samples. (The singular values of the sample matrix \mathbf{Y} are the square root of the eigenvalues of the sample covariance matrix.) This fact has been used to enable the reduced-dimension and reduced-rank STAP approaches [12]. It is also clearly related to the more recent literature of low-rank matrix completion. The proposed approach views the covariance matrix recovery problem from the perspective of low-rank matrix completion to examine whether using the low-rank assumption in the recovery procedure allows for improved matrix estimation.

4.3.2 Singular Value Thresholding (SVT) Algorithm

The algorithm for covariance estimation based on matrix completion is detailed in Algorithm 1. The inputs to the algorithm are the compressed measurements \mathbf{z} and the compression

operator $\mathcal{C}(\cdot)$. The output is an estimate of the interference covariance matrix, $\hat{\mathbf{R}}_{svt}$. The observation is iteratively modified to find a good fit with both the measurements and a low-rank condition. The iteration estimate is projected onto the larger space, where the low-rank condition is enforced using the $\text{shrink}(\cdot)$ operator. This result is compressed to give a residual with the measurements, which is used to update the estimate. By going back and forth between fidelity to the measurements and adherence to the low-rank assumption, a balance between the two can often be found.

```

input :  $\mathcal{C}, \mathbf{z}$ 
output:  $\hat{\mathbf{R}}_{svt}$ 
 $\mathbf{z}_0 \leftarrow \mathbf{z}$ 
 $r_{min} \leftarrow \infty$ 
 $k \leftarrow 0$ 
while stopping criterion not met do
     $k \leftarrow k + 1$ 
    /* compute estimate */
     $\tilde{\mathbf{Y}} \leftarrow \mathcal{C}^*(\mathbf{z}_{k-1})$ 
     $\mathbf{Y} \leftarrow \text{shrink}(\tilde{\mathbf{Y}}; \tau)$ 
     $\mathbf{r} \leftarrow \mathbf{z}_0 - \mathcal{C}(\mathbf{Y})$ 
     $\mathbf{z}_k \leftarrow \mathbf{z}_{k-1} + \delta \mathbf{r}$ 
    /* save best estimate */
     $r \leftarrow \frac{\|\mathbf{r}\|}{\|\mathbf{z}_0\|}$ 
    if  $r < r_{min}$  then
         $r_{min} \leftarrow r$ 
         $\hat{\mathbf{R}}_{svt} \leftarrow \mathbf{Y} \mathbf{Y}^H$ 
    end
end

```

Algorithm 1: Covariance Estimation via SVT

There are three aspects of this iteration that must be specified in somewhat greater detail: the threshold value τ for $\text{shrink}(\tilde{\mathbf{Y}}; \tau)$, the step size δ , and the stopping criterion. Subsequent subsections detail the approach taken for each of these.

4.3.2.1 Singular Value Threshold (τ)

This parameter balances fidelity to the observed measurements against the low-rank assumption. The threshold τ should be selected so as to preserve the non-noise interference subspace, but it should not be too large, as that would waste degrees of freedom estimating random noise.

Estimating the rank of a matrix observed with noise is a fundamental problem in array processing, communications, statistics, financial modeling, theoretical physics, and many other fields. A rich body of literature exists in the field of random matrix theory on the distribution of the eigenspectrum of Laguerre unitary ensembles, also known as Wishart random matrices [74, 75, 76], of which the sample covariance matrix is a member. One of the foundational results in this area of study is the Marčenko-Pastur law [77].

The Marčenko-Pastur function provides the expected distribution of the eigenvalues of Wishart matrices:

$$dy(x) = \frac{1}{2\pi\nu^2} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x} \quad (49)$$

where ν^2 is the noise variance, $\lambda = \frac{N_{cs}}{N_f}$, and $\lambda_{\pm} = \nu^2(1 \pm \sqrt{\lambda})^2$. So even though the “true” eigenvalues are all equal (set by the noise level), the expected sample eigenvalues can vary greatly (depending on the number of training samples relative to the cardinality of each observation). This can be seen in Figure 43 of Section 5.2.2, which shows the expected distribution and the first 25 eigenvalues from numerical simulations of the noise-only case, and noise with barely perceptible clutter. This notable random matrix theory result (56) holds in the limiting case with fixed N_{cs}/N_f and $N_f \rightarrow \infty$, but convergence to these limiting results is rapid and they are reasonably accurate for problems of the size considered here, i.e., with dimensions in the hundreds or thousands.

In the uncompressed case, any eigenvalues above the maximum expected noise-only eigenvalues can be safely assumed to consist of structured interference and must be included in the low-rank estimate; this comparison is straightforward. In the compressed case, however, it is more complicated. The nonzero mutual coherence of the compression operator spreads the concentrated clutter energy over the full interference subspace. In doing so, the eigenvalues are muddled: the eigenvalues of the clutter dimensions are reduced and the eigenvalues of the noise dimensions are raised. Thus a direct comparison between the observed and expected eigenvalues is not possible. Instead, we estimate the rank of the interference in Algorithm 2 with the observed eigenvalues \mathbf{a} , the Marčenko-Pastur expected eigenvalues \mathbf{b} (as in (56)), and the isolation η . Based on the estimated rank, the singular value threshold is chosen so that all eigenvalues after the estimated rank are extinguished

and the remaining rank is the same as the estimate.

To do so, this approach tries to reverse the energy spreading associated with the compression operation by sequentially calculating the energy that would be spread by the largest eigenvalue (\mathbf{a}'_1/η) and subtracting that from the smaller eigenvalues until the remaining smaller eigenvalues fall below the Marčenko-Pastur distribution. This tends to slightly overestimate the rank, as the requirement that all eigenvalues fall below the Marčenko-Pastur distribution is quite stringent. Because there is greater risk in slightly underestimating the clutter rank than in overestimating, this balance is warranted. The procedure is detailed in Algorithm 2, and the results of applying the algorithm are illustrated in Figure 15. The successive modification of \mathbf{a}' yields the estimate of the true rank of the interference from the compressed covariance information. Figure 16 shows how this approach performs over a broader parametric space. As the clutter eigenvalues bleed into the noise eigenvalues, either because of low CNR or because of the wide rank of the clutter, performance starts to degrade.

```

input :  $\tilde{\mathbf{Y}}, \mathcal{C}, N_{cs}, N_f, u$ 
output:  $\tau$ 
 $\hat{r} \leftarrow 0$ 
 $\mathbf{a} \leftarrow \text{eig}(\tilde{\mathbf{Y}})$ 
 $\mathbf{a}' \leftarrow \mathbf{a}$ 
 $\mathbf{b} \leftarrow \text{Marcenko-Pastur}(N_{cs}, N_f)$ 
 $\mathbf{G} \leftarrow \mathcal{C}^*(\mathcal{C}(\mathbf{I}_{N_{cs}}))$ 
 $\eta \leftarrow \sqrt{\frac{\pi}{2} \frac{N_{cs}}{u}}$ 
while any( $\mathbf{a}' \succ \mathbf{b}$ ) do
     $\hat{r} \leftarrow \hat{r} + 1$ 
     $\mathbf{a}' \leftarrow \mathbf{a}'_{2\dots n} - \mathbf{a}'_1/\eta$ 
     $\mathbf{b} \leftarrow \mathbf{b}_{2\dots n}$ 
end
 $\tau \leftarrow \mathbf{a}'_{\hat{r}+1}$ 

```

Algorithm 2: SVT Rank Determination and Threshold Estimation

4.3.2.2 Residual Step Size

Picking a larger step size can speed convergence in some cases but can lead to overshoot and instability. The approach taken in the SVT algorithm (Algorithm 2) is to update the step size by comparing the residual on one iteration to the that from the previous. If the

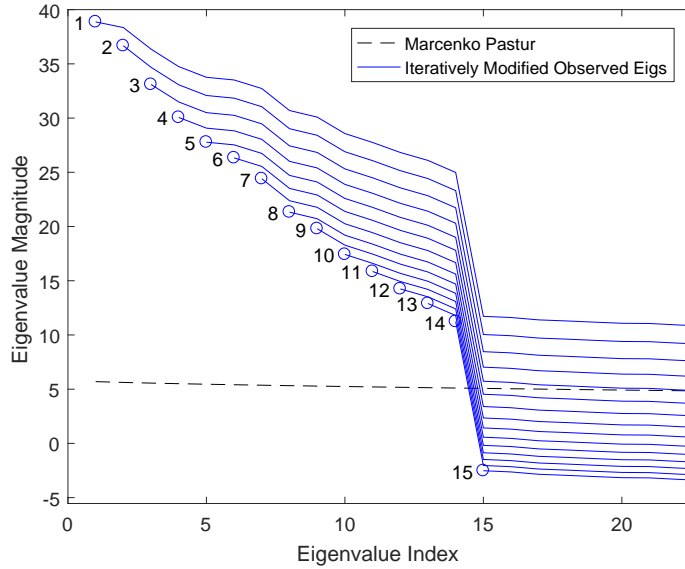


Figure 15: Iteratively modified compressed sample covariance eigenvalues (shown in blue). As Algorithm 2 modifies the eigenvalues they eventually fall below the noise-only prediction line, producing the rank estimate. In this case the estimate is accurate, as the true interference rank is 14.

angle between these two residuals is small (indicating consistency in direction of update), then increment the step size. However, if the angle is large (indicating overshooting or other large changes in the direction of the update), reduce it. With the residual at iteration k defined as $\mathbf{r}_k = \mathbf{z}_0 - \mathcal{C}(\mathbf{Y}_k)$, the angle between successive residuals is

$$\theta_k = \arccos\left(\frac{\mathbf{r}_k^H \mathbf{r}_{k-1}}{\|\mathbf{r}_k\|_2 \|\mathbf{r}_{k-1}\|_2}\right) \quad (50)$$

and the step-size update equation is

$$\delta_k = \frac{\delta_{k-1}}{\frac{19}{2} \left(\frac{\theta_k}{\pi}\right)^2 + \frac{1}{2}}. \quad (51)$$

This relationship is illustrated in Figure 17.

4.3.2.3 Stopping Criterion

The only information available to the algorithm is the path of the residual and the rank of the solution. But empirically, the accuracy of the estimate may continue to improve even when the residual has apparently stalled. In order to bias the algorithm slightly

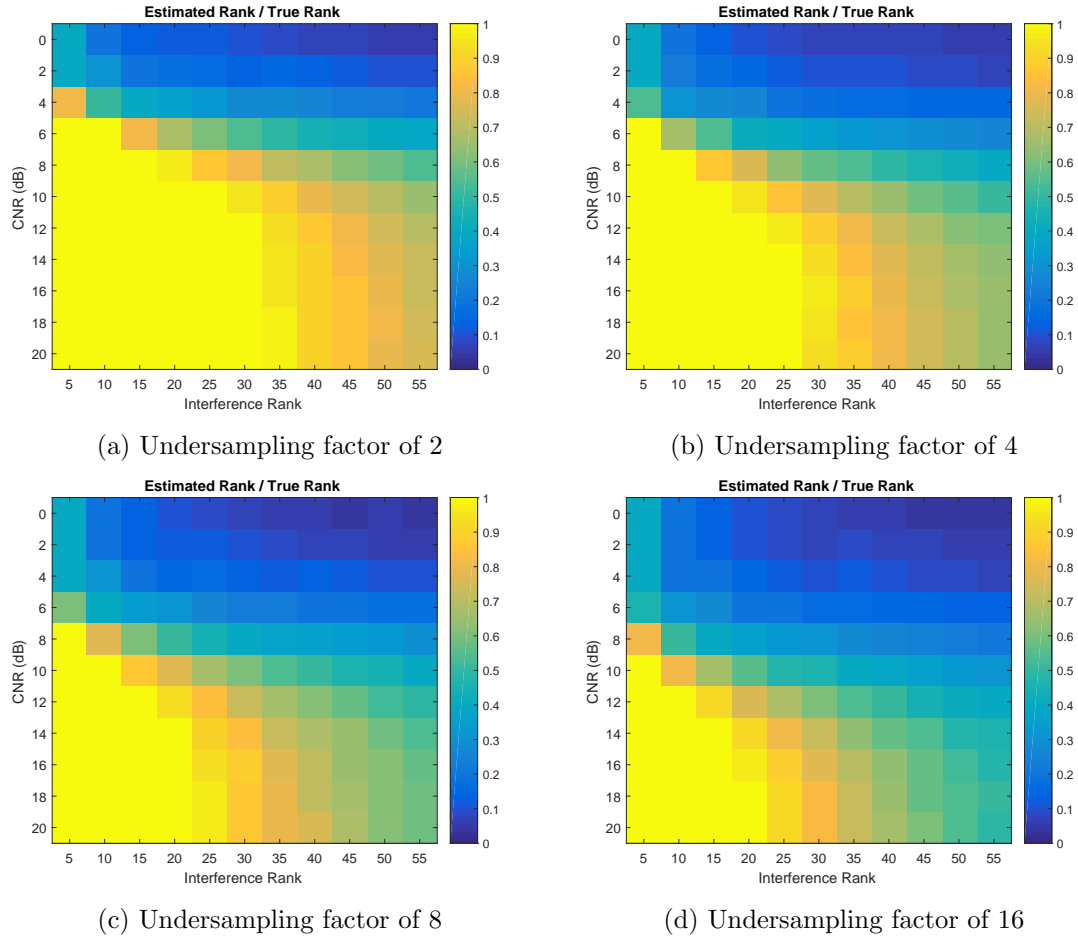


Figure 16: Performance of SVT rank estimation technique over a parametric evaluation space for different undersampling factors. For these cases, the interference rank structure is structure 1 (simple step function), the number of channels is 8, number of pulses is 128, and number of fast-time samples is 512.

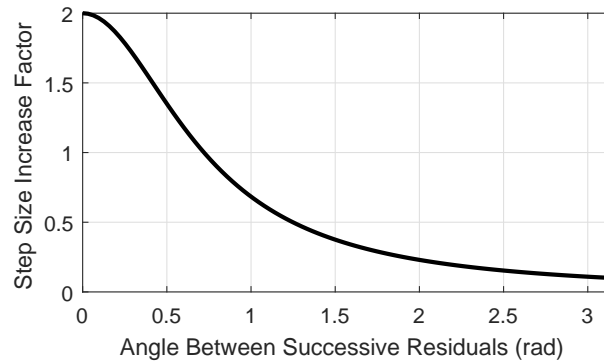


Figure 17: The step size update function which uses a measure of consistency between the calculated residual and the previous residual. If the steps are in the same direction, the step size is increased to speed convergence. If the direction between steps changes significantly, the step size is reduced to improve stability.

toward continuing execution we compute an adjusted residual $r'_k = r_k/(1 + \varepsilon)^k$, with the value $\varepsilon = 0.02$ used in the following experiments. If the maximum number of iterations is reached, or if the adjusted residual fails to improve in m of the last n iterations, the stopping criterion is met. We use $m = 8$ and $n = 11$.

4.4 Results

We evaluate the effectiveness of three covariance matrices: the sample matrix inverse (SMI) from full data $\hat{\mathbf{R}}_{full}$, SMI from compressed data $\hat{\mathbf{R}}_{comp}$, and the SVT estimate from compressed data $\hat{\mathbf{R}}_{svt}$. For bounding comparison purposes we also include the estimates resulting from using the true covariance matrix and from no covariance matrix (identity).

4.4.1 Evaluation Criteria

There are a number of potential metrics to evaluate the accuracy of a covariance matrix estimate. In the application under consideration, we are less interested in the accuracy of the estimate itself (as compared to the true covariance matrix) than we are in the ability of the estimate to perform some function, in this case filtering of the interference and detection of targets. It is for this reason that we propose to evaluate the covariance matrix estimates using detectability metrics for targets embedded in the interference after filtering using the inverse covariance matrix. The primary metrics for target detectability that we consider include probability of detection, probability of false alarm, and signal-to-interference-plus-noise ratio (SINR). Finally, we evaluate the matrices using detection from uncompressed data; this is done so that all covariance estimates (those resulting from compressed and uncompressed data) can be evaluated using a common set of criteria.

4.4.2 Synthetic Low-Rank Covariance Matrices

We test four different spectrally structured synthetic data matrices to better understand the performance of the SVT algorithm under various conditions. For an input rank width r :

1. The first decay function has r equal eigenvalues.

Table 1: Parameters used for synthetic structured interference experiments

Parameter	Value
Range Bins	100 – 500
Pulses	128
Spatial Channels	8
Clutter Spectral Decay Width	5 – 20
Clutter to Noise Ratio	0 – 40 dB
Signal to Interference Ratio	0 dB

2. The second decay function has r eigenvalues with unity magnitude, r with magnitude $1/4$, and r with magnitude $1/16$.

3. The third decay function has eigenvalues decaying, with the i -th eigenvalue having magnitude

$$\sigma_i = \frac{\exp(-(35i/r)^{7/10})}{\exp(-(35/r)^{7/10})} \quad (52)$$

4. The fourth decay function has eigenvalues decaying, with the i -th eigenvalue having magnitude

$$\sigma_i = \begin{cases} 1, & i \leq r \\ \frac{\exp(-(35(i-r)/r)^{7/10})}{\exp(-(35/r)^{7/10})}, & \text{otherwise} \end{cases} \quad (53)$$

These four synthetic interference structures are illustrated in Figure 18.

This set of spectral structures includes two with exponential decay in the true eigenvalues of the interference. This condition is included in an attempt to represent true sensing problems in which the interference does not vanish. Selected parameters used for these simulations are provided in Table 2.

To build up to the performance summary statistics we first include some detail for a particular example of generating these estimates and using them to filter interference. In this case, the true interference structure is structure 4, shown in Figure 18. The number of snapshots is 500, the number of spatial channels is 8, the number of pulses is 128, the CNR is 15 dB, and the rank width parameter is 10. The various covariance estimates exhibit varying spectral structure and are shown in Figure 19. As described in Section 4.3, the SVT estimate is generated by iteration. Figure 20 shows how the estimate evolves over those iterations, with the accuracy improving as a function of iteration. Finally, Figure 21

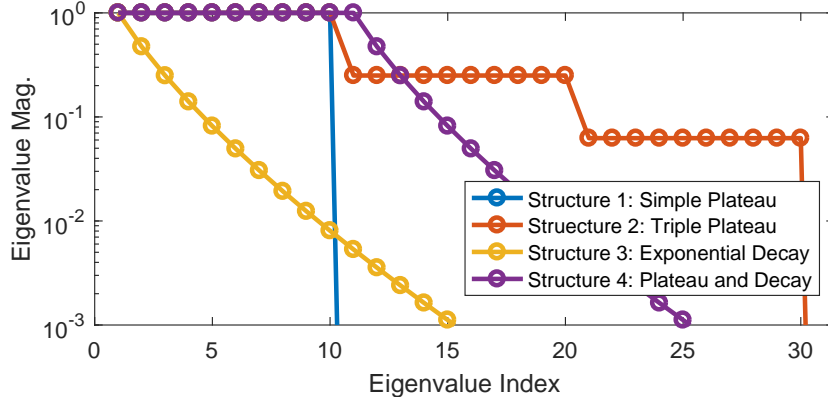


Figure 18: Synthetic structured interference eigenvalue decay functions. These functions are shown with width parameter of 10. For other values, of this parameter, the spectral structure is stretched or compressed proportionally.

shows the detection statistics after using the estimates to filter the interference with targets placed in random locations in the sampled space. In this case, the SVT estimate provides better, more consistent performance than the compressed SMI.

To evaluate over a broader problem space we conducted a parametric sweep over several relevant problem input variables: under-sampling factor, clutter-to-noise ratio (CNR), number of range samples (snapshots), and clutter eigenvalue decay function. These results are all generated using $N_a = 8$ spatial channels and $N_s = 128$ pulses. The results of this evaluation are shown the the following figures.

Figure 22 shows how performance varies as the under-sampling factor (defined in (42)) varies. It is evident that for low and moderate under-sampling factors the SVT estimator is able to perform as well as the full-data SMI, but under these same conditions the performance of compressed SMI falls off sharply. If the experiment is limited to favorable interference conditions (low rank, simple structure), the SVT estimate performs very well, out to an undersampling factor of 20, as shown in Figure 23.

Figure 24 shows how performance varies as the clutter-to-noise ratio (CNR) varies. For low CNR, when white noise is the dominant interference source, performance cannot be significantly improved by attempting to estimate a covariance. In these cases the compressed SMI harms performance (versus using no covariance matrix at all). In high clutter-to-noise

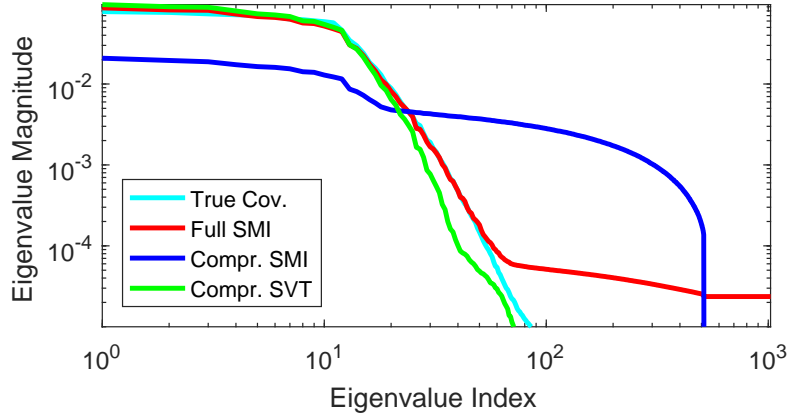


Figure 19: The spectral decay structures of the various covariance estimates of the synthetic structured interference. The SVT estimate more closely matches the truth than does the compressed SMI which uses the same data as input.

ratio cases, the covariance structure is easy to identify and the clutter can easily be nulled. In all cases the SVT estimate improves performance over the compressed SMI.

Figure 25 shows how performance varies as the number of pre-compression range bins varies. It shows that the SVT estimate is best able to take advantage of a longer observation interval to improve the estimation accuracy, even if those range bins will be compressed to reduce the number samples, a larger ambient dimension improves the estimation performance. This is somewhat challenging to at least one motivation for reducing the number of samples: non-stationarity interference forcing a reduced range sample space. But for cases in which acquiring the samples themselves is a greater challenge than inherent interference constraints, this result shows that the SVT estimator could provide improved performance.

Figure 26 shows how performance varies over the different clutter eigenspectra shown in Figure 18. Figure 27 shows how performance varies as a function of the rank of the clutter (or the rate of decay in the clutter eigenvalue magnitudes, for those with exponential decay).

4.4.3 High-Fidelity Simulated Clutter

To test against more realistic data we simulate measurements in the Adaptive Sensor Prototyping ENvironment (ASPENTM) tool developed at Georgia Tech Research Institute’s Sensors and Electromagnetic Applications Laboratory. ASPEN is a high fidelity clutter

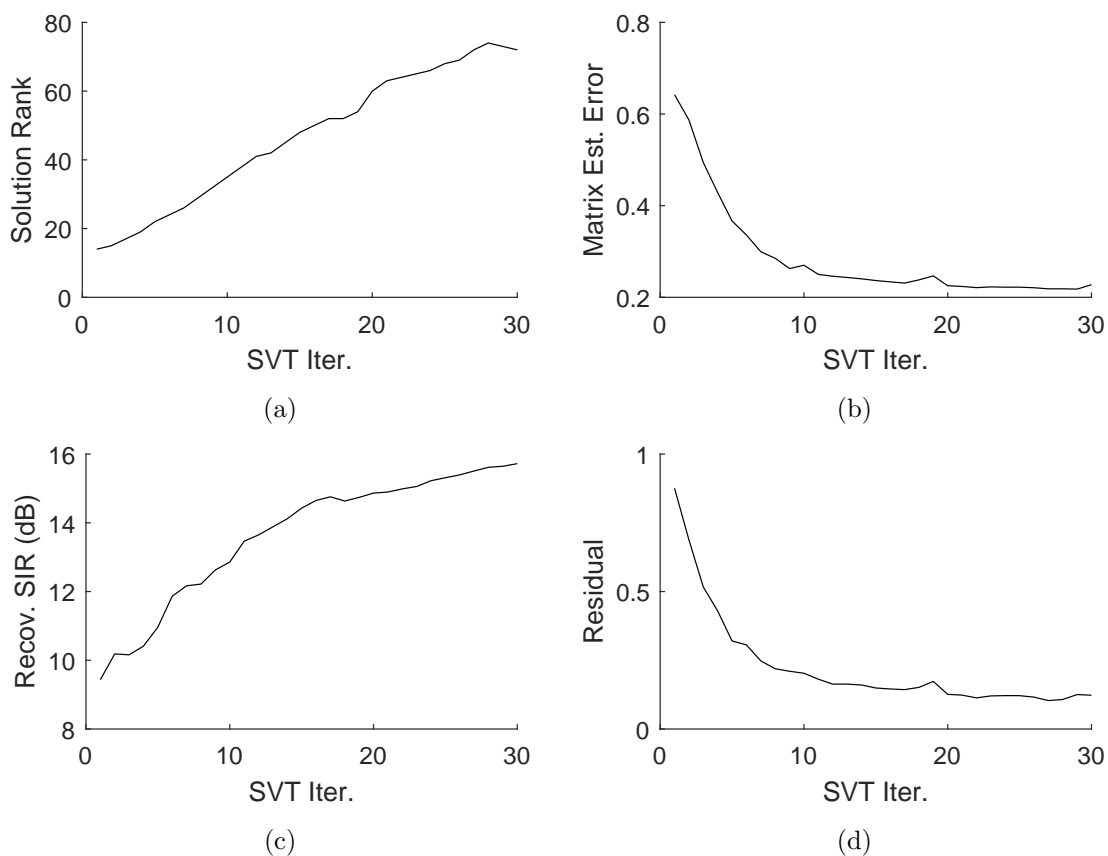


Figure 20: Example iteration results for the SVT algorithm operating on a synthetic structured interference. (a) Estimate of the interference rank increases as the SVT iteration progresses, which confirms that the resultant estimate of the covariance matrix improves. Improvement shown for three metrics: (b) direct matrix estimation error, (c) signal-to-interference ratio for a point target embedded in the interference (using the estimate as a filter), and (d) optimization residual.

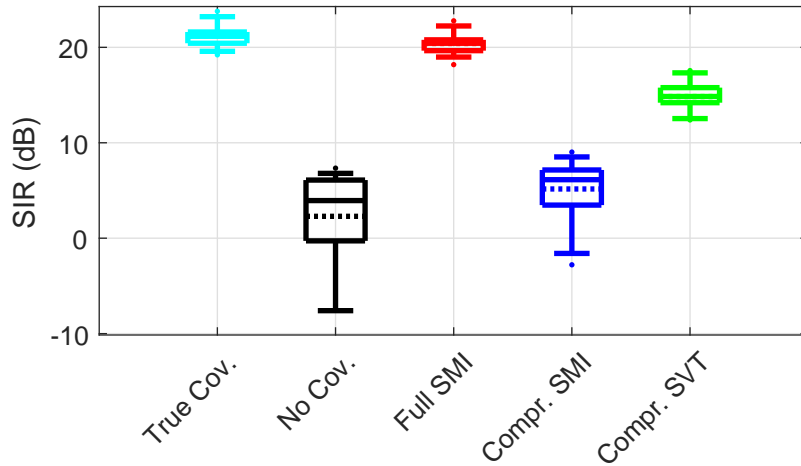


Figure 21: Example target recovery results for the various covariance estimates on synthetic structured interference. After making a set of estimates of the covariance matrix, a number of targets are embedded in the interference, each of the resultant adaptive filters applied, and the signal-to-interference ratio recorded. The statistics of one such run are shown in a box and whisker plot. The box contains the 25%-75% quantiles of the results while the whiskers contain the 5%-95% quantiles.

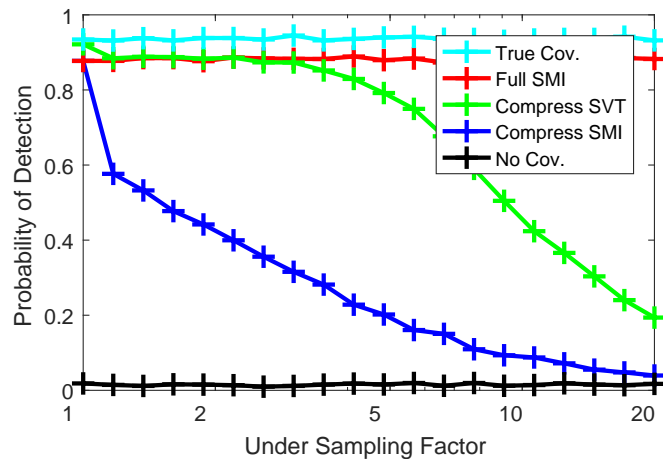


Figure 22: Average probability of detection as a function of data under-sampling factor for cases of synthetic structured interference. As the under-sampling factor increases and less data is available for the two compressed estimates, the accuracy of those estimates degrades. Notably, the SVT estimate maintains much better performance than compressed SMI as the USF increases.

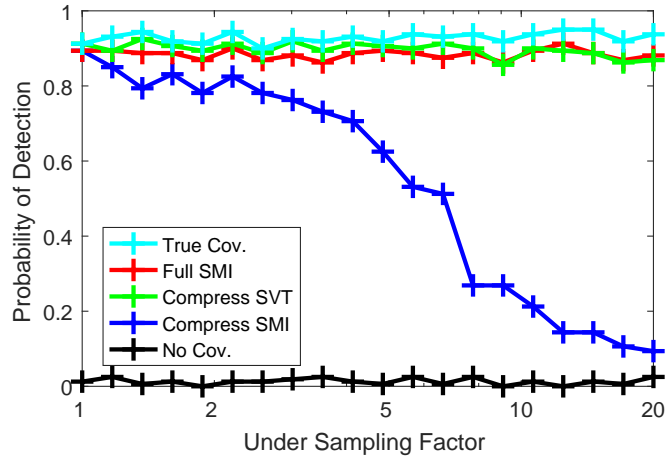


Figure 23: Average probability of detection as a function of data under-sampling factor for cases with favorable interference structure. The SVT estimate based on the compressed data performs better than the diagonally loaded estimate based on the full data and nearly as well as the true covariance matrix. The interference rank is 5 and the structure is simple (structure ID = 1).

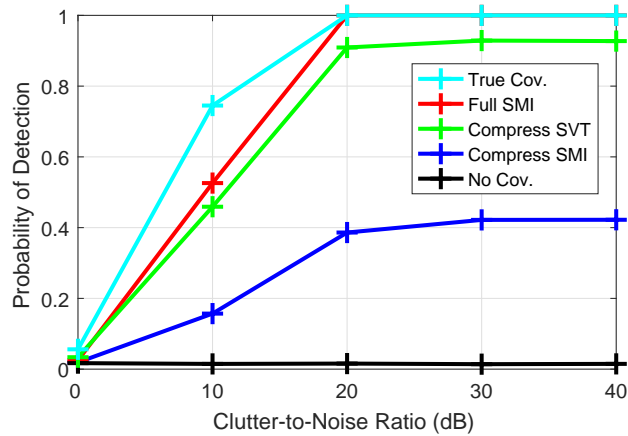


Figure 24: Average probability of detection as a function of CNR for cases of synthetic structured interference. At low clutter-to-noise ratios, noise is the dominant interference source. But as clutter becomes dominant, the structure of the interference can be estimated and used to develop a usable adaptive filter. Total signal-to-interference ratio is constant.

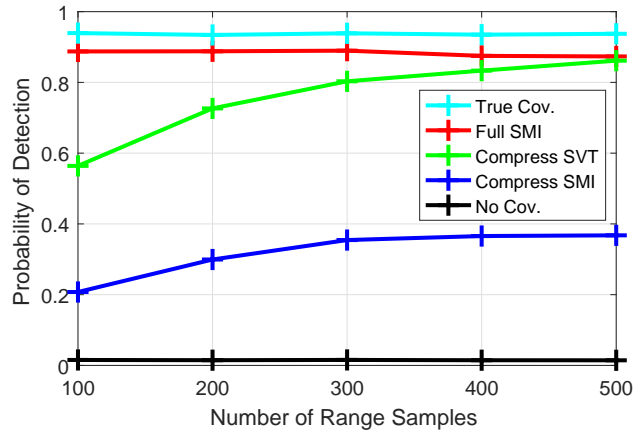


Figure 25: Average probability of detection as a function of number of range samples for cases of synthetic structured interference. As the number of range samples (snapshots) increases the two compressed estimates, which use compressed data (compressed snapshots), both improve in estimation accuracy.

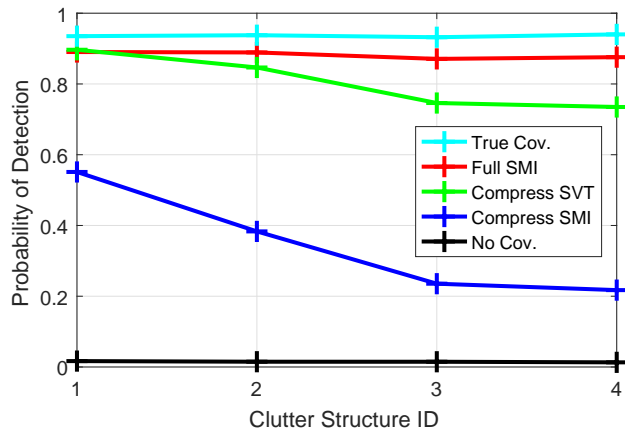


Figure 26: Average probability of detection as a function of interference structure for cases of synthetic structured interference. The performance of various estimators is tested with four different clutter structures, having the four eigenspectra illustrated in Figure 18. The performance of the estimators varies as a function of the spectra, with lower rank clutter being easier to filter out and higher rank clutter being more difficult.

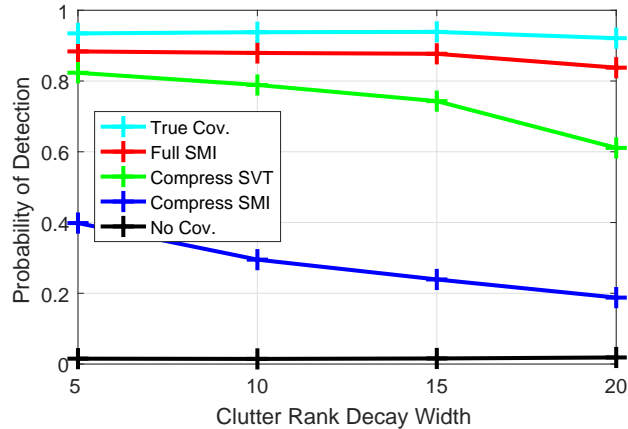


Figure 27: Average probability of detection as a function of interference rank for cases of synthetic structured interference. As the clutter becomes higher rank it is more difficult to filter out.

modeling simulation that supports flexible sensor definitions. We model an 8-channel array, with a 64-pulse coherent processing interval, and 512 range samples per pulse.

These parameters yield a data-cube of 262,144 samples, and a 512×512 full-size covariance matrix. As such, the Reed Mallet Brennan rule [26] cannot be satisfied even if all the data outside the range cell under test are used as training data.

The spectral structure of the clutter in this dataset (Figure 28) shows that the low-rank approximation holds, but not exactly. The first 40 eigenvalues (10% of the total degrees of freedom) contain 98% of the energy of the true interference. This example is instructive, but the precise nature of the spectral decay in an operational system will be a function of transmitted beamwidth, waveform bandwidth, local topography, and ground cover.

As in the synthetic interference cases, the SVT iteratively improves the estimate of the interference. The improvement over iteration is shown in Figure 29. Figure 30 shows covariance estimate eigenvalue decay for the various estimators and for the true interference. The deviation of the compressed SMI from the true interference is evident, with energy spilling to much higher rank. Figure 31 shows the angle-Doppler images generated after filtering with the various covariance estimates. The SVT estimate is shown to provide a $4\times$ improvement in SINR over the compressed SMI. Figure 32 shows the same angle-Doppler images but for a lower-rank estimation problem. In this case the interference has the same

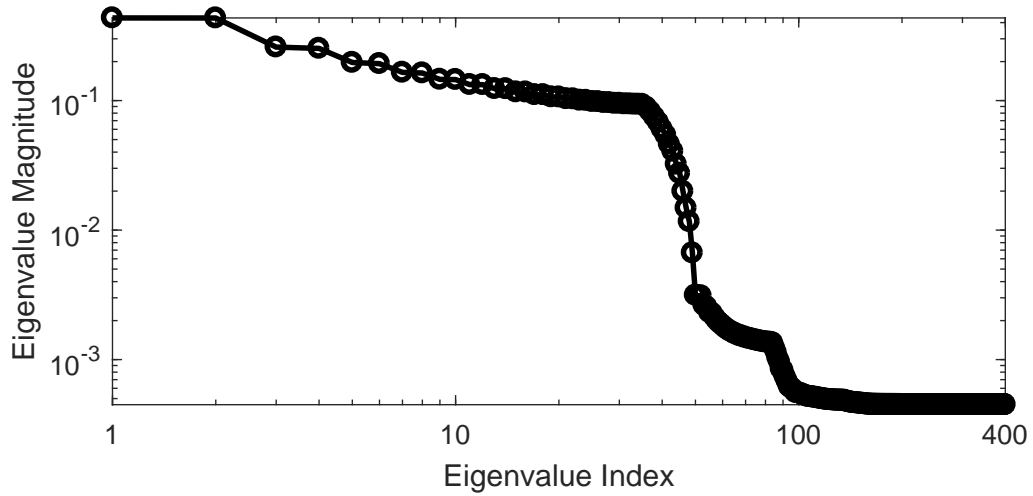


Figure 28: The true eigenspectrum of the ASPEN-generated clutter. It exhibits concentration of spectral structure but is not strictly low-rank.

magnitude as Figure 31 but is $3\times$ as compact. In this case the SVT demonstrates an even greater improvement over the compressed SMI, $10\times$ improvement in SINR.

4.4.4 Flight Test Data

We seek to validate the effectiveness of the proposed approach on live data, subject to sensor calibration irregularities, signal processing losses, and clutter inhomogeneity. To do so we processed sixteen data logs collected as part of the Multi-Channel Airborne Radar Measurements (MCARM) [78, 79] data. This data collection program was conducted by the United States Air Force Research Laboratory and Northrop-Grumman using a 24-channel L-band radar operating in coherent pulse-Doppler mode at various pulse-repetition frequencies. These 24 receiver channels recorded sum and difference data as well as the signals over a 2×11 planar array. The transmitted waveforms were $50 \mu\text{s}$, 1 MHz LFM chirps, with 630 fast-time samples per pulse. Each CPI was composed of 128 pulses with a PRF of 1,984 MHz. These parameters provide a good baseline for comparison for various processing techniques.

For the purposes of this work we will only evaluate one spatial dimension (azimuth). This data was primarily collected over rural Maryland. As such it presents some variety of terrain

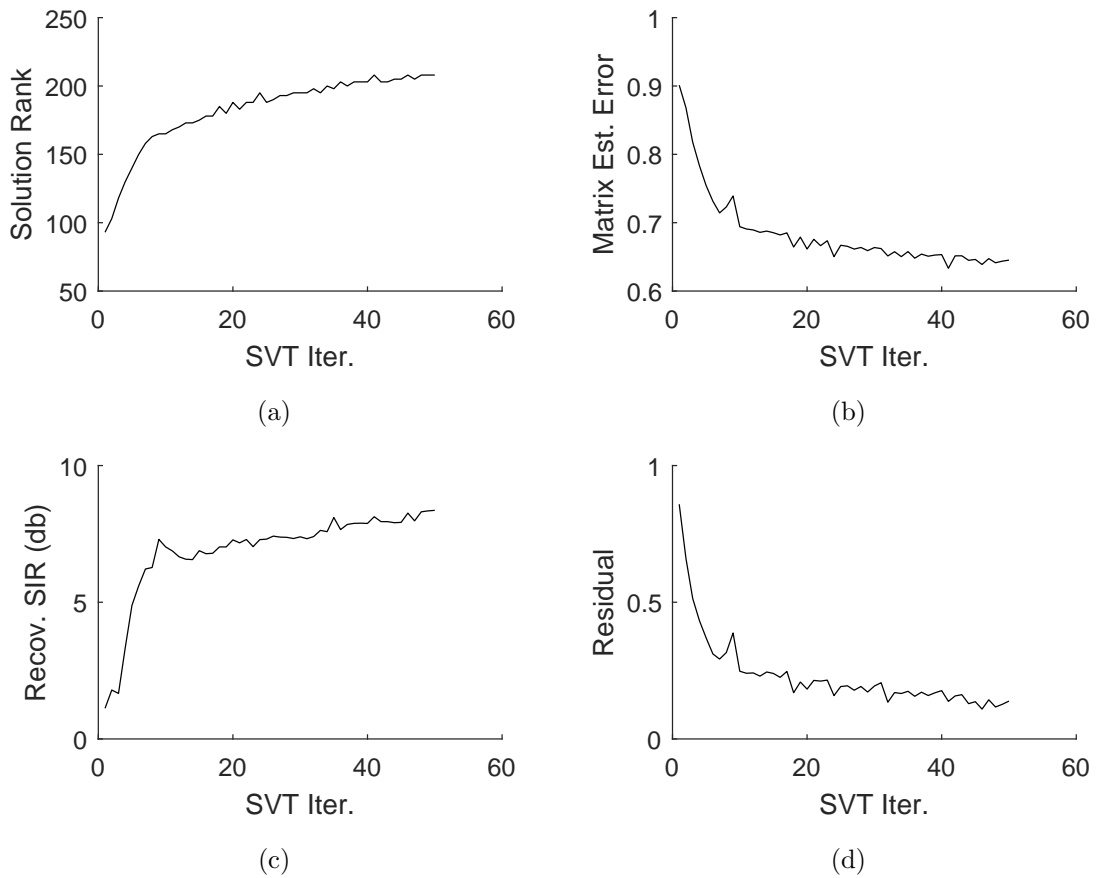


Figure 29: Example iteration results for the SVT algorithm operating on high-fidelity simulated clutter data. (a) Estimate of the interference rank increases as the SVT iteration progresses, which confirms that the resultant estimate of the covariance matrix improves. Improvement shown for three metrics: (b) direct matrix estimation error, (c) signal-to-interference ratio for a point target embedded in the interference (using the estimate as a filter), and (d) optimization residual.

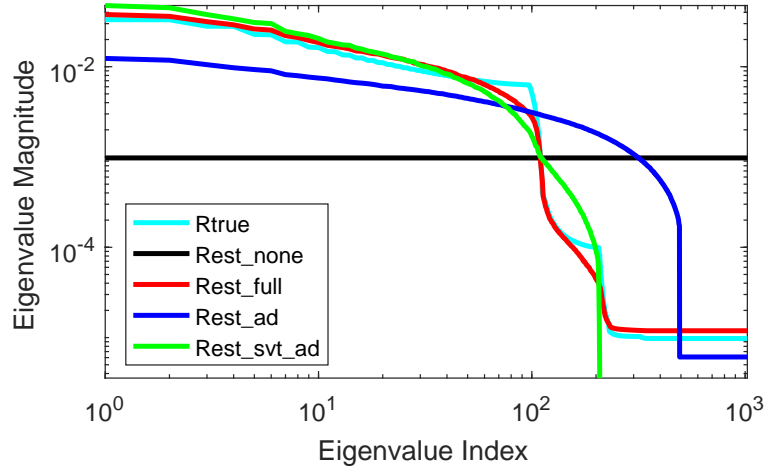


Figure 30: The spectral decay structures of the various covariance estimates for high-fidelity simulated clutter. The SVT estimate more closely matches the truth than does the compressed SMI which uses the same data as input.

Table 2: Parameters for MCARM data collection experiments

Parameter	Value
Range Bins	630
Pulses	128
Spatial Channels	11

and in the number of targets of opportunity. Summary plots of the data record RD050575 are shown in Figures 33–35. One can observe that the data displays some inconsistencies as a function of range, which are part of the motivation for using shorter data records as training data. One can also observe that the clutter does result in significant degradation of target detection absent any signal processing to remove its effects.

This data does not contain ground truth for either interference statistics or target locations. As such, we present the covariance estimates from the compressed data compared to that from the full data. For the purposes of evaluating target recovery, we extract presumed target locations by identifying the five bins in the data-cube that experienced the largest increase in relative amplitude by applying the STAP filter based on the full data SMI.

Figure 36 shows an example processing result. The clutter ridge slices through the angle-Doppler map per the collection geometry and timeline. The target is not visible in unfiltered case as it is largely covered by the clutter. However, in the full SMI case the

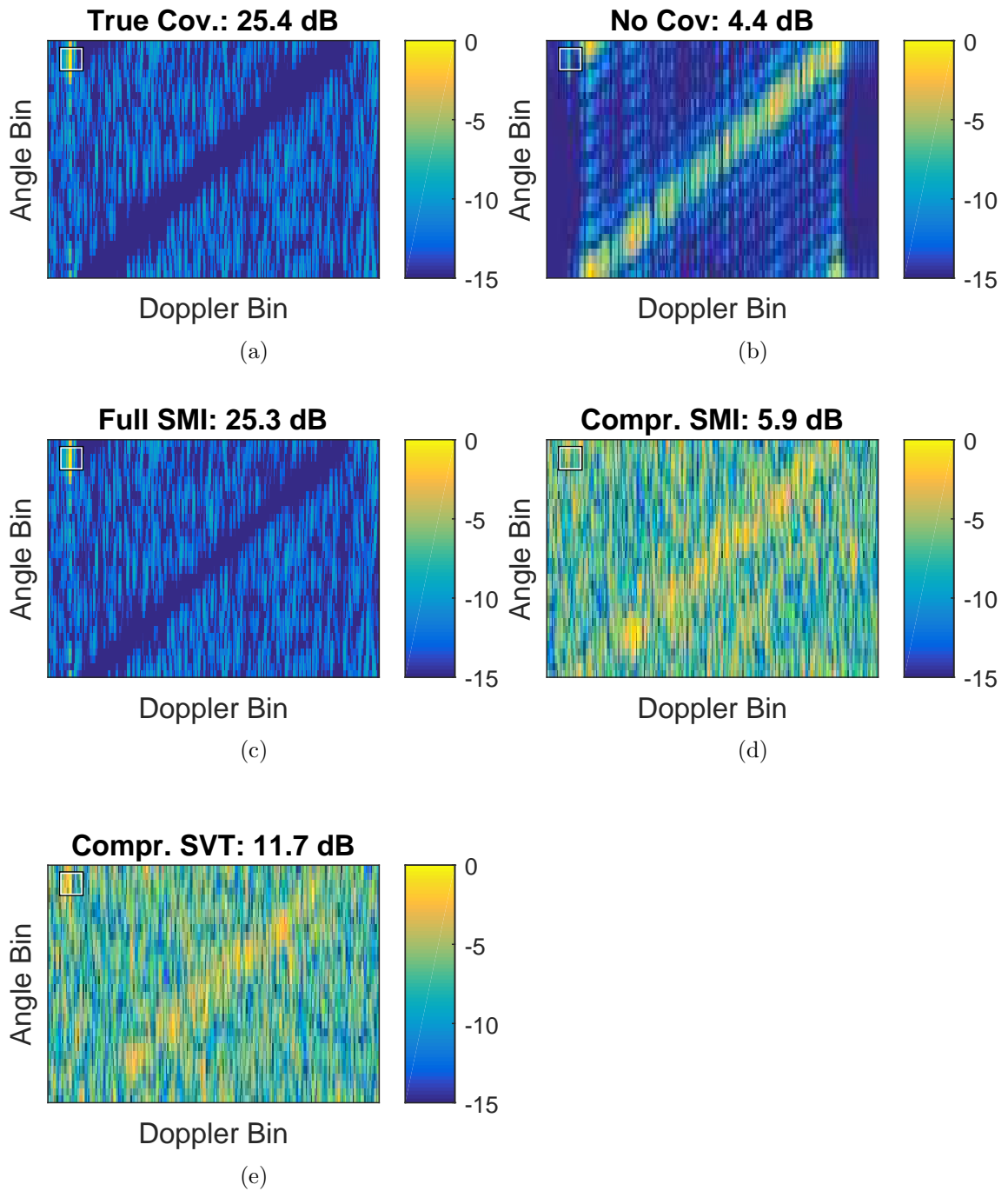


Figure 31: Angle-Doppler map results from each of the tested estimators using high-fidelity simulated clutter data with embedded target signal. The location of the true target is indicated with the square and signal-to-interference-plus-noise indicated in the subfigure titles.

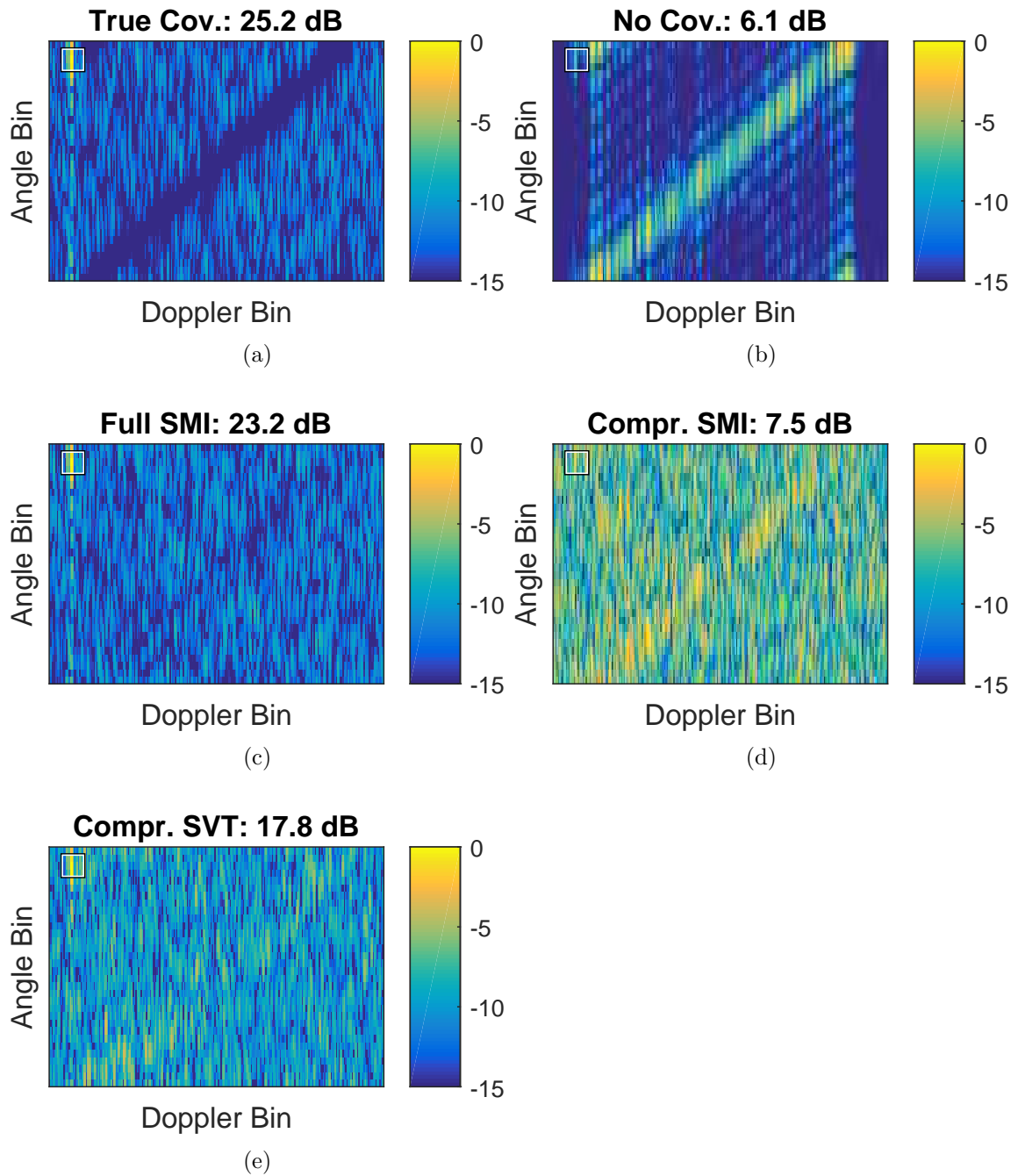


Figure 32: Angle-Doppler map results from each of the tested estimators using artificially-rank-reduced simulated clutter data with embedded target signal. The location of the true target is indicated with the square and signal-to-interference-plus-noise indicated in the subfigure titles.

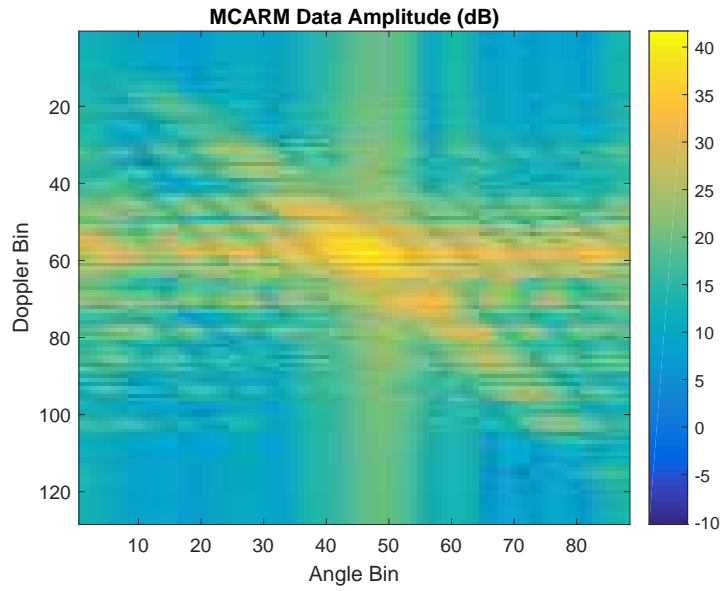


Figure 33: MCARM data record RD050575 averaged angle-Doppler map

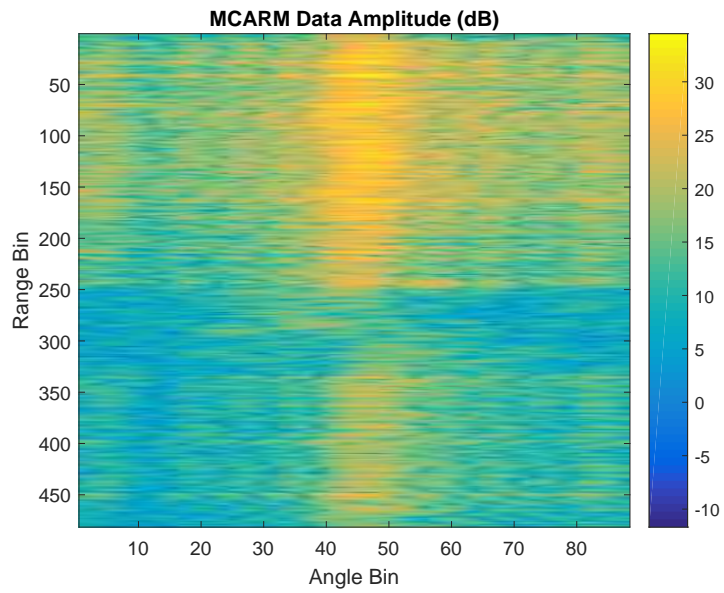


Figure 34: MCARM data record RD050575 averaged angle-range map

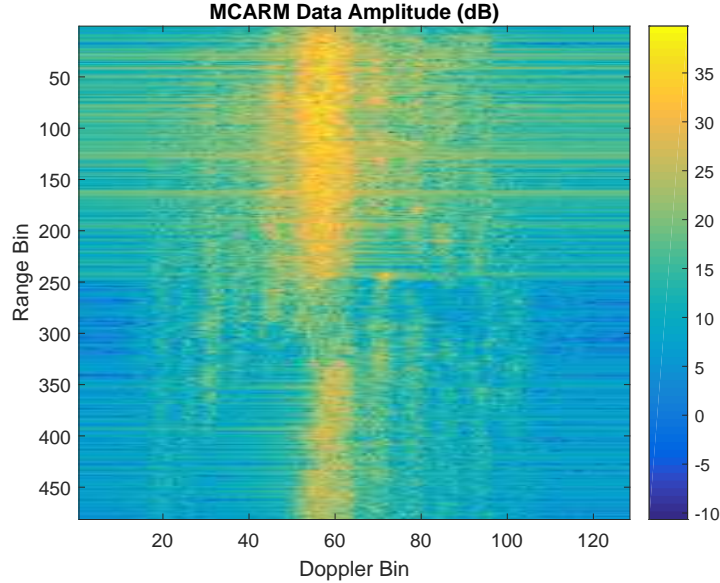


Figure 35: MCARM data record RD050575 averaged range-Doppler map

target is easily identifiable. Using compressed data, the SVT estimate shows better nulling of the clutter and preservation of target energy. Estimated interference eigenspectra are shown in Figure 37. It is evident that the SVT estimate provides a result much closer to the full data collection that does the diagonally loaded SMI based on compressed data.

One item of interest is the fact that the compressed SMI seems to perform worse than no covariance at all. Although this seems to be true, with a more optimal diagonal loading coefficient, this performance degradation could be eliminated.

4.5 Conclusion

The covariance matrix of interference in airborne GMTI radar can be estimated from compressed measurements with better accuracy using the iterative SVT approach than with compressed SMI, even with diagonal loading. This result holds when tested on flight test data, high fidelity clutter simulation data, and in synthetic structured interference. More compact interference structures lend themselves to better estimation by this method, as they better conform to the low-rank assumption. In the limit of very narrow interference, the SVT estimator performs as well as the true covariance matrix, even for undersampling ratios as high as 20. By contrast, for more complex interference structures performance is

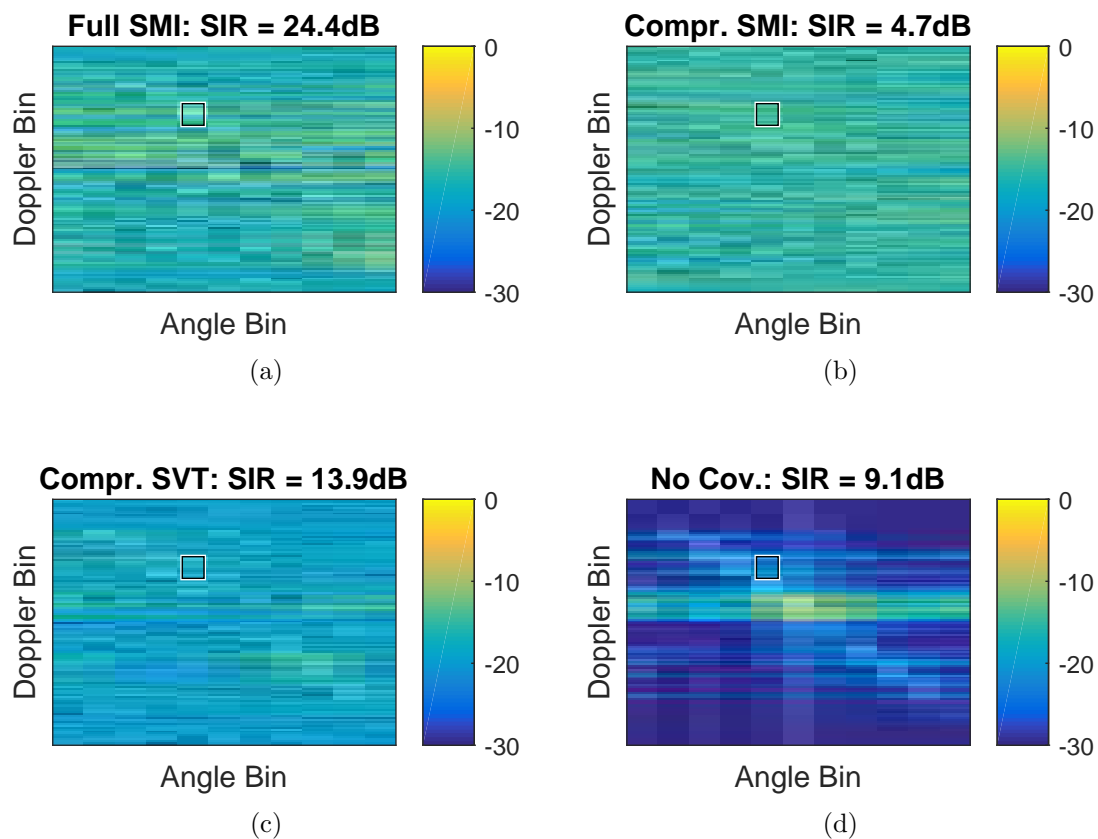


Figure 36: Angle-Doppler maps for range bin 363 in MCARM data record RD050575. The undersampling factor is 2 for the compressed cases. The location of a presumed target is indicated in each plot with a square. This target location is used to provide the indicated signal-to-interference ratio.

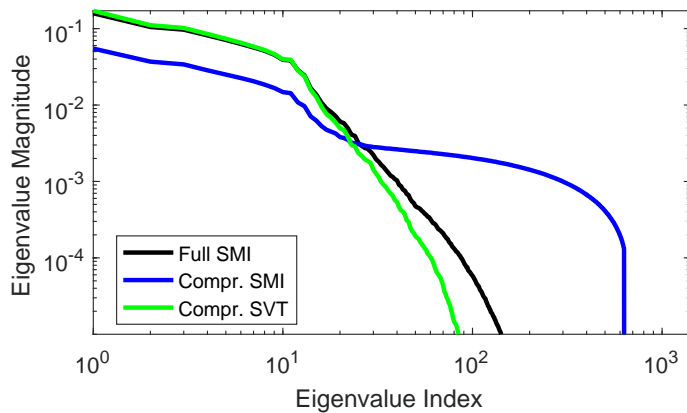


Figure 37: The eigenspectra of the tested estimates of the interference covariance for MCARM data record RD050575. The undersampling factor is 2 for the compressed cases.

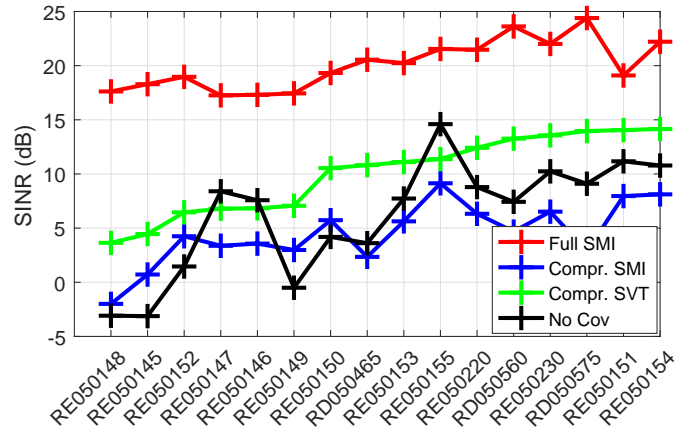


Figure 38: The performance of the tested estimators for each MCARM log file. Certain files feature more prominent targets in the observation space and targets vary in their location relative to the clutter ridge. The results shown use an undersampling factor of 2.

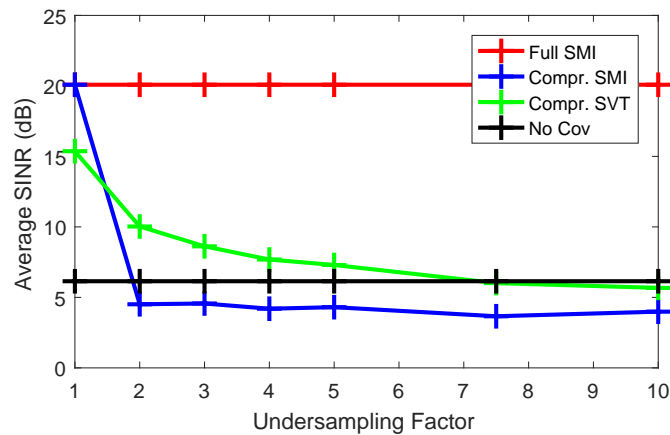


Figure 39: The performance of the tested estimators on the MCARM dataset as a function of undersampling factor. Results are averaged over all MCARM log files. In all undersampled cases the SVT estimate improves on the diagonally-loaded compressed SMI estimate.

not much better than the simple compressed sample covariance matrix.

In general, the techniques of low-rank matrix estimation and random matrix theory lend themselves very well to the adaptive filtering problem in radar and can be thought to have allowed gains in cases of compressed samples as well as in the standard, uncompressed, case.

CHAPTER V

APPLYING RANDOM MATRIX THEORY IN ADAPTIVE RADAR

This chapter deals with the application of random matrix theory to adaptive radar filtering. This topic was introduced briefly as a tool in Section 4.3.2.1, but will be developed more directly in this chapter. First, we cast the problem of covariance estimation in a setting that can be made precise for the application of this theory. Next, applicable portions of that body of theory are introduced. Finally, two applications of that body to the problem at hand are developed. The first is the inversion of the sample covariance spectrum to estimate the true eigenspectrum under the assumption of a parametrically-defined spectral structure. The second application is the selection of diagonal-loading and rank-reduction for STAP.

5.1 Problem Formulation

The motivating problem is pulse-Doppler radar. The goal is to estimate statistics of structured interference in order to apply a whitening filter and then to detect targets therein. Each of N_c receivers receives N_s pulses (slow time) with n samples per pulse (fast time). If we treat the data as n members of a $m = N_c N_s$ -dimension space, where n and m are on the same order, we would write

$$\mathbf{Y} = [\mathbf{y}_1 | \dots | \mathbf{y}_n] \in \mathbb{C}^{m \times n}. \quad (54)$$

The space-time interference is approximately stationary over these n samples. Therefore, putting aside any target energy, each of these vectors is approximately equal to a sample from an unknown normal distribution $\mathbf{y}_i \sim \mathcal{CN}(\mathbf{0}_m, \mathbf{R}_y)$. This interference distribution has both clutter and noise components: the noise contributes a diagonal component to the covariance matrix, while the clutter contributes a component with low- or approximately-low-rank structure as a result of the sensing geometry and timeline. So the unknown, true

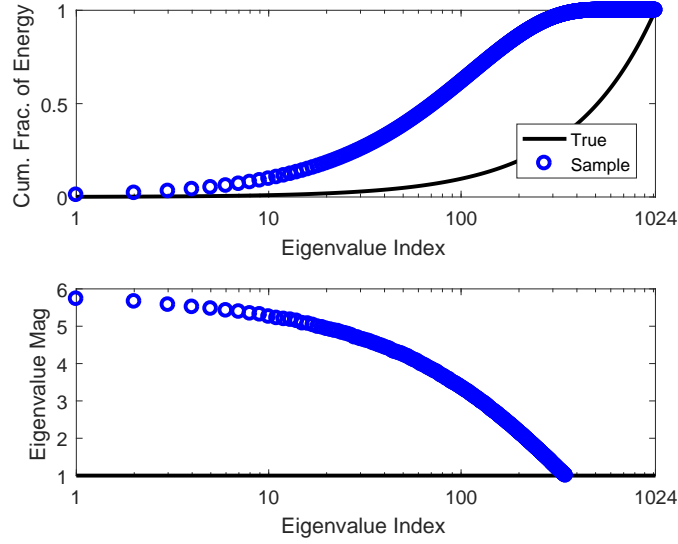


Figure 40: A set of sample eigenvalues is shown for a noise-only distribution, with many spread substantially from the true distribution. The parameters are $m = 1024$, $n = 512$, $k = 0$, $\nu = 1$.

covariance is the sum of the noise and clutter covariances

$$\mathbf{R}_y = \mathbf{R}_n + \mathbf{R}_c = \nu^2 \mathbf{I}_m + \mathbf{V} \mathbf{D} \mathbf{V}^H \quad (55)$$

where ν^2 is the measurement noise variance, $\mathbf{V} \in \mathbb{C}^{m \times k}$ with iid complex normal entries and unit-norm rows, \mathbf{D} is diagonal with entries $\sigma_1^2, \dots, \sigma_k^2$, and k is the clutter rank ($k < m$). One may assume that clutter is stronger than noise (necessitating adaptive filtering), though this may be quite variable. The sample covariance matrix provides a simple estimate of the true covariance as $\hat{\mathbf{R}}_y = (1/n) \mathbf{Y} \mathbf{Y}^H = (1/n) \mathbf{R}_y^{\frac{1}{2}} \mathbf{X} \mathbf{X}^H \mathbf{R}_y^{\frac{1}{2}}$, where $\mathbf{X} \in \mathbb{C}^{m \times n}$ and $\mathbf{X}_{i,j} \sim \mathcal{CN}(0, 1)$. As $n/m \rightarrow \infty$ this estimate will converge to the true distribution \mathbf{R}_y . But with limited data ($n \lesssim m$) the estimate can diverge substantially from the true covariance. This effect is illustrated in Figures 40 and 41, with sample eigenvalues falling far from the eigenvalues of the true distribution. This spread of the eigenvalues inhibits accurate identification of signal-bearing dimensions of the data. Eigenvectors with a projection into the clutter subspace may have small eigenvalues, as shown in Figure 42.

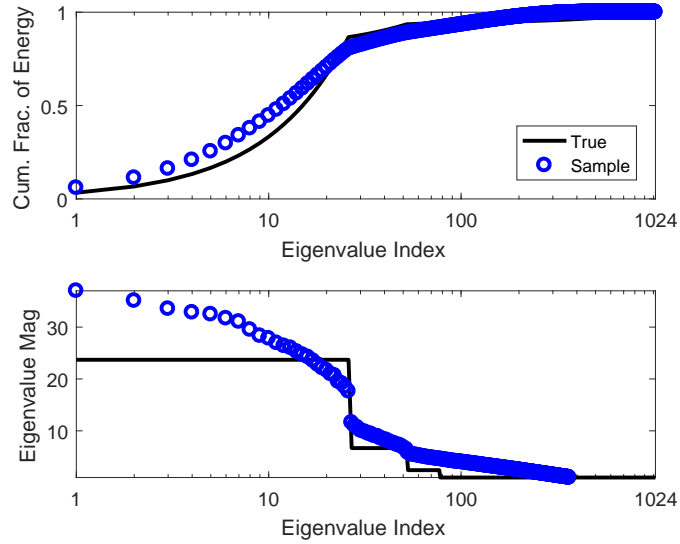


Figure 41: A set of sample eigenvalues is shown for a noise and clutter distribution, with many spread substantially from the true distribution. Clutter rank $k = 77$, clutter structure ID is 2, degrees of freedom $m = 1024$, number of observations $n = 512$, noise variance $\nu = 1$.

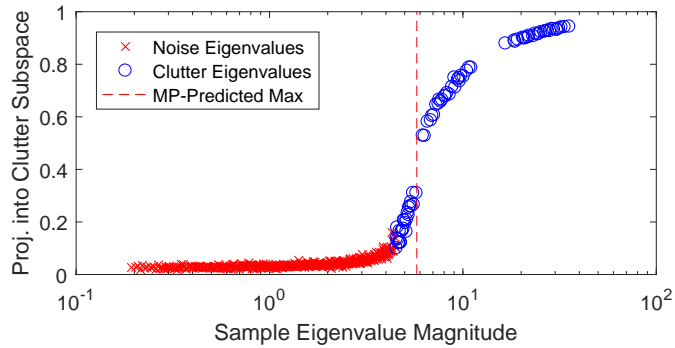


Figure 42: Projection of eigenvectors into the clutter subspace. Some smaller eigenvalues still correspond to eigenvectors with significant clutter component. As in Figure 41, clutter rank $k = 77$, clutter structure ID is 2, degrees of freedom $m = 1024$, number of observations $n = 512$, noise variance $\nu = 1$.

Table 3: The ensembles of classical random matrix theory. The commonly used Gaussian ensembles (Gaussian orthogonal ensemble, Gaussian unitary ensemble, and Gaussian symplectic ensemble) have been identified, showing their membership in this broader structure.

	Orthonormal (Real, $\beta = 1$)	Unitary (Complex, $\beta = 2$)	Symplectic (Quaternion, $\beta = 4$)
Hermitte (Wigner)	GOE	GUE	GSE
Laguerre (Wishart)	-	-	-
Jacobi (MANOVA)	-	-	-

5.2 Random Matrix Theory

5.2.1 RMT Overview

Random matrix theory is the study of matrix-valued random variables. In some cases, very strong statements about the characteristics of these random matrices can be made. This is especially true of the eigenspectrum of random matrices with certain structures. The classical building blocks of random matrix theory are the items in Table 3. This decomposition was introduced in [80] and [74] and has been elaborated more recently in [75] and [76]. The column specifies the field from which the ensemble is drawn. For the Hermitian row these would be:

- Gaussian Orthonormal Ensemble (GOE): symmetric $n \times n$ matrix generated by $(\mathbf{A} + \mathbf{A}^T)/2$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{A}_{i,j} \sim \mathcal{N}(0, 1)$
- Gaussian Unitary Ensemble (GUE): Hermitian $n \times n$ complex matrix generated by $(\mathbf{A} + \mathbf{A}^H)/2$ where $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{A}_{i,j} \sim \mathcal{CN}(0, 1)$
- Gaussian Symplectic Ensemble (GSE): dual-symmetric $n \times n$ quaternion matrix generated by $(\mathbf{A} + \mathbf{A}^D)/2$ where $\mathbf{A} \in \mathbb{Q}^{n \times n}$, $\mathbf{A}_{i,j} \sim \mathcal{QN}(0, 1)$, and A^D is the quaternion conjugate transpose.

The row specifies the structure of the ensemble. For the real case the entries are:

- Hermitian: $(\mathbf{A} + \mathbf{A}^T)/2$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{A}_{i,j} \sim \mathcal{N}(0, 1)$
- Laguerre: $\mathbf{A} \mathbf{A}^T$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{A}_{i,j} \sim \mathcal{N}(0, 1)$

- Jacobi: $\mathbf{A} \mathbf{A}^T (\mathbf{A} \mathbf{A}^T + \mathbf{B} \mathbf{B}^T)^{-1}$ where $\mathbf{A} \in \mathbb{R}^{m \times n_1}$, $\mathbf{B} \in \mathbb{R}^{m \times n_2}$ and $\mathbf{A}_{i,j}, \mathbf{B}_{i,j} \sim \mathcal{N}(0, 1)$

The complex Wishart case (LUE) is precisely the model in the case of adaptive radar, as well as others such as communication channel estimation, and system response modeling. There are many results describing the expected eigenvalues for matrices of this type.

5.2.2 Marčenko-Pastur

Results describing the distribution of eigenvalues can be divided into the global and local regimes. The global results provide a spectral measure, or distribution, of the eigenvalues overall. This can be thought of as a limiting distribution, about which there will be some fluctuation for realized finite matrices. The famous Wigner semi-circle law is a global-regime distribution that applies to Wigner matrices. For Wishart matrices, which are of interest in covariance estimation, the global distribution was provided by Vladimir Marčenko and Leonid Pastur, Soviet mathematical physicists [77]. Their eponymous law states that as $n \rightarrow \infty$, with $\frac{n}{m}$ fixed, the distribution of the eigenvalues of a Wishart matrix converges to a specified probability distribution with density:

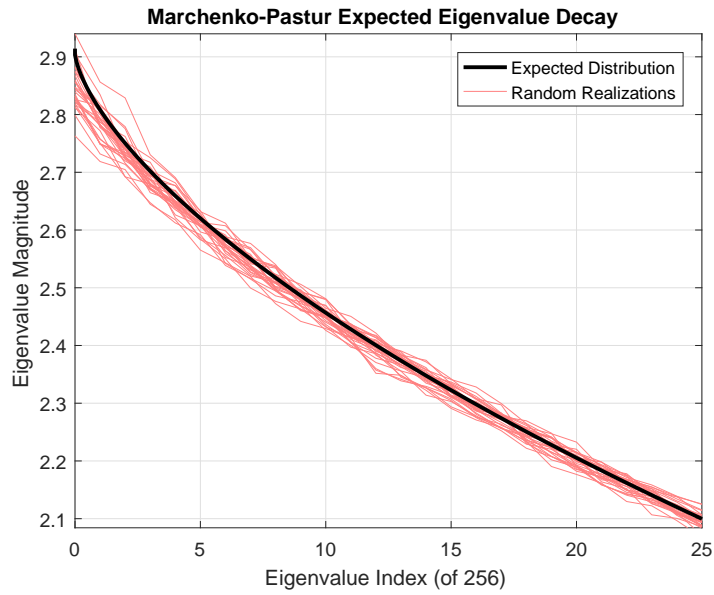
$$d(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x} \quad (56)$$

where $\lambda = \frac{m}{n}$ and $\lambda_{\pm} = \sigma^2(1 \pm \sqrt{\lambda})^2$.

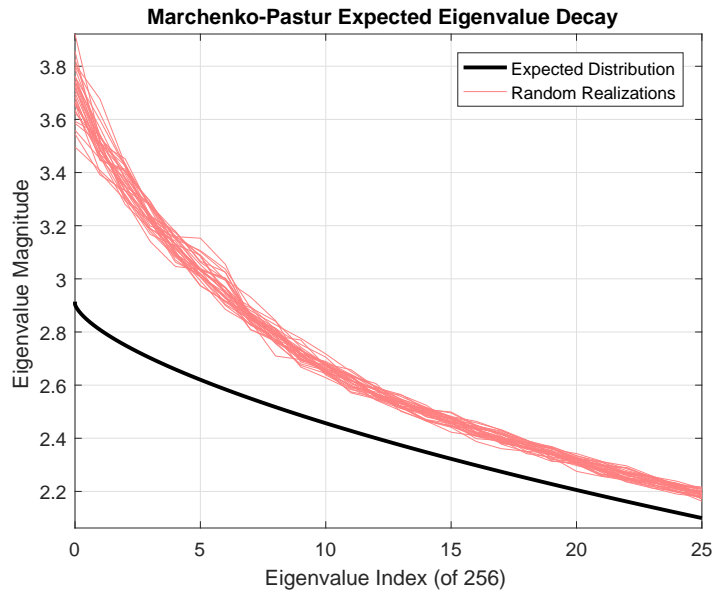
So even though the “true” eigenvalues are all equal, the expected sample eigenvalues vary greatly. This can be seen in Figure 43a. As a description of a null distribution this is accurate and converges quickly. But Marčenko-Pastur does not provide a description of the signal-bearing sample eigenvalues, as shown in Figure 43b. The distribution of sample eigenvalues of the clutter-corrupted case will be developed more in Section 5.2.4.

5.2.3 Tracy-Widom

Whereas Marčenko-Pastur provided a global distribution on the eigenvalues of the sample covariance matrix, the Tracy-Widom function provides a local distribution for the maximum eigenvalue of that matrix (also in the limiting case with fixed n/m and $n \rightarrow \infty$). For the



(a) Noise-only case



(b) Noise plus low-amplitude rank-10 interference

Figure 43: The Marčenko-Pastur distribution of noise-only eigenvalues. (a) In the noise-only case, this prediction is accurate. (b) When a low-amplitude, rank-10 interference is added to the noise, the sample eigenvalues are raised. This example uses 512 snapshots of a 256-length random vector; in the limit with infinitely many snapshots all eigenvalues would converge to unity in the noise-only case.

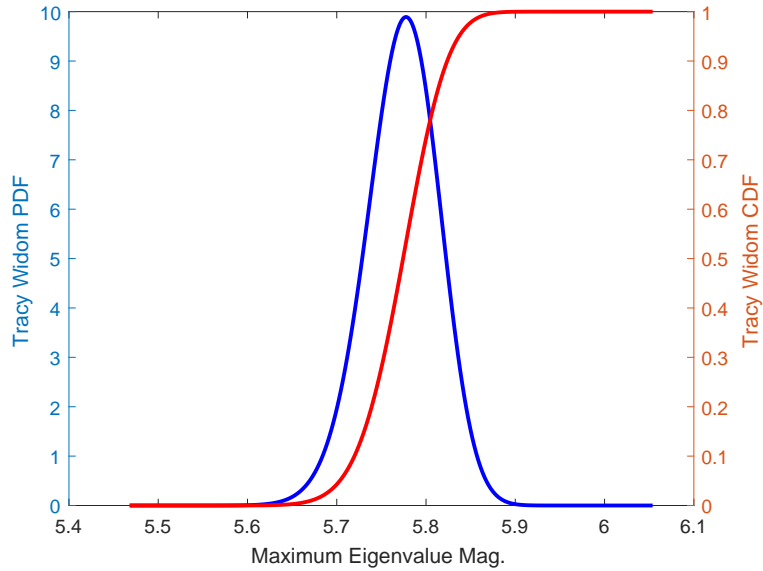


Figure 44: The Tracy-Widom function gives the distribution of the largest eigenvalue.

complex Wishart case, this function is [81], [82]:

$$F_2(x) = \exp\left(-\int_x^\infty (y-x)q^2(y)dy\right) \quad (57)$$

where $q(\cdot)$ is the Painlevé II solution to the following ODE:

$$\frac{d^2q}{dx^2} = 2q^3 + xq \quad (58)$$

with the solution boundary condition specified by the Airy function:

$$q(x) \sim \text{Ai}(x) \text{ as } x \rightarrow \infty \quad (59)$$

There is not a closed form representation of the function $q(x)$, but (58) can be solved using ODE solvers of various types. A useful framework for deriving a numerical solution for $q(x)$ can be found in [83]. An example of the distribution can be seen in Figures 44 and 45. These calculations exploit the MATLAB `ode45` solver and the `airy` special function.

In a signal detection setting, the Tracy-Widom function can be used to build a sufficient statistic for accepting or rejecting a null hypothesis that a set of data samples contains only noise. In order to develop a test for the non-noise subspace, suppose:

- An observed sample covariance matrix produces eigenvalues $\lambda_1 \dots \lambda_m$.

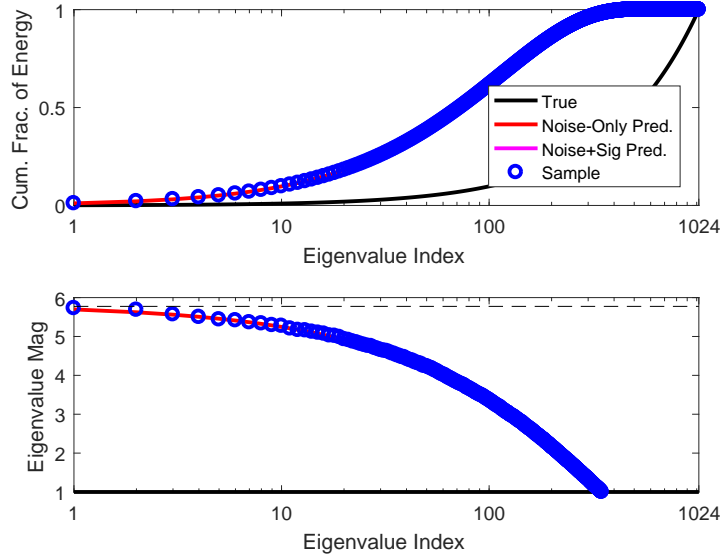


Figure 45: The sample eigenvalues verify two different RMT-derived predictions. The Marčenko-Pastur prediction is shown in red, which closely matches the sample values. And the Tracy-Widom 95% confidence test for the maximum noise-only eigenvalue is shown as a dashed line, which closely matches the maximum sample value.

- The expected eigenvalues produced by Marčenko-Pastur are $\tilde{\lambda}_1 \dots \tilde{\lambda}_m$
- The Tracy-Widom function giving the cumulative distribution function for the largest eigenvalue is $F_{\lambda_1}(\cdot)$ with quantile function $F_{\lambda_1}^{-1}(\cdot)$

The simplest hypothesis test for a non-noise subspace is to count how many eigenvalues exceed some threshold based on $F_{\lambda_1}(\cdot)$ and a confidence level. For example, any $\lambda_i > F_{\lambda_1}^{-1}(.99)$ can be confidently identified as signal+noise.

5.2.4 Non-Null Stieltjes Transform Distribution

The Marčenko-Pastur and Tracy-Widom distributions both give descriptions of the spectrum of the noise-only Wishart matrix. Silverstein, with several co-authors, developed the theory for predicting the distribution in the signal-present (non-identity covariance) case [84]. Their approach extends that of Marčenko-Pastur. Given a true distribution of eigenvalues and a sampling ratio, the broad steps for computing the spectral distribution are:

1. Use the Stieltjes transform to map the eigenvalue distribution onto the complex half

plane \mathbb{C}^+ .

2. Establish convergence in the Stieltjes transform domain (m/n fixed, $m, n \rightarrow \infty$).
3. Compute the inverse Stieltjes transform integral to produce the distribution.

Silverstein's primary contribution is to establish convergence of the limiting spectral distribution in the Stieltjes transform domain. For any PDF $G(\cdot)$ the Stieltjes transform is:

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \quad (60)$$

With inverse:

$$G([a, b]) = \frac{1}{\pi} \lim_{\eta \rightarrow 0} \int_a^b \text{Im } m_G(\xi + i\eta) d\xi \quad (61)$$

It can also be expressed using the resolvent of a random Hermitian matrix \mathbf{A} with spectral PDF $G(\cdot)$:

$$m_G(z) = \frac{1}{m} \text{tr}((\mathbf{A} - z\mathbf{I})^{-1}) \quad (62)$$

Terry Tao states that the Stieltjes transform can be viewed as “a complexification of the spectral measure. . . . As such, it neatly packages the spectral information in a way that can be easily manipulated by the methods of complex analysis [85].”

For the case under consideration, if $H(\cdot)$ is taken to be the true spectral CDF, the Stieltjes transform $m(\cdot)$ can be computed by solving the following equation:

$$m(z) = m \quad \text{s. t.} \quad f(m) = m + \left(z - c \int \frac{t dH(t)}{1 + tm} \right)^{-1} = 0 \quad (63)$$

with gradient

$$\frac{df(m)}{dm} = 1 - \frac{c \int \frac{t^2 dH(t)}{(1+tm)^2}}{\left(z - c \int \frac{t dH(t)}{1+tm} \right)^2} \quad (64)$$

and the eigenvalue density is equal to

$$d(z) = \frac{1}{c\pi} \text{Im}(m(z)), \quad z \in \mathbb{R}^+. \quad (65)$$

Calculating the distribution takes some care. The basic outline of the procedure is to first identify the support of the distribution and then to calculate the distribution in that

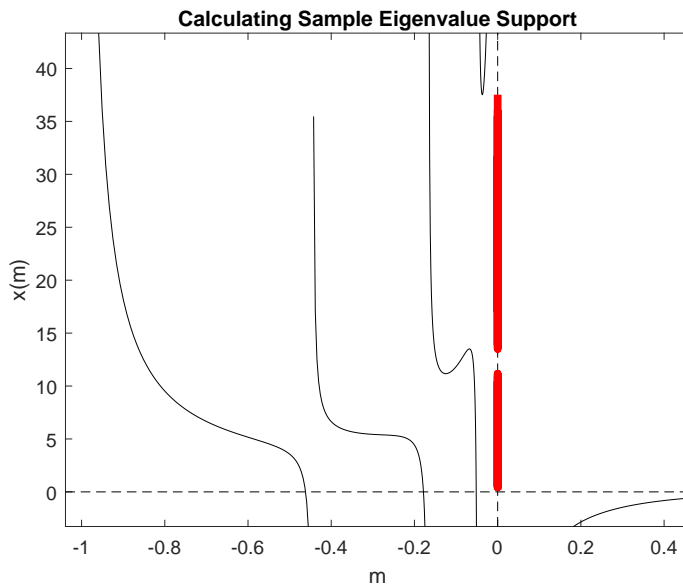


Figure 46: Plot of $x(m)$ which is used to calculate the support of the sample eigenvalue distribution. The red markers indicate the positive range over which $x(m)$ is non-increasing and represents the support of the sample eigenvalue distribution. Here $H(t) = \frac{7}{8}\delta(t-1) + \frac{1}{24}\delta(t-2.25) + \frac{1}{24}\delta(t-6) + \frac{1}{24}\delta(t-21)$ and has half as many samples as the dimension of the distribution.

domain. The support of the distribution is the set of positive real numbers for which

$$x(m) = -\frac{1}{m} + c \int \frac{t dH(t)}{1+tm} \quad (66)$$

is not increasing. A plot of this calculation is shown in Figure 46 for a particular $H(t)$. Then calculating the distribution itself requires solving (63) and applying the transformation in (65). No closed form solution to (63) exists but it is differentiable, convex, and has a unique solution in \mathbb{C}^+ .

Newton's method can be used to solve (63) for $m(z^*) \in \mathbb{C}^+$ for any $z^* \in \text{supp}(d(\cdot))$. We use a stopping criterion of $|m_i - m_{i-1}|/|m_{i-1}| < 10^{-6}$ for iteration i . In the absence of any other solution, a starting guess of $-10 + 10j$ is chosen. But typically the density is obtained sequentially for a large number of sample points in $\text{supp}(d(\cdot))$. If the last two consecutive points yielded valid solutions, a linear extrapolation from the previous two solutions is used as the starting point. The number of iterations needed to converge is typically small for cases with a good starting guess. This number of iterations needed to converge is shown in

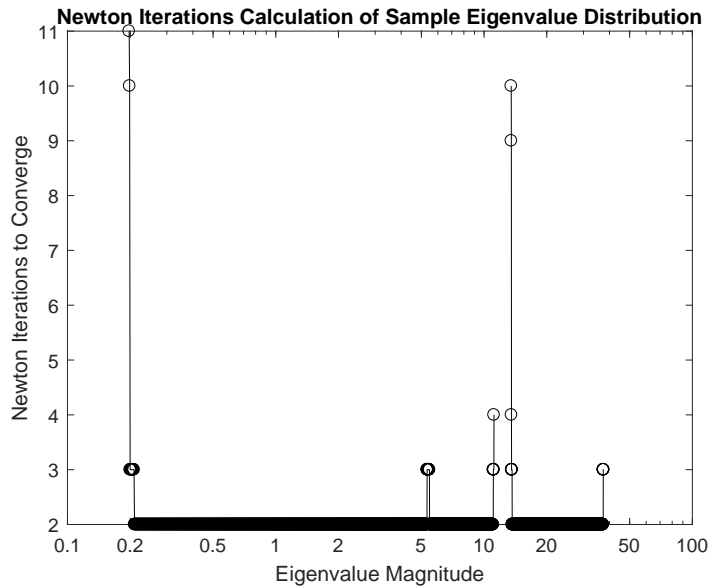


Figure 47: The number of Newton iterations to converge to a solution of (63). This number is typically small, since the solution to the previous two iterations are used to linearly predict a starting guess. Here this example uses synthetic interference structure 4 and $\text{CNR} = 13$ dB and has half as many samples as the dimension of the distribution.

Figure 47.

Once the support and density have been calculated, a simple consistency check can be made by integrating the density $d(x)$ to form a cumulative density $D(x) = \int_0^x d(\tau)d\tau$. If $|D(\infty) - 1| > \alpha$, the result is rejected as inconsistent. A sample result of these calculations is illustrated in Figures 48 and 49 showing the prediction and its accuracy.

These calculations are nontrivial and for problems with large eigenvalues very fine sampling is required. To provide an efficient solver the algorithm was implemented in Julia [86], a dynamic, just-in-time compiled language focused on numerical and scientific computing. This language advertises high performance in non-vectorized, loop-intensive workloads like this one. Indeed, in this case the Julia implementation provided an approximate $5 \times$ speedup relative to a reference MATLAB implementation. This implementation proved efficient and able to compute a distribution estimate under reasonable parameter choices in approximately 10 milliseconds. The source code for this solver has been published in the open source [87] and is provided in Appendix A of this thesis.

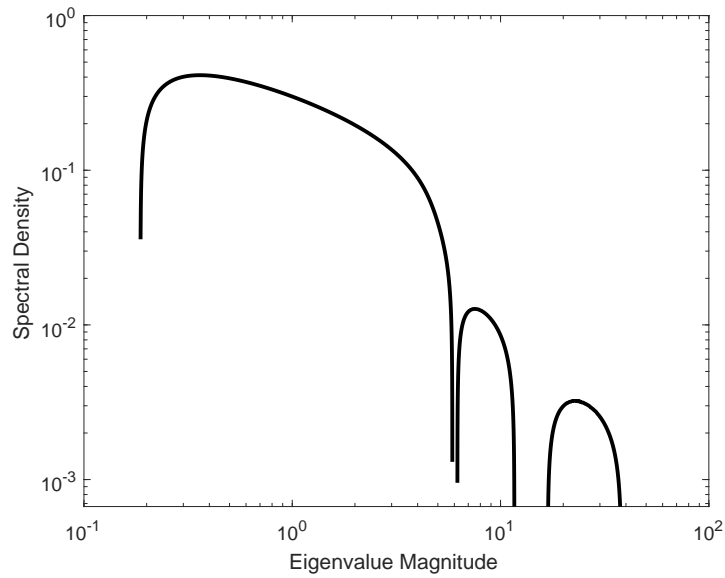


Figure 48: The Stieltjes transform-predicted sample eigenvalue distribution. This calculation predicts three groups of eigenvalues.

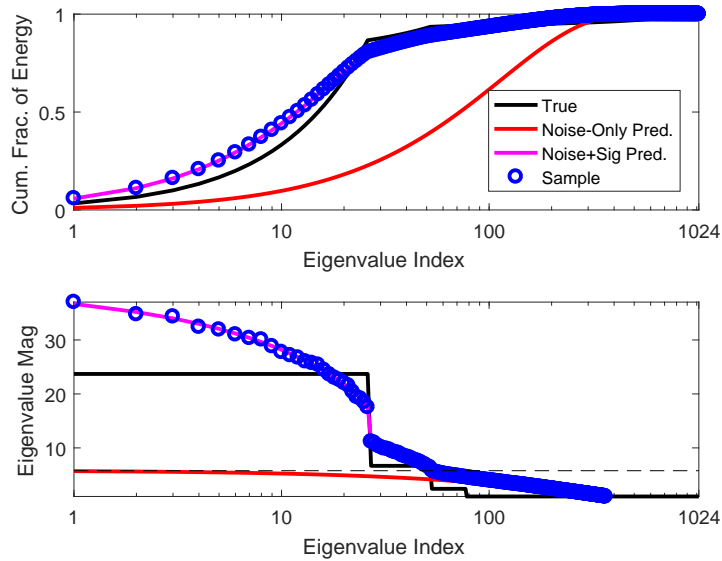


Figure 49: The Stieltjes transform-predicted sample eigenvalue distribution (magenta line) shown with a sample (blue circles). The calculation based on Stieltjes transform accurately predicts the eigenvalues of the sample.

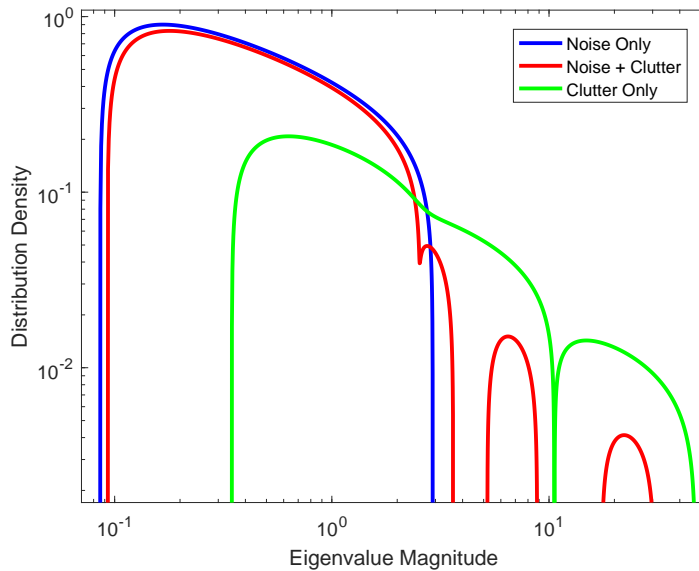


Figure 50: The spectral density of noise, clutter, and the sum (noise+clutter). The sum is not a linear combination of the components.

5.2.5 Characterizing Detectability

If one has knowledge of the true distribution of the signal (or interference to be estimated) the distribution of sample eigenvalues can be readily calculated using the tools described in this section. This is, of course, not a viable estimation technique as one does not have knowledge of the signal in that case. But it does have value as an analytic technique for predicting performance and analyzing behavior under varying conditions.

The noise-only (Marčenko-Pastur) spectral PDF, $f_n(\cdot)$, and the non-null noise + clutter spectral PDF, $f_c(\cdot)$, calculated through the Stieltjes transform procedure, can be compared. The noise, clutter, and noise + clutter distributions, in PDF and CDF, can be seen in Figures 50 and 51. These figures make clear that the sum distribution not a linear combination of the noise and clutter components. But it is also clear that a large portion of the sum distribution lies beyond the limit of the noise distribution. The relationship between the the noise and sum distributions can be analyzed to characterize the detectability of the clutter.

For any given eigenvalue magnitude, x , the PDFs can be evaluated and a fractional

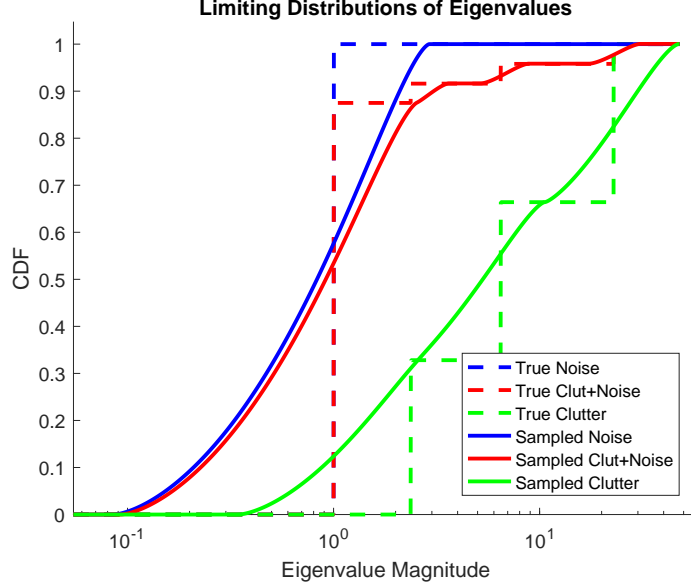


Figure 51: The cumulative spectral density of noise, clutter, and the sum (noise+clutter). The sum is not a linear combination of the components.

density

$$f_f(x) = \frac{f_c(x)}{f_n(x)} \quad (67)$$

can be computed. An example of this fractional density is shown in Figure 52. It can be used to derive decision criteria for whether a given eigenvalue is signal-bearing or not. In particular, a threshold T_f can be selected for the fractional density in which those eigenvalues with a fractional density greater than T_f are judged to be information-bearing and the others merely noise. The fraction of eigenvalues identified as signal-bearing can be plotted as a function of that fractional density threshold—an example plot is shown in Figure 53. In cases where the clutter is well-separated from the noise, the dimensions will be easy to detect. In this case, the two distributions are not well-separated and for any reasonable detection threshold (> 0.5) fewer than all the clutter eigenvalues are detected.

The identifiability of a signal can be characterized by the distributions. Suppose there is a noise-only sample eigenvalue distribution with Tracy-Widom function $F_{\lambda_1}(\cdot)$ and inverse $F_{\lambda_1}^{-1}(\cdot)$, and assume there is a noise+clutter sample eigenvalue distribution with PDF $d(\cdot)$, as in (65), and CDF $D(\cdot)$ with inverse $D^{-1}(\cdot)$. Suppose, furthermore, that the true clutter rank is k and that the observation dimension is N_{cs} and the number of observations is N_f .

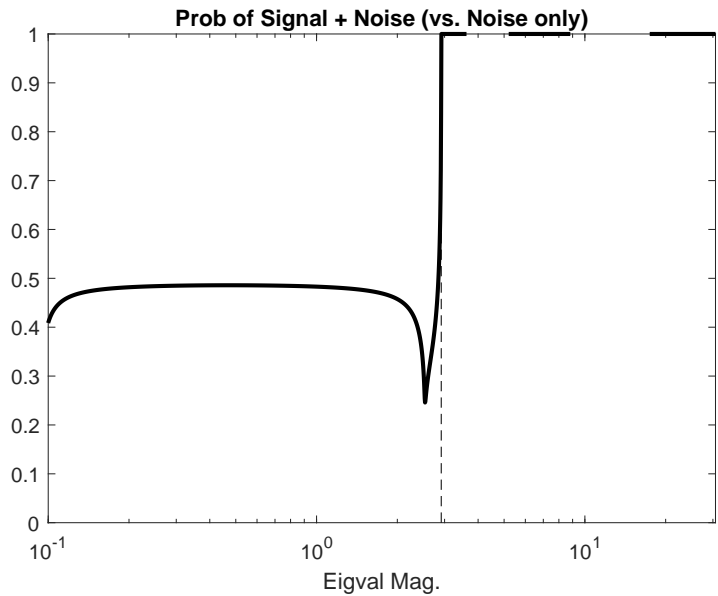


Figure 52: Identifiability as a function of eigenvalue magnitude. Large eigenvalues can be identified as belonging to the Noise + Clutter distribution but smaller ones could have come from either the Noise-only or Noise+Clutter distributions. The black dashed line shows the Marčenko-Pastur maximum eigenvalue, so any eigenvalue above that magnitude is unlikely to be from from the Noise-only distribution.

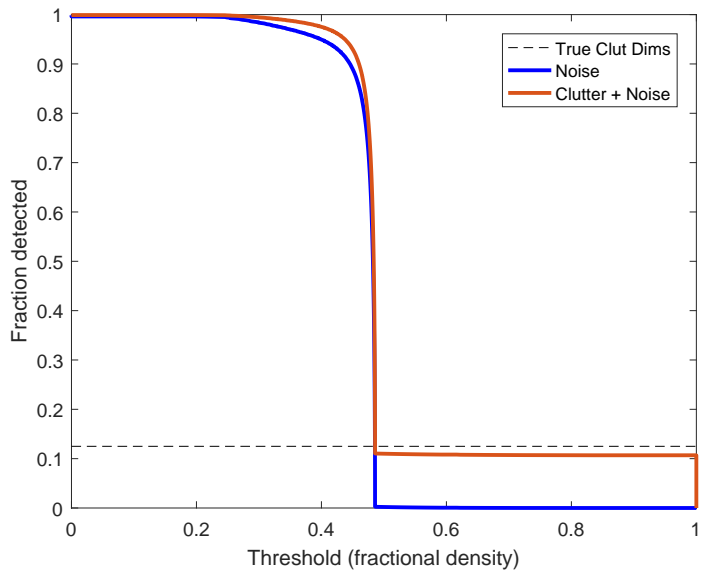


Figure 53: The fraction of dimensions identified as signal-bearing as a function of the decision threshold for both the Noise-only and Noise + Clutter distributions. For reasonable thresholds (> 0.5) fewer dimensions than the true number of clutter dimensions are detectable.

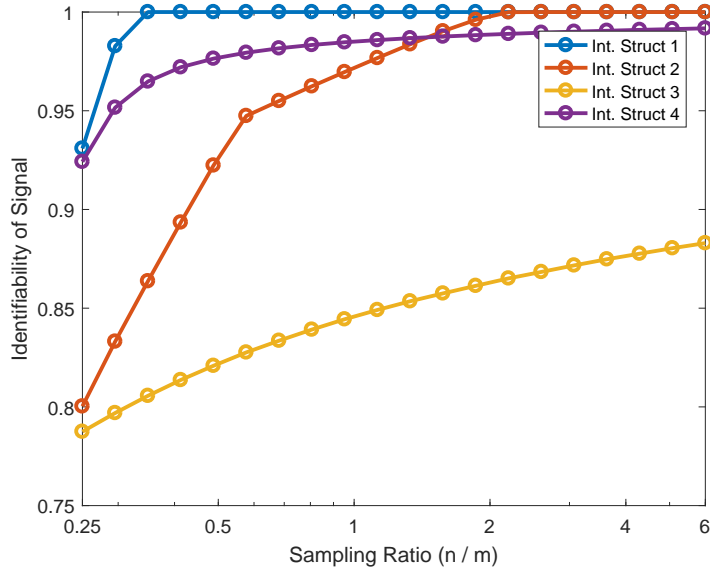


Figure 54: Identifiability as a function of sampling ratio. A higher sampling ratio indicates more samples per estimated dimension.

Then the fraction of identifiable clutter dimensions is

$$I_D = \frac{1 - D\left(F_{\lambda_1}^{-1}(.5)\right)}{\frac{k}{N_{cs}} \max\left(\frac{N_{cs}}{N_f}, 1\right)}. \quad (68)$$

If the true clutter eigenvalues are $\sigma_1, \dots, \sigma_k$ with $\sigma_{i+1} < \sigma_i$ and $s_n = \sum_{i=1}^n \sigma_i$ is the cumulative sum series, then the fraction of identifiable clutter energy is

$$I_E = \frac{s_n}{s_k}, \quad n = \lfloor I_D k \rfloor \quad (69)$$

This identifiability of the clutter energy is plotted in Figures 54 and 55. These plots, which are based on the same interference structures defined in Chapter 4, illustrate the increasing clutter identifiability as a function of both samples used and CNR.

Identifiability of the clutter space is a necessary but not sufficient condition to effective filtering. In addition to this identifiability metric, evaluation of STAP performance as a function of the number of samples used to estimate the interference statistics shows that performance is severely degraded with the number of training samples is close to the number of degrees of freedom. This effect is shown in Figure 56. The severely degraded performance at that point can be mitigated by diagonal loading, but performance is very sensitive to the

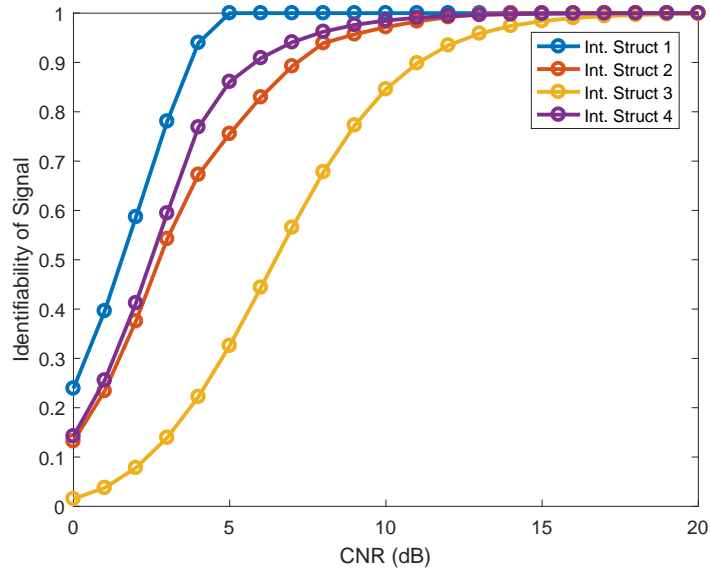


Figure 55: Identifiability as a function of clutter-to-noise ratio. Higher CNR separates the clutter eigenvalues from those of the noise and allows them to be detected an isolated more easily.

selection of diagonal loading parameter, and optimal selection is not straightforward. One may compare Figure 56 to the Reed-Mallett-Brennan rule [26], which states that twice as many samples as dimensions are needed to give less than 3 dB of SINR loss. This result is confirmed by the SMI result curve. But these curves also show that less than 3 dB of SINR performance can be achieved using far fewer samples through rank reduction. In Section 5.3, a method is provided for selecting the reduced rank and diagonal loading in order to achieve this improvement.

This effect can be explained by the distributions of sample covariance eigenvalues predicted by the Stieljes transform approach. Figure 57 shows the predicted distributions for several different numbers of samples. This plot makes it clear that as the number of samples nears the number of degrees of freedom, the likelihood of finding small eigenvalues increases dramatically, which drives sensitivity in the interferences estimate. The separation between the noise and clutter is increasing as the number of samples increases, and the identifiability metric above increases as well. Furthermore, the observed performance is driven not only by the identifiability, but also by the noise sensitivity.

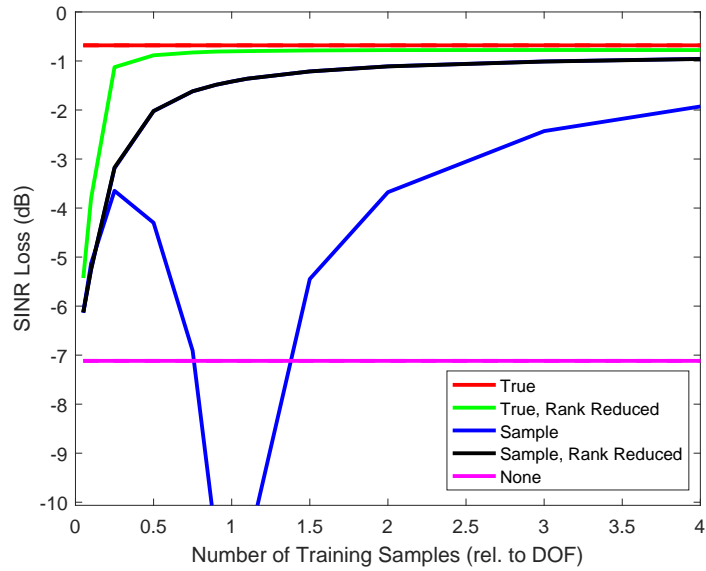


Figure 56: Performance of the adaptive filter as a function of the number of training samples. When the number of samples is close to the sample size the performance of SMI decreases. Rank reduction maintains good performance through this space. Parameters include interference structure ID of 4 and a CNR of 8 dB.

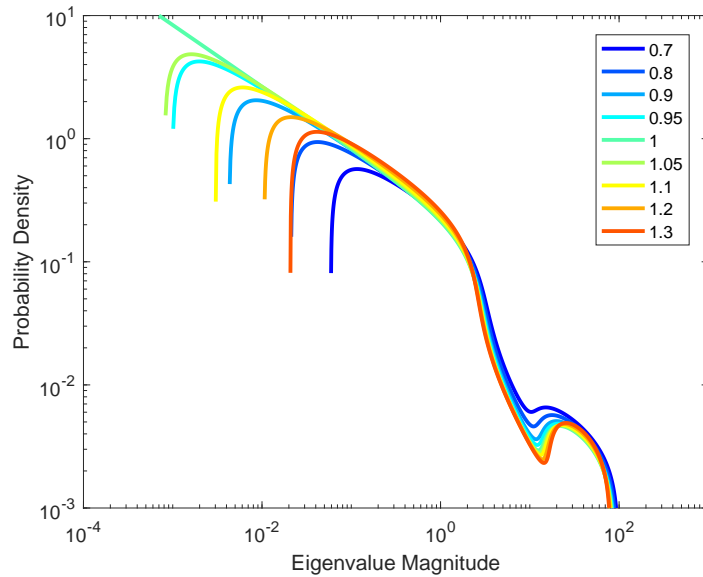


Figure 57: The predicted sample eigenvalue densities for several different training sample sizes. When the number of samples is close to the sample size the probability of small eigenvalues increases. These small eigenvalues inhibit stable matrix inversion. Parameters include interference structure ID of 4 and a CNR of 8 dB.

An important takeaway from this analysis is the performance trough that exists around $n = m$. This trough can be avoided by appropriately formulating a reduced-rank estimate, as will be shown in Section 5.3. But even without using rank reduction, there is performance to be exploited to the left of the trough. This performance region could be useful in cases of interference inhomogeneity or computational constraints. Typically STAP practitioners have worked to stay to the right of the trough, rightly avoiding that performance degradation and that sensitivity to diagonal loading selection.

5.3 Application to STAP Parameter Selection

5.3.1 Problem Statement

Two common processing approaches in adaptive filtering are diagonal loading and dimensionality reduction. Diagonal loading modifies the sample covariance matrix by adding a scaled identity component [73]. The straightforward sample covariance matrix was $\hat{\mathbf{R}}_y = (1/n)\mathbf{Y}\mathbf{Y}^H$. The diagonally loaded sample covariance matrix is

$$\hat{\mathbf{R}}_d = \hat{\mathbf{R}}_y + \delta \mathbf{I}_m.$$

Data dependent reduced rank STAP uses an eigen-decomposition of the sample covariance matrix to improve stability of the distribution estimate [12]. If $\hat{\mathbf{R}}_y$ has eigen-decomposition $\hat{\mathbf{R}}_y = \sum_{i=1}^m \mathbf{v}_i d_i \mathbf{v}_i^H$ with $d_i > d_{i+1}$ then the reduced rank estimate can be expressed as

$$\hat{\mathbf{R}}_r = \sum_{i=1}^p \mathbf{v}_i d_i \mathbf{v}_i^H, \quad p < m$$

Then the techniques can be combined

$$\hat{\mathbf{R}}_{dr} = \delta \mathbf{I}_m + \sum_{i=1}^p \mathbf{v}_i d_i \mathbf{v}_i^H.$$

These techniques have found wide use because of their simplicity and effectiveness. But the selection of p and δ have not been rigorously studied nor have optimal selection criteria been determined.

Using the Marčenko-Pastur distribution as a starting point we will propose a means of determining δ and p to accurately estimate the covariance matrix and filter the interference. Because the Marčenko-Pastur distribution is simple to compute and gives a central estimate

of the noise-only spectrum, it can be used to easily count eigenvalues that are definitely information-bearing. To this end, define the Marčenko-Pastur rank as the number of sample eigenvalues greater than maximum eigenvalue predicted by the Marčenko-Pastur law, as defined in (56).

5.3.2 Results

To test the selection of δ and p we developed a number of synthetic structured interference models. These interference models represent low-rank or approximately-low-rank interference using a series of decay functions, shown in Figure 18. Together with a clutter-to-noise-ratio (CNR) these decay functions can be used to specify the interference distribution as defined in (55). Samples from this distributions are generated to be used as training data for estimating the covariance, using a variety of estimators which will be enumerated in the next subsection. Each of these covariance estimates can be inverted to form an estimated whitening filter. Next, additional samples of the interference can be generated with point targets embedded therein. The test samples are processed using each of the estimated filters and four performance metrics are calculated: covariance matrix Frobenius norm error, precision matrix Frobenius norm error, average signal-to-interference-plus-noise ratio (SINR) loss, and average sample target signal-to-interference ratio.

5.3.2.1 Tested Estimators

A variety of estimators were tested to provide results and context.

- True: The true full interference covariance
- True, rank reduced: Hard thresholding applied to the true interference covariance eigenvalues (taking the p largest true eigenvalues)
- Singular Value Thresholding (SVT): A technique based on an approach proposed in [58], which was developed for estimating the covariance from compressed measurements, as in compressed sensing radar
- OptShrink: The previously referenced method that calculates the shrinkage coefficients given an assumed rank p

- Soft thresholding: The sample covariance matrix rank is reduced by shrinkage of the eigenvalues that leaves only p nonzero eigenvalues
- Hard thresholding: The sample covariance matrix rank is reduced by setting all but the p largest eigenvalues to zero
- No covariance: $\hat{\mathbf{R}} = \mathbf{I}_m$

To each of these estimates we add the diagonal loading $\hat{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}_m$ before inverting and evaluating a number of test cases to measure performance.

5.3.2.2 Figures of Merit

We will examine the behavior of the estimates as a function of δ and p using three figures of merit: covariance matrix accuracy, precision matrix accuracy, and post-filtering signal-to-interference ratio.

A natural figure of merit for the adaptive filtering estimation problem is the accuracy of the estimate of the interference covariance matrix. This is the norm of the error of the estimate:

$$e_c(\hat{\mathbf{R}}) = \left\| \hat{\mathbf{R}} - \mathbf{R}_y \right\|_F \left\| \mathbf{R}_y \right\|_F^{-1}. \quad (70)$$

Another potential figure of merit is the norm error of the inverse of the covariance matrix. It is defined as:

$$e_p(\hat{\mathbf{R}}) = \left\| \hat{\mathbf{R}}^{-1} - \mathbf{R}_y^{-1} \right\|_F \left\| \mathbf{R}_y^{-1} \right\|_F^{-1}. \quad (71)$$

Because the inverse is used to generate the adaptive filter, this metric is more directly tied to the application.

A third figure of merit is the SINR loss. This is the average loss introduced by error in the estimated interference covariance. In this case, with point targets and unit-variance noise, this can be expressed simply as

$$s_l(\hat{\mathbf{R}}) = \frac{1}{m} \sum_{i=1}^m \left(\frac{\left(\hat{\mathbf{R}}^{-1} \right)_{i,i}^2}{\left(\hat{\mathbf{R}}^{-1} \mathbf{R}_y \hat{\mathbf{R}}^{-1} \right)_{i,i}} \right). \quad (72)$$

The final figure of merit is also closely aligned to the problem at hand. The signal-to-interference ratio of an estimate \mathbf{a} is calculated using the location of the true target, at index k , as:

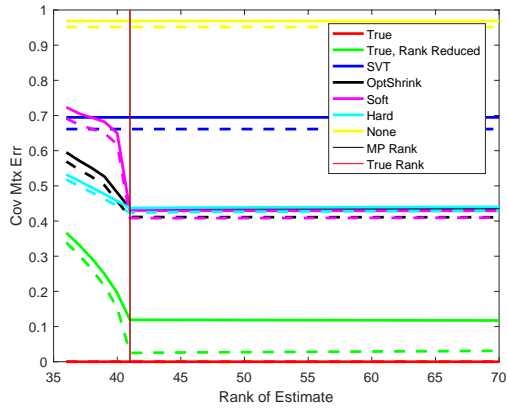
$$s_m(\mathbf{a}) = \mathbf{a}_k^2 / \text{var}_{i, i \neq k} \{\mathbf{a}_i\}. \quad (73)$$

5.3.2.3 Simulation Results

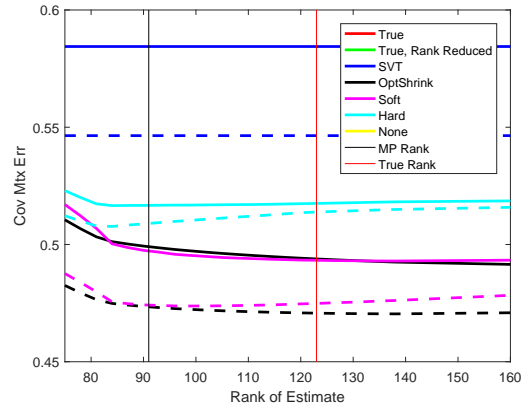
For all cases, the simulated parameters include the following: 512 range samples, 8 channels, 128 pulse CPI, CNR of 8 dB, and SIR of 2 dB. The covariance matrix accuracy results are shown in Figure 58. Apart from methods with access to truth, soft-thresholding and OptShrink provide the best performance. But this performance requires optimal selection of the diagonal loading coefficient, which can be difficult to predict. The results that only use the minimal diagonal loading (shown in solid lines) perform somewhat worse by this metric. The performance advantage of the soft threshold estimator in the minimal diagonal loading case points to its robustness to diagonal loading selection.

The precision matrix accuracy results are shown in Figure 59. In this case, all results with only minimal diagonal loading performed very poorly, because a small diagonal does not improve the condition number of the matrix and small errors in the estimate can lead to large errors in the inverse. This fact is evidenced by observing that no solid lines appear in the figure. The OptShrink method provides somewhat better performance than the soft thresholded estimator, especially at higher ranks.

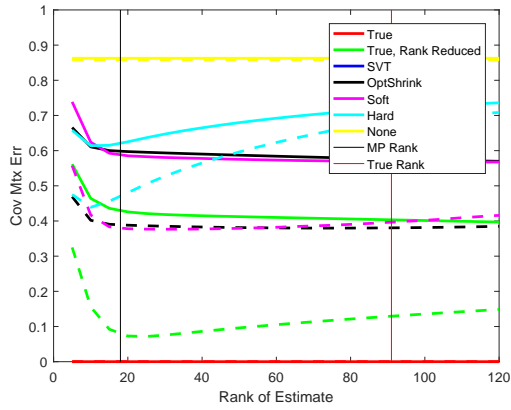
The average SINR loss, shown in Figure 60, puts the results in a context more relevant to the adaptive filtering practitioner. Since the vertical axis is loss, 0 dB at the top is best. An empirical average target signal-to-interference ratio, is the fourth and final result which is shown in Figure 61. These results both show that if one does not have access to the optimal diagonal loading, the most direct way to optimize performance is to use minimal diagonal loading (solid lines) along with the the Marčenko-Pastur rank. In that case the soft threshold, hard threshold, and Optshrink provide nearly identical performance. This selection provides near optimal results without undue downside risk associated with selecting these parameters through some other heuristic.



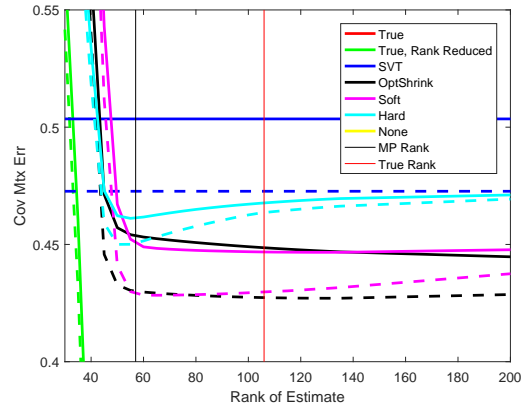
(a) Interference structure 1



(b) Interference structure 2

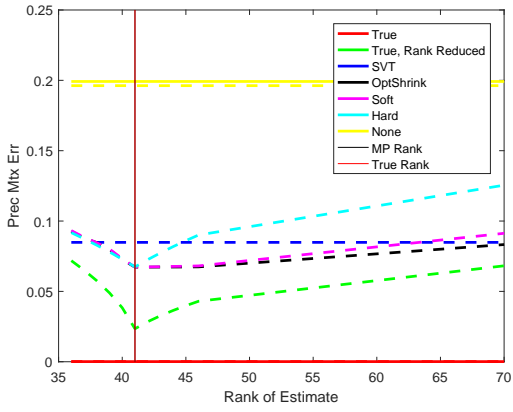


(c) Interference structure 3

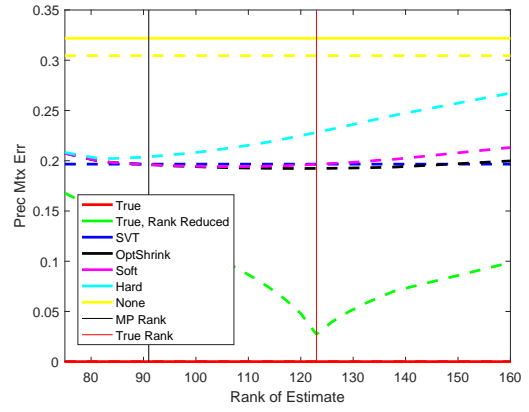


(d) Interference structure 4

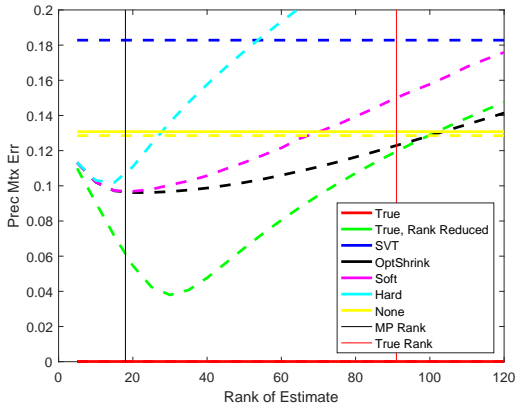
Figure 58: Covariance matrix error relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity.



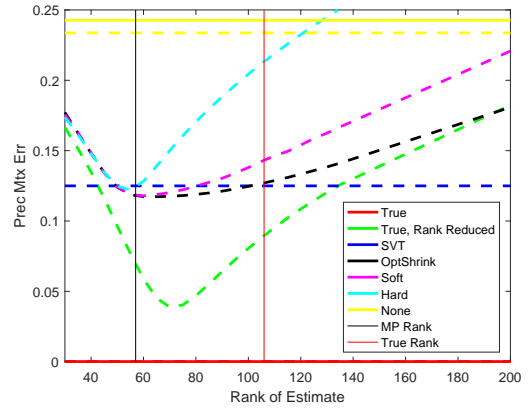
(a) Interference structure 1



(b) Interference structure 3

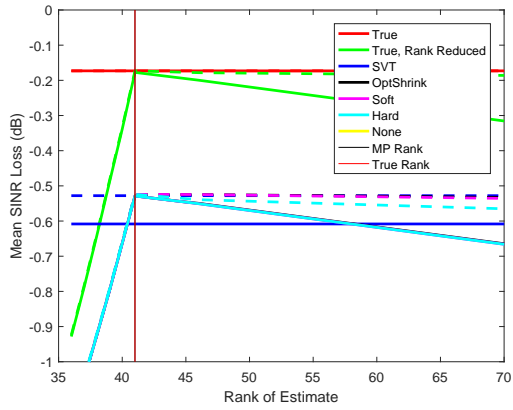


(c) Interference structure 3

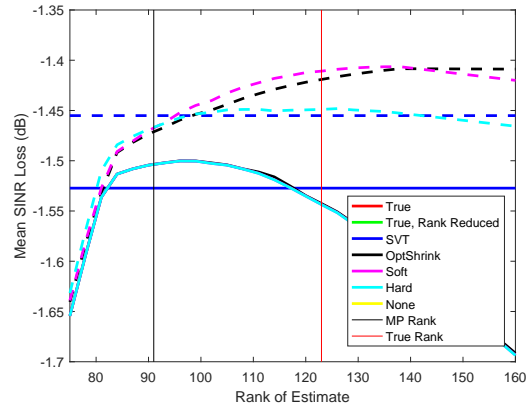


(d) Interference structure 4

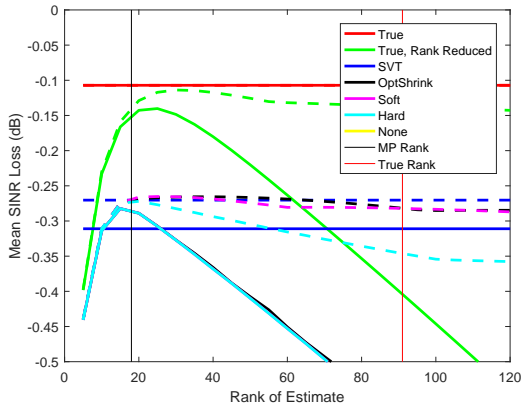
Figure 59: Precision matrix error relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large errors have been cropped for clarity.



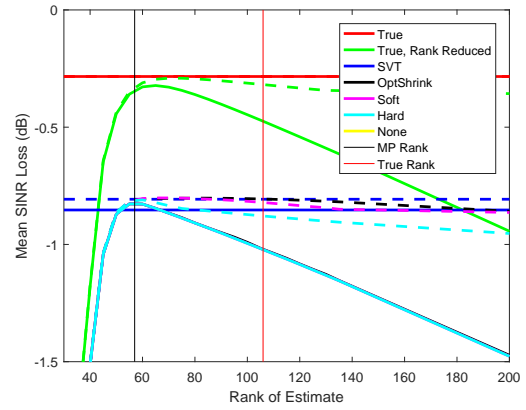
(a) Interference structure 1



(b) Interference structure 2

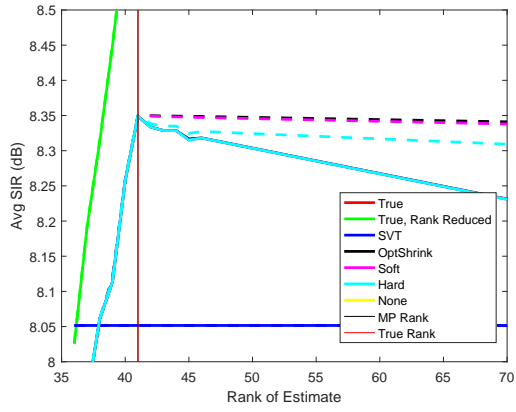


(c) Interference structure 3

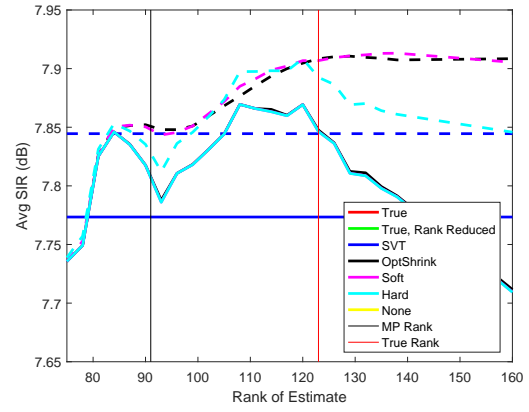


(d) Interference structure 4

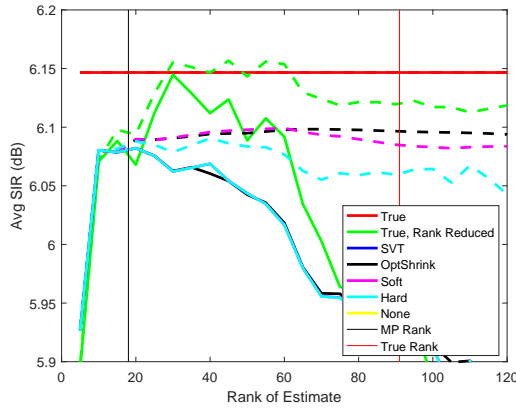
Figure 60: Mean SINR loss relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity.



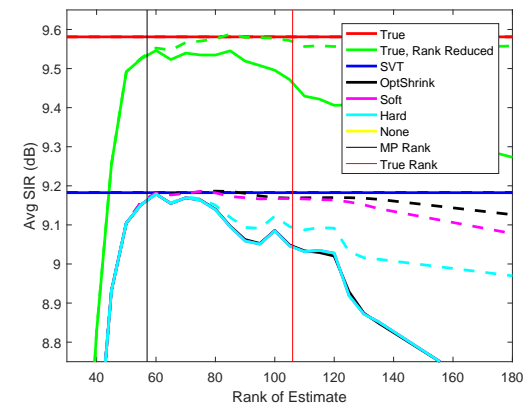
(a) Interference structure 1



(b) Interference structure 2



(c) Interference structure 3



(d) Interference structure 4

Figure 61: Signal-to-interference ratio relative to the rank of the estimate. Solid lines show the results obtained with minimum diagonal loading ($\delta = 10^{-7}$) and the dashed lines show results with the best diagonal loading at each evaluated rank. Some results with large and small errors have been cropped for clarity.

5.3.3 Summary

Based on results like that in Figure 42, which showed that even eigenvalues smaller than the largest expected noise-only eigenvalue can have significant projection into the clutter subspace, one might expect performance to be quite poor when those eigenvalues are not included in the reduced rank used to estimate the covariance. But this intuition, when tested empirically, appears not to be true. The information in those eigenvalues, though signal bearing, does not significantly improve the filtered estimate in the diagonally-loaded reduced-rank STAP estimation setting.

Filtered estimator accuracy is a slightly different figure of merit than either covariance matrix accuracy or precision matrix accuracy. Practitioners interested in this specific use case for the covariance matrix may gain some insight into performance from covariance and precision matrix accuracy, but they are not direct predictors of filter SIR.

Analysis of results from structured interference covariance estimation shows that filtered target estimates can be nearly optimized by selecting the rank of the filter equal to the number of eigenvalues that exceed the Marčenko-Pastur maximum

$$p = \text{count}_i(d_i > \sigma^2(1 + \sqrt{m/n})^2),$$

and setting the diagonal loading term to be $\delta = 10^{-8}$. There are slight gains to be achieved if the correct combination of rank (above the MP rank) and diagonal loading parameter is identified. On the other hand, downside risk exists in trying to select these parameters since a poor choice can degrade performance.

5.4 “Inverting” Sample Eigenvalues

It is clear that there exists a deterministic relationship between the true eigenvalues of a sampled distribution and the distribution of eigenvalues of a set of samples therefrom. A natural question is whether the true eigenvalues, or some characteristics of the true eigenvalues, can be inferred from the sample eigenvalues. There is, of course, not an invertible relationship between the two. But if a parameterization of the true eigenvalue distribution is assumed, the parameters may be accurately estimated using the forward model described

in the previous section.

5.4.1 Problem Statement

As before, $\mathbf{Y} = [\mathbf{y}_1 | \dots | \mathbf{y}_n] \in \mathbb{C}^{m \times n}$ where each of these n vectors is a sample from an m -dimensional normal distribution, $\mathbf{y}_i \sim \mathcal{CN}(\mathbf{0}_m, \mathbf{R}_y)$. The sample covariance matrix is $\hat{\mathbf{R}}_y = n^{-1} \mathbf{Y} \mathbf{Y}^H$ and the sample eigenvalues of $\hat{\mathbf{R}}_y$ are $S = \{s_i\}$, $i = 1, \dots, l$, where $l = \min(m, n)$, so as to ignore the unobservable eigenvalues when $n < m$. The true interference distribution has both clutter and noise components with covariance $\mathbf{R}_y = \mathbf{R}_n + \mathbf{R}_c = \nu^2 \mathbf{I}_m + \mathbf{V} \mathbf{D} \mathbf{V}^H$ where ν^2 is the measurement noise variance, $\mathbf{V} \in \mathbb{C}^{m \times k}$ with iid complex normal entries and unit-norm rows, \mathbf{D} is diagonal with entries $\sigma_1^2, \dots, \sigma_k^2$, and $k \leq m$. The eigenvalues of the true covariance matrix, \mathbf{R}_y are $\Lambda = \{\lambda_i\}$, $i = 1, \dots, l$, where each λ_i is:

$$\lambda_i^2 = \begin{cases} \nu^2 + \sigma_i^2, & i = 1, \dots, k, \\ \nu^2, & i > k \end{cases}. \quad (74)$$

These true interference eigenvalues can in turn be specified by a set of parameters $P = \{p_i\}$, $i = 1, \dots, L$, with this relationship denoted as $\Lambda(P)$. For example, see the four different decay functions in Figure 18 (on p. 57) which can be specified by the clutter-to-noise ratio and the clutter rank width.

From this true set of eigenvalues, a distribution of the observed sample eigenvalues can be computed using (65). This sample eigenvalue cumulative distribution $R(\cdot; \Lambda(P))$ can be computed for any parameterization. As a cumulative spectral distribution, R is a monotonically increasing function from the domain of positive real numbers to the range of reals in $[0, 1]$: $R(\cdot; \Lambda) : \mathbb{R}^+ \rightarrow \mathbb{R} \in [0, 1]$, $\frac{dR(x; \Lambda)}{dx} \geq 0$.

The objective then is to find the set of parameters that best matches the observed data, according to some distance measure D .

$$P^* = \arg \min_P D(S, R(\cdot; \Lambda(P))) \quad (75)$$

We define D as the area between the predicted and observed eigenvalue CDFs. This is proportional to the average distance of the observed eigenvalues from the predicted curve.

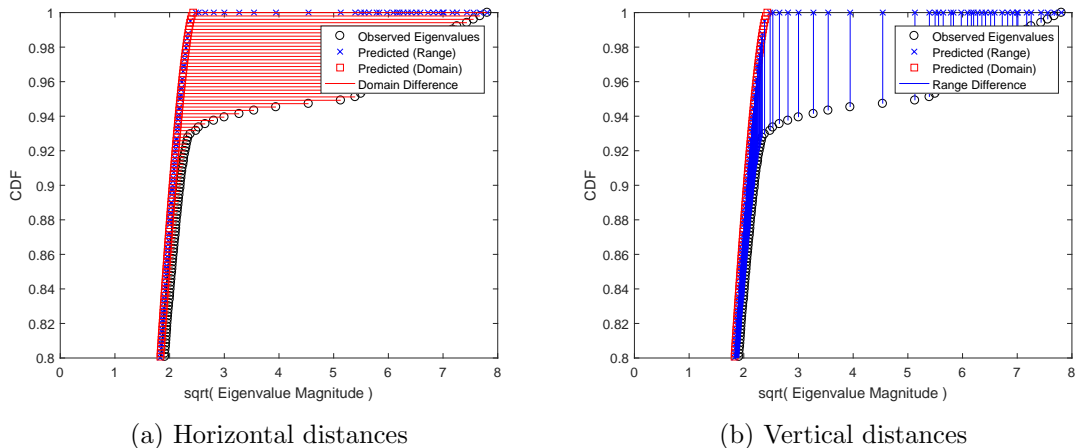


Figure 62: The two components that form the distance metric used as the minimization objective. The horizontal and vertical distances are indicated on the CDF plots.

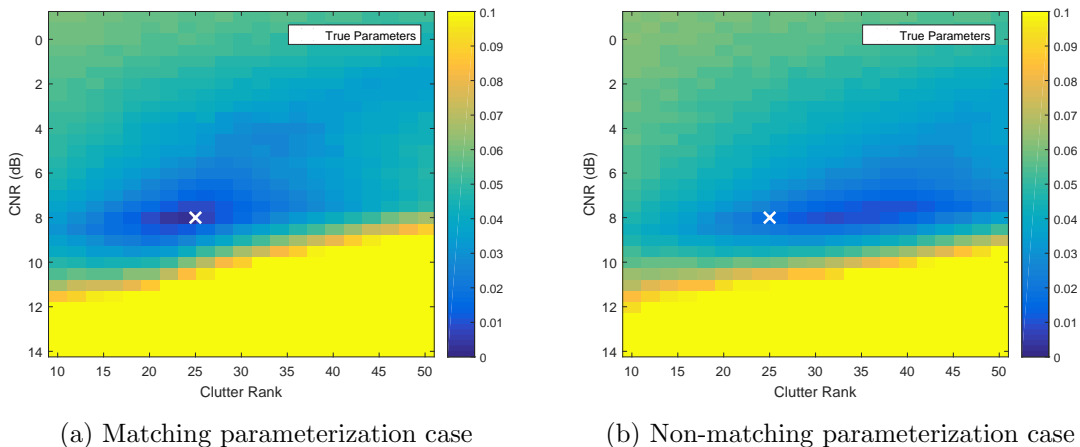


Figure 63: Distance between the observed sample eigenvalues and the predicted sample eigenvalues as a function of clutter rank and CNR.

An example of this area is shown in Figure 62, evaluated using both domain and range sampling.

The optimization surface is shown in Figure 63 for two cases. In the case where the assumed eigenvalue decay function matches the true decay, the minimum of the objective closely matches the true interference parameters; in the other, it does not.

The objective functional for this optimization is not necessarily convex. In addition, when one of the parameters is the true clutter rank, it includes an integer-valued argument. For optimization purposes we wish to allow the estimated parameters to be selected from

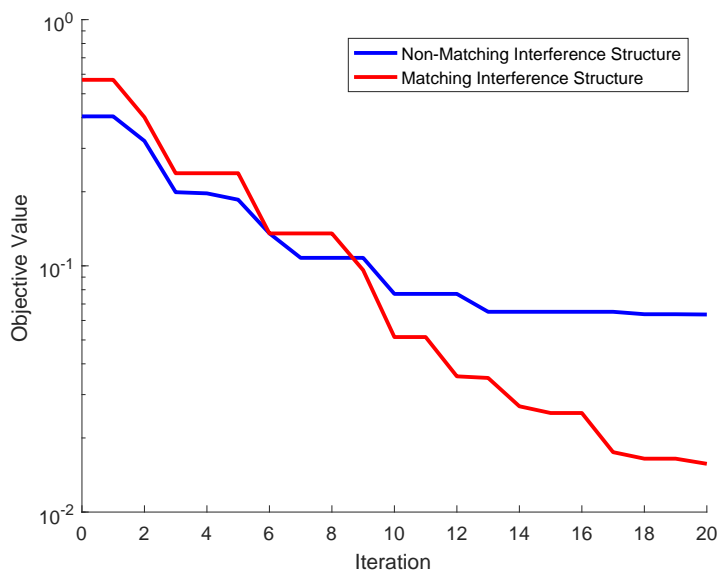


Figure 64: The optimization objective as a function of iteration. The objective is minimized effectively when the assumed parameterization matches the true structure (red case). However, when the model does not match the true structure the algorithm does not converge well (blue case). The objective here is the area between the predicted CDF and the observed eigenvalue CDF.

the set of real numbers. Since the rank of the interference can only take on integer values, we evaluate the objective twice, once for the floor of the argument and once for the ceiling, and then perform linear interpolation to calculate the result. To perform the optimization, a simple Nelder-Mead simplex algorithm [88] is used, an implementation of which is available in the MATLAB function `fminsearch`. The convergence rate for an example problem is shown in Figure 64. This algorithm is not well suited to optimization in high dimensions and can fail to converge in reasonable time with more than 4 optimization variables.

5.4.2 Applicability to Interference Estimation

If a tight parameterization of the interference structure can be identified, those parameters may be estimated from the observed eigenvalues. Uncertainties in the parameterization can inhibit convergence or give false solutions. Thus the utility of such an estimate of the true interference eigenvalues is limited. In fact, even exactly knowing the true eigenvalues is not of much benefit if it is separated from the eigenvectors. As will be shown in Section 5.3, the optimal number of dimensions to use if constructing the adaptive filter is not

necessarily equal to the number of non-noise dimensions. In other applications knowledge of the parameters that control the distribution may be of greater value. For instance in a principal components model order selection problem, a tighter estimate on the eigenvalues themselves could help bound an the optimal number of controlling parameters.

5.5 Conclusion

Random matrix theory provides powerful tools for predicting some characteristics of the sample covariance matrix. Though this theory has not been substantially developed for the compressed sensing case, it has much to say for the standard sampling case. These results have a number of uses:

- Characterizing the observability of interference dimensions embedded in noise
- Informing selection of number of dimensions used in rank-reduced STAP to maximize performance without incurring substantial performance risk
- Allowing estimation of true interference structure from samples in cases with parametrically-defined interference distribution

There are natural next steps in these areas as well. The extension of RMT to the compressed observation case would serve many working in low-rank matrix estimation. And the exploitation of existing RMT results in adaptive filtering could yield notable results in this fields.

CHAPTER VI

CONCLUSION

6.1 *Contribution*

The primary contributions of this research are:

1. Demonstration of the robustness of CS recovery in radar setting to small errors in the sensing model estimate
 - Presented at the Asilomar Conference on Signals, Systems, and Computers [20]
2. Development of an approach to exploiting the interference covariance as an adaptive filter as part of the compressed sensing optimization problem
 - Published in IEEE Transactions on Aerospace [15] and Electronic Systems and presented at the Asilomar Conference on Signals, Systems, and Computers [14]
3. Demonstration of the detection performance of the soft-thresholded matched filter as an alternative to computationally-complex convex optimization approaches in compressed sensing settings
 - Presented at the International Workshop for Compressed Sensing Applied to Radar [89]
4. Adaptation of the singular-value thresholding approach for estimating the interference covariance matrix, and demonstration of its performance on credible set of data
 - Initially presented to NATO specialists' meeting [90] and submitted to IEEE Transactions on Aerospace and Electronic Systems [91]
5. Proposal of an RMT-based rule for selecting the active rank for reduced-rank STAP which improves on other selection criteria while minimizing performance risk associated with improper selection

- Submitted to the Asilomar Conference on Signals, Systems, and Computers [92]
6. Implementation of a calculator of high-dimension sample covariance eigenvalue distributions (in the high-performance Julia language) which is suitable for use in a variety of fields including signal processing, telecommunications, statistical inference, and finance
- Published as open-source on GitHub [87] and included as an appendix to this document

6.2 Summary

The contributions developed in this work fit into the rich field of adaptive filtering in radar for detection of targets in structured interference, such as clutter. The history of this topic goes back to the earliest days of radar sensing; a fundamental advance in the ability to perform this filtering was the development of multi-channel array radars and space-time adaptive processing (STAP). Improvements to this technique have led to the state of the art in this field. In the time since STAP was developed, the field of compressed sensing has come to maturity and has changed the way sampling requirements are conceptualized. Two trends make compressed sensing and radar a well-suited match: radars move to smaller platforms, making wideband sampling more challenging, and computational resources become less expensive in cost and power. But the topics of adaptive filtering and compressively sampled radar have not been significantly examined.

To fill this gap, this work presents three related topics of study. First, the incorporation of the interference covariance matrix into the compressed sensing optimization framework is presented. This work shows the utility of accurate interference statistics in the compressed setting, extending the concepts of STAP into this domain. The proposed algorithm to accomplish this estimation, termed *covariance-aware compressed sensing*, uses the covariance matrix to warp the ℓ_2 ball of the objective to the ellipsoid specified by the error covariance matrix. This allows improved detection of targets in clutter. However, simple compressed STAP gives similar detection performance at reduced computational complexity.

Second, the estimation of the interference statistics from compressed measurements is

examined. By using the low-rank nature of the structured interference the statistics of the interference can be successfully estimated from compressed measurements. Building on the literature of matrix completion, the *singular value thresholding* algorithm is proposed for use in the adaptive filtering context. This technique is tested on synthetic structured interference, high-fidelity simulated clutter, and flight-collected radar data. In all cases, the SVT algorithm improves performance over sample matrix inversion.

Third, and finally, the underlying structure of the sample covariance estimates from the field of random matrix theory (RMT) is brought to bear on the topic. From this field, the Marčenko-Pastur, Tracy-Widom, and Stieltjes transform technique for estimating sample covariance statistics all illuminate the topic of adaptive filtering. These results were incorporated into the SVT algorithm and provide two additional results: the reliable selection of rank for reduced rank SMI and estimation of clutter parameters.

6.3 Future Work

There is much left to be done to bring the concept of an airborne compressive sampling radar to operational realization.

- A primary challenge associated with bringing these developments to widespread use is the realization of a practical compressive sampler. Although strides have been made in this regard, the hardware challenges are significant and remain a topic of active study.
- Another area for fruitful future work would be a closer connection of random matrix theory and compressed sensing. In particular, extension of the Marčenko-Pastur and the Stieltjes transform forward model results to the compressed setting would be of use to those working in various branches of low-rank matrix estimation.
- The results of random matrix theory have not been sufficiently integrated into the fields of adaptive filtering, and particularly of STAP. The results in Section 5.3 begin to address this topic, but additional work relating sample spectral distributions to

performance metrics and to processing approaches (in both compressed and uncompressed settings) would be of value.

- And computational complexity must be carefully managed for real-time operation. One avenue for advancement here may be bounding performance of estimators after a limited number of iterations, as opposed to the accuracy at convergence.

APPENDIX A

IMPLEMENTATION OF SPECTRAL DISTRIBUTION CALCULATION

The following Julia code was used to calculate the distributions of eigenvalues of sample covariance matrices in the non-null case. This code is also available as open-source on GitHub [87].

```
function sampCovEigDist(evweight, c, zvec2in=0)
# (density, zvec2, cdf) = getDensity(evweight, c, zvec2in=0)
# This function computes the expected sample spectral distribution
#
# Inputs:
# - evweight: a two-column matrix in which the first column is (samples of) the
#   true eigenvalue distribution
#   and the second column is the mass at each corresponding eigenvalue.
# - c: the sampling ratio (number of samples to dimensionality of the random
#   variable)
# - zvec2in (optional): the points at which the return distribution will be
#   discretized
# Outputs:
# - density: the eigenvalue density
# - zvec2: the eigenvalue domain at which the density is specified
# - cdf: a cumulative sum of the density
#
# Uses Stieltjes transform convergence and inversion
# Ref: Bai and Silverstein, "No Eigenvalues Outside the Support of the Limiting
#   Spectral Distribution of Large Dimensional Sample Covariance Matrices"
#   Ann. Prob., 26 (1), 316-345.
#   https://projecteuclid.org/euclid.aop/1022855421

nMbar = 5000 # sufficient for most
nZvec = 5000

evweight[:,1] = evweight[:,1]/sum(evweight[:,1])
evweight = round(evweight,6);

#generate functions to compute common integrals
computeInteg,computeIntegSq = makeIntegralFuncs(evweight)

# find start and end-points where intervals where eigenvalues live
intervalStarts,intervalEnds = getIntervals(nMbar,evweight,computeInteg)

zvec2::Vector{Float64} = []
if length(zvec2in)>1
    zvec2 = zvec2in;
    nZvec = length(zvec2)
else
    l1 = 3;
    zvec2 = linspace(0,(1.01*intervalEnds[end]).^(1/l1),nZvec+1).^l1;
    zvec2 = zvec2[2:end];
end

density = zeros(nZvec);
wentBadly = false
guess0 = -5 + 10.0im
guess = 0.0im
fminres = 0.0im
fminresOld = 0.0im
cfact = 1/min(c,1)/pi

# Calculate which points in eigenvalue sample space are in the support
```

```

islt = zeros(size(zvec2))
ielt = zeros(size(zvec2))
intervalStarts99::Vector{Float64} = intervalStarts*.99
intervalEnds101::Vector{Float64} = intervalEnds*1.01
@inbounds for iz in eachindex(zvec2)
    for ii1 in eachindex(intervalStarts99)
        if zvec2[iz]>intervalStarts99[ii1]
            islt[iz] += 1
        end
        if zvec2[iz]>intervalEnds101[ii1]
            ielt[iz] += 1
        end
    end
end

#Main Loop: loop over the domain
zast::Float64 = 0
@inbounds for iz in eachindex(zvec2)
    zast = zvec2[iz]
    if islt[iz]>ielt[iz]
        # Define the objective function and its derivative
        @inline function mobj(m::Complex{Float64})
            funcout = m + (zast-cumputeInteg(m))^-1
        end
        @inline function mobjDiff(m::Complex{Float64})
            funcout = 1 - cumputeIntegSq(m) * (zast-cumputeInteg(m))^-2.0
        end

        # Make the guess using linear prediction from last two results
        if !wentBadly && imag(fminres)>1e-8
            guess = fminres+fminres-fminresOld
        else
            guess = guess0
        end

        # Perform the optimization to find the density at this point
        fminresOld = fminres
        fminres = newton(mobj, mobjDiff, guess);
        density[iz] = abs(imag(fminres))*cfact

        # Check the result to ensure consistency
        z_recalc = -1/fminres + cumputeInteg(fminres);
        z_resid = abs(z_recalc - zast);
        if z_resid > 1e-3 && z_resid/abs(zast) > 1e-3
            wentBadly = true;
            print("some divergence in z: $(z_resid/abs(zast))");
        else
            wentBadly = false
        end
    end
end

# Calculate CDF using trapezoidal integration
cdf = trapz(zvec2,density)
if cdf[end]>1.02 || cdf[end]<1/1.02
    warn("cdf was $(cdf[end])!\n")
end
cdf = cdf/cdf[end]

return density, zvec2, cdf
end

function newton(f::Function, df::Function, x0::Complex{Float64}, tol=1e-6,
    kmax=100)
    # Textbook Newton solver
    x::Complex{Float64} = 0.0
    fx0::Complex{Float64} = 0.0
    dfx0::Complex{Float64} = 0.0
    ex::Float64 = tol + 1
    k::Int16 = 0;
    while ex > tol && k <= kmax
        k += 1
        fx0 = f(x0)
        dfx0 = df(x0)
        x = x0 - fx0 / dfx0
        ex = abs( x - x0 )/abs(x)
        x0 = x
    end
end

```

```

end
return x
end

function trapz(x,y)
# Trapezoid integration
area = zeros(size(x))
for i in 1:lx-1
    area[i+1] = area[i] + (x[i+1] - x[i]) * (y[i+1] + y[i]) * .5
end
return area
end

function makeMbarVec(evweight,nMbar)
startm = min(log10(1/maximum(evweight[:,2])/10),-3)
endm = max(-log10(1/minimum(evweight[:,2])),3)
mbarvecHalf = logspace(startm,endm,Int(nMbar/2))
mbarvec = sort([-mbarvecHalf;mbarvecHalf])
end

function makeIntegralFuncs(evweight)
# Generate the functions that will compute commonly used integral and
# its partial derivative
ut::Array{Float64,1} = unique(evweight[:,2]);
ht::Array{Float64,1} = zeros(size(ut))
for iu in eachindex(ut)
    ixEv = evweight[:,2].== ut[iu]
    ht[iu] = sum(evweight[ixEv,1])
end

uh::Array{Float64,1} = ut.*ht;
uhSq::Array{Float64,1} = (ut.^2) .* ht;
@inbounds @inline function computeInteg(mbar)
    funcsum::Complex{Float64} = 0
    @inbounds for ii2 in eachindex(ut)
        funcsum += uh[ii2]/(1+ut[ii2]*mbar);
    end
    funcsum *= c
    return funcsum
end
@inbounds @inline function computeIntegSq(mbar)
    funcsum::Complex{Float64} = 0
    @inbounds for ii3 in eachindex(ut)
        funcsum += uhSq[ii3]/(1+ut[ii3]*mbar)^2;
    end
    funcsum *= c
    return funcsum
end

return computeInteg,computeIntegSq
end

function getIntervals(nMbar,evweight,computeInteg)
# find start and stop values for each interval in which
# the spectral distribution will have non-zero mass
mbarvec = makeMbarVec(evweight,nMbar)

zvec = zeros(size(mbarvec))
for i1 in eachindex(mbarvec)
    mbar = mbarvec[i1];
    zvec[i1] = -1/mbar + computeInteg(mbar);
end

eigInds::Array{Int32,1} = zeros(size(evweight,1))
evDiff = 0.0
for uu in 1:size(evweight,1)
    evInv::Float64 = -1/evweight[uu,2]
    minval = 9999.9
    minind = 0
    for iu = eachindex(mbarvec)
        evDiff = abs(evInv-mbarvec[iu])
        if evDiff < minval
            minval = abs(evInv-mbarvec[iu])
            minind = iu
        end
    end
    eigInds[uu] = minind
end
end

```

```

isPeak = unique(eigInds);
diffzvec = diff(zvec)
picks = diffzvec .> 0
picks[isPeak-1] = false
picks[isPeak] = false
picks[isPeak+1] = false
findPicks = find(picks)
dfpUp = find(diff([-Inf;findPicks]).>1)
dfpDown = find(diff([findPicks;Inf]).>1)
zvs = zvec[findPicks[dfpUp]]
zve = zvec[findPicks[dfpDown]]
complementOfSupport = sort([zvs zve],1);
intervalStarts = complementOfSupport[1:end-1,2];
intervalEnds = complementOfSupport[2:end,1];
if length(intervalStarts)>0
    intLength = intervalEnds-intervalStarts
    intBreak = intervalStarts[2:end]-intervalEnds[1:end-1]
    if length(intBreak)>0
        intervalStarts = intervalStarts[[true;intBreak.>0]]
        intervalEnds = intervalEnds[[intBreak.>0;true]]
    end
end
intervalStarts[intervalStarts.<0]=0;
if any( intervalStarts[2:end] .<= intervalEnds[1:end-1] )
    dse = intervalStarts[2:end] - intervalEnds[1:end-1]
    takeStarts = [1;find(dse.>0)+1];
    takeEnds = [find(dse.>0);length(intervalStarts)];
    intervalStarts = intervalStarts[takeStarts];
    intervalEnds = intervalEnds[takeEnds];
end
return intervalStarts,intervalEnds
end

```

REFERENCES

- [1] BROWN, L., *A Radar History of World War II: Technical and Military Imperatives*. Institute of Physics Publishing, 1999.
- [2] DICKEY, F. R., LABITT, M., and STAUDAHER, F. M., “Development of airborne moving target radar for long range surveillance,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 27, pp. 959–972, Nov 1991.
- [3] GOETZ, L. P. and ALBRIGHT, J. D., “Airborne pulse-doppler radar,” *IRE Transactions on Military Electronics*, vol. MIL-5, pp. 116–126, April 1961.
- [4] BRENNAN, L. and REED, L., “Theory of adaptive radar,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 9, pp. 237–252, March 1973.
- [5] RICHARDS, M. A., *Fundamentals of Radar Signal Processing*. New York: McGraw-Hill, 2005.
- [6] RIHACZEK, A., “Radar signal design for target resolution,” *Proceedings of the IEEE*, vol. 53, pp. 116–128, Feb 1965.
- [7] DELONG, D. and HOFSTETTER, E., “On the design of optimum radar waveforms for clutter rejection,” *Information Theory, IEEE Transactions on*, vol. 13, pp. 454–463, July 1967.
- [8] BENEDETTO, J., KONSTANTINIDIS, I., and RANGASWAMY, M., “Phase-coded waveforms and their design,” *Signal Processing Magazine, IEEE*, vol. 26, pp. 22–31, Jan 2009.
- [9] SKOLNIK, M., *Radar Handbook, Third Edition*. New York: McGraw-Hill Professional, 2008.
- [10] SWERLING, P., “Probability of detection for fluctuating targets,” *Information Theory, IRE Transactions on*, vol. 6, pp. 269–308, April 1960.
- [11] MARCUM, J., “A statistical theory of target detection by pulsed radar,” *Information Theory, IRE Transactions on*, vol. 6, pp. 59–267, April 1960.
- [12] GUERCI, J., GOLDSTEIN, J., and REED, I., “Optimal and adaptive reduced-rank stap,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 36, pp. 647–663, April 2000.
- [13] YINGXI, Z., ZHIMING, H., and ZHULIN, Z., “Design of the high-powered digital pulse compression real-time processing system based on ADSP-TS203,” in *Radar, 2006. CIE '06. International Conference on*, pp. 1–4, October 2006.
- [14] TUUK, P. and MARPLE, S., “Compressed sensing radar amid noise and clutter,” in *Signals, Systems and Computers, Record of the Forty-Sixth Asilomar Conference on*, pp. 446–450, 2012.

- [15] TUUK, P. and MARPLE, S., “Compressed sensing radar amid noise and clutter using interference covariance information,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, pp. 887–897, April 2014.
- [16] SHNIDMAN, D., “Radar detection in clutter,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, pp. 1056–1067, July 2005.
- [17] SHNIDMAN, D., “Generalized radar clutter model,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, pp. 857–865, July 1999.
- [18] BILLINGSLEY, J., FARINA, A., GINI, F., GRECO, M., and VERRAZZANI, L., “Statistical analyses of measured radar ground clutter data,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, pp. 579–593, April 1999.
- [19] ANASTASSOPOULOS, LAMPROPOULOS, G., DROSOPOULOS, A., and REY, N., “High resolution radar clutter statistics,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, pp. 43–60, January 1999.
- [20] TUUK, P. and SHARMA, A., “Sensor calibration errors in compressive distributed-aperture radar sensing,” in *Signals, Systems and Computers, Record of the Forty-Fifth Asilomar Conference on*, pp. 2117–2121, 2011.
- [21] LIPOR, J. and BALZANO, L., “Robust blind calibration via total least squares,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4244–4248, May 2014.
- [22] MCAULAY, R., “A theory for optimal MTI digital signal processing part i. receiver synthesis,” Tech. Rep. 1972-14, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, Massachusetts, February 1972.
- [23] WARD, J., “Space-time adaptive processing for airborne radar,” Tech. Rep. 1015, MIT Lincoln Laboratory, Lexington, MA, USA, 1994.
- [24] MELVIN, W., “A STAP overview,” *Aerospace and Electronic Systems Magazine, IEEE*, vol. 19, pp. 19–35, January 2004.
- [25] GUERCI, J., *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.
- [26] REED, I., MALLETT, J., and BRENNAN, L., “Rapid convergence rate in adaptive arrays,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 10, pp. 853–863, Nov 1974.
- [27] PECKHAM, C., HAIMOVICH, A., AYOUB, T., GOLDSTEIN, J., and REID, I., “Reduced-rank STAP performance analysis,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 36, pp. 664–676, April 2000.
- [28] FA, R., DE LAMARE, R., and WANG, L., “Reduced-rank STAP schemes for airborne radar based on switched joint interpolation, decimation and filtering algorithm,” *Signal Processing, IEEE Transactions on*, vol. 58, pp. 4182–4194, August 2010.
- [29] HUBER, P. J., “Robust estimation of a location parameter,” *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.

- [30] MARONNA, R. A., “Robust M -estimators of multivariate location and scatter,” *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, 1976.
- [31] TYLER, D. E., “A distribution-free M -estimator of multivariate scatter,” *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.
- [32] CHEN, Y., WIESEL, A., and HERO, A. O., “Robust shrinkage estimation of high-dimensional covariance matrices,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 4097–4107, Sept 2011.
- [33] NADAKUDITI, R. R., “Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage,” *Information Theory, IEEE Transactions on*, vol. 60, pp. 3002–3018, May 2014.
- [34] KANG, B., MONGA, V., and RANGASWAMY, M., “Rank-constrained maximum likelihood estimation of structured covariance matrices,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, pp. 501–515, January 2014.
- [35] CANDÈS, E., ROMBERG, J., and TAO, T., “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [36] DONOHO, D., “Compressed sensing,” *Information Theory, IEEE Transactions on*, vol. 52, pp. 1289–1306, April 2006.
- [37] CANDÈS, E. J., ROMBERG, J. K., and TAO, T., “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [38] CANDÈS, E. and ROMBERG, J., “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, pp. 969–985, April 2007.
- [39] DAVENPORT, M. A., DUARTE, M. F., WAKIN, M. B., LASKA, J. N., TAKHAR, D., KELLY, K. F., and BARANIUK, R. G., “The smashed filter for compressive classification and target recognition,” in *Proc. Computational Imaging 2007*, International Society for Optics and Photonics, 2007.
- [40] STROHMER, T. and FRIEDLANDER, B., “Some theoretical results for compressed MIMO radar,” in *Signals, Systems and Computers, Record of the Forty-Fifth Asilomar Conference on*, pp. 739–743, November 2011.
- [41] PETROPULU, A., YU, Y., and POOR, H., “Distributed MIMO radar using compressive sampling,” in *Signals, Systems and Computers, Record of the Forty-Second Asilomar Conference on*, pp. 203–207, October 2008.
- [42] CHEN, C.-Y. and VAIDYANATHAN, P., “Compressed sensing in MIMO radar,” in *Signals, Systems and Computers, Record of the Forty-Second Asilomar Conference on*, pp. 41–44, 2008.
- [43] POTTER, L., ERTIN, E., PARKER, J., and CETIN, M., “Sparsity and compressed sensing in radar imaging,” *Proceedings of the IEEE*, vol. 98, pp. 1006–1020, June 2010.

- [44] ANITORI, L., MALEKI, A., OTTEN, M., BARANIUK, R., and HOOGEBOOM, P., “Design and analysis of compressed sensing radar detectors,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 4, pp. 813–827, 2013.
- [45] GURBUZ, A., MCCLELLAN, J., and SCOTT, W., “A compressive sensing data acquisition and imaging method for stepped frequency GPRs,” *Signal Processing, IEEE Transactions on*, vol. 57, pp. 2640–2650, July 2009.
- [46] SELESNICK, I., PILLAI, S., LI, K. Y., and HIMED, B., “Angle-doppler processing using sparse regularization,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 2750–2753, March 2010.
- [47] LE, B., RONDEAU, T., REED, J., and BOSTIAN, C., “Analog-to-digital converters,” *Signal Processing Magazine, IEEE*, vol. 22, pp. 69–77, November 2005.
- [48] BARANSKY, E., ITZHAK, G., WAGNER, N., SHMUEL, I., SHOSHAN, E., and ELДАР, Y., “Sub-Nyquist radar prototype: Hardware and algorithm,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, pp. 809–822, April 2014.
- [49] TROPP, J., LASKA, J., DUARTE, M., ROMBERG, J., and BARANIUK, R., “Beyond Nyquist: Efficient sampling of sparse bandlimited signals,” *Information Theory, IEEE Transactions on*, vol. 56, pp. 520–544, January 2010.
- [50] ZHEN, L., XIZHANG, W., and XIANG, L., “CS-based moving target detection in random PRI radar,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 7476–7479, July 2012.
- [51] CARIN, L., “On the relationship between compressive sensing and random sensor arrays,” *Antennas and Propagation Magazine, IEEE*, vol. 51, pp. 72–81, October 2009.
- [52] CARIN, L., LIU, D., and GUO, B., “Coherence, compressive sensing, and random sensor arrays,” *IEEE Antennas and Propagation Magazine*, vol. 53, pp. 28–39, Aug 2011.
- [53] VAIDYANATHAN, P. P. and PAL, P., “Sparse sensing with co-prime samplers and arrays,” *Signal Processing, IEEE Transactions on*, vol. 59, pp. 573–586, Feb 2011.
- [54] CANDÈS, E. J. and RECHT, B., “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [55] CANDÈS, E. J. and PLAN, Y., “Matrix completion with noise,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [56] GROSS, D., “Recovering low-rank matrices from few coefficients in any basis,” *Information Theory, IEEE Transactions on*, vol. 57, pp. 1548–1566, March 2011.
- [57] CANDÈS, E. and TAO, T., “The power of convex relaxation: Near-optimal matrix completion,” *Information Theory, IEEE Transactions on*, vol. 56, pp. 2053–2080, May 2010.
- [58] CAI, J.-F., CANDÈS, E. J., and SHEN, Z., “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

- [59] SEN, S., “Low-rank matrix decomposition and spatio-temporal sparse recovery for stap radar,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, pp. 1510–1523, Dec 2015.
- [60] CHEN, Y., CHI, Y., and GOLDSMITH, A. J., “Exact and stable covariance estimation from quadratic sampling via convex programming,” *Information Theory, IEEE Transactions on*, vol. 61, pp. 4034–4059, July 2015.
- [61] BAHMANI, S. and ROMBERG, J., “Near-optimal estimation of simultaneously sparse and low-rank matrices from nested linear measurements,” *Information and Inference: A Journal of the IMA*, vol. 5, no. 3, p. 331, 2016.
- [62] BAI, L., ROY, S., and RANGASWAMY, M., “Compressive radar clutter subspace estimation using dictionary learning,” in *Radar Conference (RADAR), 2013 IEEE*, pp. 1–6, April 2013.
- [63] SUN, K., ZHANG, H., LI, G., MENG, H., and WANG, X., “A novel STAP algorithm using sparse recovery technique,” in *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, vol. 5, pp. V–336–V–339, July 2009.
- [64] SUN, K., MENG, H., WANG, Y., and WANG, X., “Direct data domain {STAP} using sparse representation of clutter spectrum,” *Signal Processing*, vol. 91, no. 9, pp. 2222 – 2236, 2011.
- [65] POURKAMALI-ANARAKI, F., “Estimation of the sample covariance matrix from compressive measurements,” *IET Signal Processing*, vol. 10, no. 9, pp. 1089–1095, 2016.
- [66] YANG, Z., DE LAMARE, R., and LI, X., “L1-regularized STAP algorithms with a generalized sidelobe canceler architecture for airborne radar,” *Signal Processing, IEEE Transactions on*, vol. 60, pp. 674–686, February 2012.
- [67] BECKER, S., CANDÈS, E., and GRANT, M., “Templates for convex cone problems with applications to sparse signal recovery,” *Mathematical Programming Computation*, vol. 3, pp. 165–218, 2011.
- [68] BECKER, S., CANDÈS, E., and GRANT, M., “TFOCS templates for first-order conic solvers,” April 2012. <http://tfocs.stanford.edu/>.
- [69] AUSLENDER, A. and TEBoulLE, M., “Interior gradient and proximal methods for convex and conic optimization,” *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 697–725, 2006.
- [70] HAUPT, J., CASTRO, R., and NOWAK, R., “Distilled sensing: Adaptive sampling for sparse detection and estimation,” *Information Theory, IEEE Transactions on*, vol. 57, pp. 6222–6235, September 2011.
- [71] RUBNER, Y., TOMASI, C., and GUIBAS, L. J., “The earth mover’s distance as a metric for image retrieval,” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [72] TREICHLER, J., DAVENPORT, M., LASKA, J., and BARANIUK, R., “Dynamic range and compressive sensing acquisition receivers,” in *Proc. 7th U.S. / Australia Joint Workshop on Defense Applications of Signal Processing (DASP)*, July 2011.

- [73] CARLSON, B. D., “Covariance matrix estimation errors and diagonal loading in adaptive arrays,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, pp. 397–401, Jul 1988.
- [74] DYSON, F. J., “The threefold way. algebraic structure of symmetry groups and ensembles in quantum mechanics,” *Journal of Math. Physics*, vol. 3, no. 6, pp. 1199–1215, 1962.
- [75] DUMITRIU, I. and EDELMAN, A., “Matrix models for beta ensembles,” *Journal of Mathematical Physics*, vol. 43, no. 11, pp. 5830–5847, 2002.
- [76] ANDERSON, G., *An introduction to random matrices*. Cambridge: Cambridge University Press, 2010.
- [77] MARČENKO, V. A. and PASTUR, L. A., “Distribution of eigenvalues for some sets of random matrices,” *Math. of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.
- [78] BABU, B. N. S., TORRES, J. A., and MELVIN, W. L., “Processing and evaluation of multichannel airborne radar measurements (MCARM) measured data,” in *Phased Array Systems and Technology, 1996., IEEE International Symposium on*, pp. 395–399, Oct 1996.
- [79] HIMED, B. and MELVIN, W. L., “Analyzing space-time adaptive processors using measured data,” in *Signals, Systems and Computers, Record of the Thirty-First Asilomar Conference on*, vol. 1, pp. 930–935 vol.1, Nov 1997.
- [80] WIGNER, E. P., “On the distribution of the roots of certain symmetric matrices,” *Annals of Math.*, vol. 67, no. 2, pp. 325–327, 1958.
- [81] TRACY, C. A. and WIDOM, H., “On orthogonal and symplectic matrix ensembles,” *Communications in Math. Physics.*, vol. 177, no. 3, pp. 727–754, 1996.
- [82] NADAKUDITI, R. R. and EDELMAN, A., “Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples,” *Signal Processing, IEEE Transactions on*, vol. 56, pp. 2625–2638, July 2008.
- [83] EDELMAN, A. and PERSSON, P., “Numerical Methods for Eigenvalue Distributions of Random Matrices,” *ArXiv Mathematical Physics e-prints*, Jan. 2005.
- [84] SILVERSTEIN, J. and CHOI, S., “Analysis of the limiting spectral distribution of large dimensional random matrices,” *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295 – 309, 1995.
- [85] TAO, T., “Intuitive understanding of the Stieltjes transform.” MathOverflow, 2011. URL:<http://mathoverflow.net/q/79129> (version: 2011-10-25).
- [86] BEZANSON, J., KARPINSKI, S., SHAH, V. B., and EDELMAN, A., “Julia: A fast dynamic language for technical computing,” *CoRR*, vol. 1209.5145, 2012.
- [87] TUUK, P. B., “Sample covariance eigenvalue distribution.” <https://github.com/petertuuk/sampCovEigDist>, May 2017.

- [88] LAGARIAS, J. C., REEDS, J. A., WRIGHT, M. H., and WRIGHT, P. E., “Convergence properties of the Nelder–Mead simplex method in low dimensions,” *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [89] TUUK, P. and MCCLELLAN, J., “Detection performance from compressed measurements,” in *Compressed Sensing Applied to Radar, Second International Workshop on*, 2013.
- [90] TUUK, P., “Adaptive filtering in CS for structured interference suppression,” in *NATO Specialists’ Meeting SET-213: Compressive Sensing for Radar/SAR and EO/IR Imaging*, May 2014.
- [91] TUUK, P. B. and MCCLELLAN, J. H., “Estimating clutter covariance from compressed measurements with low-rank matrix approximation,” *Submitted to Aerospace and Electronic Systems, IEEE Transactions on*, 2017.
- [92] TUUK, P. B. and MCCLELLAN, J. H., “Using random matrix theory to improve radar space-time adaptive processing,” in *Submitted to Asilomar Conference on Signals, Systems and Computers*, 2017.

VITA

Peter B. Tuuk was born and raised in Grand Rapids, Michigan, earning the rank of Eagle from the Boy Scouts of America and graduating from Grand Rapids Christian High School. He attended Calvin College, where he was named captain of the men's swimming team and earned a Bachelor of Science degree in electrical engineering. Subsequently he entered graduate study at the Georgia Institute of Technology with the support of an Institute President's Fellowship, earning the degree of Master of Science in Electrical and Computer Engineering. Since then, as a doctoral candidate and a research engineer at Georgia Tech Research Institute, he has worked in the fields of compressed sensing for radio frequency applications, adaptive signal processing in sample-starved conditions, multi-target tracking, and multi-sensor fusion. He has given invited and keynote presentations in the United States and abroad on these topics. Primary publication venues for this work include the IEEE Transactions on Aerospace and Electronic Systems and the Asilomar Conference on Signals, Systems, and Computers. He expects to earn the Doctor of Philosophy degree in electrical and computer engineering in August 2017.