

Contributions to Robust Adaptive Signal Processing

Gregory Neumann Schoenig

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Lamine Mili, Chair
J. Scott Goldstein
Amir I. Zaghloul
A. A. (Louis) Beex
Daniel Spitzner
Michael Picciolo

April 12, 2007

Keywords: Adaptive Signal Processing, Robust, SINR Convergence, GM-estimator

Copyright 2003-2007, Gregory N. Schoenig

Contributions to Robust Adaptive Signal Processing

Gregory N. Schoenig

(ABSTRACT)

Classical adaptive signal processors typically utilize assumptions in their derivation. The presence of adequate Gaussian and independent and identically distributed (i.i.d.) input data are central among such assumptions. However, classical processors have a tendency to suffer a degradation in performance when assumptions like these are violated. Worse yet, such degradation is not guaranteed to be proportional to the level of deviation from the assumptions. This dissertation proposes new signal processing algorithms based on aspects of modern robustness theory, including methods to enable adaptivity of presently non-adaptive robust approaches. The contributions presented are the result of research performed jointly in two disciplines, namely robustness theory and adaptive signal processing. This joint consideration of robustness and adaptivity enables improved performance in assumption-violating scenarios – scenarios in which classical adaptive signal processors fail. Three contributions are central to this dissertation. First, a new adaptive diagnostic tool for high-dimension data is developed and shown robust in problematic contamination. Second, a robust data-pre-whitening method is presented based on the new diagnostic tool. Finally, a new suppression-based robust estimator is developed for use with complex-valued adaptive signal processing data. To exercise the proposals and compare their performance to state-of-the-art methods, data sets commonly used in statistics as well as Space-Time Adaptive Processing (STAP) radar data, both real and simulated, are processed, and performance is subsequently computed and displayed. The new algorithms are shown to outperform their state-of-the-art counterparts from both a signal-to-interference plus noise (SINR) convergence rate and target detection perspective.

To Mom, Dad, and Jackie

Acknowledgments

I wish to express sincere appreciation and gratitude to my Ph.D. committee for their advisement and effort throughout my coursework and research. I thank Dr. Lamine Mili for enabling my Ph.D. research through selfless years of teaching, research ideas and direction, and constant encouragement. I thank Dr. Amir Zaghloul for meeting with a forward-planning M.B.A. student in 2002 and for supporting my initial Ph.D. studies at Virginia Tech. I thank Dr. A.A. (Louis) Beex and Dr. Daniel Spitzner for taking the time to be on my committee and for providing perspective and information related to my research in their respective areas of expertise. I thank Dr. Scott Goldstein for his advocacy, research ideas and discussion, and for providing an excellent research environment at Science Applications International Corporation (SAIC). Finally, I thank Dr. Michael Picciolo for taking me under his wing at SAIC and spending countless hours sharing with me his knowledge of adaptive signal processing and research concepts.

I would also like to thank my coworkers at SAIC who have been supportive and inspirational to my research. In particular, I thank Mr. Michael Tinston for years of productive partnership and for allowing me to pick his brain on adaptive signal processing (Many of our best technical discussions were done on an airplane). I also thank Mr. Peter Johnson for his support and for offering his lunch hour to provide me a healthy meal and a laugh when needed.

Finally, I am eternally grateful to my parents John and Barbara Schoenig and my brother Jackie Schoenig for seeing me through the thick and thin of maturing as a researcher and as an individual. Without our supportive family bond, the Ph.D. degree would have been nearly impossible and I would not be who I am today.

Contents

1	Introduction	2
1.1	Literature Review and Prior Work	3
1.1.1	Outlier Diagnostics and the Projection Statistics Algorithm	3
1.1.2	Data Pre-whitening	6
1.1.3	GM-Estimation	7
1.2	Summary of Novel Contributions	8
1.2.1	The Adaptive Complex Projection Statistics (ACPS) Algorithm	9
1.2.2	A Robust Data Pre-whitener Based on ACPS	10
1.2.3	The GM-estimator for Complex-valued Data	12
1.3	Organization	12
2	Statistical and Adaptive Signal Processing	14
2.1	Stochastic Process Characterization	15
2.1.1	The Mean-value	16

2.1.2	The Cross-correlation and Auto-correlation Matrices	17
2.1.3	The Covariance and Autocovariance Matrices	18
2.1.4	Probability Density Function (PDF)	19
2.1.5	Independence, Correlation, and Orthogonality	20
2.1.6	Stationarity	21
2.1.7	Ergodicity	21
2.2	Statistical Signal Processing	23
2.2.1	The Wiener Filter	23
2.2.2	The Wiener Filter Innovations Process	33
2.2.3	The Distinction Between Least-Squares and the Wiener Filter	35
2.3	Adaptive Signal Processing	36
2.3.1	Adaptive Beamforming	37
2.4	Application: Space-Time Adaptive Processing (STAP) Radar	39
2.4.1	STAP Overview	41
2.4.2	Array-Vector Mathematics and Weight Vector Computation	43
2.4.3	Performance Measures	45
2.4.4	Improving SINR Convergence	47
2.4.5	Nonhomogeneity Detection in STAP	48
2.5	Consideration of Rank	53

3	Classical and Robust Estimation	55
3.1	Approaches to Handling Non-homogeneous Observations	56
3.1.1	Diagnostic Approach	56
3.1.2	Suppression / Accommodation Approach	57
3.2	Basic Statistical Estimators	58
3.2.1	Estimators of Location	59
3.2.2	Estimators of Dispersion	60
3.2.3	Estimation of Scatter	62
3.3	Classical (Parametric) Estimation Theory	62
3.3.1	Maximum Likelihood Estimators	63
3.3.2	Consistency	67
3.3.3	Fisher Information	67
3.3.4	The Cramèr-Rao Lower Bound	68
3.3.5	Efficiency	68
3.4	Robustness Theory	69
3.4.1	Robustness Concepts	72
3.4.2	Breakdown Point (BP)	72
3.4.3	Influence Function (IF)	74
3.4.4	The M-estimator	77
3.4.5	The GM-estimator	80

3.4.6	The S-estimator	83
3.4.7	The Minimum Covariance Determinant (MCD)	85
3.4.8	The Tradeoff Between Robustness and Efficiency	85
4	The Adaptive Complex Projection Statistics (ACPS) Algorithm	87
4.1	Projection Statistics for Real-valued Data	89
4.2	Projection Statistics for Complex-Valued Data	90
4.3	An Adaptive Threshold for Complex Projection Statistics	95
4.3.1	Choice of Location Estimator	96
4.3.2	Choice of Scale Estimator	97
4.3.3	Proposal of a New Statistical Test Based on CPS	98
4.4	Application and Results	99
4.4.1	SINR Convergence	101
4.4.2	STAP Application Example	112
4.4.3	Consideration of a Constant False Alarm Rate (CFAR)	120
4.5	Summary	121
5	Robust Data Pre-whitening using ACPS	122
5.1	Motivation	122
5.2	Classical Data Pre-whitening	126
5.3	Robust Data Pre-whitening based on ACPS	132

5.3.1	Procedure	132
5.3.2	STAP Example	135
5.4	Summary	139
6	The GM-estimator for Complex-Valued Data	140
6.1	The Choice of $\rho(\mathbf{r}) = \mathbf{r}^2$	142
6.2	The CGM-estimator Weighting Function	142
6.3	Regression Equalities for Complex Data	144
6.4	Simulation	146
6.4.1	Real Data	147
6.4.2	Complex Data	150
6.5	Summary	152
7	Conclusions and Future Work	153
A	Complete Results for ACPS and Robust Data Pre-whitening SINR Convergence Simulations	155
A.1	2 Dimensions	156
A.2	6 Dimensions	161
A.3	10 Dimensions	165
A.4	20 Dimensions	170
A.5	30 Dimensions	174

A.6	40 Dimensions	179
-----	-------------------------	-----

List of Figures

2.1	The Unconstrained Wiener Filter	25
2.2	The Constrained Wiener Filter with Equivalent Vector \mathbf{f}	30
2.3	The Wiener Filter with its Innovations Process	34
2.4	A Spatial Beamforming Geometry and Architecture	38
2.5	An Airborne STAP Radar Scenario (Used with permission [1])	42
2.6	The STAP Datacube with N Antenna Elements, M Pulses, and L Range Gates (Used with permission [1])	43
2.7	The Airborne STAP Spectrum (Used with permission [1])	44
3.1	Asymptotic Maximum Bias $b(\epsilon)$ as a Function of ϵ	77
3.2	The Huber ρ and ψ Functions ($b = 2$)	80
3.3	The Bisquare (Biweight) ρ and ψ Functions ($b = 2$)	85
4.1	Histogram of the Raw Projection Statistics	96
4.2	Histogram of the ACPS Values. Location and Scale have been Robustly Re- moved.	98

4.3	6 Dimension GIP Results, 1 Iteration, 1 Outlier	103
4.4	6 Dimension R-GIP Results, 10 Iterations, 1 Outlier	103
4.5	6 Dimension CPS Results, 1 Outlier	104
4.6	6 Dimension ACPS Results, 1 Outlier	104
4.7	6 Dimension GIP Results, 1 Iteration, 3 Outliers	105
4.8	6 Dimension R-GIP Results, 10 Iterations, 3 Outliers	106
4.9	6 Dimension CPS Results, 3 Outliers	106
4.10	6 Dimension ACPS Results, 3 Outliers	107
4.11	20 Dimension GIP Results, 3 Outliers	107
4.12	20 Dimension R-GIP Results, 10 Iterations, 3 Outliers	108
4.13	20 Dimension CPS Results, 3 Outliers	108
4.14	20 Dimension ACPS Results, 3 Outliers	109
4.15	STAP residue performance vs. range cell with no outlier detection. Two injected targets at range cells 307 and 310. A detection threshold at ~ 45 dB makes target detection possible.	116
4.16	STAP residue performance vs. range cell using inner-product outlier detection. Two injected targets at range cells 307 and 310. Reliable target detection.	116
4.17	STAP residue performance vs. range cell using GIP outlier detection. Two injected targets at range cells 307 and 310. Reliable target detection.	117
4.18	STAP residue performance vs. range cell using ACPS NHD. Two injected targets at range cells 307 and 310. Reliable target detection.	117

4.19	STAP residue performance vs. range cell with no NHD used. No reliable detection of targets.	118
4.20	STAP residue performance vs. range cell using simple power-based NHD. Detection of targets compromised by non-robust residuals.	118
4.21	STAP residue performance vs. range cell using GIP-based NHD Detection of targets compromised by non-robust residuals.	119
4.22	STAP residue performance vs. range cell using ACPS NHD. Reliable detection of targets.	119
5.1	Covariance Visualization for a Three-Dimensional Pure Gaussian Process . .	123
5.2	Covariance Visualization for Finite Samples Drawn From \mathbf{F}_x	124
5.3	Covariance Visualization for Finite Samples Drawn From \mathbf{F}_x . Single Contaminated Data Value.	125
5.4	Two-dimensional Data Before Classic Pre-whitening. 97.5% Confidence Ellipse Shown.	129
5.5	97.5% Confidence Ellipses Shown for Non-robust (solid) and Robust (dashed) Outlier Detection Following Classical Pre-whitening.	130
5.6	97.5% Confidence Ellipses Shown for Non-robust (solid) and Robust (dashed) Outlier Detection Following Robust Pre-whitening.	133
5.7	20 Dimension GIP Results. 3 Outliers.	136
5.8	20 Dimension R-GIP Results. 10 Iterations, 3 Outliers.	136
5.9	20 Dimension CPS Results. 3 Outliers.	137

5.10	20 Dimension ACPS Results. 3 Outliers.	137
6.1	LS and CGM-estimator Residual Performance for Coleman Data	148
6.2	Complex GM Estimator Residuals (dashed) versus LS Residuals (solid) for Simulated Six-dimension Complex Data. Three Injected Targets (CUT's 10,13,16).	150
6.3	Complex GM Estimator Output Residuals versus LS Output Residuals for Simulated Six-dimension Complex Data. Three Injected Targets (CUT's 10,13,16).	152
A.1	2 Dimension GIP Results, 1 Iteration, 3 Outliers	156
A.2	2 Dimension R-GIP Results, 10 Iterations, 3 Outliers	157
A.3	2 Dimension CPS Results, 3 Outliers	157
A.4	2 Dimension ACPS Results, 3 Outliers	158
A.5	2 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	158
A.6	6 Dimension GIP Results, 1 Iteration, 3 Outliers	161
A.7	6 Dimension R-GIP Results, 10 Iterations, 3 Outliers	162
A.8	6 Dimension CPS Results, 3 Outliers	162
A.9	6 Dimension ACPS Results, 3 Outliers	163
A.10	6 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	163
A.11	10 Dimension GIP Results, 1 Iteration, 3 Outliers	166
A.12	10 Dimension R-GIP Results, 10 Iterations, 3 Outliers	166

A.13 10 Dimension CPS Results, 3 Outliers	167
A.14 10 Dimension ACPS Results, 3 Outliers	167
A.15 10 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	168
A.16 20 Dimension GIP Results, 1 Iteration, 3 Outliers	170
A.17 20 Dimension R-GIP Results, 10 Iterations, 3 Outliers	171
A.18 20 Dimension CPS Results, 3 Outliers	171
A.19 20 Dimension ACPS Results, 3 Outliers	172
A.20 20 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	172
A.21 30 Dimension GIP Results, 1 Iteration, 3 Outliers	175
A.22 30 Dimension R-GIP Results, 10 Iterations, 3 Outliers	175
A.23 30 Dimension CPS Results, 3 Outliers	176
A.24 30 Dimension ACPS Results, 3 Outliers	176
A.25 30 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	177
A.26 40 Dimension GIP Results, 1 Iteration, 3 Outliers	179
A.27 40 Dimension R-GIP Results, 10 Iterations, 3 Outliers	180
A.28 40 Dimension CPS Results, 3 Outliers	180
A.29 40 Dimension ACPS Results, 3 Outliers	181
A.30 40 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers	181

List of Tables

4.1	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	110
4.2	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	110
4.3	10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	110
4.4	20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	111
4.5	30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	111
4.6	40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	111
4.7	Scenario 1 Injected Target Outlier Levels	112
4.8	Scenario 2 Injected Target Outlier Levels	113
5.1	Outlier Diagnostic Statistics for MD, GIP, and ACPS	131

5.2	Outlier Diagnostic Statistics on Robustly Pre-whitened Data	134
5.3	20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	138
5.4	30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	139
5.5	40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	139
6.1	Residuals for Coleman Data (Real: 1-20; Imaginary: 21-40 not shown.) . . .	149
6.2	Residuals for Input Data (Real: 1-30; Imaginary: 31-60)	151
A.1	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	159
A.2	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	159
A.3	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	159
A.4	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	159
A.5	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	160
A.6	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	160

A.7	2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	160
A.8	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	161
A.9	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	164
A.10	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	164
A.11	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	164
A.12	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	164
A.13	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	165
A.14	6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	165
A.15	10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	165
A.16	10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	168
A.17	10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	168

A.18 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	169
A.19 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	169
A.20 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	169
A.21 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	169
A.22 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	170
A.23 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	173
A.24 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	173
A.25 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	173
A.26 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	173
A.27 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	174
A.28 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	174

A.29 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	174
A.30 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	177
A.31 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	177
A.32 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	178
A.33 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	178
A.34 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	178
A.35 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	178
A.36 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level	179
A.37 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level	182
A.38 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level	182
A.39 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level	182

A.40 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level	182
A.41 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level	183
A.42 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level	183

List of Acronyms

ACPS	Adaptive Complex Projection Statistics
ADT	Average Decision Threshold
ARE	Asymptotic Relative Efficiency
BP	Breakdown Point
CDF	Cumulative Density Function
CFAR	Constant False-Alarm Rate
CPS	Complex PS
CRLB	Cramèr-Rao Lower Bound
CUT	Cell Under Test
FIR	Finite Impulse Response
GES	Gross Error Sensitivity
GIP	Generalized Inner Product
GSM	Global System for Mobile communications
ICA	Independent Component Analysis
IF	Influence Function
i.i.d.	Independent and Identically Distributed
IP	Influence of Position
IR	Influence of Residual
LMS	Least-Median-Square
LS	Least Squares
LSH	Length of the Shortest Half
MAD	Median Absolute Deviation from the median
MCARM	Multi-Channel Airborne Radar Measurement
MCD	Minimum Covariance Determinant
MD	Mahalanobis Distance
ML	Maximum Likelihood
MMSE	Minimum Mean-Squared Error
MSE	Mean-Squared Error
MWF	Multistage Wiener Filter
OS-CFAR	Order Statistics CFAR
PCA	Principal Component Analysis
PDF	Probability Density Function
PS	Projection Statistics
R-GIP	Reiterative GIP
RMB	Reed-Mallett-Brennan
SINR	Signal-to-Interference plus Noise Ratio
SMI-STAP	Sample Matrix Inversion STAP
STAP	Space-Time Adaptive Processing
ULA	Uniform Linear Array
WCDMA	Wide-band Code Division Multiple-Access

Chapter 1

Introduction

Classical statistical signal processors optimize their detection performance through a structure that maximizes the amount of signal energy that can be separated from the observed interference and noise processes, which additively create the waveform observed at a receiver. Specifically, these processors ensure the optimal probability of detection P_d is attained through the maximization of the signal to interference plus noise ratio (SINR).

In cases where only finite data is observed, adaptive versions of the asymptotic processor must be employed. Some adaptive signal processors are proven to approach the optimal solution when the assumptions hold and an adequate number of observations exist from which to estimate the interference and noise that competes with the signal of interest. However, assumptions are made in predicting a convergence to the optimal solution. Most notably, it is often assumed that the interference and noise observations observed at the receiver are Gaussian and independent and identically distributed (i.i.d.).

Conventional adaptive processors have a tendency to suffer a degradation in performance when assumptions such as these are violated. The degradation is not guaranteed to be proportional to the level of deviation from the assumptions. In fact, a small deviation in the assumptions can easily impact adaptive processor performance in a large and negative manner, rendering the classical adaptive processor deficient in its performance.

In this dissertation, robustness theory is extended and novel contributions are made which allow adaptive signal processors to exhibit a more robust performance in the presence of observations that violate the assumptions fundamental to such processors.

1.1 Literature Review and Prior Work

The contributions of this dissertation involve three major topics. These are 1) outlier diagnostics and the projection statistics (PS) algorithm, 2) data pre-whitening, and 3) GM-Estimation. The prior art for each of these areas as relevant to this dissertation is now summarized.

1.1.1 Outlier Diagnostics and the Projection Statistics Algorithm

The projection statistics (PS) algorithm for real-valued data has been developed and published in the literature. PS are able to detect outliers in multiple dimensions in a set of otherwise uncontaminated data vectors. It was developed to replace the Mahalanobis dis-

tance, which is not robust in the presence of multiple outliers.

A powerful multivariate projection method for real-valued data was independently developed by Stahel [2] in 1981 and Donoho [3] in 1982. Stahel and Donoho proposed the projection of multi-dimensional input data points onto all possible one-dimensional directions. While the one-dimensional projection concept was a valuable contribution, the prosecution of *all possible one-dimensional directions* presented a serious computational issue to any adopter of the new algorithm. In 1982, Gasko and Donoho proposed a revised algorithm aimed at investigating only those directions that originate at the coordinate-wise median and pass through one of the input data points [4]. The maximum of these projections for each input data point is termed the *projection statistic* for that data point, and represents the worst one-dimensional projection of that data point in consideration of the rest of the data points [5]. The last major contribution to the PS algorithm came in 1991 when Rousseeuw and Van Zomeren [6] performed simulations to determine appropriate cut-off values of the PS algorithm. The PS cut-off value is simply the threshold above which one declares that an outlier exists. This work was a follow-up to their 1990 work [7]. It is clear that Rousseeuw and Van Zomeren were implementing the new technique and discovering its subtleties in 1990, and working toward the goal of cut-off value determination, which they achieved in 1991.

The version of the projection algorithm, as completed by Rousseeuw and Van Zomeren, has been employed throughout the 1990's in the high-dimension application of power system state estimation. For example, Mili et al. [8] initiated a structured nonlinear regression

technique based on PS that works well in the sparse, large-scale matrices encountered in the application. This work was paralleled by the structured linear regression work and introduction and analysis of the D-estimator by Mili and Coakley [9]. The projection algorithm was termed projection statistics (PS) by this time. Important to this dissertation is the fact that the developed estimators utilized real-valued data exclusively, and not complex-valued data.

The need for robust statistical tests in adaptive signal processing, particularly in STAP radar, is evident from the volume of works found in the adaptive radar signal processing literature (for example, see [10] [11] [12] [13] [14]). The Generalized Inner Product (GIP) [15] [16] [17] [18] [19] is the most well-documented and only patented [20] outlier (termed *nonhomogeneity* in STAP) detection tool employed in STAP radar today.

However, the GIP statistical test, which is based on a zero-mean Mahalanobis distance, suffers from a fatal statistical flaw when presented with multiple outliers. Namely, it uses the sample covariance matrix, a non-robust estimator, as part of the determination of the test statistic for a STAP range cell. It is well-documented that non-robust estimators like the Mahalanobis distance (and thus the the GIP) suffer from the *masking effect* [21] [22] [7] [6] [8] [23] [24]. This effect refers to an estimator being unable to diagnose the presence of multiple outlying data points.

1.1.2 Data Pre-whitening

The pre-whitening of data is a fundamental and widespread operation in many signal processing applications. The utility of pre-whitening is mainly due to the fact that many signal processing operations require uncolored (white) data as an input (i.e., the input data are required to be uncorrelated).

The term *pre-whitening* is used to describe a data filter capable of producing a statistically uncorrelated output signal from an input signal that may be colored, or present a statistical dependence between data samples. Pre-whitening is a powerful yet simple tool in signal processing, and is often underutilized. Initial discussions concerning the need for data pre-whitening along with some early approaches date back to a 1958 work by Blackman and Tukey [25]. Several approaches to robust pre-whitening have been proposed since this time. The most relevant approaches with respect to this dissertation are now presented.

Efron et al. [26] [27] and Efron and Jeon [28] [29] present an up-front pre-whitening filter concept; however, this approach is not based on robustness theory and is not adaptive. The lack of adaptivity is evidenced by a fixed outlier diagnostic threshold of 3σ , where σ is the standard deviation of the observations. Also, the simulations presented in this work do not address higher dimensions as required by modern adaptive signal processing applications.

Mollah et al. [30] developed a robust pre-whitening technique for independent component analysis (ICA), and use an influence function approach. The draft manuscript shows that the new technique performs well in the presented two-dimension simulations. The ap-

proach however is not shown to apply to more challenging scenarios that employ data in high-dimension, and experience *masking* outliers.

1.1.3 GM-Estimation

The GM-Estimator, which stands for "Generalized M-estimator", has two primary forms common in the literature. The first was developed in an unpublished memorandum by Mallows in 1975 [31], and the second was suggested by Schweppe, Wildes, and Rom in 1970 [32].

The GM-estimator's primary advantageous trait is its bounded influence function. In situations where an M-estimator may behave poorly, the GM-estimator ensures that both the influence of position and the influence of residual are bounded. A bounded influence is desirable for use in adaptive signal processing, since outliers are common in practice. Classical processors have no such bounded influence. Thus, a single outlier can, in theory, render the adaptive processor output incorrect and perhaps even useless. Even worse, more subtle outliers can corrupt the output without a large warning sign.

The GM-estimator has previously derived for use with real-valued data. A GM-estimator for complex-valued data has not been derived in the literature.

1.2 Summary of Novel Contributions

The novel contributions in this dissertation are

1. The adaptive complex projection statistics (ACPS) algorithm,
2. The robust data pre-whitener based on ACPS, and
3. The GM-estimator for complex-valued data.

These contributions are the result of research performed jointly in robustness theory and adaptive signal processing. The origins of these fields are strikingly intertwined. However, as new developments emerge in each area, the results and conclusions do not always cross discipline boundaries. A number of factors contribute to this lack of communication, including generally separate sets of publications and conferences, the time investment required to learn the nomenclature of another discipline, and the time and experience required to truly understand the major principles that, in contrast, are second-nature to a seasoned researcher in that discipline. In short, the lack of cross-pollination between robustness and adaptive radar signal processing has left potential performance improvements and insights unexploited.

The contributions in this dissertation are a direct result of developing this missed potential into new and robust adaptive signal processing methods. Concepts from the robustness literature are developed for use within an adaptive signal processing context, resulting in novel adaptive signal processors that are more robust than the state-of-the-art yet are com-

putationally efficient. The novel contributions are now summarized.

1.2.1 The Adaptive Complex Projection Statistics (ACPS) Algorithm

The Adaptive Complex Projection Statistics (ACPS) algorithm presented herein extends the applicability of a powerful projection-based algorithm independently proposed by Stahel [2] and Donoho [3] in 1981 and 1982, respectively. This original algorithm has been further studied and is commonly known as the projections statistics (PS) algorithm. The PS algorithm is shown to robustly diagnose outliers in high-dimension data sets, and it does so with a low computational requirement. The PS algorithm currently depends on a fixed threshold value proposed in [6], based on which an observation is either kept or discarded as an outlier.

The attractive properties of the PS algorithm, when combined with the stated need for and prior work addressing complex-valued, high-dimension outliers in adaptive signal processing, constituted the motivation to develop a PS algorithm for complex-valued data. Also, any new algorithm for adaptive signal processing should operate in an adaptive fashion commensurate with the non-stationary aspect of real-world environments. Hence, a requirement for adaptivity was also noted.

The net result of the GIP and its inherent masking property with respect to STAP radar is as follows. Under realistic radar scenarios in which multiple targets are present in STAP training data, the GIP's inherent masking property is shown to degrade detection

performance, while the author’s PS-based algorithm for complex-valued data performs well under the identical scenario [23] [24]. A similar advantage of a PS-based method over GIP for use in STAP radar was shown in two-dimensions, as well [33].

No published work addressing a PS algorithm for complex-valued data exists other than those written by or involving the author [34] [33] [23] [24]. Further, the latter primary-authorships [23] [24] specifically address the problem of high-dimension complex-valued outliers such as those expected in STAP radar processing.

In this dissertation, the ACPS algorithm is exercised using STAP radar data. A greatly improved level of robustness to multiple outliers in high dimension is shown. However, it is important to note that the ACPS algorithm presented in this dissertation can apply to applications other than radar. It will especially be suitable for applications in high-dimension and those requiring the mitigation of outliers.

1.2.2 A Robust Data Pre-whitener Based on ACPS

Instantiations of data pre-whitening are numerous. The Wiener filter, for example, intrinsically performs a data pre-whitening step to generate the innovations sequence from which it then searches for a signal of interest. A second example of pre-whitening is a state-of-the-art test for outliers in STAP radar, namely the GIP presented in Section 2.4.5. The GIP performs an intrinsic conventional pre-whitening step in its determination of the diagnostic statistic for each observation. Thus, *by its very definition*, the GIP allows outliers to affect

the statistical test used to discover the outliers. Evidence of this is shown in Chapter 4.

Classical data pre-whitening can fail to produce appropriately-whitened data if the original data contains multiple outliers – especially those in positions that considerably inflate the covariance yielded by the estimator fundamental to the pre-whitener. In such scenarios, the conventionally prewhitened data’s covariance ellipsoid will not be representative of the bulk of the data. This poses potentially grave consequences for the performance of processors tasked to operate on the whitened data for applications such as detection of a target or demodulation of symbols in data communication.

This dissertation presents a new pre-whitening procedure for complex data, based on the ACPS algorithm, that is highly robust to multiple outliers in high dimension. Once the outliers are diagnosed and removed, we can safely employ standard adaptive signal processors in scenarios previously unsuitable for such processing.

In contrast to the prior work on robust data pre-whitening, the robust pre-whitening contribution in this dissertation is based on the robustness literature’s insight into the robust detection of multiple high-dimension outliers. The robust data pre-whitener proposed in this dissertation is based on ACPS from Chapter 4, which is shown to identify multiple high-dimension outliers that contaminate observed complex-valued data, regardless of their position.

1.2.3 The GM-estimator for Complex-valued Data

In order to create a robust estimator of the weight vector for adaptive signal processing applications that enjoys both the bounding of residual and position influence, a GM-estimator that operates on complex-valued data is required. As such, the final contribution of this dissertation is the derivation of a new GM-estimator for complex-valued data.

A few works by Ollila [35] [36] employ an M-estimator for complex-valued data. However, as discussed in Section 1.2.3, an M-estimator is shown in the robustness literature as having an unbounded influence of position which reveals its incapacity to handle bad leverage points, which are bad data points whose projection on the design space are outliers. In contrast, a GM-estimator enjoys a bounded influence of both residual *and* position and consequently can withstand all types out outliers, including bad leverage points.

Such a bounded total influence is desirable in adaptive signal processing applications. In this dissertation, the GM-estimator developed for real-valued data is extended for use with complex-valued data in adaptive signal processors.

1.3 Organization

In Chapter 2, statistical signal processing concepts are conveyed, including both optimal and adaptive approaches. The application of statistical signal processing to STAP radar is also presented. Relevant background material from the statistics literature is presented in

Chapter 3. This includes the the diagnostic and suppression/accommodation approaches to handling outlier data snapshots, statistical estimators used throughout the dissertation, classical and robustness theory concepts, and finally multivariate outlier accommodation estimators. In Chapter 4, the classical PS algorithm for real-valued data is extended into a new complex form and developed to be adaptive for multiple dimensions. The new algorithm is termed adaptive complex projection statistics (ACPS). The ACPS method is demonstrated in a STAP radar scenario in which SINR convergence and target identification performance are exhibited. In Chapter 5, a robust pre-whitening filter based on ACPS is developed, and examples of improved performance are presented. The new pre-whitening filter is shown to expose outlier data snapshots more robustly than the state of the art, resulting in faster SINR convergence performance than the state of the art. Constant false-alarm rate (CFAR) considerations are also discussed. Chapter 6 presents the proposed GM-estimator for complex-valued data, termed the CGM-estimator. The classical real-valued GM-estimator is extended to handle multidimensional complex-valued data as required in modern signal processing, and the new estimator is used in place of classical methods to develop a weight vector that is more robust to outlier data snapshots. Finally, conclusions are drawn and future research work is outlined in Chapter 7.

Chapter 2

Statistical and Adaptive Signal Processing

Statistical signal processing refers to signal processing applications in which the detection and/or estimation of parameters are performed. Statistical signal processing has its roots in the areas of probability, statistics, linear algebra, signals and systems theory, and digital signal processing. Many diverse disciplines and applications employ statistical signal processing; from radar and sonar to the most modern of wireless communications applications such as WCDMA, the third-generation Global System for Mobile (GSM) modulation scheme, and WiMax, the latest attempt at a more ubiquitous high-speed wireless data standard.

Statistical signal processing provides an optimal solution to a problem given a set of assumptions. Unfortunately, when employed in a real-world environment, the optimal solution

designed for i.i.d. Gaussian samples can often fail due to deviations from the assumptions. Situations such as these are where robust *adaptive* signal processing becomes vital. Adaptive signal processing, in general, extends concepts from the optimal solution under ideal assumptions onto the set of real-world environments about which our knowledge is usually quite limited and must be estimated in real-time or near real-time. The results in this dissertation are based on both statistical and adaptive signal processing concepts and robustness theory as applied to the STAP radar application, but are extensible to many other application areas.

Before discussing statistical and adaptive signal processing, a brief overview of applicable random process characterizations will be discussed.

2.1 Stochastic Process Characterization

Prior to presenting more advanced concepts, the fundamentals of estimating the mean value, correlation, covariance, autocorrelation, and autocovariance of a random process are now presented. The primary references for this discussion are works by Hayes [37] and Haykin [38]. Other quality works such as [39] exist on this subject.

Random processes and samples thereof in this section are assumed to be complex-valued unless otherwise noted.

2.1.1 The Mean-value

Let \mathbf{X} be a random process and let \mathbf{x} be a vector containing L samples of \mathbf{X} , yielding

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_L \end{bmatrix} . \quad (2.1)$$

of the random process is observed. The mean value of \mathbf{x} , termed $\mu_{\mathbf{x}}$, is defined as the expected value of \mathbf{x} , that is

$$\mu_{\mathbf{x}} = E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_L] \end{bmatrix} . \quad (2.2)$$

Assuming ergodicity (discussed shortly), the estimator of the mean, $\mu_{\mathbf{x}}$, based on the ensemble, termed the ensemble average, can be conveniently replaced by the time average.

For the discussion that follows, let \mathbf{Y} be a random process and let \mathbf{y} contain L samples of \mathbf{Y} , and have a mean of $\mu_{\mathbf{y}}$.

2.1.2 The Cross-correlation and Auto-correlation Matrices

The cross-correlation matrix of \mathbf{x} and \mathbf{y} is defined as

$$\mathbf{K}_{\mathbf{xy}} = E[\mathbf{xy}^H], \quad (2.3)$$

where $(\cdot)^H$ refers to the Hermitian transpose (transpose-complex-conjugate) operator. Each element of the $L \times L$ matrix $\mathbf{K}_{\mathbf{xy}}$ is given by

$$r(i, j) = E[\mathbf{x}_i \mathbf{y}_j^*] \quad (2.4)$$

where i and j are absolute matrix positions column-wise and row-wise, respectively, and $(\cdot)^*$ indicates conjugation of a complex quantity. Assuming the random variables \mathbf{x} and \mathbf{y} are standardized, the quantity $r(i, j)$ represents the correlation coefficient.

Also of use in this dissertation is the *autocorrelation matrix*, which is defined as

$$\mathbf{K}_{\mathbf{x}} = E[\mathbf{xx}^H], \quad (2.5)$$

where once again

$$r(i, j) = E[\mathbf{x}_i \mathbf{x}_j^*] \quad (2.6)$$

represents the correlation coefficient assuming \mathbf{x} is standardized.

2.1.3 The Covariance and Autocovariance Matrices

The correlation and autocorrelation matrices in the previous section do not account for the mean values of the random variables \mathbf{x} and \mathbf{y} . The covariance matrix, on the other hand, does this accounting. The mathematics required to include the mean value and derive the covariance matrix $\mathbf{R}_{\mathbf{xy}}$ are straightforward and are given by

$$\mathbf{R}_{\mathbf{xy}} = E[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})^H] \quad (2.7)$$

$$= E[\mathbf{xy}^H - \mu_{\mathbf{y}}\mathbf{x}^H - \mathbf{x}\mu_{\mathbf{y}} - \mu_{\mathbf{x}}\mu_{\mathbf{y}}^H] \quad (2.8)$$

$$= \mathbf{K}_{\mathbf{xy}} - \mu_{\mathbf{x}}\mu_{\mathbf{y}}^H. \quad (2.9)$$

As is easily observed from (2.9), if the sample vectors \mathbf{x} and \mathbf{y} have a mean value of zero, the covariance is equal to the correlation, as in

$$\mathbf{R}_{\mathbf{xy}} = \mathbf{K}_{\mathbf{xy}}. \quad (2.10)$$

In a similar fashion, the *autocovariance matrix* $\mathbf{R}_{\mathbf{x}}$ can be derived as

$$\mathbf{R}_{\mathbf{x}} = E[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})^H] \quad (2.11)$$

$$= \mathbf{K}_{\mathbf{x}} - \mu_{\mathbf{x}}\mu_{\mathbf{x}}^H. \quad (2.12)$$

As is easily observed from (2.12), a zero-mean process has an autocovariance equal to its autocorrelation. Formally, this fact is given by

$$\mathbf{R}_{\mathbf{x}} = \mathbf{K}_{\mathbf{x}}. \quad (2.13)$$

2.1.4 Probability Density Function (PDF)

Consider once again the dimension $L \times 1$ sample vector \mathbf{x} of the random process \mathbf{X} . The values taken on by \mathbf{x} can be said to follow a *probability distribution*, characterized by a *probability density function* (PDF) denoted by $f_{\mathbf{x}}(\mathbf{x})$.

For example, suppose that the values of \mathbf{x} are distributed according to the multidimensional complex Gaussian distribution. This implies that the values taken on by \mathbf{x} follow the multidimensional complex Gaussian PDF,

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\pi^L |\mathbf{R}_{\mathbf{x}}|} e^{-(\mathbf{x}-\mu_{\mathbf{x}})^H \mathbf{R}_{\mathbf{x}}^{-1} (\mathbf{x}-\mu_{\mathbf{x}})}, \quad (2.14)$$

where $|\cdot|$ denotes the matrix determinant. The PDF of a signal processor's input data is many times *assumed* to be of a certain type or class. This brings up the valid question, "Well, what happens when the data are not Gaussian?" This dissertation provides new tools to aid the conventional applied signal processor in detection of data points that are *outlying* with respect to the bulk of the data (which may very well be Gaussian-, or other-, distributed). In other words, the outliers are data points that do not follow $f_{\mathbf{x}}(\mathbf{x})$. More will be said on

this topic later.

2.1.5 Independence, Correlation, and Orthogonality

Assume the random variables \mathbf{x} and \mathbf{y} exist, as in the last section. The random variables \mathbf{x} and \mathbf{y} are statistically *independent* if their joint PDF can be written as the products of the marginal PDFs, that is

$$f_{\mathbf{x},\mathbf{y}}(\alpha, \beta) = f_{\mathbf{x}}(\alpha)f_{\mathbf{y}}(\beta). \quad (2.15)$$

If (2.15) does not hold, other properties may still be ascertained. For instance, if we have

$$E[\mathbf{x}\mathbf{y}^H] = E[\mathbf{x}]E[\mathbf{y}^H], \quad (2.16)$$

the random variables \mathbf{x} and \mathbf{y} are said to be *uncorrelated*. Statistical independence is a stronger condition than this, except at the Gaussian distribution where they are equivalent.

A final relevant property that two random variables may have is *orthogonality*. The random variables \mathbf{x} and \mathbf{y} are orthogonal if their inner product is zero, as in

$$E[\mathbf{x}^H \mathbf{y}] = E[\mathbf{y}^H \mathbf{x}] = 0. \quad (2.17)$$

2.1.6 Stationarity

A random variable \mathbf{x} is said to be wide-sense stationary (WSS) if its first and second moments do not vary with time. This implies that the mean $\mu_{\mathbf{x}}$ is a constant, and that the autocorrelation function $r(i, j)$ is only a function of $(i - j)$. Formally, this can be stated as

$$r(i, j) = r(j - i). \quad (2.18)$$

This property also implies that for a zero-mean, WSS random process, the covariance matrix is conjugate symmetric. Following the mathematics, this means that

$$\mathbf{R}_{\mathbf{x}} = \mathbf{R}_{\mathbf{x}}^H. \quad (2.19)$$

WSS is a relaxation of strict sense stationarity (SSS), in which all joint and marginal PDFs are time-invariant.

2.1.7 Ergodicity

An important concept for this dissertation, from a high level, is *ergodicity*. A process is termed *ergodic* in the sense of a statistical property if the time average of the samples approaches the ensemble average.

For example, a random process that is ergodic in the mean has a time-average that represents the ensemble average. In other words, a process is ergodic in the mean sense if

the mean value of a snapshot of that process is equal to the true mean of that process for all snapshots for time all t , i.e. $-\infty < t < \infty$.

Ergodicity is an implicit assumption made in adaptive processing applications that estimate statistical parameters to gain insight into the traits of that process. The reason ergodicity is important to consider is that in practice, ensemble averages are often difficult or impossible to obtain due to the fact that typically, only one realization of the process is observed.

STAP radar sensor arrays jointly consider both the temporal domain and the spatial domain. In STAP, spatial ergodicity is simply termed *homogeneity*. As with temporal ergodicity, spatial ergodicity is assumed in sensor array processors in order to realize a practical system.

In summary, adaptive processors that employ spatial and temporal degrees of freedom typically assume that temporally-determined statistics approach their ensemble averages, and that in like manner, spatially-determined statistics approach their ensemble average. Nonetheless, the properties of temporal and spatial ergodicity need to be constantly considered and evaluated in any analysis or advancement in processing data generated by a random process.

2.2 Statistical Signal Processing

Statistical signal processing has its roots in the eighteenth century, with notable contributions made throughout the early and mid twentieth century. Of note to this dissertation is Fisher's work [40], in which the notions of the maximum likelihood solution, efficiency, consistency, and the concept of *information* is introduced. Also of equal interest is the Wiener filter, first developed by Norbert Wiener during World War II [41]. The purpose of his seminal work was to unite the then-largely separate fields of time series statistics and communication engineering [41]. The Wiener filter depends on the various characterizations of a random processes discussed in the previous sections.

The primary reference for the discussion that follows is by Haykin [38], however many informative references on the subject are available, e.g., [42], [39], and [43].

2.2.1 The Wiener Filter

The Wiener filter is a fundamental building block in linear adaptive filtering [38]. It is the most fundamental linear filter for WSS processes [44]. As discussed in Section 2.1.6, A WSS process' first two moments remain constant over all realizations of the process.

In contrast with the standard white-noise matched filter, the Wiener filter allows for coloration in the received data through the use of an autocovariance matrix, to be termed simply the *covariance matrix* from here on, to match the literature on the subject. The Wiener Filter and the white-noise matched filter are equivalent when the covariance matrix

of the process equals the identity matrix, implying that the process is uncolored, or has no correlation from sample to sample.

In this dissertation, all use of the Wiener filter will be in the discrete-time sense. The discrete-time Wiener filter is conducive for modern applications as it allows for straightforward implementation in digital hardware, firmware, or software.

The Robust Wiener Filter in Spectrum Estimation

Early work with the Wiener filter was geared toward spectrum estimation. Of special interest to this dissertation is the pioneering work by Kassam and Poor and their colleagues on robust Wiener filtering. For several years in the early 1980's, [45] [46] [47] [48], research concerning robust Wiener filtering was published. The derivations and applications were mainly in the continuous time domain and were focused on single and multiple-input Wiener filters that were robustified for certain classes of spectral signals.

Today, the Wiener filter and its applications (especially STAP as studied in this dissertation) are geared toward discrete-time processes that can be computed quickly thanks to modern computing power and form factors. Nonetheless, the robust Wiener filtering work by Kassam and Poor and their colleagues in the spectral-domain paved the way for widespread understanding of the need for robustness and thus acceptance of the type of robust time-domain adaptive processing performed in this dissertation.

The Unconstrained Wiener Filter

The most basic form of the discrete-time Wiener filter is called the unconstrained Wiener filter, displayed in Figure 2.1.

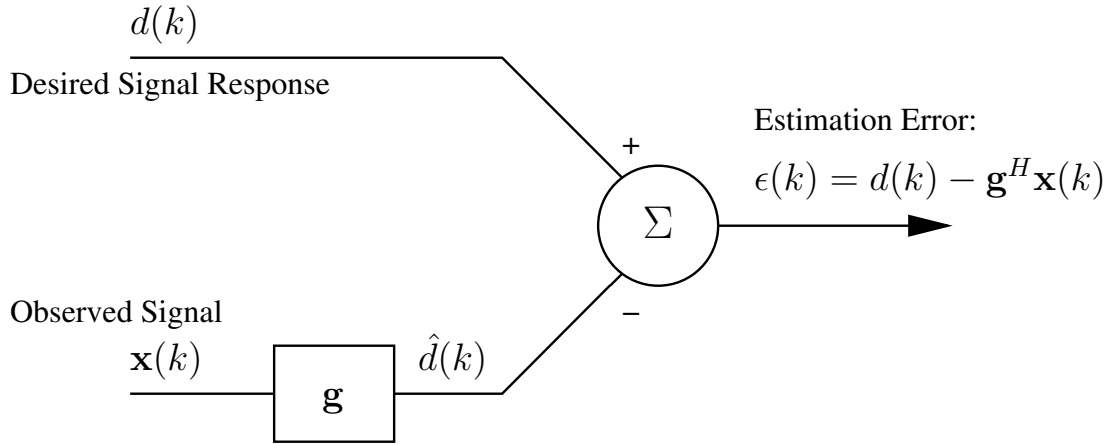


Figure 2.1: The Unconstrained Wiener Filter

The goal is to estimate the desired process $d(k)$ (a scalar) from the L -dimensional observed data $\mathbf{x}(k)$. For this discussion, $d(k)$ and $\mathbf{x}(k)$ are assumed to be WSS, zero-mean, and complex random processes. The weights \mathbf{g} are applied to k^{th} realization of \mathbf{x} (i.e., applied to $\mathbf{x}(k)$) to create a scalar estimate for time k , and the difference between the desired and estimated scalars at time k is termed the error $\epsilon(k)$ (also a scalar). Mathematically, this is given by

$$\epsilon(k) = d(k) - \mathbf{g}^H \mathbf{x}(k). \quad (2.20)$$

Presented here is the filter diagram and literal interpretation of the unconstrained Wiener

filter. The manner in which the weights \mathbf{g} are chosen depends upon the objective of the filter.

The Minimum Mean Square Error Objective

The Wiener filter's objective is to minimize the mean-squared error (MSE). The reasons for focusing on the MSE as a performance metric lie in the mathematical tractability of MSE minimization as well as the existence of a unique minimum of the MSE cost function [38]. Under the assumptions of $d(k)$ and $\mathbf{x}(k)$ being jointly WSS and Gaussian, the Wiener filter is an optimal filter.

The equations that lead to the unique, minimum mean-squared error of the unconstrained Wiener filter will now be derived. Formally, we have

$$\min_{\mathbf{g}} E \left[|\epsilon(k)|^2 \right] = \min_{\mathbf{g}} E \left[\left(d(k) - \mathbf{g}^H \mathbf{x}(k) \right) \left(d(k) - \mathbf{g}^H \mathbf{x}(k) \right)^H \right] \quad (2.21)$$

$$= \min_{\mathbf{g}} E \left[d(k)d^*(k) - \mathbf{g}^H \mathbf{x}(k)d(k) - d(k)\mathbf{x}^H(k)\mathbf{g} - \mathbf{g}^H \mathbf{x}(k)\mathbf{x}^H(k)\mathbf{g} \right] \quad (2.22)$$

$$= \min_{\mathbf{g}} E \left[\sigma_d^2 - \mathbf{g}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{g} + \mathbf{g}^H \mathbf{R}_x \mathbf{g} \right]. \quad (2.23)$$

The goal is to minimize the argument of the expectation operator J , that is

$$J = \sigma_d^2 - \mathbf{g}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{g} + \mathbf{g}^H \mathbf{R}_x \mathbf{g}. \quad (2.24)$$

To this end, we take the complex gradient [49] [38] of J with respect to the weight vector \mathbf{g}

and set it equal to zero, yielding

$$\nabla_{\mathbf{g}}(J) = \nabla_{\mathbf{g}} \left[\sigma_d^2 - \mathbf{g}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{g} + \mathbf{g}^H \mathbf{R}_x \mathbf{g} \right] = 0. \quad (2.25)$$

In [38], it is noted that $\nabla_{\mathbf{g}}(J)$ is equivalent to twice the partial derivative $\partial J / \partial \mathbf{g}^*$. This equality allows for straightforward processing of (2.25), as in

$$\nabla_{\mathbf{g}}(J) = 2(-\mathbf{r}_{xd} + \mathbf{R}_x \mathbf{g}) = 0. \quad (2.26)$$

Solving for the weight vector \mathbf{g} , we obtain, under the usual assumptions, the following *optimal* solution:

$$\mathbf{g}_{opt} = \mathbf{R}_x^{-1} \mathbf{r}_{xd}. \quad (2.27)$$

This is a matrix form of the Wiener-Hopf equations. Applying the optimal Wiener weight vector \mathbf{g}_{opt} to the input $\mathbf{x}(k)$, we obtain the best estimate of $d(k)$ from $\mathbf{x}(k)$, which is given by

$$y_{opt}(k) = \mathbf{g}_{opt}^H \mathbf{x}(k) \quad (2.28)$$

$$= \mathbf{r}_{xd}^H \mathbf{R}_x^{-1} \mathbf{x}(k). \quad (2.29)$$

The resulting filter error is optimal and is written as

$$\epsilon_{opt} = d(k) - y_{opt}(k). \quad (2.30)$$

Given this optimal filter error $\epsilon_{opt}(k)$, the *minimum* MSE can be determined as

$$\text{MMSE}_{opt} = E \left[|\epsilon_{opt}(k)|^2 \right] \quad (2.31)$$

$$= E \left[(d(k) - y_{opt}(k)) (d(k) - y_{opt}(k))^H \right] \quad (2.32)$$

$$= \sigma_d^2 + \mathbf{r}_{\mathbf{x}d}^H \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{r}_{\mathbf{x}d}. \quad (2.33)$$

It is important to note that the derivation in (2.33) depends upon the *Minimum Mean-Square Error (MMSE) principal of orthogonality* [37] [38] [44]. This principal states that due to the structure of the Wiener filter, the error signal $\epsilon(k)$ and the input data $\mathbf{x}(k)$ are uncorrelated. Thus, two of the terms that fall out of the multiplication in (2.32) have an expected value of zero, leading to the rather clean result in (2.33). Reference [37] provides an excellent treatment of the aforementioned principle of orthogonality, and [44] for a compact geometrical interpretation the same.

The Constrained Wiener Filter

A desired signal $d(k)$ is not always available *a priori*, and in such a case must be chosen based on applying constraint(s) to the solution. The constrained Wiener filter formulation

exists just for such a purpose.

Allow linear constraints to be placed on a new weight vector \mathbf{f} , and allow each of P constraints to be expressed in terms of a length L column vector, which are the columns of an $L \times P$ constraint matrix \mathbf{C} . Also allow each constraint to have a gain associated with it, with each of the L constraint gains placed in a gain vector \mathbf{v} . These constraints are placed on the weight vector \mathbf{f} as given by

$$\mathbf{C}^H \mathbf{f} = \mathbf{v}. \quad (2.34)$$

Next, determine a *blocking matrix* \mathbf{B} such that $\mathbf{C}^H \mathbf{B} = \mathbf{0}$, where $\mathbf{0}$ is the $P \times P$ zero matrix. Finally, determine a fixed $L \times 1$ constraint vector \mathbf{w}_c such that $\mathbf{w}_c = \mathbf{C}(\mathbf{C}^H \mathbf{C})^{-1} \mathbf{v}$. Using the vector \mathbf{v} and the matrix \mathbf{B} as pre-processors for an observation signal $\mathbf{u}(k)$, the analysis of the constrained Wiener filter becomes a straightforward application of the derivation presented for the unconstrained case, given a few subsequent aggregations of the observation and pre-processing elements.

Figure 2.2 shows the filter diagram of this formulation. The input data is now termed $\mathbf{u}(k)$ to preserve the nomenclature from the unconstrained case. The constraint vector \mathbf{w}_c and blocking matrix \mathbf{B} pre-process the observed signal in a way that satisfies the imposed constraint(s).

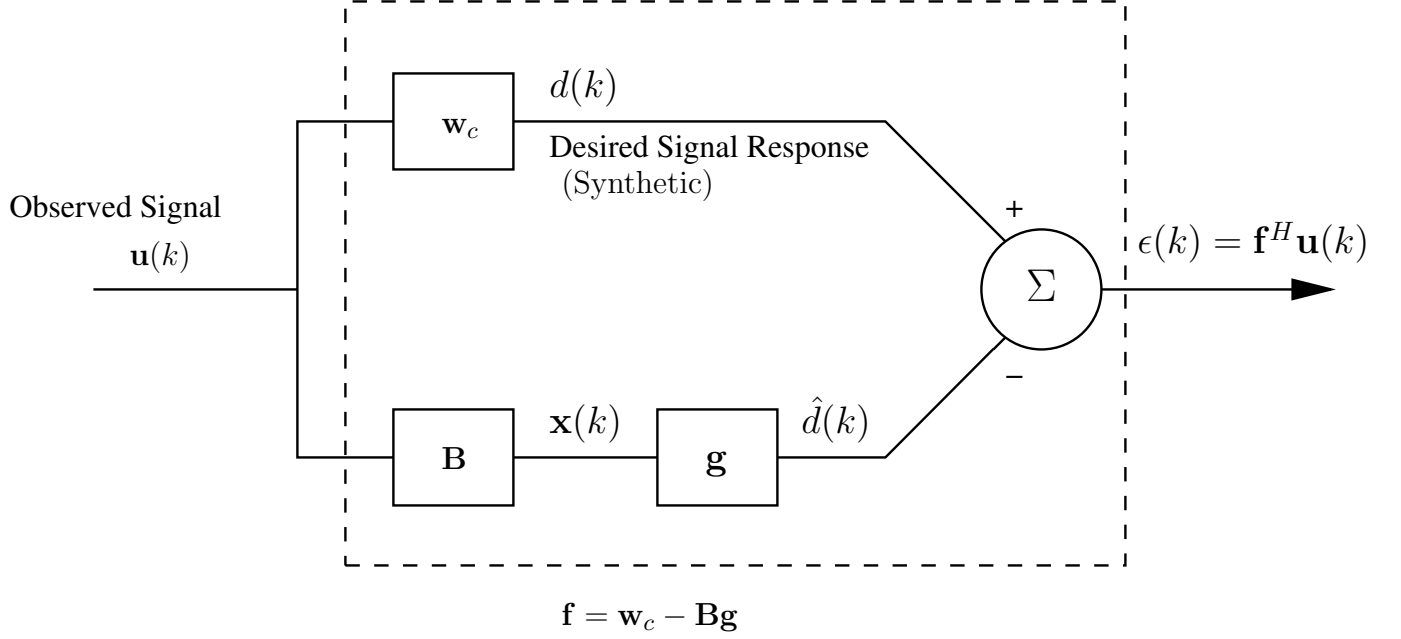


Figure 2.2: The Constrained Wiener Filter with Equivalent Vector \mathbf{f}

The MMSE Objective Revisited In like manner to Section 2.2.1, the constrained Wiener filter has the objective of minimizing the MSE between the observation and the Wiener weight-based estimate of the observation.

The development of the MSE cost function and MMSE expression can be expressed in similar form using intelligent selection of variables and quantities. First, we develop the processed signals $d(k)$ and $\mathbf{x}(k)$ based on the construction of the filter, as in

$$d(k) = \mathbf{w}_c^H \mathbf{u}(k), \quad (2.35)$$

and

$$\mathbf{x}(k) = \mathbf{B}^H \mathbf{u}(k). \quad (2.36)$$

Next, we develop the variance of $d(k)$ to be

$$\sigma_d^2 = \mathbf{w}_c^H \mathbf{R}_u \mathbf{w}_c, \quad (2.37)$$

where \mathbf{R}_u is the clairvoyant covariance matrix of the process generating the observations $\mathbf{u}(k)$. Subsequently, we determine the cross-correlation between the matrix-processed observations $\mathbf{u}(k)$ and its counterpart $d(k)$,

$$\mathbf{r}_{xd} = \mathbf{B}^H \mathbf{R}_u \mathbf{w}_c. \quad (2.38)$$

Finally, we determine the clairvoyant covariance of the input to the unconstrained weights \mathbf{g} , given by

$$\mathbf{R}_x = \mathbf{B} \mathbf{R}_u \mathbf{B}^H. \quad (2.39)$$

With these quantities in hand, the optimal MSE is given by

$$\min_{\mathbf{g}} E \left[|\epsilon(k)|^2 \right] = \min_{\mathbf{g}} E \left[\left| \mathbf{f}^H \mathbf{u}(k) \right|^2 \right] \quad (2.40)$$

$$= \min_{\mathbf{g}} \mathbf{f}^H \mathbf{R}_u \mathbf{f}. \quad (2.41)$$

Substituting for the composite weight vector $\mathbf{f} = \mathbf{w}_c - \mathbf{B}\mathbf{g}$, the MMSE for this filter is further developed as

$$\min_{\mathbf{g}} \left(\sigma_d^2 - \mathbf{g}^H \mathbf{r}_{xd} - \mathbf{r}_{xd}^H \mathbf{g} + \mathbf{g}^H \mathbf{R}_x \mathbf{g} \right). \quad (2.42)$$

As is evident, the MMSE of the constrained Wiener filter in (2.42) is in an analogous form to that of the unconstrained Wiener filter in (2.33). This feature of the constrained Wiener filter's MMSE ties in with the fact that its filter structure (shown in Figure 2.2) is comprised of an unconstrained Wiener filter preceded by the constraint preprocessors \mathbf{w}_c and \mathbf{B} .

Example: The Minimum-Variance Distortionless Response Beamformer A relevant processor based on this scheme is the *minimum-variance distortionless response* (MVDR) beamformer. The MVDR beamformer is a special case of the more general class of *linearly constrained minimum-variance* (LCMV) filters.

Consider a spatial filtering application employing an L -element uniform linear array (ULA) (using standard half-wavelength spacing). Assume that a desired sinusoid, when present, is known to have an electrical angle θ_o (direction of arrival) with respect to the array bore-sight. In this case, a desirable constraint for a Wiener filter, tasked to detect this target sinusoid, is to provide a gain of 1 (0 dB) in the expected direction of the sinusoid, when present. Formally, this is given by

$$\mathbf{C} = \mathbf{s}(\theta_o) = \left[1, e^{-j\theta_o}, \dots, e^{-j(L-1)\theta_o} \right], \quad (2.43)$$

and

$$\mathbf{v} = 1, \quad (2.44)$$

where \mathbf{s} is termed the *steering vector*. With this constraint imposed, the optimal MVDR weight vector can be developed using the method of Lagrange multipliers and is given by

$$\mathbf{g}_{MVDR} = \frac{\mathbf{R}^{-1}\mathbf{s}(\theta_o)}{\mathbf{s}^H(\theta_o)\mathbf{R}^{-1}\mathbf{s}(\theta_o)}, \quad (2.45)$$

where \mathbf{R} is the interference-plus-noise covariance matrix. In a nutshell, the MVDR filter *minimizes the variance* of signals not originating at the electrical angle θ_o , while simultaneously passing sinusoids originating at electrical angle θ_o with a gain of 1, otherwise known as a *distortionless response* to the signal of interest.. Hence the appropriate name of this common but powerful filter: *MVDR*.

2.2.2 The Wiener Filter Innovations Process

In order for the presented research to be germane to Wiener filtering in applications such as STAP radar, the outliers considered in the research must be in line with those expected in a practical setting. The robust statistics literature concerns itself with two primary outlier classes. These are *additive outliers* and *innovations outliers*. Additive outliers, as the name suggests, are simply outliers added to the observations at the input to the filter. How-

ever, to determine where the Wiener filter's innovations information is obtained, a different perspective of the Wiener filter must now be considered.

This alternate perspective of the Wiener filter is obtained by decomposing the filter into a pre-whitening component and a subsequent white-noise filter component. The optimal Wiener weight solution presented in (2.27) can be restated as

$$\mathbf{g}_{opt} = \mathbf{R}_x^{-\frac{1}{2}} \mathbf{R}_x^{-\frac{1}{2}} \mathbf{r}_{xd}. \quad (2.46)$$

With respect to applying Wiener weights to the observations, the resulting block diagram of this expansion can be viewed as shown in Figure 2.3.

The vector \mathbf{v} , a result of the observations \mathbf{x} undergoing a pre-whitening, is considered the innovations process. Frequency-domain versions of this concept are found in [50] and [51]. However, for the purpose of this dissertation, the time series perspective has been developed.

Given the straightforward outlier phenomenology associated with a single aggregated Wiener filter input as considered in this dissertation, the outliers considered will be additive since we observe \mathbf{x} and not \mathbf{v} .

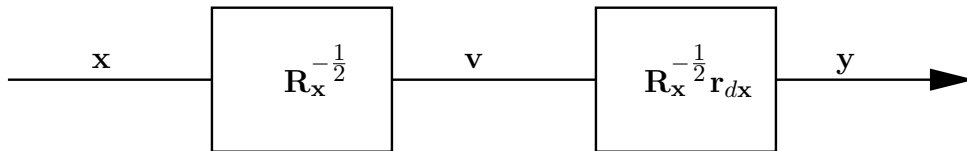


Figure 2.3: The Wiener Filter with its Innovations Process

2.2.3 The Distinction Between Least-Squares and the Wiener Filter

The discrete-time Wiener filter equations appear in a form familiar to many readers. The Wiener weight solution indeed appears similar to a Least-squares solution. There are however a few important distinctions between Wiener filter and Least-squares. First, the method of least-squares does not specifically address the stochastic nature or statistical assumptions [51] [52]. There is no expectation operator required, and there is no cross-correlation involved.

Least-squares is a procedure that creates weights based on the best linear fit to the data. Second, the data used in the least-squares problem is all least-squares uses to find a solution. In Wiener filtering, on the other hand, independent training and testing data is often used [51]. Specifically, the Wiener filter will use one set of data to develop a covariance matrix and generate Wiener weights that satisfy the MMSE criteria, while the real environment under test may not be homogenous with the training data. The critical assumption in this method is that the training data and real-world are homogeneous, which is not always the case.

If the real-world data and training data are the same data, as is the case with STAP often times, then Least-squares and Wiener filtering will provide a similar solution.

2.3 Adaptive Signal Processing

The distinction between the data-focused method of least squares and the stochastically-based Wiener filter is an excellent context from which to introduce adaptive signal processing.

In practice, observed data may not (and usually does not) perfectly follow the statistical assumptions supposed for a given problem or class of problems. Given this unavoidable truth, it is required in practice to adaptively estimate certain statistical quantities, such as the covariance matrix of a random process, that are not known *a priori*. Primarily of concern in this research (and also discussed from a statistical perspective in Chapter 3) is the sample covariance matrix (SCM) $\hat{\mathbf{R}}_x$,

$$\hat{\mathbf{R}}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H, \quad (2.47)$$

where \mathbf{x}_k is the k^{th} $N \times 1$ observation vector of process \mathbf{X} , and N is the number of degrees of freedom of the observation system.

The estimator given by (2.47) allows the use of statistical signal processing tools such as the (asymptotic) Wiener filter presented earlier.

The ramifications of substituting an estimated quantity for an actual quantity must be considered prior to employment of an application based on such a substitution in a derived statistical signal processor or filter. In this case, the SCM's statistical properties (again, discussed in Chapter 3) and probability distribution expectations must be known in order to ensure the estimator is efficient and generally appropriate for use. Deviations

from statistical assumptions can have dramatic effects on statistical performance, as will be evident later. It is desired that adaptive signal processors are made robust, so that the end users of adaptive signal processors – be they military, civilian, or private entities – have their respective expectations of performance met in a consistent and reliable fashion.

2.3.1 Adaptive Beamforming

In many adaptive signal processing applications, such as the STAP applications considered herein, vector processing is performed. As opposed to an optimal beamformer, where the statistics of the interference and noise environment are known explicitly and *a priori*, the adaptive beamformer is tasked with estimating these statistics from the received data instead of knowing them omnisciently.

An example of an adaptive spatial beamforming application, the uniform linear array (ULA) adaptive beamformer, is depicted in Figure 2.4.

By electrically sampling the elements of the ULA, which are spaced at a distance d apart, the angle of incidence θ can be determined via the geometry of the ULA. If the target is assumed to be a sinusoidal emitter with wavelength λ , the expected waveform as a function of θ is presented as *steering vector* responses s_n given by

$$s_n = e^{j2\pi(n-1)\frac{d}{\lambda}\sin\theta}. \quad (2.48)$$

The reason that s_n can be described in this manner is because a sinusoidal target is assumed

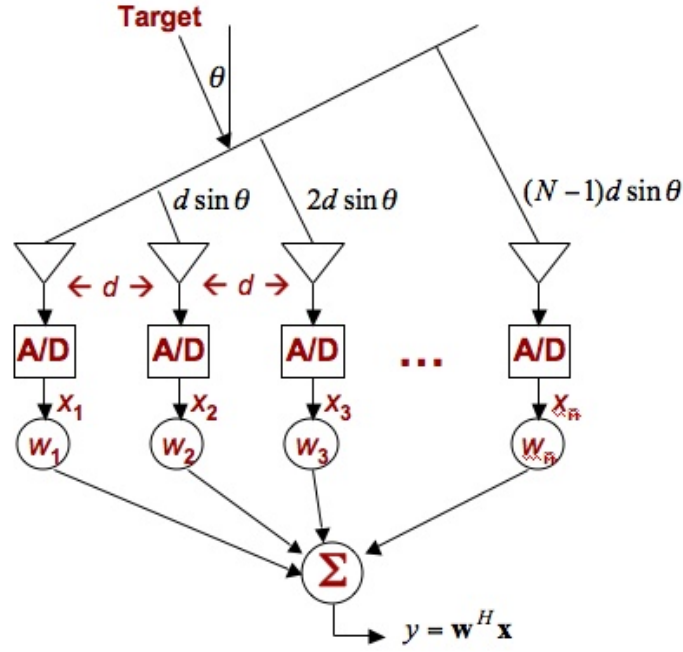


Figure 2.4: A Spatial Beamforming Geometry and Architecture

to have a linear phase progression across the ULA. Practically, this is not always the case, however for the current discussion this explanation suffices. Consequently, steering vector \mathbf{s} for a given or hypothesized incident angle θ can be presented as

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix} = \begin{bmatrix} e^0 \\ e^{j2\pi \frac{d}{\lambda} \sin \theta} \\ \vdots \\ e^{j2\pi (N-1) \frac{d}{\lambda} \sin \theta} \end{bmatrix}. \quad (2.49)$$

The response \mathbf{x} is expected to be \mathbf{s} under a Gaussian i.i.d. noise and interference profile. Thus, to search for a particular target at angle θ , one can set $\mathbf{w} = \alpha \mathbf{s}$ (where α can be set based on expected SINR) and look for a peak in the output beamformer residuals $y = \mathbf{s}^H \mathbf{x}$.

If a peak is observed, the signal of interest can be deemed present. This is essentially a matched filter, but in the spatial domain.

Analogous to this is the concept of a temporal beamformer, otherwise termed a matched filter. Under the same conditions (Gaussian and i.i.d.), a temporal steering vector $\mathbf{t}(\tau)$ can be developed to search for a target at *time* τ (rather than angle θ). The same principle applies; a peak indicated a detection of the signal of interest in the temporal domain.

In STAP, spatial and temporal domains are considered jointly. The STAP space-time steering vector is defined as the Kronecker product of the spatial and temporal steering vectors, as given by

$$\mathbf{s}_{ST} = \mathbf{s}(\theta) \otimes \mathbf{t}(\tau). \quad (2.50)$$

The resulting space-time steering vector \mathbf{s}_{ST} is constructed to search for a hypothesized target at angle θ *and* temporal offset (Doppler time shift) τ . The space-time steering vector is employed heavily in this dissertation's STAP simulations.

2.4 Application: Space-Time Adaptive Processing (STAP)

Radar

Since the beginning of the twentieth century, *RA*dio *D*etection *A*nd *R*anging (*RADAR*) has been employed and improved in order to detect and determine positional attributes of objects

of interest using radio waves rather than mere visual, human scanning. Largely credited with the earliest incarnation of radar is Christian Hülsmeyer, who in 1904 received German and British patents for his *telemobiloscope*, which was able to detect the presence of ship traffic when visual conditions such as heavy fog prohibited such information from being gathered. However, Hülsmeyer's telemobiloscope could not perform ranging; it was solely a detection and direction-finding device. The first real radars, ones that performed both ranging and detection, were later developed by American, British, German and French scientists around the time of the second World War.

The basic idea behind the modern radar is that it transmits a directional burst of electromagnetic waves in hopes that objects may be *detected* and *ranged* using the transmitted radio waves which are received by the radar's receive antenna. In basic pulsed radar for insistence, time delay from transmission to reception can be used to determine the distance (range). If the directional transmit antenna is scanned, the line of bearing can be determined by estimating the scan angle that produces the highest return of the transmitted waveform. In addition to single waveforms, multiple waveforms can be transmitted in order to determine a phase relationship between the multiple returns from a moving target, thus allowing the target closing rate to be determined.

Certain radars can also employ multiple receive antenna elements from which a line of bearing can be determined not by mechanical scanning of an antenna, but rather by electronic beamforming using the multiple receiver elements. Unfortunately, in both temporal-only and spatial-only applications of radar, unwanted radar returns from the ground, buildings, and

other non-target elements typically mix with any desired target returns (e.g., those from an enemy vehicle). Such undesired returns are known as *clutter*.

2.4.1 STAP Overview

In the early 1970's, powerful seminal works by Brennan and Reed [53] and Reed, Mallett, and Brennan [54] considered the joint use of spatial and temporal degrees of freedom. Resulting processors are generally termed "STAP". The joint consideration of spatial and temporal aspects of the radar problem include benefits such as a drastically increased probability of detection given some simple assumptions regarding the clutter, but also introduces drawbacks such as the computational burden required to run in real-time. The computing power advances since the 1970's have drastically improved the latter drawback, however it is still a serious issue to consider in designing any STAP-based system.

Many modern STAP applications are geared toward airborne radar. The airborne STAP scenario, depicted in Figure 2.5, will be considered in this dissertation. However, STAP concepts can be extended to other forms of radar as well as STAP for other domains.

A fundamental and widely-cited work on this topic is by Ward [55], from which some background discussion will now be considered. Other excellent STAP references include [56], [57], [58], and [59].

As illustrated in Figure 2.6, received STAP data is generally viewed as composing a *datacube*. This datacube contains each sample from the N antenna elements for each of the

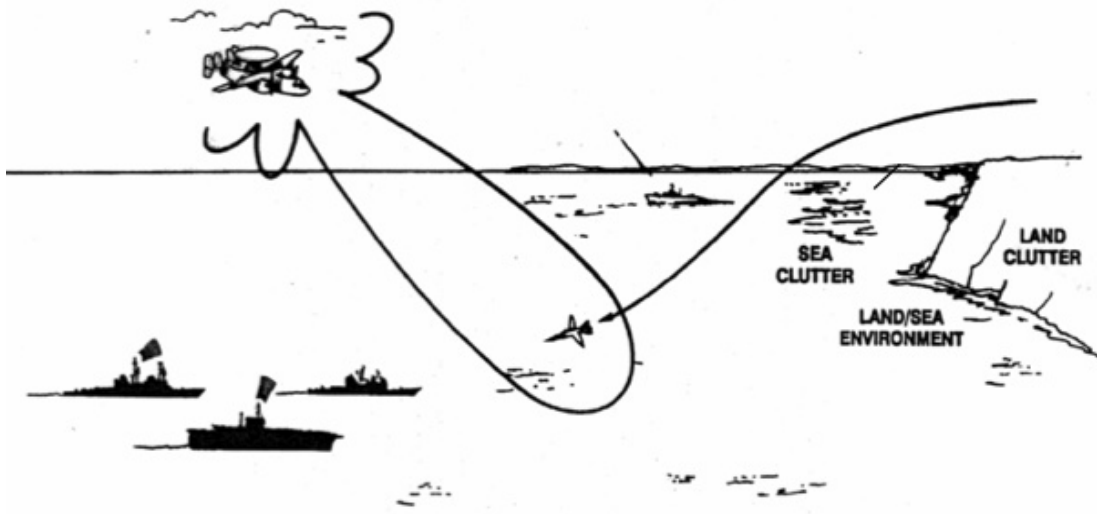


Figure 2.5: An Airborne STAP Radar Scenario (Used with permission [1])

M pulses used in the STAP scenario, and it does so for all L range bins in the STAP scenario - those representing close ranges to those furthest away.

The reason that STAP is desirable compared to spatial- or temporal-only processing is the joint consideration of spatial and temporal degrees of freedom. Rather than mainlobe clutter taking away from the ability to detect near-boresite or near-zero Doppler targets in the spatial-only or temporal-only domain, respectively, STAP opens up a detection *plane* in which many more types of targets can be identified. The STAP detection plane is only hindered by a clutter ridge; otherwise, targets can theoretically be seen where they otherwise would not if a one-dimensional filter was used.

The drawback to joint consideration of space and time is the increase in computing time. If 10 antenna elements are employed in a ULA adaptive beamformer, and coloration

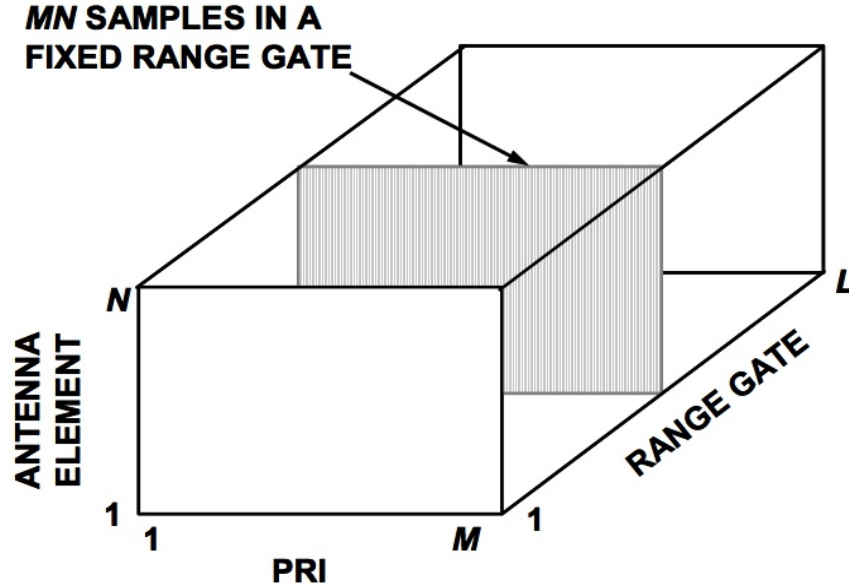


Figure 2.6: The STAP Datacube with N Antenna Elements, M Pulses, and L Range Gates (Used with permission [1])

in the observations is now assumed, the covariance matrix size is 10×10 . The situation is similar with a 10-tap temporal filter. The STAP application, however, requires that $10 \times 10 = 100$ degrees of freedom are considered, and the resulting space-time covariance matrix to be estimated is 100×100 . Reduced-rank processing can help eliminate this burden, although at some cost in and of itself. More is discussed in Section 2.5.

2.4.2 Array-Vector Mathematics and Weight Vector Computation

A fundamental result from the seminal STAP research involves the Sample Matrix Inversion (SMI) STAP processor [54, 55]. This processor provides an adaptive solution to the optimal Wiener weights for STAP, which is given by the optimal weight \mathbf{w}_{OPT} that produces the

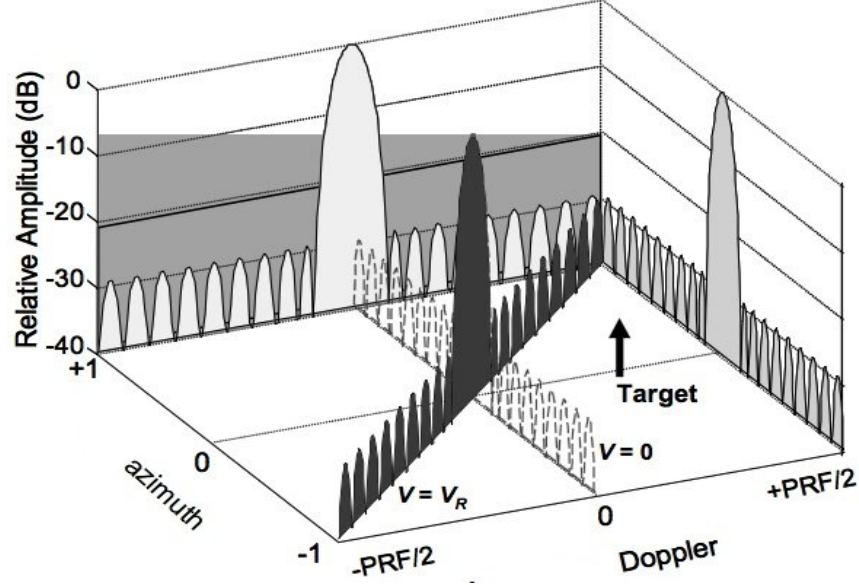


Figure 2.7: The Airborne STAP Spectrum (Used with permission [1])

maximum output SINR. This optimal weight is expressed as

$$\mathbf{w}_{OPT} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}}, \quad (2.51)$$

where \mathbf{s} is the known steering vector for the target of interest, \mathbf{R} is the true covariance matrix that describes the statistics of the background clutter, jamming and noise for the range cell under test (CUT). In practice, optimal SINR performance is not achievable because an infinite number of stationary, i.i.d. and Gaussian training data snapshots must exist for an estimate of \mathbf{R} to approach the true \mathbf{R} , assuming ergodicity.

In [54], it was determined that if a STAP processor has N space-time degrees of freedom, then $K = 2N - 3$ stationary i.i.d. Gaussian training data snapshots \mathbf{x}_k are required for near optimal convergence performance when using (2.47) to develop the maximum likelihood

covariance matrix estimate $\hat{\mathbf{R}}_{SMI}$ in lieu of the optimal (asymptotic) covariance \mathbf{R} given by (2.51), as in

$$\hat{\mathbf{R}}_{SMI} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H, \quad (2.52)$$

where \mathbf{x}_k is an $N \times 1$ snapshot vector. The resulting adaptive weight vector is

$$\hat{\mathbf{w}}_{SMI} = \frac{\hat{\mathbf{R}}_{SMI}^{-1} \mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}_{SMI}^{-1} \mathbf{s}}. \quad (2.53)$$

Specifically, the use of \mathbf{w}_{SMI} yields an average SINR performance within 3 dB of the optimal performance using $K = 2N - 3$ training snapshots for *any* true covariance matrix \mathbf{R} if the assumptions of the data processed by the estimator $\hat{\mathbf{R}}_{SMI}$ are strictly satisfied.

2.4.3 Performance Measures

In this section, the key STAP performance measures are presented. These are signal-to-interference-plus-noise ratio (SINR) and mean-squared error (MSE).

Signal-to-Interference Plus Noise Ratio (SINR)

The maximization of output SINR is shown to maximize the probability of target detection when i.i.d. Gaussian noise and interference are competing with the target signal [53]. Thus, the SINR is an important metric in adaptive filtering.

The SINR is accurately described by its title. Namely, it is the ratio of the observed

signal power to the observed power of the noise and interference combined. Formally, this is given by

$$SINR = \frac{|\hat{\mathbf{w}}^H \mathbf{s}|^2}{\hat{\mathbf{w}}^H \hat{\mathbf{R}}_X \hat{\mathbf{w}}}, \quad (2.54)$$

where $\hat{\mathbf{w}}$ is the sample (SMI-developed) weight vector from (2.53), and $\hat{\mathbf{R}}_X$ represents the sample covariance matrix developed for the observations $\mathbf{X} = [x_1, x_2, \dots, x_k]$, and \mathbf{s} is the steering vector of interest.

Mean-Squared Error (MSE)

The optimal mean-square error is expressed as

$$MSE_{opt} = \mathbf{g}_o^H \mathbf{R} \mathbf{g}_o, \quad (2.55)$$

where the subscript 'o' refers to the optimal weight vector for the clairvoyant covariance matrix \mathbf{R} . A realizable MSE in simulation is

$$MSE_{realizable} = \hat{\mathbf{g}}^H \mathbf{R} \hat{\mathbf{g}}. \quad (2.56)$$

A possibly misleading but often-implemented MSE is given by

$$MSE_{fictional} = \hat{\mathbf{g}}^H \hat{\mathbf{R}} \hat{\mathbf{g}}. \quad (2.57)$$

The reason this MSE is misleading is that the training data for $\hat{\mathbf{R}}$ is the same data used to create $\hat{\mathbf{g}}$. This estimator for MSE has been shown to lead to a higher-than predicted SINR (and equivalently under the i.i.d. Gaussian assumption, a lower-than predicted MSE) than is truly realizable under the interference and noise environment [52][60]. While some work in this area has been performed [61], a deeper investigation into the matter is warranted.

2.4.4 Improving SINR Convergence

Fast convergence in SINR is a fundamental goal of adaptive processors, as data are typically scarce and must be processed quickly. In the simulation presented here, the average SINR is computed by averaging the SINR over M Monte Carlo runs as given by

$$\text{Avg SINR} = \frac{1}{M} \frac{1}{\text{SINR}_{opt}} \sum_{i=1}^M \frac{|\hat{\mathbf{w}}_i^H \mathbf{s}|^2}{\hat{\mathbf{w}}_i^H \hat{\mathbf{R}}_i \hat{\mathbf{w}}_i}, \quad (2.58)$$

where SINR_{opt} is a normalizing factor that allows the asymptotic SINR to be exactly 0 dB.

It is the reciprocal of the MMSE, which is

$$\text{MMSE} = \mathbf{w}_{opt}^H \mathbf{R}_{opt} \mathbf{w}_{opt}, \quad (2.59)$$

where the subscript *opt* refers to the optimal, or asymptotic value of the subscripting quantity. Thus, the asymptotic (optimal) SINR is given by

$$\text{SINR}_{opt} = \frac{1}{\mathbf{w}_{opt}^H \mathbf{R}_{opt} \mathbf{w}_{opt}} = \mathbf{s}^H \mathbf{R}_{opt}^{-1} \mathbf{s} \quad (2.60)$$

for a target power of unity (assumed here). In the case where the data is in *canceler form* (see Picciolo [1], Picciolo et al.[33], and Picciolo and Gerlach [62] for recent thorough work in canceler form), then $\mathbf{s} = [1, 0, 0, \dots, 0]^T$ and thus the optimal SINR (and MMSE) can be found via the (1,1) position of the covariance matrix of the canceler-form data.

2.4.5 Nonhomogeneity Detection in STAP

Detection of statistical outliers in STAP is termed Nonhomogeneity detection (NHD). NHD techniques typically involve the computation of a distance metric for each training data snapshot in comparison to other training data. Several distance metrics are popular for parameter estimation in a wide range of applications. These include the Kolmogorov-Smirnov distance [63], the Kullback-Leibler distance [63], and the Mahalanobis distance [64]. In this section, the state-of-the-art statistical tests used in STAP to detect nonhomogeneities (outliers) are presented.

Inner-Product Power Test

A test based on simple power estimation is used as a baseline comparison. If the complex inner product of a snapshot with itself exceeds a threshold, the snapshot is declared an outlier. Failure of such a simple test to detect outlier snapshots in the 0 dB – 5 dB level above the interference level will highlight the need for more sophisticated NHD techniques.

The power test statistic P_{CUT} is the complex inner-product of the CUT snapshot \mathbf{x}_{CUT} , namely

$$P_{CUT} = \mathbf{x}_{CUT}^H \mathbf{x}_{CUT}. \quad (2.61)$$

This value is then normalized by the mean of the powers of all range cells in the STAP training data window. The result is then compared to a threshold,

$$P_{CUT} > threshold \quad , \quad (2.62)$$

and the snapshots corresponding to P_{CUT} values exceeding the threshold are removed from the training data. Note that after the outlier removal process, in the subsequent SMI STAP covariance estimation using (2.52), the value of K is reduced by the number of excised training snapshots.

In [65], it is well-summarized why the inner-product power test is generally not a practical approach for NHD in adaptive filtering. A primary reason involves the expected value

of the inner product in (2.61), namely

$$E[P_{CUT}] = \text{trace}(\mathbf{R}_i). \quad (2.63)$$

The issue is that two cells under test, say \mathbf{x}_{CUT1} and \mathbf{x}_{CUT2} , can be drawn from random processes exhibiting two extremely different covariance matrices, but happen to have the same trace. Thus the inner-product power test is not of much value in adaptive processing except for in CFAR processors that utilize a scalar version of the inner-product.

Generalized Inner Product (GIP)

Based on the Mahalanobis distance, the GIP [15] is a technique used to detect outlying points in multidimensional data. The GIP determines whether a particular data vector \mathbf{x} has the same (or similar) covariance to that used in the GIP computation. Using the sample covariance matrix $\hat{\mathbf{R}}$, the GIP is computed as

$$GIP_i = \mathbf{x}_i^H \hat{\mathbf{R}}^{-1} \mathbf{x}_i. \quad (2.64)$$

The value of GIP_i is then compared to a threshold to decide if \mathbf{x} is an outlier with respect to $\hat{\mathbf{R}}$. The 97.5 percentile value of the χ^2 distribution with N degrees of freedom, where N is the number of dimensions of \mathbf{x} , is used as the threshold. The fact that GIP values are distributed as χ_N^2 is shown in [16].

In practice, training data snapshots for STAP often are scarce, and thus obtaining performance within a specified tolerance from optimal is challenging. Further, strong targets, clutter discretizes, sudden changes in clutter distribution, and glint are examples of nonhomogeneities that can corrupt this scarce set of otherwise homogeneous training data samples. The incorporation of such nonhomogeneities in the training data leads to a biased covariance matrix estimate $\hat{\mathbf{R}}_{SMI}$, and the resulting SMI-STAP weight vector \mathbf{w}_{SMI} formed in (2.53) may not yield a processor with average SINR performance within 3 dB of optimal for a reasonable number of samples K . In fact, the SINR performance can be far from optimal in many cases, and take considerably more than $2N$ samples to approach a near optimal SINR. This degraded SINR performance generally results in a lower probability of target detection P_d .

Successful detection and excision of nonhomogeneous training snapshots often is a critical step to maintaining adequate SINR performance in STAP radar systems. Techniques to perform NHD for STAP have been widely studied. The GIP [15, 62] is a commonly-used NHD technique capable of detecting outliers in certain situations. However, as will be shown empirically in Chapter 4, this technique can fail in the presence of strong outliers, in particular those snapshots having powers in the range 0 dB to 5 dB above the interference power. Such nonhomogeneities are strong, but not strong enough to be detected reliably by a simple power test and thus are problematic to training data selection. Lower power outliers are not examined in this work.

In particular, the GIP method of NHD can suffer from the *masking effect* when more

than one outlier is present, as noted in the literature [8]. The masking effect occurs when non-robust estimators are used in the process of detecting multiple outlier snapshots. Because multiple outliers exist, it is not generally possible to exclude them all from the training data when performing any NHD for a particular test snapshot and the NHD test may therefore become biased. The underlying covariance estimator used in GIP statistic calculations forms a sample covariance matrix that has a zero breakdown point (i.e. in theory, a single outlier can unboundedly bias the estimator), thus implying that the GIP statistic is non-robust.

It can be shown that the GIP is equal to the square of the Mahalanobis distance for zero-mean data. The GIP statistic is a scalar quantity that may be compared to a threshold to determine if the corresponding N -dimensional snapshot vector is an outlier with respect to the training data. Specifically, for a test data vector \mathbf{x}_{CUT} , the remaining training data samples (except for three “guard” samples on either side of \mathbf{x}_{CUT} for the simulations in this work) are used to form the sample covariance matrix $\hat{\mathbf{R}}_x$ given by (2.52). To test if \mathbf{x}_{CUT} is an outlier, the GIP statistic is formed as

$$GIP_{CUT} = \mathbf{x}_{CUT}^H \hat{\mathbf{R}}_x^{-1} \mathbf{x}_{CUT} \quad . \quad (2.65)$$

The resulting value is guaranteed to be a scalar since $\hat{\mathbf{R}}_x$ (and thus $\hat{\mathbf{R}}_x^{-1}$) is positive semi-definite [66]. The resulting statistic is compared to a threshold as

$$GIP_{CUT} > threshold \quad , \quad (2.66)$$

and snapshots with a GIP statistic exceeding the threshold are removed from the training data in subsequent SMI STAP processing.

Reiterative (Multi-pass) GIP

The GIP has been extended to a multi-pass version by one of its inventors [65]. Essentially, the technique in the previous section is iterated to improve convergence. Even though the GIP is not robust, the reiterative, or multi-pass version of the GIP does present improved performance on average when compared to one-pass GIP.

There is also a version of Multi-pass GIP that employs a new adaptive threshold [67]. However, as will be seen in Chapter 4 and 5 of this dissertation, the Mahalanobis distance and thus the GIP are prone to the masking effect and thus may present poor outlier diagnostic performance. In such situations, the threshold value, be it adaptive or not, is *irrelevant* due to the lack of a gap between statistics of the outliers and the statistics of the bulk of the data.

2.5 Consideration of Rank

A fully-adaptive STAP radar employs all $L = M \times N$ degrees of freedom to compute the STAP weight vector. In doing so, the received $L \times 1$ STAP snapshots \mathbf{x}_i are used to create a sample covariance matrix of dimension $L \times L$. As the inversion of a matrix of this size is computationally intensive, specifically on the order of $O(L^3)$, and also due to the lack of

enough training data for the high-dimension full-rank STAP case, fully-adaptive STAP is at times not practical [55].

To enable feasibility in large-dimension sensor array problems, an area of research on *partially-adaptive* STAP has been active since the late 1970's. Partially-adaptive STAP is focused on processing in a dimension $P \ll L$ and obtaining results similar to (or better than, in cases) full-rank STAP. Early work by Morgan [68], Van Veen and Roberts [69], and Van Veen [70] were precursors to the breakthrough work by Goldstein and Reed [71] [72] [73] and Goldstein et al. [74] [75] [76] at the very end of the twentieth century. The Multistage Wiener Filter (MWF), a main contribution of Goldstein in his Ph.D. thesis [44], has become the focus of many recent works in radar, communications, and global positioning interference mitigation. The MWF allows for superior detection and false-alarm performance with a lower sample support requirement compared to that of conventional reduced-rank techniques such as those based on principal component analysis (PCA). In addition, the MWF is able to attain such performance under a broad range of interference environments.

The focus of this author's dissertation is on new and novel statistical methods for handling a broad range of input data characteristics and nonhomogeneities. Thus, the full-rank SMI-STAP processor is used for ease of explanation and as an exemplar application. Reduced-rank versions of nonhomogeneity detection have, in recent years, become an area of active research (e.g., see [77]). Reduced-rank versions of the methods developed in this dissertation are left as a future and fruitful area of research.

Chapter 3

Classical and Robust Estimation

It is perfectly proper to use both classical and robust/resistant methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard. (John Tukey, 1975)

The requisite robustness theory, concepts, and approaches relating to this dissertation are presented in this chapter. First, the two primary schools of thought regarding the observation and handling of non-homogeneous observations are presented. Then, simple estimators of location, scale, and covariance are defined. Next, the two major estimation methodologies, *Classical* (also known as Parametric) [40] [78] [21] and *Robustness* (sometimes known as 'Approximate-parametric') [78] [21] [22], are presented along with important quantities and qualities of each. Finally, the state-of-the-art methods and algorithms for estimating multivariate fit are presented. These methods and algorithms will be used as a basis for

comparison to the novel work presented later in this dissertation.

3.1 Approaches to Handling Non-homogeneous Observations

There are two approaches considered by modern statisticians and engineers when non-homogeneous observations must be handled in a given problem or system; the *Diagnostic Approach* and the *Outlier Accommodation* approach [79]. Each approach has relative advantages and disadvantages when compared to the other. These aspects are presented next.

3.1.1 Diagnostic Approach

The most straightforward and usually most practical approach when an outlying data point has been declared is to simply discard the data point. The general term for this is the diagnostic approach – a term which makes sense, as the only intelligent step performed is the diagnosis; no steps are taken to accommodate the data point in any way.

The main advantage of this approach is the straightforward application to almost any conventional estimator or processor. When the diagnostic approach is chosen, the estimator need not change at all; it simply needs to ensure that declared outliers are not included in the estimation of a given quantity. Examples of the diagnostic approach in action can be

found in [6] [8] [16] [23] [24] [33] [34] [15], among many others.

A primary disadvantage of this approach is that unless a given estimator is tuned precisely for the remaining data, which may contain outliers whose test statistic was close but not exceeding the given tolerance (or threshold), the estimator may still suffer significant degradation in performance. For example, estimators requiring normality of input data, but given only approximately normal data, will not necessarily yield a performance approximate of that expected under normality [79]. To solve a problem more holistically, another approach can be used and enhanced results for the problem of interest experienced.

3.1.2 Suppression / Accommodation Approach

The more complicated but perhaps more attractive approach to handling outliers involves an estimator continuing to use a (likely altered, or accommodated) version of the statistically unfit data points. In other words, outlying data points are no longer excised; rather, they are included in the estimation of a statistical quantity.

A principal reference on this matter is Barnett and Lewis' 1978 book [80]. In their work, they raise an important consideration regarding statistical tests for outlying data values. Namely, if the initial statistical model is flawed in some way, outliers determined by a statistical test may in fact not be outliers at all. Rather, they could be a sign that the underlying statistical model is incorrect for use in observing the stochastic process in question, and that a statistically respectable technique to include these outliers should be

developed.

Agreeing with Barnett and Lewis was Atkinson [81] [82] [83] [84], who chose to use the term *transformation* rather than accommodation. Atkinson's persistence on the matter of transformation throughout the 1980's was well-received in reviews of his work by famous statisticians at the time. Notably, [81] contains comments from top statisticians reviewing Atkinson's work and suggestions. Atkinson's is commended for raising the issue in such a rigorous manner, and it is agreed that underlying assumptions are overlooked all-too-often in the processing of observed data.

Unfortunately, the assumptions that underly signal processing kernels in a wide array of fields, including adaptive radar signal processing, often remain overlooked. Careful consideration must be paid to this matter, or a risk in performance will continue to exist.

3.2 Basic Statistical Estimators

This section presents common estimators used in the fields of statistics and adaptive signal processing.

As a model for use in this chapter, let $\mathbf{x} = (x_1, x_2, x_3, \dots, x_K)^T$ represent a finite-length vector of K scalar observations of a random process. Given these observations, we will now estimate certain quantities that describe the random process.

3.2.1 Estimators of Location

The first trait of \mathbf{x} we wish to estimate is its location. A location estimate can be thought of as a measurement of the *center* of \mathbf{x} , or the value that best represents the middle point of the data. While location may seem a simple concept, there are many ways to estimate location. Some location estimators are well suited to a particular type of observation, while poorly suited to estimate the location of any other observation type. Other location estimators are not suited to a particular application, but are more widely applicable to many types of observations.

The Sample Mean

The sample mean of a set of numbers is also known as the average of the numbers. For a set of K numbers $\{x_1, x_2, \dots, x_K\}$, the sample mean \bar{x} is computed as

$$\bar{x} = \frac{1}{K} \sum_{k=1}^K x_i. \quad (3.1)$$

The sample mean is the Maximum-Likelihood estimate of location if the samples x_i are drawn from a Gaussian distribution [40]. This estimator is asymptotically efficient for an underlying Gaussian distribution. (The concept of statistical efficiency will be discussed shortly.)

The Sample Median

The median is an order statistic. Given a set of numbers $\{x_1, x_2, \dots, x_K\}$, the data are first ordered from least to greatest value. Once ordered, the median is the central value. Let $m = \lfloor K/2 \rfloor + 1$ where $\lfloor \cdot \rfloor$ refers to the floor, or integer part of the expression. If K is odd, then the sample median is simply x_m . If K is even, then the sample median is the average of the two middle values, $(x_{m-1} + x_m)/2$. In summary, the sample median is given by

$$x_{med} = \begin{cases} x_m & , K \text{ odd} \\ (x_{m-1} + x_m)/2 & , K \text{ even} \end{cases} . \quad (3.2)$$

The sample median is known to be the Maximum Likelihood estimator of location when the input data are Laplacian-distributed [85].

3.2.2 Estimators of Dispersion

In the previous section, estimators designed to determine the location of a data set were presented and discussed. In this section, we now discuss estimators designed to gauge how variable a process is. Such estimators are termed estimators of *scale* or *dispersion*. The sample standard deviation is regarded as the classical estimator of scale. The sample standard deviation, denoted as σ , is defined as the square root of the sample variance,

$$\sigma^2 = \frac{1}{K} \sum_{i=1}^K (x_i - \bar{x})^2, \quad (3.3)$$

where \bar{x} is the sample mean. It is important to note that certain probability distributions have undefined second moments, in which case (3.3) will be undefined. An obvious example is the Cauchy distribution, for which exactly zero moments are defined. In the event that a set of data are expected to have this issue, robust estimators of scale such as the median absolute deviation from the median (MAD), are more appropriate.

The Median Absolute Deviation from the Median (MAD)

The MAD is a robust estimator of scale. The MAD is defined as [21]

$$\text{MAD} = 1.4826 \tilde{b}_K \text{med}_i |x_i - \text{med}_j(x_j)|, \quad (3.4)$$

where the factor 1.4826 ensures Fisher consistency for Gaussian x_i values, i.e. so the MAD approaches σ for infinite sample support, and where \tilde{b}_K is a correction factor for unbiasedness that accounts for finite sample support bias. This value is tabulated for $K \leq 9$, and given by

$$\tilde{b}_K = \frac{m}{m - 0.8} \quad (3.5)$$

for $K > 9$.

3.2.3 Estimation of Scatter

Scatter is measured by a covariance matrix from which correlation information about the observations is derived. The concept of scatter, and the covariance matrix in particular, are widely used for this dissertation.

The true covariance matrix (otherwise known as the *asymptotic* covariance matrix, or the *clairvoyant* covariance matrix) \mathbf{R} of any stochastic process can only be determined if the entire process is observed over all time, and if the process is wide-sense stationary (stationary from both the mean and variance perspective).

In the event that an infinite amount of data is not readily at hand (practically, this is always the case), an estimator of the true covariance must be employed. The maximum likelihood covariance matrix estimate of a Gaussian process \mathbf{X} , labeled $\hat{\mathbf{R}}_x$, is

$$\hat{\mathbf{R}}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H \quad , \quad (3.6)$$

where \mathbf{x}_k is the k^{th} $N \times 1$ snapshot vector of process \mathbf{X} , and N is the number of dimensions of the observation system. The concept of ML is discussed in Section 3.3.1.

3.3 Classical (Parametric) Estimation Theory

Classical or parametric estimation theory was introduced by Fisher in the early 1900's, with what is considered to be his capstone paper published in 1925 [40], as discussed in Section 2.2.

The fundamental assumption made in parametric estimation theory is that the probability density function (PDF) and cumulative distribution function (CDF) are known a priori for a given random variable. Efficient estimators for a given distribution and estimation quantity are then said to be optimal for a random variable with exactly that PDF and CDF.

The discussion below follows that of [85], and includes some concepts presented in [86], with the most relevant concepts for this dissertation emphasized and discussed in more detail.

3.3.1 Maximum Likelihood Estimators

Maximum likelihood (ML) estimation, introduced by Fisher [40], enforced some needed rigor in the estimation process. Under this approach, estimators for desired quantities could now be designed for particular distributions of interest.

Suppose a parameter with true values θ of a set of received data $\mathbf{z} = [z_1, z_2, \dots, z_m]^T$ is desired to be estimated. The ML estimator $\hat{\theta}_{ML}$ minimizes the negative log of the likelihood function, which is formally given by

$$\hat{\theta}_{ML} = \min_{\theta} \sum_{i=1}^m -\ln f(z_i; \theta). \quad (3.7)$$

Consider the case of location estimation, estimators for which were presented in Section

3.2.1. The *residual* of location at each point i is given by

$$r_i = z_i - \theta. \quad (3.8)$$

For the location case and assuming $\sigma = 1$ without loss of generality, the PDF of z_i can be written as a function of the residuals r_i . Formally, this is given by

$$f(z_i; \theta) = f(z_i - \theta) = f(r_i). \quad (3.9)$$

It follows that the ML estimator for the location case is given by

$$\hat{\theta}_{ML} = \min_{\theta} \sum_{i=1}^m -\ln f(r_i). \quad (3.10)$$

The ML estimator is then determined through the minimization of the objective function

$$J(\theta) = \sum_{i=1}^m -\ln f(r_i) = \sum_{i=1}^m \rho(r_i), \quad (3.11)$$

where $\rho(r_i)$ is the general form of the function kernel; in this case, $\rho(r_i) = -\ln f(r_i)$. The minimum of $J(\theta)$ is found by taking the derivative of $J(\theta)$ and setting it equal to zero, yielding

$$\frac{dJ(\theta)}{d\theta} = \frac{d}{d\theta} \left(\sum_{i=1}^m \rho(r_i) \right) = \sum_{i=1}^m \frac{d\rho(r_i)}{dr} \frac{dr}{d\theta} = \sum_{i=1}^m \frac{d\rho(r_i)}{dr} \times -1 = 0. \quad (3.12)$$

The $\psi(r)$ function, referred to as the *score function* in its negative form, is then defined as

$$\psi(r) = \frac{d\rho(r)}{dr} = \frac{-d \ln f(r)}{dr} = -\frac{f'(r)}{f(r)}. \quad (3.13)$$

This our minimization problem is now the solution to

$$\frac{dJ(\theta)}{d\theta} = \sum_{i=1}^m \psi(r_i) = 0. \quad (3.14)$$

(3.14) is termed the *likelihood equation* in the ML construct. Assuming the distribution of r_i is Gaussian with $\sigma = 1$, as in

$$f(r) = \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}}, \quad (3.15)$$

then the ML estimator is derived by first determining $\rho(r)$ as

$$\rho(r) = -\ln f(r) = \ln \sqrt{2\pi} + \frac{r^2}{2}. \quad (3.16)$$

Without loss of generality, the constant term is neglected as it will be zeroed by the upcoming differentiation. Our cost function thus becomes

$$\frac{dJ(\theta)}{d\theta} = \frac{d}{d\theta} \left(\sum_{i=1}^m \rho(r_i) \right) = 0. \quad (3.17)$$

Next, the $\psi(r)$ function is determined as in (3.13), namely

$$\psi(r) = \frac{d\rho(r)}{dr} = r. \quad (3.18)$$

So, the likelihood equation to be solved to find the ML location estimator under Gaussianity is given by

$$\sum_{i=1}^m \psi(r_i) = \sum_{i=1}^m r_i = 0. \quad (3.19)$$

Substituting for r_i and solving for $\hat{\theta}_{ML}$, the solution is given by

$$\sum_{i=1}^m (z_i - \hat{\theta}_{ML}) = 0 \quad (3.20)$$

$$\Rightarrow \sum_{i=1}^m (z_i) - m\hat{\theta}_{ML} = 0 \quad (3.21)$$

$$\hat{\theta}_{ML} = \frac{1}{m} \sum_{i=1}^m z_i. \quad (3.22)$$

Thus, the ML estimate of location for Gaussian-distributed observations is the sample mean, which was presented in Section 3.2.1. Using the same ML technique but assuming Laplacian-distributed observations, the ML estimate of location is the sample median, also presented in Section 3.2.1.

It is important to note under the ML construct that there is an explicit assumption of the underlying distribution. This requires that it be exactly known *a priori*. Under

this assumption, the ML estimator is asymptotically efficient. (The concept of efficiency is presented shortly.)

3.3.2 Consistency

An estimator is said to be *consistent* if it provides the true value of the parameter given infinite sample support. Mathematically this can be stated as

$$\lim_{m \rightarrow \infty} \hat{\theta}_m = \theta, \quad (3.23)$$

where m is the number of observations used in the estimate.

3.3.3 Fisher Information

In Fisher's seminal work [40], a random process Z is said to contain a certain amount of *information* about a variable of interest θ . This information is termed the Fisher Information, and denoted as $I_f(\theta)$. Another perspective can be gained by considering the maximum amount of information that can be asymptotically garnered about θ from \mathbf{Z} . This is the essence of the concept of Fisher information. Formally, the Fisher information is given by

$$I_f(\theta) = E \left(\left[\frac{\partial}{\partial \theta} \ln f(z; \theta) \right]^2 \right). \quad (3.24)$$

3.3.4 The Cramèr-Rao Lower Bound

The inverse of $I_f(\theta)$ is commonly referred to as the Cramèr-Rao lower bound (CRLB). The CRLB is the bound on the variance one can expect if an unbiased estimator for θ is employed. Under this assumption, the asymptotic variance of $\hat{\theta}$ is guaranteed to be no smaller than the value dictated by the CRLB. Mathematically, the relationship between the asymptotic variance, the CRLB, and the Fisher information is given by

$$\text{var}(\sqrt{m}\hat{\theta}_m) \geq CRLB = \frac{1}{I_f(\theta)}, \quad (3.25)$$

where the left-most quantity is the normalized sample variance.

3.3.5 Efficiency

An estimator's efficiency refers to how well its variance converges to the minimum possible variance. The better the efficiency of an estimator for a given distribution, the better that estimator is suited to measure the variable of interest under the assumed distribution.

In the finite sample case, the efficiency is a measure of comparison between the CRLB and the normalized sample variance. Sample efficiency is expressed as

$$e_m = \frac{I_f^{-1}(\theta)}{\text{var}(\sqrt{m}\hat{\theta}_m)}. \quad (3.26)$$

If the ratio is exactly equal to 1, the estimator is said to be efficient.

Asymptotic efficiency is defined as the efficiency of a Fisher-consistent estimator when the sample support approaches infinity. Mathematically this is given by

$$\lim_{m \rightarrow \infty} \text{var}(\sqrt{m}\hat{\theta}_m) = I_f^{-1}(\theta) = \lim_{m \rightarrow \infty} \frac{I_f^{-1}(\theta)}{\text{var}(\sqrt{m}\hat{\theta}_m)}. \quad (3.27)$$

Asymptotic relative efficiency (ARE) is a metric used to compare the efficiency of two consistent estimators $\hat{\theta}'_m$ and $\hat{\theta}''_m$. The comparison is formally stated as

$$e_{\hat{\theta}'_m, \hat{\theta}''_m} = \lim_{m \rightarrow \infty} \frac{\text{var}(\sqrt{m}\hat{\theta}'_m)}{\text{var}(\sqrt{m}\hat{\theta}''_m)}. \quad (3.28)$$

3.4 Robustness Theory

Robustness theory is a theoretical framework by which estimators are not designed around strict statistical assumptions as was the case in ML estimation theory discussed in Section 3.3.

The first theoretical approach to robustness is credited to Huber [87]. Considering the comments of statisticians such as Pearson, Box, and Tukey [88] regarding the need for robust procedures, Huber published a groundbreaking work in the the Annals of Mathematical Statistics entitled "Robust estimation of a location parameter" in 1964. In essence, Huber proposed that instead of a strict probability model, a *neighborhood* of such a distribution should be considered. Then, the best estimator *over the neighborhood* of that distribution,

in a minimax sense, be used instead of the ML estimator. With this work, Huber brought the suggestion for robust procedures made by Tukey et al. [88] to the mainstream.

The second key contributor to early robustness theory is Hampel, whose 1968 Ph.D. thesis [89] presented the *influence curve* (later developed into the *influence function*) and *breakdown point*. The majority of Hampel's thesis turned in to what are considered seminal robustness theory articles on *infinitesimal robustness* [90] [91]. It was at this point that the three important robustness concepts were documented: 1) *qualitative* robustness, 2) the *influence curve* (IC) (later called the influence function (IF) due to the study and processing of higher-dimension data), and 3) the *breakdown point*.

The qualitative robustness of an estimator refers to the effect a small perturbation of input data has on the estimator. The IC (later called the IF), on the other hand, measures the effects that infinitesimal perturbations to the input data distribution have on an estimator. Finally, the breakdown point (BP) measures exactly what percentage of input distribution contamination it takes to cause the estimator stop working. To this day, the BP is one of the most important measures of an estimator's global robustness. The BP will be utilized in this dissertation to assess the global robustness of the proposed estimators.

Since this early work on robustness, other important contributions to the field have been made. Of note to the research herein, Ronchetti [92], Rousseeuw [93], and Rousseeuw and Ronchetti [94] presented a generalization of the robust estimation approach to that of a statistical test.

While Huber and Hampel acted on the outcry of researchers such as Tukey and invented the robustness framework known to us today, it is well-documented that other great minds from prior periods realized the need to handle observations that do not fit the assumed parametric model. Ever since mathematicians, scientists and engineers have collected and analyzed data using well-defined procedures, the need for robustness has existed. Boscovich in 1755 is on record as rejecting outliers. In 1777, Bernoulli refers to Boscovich's outlier rejection as common practice in astronomy [21]. Even the 1805 seminal work on the method of least-squares by Legendre contained a reference to outlier rejection. However, much of this very early work focused on the proving of least squares, rather than the development of techniques that are systematically robust or resistant (Tukey used the terms interchangeably) to outliers. The first mathematically-based work on outlier acknowledgment is considered to be that of Laplace in his 1818 work describing the distribution of the median [95].

The motivation for robustness theory applied to adaptive signal processing applications includes the susceptibility of commonly-employed ML estimators, such as the sample mean and sample covariance, to data points not fitting the parametric model assumptions that govern the optimality of such estimators. As will be explained shortly, these classical estimators have a low or sometimes zero BP – an undesirable trait from both a robustness perspective and an adaptive processing perspective.

3.4.1 Robustness Concepts

Hampel [90] [91] proposed three aspects of an estimator's robustness. They are:

Qualitative Robustness The qualitative robustness of an estimator is, simply stated, the extent to which an estimator is affected by a small deviation from the assumptions underlying the estimation problem. A qualitatively robust estimator suffers small consequences (error in estimation) when the deviation from the assumptions is small.

Global Robustness The global robustness of an estimator is quantified by its breakdown point (discussed shortly). A globally robust estimator can handle large deviations in the input data.

Local Robustness Local robustness is measured in terms of the influence function and the change-of-variance function (discussed shortly). The local robustness analyzes the effects if infinitesimal bias and variance perturbations.

3.4.2 Breakdown Point (BP)

The BP is the primary metric that measures the global robustness of an estimator.

The breakdown point of an estimator refers to the contamination percentage a data set may have and still allow the estimator to function well. By contamination, we refer to data values not homogeneous with the majority. Such data points are otherwise known as

outliers.

Consider m 'good' data points, $\mathbf{z} = [z_1, z_2, \dots, z_m]^T$. As before, suppose that $\hat{\theta}_m$ be an estimate of θ from \mathbf{z} . Now, suppose that f of the m data values are replaced with an arbitrarily large value, rather than the 'good' value previously discussed, and call this new sample \mathbf{z}' . Finally, let $\hat{\theta}'_m$ be the estimator of θ processed from \mathbf{z}' .

A few basic definitions are now given. First, the *fraction of contamination* is $\epsilon = f/m$. Next, the *maximum bias* caused by the contamination of these f points is given by

$$b_{max} = \sup \left| \hat{\theta}_m - \hat{\theta}'_m \right|. \quad (3.29)$$

The *breakdown point* is then given by

$$\epsilon^* = \max \left\{ \epsilon = \frac{f}{m}; b_{max} \text{ finite} \right\}. \quad (3.30)$$

As robust estimators focus on the majority of data points, it follows that the maximum breakdown is 50%. The sample mean is a perfect example of such a high-breakdown estimator in the location case. Regardless of probability distribution of the original m 'good' data values, up to 50% of these can be contaminated and the estimator is not affected.

Other estimators, such as those that work in high dimensions, may have a breakdown point significantly lower than 50%. However, the mere fact that an estimator has a nonzero breakdown point, coupled with an understanding of the BP concept in the first place, is a

great step in the direction of robust processing.

3.4.3 Influence Function (IF)

The IF is the first of two primary measures of local robustness at a given probability distribution. Specifically, the IF determines the robustness of bias in the presence of a single outlier.

Finite-sample IF

The finite sample IF of the location estimator $\hat{\theta}_m$ at the distribution F will be presented as a first step. Assume $m - 1$ observations from distribution F are observed, $\{z_1, z_2, \dots, z_{m-1}\}$. Now allow a contaminated point z enter the observations, rendering our complete observation to be $\{z_1, z_2, \dots, z_{m-1}, z\}$. Allow $\hat{\theta}_m$ to indicate the estimate of θ processing all m points, where $\hat{\theta}_{m-1}$ processes only the first $m - 1$ points. The IF is then given by

$$IF(z; \hat{\theta}_m, F) = m \left[\hat{\theta}_m(z_1, \dots, z_{m-1}, z) - \hat{\theta}_{m-1}(z_1, \dots, z_{m-1}) \right]. \quad (3.31)$$

In plain text, the IF literally measures the influence that z has on the estimator $\hat{\theta}$.

Gross Error Sensitivity (GES) The GES is an important local robustness metric derived from the IF. It is the supremum of the IF over all possible values of z (the contaminant).

The GES is formally given by

$$\gamma^* = \sup |IF(z)|. \quad (3.32)$$

Depending on the GES is a critical quantity called the *maximum bias*. The maximum bias is given by

$$b_{max} = \epsilon \gamma^*. \quad (3.33)$$

An estimator that is desirable usually has bounded GES, and thus a bounded maximum bias. Estimators with these traits are termed *B-robust* (the 'B' is for 'Bias').

Asymptotic IF

For the asymptotic case, the number of observations m goes to infinity. In this case, we assume an ϵ -contaminated probability distribution, given as follows. Denote an uncontaminated probability distribution as F , and a probability point mass of 1 at arbitrary location z , denoted Δ_z . The contaminated distribution G is modeled as

$$G = (1 - \epsilon)F + \epsilon\Delta_z. \quad (3.34)$$

Under certain regularity conditions [21], the asymptotic IF can then be derived as the partial derivative with respect to the contamination factor ϵ of the functional form of $\hat{\theta}$ at G , or

$\hat{\theta}(G)$, evaluated for $\epsilon = 0$. Formally, this is given by

$$IF(z; \hat{\theta}_m, F) = \left. \frac{\partial \hat{\theta}(G)}{\partial \epsilon} \right|_{\epsilon=0}. \quad (3.35)$$

This is equivalent to the familiar differentiation limit of

$$\lim_{\epsilon \rightarrow 0} \frac{\hat{\theta}(G) - \hat{\theta}(F)}{\epsilon}, \quad (3.36)$$

since F and G differ by an arbitrarily small value of ϵ , and $F = G$ when $\epsilon = 0$. Thus, the asymptotic IF describes the local behavior of $\hat{\theta}$ in an arbitrarily close neighborhood of F . This is a primary example of why robustness-based estimators are considered to be *approximately parametric*. The asymptotic bias is determined by

$$b(\epsilon) = \left| \hat{\theta}(G) - \hat{\theta}(F) \right|. \quad (3.37)$$

Figure 3.1 displays an *asymptotic maximum bias curve* for the example of the sample median. This curve relates several of the aforementioned robustness concepts and measures in one holistic way. The qualitative robustness of the estimator can be gauged by the continuity of the curve at precisely $\epsilon = 0$. The GES, γ^* is the slope of the tangent of the maximum bias curve at $\epsilon = 0$. Finally, the BP ϵ^* is located on the ϵ axis where the curve asymptotically deviates to infinity.

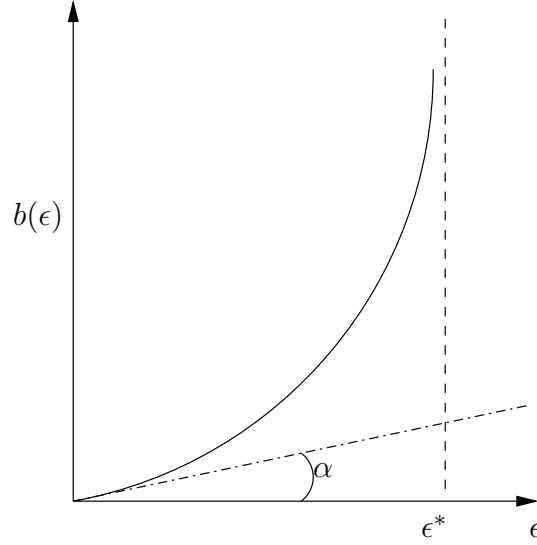


Figure 3.1: Asymptotic Maximum Bias $b(\epsilon)$ as a Function of ϵ

3.4.4 The M-estimator

Huber [87] [78] defined a new class of estimators in 1964 termed *M-estimators*, which stands for *Generalized Maximum Likelihood Estimators*. Instead of strict parametric assumptions, the new class of estimator assumes *approximate* parametric models

For the discussions that follow, consider the standard linear regression model of the matrix form

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}, \quad (3.38)$$

where $\mathbf{z} = [z_1, z_2, \dots, z_K]^T$ is the desired response, $\mathbf{e} = [e_1, e_2, \dots, e_K]^T$ is the error, $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$ is the weighting vector for the regression for which we would like to solve,

and \mathbf{H} is the matrix of observed data such that

$$\mathbf{H}^T = \begin{bmatrix} | & | & | & | \\ \mathbf{h}_1 & \mathbf{h}_2 & \dots & \mathbf{h}_K \\ | & | & | & | \end{bmatrix}. \quad (3.39)$$

The residual vector is defined as

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\mathbf{x}, \quad (3.40)$$

and the i th component of \mathbf{r} is given by

$$r_i = z_i - \mathbf{h}_i^T \mathbf{x}. \quad (3.41)$$

Given this model, the formal definition of an M-estimator involves the minimization of an objective function $J(\mathbf{x})$, given by

$$\min_{\mathbf{x}} J(\mathbf{x}) = \sum_{i=1}^m \rho\left(\frac{r_i}{\sigma_i}\right), \quad (3.42)$$

where the argument of $\rho()$ now includes a scaling factor σ_i , assumed estimated or known *a priori*. The minimization is determined in similar fashion to Section 3.3.1; by taking the partial derivative of $J(\mathbf{x})$ with respect to \mathbf{x} (again, this assumes a well-behaved $\rho()$ function

[21] including differentiability), and setting equal to zero, as in

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^m \frac{1}{\sigma_i} \frac{\partial \rho\left(\frac{r_i}{\sigma_i}\right)}{\partial \left(\frac{r_i}{\sigma_i}\right)} \left(\frac{\partial r_i}{\partial \mathbf{x}}\right)^T = 0. \quad (3.43)$$

To simplify this, it is easily obtained from the regression in (refeq:regress2) that $\partial r_i / \partial \mathbf{x} = -\mathbf{h}_i^T$. Also, define the scalar function $\psi(u) = \partial \rho(u) / \partial u$. Given this, (3.43) becomes

$$\sum_{i=1}^m \psi\left(\frac{r_i}{\sigma_i}\right) \frac{1}{\sigma_i} \mathbf{h}_i = 0. \quad (3.44)$$

An important insight into the M-estimator can be had if one allows $\rho(r) = -\ln f(r)$, setting $\sigma = 1$ without loss of generality. In doing so, $\psi(r) = -f'(r)/f(r)$, and thus we obtain the standard ML estimator as presented in Section 3.3.1. Hence, the ML estimator is considered a special case of the M-estimator.

However, the power of the M-estimator is derived from its flexibility in choosing $\rho(r)$. For instance, Huber's first proposal for $\rho(r)$ and $\psi(r)$, respectively given by

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & \text{if } |r| \leq b \\ b|r| - \frac{1}{2}b^2 & \text{if } |r| > b \end{cases} \quad (3.45)$$

and

$$\psi(r) = \begin{cases} r & \text{if } |r| \leq b \\ b \operatorname{sign}(r) & \text{if } |r| > b \end{cases}, \quad (3.46)$$

allows for down-weighting, thereby bounding their influence. For $b = 2$, the Huber ρ and ψ functions are shown in Figure 3.2. Huber's function was part off the first of three classes of M-estimators. The first class contained convex $\rho(r)$ functions. The other two classes contain non-convex $\rho(r)$ functions. Depending on the application, a particular ψ -function is chosen to achieve desired efficiency and robustness properties.

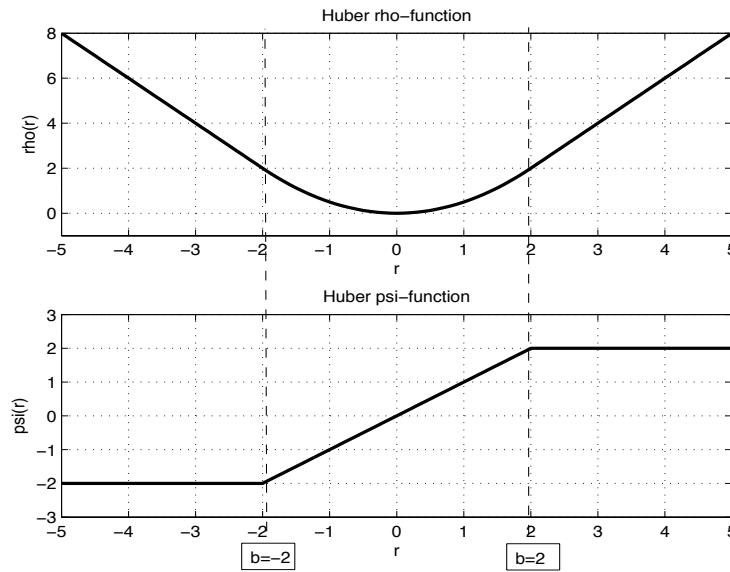


Figure 3.2: The Huber ρ and ψ Functions ($b = 2$)

3.4.5 The GM-estimator

If one examines (3.44) for a moment, it is clear that the influence of residual (IR) values r_i on the solution are bounded by the $\psi(r)$ function, while the influence of the data vectors are not bounded in any way. Thus, for a single outlying observation \mathbf{h}_i associated with a leverage point, it is easily seen that the M-estimator may very well break down. This issue

is known as an unbounded *influence of position* (IP).

Once the unbounded IP of an M-estimator was discovered and became enough of a serious issue, several proposals were made to bound both the IR and the IP. Mallows [31] proposed a new estimator that bounded both the IP and IR, but did so separately – a non-ideal solution to the problem. Schweppe [32], on the other hand, realized the need to inter-relate IR and IP such that the total influence is guaranteed to be bounded provided that the $\psi(r)$ function is bounded. Schweppe’s version of the GM-estimator is also known as the ”Hampel-Krasker-Welsch” estimator [96]. This type of GM-estimator also was shown to have a positive BP by Maronna [97] for the same (very reasonable) assumption.

In this dissertation, the Schweppe-type estimator will be exclusively considered due to the desire for a bounded influence of both residual and position, jointly. This GM-estimator is formally defined as the solution to

$$\min_{\mathbf{x}} J(\mathbf{x}) = \sum_{i=1}^m w^2(\mathbf{h}_i) \rho \left(\frac{r_i}{\sigma_i w(\mathbf{h}_i)} \right). \quad (3.47)$$

This minimization is solved using partial derivatives as in

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = \sum_{i=1}^m \frac{w(\mathbf{h}_i)}{\sigma_i} \frac{\partial \rho \left(\frac{r_i}{\sigma_i w(\mathbf{h}_i)} \right)}{\partial \left(\frac{r_i}{\sigma_i w(\mathbf{h}_i)} \right)} \left(\frac{\partial r_i}{\partial \mathbf{x}} \right)^T = 0. \quad (3.48)$$

To simplify this, it is easily obtained from (3.41) that $\partial r_i / \partial \mathbf{x} = -\mathbf{h}_i^T$. Also, define the scalar

function $\psi(u) = \partial\rho(u)/\partial u$. Given these pieces of information, (3.48) can be re-stated as

$$\sum_{i=1}^m \frac{w(\mathbf{h}_i)}{\sigma_i} \mathbf{h}_i \psi\left(\frac{r_i}{\sigma_i w(\mathbf{h}_i)}\right) = 0. \quad (3.49)$$

The weight function of the observations $w(\mathbf{h}_i)$ is given by [98] [8]

$$w(\mathbf{h}_i) = \min\left(1, \frac{b^2}{PS_i^2}\right), \quad (3.50)$$

where PS_i is the projection statistic computed for observation i , and b is the 97.5-percentile of the chi-square distribution for projection statistics, given by

$$b = \chi_{N,.975}^2, \quad (3.51)$$

where N is the number of dimensions of the observations. It is important to note that this weighting scheme is used for low-redundancy applications such as power system state estimation. If redundancy is high, the recommended weight function is given by

$$w(\mathbf{h}_i) = \min\left(1, \frac{b}{PS_i^2}\right). \quad (3.52)$$

The Schweppe-type GM-estimator enjoys a bounded influence. However, this comes at the cost of a lower breakdown point in high dimension applications [79]. That said, any robustness added to a non-robust estimator (such as is used in the conventional STAP

radar application) is an improvement. Furthermore, an increase in data redundancy through an increase in pulse compression bandwidth, for example, will aid in any breakdown point issue the GM-estimator may present. The GM-estimator is the focus of Chapter 6 of this dissertation; specifically, a GM-estimator for STAP radar that handles complex-valued data is developed.

3.4.6 The S-estimator

The S-estimator is a class of estimators based on a robust residual scale [79][21]. Any scale-equivariant estimator of scale $s()$ may be used in the S-estimator, i.e., $s(\lambda r_1, \dots, \lambda r_n) = |\lambda|s(r_1, \dots, r_n)$. For example, if the estimator of this scale is chosen to be the median, then such an S-estimator specifically becomes the least median of squares (LMS) estimator.

Speaking generally, let $\hat{\sigma} = \hat{\sigma}(\mathbf{r}(\beta))$ be a scale-equivariant robust scale estimator with argument β of the residual value vector \mathbf{r} from (3.40). A regression estimate based on this robust scale is defined as

$$\hat{\beta} = \arg \min_{\beta} \hat{\sigma}(\mathbf{r}(\beta)). \quad (3.53)$$

It represents an estimator based on robust residual scale.

An S-estimate is a special case of this regression estimate. For instance, by defining

$\hat{\sigma}(\mathbf{r})$, an M-estimator of scale is given by

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_i}{\hat{\sigma}}\right) = \delta, \quad (3.54)$$

where δ is a tuning parameter. As with M- and GM-estimators, the ρ -function can be chosen to meet robustness requirements, implying that it must be bounded.

As recommended in [79], the S-estimator used in this work is based on the bisquare ρ function. Without loss of generality, allow $\sigma = 1$. The bisquare ρ -function is then defined as

$$\rho(r) = \begin{cases} \left[1 - \left[1 - \left(\frac{r}{b}\right)^2\right]^3\right] & \text{if } |r| \leq b \\ 1 & \text{if } |r| > b \end{cases}. \quad (3.55)$$

Its derivative is given by

$$\psi(r) = \begin{cases} r \left[1 - \left(\frac{r}{b}\right)^2\right]^2 & \text{if } |r| \leq b \\ 0 & \text{if } |r| > b \end{cases}. \quad (3.56)$$

For $b = 2$, the bisquare (biweight) ρ and ψ functions are shown in Figure 3.3 as a function of r . S-estimators generally yield a high BP, even in larger dimensions. Since the S-estimator has the form of the M-estimator, its asymptotic variance and IF are the same as that of the M-estimator [21]. The S-Estimator is used in Chapter 4 to obtain a robust yet computationally efficient dispersion estimate for the Adaptive Complex Projection Statistics (ACPS) algorithm.

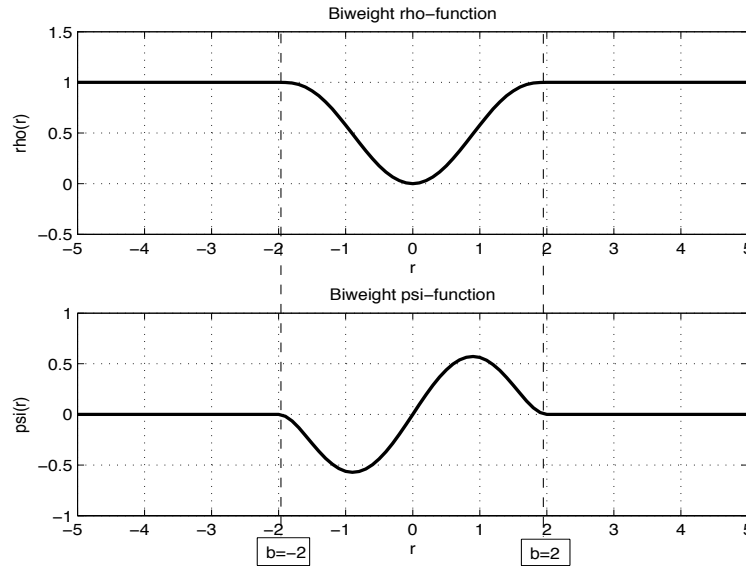


Figure 3.3: The Bisquare (Biweight) ρ and ψ Functions ($b = 2$)

3.4.7 The Minimum Covariance Determinant (MCD)

From a high level, the MCD is self-explanatory. It is defined as the smallest covariance determinant sought out of a subset h of the k observations. Because the data subset associated with the MCD is from the bulk of the data set, this estimator is robust to outliers. The specific MCD implemented to perform MCD simulations is the FAST-MCD algorithm of Rousseeuw and Van Driessen [99].

3.4.8 The Tradeoff Between Robustness and Efficiency

High breakdown comes at a cost. The price paid is typically a loss of efficiency at the Gaussian distribution. For example, suppose an i.i.d. sample of size m is given. Then, suppose it is desired to determine the location parameter based on this sample. Finally,

assume the the sample mean estimator $\hat{\mu}$ is employed. It turns out that the sample mean $\hat{\mu}$ is well-known to be the best estimator of location at the Gaussian distribution, since its variance attains the Cramèr-Rao lower bound, yielding an efficiency of 100%.

Unfortunately, it only takes one infinitesimal perturbation to carry $\hat{\mu}$ over all bounds, signifying a breakdown point of zero. To fix this robustness problem, one may choose to employ a *more robust*, but *less efficient* estimator of location such as the sample median, $\hat{\theta}_{med}$. Under this scenario, robustness is greatly improved, at the expense of efficiency, which is only 63% for the median at the Gaussian distribution. Similar *robustness-efficiency* tradeoffs are common when designing the solution to an estimation problem.

Chapter 4

The Adaptive Complex Projection Statistics (ACPS) Algorithm

The robustness literature contains an array of diagnostic tools, each having respective strengths and weaknesses, that can be used to detect outliers in observed sets of data. In adaptive signal processing applications, the needs to *robustly* yet *quickly* determine a solution are paramount. Computing limitations and a need to handle high-dimensional data are also frequently present.

As such, a high-breakdown, low-computation diagnostic test for outliers in multiple dimensions in adaptive signal processing applications is warranted. An algorithm independently developed by Stahel [2] in 1981 and Donoho [3] in 1982, commonly termed *Projection Statistics*, became a major focus of this research work, as its traits are fundamentally in

line with the presented adaptive signal processing requirements for a new, highly robust diagnostic tool to detect outliers quickly for practical, on-line applications (as opposed to applications which have the luxury of time commensurate with off-line processing.) The PS algorithm is highly robust while being computationally attractive compared to other outlier diagnostics approaches.

The original algorithm proposed by Stahel and Donoho at first involved the projection of observed multi-dimensional input data points onto *all* possible one-dimensional directions. While this approach is a valuable contribution, the prosecution of all directions presents a serious practicality issue. So, in 1982, Gasko and Donoho proposed a revised algorithm that investigates only those directions that originate at the coordinate-wise median and pass through one of the input data points [4]. The maximum of these projections for each input data point is termed the *projection statistic* for that data point, and represents the worst one-dimensional projection of that data point in consideration of the rest of the data points [5]. The last major contribution to the PS algorithm came in 1991 when Rousseeuw and Van Zomeren [6] performed simulations to determine appropriate cut-off values of the PS algorithm. The PS cut-off value is simply the threshold above which one declares that an outlier, defined as a data point not predicted by the Gaussian distribution, exists. This work was a follow-up to their 1990 paper [7]. It is clear that Rousseeuw and Van Zomeren were implementing the new technique and working toward the goal of cut-off value determination which they achieved in 1991.

The PS algorithm as completed by Rousseeuw and Van Zomeren, has been used through-

out the 1990's in the challenging application of structured non-linear regression in high dimensions as exemplified by power system state estimation problems [8].

The state-of-the-art PS algorithm has two aspects that do not permit its ready use in adaptive signal processing applications. First, the state-of-the-art PS algorithm utilizes real-valued data exclusively – not the complex-valued data as is used in the adaptive signal processing applications under consideration in this dissertation such as STAP. Second, the threshold setting for PS applications is not adaptively determined, and is thus not appropriate for most signal environments in which changes in average power level, for instance, can be a constantly shifting quantity. Compounding this threshold determination is the fact that PS is a non-parametric algorithm; thus, its output distribution is not tractable such as that of the GIP.

This chapter addresses these two critical aspects of PS and proposes a revised algorithm, adaptive complex projection statistics (ACPS), in order to allow for the ready use of a PS-based algorithm in real adaptive signal processing systems such that the diagnostic power of PS can be leveraged as never before possible.

4.1 Projection Statistics for Real-valued Data

Projection Statistics (PS), as most recently documented in [6], is an algorithm for determining whether a particular N-dimensional data snapshot is an outlier with respect to a set of data snapshots. The PS technique is non-parametric and thus a tractable proof of the

output statistical distribution is not possible. However, Monte Carlo analysis was performed in [6] to determine a usable threshold for a given set of assumptions.

The PS algorithm for real-valued data is as follows:

Given K data vectors \mathbf{x}_l , $l = 1, \dots, K$, compute the coordinate-wise median \mathbf{m} . Then, for each of the K directions $\mathbf{v}_l = \mathbf{x}_l - \mathbf{m}$, repeat:

1. Project the other observations on to \mathbf{v}_l , $y_i = \mathbf{x}_i \mathbf{v}_l^T$.
2. Calculate location and scale estimates $L = L(y_1, \dots, y_K)$ and $S = S(y_1, \dots, y_K)$.
3. Standardize the projection values: $z_i = \frac{|y_i - L|}{S}$, $i = 1, \dots, K$
4. Find the maximum of these standardized projections, i.e., $PS_i = \max(z_i)$. This is the projection statistic value for the data vector \mathbf{x}_l . To determine if \mathbf{x}_l is an outlier, we again must use a threshold. Rousseeuw et. al.[6] showed that for Gaussian input data snapshots, the PS_i values are approximately $\chi_{N,97.5\%}^2$ for sample support values of $5K$ or more. The distribution of PS values is poorly represented by the χ_N^2 when the redundancy is less than $5N$, particularly in higher dimensions. We propose in Section 4.3.3 a statistical test to deal with the lack of distributional knowledge for outlier diagnostics.

4.2 Projection Statistics for Complex-Valued Data

A complex data value z_i involves both a real and imaginary component, as in

$$z_i = x_i + jy_i, \quad (4.1)$$

where x_i is the real component and y_i is the imaginary component, and j represents the imaginary dimension with a value of $j = \sqrt{-1}$. In vector form, this becomes

$$\mathbf{z} = \mathbf{x} + j\mathbf{y}, \quad (4.2)$$

where \mathbf{x} and \mathbf{y} are vectors of the real and imaginary components of \mathbf{z} , respectively.

A complex data value can be viewed as a representation of two dimensions, from either the diagnostic or accommodation/suppression statistical viewpoint. In contrast, the association of x_i and y_i is *critical* in adaptive signal processing applications, as phase information is preserved via this association.

From the statistical viewpoint, these values are to be treated as separate statistical dimensions from which to help determine the statistical fit of z_i or a vector \mathbf{z} of z_i values. As such, a complex data vector $\mathbf{z} = [z_1, z_2, \dots, z_n]^T = \mathbf{x} + j\mathbf{y}$ will be treated according to [100] in this development as

$$\zeta = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}. \quad (4.3)$$

This formulation of complex data allows for full exploitation of the information offered by such quantities, and its dimensionality works well with the robustness concepts developed herein. This formulation of complex data came after much research and experimenting with the various ways to accomplish the goals of accepting complex-valued observations and

robustly performing the adaptive signal processing. This formulation is superior to that used in the author's earlier work (e.g., [23] [24]) in that it is more robust, yields higher performance, and is statistically sound from an algorithm dimensionality perspective. Early approaches to the research in this dissertation yielded excellent results in N-Dimensions, however a lack of exploitation of the complex data's structure allowed for the continued improvement and ACPS proposal contained herein.

To form PS, assume there are K complex input data snapshots \mathbf{x}_k , $k= 1, \dots, K$, each of dimension $2MN \times 1$. The n th element of snapshot k shall be denoted x_{nk} . First, compute the coordinate-wise complex sample median of the K input data snapshots defined here to be

$$\mathbf{m} = \begin{bmatrix} \text{med}_{k=1,\dots,K}(x_{1k}) \\ \vdots \\ \text{med}_{k=1,\dots,K}(\{x_{Nk}\}) \end{bmatrix}. \quad (4.4)$$

The $2MN \times 1$ vector \mathbf{m} represents a robust estimate of the complex *location* coordinates of the cloud of $2MN$ -dimensional "unfolded" complex snapshots. This location estimate is subtracted from each snapshot to obtain the centralized vectors \mathbf{u}_k ,

$$\mathbf{u}_k = \mathbf{x}_k - \mathbf{m}, \quad k = 1, \dots, K. \quad (4.5)$$

Next, \mathbf{u}_k is normalized to obtain \mathbf{v}_k as,

$$\mathbf{v}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|_2} = \frac{\mathbf{u}_k}{\sqrt{u_{1k}^2 + \dots + u_{Nk}^2}}, \quad k = 1, \dots, K. \quad (4.6)$$

We now determine the projection of each input data snapshot \mathbf{x}_k in each direction \mathbf{v}_k ,

$$\begin{aligned} z_{1k} &= \mathbf{x}_1^T \mathbf{v}_k \\ z_{2k} &= \mathbf{x}_2^T \mathbf{v}_k \\ &\vdots \\ z_{Kk} &= \mathbf{x}_K^T \mathbf{v}_k \end{aligned} \quad k = 1, \dots, K \quad (4.7)$$

The $K \times 1$ vectors $\mathbf{z}_k = [z_{1k}, z_{2k}, \dots, z_{Kk}]^T$ represent the vector of projections for input data snapshot k .

For each vector \mathbf{z}_k , we next compute the robust location and scale estimates using the median and MAD respectively,

$$z_{med_k} = \underset{n=1, \dots, K}{med} [z_{nk}] \quad (4.8)$$

$$MAD_k = 1.4826 \left[\underset{n=1, \dots, K}{med} |z_{nk} - z_{med_k}| \right], \quad (4.9)$$

where $|*|$ denotes absolute value. Next we compute the standardized projections,

$$p_{nk} = \frac{|z_{nk} - z_{med_k}|}{MAD_k}, \quad n = 1, \dots, K \quad (4.10)$$

and repeat (4.8) – (4.10) for all k values ($k=1,\dots,K$). Finally, the K projection statistic values are determined as

$$PS_n = \max_{k=1,\dots,K} \{p_{nk}\}, \quad n = 1, \dots, K \quad (4.11)$$

The projection statistic values for each input snapshot \mathbf{x}_k , $k = 1, \dots, K$ is given by the corresponding PS_n for $n = 1, \dots, K$. If a PS_n crosses the threshold, then the associated data snapshot is diagnosed as an outlying observation. As is the case with GIP statistics, if the \mathbf{x}_k data are Gaussian-distributed, then the resulting PS_n approximately follow a Chi-squared distribution [6], resulting in a threshold value of 50.725 for the 33 complex degrees of freedom (DoF), or dimension, scenario studied in the STAP scenario presented later in this chapter. This scenario inserts multiple outlying data points (modeled as strong target snapshots) into the otherwise near-Gaussian training data. Thus, the \mathbf{x}_k data are not completely homogeneous. It follows that the approximate Chi-squared distribution cannot be guaranteed when these multiple outliers cause the data to violate the Gaussian assumption. Thus, the threshold must be adjusted manually. Methods to adjust the PS threshold in a data-driven manner are presented later in this section. The non-adaptive version of PS for complex data as presented here is used in the simulations for comparison, and is termed CPS.

4.3 An Adaptive Threshold for Complex Projection Statistics

In this section, the 'Adaptive' component of ACPS is described. The raw Projection Statistics values observed are skewed and not centered. Both the skewness level and location of the PS values, much more so than other diagnostic values such as the GIP, vary with the number of dimensions involved. This was roughly shown in [16] for the GIP and in [6] for the original real PS algorithm.

In [6], the threshold recommendations made for PS were based on Monte Carlo simulations. There, it was shown that the threshold recommendation is not a good choice for higher dimensions due to heavy tails. Hence, a threshold that adapts to the distribution characteristics of the PS, while still preserving the PS's powerful ability to identify outliers, is required for operation of the new ACPS technique in practice.

It is desirable for diagnostic statistics to present a separation between the bulk of observed data and any outlying data points, and to be robust to the masking effect when doing so. PS-based algorithms possess this desirable property. As shown in [6], the output distribution threshold for diagnosing an outlier varies with the number of data dimensions. For PS specifically, the Chi-squared distribution of PS does not hold true for all dimensions, as evidenced by the Q-Q plots performed in [6] and replicated by the author of this dissertation.

A representative set of PS values for a contaminated 20-dimensional complex data set

(taken from the simulations in Section 4.4.1 is shown in Figure 4.1. The gap between the outlier and the bulk of the data is a desirable feature, and should be preserved by any statistical transformation proposed. To transform the raw PS values into an adaptively usable statistic, we must wisely choose both location and scale estimators that will allow for skewness as well as be highly efficient in the other-than-outliers Gaussian data.

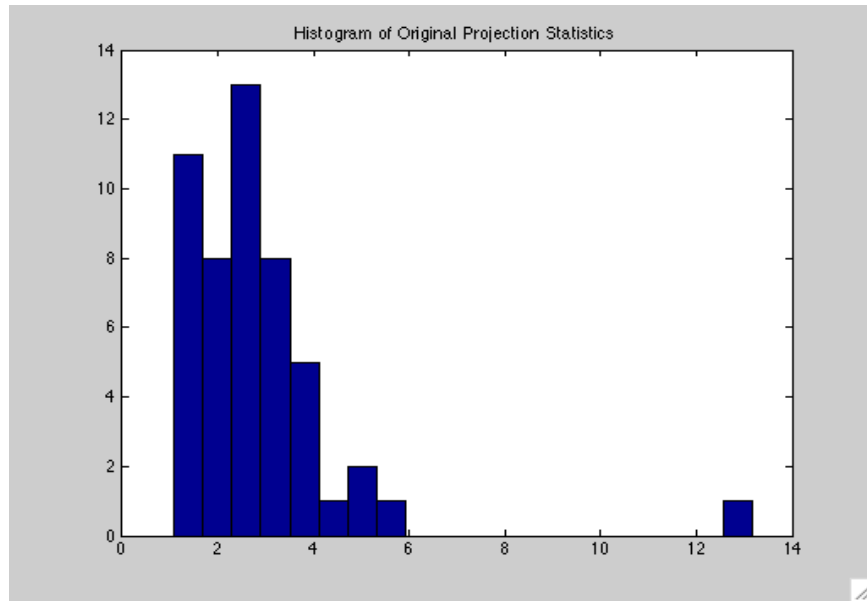


Figure 4.1: Histogram of the Raw Projection Statistics

4.3.1 Choice of Location Estimator

Due to the skewness of the distribution of observed PS values, a location estimate robust to this skew is desired. The mean of the length of the shortest half (LSH) of the data, known as the *shorth*, is a very robust estimator of location [21]. Thus, the location parameter proposed for use in the ACPS test statistic is the *shorth*.

4.3.2 Choice of Scale Estimator

The S-estimator and Q-estimator of scale initiated by Rousseeuw et al [22] [101] are considered here. Unlike other scale estimators, neither the S-estimator nor the Q-estimator require the use of a location estimator. We analyze their respective strengths and weaknesses and evaluate their performance for meeting our goals.

As shown in [101], S and Q estimators are 58% and 82% efficient at the Gaussian distribution, respectively. Both of these values are superior to the MAD's 37% efficiency. While this theoretical property is appropriate for consideration, the efficiency for finite sample sizes is difficult to compare. While the Q-Estimator enjoys a smooth asymptotic IF, the S-Estimator does not. Furthermore, the Q-Estimator loses efficiency faster in small sample sizes than the S-Estimator, however these problematic sample sizes of the Q-estimator are generally on the low end of sample sizes employed in practical adaptive signal processing. Despite these advantages of the Q-estimator, a low computational complexity is the most pressing requirement. Thus, our choice settles on the S-estimator for adaptive signal processing applications due to the S-estimator's lower computational complexity.

4.3.3 Proposal of a New Statistical Test Based on CPS

Thus, the adaptive complex projection statistic values, denoted as $ACPS_i$, that we use for outlier detection are given by

$$ACPS_i = \frac{|PS_i - \text{shorth}\{\mathbf{PS}\}|}{SC} \quad (4.12)$$

where $ACPS_i$ is ACPS value for observation i , \mathbf{PS} is the vector of PS values for the observations, and PS_i refers to the PS value for observation i , and SC is the S-Estimator of scale of \mathbf{PS} . Based on the data shown in Figure 4.1, the statistic proposed in (4.12) presents the new histogram shown in Figure 4.2.

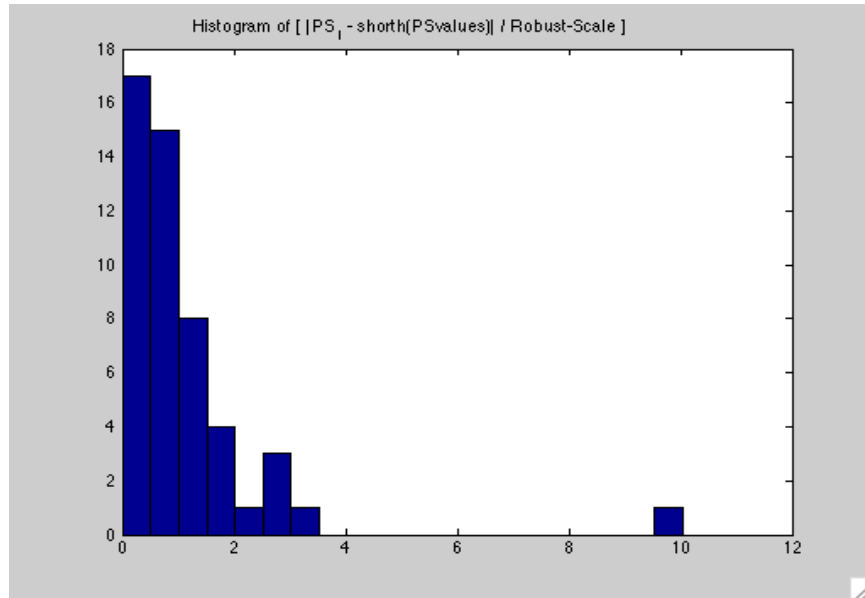


Figure 4.2: Histogram of the ACPS Values. Location and Scale have been Robustly Removed.

Diagnosis of Outliers using ACPS Observation i is diagnosed as an outlier if

$$ACPS_i > 2 * MAD(\mathbf{ACPS}), \quad (4.13)$$

where \mathbf{ACPS} is the vector of all ACPS values. Otherwise, observation i is declared as a good data point. The cut-off value of $2 * MAD(\mathbf{ACPS})$ is used to obtain adaptivity across multiple values of DoF N *as well as* the varied power levels that may be observed.

ACPS Algorithm Summary The ACPS algorithm developed and utilized in this dissertation is summarized in Table 1. As will be demonstrated, the ACPS algorithm makes for a readily useful, highly robust, and computationally efficient outlier diagnostics technique applicable to adaptive signal processing applications.

4.4 Application and Results

In this section, the ACPS algorithm is exercised and performance is demonstrated. First, the convergence of the SMI STAP processor SINR is gauged using simulated data and compared to the state-of-the-art. Second, STAP scenarios based on real MCARM [102] data are presented and performance of ACPS is compared to the state-of-the-art.

Algorithm 1 Standardized Projection Statistics

Input: A set of K observed P -dimensional complex data vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K]$.

Complex element p of observation k is denoted as x_{pk} . {In STAP, $P = MN$.}

Output: ACPS value vector **ACPS** and outlier list **out_a**.

A. Create a new observation matrix $\tilde{X} = [\Re\{\mathbf{X}\}; \Im\{\mathbf{X}\}]$. Each observation is now $L=2P=2MN$ long, and $\tilde{\mathbf{X}}$ is of size $L \times K$. Real-valued element l of observation k is denoted as \tilde{x}_{lk} .

B. Compute the coordinate-wise median $\mathbf{m} = \left[\text{med}_k(\tilde{x}_{1k}), \dots, \text{med}_k(\tilde{x}_{Lk}) \right]$.

C. Compute K one-dimensional directions $\mathbf{u}_k = \tilde{\mathbf{x}}_k - \mathbf{m}$.

D. Standardize the \mathbf{u}_k values via $\mathbf{v}_k = \mathbf{u}_k / \|\mathbf{u}_k\|$.

E. For each \mathbf{v}_i , $i = 1, \dots, K$, compute the standardized projections of the other observations onto \mathbf{v}_i :

- 1) Project all data onto \mathbf{v}_i by computing $\mathbf{z}_i = \mathbf{X}^T \mathbf{v}_i$.
- 2) Robustly localize the \mathbf{z}_i values by computing $m_i = \text{med}(z_{1i}, z_{2i}, \dots, z_{Ki})$.
- 3) Robustly scale the \mathbf{z}_i values by computing the MAD, as in $\hat{\sigma}_i = 1.4826 \text{ med}|\mathbf{z}_i - m_i|$.
- 4) Compute the standardized projection vector $\mathbf{s}_i = |\mathbf{z}_i - m_i| / \hat{\sigma}_i$.

F. Determine the Projection Statistic for each observation k , $k = 1, \dots, K$ by finding the maximum 1-D projection value, i.e., $PS_k = \max_i(\mathbf{s}_{ik})$.

G. Determine SC , the S-estimator's robust scale estimate of **PS**, the vector of PS_k values. Also determine $\text{shorth}(\mathbf{PS})$.

H. Robustly localize and scale the PS_k values by assigning $\mathbf{ACPS} = |\mathbf{PS} - \text{shorth}(\mathbf{PS})| / SC$.

I. Return the statistics $\mathbf{ACPS} = [a_1, a_2, \dots, a_K]$ and the declared outliers $\mathbf{out}_a = \text{find}(\mathbf{ACPS} > 2 * \text{MAD}(\mathbf{ACPS}))$.

4.4.1 SINR Convergence

Before applying the proposed ACPS technique to target detection in a realistic environment, we first will analyze its fundamental SINR convergence performance in controlled, simulated Gaussian data with known injected target characteristics. Targets are the most problematic of outliers when not in the STAP range cell under test, but in the training data.

SINR convergence as a function of sample support in simulated colored Gaussian data is shown in the following graphs. The colored Gaussian data simulation was set up as follows.

1. Gaussian data generated (with a consistent seed to eliminate number generator effects)
2. 3 jamming interferers (each 30 dB above data floor. Jammer angles $\theta_i = [-60, 80, 110]$ (deg), $i = 1 \dots 3$)
3. 3 outliers (with one 1-outlier case presented first to demonstrate performance with no masking effect)
4. 50 MC trials for each outlier power of each scenario
5. Both 6 and 20 complex data dimension scenarios are presented
6. Static Threshold (where applicable) – from robustness literature [6] and GIP literature [16]
 - For 6 DoF, threshold = $\sqrt{\chi_{6, 97.5\%}^2} = 3.8012$
 - For 20 DoF, threshold = $\sqrt{\chi_{20, 97.5\%}^2} = 5.8455$

7. SMI STAP processing is performed using Equations 2.52 - 2.53 and empirical output SINR recorded for each MC trial
8. SINR MC results are then averaged and presented in each graph for each target power level

This convergence analysis is performed on the standard GIP, Reiterative (multi-pass) GIP, non-adaptive Complex PS (CPS) , and the new algorithm ACPS. In the simulations presented here, the average SINR is computed by averaging the SINR over M Monte Carlo runs using (2.58), restated here for convenience,

$$\text{Avg SINR} = \frac{1}{M} \frac{1}{\text{SINR}_{opt}} \sum_{i=1}^M \frac{|\hat{\mathbf{w}}_i^H \mathbf{s}|^2}{\hat{\mathbf{w}}_i^H \hat{\mathbf{R}}_i \hat{\mathbf{w}}_i}, \quad (4.14)$$

where SINR_{opt} is a normalizing factor that allows the asymptotic SINR to be exactly 0 dB. SINR_{opt} can be approximated by the reciprocal of the MMSE from (2.59),

$$\text{MMSE} = \mathbf{w}^H \mathbf{R} \mathbf{w}, \quad (4.15)$$

for a target power of unity (assumed here).

An initial single target-outlier test is performed to demonstrate the functionality of the GIP, Reiterative GIP, CPS, and ACPS. As shown in Figures 4.3 through 4.6, all statistical tests perform well, as is expected with a single target outlier which causes no masking effect.

The first multiple target-outlier test is performed in low dimension. Six complex di-

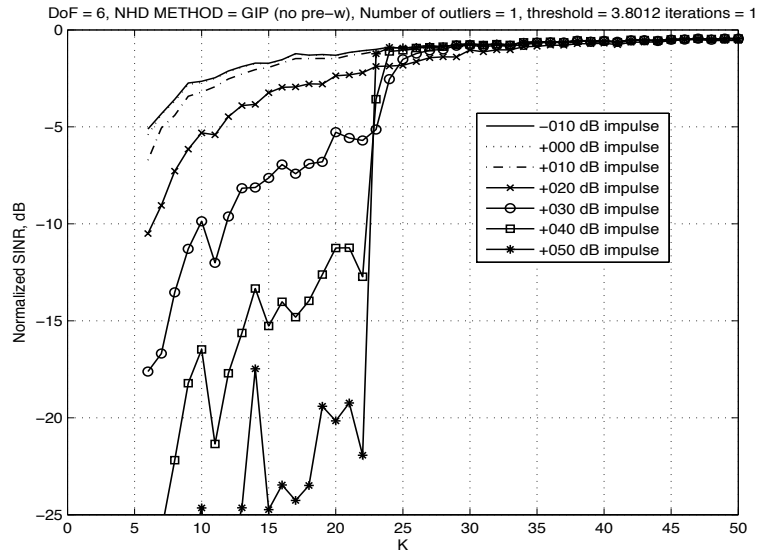


Figure 4.3: 6 Dimension GIP Results, 1 Iteration, 1 Outlier

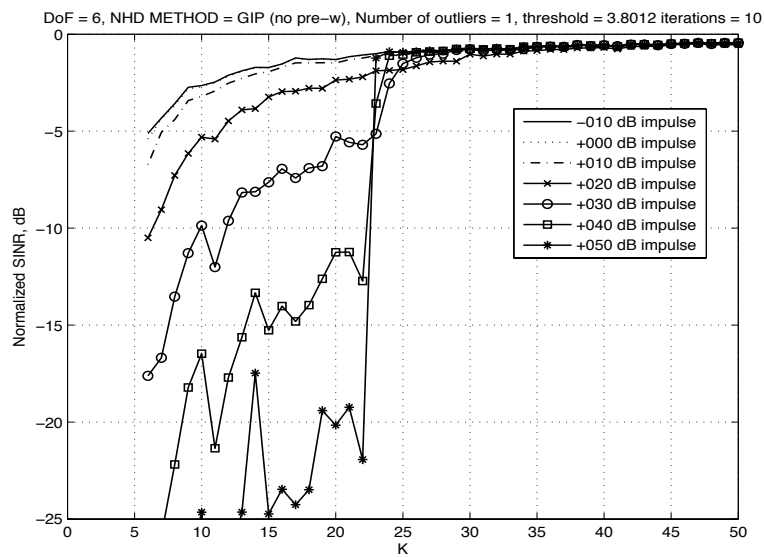


Figure 4.4: 6 Dimension R-GIP Results, 10 Iterations, 1 Outlier

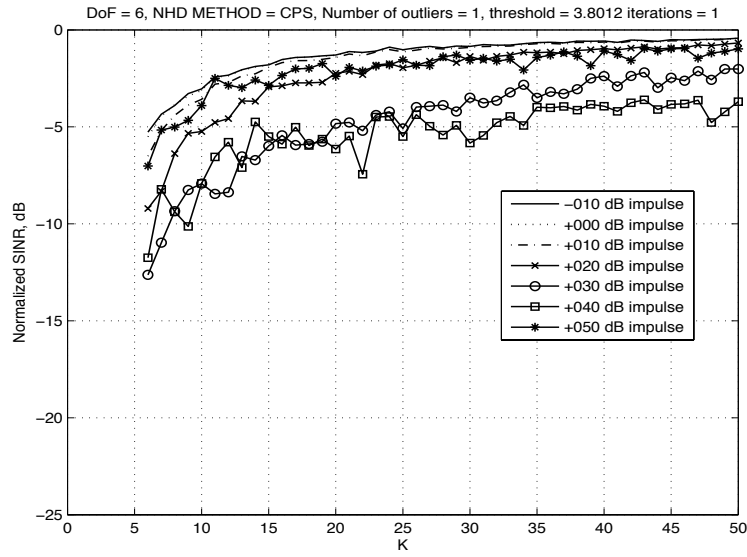


Figure 4.5: 6 Dimension CPS Results, 1 Outlier

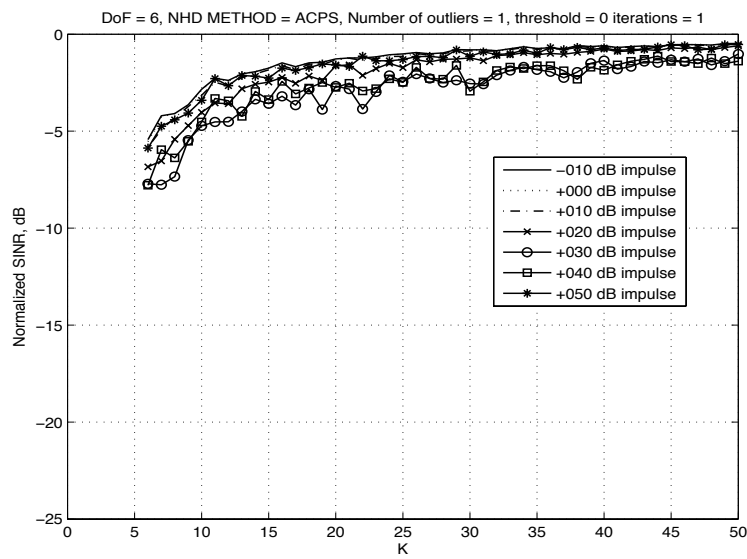


Figure 4.6: 6 Dimension ACPS Results, 1 Outlier

mensions are employed, and the diagnostic GIP test is applied along with CPS and ACPS, with results presented in Figures A.6 through A.9.

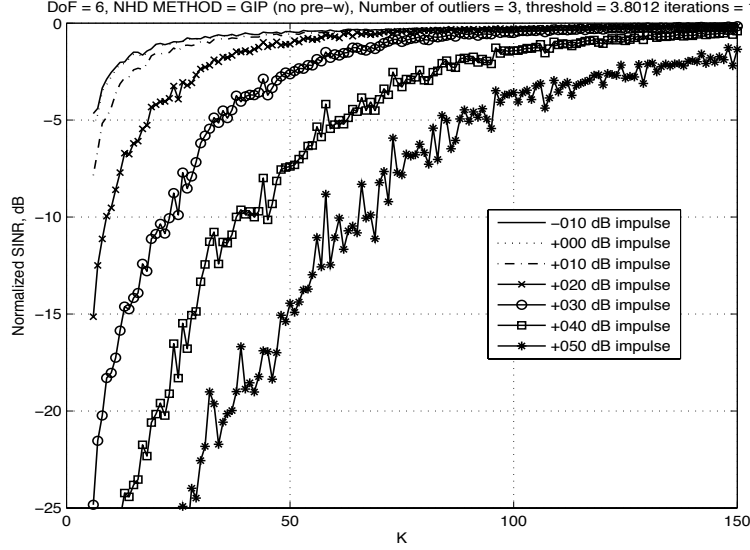


Figure 4.7: 6 Dimension GIP Results, 1 Iteration, 3 Outliers

As is seen from the figures, the STAP processor using the ACPS diagnostic test has the fastest SINR convergence as a function of sample support K for the given colored Gaussian data environment when compared to the state-of-the-art GIP diagnostic test, even when the GIP diagnostic test performs multiple iterations.

For a test of robustness in higher dimensions, those on the order of magnitude with those used in the MCARM STAP scenario, the number of complex dimensions is increased to 20, with no change in the other simulation parameters. Results of this more challenging scenario are presented in Figures A.16 through A.19.

As is seen from Figures A.16 through A.19, once again the STAP processor using the

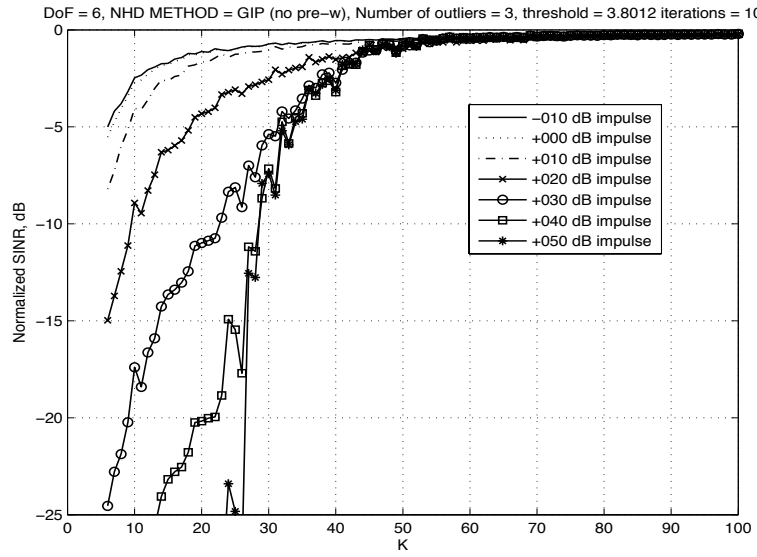


Figure 4.8: 6 Dimension R-GIP Results, 10 Iterations, 3 Outliers

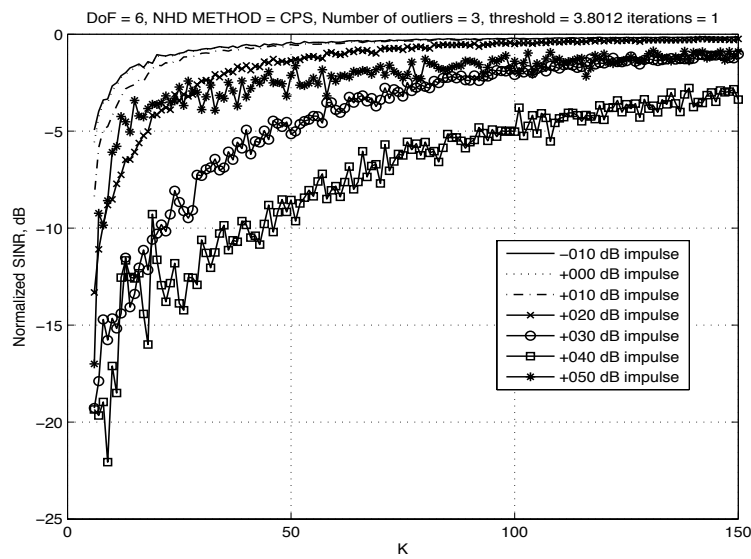


Figure 4.9: 6 Dimension CPS Results, 3 Outliers

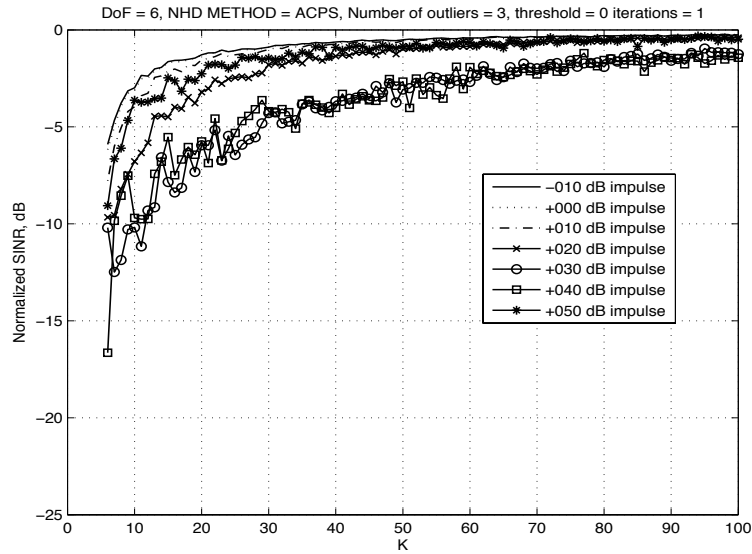


Figure 4.10: 6 Dimension ACPS Results, 3 Outliers

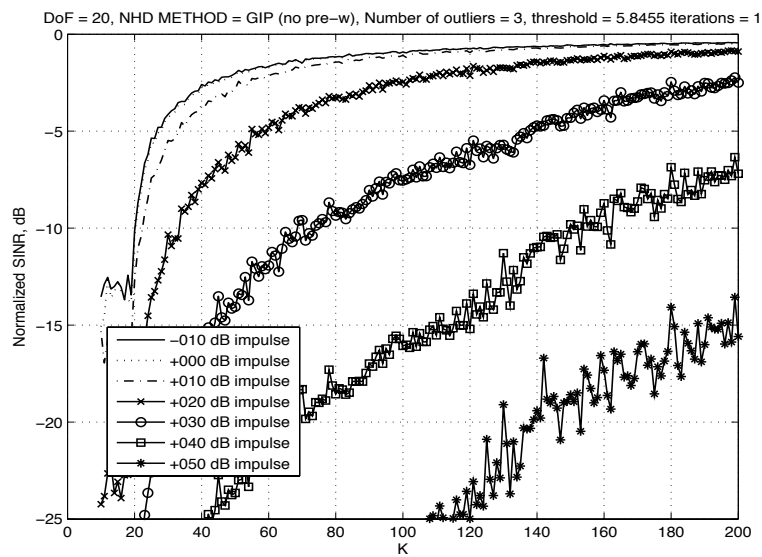


Figure 4.11: 20 Dimension GIP Results, 3 Outliers

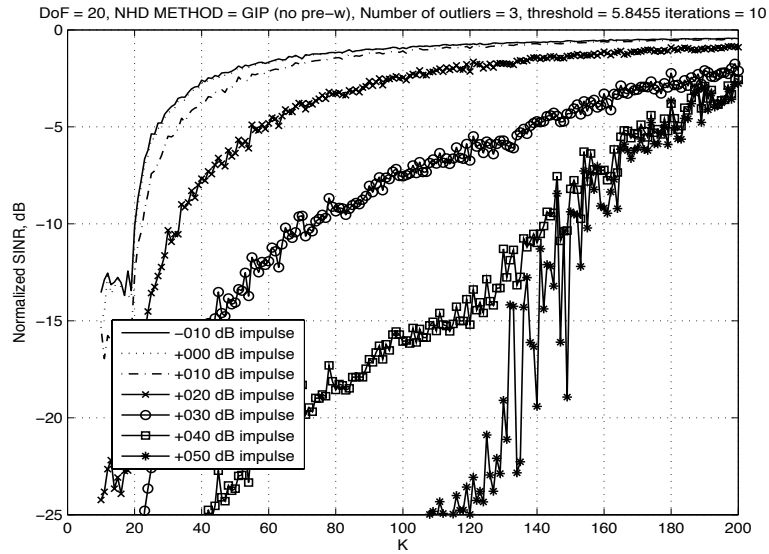


Figure 4.12: 20 Dimension R-GIP Results, 10 Iterations, 3 Outliers

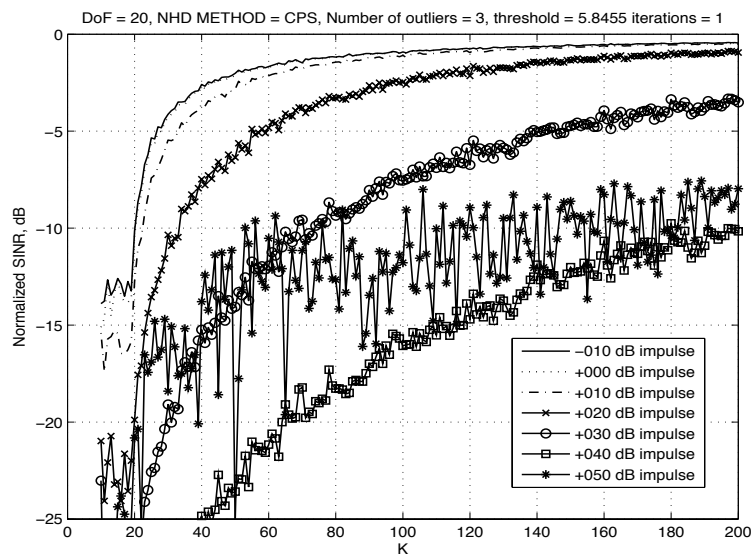


Figure 4.13: 20 Dimension CPS Results, 3 Outliers

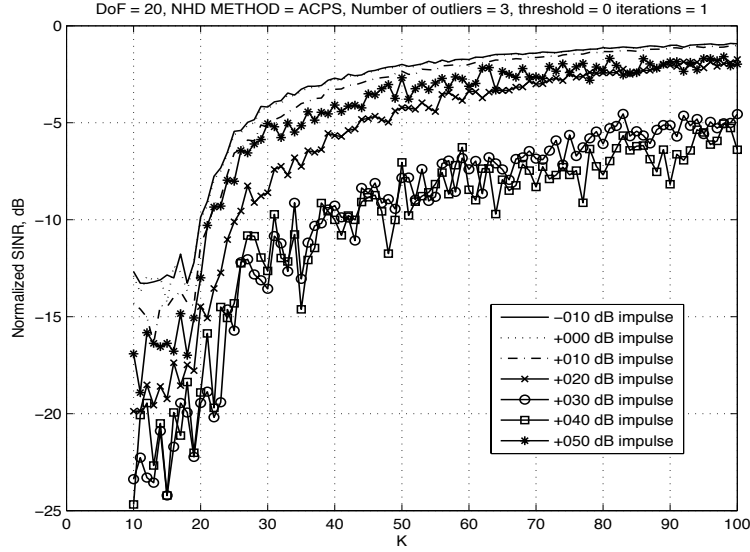


Figure 4.14: 20 Dimension ACPS Results, 3 Outliers

ACPS diagnostic test has the fastest SINR convergence as a function of sample support for the given colored Gaussian data environment when compared to the state-of-the-art GIP diagnostic test.

Tables of complete results for 2, 6, 10, 20, 30, and 40 dimensions are shown next for the 20 dB outlier power level. The metric developed for presentation here is a ratio of the number of observations required to reach 20 dB for the given method, divided by the number of observations predicted for convergence within 3 dB of the optimal, given by the rule from Reed, Mallett, and Brennan's seminal work in [54]. Formally, the metric is given by

$$\kappa_i = \frac{K_{sim}}{K_{RMB}}, \quad (4.16)$$

where K_{sim} is the empirically simulated number of observations, and $K_{RMB} = 2MN - 3$

is the rule derived in [54]. Detailed results for all power levels and dimensions are given in Appendix A.

Table 4.1: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000

Table 4.2: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.667	0.778	1.000	2.111	3.222
R-GIP	0.667	0.667	1.111	2.111	3.000
CPS	0.667	0.667	0.889	2.111	3.222
ACPS	0.667	0.667	0.667	1.444	2.444

Table 4.3: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.588	0.706	1.118	1.882	2.706
R-GIP	0.588	0.706	1.118	1.882	2.706
CPS	0.588	0.706	1.000	1.882	2.706
ACPS	0.588	0.647	0.941	1.588	2.412

Given these SINR convergence results, it is expected that the ACPS algorithm will perform in a superior manner when employed in a realistic STAP scenario in which multiple

Table 4.4: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.541	0.649	0.919	1.486	2.378
R-GIP	0.541	0.649	0.919	1.486	2.378
CPS	0.541	0.649	0.919	1.486	2.378
ACPS	0.270	0.568	0.703	1.189	1.811

Table 4.5: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.544	0.614	0.789	1.333	1.982
R-GIP	0.544	0.614	0.789	1.333	1.982
CPS	0.544	0.614	0.789	1.333	1.982
ACPS	0.228	0.561	0.702	1.123	1.632

Table 4.6: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.532	0.597	0.740	1.299	1.857
R-GIP	0.532	0.597	0.740	1.299	1.857
CPS	0.532	0.597	0.740	1.299	1.857
ACPS	0.455	0.545	0.649	1.052	1.506

outliers (targets) are present in the training data. This demonstration is performed in the following section.

4.4.2 STAP Application Example

The results presented herein use the MCARM Mission 575 data [102] as a clutter background, with two sidelobe barrage noise jammers simulated at 20 dB power above the estimated thermal noise floor of the MCARM data. The scenario uses eleven spatial DoF from a linear array and three temporal DoF, yielding a 33-DoF vector (complex) for each training sample. For estimating $\hat{\mathbf{R}}_{SMI}$ we use a subset of 80 training snapshots chosen symmetrically about the center seven cells made up of the CUT and three guard cells on either side of the CUT. The three guard cells and the CUT are excluded from the training data used in the formation of $\hat{\mathbf{R}}_{SMI}$ to avoid target leakage into the ideally target-free (i.e., outlier-free) training data set.

Two outlier scenarios are studied. In the first scenario, two target outliers are injected into the MCARM 575 clutter and jamming training snapshots. The target powers relative to the clutter and jamming level for this scenario are summarized in Table 1.

Table 4.7: Scenario 1 Injected Target Outlier Levels

Target Range Cell	Target Power Relative to Clutter/Jamming
307	0 dB
310	0 dB

Since range cells 307 and 310 are within the guard cell band of each other (i.e., within 3 range cells of the CUT), this is a benign multi-outlier scenario due to the guarantee of mutual exclusion from each other's CUT training data selection.

In the second scenario, many outliers (nine in total) with powers ranging from 0 dB to

5 dB above the combined clutter and jamming interference power are injected into several range cells in a random fashion. The target powers relative to clutter and jamming are summarized in Table 2.

Table 4.8: Scenario 2 Injected Target Outlier Levels

Target Range Cell	Target Power Relative to Clutter/Jamming
260	0 dB
270	5 dB
286	3 dB
290	0 dB
297	4 dB
307	0 dB
310	0 dB
312	4 dB
320	3 dB

All of these outlier snapshots have the form of a target-matched space-time steering vector \mathbf{s} . Target outliers have been shown to adversely affect convergence of linearly constrained adaptive beamformers [62]. They degrade SMI STAP radar performance by biasing $\hat{\mathbf{R}}_{SMI}$ and thus \mathbf{w}_{SMI} . Therefore, the detection and removal of these outliers prior to the formation of $\hat{\mathbf{R}}_{SMI}$ is highly desirable.

The respective STAP filter residual powers versus range are shown. The goal is to attain the maximum STAP residual SINR performance by obtaining the highest residual levels in the range cells of injected target outliers, and to simultaneously have low residual levels in adjacent range cells that do not contain targets, nominally referred to here as SINR.

Under all three NHD methods, if a snapshot is determined to be an outlier, that outlier snapshot is excised from the training data and the number of training data samples is reduced by one for each excised outlier snapshot. For comparison, SMI STAP performance is first shown for the case when detection of outliers in the training data is not attempted, i.e. when

no NHD is used.

Figures 4.15 through 4.18 correspond to Scenario 1 (two target outliers injected) and Figures 4.19 through 4.22 correspond to Scenario 2 (nine target outliers injected). Scenario 1 is meant to demonstrate performance of the three NHD algorithms under a benign multi-outlier scenario, and Scenario 2 demonstrates the performance of the NHD techniques under a more stressing outlier scenario.

Scenario 1 Results (Two Outliers) Figure 4.15 shows (as a baseline) that not using any form of NHD results in poor SINR. The outliers significantly affect the sample covariance matrix used in the SMI technique. The strong target outliers are visible with 5 – 8 dB SINR, but the multiple outliers appear to raise the local residue level by approximately 40 dB, masking typical smaller target powers expected in practice at long ranges. Thus, overall performance is poor.

Figure 4.16 shows the SMI-STAP performance when the snapshot power test is used. The outliers have again contaminated the sample covariance matrix, however slightly improved performance over the no-NHD case in Fig. 1 is seen by a smaller width in the local residue level. Performance is still poor due to this condition which reduces detection sensitivity over the higher noise region.

In Figure 4.17, the SMI-STAP performance is shown using the GIP NHD technique to remove outliers from the training data. The SMI-STAP processor once again indicates the injected targets in this particular scenario. However, the residual noise floor still is raised over most of the range cells, indicating that GIP NHD had trouble removing the outliers from the training.

In Figure 4.18, the SMI-STAP performance is shown when using the ACPS NHD tech-

nique for outlier removal. The processor clearly indicates both targets and the STAP residual noise floor is not affected nearly as much as for the case of GIP. The outliers have much less effect on the ACPS NHD test compared to the other two tests shown.

Scenario 2 Results (Nine Outliers) In Figure 4.19, we show that not using any form of NHD results in poor SMI-STAP residuals under the nine-outlier case. The outliers have significantly affected the sample covariance matrix used in the SMI technique, resulting in a raised residual noise floor over most of the range cells.

Figure 4.20 shows the SMI-STAP performance when the snapshot power test is used. This technique does not identify the injected targets in the training data surrounding each CUT, which leads to undesirable detection performance. This is due to the multiple outliers contaminating the average power estimate, which uses the sample mean of the powers from each range bin which is a non-robust estimator of the power level of the true interference background.

Figure 4.21, the SMI-STAP residual performance is shown using the GIP NHD technique to remove outliers. The SMI-STAP processor once again does not clearly indicate the injected targets since the GIP uses a sample covariance matrix, which is generally known to suffer from the masking effect in the presence of numerous outliers.

In Figure 4.22, the SMI-STAP performance is shown when using the ACPS NHD technique for outlier removal. The processor clearly indicates all nine targets when trained on the PS-cleaned training data. The ACPS NHD technique is shown to be robust, as multiple outliers have less effect on the test compared to the other two tests shown in Figs. 4.19 through 4.21 5 – 7. Clearly, ACPS NHD is the superior NHD technique based on the studied scenarios.

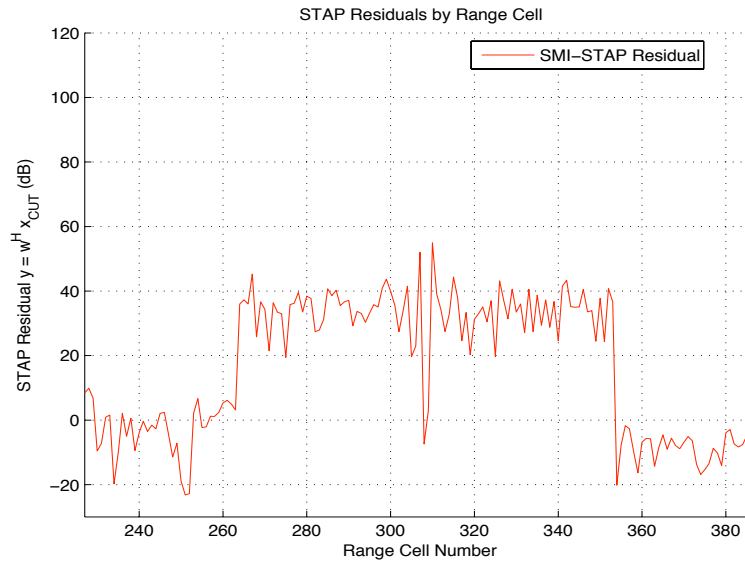


Figure 4.15: STAP residue performance vs. range cell with no outlier detection. Two injected targets at range cells 307 and 310. A detection threshold at ~ 45 dB makes target detection possible.



Figure 4.16: STAP residue performance vs. range cell using inner-product outlier detection. Two injected targets at range cells 307 and 310. Reliable target detection.



Figure 4.17: STAP residue performance vs. range cell using GIP outlier detection. Two injected targets at range cells 307 and 310. Reliable target detection.

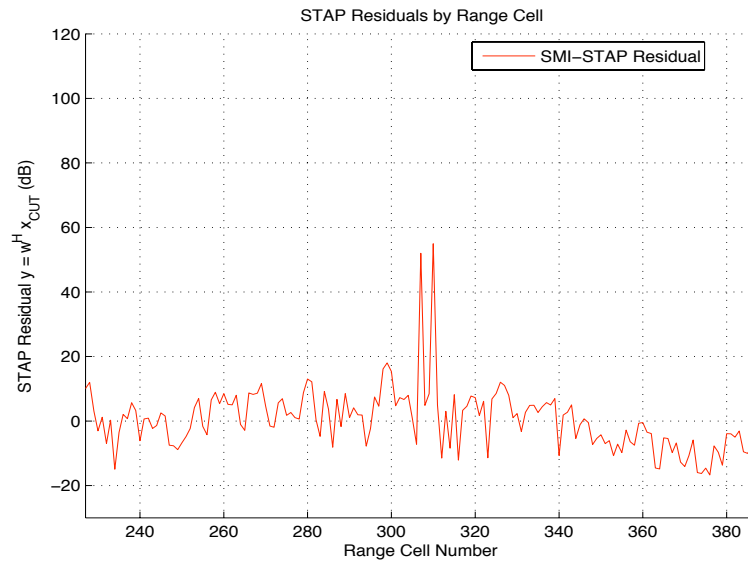


Figure 4.18: STAP residue performance vs. range cell using ACPS NHD. Two injected targets at range cells 307 and 310. Reliable target detection.

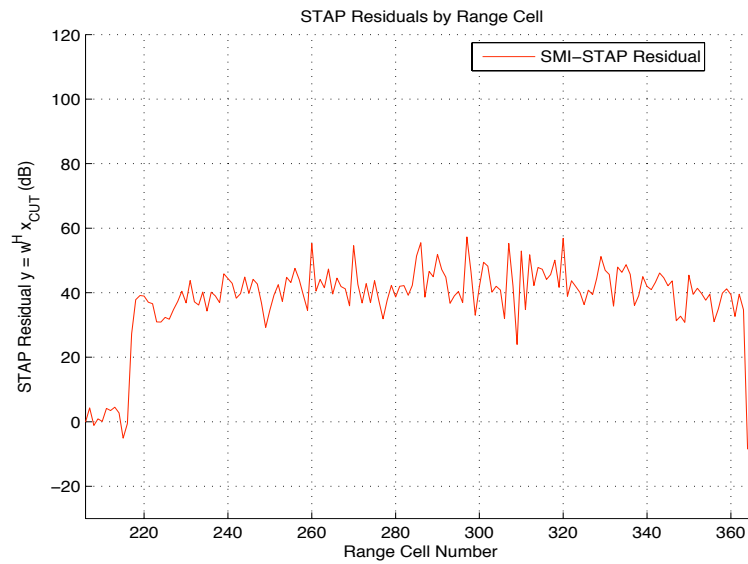


Figure 4.19: STAP residue performance vs. range cell with no NHD used. No reliable detection of targets.

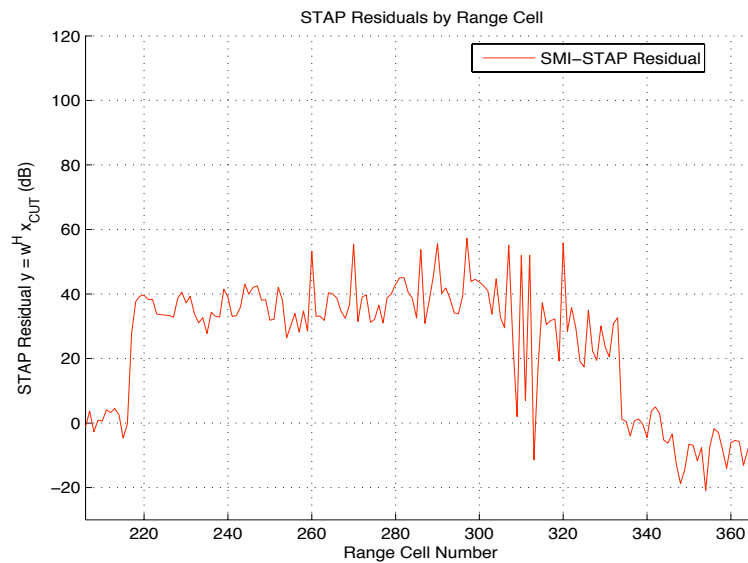


Figure 4.20: STAP residue performance vs. range cell using simple power-based NHD. Detection of targets compromised by non-robust residuals.

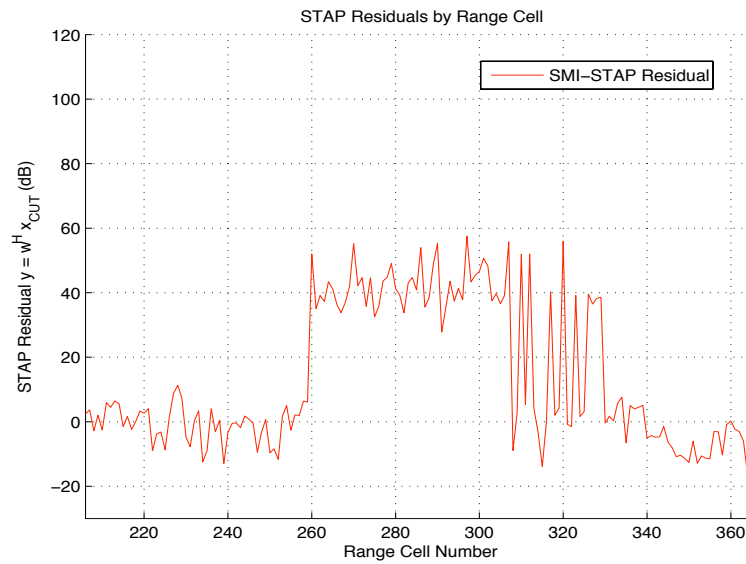


Figure 4.21: STAP residue performance vs. range cell using GIP-based NHD Detection of targets compromised by non-robust residuals.

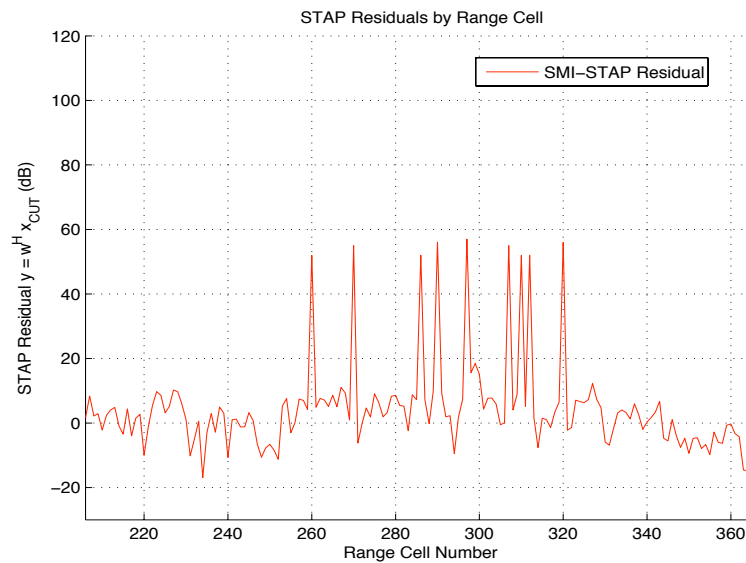


Figure 4.22: STAP residue performance vs. range cell using ACPS NHD. Reliable detection of targets.

4.4.3 Consideration of a Constant False Alarm Rate (CFAR)

In adaptive signal processors, the probability of false alarm P_{fa} is related to the detection threshold as well as the interference and noise power observed at the receiver. Due to the false alarm rate's dependence on interference and noise power, a class of processors termed *Constant False Alarm Rate* (CFAR) have been developed. A processor that enjoys a CFAR property is one that is typically designed using P_{fa} as the initial design parameter, based on which most other parameters are determined. The consideration of P_{fa} is critical for practicality of any adaptive processor responsible for detection, such as an adaptive radar. If false alarms occur too frequently, the confidence in the radar system approaches zero and eventually such a radar will be turned off by leadership or even the operator.

The adaptive processing literature is ripe with CFAR contributions since the mid 1970's, including [54] [103] [104] [55] [76]. Increasingly effective ways to choose an adaptive threshold have been proposed since that time. A limited number of these contributions included robust estimators as a part of the CFAR threshold determination. Unfortunately, even these CFAR methods did not employ the most powerful and effective estimators available in the robustness literature.

An appealing avenue of research is the analysis of order statistics CFAR (OS-CFAR). The OS-CFAR class of processors generally estimate the clutter power more robustly than the original average-based CFAR scaling factor in the classic works on CFAR such as [104] and [105].

Since the ACPS algorithm is based on the median, an order statistic, it is likely that a CFAR property can be developed for ACPS. Specifically, Röhling's average decision threshold (ADT) [103] has been linked to the functional form of the influence function (IF) of the clutter power estimator employed in OS-CFAR [106] [107]. Further research on this path appears

desirable, and may yield a CFAR assessment of the ACPS.

4.5 Summary

The ACPS algorithm proposed herein has been exercised in both a controlled SINR convergence simulation as well as a STAP target detection scenario based on real MCARM data. ACPS as an NHD technique is shown to improve adaptive performance in SMI-STAP when compared to other NHD techniques, including the state-of-the-art GIP statistical test.

Chapter 5

Robust Data Pre-whitening using ACPS

5.1 Motivation

As discussed in this dissertation, assumptions are critical in ensuring the expected performance from an adaptive processor. For instance, the optimal covariance matrix for a three-dimension i.i.d. multivariate Gaussian random process \mathbf{F}_x with zero mean and unit variance with three degrees of freedom has the covariance matrix

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.1)$$

which is of course the identity matrix. Graphically, the covariance has the spheroid structure as shown in Figure 5.1.

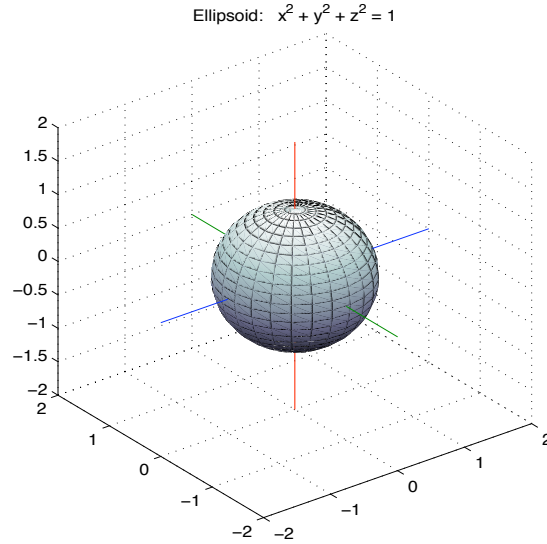


Figure 5.1: Covariance Visualization for a Three-Dimensional Pure Gaussian Process

While this is the ideal case, the reality of adaptive signal processing is that only a finite number of samples are observed. Even assuming that all observed samples are homogeneous (i.e., there are no observations generated by a process with a different distribution than \mathbf{F}_x), deleterious effects on the covariance matrix and ellipsoid are hard to avoid. For example, assume six samples of dimension three are drawn from \mathbf{F}_x ,

$$\mathbf{X} = \begin{bmatrix} -0.4293 & -0.4650 & 2.1122 & 1.0378 & 0.3155 & 1.9574 \\ 0.0558 & 0.3710 & -1.3573 & -0.3898 & 1.5532 & 0.5045 \\ -0.3679 & 0.7283 & -1.0226 & -1.3813 & 0.7079 & 1.8645 \end{bmatrix}. \quad (5.2)$$

While the covariance matrix assuming an infinite number of observations from \mathbf{F}_x is shown

in (5.1), the sample covariance matrix for the finite observations in (5.2) is given by

$$\hat{\mathbf{R}}_{\mathbf{x}} = \begin{bmatrix} 1.6450 & -0.3317 & 0.0165 \\ -0.3317 & 0.8003 & 0.7027 \\ 0.0165 & 0.7027 & 1.2662 \end{bmatrix}. \quad (5.3)$$

The ellipsoid representative of this covariance is shown in Figure 5.2.

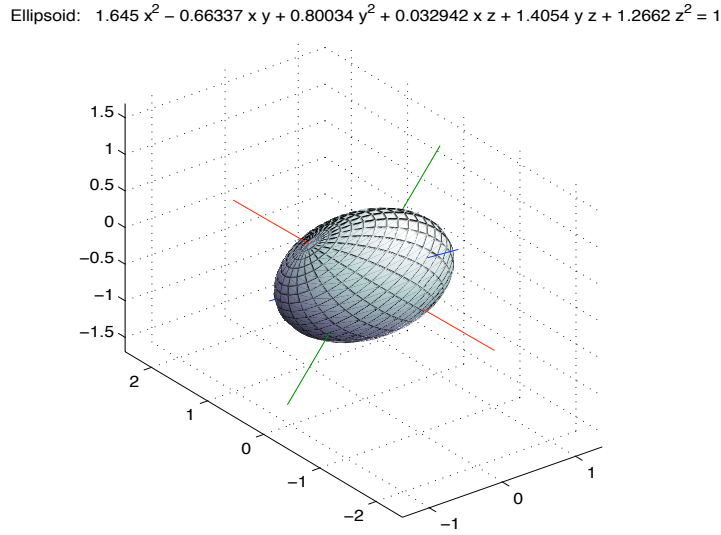


Figure 5.2: Covariance Visualization for Finite Samples Drawn From \mathbf{F}_x .

In addition to the effects of a finite number of observations, real-world signal processors are often presented with outliers in the observed data. In this case, an outlier is an observed data point that is generated by a random process other than \mathbf{F}_x . Let us now examine the effects of observed outliers and how such outliers present an ellipsoid deviant from that which is assumed under \mathbf{F}_x . Assume that \mathbf{X} in (5.2) is contaminated by a strong observation in data values (2,4) and (3,4), which could, for example, represent auxiliary antenna elements presenting higher voltages due to strong interference at the time of observation 4. The new

observation matrix is then given by

$$\mathbf{X} = \begin{bmatrix} -0.4293 & -0.4650 & 2.1122 & 1.0378 & 0.3155 & 1.9574 \\ 0.0558 & 0.3710 & -1.3573 & 5.000 & 1.5532 & 0.5045 \\ -0.3679 & 0.7283 & -1.0226 & 5.000 & 0.7079 & 1.8645 \end{bmatrix}, \quad (5.4)$$

with the sample covariance of \mathbf{X} now given by

$$\hat{\mathbf{R}}_{\mathbf{x}} = \begin{bmatrix} 1.6450 & 0.6006 & 1.1203 \\ 0.6006 & 4.9417 & 4.7796 \\ 1.1203 & 4.7796 & 5.1148 \end{bmatrix}. \quad (5.5)$$

The resulting ellipsoid is even more dissimilar to the assumed ellipsoid for the process \mathbf{F}_x , as depicted in Figure 5.3.

Ellipsoid: $1.645 x^2 + 1.2012 x y + 4.9417 y^2 + 2.2405 x z + 9.5593 y z + 5.1148 z^2 = 1$

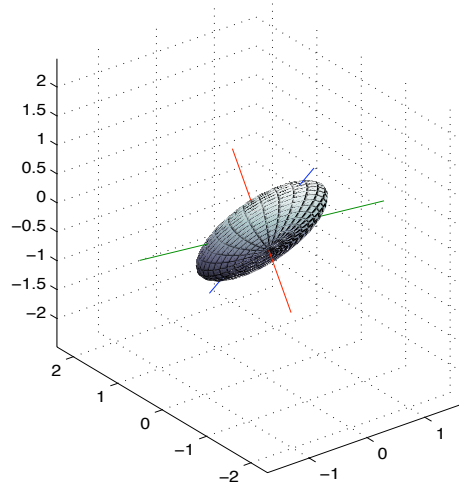


Figure 5.3: Covariance Visualization for Finite Samples Drawn From \mathbf{F}_x . Single Contaminated Data Value.

In summary, the assumptions involved in adaptive signal processing are rarely, if ever,

representative of the observations, and can lead to a covariance structure that is not representative of the assumed distribution. A three-dimensional example has been utilized herefor visualization purposes. However, it is important to note that the effects described occur in higher dimensions, as well.

The method of *data pre-whitening* has the effect of turning the covariance ellipsoid into a more spherical shape. As is evident from the figures, if this covariance matrix is used as observed without mitigation or investigation of outliers, unexpected performance may result. The process of data pre-whitening is at times used to lessen the impact of such deviant observations. As will be demonstrated however, classical data pre-whitening is not guaranteed to help the problem, and may even further impact performance negatively.

A robust data pre-whitening method is presented in this chapter. The new algorithm incorporates the powerful outlier diagnosis capability of ACPS into a data pre-whitening operation. This allows for objective and adaptive diagnosis of outlier observations before they can degrade the performance of classical data-prewhitening. Further, the presented method benefits from robust identification of multiple outliers post-whitening, and outperforms the compared techniques. The SINR convergence measure of effectiveness is used once again.

5.2 Classical Data Pre-whitening

The term *pre-whitening* is typically used to describe a data filter capable of producing a statistically uncorrelated output signal from an input signal that may be colored. The effects of pre-whitening include the flattening of spectral components of the input data exhibiting linear dependencies. Such a feature allows for the improved detection of and distinction of desired signals that may have small amplitudes compared to higher-powered colored interference. Pre-whitening is a powerful yet simple tool in signal processing, and is

often underutilized [25].

Besides STAP, the applications of pre-whitening are numerous. For example, digital cellular phones do not directly transmit a user's digitized voice, as this is costly in terms of bandwidth. Instead, a pre-whitening FIR filter is typically created and applied to the speech signal to flatten its spectrum. The speech signal can then be reconstructed at the receiver by transmitting just the power level of the *whitened* signal along with the pre-whitening filter coefficients. This technique requires less information bandwidth and allows cellular carriers to pack more users into their allotted spectrum.

Initial discussions concerning the need for data pre-whitening and early approaches dates back to Blackman and Tukey (1958) [25]. The focus of the following pre-whitening discussion is targeted for application to adaptive processing. Pre-whitening of data is performed as follows. Suppose that K original N -dimensional snapshots \mathbf{x}_i are arranged in columns of a matrix $\mathbf{\Psi}$, as in

$$\mathbf{\Psi} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_K]. \quad (5.6)$$

The sample covariance matrix of the original snapshots \mathbf{x}_i is then constructed using (3.6) as in

$$\hat{\mathbf{R}}_{\chi} = \mathbf{\Psi} \mathbf{\Psi}^H. \quad (5.7)$$

The original data are then whitened by multiplying $\mathbf{\Psi}$ by the inverse of the square root of $\hat{\mathbf{R}}_{\chi}$, yielding the whitened $\mathbf{\Psi}$ given by

$$\mathbf{\Psi}_W = \hat{\mathbf{R}}_{\chi}^{-\frac{1}{2}} \mathbf{\Psi}. \quad (5.8)$$

The columns of Ψ_W are now used as the N -dimensional snapshots to be processed. This conventional form of data pre-whitening has been applied to adaptive radar applications, including STAP [24] [33] [108] [109].

Using two dimensions for visualization purposes, it is straightforward to introduce here the problems that classical pre-whitening may present in diagnosing outliers. A simple simulation involving 25 Gaussian-distributed data points with a correlation coefficient of $\rho = .9$ is described to show an example. Outliers are placed in data points 10, 11, and 12, and marked by a circle on the Figures to follow. These data point numbers are chosen for no particular reason. In fact, there is no known dependency on data point number in this scenario. Figures 5.4 and 5.5 show the data values before and after pre-whitening, respectively.

The problems in outlier diagnostics are subtle and can be easily missed if not examined carefully. Specifically, the outliers are seen to move inward toward the coordinate-wise center of the sample data. This is expected in a pre-whitening operation. However, the outlier diagnostics are affected in a large way, evidenced in Table 5.1.

As seen in the table, the GIP and MD suffer from the masking effect, allowing several non-outlier data points to be identified as outliers. Even if an adaptive threshold was used, the statistics do not allow a cut-off to be drawn to identify the three outliers at points 10, 11, and 12. The $ACPS_i$ values are the only reliable way to detect the outliers before pre-whitening. After classical pre-whitening, none of the methods identify all outliers. The pre-whitening operation has destroyed the ability to diagnose outliers in this situation, either robustly or classically. This is the essence of the disadvantage of classical data pre-whitening and why it may cause problems.

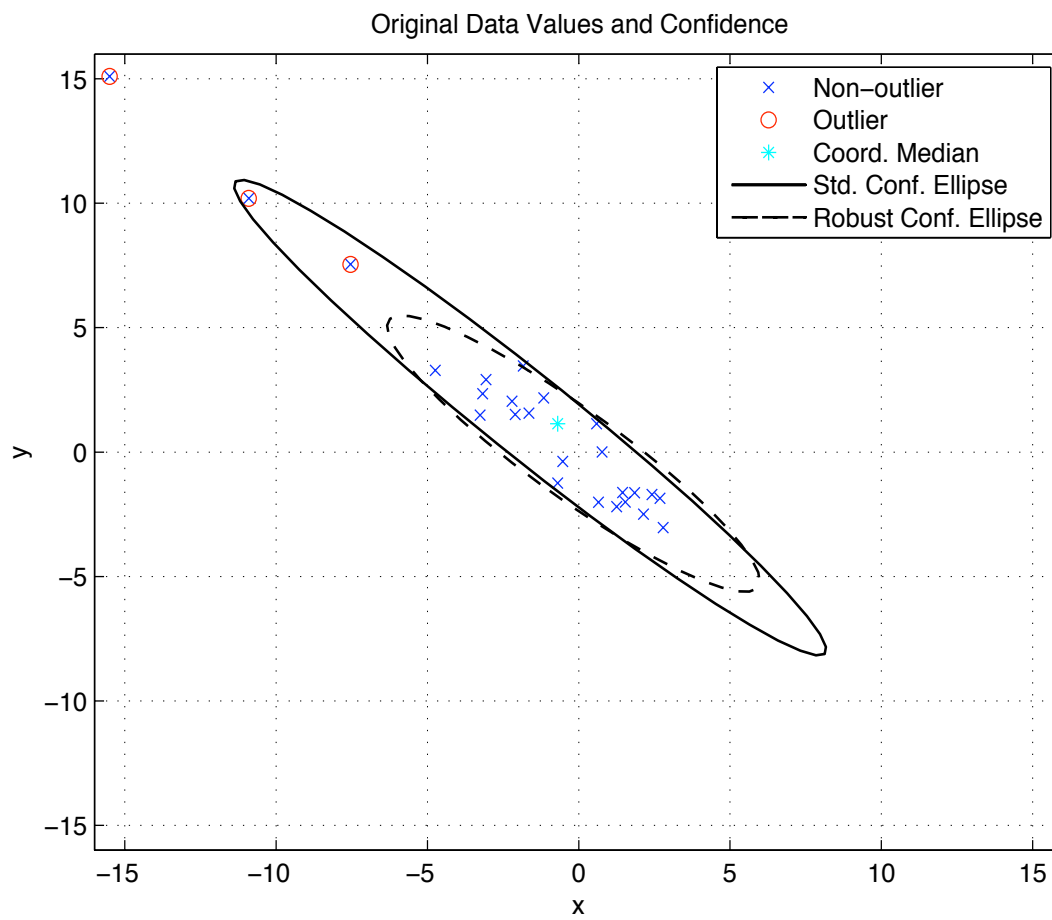


Figure 5.4: Two-dimensional Data Before Classic Pre-whitening. 97.5% Confidence Ellipse Shown.

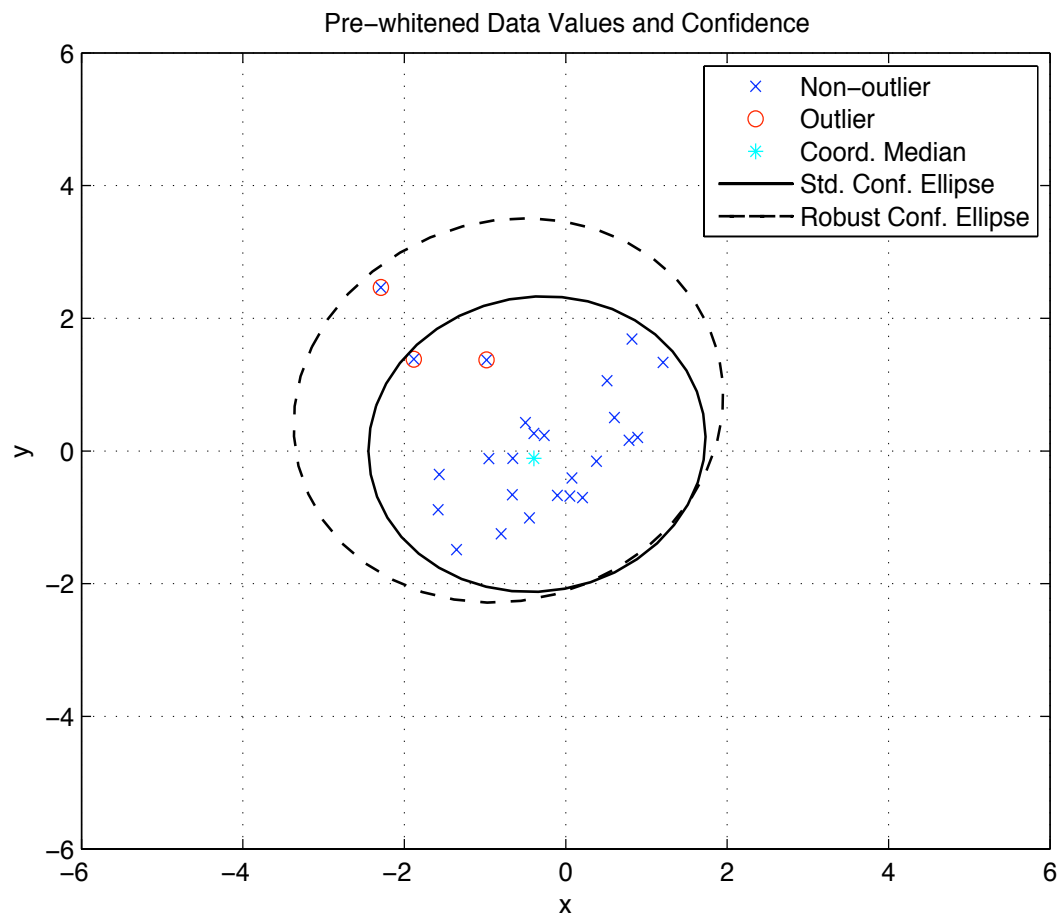


Figure 5.5: 97.5% Confidence Ellipses Shown for Non-robust (solid) and Robust (dashed) Outlier Detection Following Classical Pre-whitening.

Table 5.1: Outlier Diagnostic Statistics for MD, GIP, and ACPS

i	Before Classic Pre-whitening			After Classic Pre-whitening			remarks
	MD_i	GIP_i	$ACPS_i$	MD_i	GIP_i	$ACPS_i$	
1	0.696	0.477	0.825	0.696	0.477	0.187	
2	1.255	1.269	0.995	1.255	1.269	0.513	
3	0.714	0.177	0.476	0.714	0.177	0.016	
4	0.844	0.486	1.277	0.844	0.486	0.375	
5	3.967	<u>3.664</u>	0.573	3.967	<u>3.664</u>	2.343	
6	0.025	0.130	0.472	0.025	0.130	0.755	
7	1.794	0.871	1.251	1.794	0.871	0.928	
8	0.145	0.462	0.195	0.145	0.462	0.640	
9	1.193	0.646	0.649	1.193	0.646	0.496	
10	10.379	<u>11.795</u>	<u>15.059</u>	10.379	<u>11.795</u>	<u>6.371</u>	Injected Outlier 1
11	4.531	<u>5.693</u>	<u>9.724</u>	4.531	<u>5.693</u>	<u>3.879</u>	Injected Outlier 2
12	2.144	<u>2.956</u>	<u>6.221</u>	2.144	<u>2.956</u>	2.322	Injected Outlier 3
13	2.618	<u>3.423</u>	0.751	2.618	<u>3.423</u>	1.569	
14	0.447	0.959	0.925	0.447	0.959	0.183	
15	0.028	0.234	0.113	0.028	0.234	0.518	
16	1.498	0.667	0.986	1.498	0.667	0.730	
17	4.170	<u>3.364</u>	0.826	4.170	<u>3.364</u>	2.437	
18	0.679	0.910	0.691	0.679	0.910	0.172	
19	1.716	1.438	0.484	1.716	1.438	1.081	
20	1.855	2.689	2.490	1.855	2.689	0.856	
21	1.072	0.562	1.776	1.072	0.562	0.590	
22	2.027	2.289	0.829	2.027	2.289	1.170	
23	0.133	0.449	1.082	0.133	0.449	0.068	
24	0.500	0.178	0.466	0.500	0.178	0.154	
25	3.570	<u>4.211</u>	0.113	3.570	<u>4.211</u>	2.069	

5.3 Robust Data Pre-whitening based on ACPS

A key contribution of this dissertation will now be demonstrated in the form of a new robust data pre-whitener based on the proposed ACPS algorithm from Chapter 4.

5.3.1 Procedure

The procedure for robust data pre-whitening is as follows. First, the data are to be initially screened for outliers. As demonstrated in the previous section, a robust method must be used to avoid the masking effect. Otherwise, non-outlier data may be removed in error, or outliers may be missed. Once the outliers are robustly diagnosed, they are removed from the data set temporarily to enable the computation of a finite-sample robust covariance matrix. Denote this trimmed data set by $\tilde{\Psi}$. Compute the SCM using (2.52), yielding $\tilde{\mathbf{R}}$. This robust covariance is to be used in place of the normal covariance in data pre-whitening. Perform the pre-whitening operation in (5.8), yielding the robustly whitened data set $\tilde{\Psi}_{RW}$ as in

$$\tilde{\Psi}_{RW} = \tilde{\mathbf{R}}_{\tilde{H}}^{-\frac{1}{2}} \tilde{\Psi}. \quad (5.9)$$

The robust pre-whitening procedure, when applied to the scenario from Section 5.2, yields the following improved confidence ellipses and diagnostic statistics.

The diagnostic statistics based on the robustly pre-whitened data are listed in Table 5.2.

As seen in the table, the GIP and MD once again suffer from the masking effect, allowing an undesirable number of good data points to be identified as outliers along with the outliers themselves. Even if an adaptive threshold was used, the statistics do not allow a cut-off to

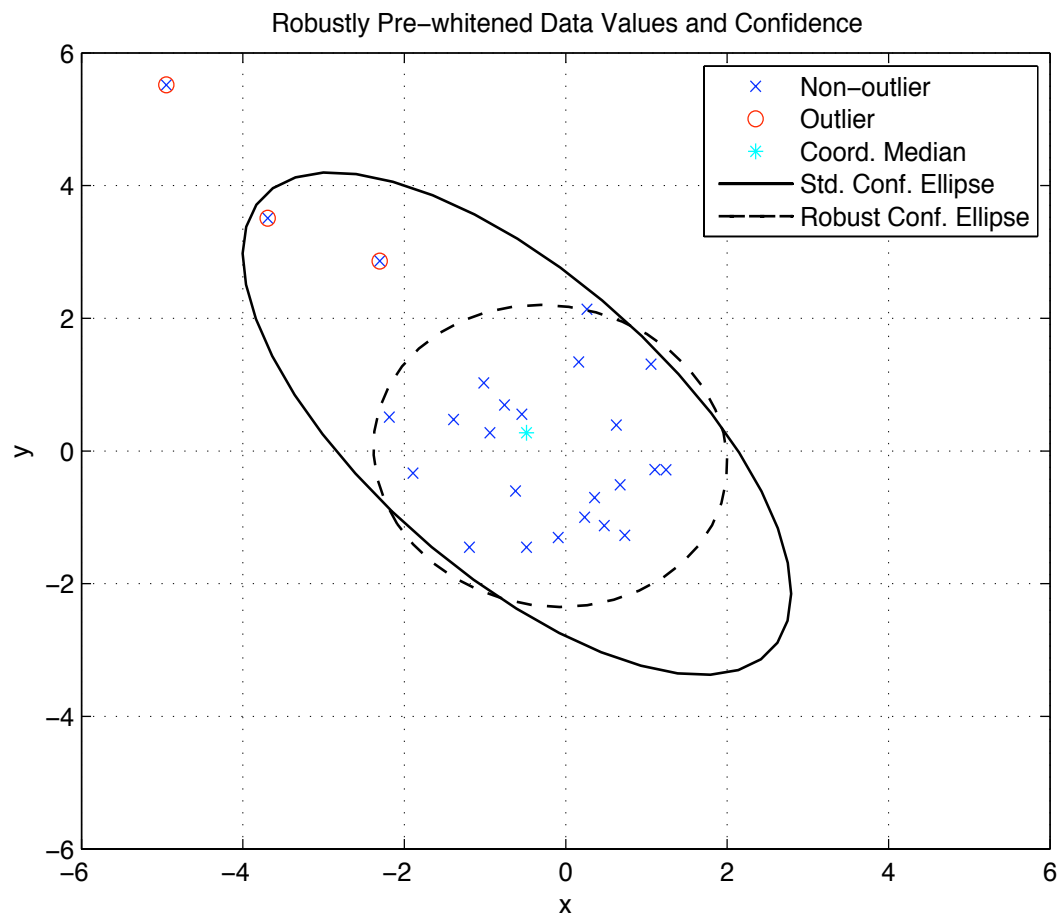


Figure 5.6: 97.5% Confidence Ellipses Shown for Non-robust (solid) and Robust (dashed) Outlier Detection Following Robust Pre-whitening.

Table 5.2: Outlier Diagnostic Statistics on Robustly Pre-whitened Data

i	MD_i	GIP_i	$ACPS_i$	remarks
1	0.696	0.477	0.075	
2	1.255	1.269	0.548	
3	0.714	0.177	0.010	
4	0.844	0.486	0.293	
5	3.967	<u>3.664</u>	<u>2.269</u>	
6	0.025	0.130	0.804	
7	1.794	0.871	1.060	
8	0.145	0.462	0.535	
9	1.193	0.646	0.526	
10	10.379	<u>11.795</u>	<u>7.539</u>	Injected Outlier 1
11	4.531	<u>5.693</u>	<u>4.728</u>	Injected Outlier 2
12	2.144	<u>2.956</u>	<u>2.768</u>	Injected Outlier 3
13	2.618	<u>3.423</u>	1.338	
14	0.447	0.959	0.106	
15	0.028	0.234	0.474	
16	1.498	0.667	0.834	
17	4.170	<u>3.364</u>	2.101	
18	0.679	0.910	0.157	
19	1.716	1.438	0.900	
20	1.855	2.689	1.172	
21	1.072	0.562	0.559	
22	2.027	2.289	1.065	
23	0.133	0.449	0.044	
24	0.500	0.178	0.170	
25	3.570	<u>4.211</u>	1.751	

be drawn to identify only the three outliers at points 10, 11, and 12. The $ACPS_i$, on the other hand, allow ready identification of the outliers, along with data point number 5 which shows large leverage in all statistical tests.

With further refinement of the adaptive threshold in the ACPS algorithm, the new statistical test can identify only the outliers at 10, 11, and 12. This is not possible with the MD or GIP statistic due to the masking effect causing increased statistic values for good data points above and beyond the statistics for the true outliers.

The robust data pre-whitening procedure is summarized in Algorithm 2.

Algorithm 2 Robust Data Pre-whitening using ACPS

Input: A set of K observed P -dimensional complex data vectors $\mathbf{\Psi} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]$.

Complex element p of observation k is denoted as h_{pk} . {In STAP, $P = MN$.}

Output: Robustly whitened data $\tilde{\mathbf{\Psi}}_{RW} = [\mathbf{h}_{RW}(1), \mathbf{h}_{RW}(2), \dots, \mathbf{h}_{RW}(K)]$

A. Perform the ACPS algorithm on the observations using on Algorithm 1, yielding the vector of ACPS values \mathbf{ACPS} and vector of identified outlier observation numbers \mathbf{out}_a .

Let $n_o = \text{length}(\mathbf{out}_a)$. Preserve the matrix $\tilde{\mathbf{\Psi}}$ used in this step.

B. Form $\tilde{\mathbf{\Psi}}$ from $\mathbf{h}_i, \forall i = 1, \dots, K, i \notin \mathbf{out}_a = [out_{a1}, \dots, out_{ano}]$.

C. Compute the robust SCM for $\tilde{\mathbf{\Psi}}, \tilde{\mathbf{R}}_{\tilde{H}} = \tilde{\mathbf{\Psi}}\tilde{\mathbf{\Psi}}^T$.

D. Whiten all observations using $\tilde{\mathbf{R}}_{\tilde{H}}$ by computing $\tilde{\mathbf{\Psi}}_{RW} = \tilde{\mathbf{R}}_{\tilde{H}}^{-\frac{1}{2}}\tilde{\mathbf{\Psi}}$. This is the desired output of robustly whitened observations.

5.3.2 STAP Example

For a test of robustness in higher dimensions, 20 complex dimensions were again employed in the SINR convergence scenario outlined in Section 4.4.1 the number of complex dimensions is increased to 20, with no change in the other simulation parameters. Results of this more challenging scenario are presented in Figures 5.7 through A.20.

As is seen from Figures 5.7 - A.20., the STAP processor using the robustly pre-whitened ACPS-based diagnostic tests exhibit the fastest SINR convergence as a function of sample

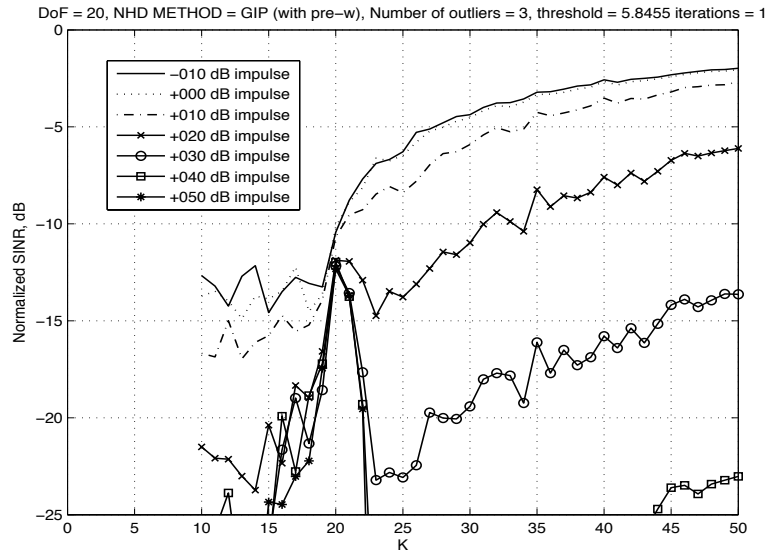


Figure 5.7: 20 Dimension GIP Results. 3 Outliers.

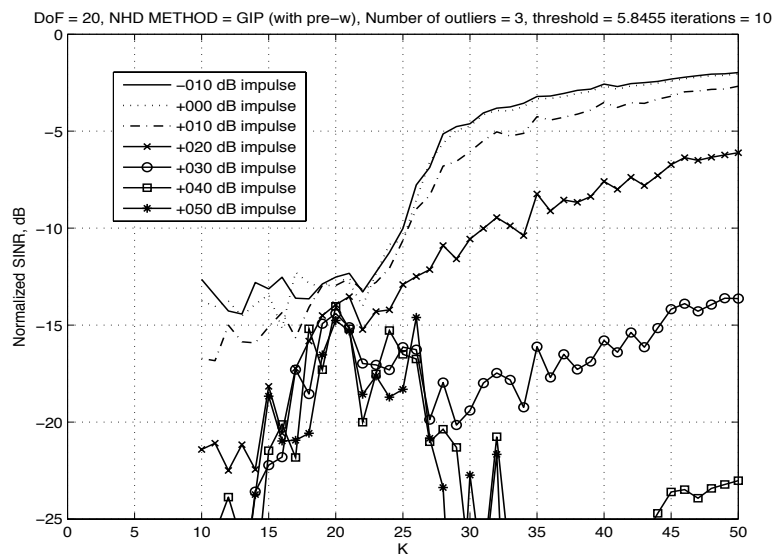


Figure 5.8: 20 Dimension R-GIP Results. 10 Iterations, 3 Outliers.

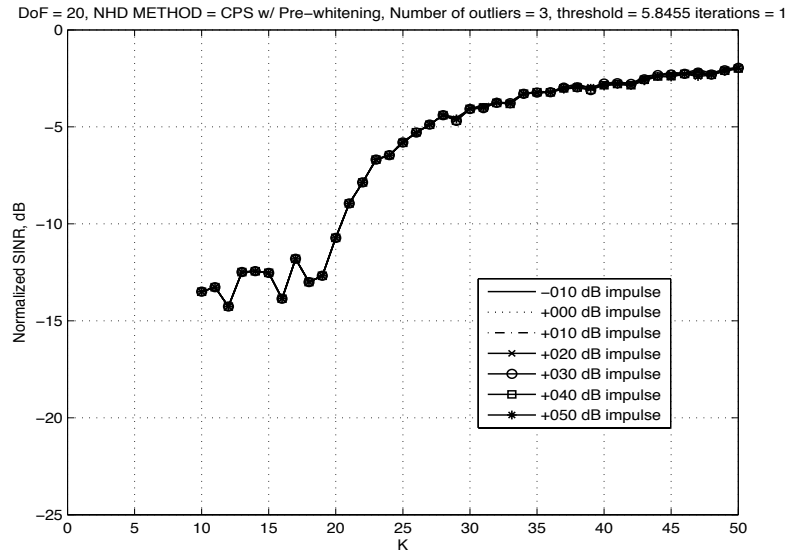


Figure 5.9: 20 Dimension CPS Results. 3 Outliers.

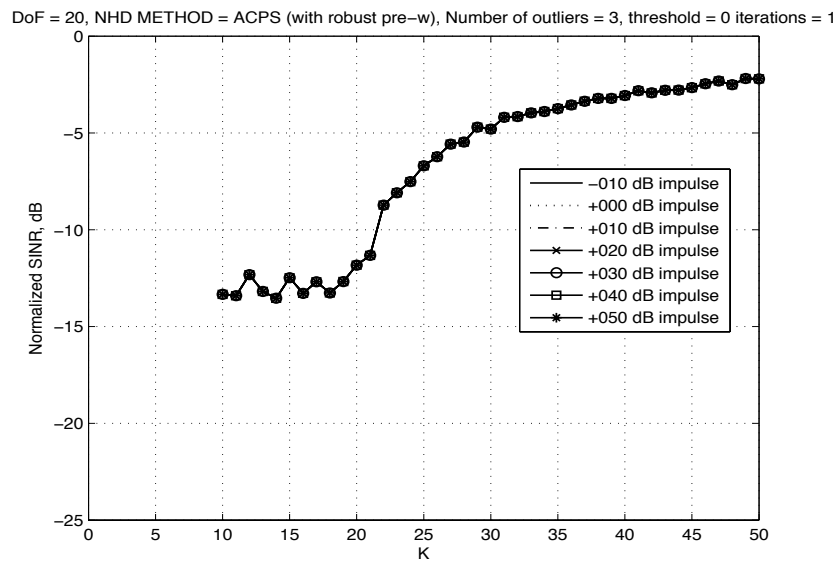


Figure 5.10: 20 Dimension ACPS Results. 3 Outliers.

support for the given colored Gaussian data environment when compared to the state-of-the-art GIP diagnostic tests (regardless of GIP iterations). The performance improvements over the GIP in 20 complex dimensions are significant. The Reed-Mallett-Brennan 3-dB rule [54] dictates that in this simulation's dimensionality, that at $K = (2MN - 3) = 37$ training samples, the SMI-STAP adaptive processor shall converge within 3dB of optimal SINR assuming uncontaminated, Gaussian-distributed observations.

Evident from the figures, only with the ACPS-based robust pre-whitening diagnostic tool can the *contaminated* data set to be processed by the SMI-STAP adaptive processor and continue to enjoy the convergence properties of an i.i.d. Gaussian observation set.

Employing 20, 30, and 40 complex dimensions, Tables 5.3 through Tables 5.5 present the SINR convergence performance measured by κ_i (as defined in (4.16) for the case of three 20 dB, target matched outliers contaminating the STAP scenario presented in Section 4.4.1. The performance of the presented robust pre-whitening method is listed in the last row of each table.

Table 5.3: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.459	0.541	0.865	>1.351	≫ 1.351
R-GIP	0.405	0.514	0.865	>1.351	≫ 1.351
CPS	0.270	0.270	0.568	0.730	1.000
ACPS	0.270	0.270	0.595	0.784	1.108

Each convergence milestone (-20 dB, -15 dB, -10 dB, -5 dB, and -3 dB) is reached first by the new methods proposed in this dissertation, namely ACPS and robust pre-whitening. In all cases, the convergence of the new robust pre-whitening method presented here is the best and attains an SINR within 3 dB of the optimal value at a point extremely close to

Table 5.4: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.544	0.614	0.789	1.333	1.982
R-GIP	0.544	0.614	0.789	1.333	1.982
CPS	0.544	0.614	0.789	1.333	1.982
ACPS	0.228	0.561	0.702	1.123	1.632
Robust PW	0.175	0.228	0.579	0.789	1.070

Table 5.5: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.532	0.597	0.740	1.299	1.857
R-GIP	0.532	0.597	0.740	1.299	1.857
CPS	0.532	0.597	0.740	1.299	1.857
ACPS	0.455	0.545	0.649	1.052	1.506
Robust PW	0.130	0.519	0.571	0.792	1.039

that predicted by the RMB rule for Gaussian data. This extremely robust performance in contamination is the desired result. Comprehensive SINR convergence results for 2, 6, 10, 20, 30 and 40 complex dimensions are presented in Appendix A.

5.4 Summary

In summary, a new data pre-whitening method based on ACPS has been proposed and demonstrated. From an SINR perspective, the new method outperforms other method, offers a dramatic improvement in SINR convergence rate to near that predicted for an *uncontaminated* set of observations, and is able to offer such improvements in an adaptive manner.

Chapter 6

The GM-estimator for Complex-Valued Data

The robustness literature is rich with various formulations of M-estimators and GM-estimators. The bounded-influence of the GM-estimator, regardless of its relatively low breakdown point in very high dimensions, is a desirable property for adaptive signal processors, in particular. The exemplar application of GM-estimators in practice is in power system state estimation, where Mili et al. [8] have initiated robust nonlinear regression using the GM-Estimator. Electric power systems have a very large number of state variables to be estimated, often in the hundreds or even thousands.

The GM-estimators in the literature exclusively deal with real-valued data. The application of GM-estimators to applications involving complex-valued data has not been developed or proposed in the published literature. The GM-estimator developed herein is derived to use complex-valued data and performance using this new estimator is determined and compared to theoretical performance prediction.

In Section 3.4.5, the GM-estimator of the desirable bounded-influence Schweppe type was shown to be the solution to the minimization of a cost function J with respect to the weight vector \mathbf{x} , given by

$$\min_{\mathbf{x}} J(\mathbf{x}) = \sum_{i=1}^m w^2(\mathbf{h}_i) \rho\left(\frac{r_i}{\sigma_i w(\mathbf{h}_i)}\right). \quad (6.1)$$

based on the regression model from (3.40),

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}, \quad (6.2)$$

with the residuals given by (3.41) and repeated here for convenience,

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\mathbf{x}, \quad (6.3)$$

$$r_i = z_i - \mathbf{h}_i^T \mathbf{x}_i. \quad (6.4)$$

The response vector \mathbf{z} , the column vectors \mathbf{h}_i of \mathbf{H}^T , and the weight vector \mathbf{x} are all real-valued per the derivation of the classical Schweppe-type GM-estimator. Also real-valued is the residual vector \mathbf{r} and standard deviation σ_i .

In this chapter, we develop the GM-estimator for complex-valued data, termed *CGM-estimator* for short. It will employ the Huber ρ and ψ functions shown in (3.45) and (3.46) and depicted in Figure 3.2 for a tuning parameter b equal to 2. In the CGM-estimator, a tuning parameter $b = 1.2$ was used, although other values of b provided similar performance, since the Huber function's gradual roll-off allows for a less sensitive tuning (as opposed to the hard-cutoffs used in the diagnostic approaches of Chapters 4 and 5).

6.1 The Choice of $\rho(\mathbf{r}) = \mathbf{r}^2$

If $\rho(\mathbf{r}) = \mathbf{r}^2$, the objective function of the GM-Estimator is equivalent to that of the Wiener filter, as in the asymptotic case, it reduces to minimizing the mean squared error. The GM-Estimator, and its new extension the CGM-estimator, are indeed generalizations of the Wiener filter used heavily in adaptive signal processing. The key advantage of the GM- and CGM-estimators is that the objective function can be chosen based on robustness and efficiency requirements. Thus the GM- and CGM-estimators can be thought of as robust versions of the Wiener filter given the appropriate $\rho()$ function is chosen for the application of interest.

6.2 The CGM-estimator Weighting Function

The weight function of observation i , $w(\mathbf{h}_i)$, for high data redundancy is given by (3.52) and restated here for convenience as

$$w(\mathbf{h}_i) = \min \left(1, \frac{b}{PS_i^2} \right), \quad (6.5)$$

where PS_i is the projection statistic computed for observation i , and b is the 97.5-percentile of the chi-square distribution for projection statistics, given by

$$b = \chi_{N, .975}^2, \quad (6.6)$$

where N is the number of dimensions of the observations. Based on the increase in SINR convergence rate shown using an adaptive threshold rather than a fixed PS cut-off value, the CGM-estimator uses a new, adaptive version of b in the computation of $w(\mathbf{h}_i)$. Specifically,

the new weight function is given by

$$w_{CGM}(\mathbf{h}_i) = \min \left(1, \frac{b_{CGM}}{PS_i^2} \right), \quad (6.7)$$

where b_{CGM} is defined similar to the new adaptive cut-off value presented in (4.13), and is given by

$$b_{CGM} = 2 \text{MAD}(\mathbf{ACPS}), \quad (6.8)$$

where \mathbf{ACPS} is the vector of ACPS values of the observations. This proposal of an adaptive cut-off value for the outlier detection in the weight computation was made in the same spirit as in the ACPS and robust pre-whitening contributions. However, since the CGM-estimator is a *suppression*-focused estimator, improvements are expected to be of a more continuous nature as outliers move away from the bulk of the observations. This is in contrast to the purely diagnostic contributions of Chapters 4 and 5 where the diagnosis of an observation, although adaptive, is either a hard rejection or a complete inclusion (will full weight) of the observation in question. One important note is that the simulations in this chapter will involve a reasonable level of data redundancy; generally greater than $5N$, due to the desire to examine the new estimator without introducing errors or effects that a low data redundancy may present. If the data redundancy for an application involving complex-valued data is low, the CGM-estimator may require an adjustment to its weighting scheme, as is the case for the standard GM-estimator (see (3.50)).

6.3 Regression Equalities for Complex Data

Consider the two-dimension complex-valued regression problem when two explanatory variables and a response variable are utilized. (6.2) can be expanded as

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e} \quad (6.9)$$

$$(6.10)$$

$$\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix} = \begin{bmatrix} \tilde{h}_{11} & \tilde{h}_{12} \\ \tilde{h}_{21} & \tilde{h}_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} + \begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{bmatrix}, \quad (6.11)$$

where $(\tilde{*})$ denotes a complex quantity, i.e., $\tilde{z}_1 = z_{1r} + jz_{1c}$. Solving this regression (ignoring the error term at this time) we obtain

$$\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \tilde{h}_{11} + \tilde{x}_2 \tilde{h}_{12} \\ \tilde{x}_1 \tilde{h}_{21} + \tilde{x}_2 \tilde{h}_{22} \end{bmatrix}. \quad (6.12)$$

Next, substitute the rectangular forms of the complex variables in (6.12) to get

$$\begin{bmatrix} z_{1r} + jz_{1c} \\ z_{2r} + jz_{2c} \end{bmatrix} = \begin{bmatrix} (x_{1r} + jx_{1c})(h_{11r} + jh_{11c}) & (x_{2r} + jx_{2c})(h_{12r} + jh_{12c}) \\ (x_{1r} + jx_{1c})(h_{21r} + jh_{21c}) & (x_{2r} + jx_{2c})(h_{22r} + jh_{22c}) \end{bmatrix}. \quad (6.13)$$

The goal is to solve for the real and imaginary components of \mathbf{z} , namely \tilde{z}_1 and \tilde{z}_2 which include the elements z_{1r} , z_{1c} , z_{2r} , and z_{2c} , in terms of the real and imaginary components of \mathbf{x} and \mathbf{H} . Toward this aim, we have

$$z_{1r} + jz_{1c} = (x_{1r} + jx_{1c})(h_{11r} + jh_{11c}) + (x_{2r} + jx_{2c})(h_{12r} + jh_{12c}) \quad (6.14)$$

$$= x_{1r}h_{11r} + jx_{1c}h_{11r} + jx_{1r}h_{11c} - x_{1c}h_{11c} + x_{2r}h_{12r} + jx_{2c}h_{12r} + jx_{2r}h_{12c} - x_{2c}h_{12c} \quad (6.15)$$

$$\Rightarrow z_{1r} = x_{1r}h_{11r} - x_{1c}h_{11c} + x_{2r}h_{12r} - x_{2c}h_{12c} \quad (6.16)$$

$$z_{1c} = x_{1c}h_{11r} + x_{1r}h_{11c} + x_{2c}h_{12r} + x_{2r}h_{12c} \quad (6.17)$$

and

$$z_{2r} + jz_{2c} = (x_{1r} + jx_{1c})(h_{21r} + jh_{21c}) + (x_{2r} + jx_{2c})(h_{22r} + jh_{22c}) \quad (6.18)$$

$$= x_{1r}h_{21r} + jx_{1c}h_{21r} + jx_{1r}h_{21c} - x_{1c}h_{21c} + x_{2r}h_{22r} + jx_{2c}h_{22r} + jx_{2r}h_{22c} - x_{2c}h_{22c} \quad (6.19)$$

$$\Rightarrow z_{2r} = x_{1r}h_{21r} - x_{1c}h_{21c} + x_{2r}h_{22r} - x_{2c}h_{22c} \quad (6.20)$$

$$z_{2c} = x_{1c}h_{21r} + x_{1r}h_{21c} + x_{2c}h_{22r} + x_{2r}h_{22c}. \quad (6.21)$$

With these identities for z_{1r} , z_{1c} , z_{2r} , and z_{2c} in hand, we now arrange the rectangular components of \mathbf{z} and \mathbf{x} according to (4.3), and then use the equalities in (6.16) through (6.21) to obtain an expanded regression equation, given by

$$\begin{bmatrix} z_{1r} \\ z_{2r} \\ z_{1c} \\ z_{2c} \end{bmatrix} = \left[\begin{array}{cc|cc} h_{11r} & h_{12r} & -h_{11c} & -h_{12c} \\ h_{21r} & h_{22r} & -h_{21c} & -h_{22c} \\ \hline h_{11c} & h_{12c} & h_{11r} & h_{12r} \\ h_{21c} & h_{22c} & h_{21r} & h_{22r} \end{array} \right] \begin{bmatrix} x_{1r} \\ x_{2r} \\ x_{1c} \\ x_{2c} \end{bmatrix}. \quad (6.22)$$

The result obtained in (6.22) shows a clear pattern to the arrangement of the real and imaginary components of the original \mathbf{h}_i vectors. This pattern allows us to simplify the result in (6.22). Assuming the intuitive extension to any number of dimensions, we propose

a transformation of the complex input data, given by

$$\tilde{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix} . \quad (6.23)$$

This transformation of the input data allows for ready use of complex-valued data in a GM-estimator, backed by knowledge of the GM-estimator's thoroughly studied IF, BP, ρ - and ψ -functions and related properties, and so forth. The resulting estimator is termed the Complex GM-estimator (CGM-estimator).

The CGM-estimator, as proposed and implemented here, offers the unique ability to individually weight the separate components of each complex number in the observation matrix. As will be demonstrated in the simulation that follows, the real and imaginary components of the residuals can provide separate information regarding the regression problem. The residuals are a function of the weighted observation matrix, and it follows that dealing with the individual real and imaginary elements can be beneficial to finding a robust regression solution in a more expedient manner than if the complex data is reduced to magnitude information and processed.

With the development of $\tilde{\mathbf{H}}$ completed, we may now exercise the resulting estimator and determine its utility.

6.4 Simulation

To determine the CGM-Estimator's viability and performance, two types of simulation will be performed. First, real-valued data from a classic data set will be sent into the CGM-estimator for residual analysis, and compared to that of standard least-squares (LS). Good

performance in this regard will act as a regression test to ensure the new formulation does not adversely impact performance on real-valued data sets.

Then, a set of STAP radar scenarios will be presented to the CGM-estimator for residual analysis as well as weight vector application to the observed radar returns. Performance in these exercises will demonstrate the utility of the CGM-estimator in contaminated multi-dimensional radar data.

6.4.1 Real Data

To check the validity of the new estimator input formulation proposed in (6.23), a familiar data set from [22] is used to exercise the estimator. The Coleman data set contains the verbal mean test score for all sixth-grade pupils for 20 New England and Mid-Atlantic region schools, along with several variables that may impact verbal test scores. These variables include staff salaries per pupil, percent of white-collar fathers, a socioeconomic deviation factor, the mean teacher verbal test score, and the mean educational level of the mothers of the pupils.

The LS and CGM methods are applied to the data set. The residuals of the real and imaginary components are plotted separately, yielding a record length of $2N = 40$ records (Real: 1-20; Imaginary: 21-40). The LS fit is shown with a solid line, and the CGM fit shown with a dashed line.

As evident in Figure 6.1, the CGM fit is only strong in magnitude at the outliers 3, 17, and 18. This is in line with the performance of the least-median of squares estimator shown in [22]. The LMS residual test is considered as the benchmark to which other estimators are compared from a statistical perspective.

The LS residuals, on the other hand, are known to point to school number 11 having an outlying response, in addition to school 3 and school 18. A robust fit using the relatively

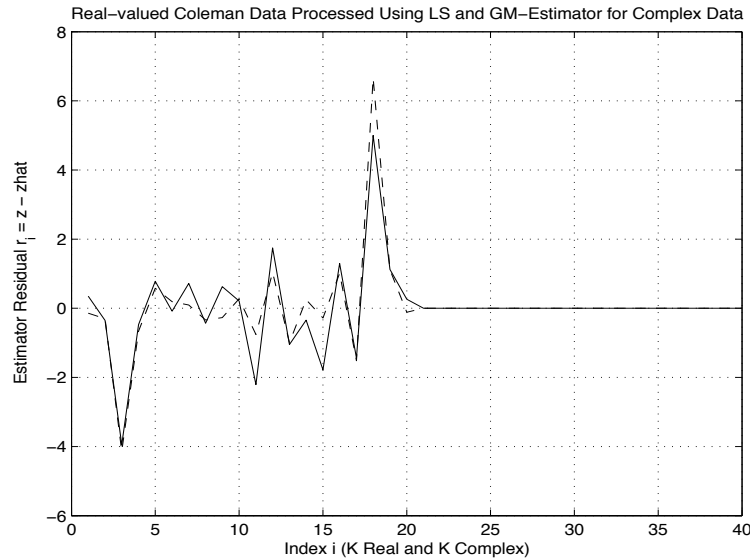


Figure 6.1: LS and CGM-estimator Residual Performance for Coleman Data

time-consuming LMS estimator shows that schools 3, 17, and 18 are the true outlying schools [22], and that school 11 is actually very in line with the bulk of the other schools. The CGM-estimator yields the same answer, but using a more computationally efficient algorithm than LMS. This combination of robustness and efficiency is the key advantage of the new approach.

The results of LS and CGM estimation are summarized in the following table, with outliers underlined. Note that in the new CGM-estimator, a duplicate set of residuals is automatically generated for the imaginary components of the observations. In the Coleman data case, no imaginary components exist for the input data, and thus the residuals were zero and are not shown for brevity.

The robust operation of the CGM-estimator on this classic data set indicates that the new estimator is capable of operating as a normal GM-estimator for real-valued data. It also suggests that the changes made to accommodate complex-valued data have not had an adverse impact on real-valued inputs.

Table 6.1: Residuals for Coleman Data (Real: 1-20; Imaginary: 21-40 not shown.)

<i>Sample</i>	<i>r_{LS}</i>	<i>r_{CGM}</i>	<i>r_{LMS}</i>
1	0.196	-0.143	-1.59
2	-0.197	-0.301	-0.01
3	-2.218	<u>-4.140</u>	<u>4.04</u>
4	-0.266	-0.660	-1.28
5	0.439	0.577	-0.01
6	-0.048	0.190	-0.01
7	0.404	0.094	0.15
8	-0.244	-0.340	0.41
9	0.351	-0.269	-0.43
10	0.118	0.272	2.00
11	-1.240	-0.760	0.11
12	0.981	1.035	1.43
13	-0.589	-0.995	-0.01
14	-0.194	0.248	0.43
15	-0.999	-0.281	-0.01
16	0.729	1.014	-0.29
17	-0.809	<u>-1.517</u>	<u>-2.82</u>
18	2.809	<u>6.661</u>	<u>5.76</u>
19	0.631	1.110	0.37
20	0.148	-0.117	-0.01

6.4.2 Complex Data

To exercise the CGM-estimator in a realistic adaptive processing scenario, a six-dimension complex-valued adaptive signal processing scenario is developed, similar to the target detection exercise in Chapter 4. Three 40 dB targets with respect to the noise floor are presented in cells 10, 13, and 16 out of 30 total cells. The contaminated data is processed by a LS Estimator and a CGM-estimator. The residuals of this processing are shown in Figure 6.2. The output residue $y_{out} = \mathbf{x}^H \mathbf{u}$, where \mathbf{x} is the estimated weight vector and \mathbf{u} is the cell

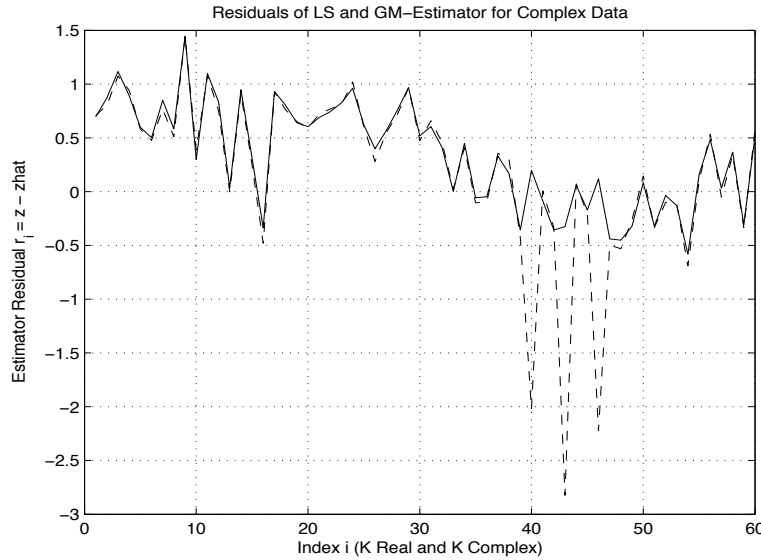


Figure 6.2: Complex GM Estimator Residuals (dashed) versus LS Residuals (solid) for Simulated Six-dimension Complex Data. Three Injected Targets (CUT's 10,13,16).

under test (CUT), is shown in Figure 6.3 for both LS (solid line) and the CGM-estimator. As is evident from the figure, the LS processor is unable to reliably detect the targets due to a weak output residual. This is due to the masking effect, and caused by multiple similar outliers (the three targets, in this case) exploiting the non-robust LS estimator. This becomes more clear in the residuals generated by the LS and CGM-estimator when computing the weight vector, shown in Table 6.2.

Table 6.2: Residuals for Input Data (Real: 1-30; Imaginary: 31-60)

<i>CUT</i> (\Re)	r_{LS}	r_{CGM}	<i>CUT</i> (\Im)	r_{LS}	r_{CGM}
1	0.695	0.709	31	0.604	0.660
2	0.870	0.801	32	0.411	0.459
3	1.117	1.074	33	0.002	0.022
4	0.893	0.943	34	0.450	0.409
5	0.598	0.583	35	-0.057	-0.104
6	0.502	0.477	36	-0.048	-0.093
7	0.850	0.767	37	0.330	0.356
8	0.584	0.511	38	0.166	0.295
9	1.444	1.443	39	-0.357	-0.429
10	0.300	0.387	40	0.198	<u>-2.024</u>
11	1.098	1.077	41	-0.085	0.007
12	0.833	0.752	42	-0.357	-0.336
13	0.036	-0.001	43	-0.326	<u>-2.824</u>
14	0.948	0.912	44	0.059	0.072
15	0.311	0.245	45	-0.171	-0.210
16	-0.337	-0.480	46	0.120	<u>-2.223</u>
17	0.931	0.927	47	-0.438	-0.498
18	0.802	0.760	48	-0.451	-0.531
19	0.641	0.652	49	-0.321	-0.264
20	0.604	0.593	50	0.083	0.146
21	0.689	0.721	51	-0.322	-0.334
22	0.744	0.769	52	-0.035	-0.099
23	0.826	0.811	53	-0.131	-0.117
24	0.961	1.020	54	-0.582	-0.691
25	0.619	0.599	55	0.161	0.104
26	0.399	0.280	56	0.485	0.534
27	0.566	0.548	57	0.032	-0.051
28	0.758	0.716	58	0.367	0.341
29	0.967	0.965	59	-0.309	-0.336
30	0.520	0.475	60	0.522	0.582

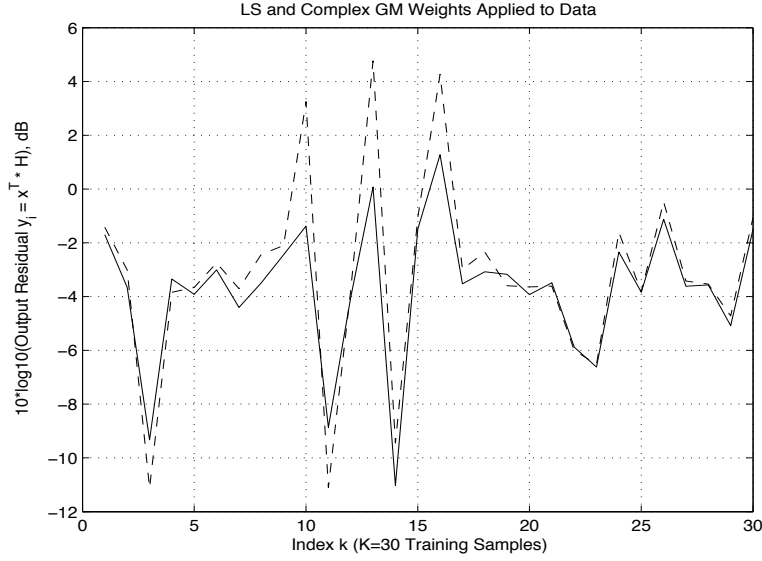


Figure 6.3: Complex GM Estimator Output Residuals versus LS Output Residuals for Simulated Six-dimension Complex Data. Three Injected Targets (CUT's 10,13,16).

6.5 Summary

In this chapter, a GM-estimator for complex-valued data has been developed for the first time. The processing of complex-valued data and the associated regression formulation has been described. Simulations using both real- and complex-valued data sets have been performed and performance was gauged. The performance of the new estimator was favorable in all trials when compared to classical estimators. Using the new estimator, adaptive signal processors utilizing complex-valued data may, for the first time, take advantage of the bounded total influence of the GM-estimator family.

Chapter 7

Conclusions and Future Work

This dissertation has proposed three main contributions to robust adaptive signal processing. In Chapter 4, the ACPS algorithm was developed and exercised. ACPS is the first projection statistics algorithm that is adaptive and handles complex-valued data. Its performance in the STAP radar application as a nonhomogeneity detector (NHD) was superior to other state-of-the-art NHD methods. As an NHD method, ACPS allowed for improved SINR convergence under contamination as well as made possible the detection of multiple targets in a contaminated STAP data set based on real MCARM data.

A robust data pre-whitening procedure was developed and exercised in Chapter 5. The procedure was tested against the identical STAP radar scenarios presented in Chapter 4. The robust pre-whitening procedure netted further improvements in SINR convergence compared to whitening of other state-of-the-art NHD methods. It does so, of course, at the expense of increased computation requirements inherent to any pre-whitening operation.

In Chapter 6, a GM-estimator for complex-valued data, the CGM-estimator, was proposed through the derivation of regression equalities and a rearrangement of the rectangular

forms of all quantities in the regression. The CGM-estimator is the first published GM-estimator that handles complex-valued data. The CGM-estimator performed well in the studied data scenarios, yielding a multipurpose estimator with inherent outlier suppression and accommodation qualities as well as a bounded total influence.

Many areas of research have been opened by this work. For instance, while the adaptive portion of the ACPS algorithm is shown to gracefully handle changes in contamination and dimensionality of the adaptive processor, the threshold and underlying estimators are by all means a first, albeit statistically-valid approach. Other estimators proposed in more recent literature may have computational advantages, robustness advantages, or a combination thereof, when compared to the *S*-Estimator and *shorth* used in the ACPS contribution.

In addition, the ACPS method, as presented, has several potential improvements – the most glaring to this researcher being the finite number of directions considered by the algorithm. With the ever-increasing computing resources available to modern adaptive signal processors, the investigation of more than the standard set of directions in PS will be beneficial. Gains in robustness and performance can be had via the investigation of twice the number of directions currently considered or possibly even more. Also, the constant false alarm rate (CFAR) properties of the ACPS and robust data pre-whitener methods are ripe for exploration. As the novel contributions in this dissertation are based on robust estimators, the CFAR properties thereof may be similar in derivation to those of the OS-CFAR variety. Several references are given in Section 4.4.3 from which an entire doctorate-level research effort could be initiated.

The joint consideration of robustness theory and adaptive signal processing has indeed yielded a set of novel contributions for robust adaptive signal processing. In addition, this work has presented several new research avenues in the fields of robustness and adaptive signal processing that are primed for exploration.

Appendix A

Complete Results for ACPS and Robust Data Pre-whitening SINR Convergence Simulations

The complete simulation results for ACPS SINR Convergence and Robust Pre-whitening SINR convergence are shown here. The results are separated by the dimensionality of the simulation starting at two complex dimensions and ending at 40 complex dimensions, with reasonable steps in between dimension choices. The convergence performance of each NHD method in contamination is presented in a convergence plot as well as a table. The contaminating outlier powers ranged from -10 dB to +50 dB in steps of 10 dB. Injected outliers are target-matched.

A.1 2 Dimensions

This case presented a simulation issue. Most values of κ_i are 3, except for in the 50 dB outlier case, in which ACPS and robust pre-whitening showed benefit compared to the other methods. 3 is the minimum possible value of κ_i for 2 dimensions, since the minimum K was set to 3 in this simulation, and the the RMB rule predicts convergence in $4 - 3 = 1$ snapshot, yielding $\kappa_i = 3/1 = 3.0$ Thus, the convergence is complete at the first simulation point for all methods yielding a value of 3 for κ_i . It is likely converge at even fewer training samples. However, this simulation was not performed.

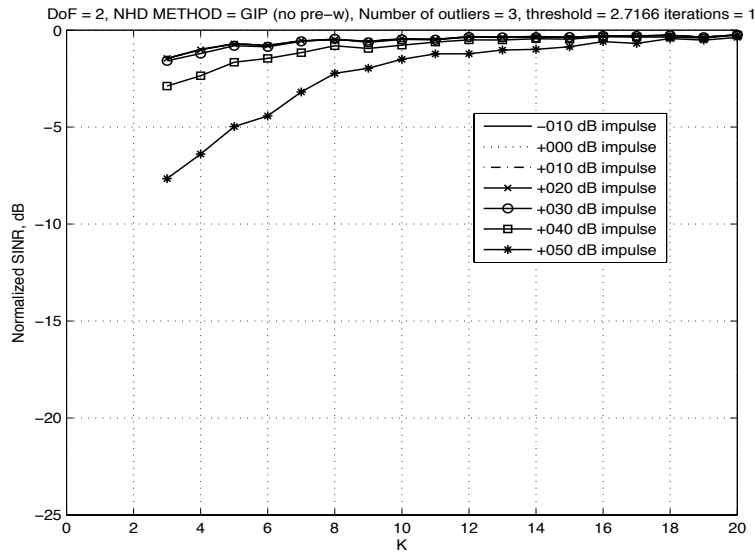


Figure A.1: 2 Dimension GIP Results, 1 Iteration, 3 Outliers

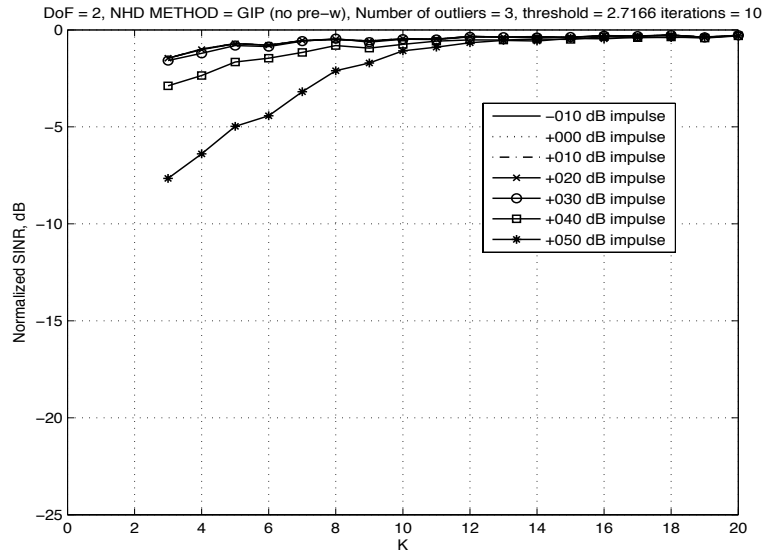


Figure A.2: 2 Dimension R-GIP Results, 10 Iterations, 3 Outliers

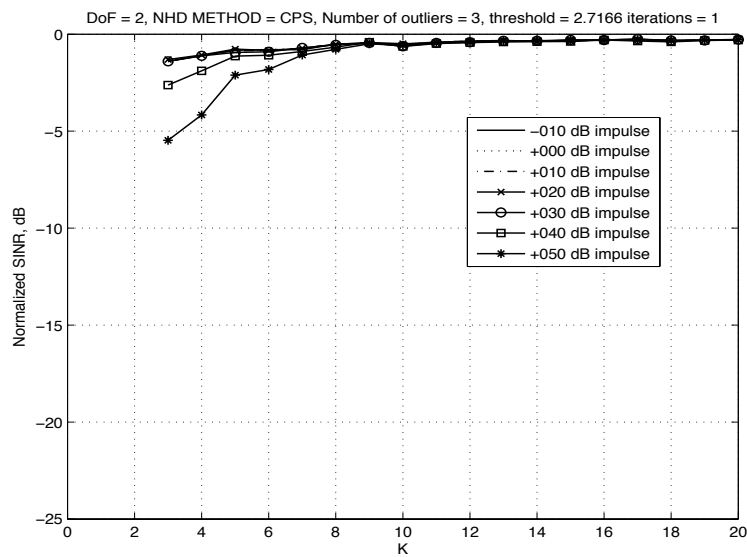


Figure A.3: 2 Dimension CPS Results, 3 Outliers

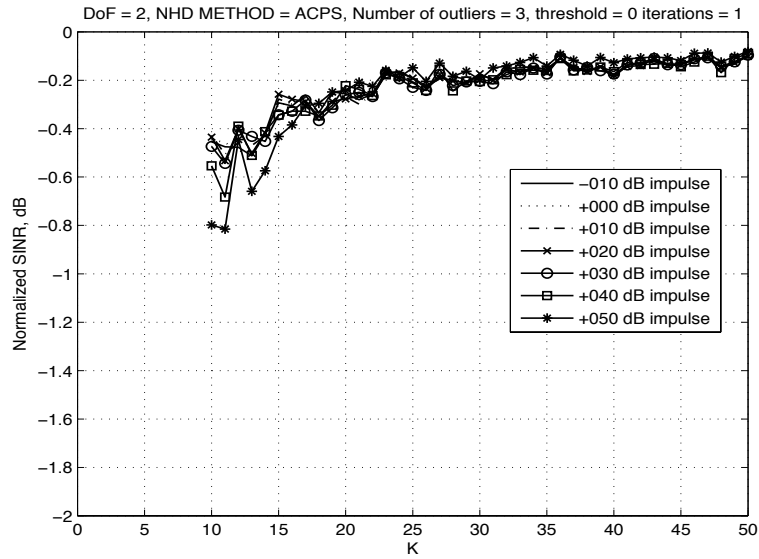


Figure A.4: 2 Dimension ACPS Results, 3 Outliers

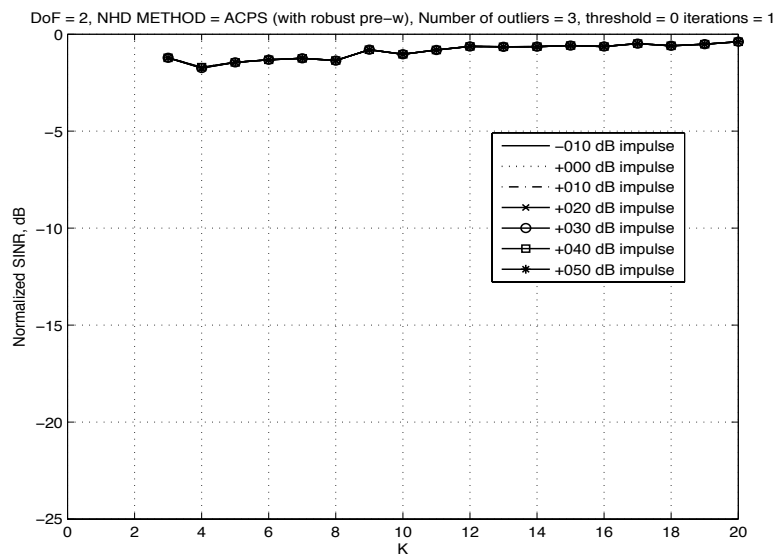


Figure A.5: 2 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.1: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.2: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.3: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.4: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.5: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.6: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.000	3.000	3.000	3.000	3.000
R-GIP	3.000	3.000	3.000	3.000	3.000
CPS	3.000	3.000	3.000	3.000	3.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

Table A.7: 2-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.000	3.000	3.000	5.000	8.000
R-GIP	3.000	3.000	3.000	5.000	8.000
CPS	3.000	3.000	3.000	4.000	5.000
ACPS	3.000	3.000	3.000	3.000	3.000
Robust PW	3.000	3.000	3.000	3.000	3.000

A.2 6 Dimensions

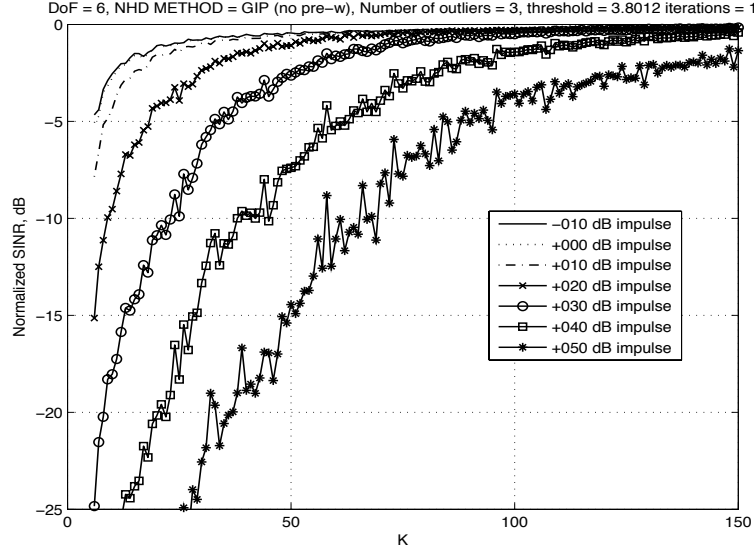


Figure A.6: 6 Dimension GIP Results, 1 Iteration, 3 Outliers

Table A.8: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.667	0.667	0.667	0.667	1.000
R-GIP	0.667	0.667	0.667	0.778	1.111
CPS	0.667	0.667	0.667	0.667	1.111
ACPS	0.667	0.667	0.667	0.778	1.111
Robust PW	0.667	0.667	0.667	1.111	1.778

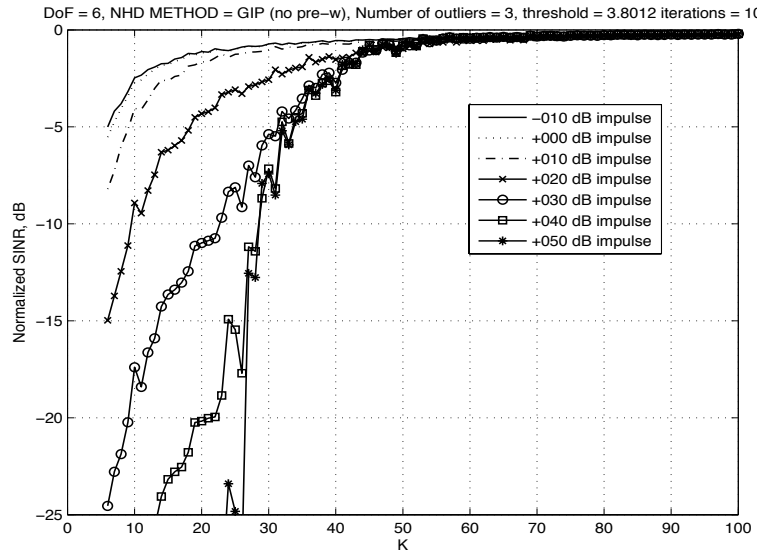


Figure A.7: 6 Dimension R-GIP Results, 10 Iterations, 3 Outliers

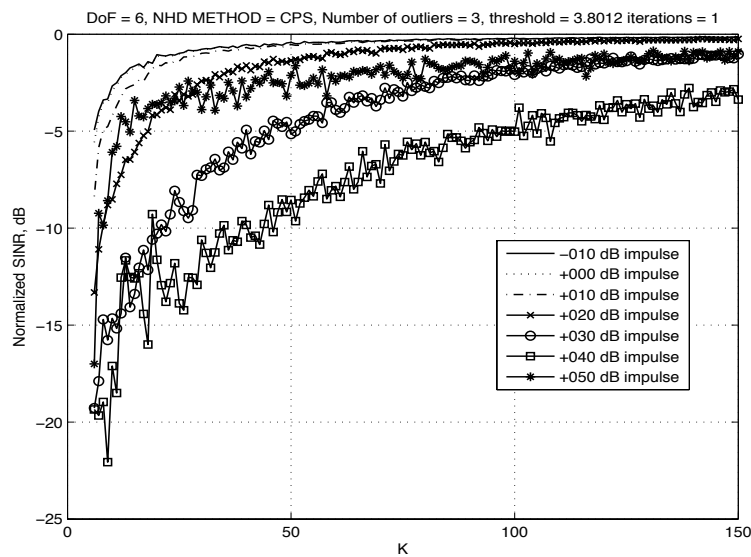


Figure A.8: 6 Dimension CPS Results, 3 Outliers

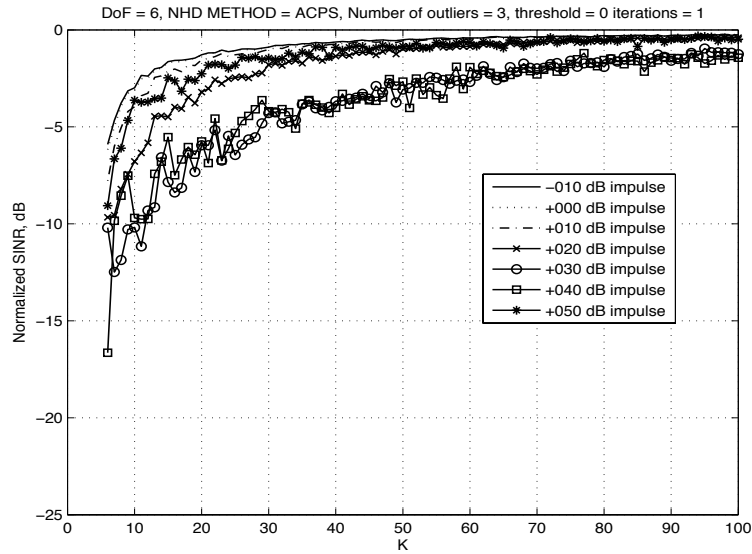


Figure A.9: 6 Dimension ACPS Results, 3 Outliers

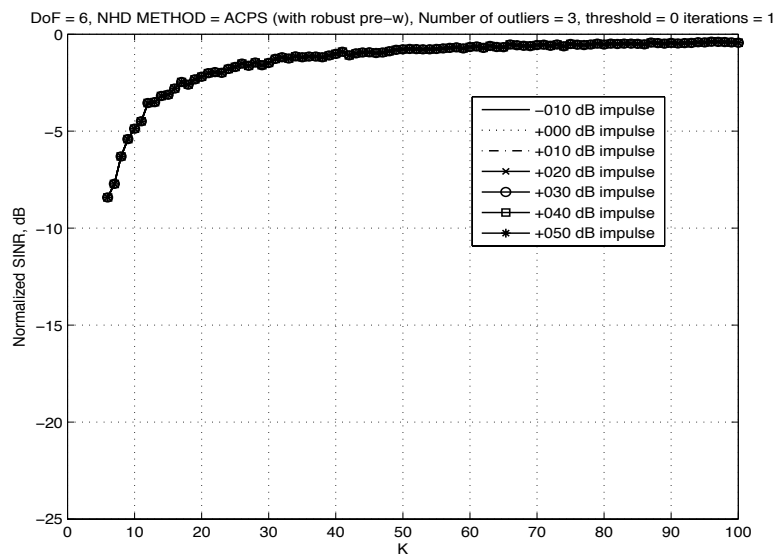


Figure A.10: 6 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.9: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.667	0.667	0.667	0.778	1.111
R-GIP	0.667	0.667	0.667	0.778	1.111
CPS	0.667	0.667	0.667	0.778	1.111
ACPS	0.667	0.667	0.667	0.778	1.222
Robust PW	0.667	0.667	0.667	1.111	1.778

Table A.10: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.667	0.667	0.667	1.000	1.444
R-GIP	0.667	0.667	0.667	1.111	1.556
CPS	0.667	0.667	0.667	0.889	1.444
ACPS	0.667	0.667	0.667	1.000	1.444
Robust PW	0.667	0.667	0.667	1.111	1.778

Table A.11: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.667	0.778	1.000	2.111	3.222
R-GIP	0.667	0.667	1.111	2.111	3.000
CPS	0.667	0.667	0.889	2.111	3.222
ACPS	0.667	0.667	0.667	1.444	2.444
Robust PW	0.667	0.667	0.667	1.111	1.778

Table A.12: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	1.000	1.444	2.667	3.667	4.889
R-GIP	1.111	1.556	2.556	3.556	4.000
CPS	0.667	0.889	2.333	4.444	7.889
ACPS	0.667	0.667	1.333	3.222	5.111
Robust PW	0.667	0.667	0.667	1.111	1.778

Table A.13: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	2.333	3.222	4.222	6.444	8.111
R-GIP	2.444	2.667	3.222	3.556	4.222
CPS	0.667	1.333	2.111	10.222	15.444
ACPS	0.667	0.778	0.778	2.444	5.333
Robust PW	0.667	0.667	0.667	1.111	1.778

Table A.14: 6-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.556	5.556	6.444	9.333	12.111
R-GIP	3.000	3.000	3.222	3.778	4.222
CPS	0.667	0.778	0.778	1.333	2.778
ACPS	0.667	0.667	0.667	1.000	1.667
Robust PW	0.667	0.667	0.667	1.111	1.778

A.3 10 Dimensions

Table A.15: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.588	0.588	0.588	0.765	1.118
R-GIP	0.588	0.588	0.588	0.765	1.118
CPS	0.588	0.588	0.588	0.765	1.118
ACPS	0.588	0.588	0.588	0.824	1.059
Robust PW	0.588	0.588	0.588	1.000	1.529

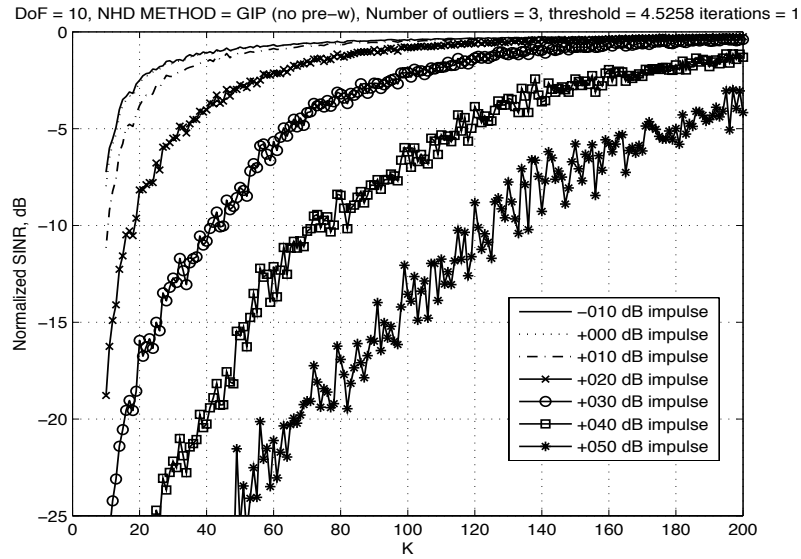


Figure A.11: 10 Dimension GIP Results, 1 Iteration, 3 Outliers

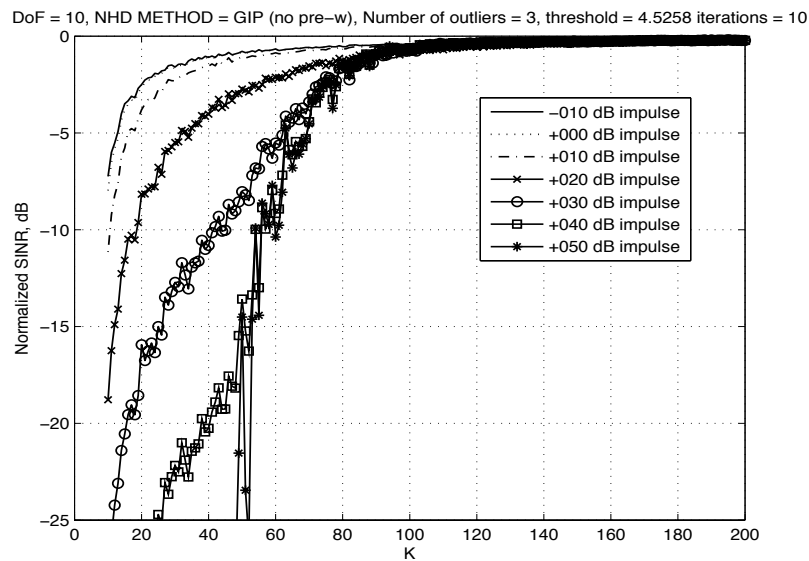


Figure A.12: 10 Dimension R-GIP Results, 10 Iterations, 3 Outliers

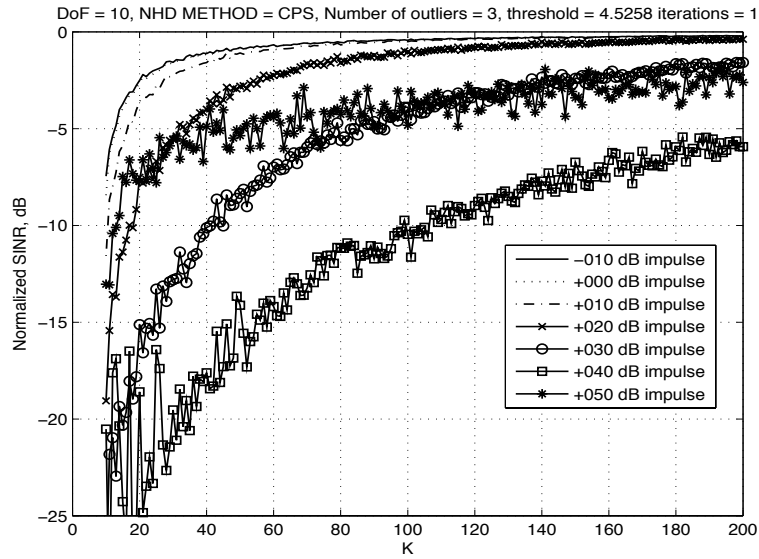


Figure A.13: 10 Dimension CPS Results, 3 Outliers

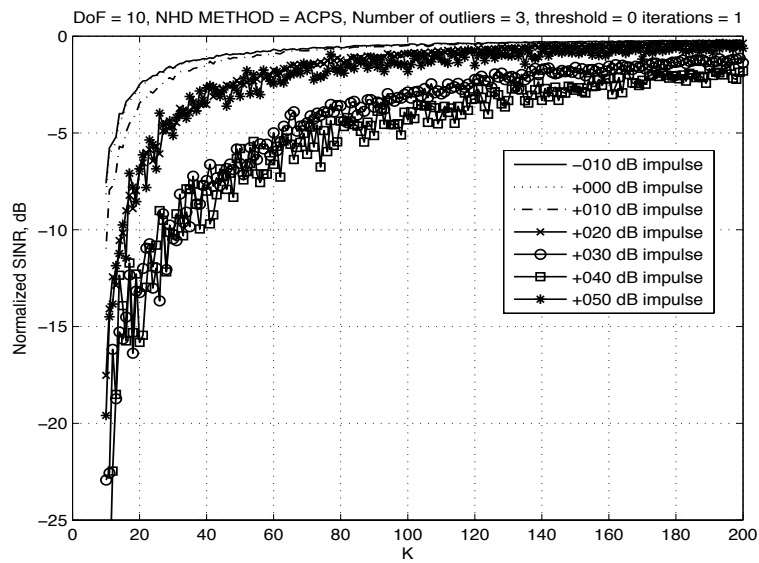


Figure A.14: 10 Dimension ACPS Results, 3 Outliers

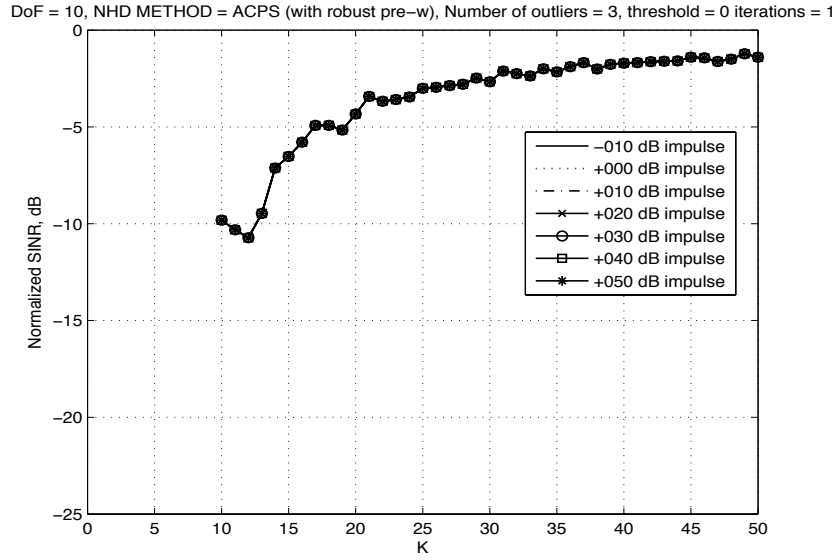


Figure A.15: 10 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.16: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.588	0.588	0.588	0.824	1.118
R-GIP	0.588	0.588	0.588	0.824	1.118
CPS	0.588	0.588	0.588	0.824	1.118
ACPS	0.588	0.588	0.588	0.824	1.118
Robust PW	0.588	0.588	0.588	1.000	1.529

Table A.17: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.588	0.588	0.647	0.941	1.471
R-GIP	0.588	0.588	0.647	0.941	1.471
CPS	0.588	0.588	0.647	1.000	1.471
ACPS	0.588	0.588	0.647	0.941	1.353
Robust PW	0.588	0.588	0.588	1.000	1.529

Table A.18: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.588	0.706	1.118	1.882	2.706
R-GIP	0.588	0.706	1.118	1.882	2.706
CPS	0.588	0.706	1.000	1.882	2.706
ACPS	0.588	0.647	0.941	1.588	2.412
Robust PW	0.588	0.588	0.588	1.000	1.529

Table A.19: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.941	1.588	2.471	3.706	5.059
R-GIP	0.941	1.588	2.471	3.706	4.235
CPS	0.824	1.471	2.471	4.647	7.529
ACPS	0.706	0.941	1.588	3.529	5.529
Robust PW	0.588	0.588	0.588	1.000	1.529

Table A.20: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	2.235	3.118	4.235	6.765	7.882
R-GIP	2.235	2.941	3.176	3.706	4.294
CPS	0.706	2.882	5.824	> 11.76	> 11.76
ACPS	0.765	0.824	1.529	4.176	6.882
Robust PW	0.588	0.588	0.588	1.000	1.529

Table A.21: 10-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.882	5.353	7.059	10.059	11.588
R-GIP	2.941	2.941	3.176	3.706	4.353
CPS	0.588	0.588	0.824	2.235	4.059
ACPS	0.588	0.647	0.882	1.529	2.235
Robust PW	0.588	0.588	0.588	1.000	1.529

A.4 20 Dimensions

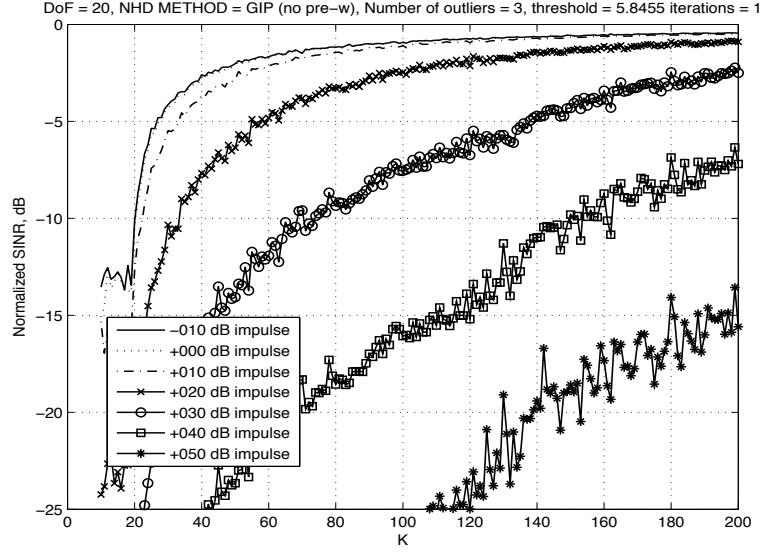


Figure A.16: 20 Dimension GIP Results, 1 Iteration, 3 Outliers

Table A.22: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.270	0.270	0.568	0.730	1.027
R-GIP	0.270	0.270	0.568	0.730	1.027
CPS	0.270	0.270	0.568	0.730	1.027
ACPS	0.270	0.270	0.568	0.757	1.054
Robust PW	0.270	0.270	0.595	0.784	1.108

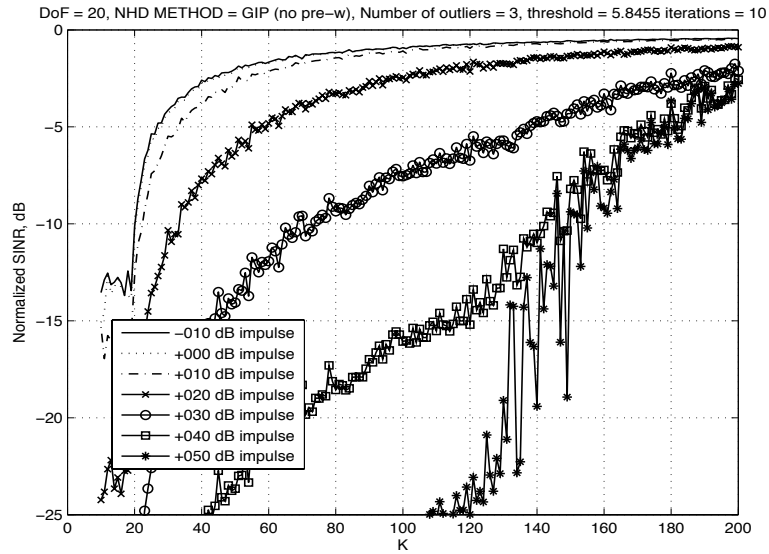


Figure A.17: 20 Dimension R-GIP Results, 10 Iterations, 3 Outliers

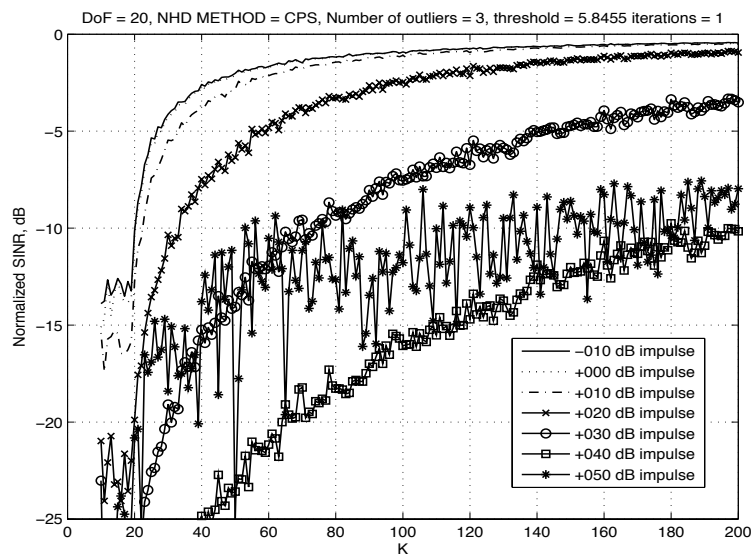


Figure A.18: 20 Dimension CPS Results, 3 Outliers

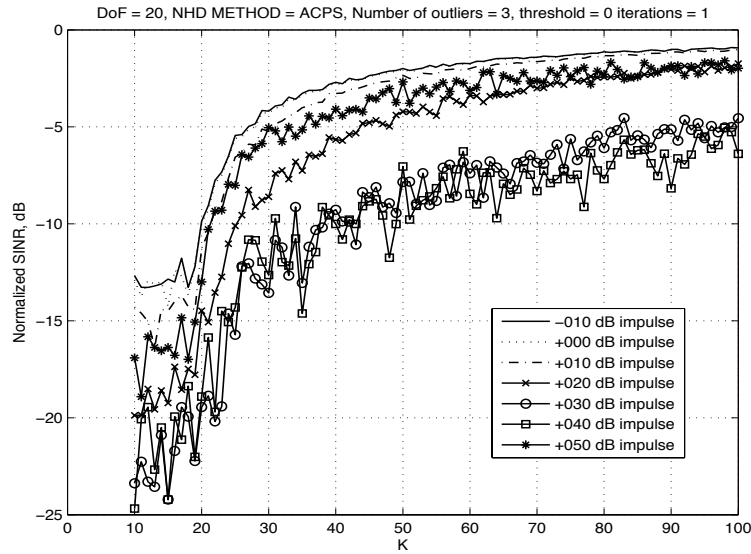


Figure A.19: 20 Dimension ACPS Results, 3 Outliers

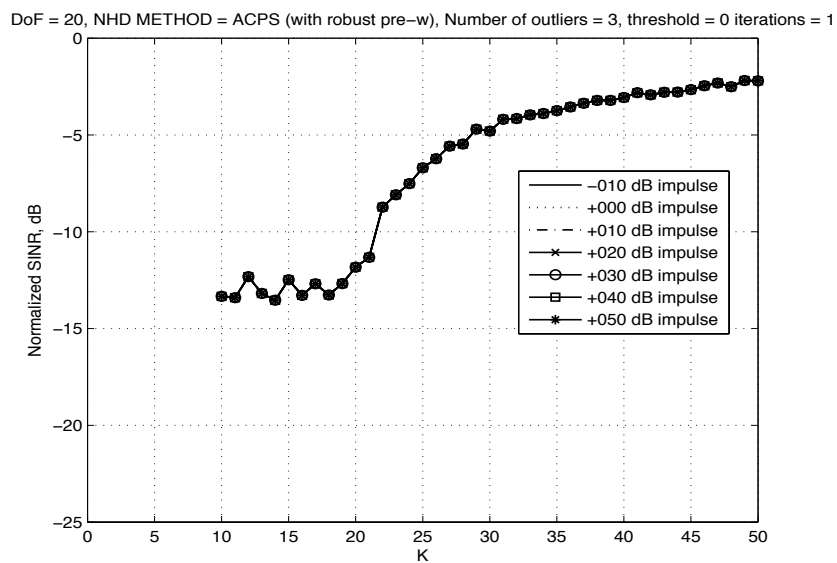


Figure A.20: 20 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.23: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.270	0.297	0.568	0.757	1.081
R-GIP	0.270	0.297	0.568	0.757	1.081
CPS	0.270	0.297	0.568	0.757	1.081
ACPS	0.270	0.270	0.568	0.757	1.054
Robust PW	0.270	0.270	0.595	0.784	1.108

Table A.24: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.270	0.541	0.622	0.919	1.297
R-GIP	0.270	0.541	0.622	0.919	1.297
CPS	0.270	0.541	0.622	0.919	1.297
ACPS	0.270	0.270	0.595	0.838	1.162
Robust PW	0.270	0.270	0.595	0.784	1.108

Table A.25: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.541	0.649	0.919	1.486	2.378
R-GIP	0.541	0.649	0.919	1.486	2.378
CPS	0.541	0.649	0.919	1.486	2.378
ACPS	0.270	0.568	0.703	1.189	1.811
Robust PW	0.270	0.270	0.595	0.784	1.108

Table A.26: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.811	1.189	1.865	3.730	4.459
R-GIP	0.811	1.189	1.865	3.730	4.459
CPS	0.811	1.189	1.865	3.757	> 5.405
ACPS	0.432	0.649	0.919	2.243	3.784
Robust PW	0.270	0.270	0.595	0.784	1.108

Table A.27: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	1.757	3.000	4.054	> 5.405	> 5.405
R-GIP	1.757	3.000	3.865	4.622	5.081
CPS	1.757	3.000	4.865	> 5.405	> 5.405
ACPS	0.270	0.595	0.811	2.486	> 5.405
Robust PW	0.270	0.270	0.595	0.784	1.108

Table A.28: 20-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	3.514	4.865	> 5.405	> 5.405	> 5.405
R-GIP	3.514	3.568	3.946	4.703	5.135
CPS	0.622	0.703	1.432	> 5.405	> 5.405
ACPS	0.270	0.459	0.595	0.919	1.378
Robust PW	0.270	0.270	0.595	0.784	1.108

A.5 30 Dimensions

Table A.29: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.175	0.193	0.561	0.737	1.000
R-GIP	0.175	0.193	0.561	0.737	1.000
CPS	0.175	0.193	0.561	0.737	1.000
ACPS	0.175	0.193	0.561	0.754	1.018
Robust PW	0.175	0.228	0.579	0.789	1.070

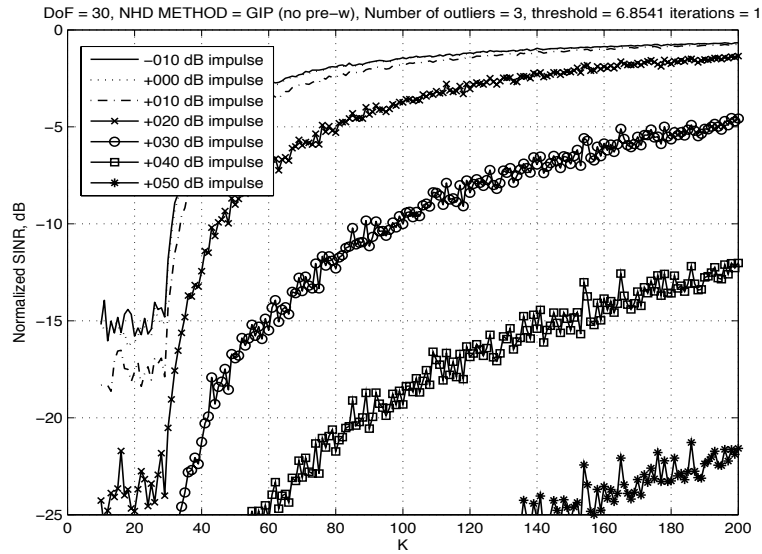


Figure A.21: 30 Dimension GIP Results, 1 Iteration, 3 Outliers

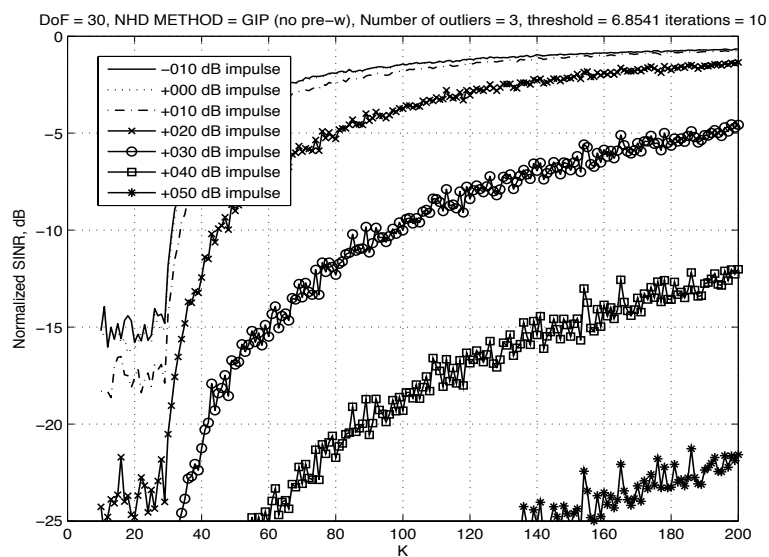


Figure A.22: 30 Dimension R-GIP Results, 10 Iterations, 3 Outliers

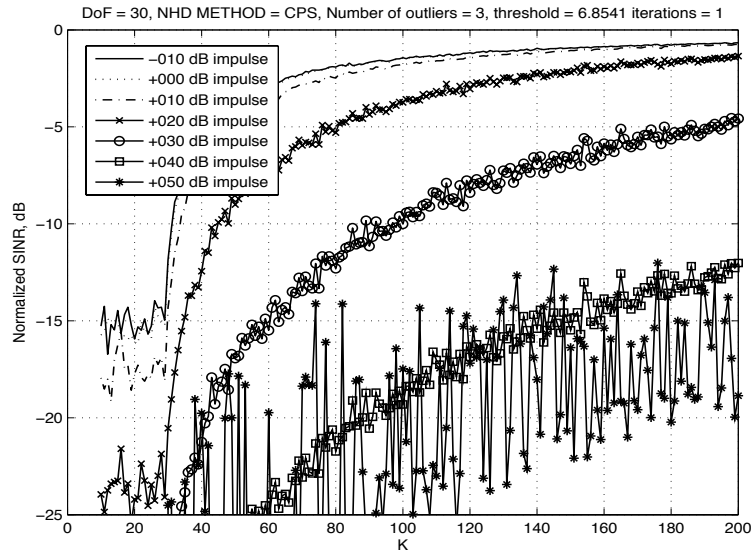


Figure A.23: 30 Dimension CPS Results, 3 Outliers

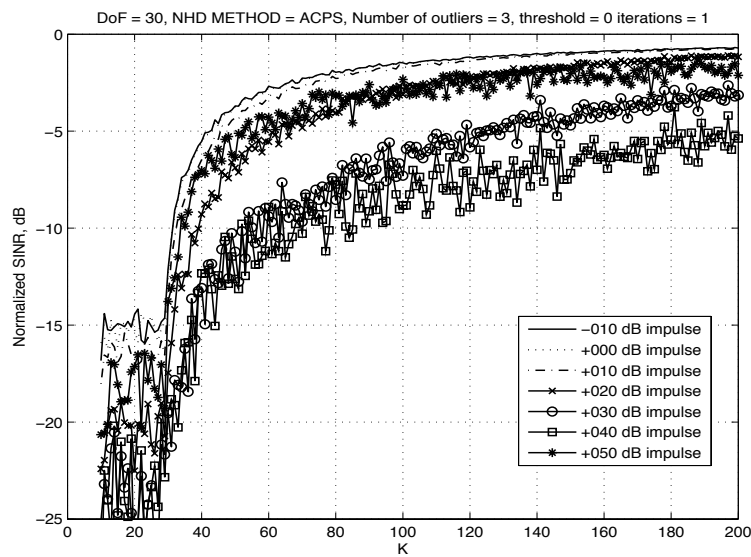


Figure A.24: 30 Dimension ACPS Results, 3 Outliers

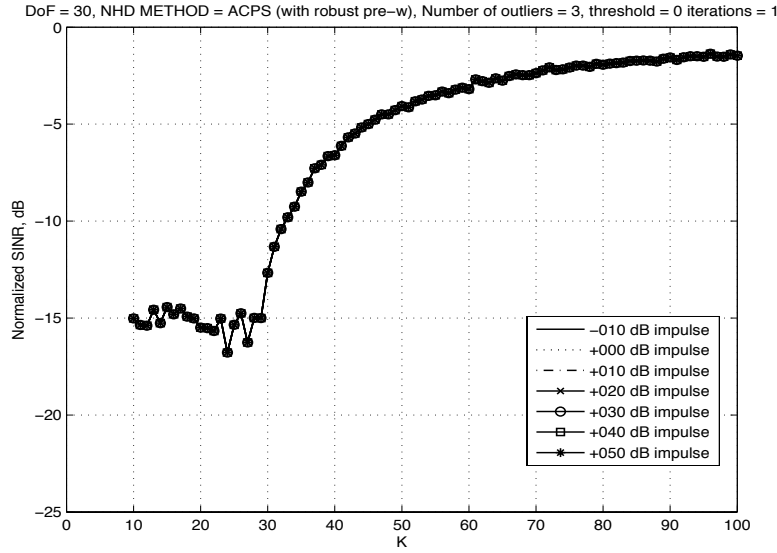


Figure A.25: 30 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.30: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.175	0.456	0.561	0.754	1.035
R-GIP	0.175	0.456	0.561	0.754	1.035
CPS	0.175	0.211	0.561	0.754	1.035
ACPS	0.175	0.228	0.561	0.772	1.053
Robust PW	0.175	0.228	0.579	0.789	1.070

Table A.31: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.175	0.526	0.596	0.860	1.175
R-GIP	0.175	0.526	0.596	0.860	1.175
CPS	0.175	0.526	0.596	0.860	1.175
ACPS	0.175	0.526	0.579	0.807	1.123
Robust PW	0.175	0.228	0.579	0.789	1.070

Table A.32: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.544	0.614	0.789	1.333	1.982
R-GIP	0.544	0.614	0.789	1.333	1.982
CPS	0.544	0.614	0.789	1.333	1.982
ACPS	0.228	0.561	0.702	1.123	1.632
Robust PW	0.175	0.228	0.579	0.789	1.070

Table A.33: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.737	1.035	1.561	3.123	> 6.667
R-GIP	0.737	1.035	1.561	3.123	> 6.667
CPS	0.737	1.035	1.561	3.123	> 6.667
ACPS	0.526	0.649	0.947	2.088	3.123
Robust PW	0.175	0.228	0.579	0.789	1.070

Table A.34: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	1.491	2.386	> 6.667	> 6.667	> 6.667
R-GIP	1.491	2.386	> 6.667	> 6.667	> 6.667
CPS	1.491	2.386	> 6.667	> 6.667	> 6.667
ACPS	0.491	0.649	0.912	2.474	> 6.667
Robust PW	0.175	0.228	0.579	0.789	1.070

Table A.35: 30-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	> 6.667	> 6.667	> 6.667	> 6.667	> 6.667
R-GIP	> 6.667	> 6.667	> 6.667	> 6.667	> 6.667
CPS	0.667	1.298	> 6.667	> 6.667	> 6.667
ACPS	0.228	0.526	0.596	0.930	1.316
Robust PW	0.175	0.228	0.579	0.789	1.070

A.6 40 Dimensions

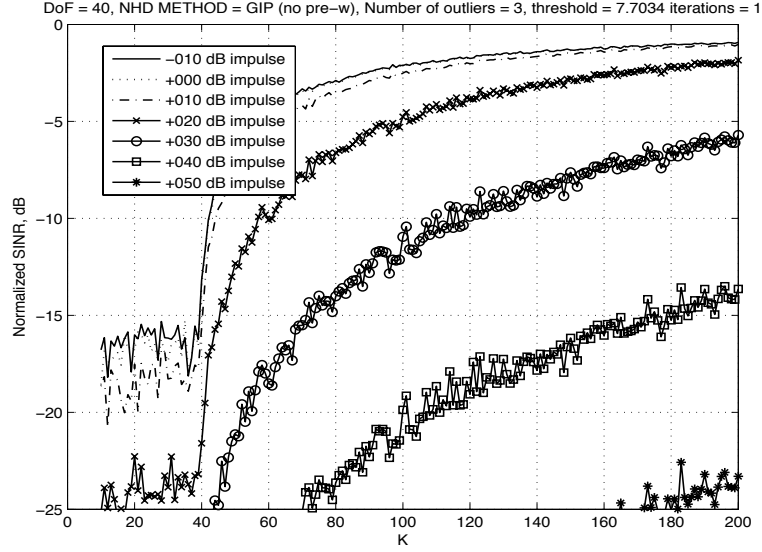


Figure A.26: 40 Dimension GIP Results, 1 Iteration, 3 Outliers

Table A.36: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; -10 dB Outlier Level

Method	-20 dB	-15 dB	-10 dB	-5 dB	-3 dB
GIP	0.130	0.519	0.558	0.727	1.013
R-GIP	0.130	0.519	0.558	0.727	1.013
CPS	0.130	0.519	0.558	0.727	1.013
ACPS	0.130	0.519	0.558	0.740	1.013
Robust PW	0.130	0.519	0.571	0.792	1.039

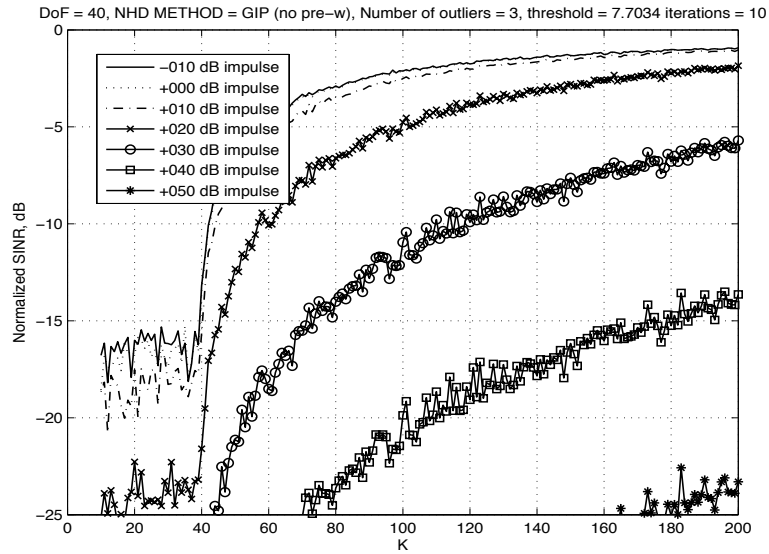


Figure A.27: 40 Dimension R-GIP Results, 10 Iterations, 3 Outliers

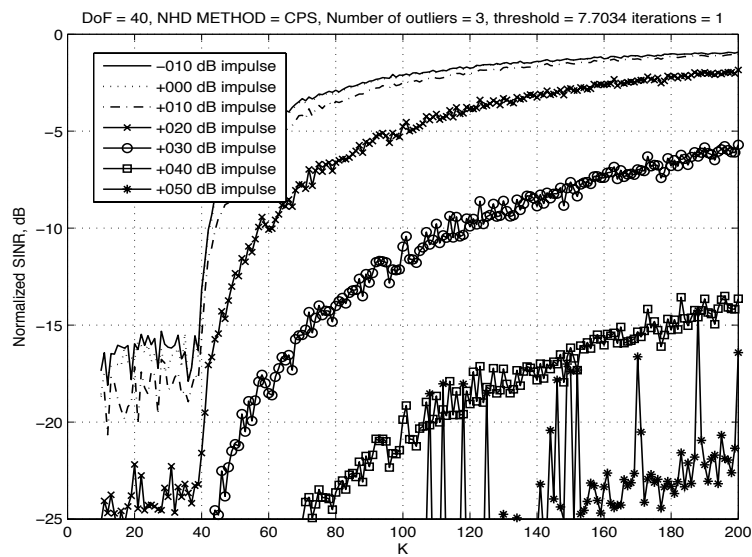


Figure A.28: 40 Dimension CPS Results, 3 Outliers

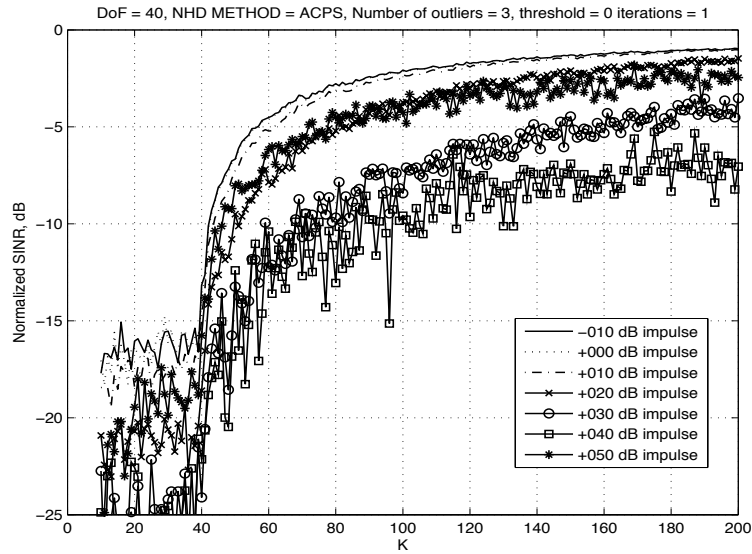


Figure A.29: 40 Dimension ACPS Results, 3 Outliers

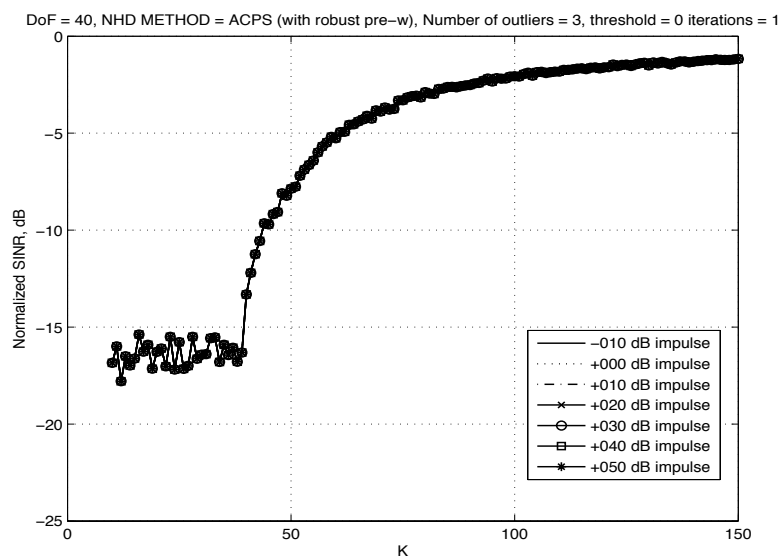


Figure A.30: 40 Dimension Robust Pre-whitening Using ACPS Results, 3 Outliers

Table A.37: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 0 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.130	0.519	0.558	0.740	1.039
R-GIP	0.130	0.519	0.558	0.740	1.039
CPS	0.130	0.519	0.558	0.740	1.039
ACPS	0.130	0.377	0.558	0.753	1.013
Robust PW	0.130	0.519	0.571	0.792	1.039

Table A.38: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 10 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.130	0.532	0.584	0.831	1.156
R-GIP	0.130	0.532	0.584	0.831	1.156
CPS	0.130	0.532	0.584	0.831	1.156
ACPS	0.130	0.519	0.571	0.805	1.117
Robust PW	0.130	0.519	0.571	0.792	1.039

Table A.39: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 20 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.532	0.597	0.740	1.299	1.857
R-GIP	0.532	0.597	0.740	1.299	1.857
CPS	0.532	0.597	0.740	1.299	1.857
ACPS	0.455	0.545	0.649	1.052	1.506
Robust PW	0.130	0.519	0.571	0.792	1.039

Table A.40: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 30 dB Outlier Level

Method	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	0.675	0.935	1.429	> 5.000	> 5.000
R-GIP	0.675	0.935	1.429	> 5.000	> 5.000
CPS	0.675	0.935	1.429	> 5.000	> 5.000
ACPS	0.545	0.597	0.766	1.805	> 5.000
Robust PW	0.130	0.519	0.571	0.792	1.039

Table A.41: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 40 dB Outlier Level

<i>Method</i>	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	1.299	2.247	> 5.000	> 5.000	> 5.000
R-GIP	1.299	2.247	> 5.000	> 5.000	> 5.000
CPS	1.299	2.247	> 5.000	> 5.000	> 5.000
ACPS	0.545	0.649	0.883	> 5.000	> 5.000
Robust PW	0.130	0.519	0.571	0.792	1.039

Table A.42: 40-Dimension SINR Convergence Performance by Outlier Diagnostic Method, Metric of κ_i ; 50 dB Outlier Level

<i>Method</i>	−20 dB	−15 dB	−10 dB	−5 dB	−3 dB
GIP	> 5.000	> 5.000	> 5.000	> 5.000	> 5.000
R-GIP	> 5.000	> 5.000	> 5.000	> 5.000	> 5.000
CPS	1.403	2.442	> 5.000	> 5.000	> 5.000
ACPS	0.260	0.532	0.610	0.935	1.545
Robust PW	0.130	0.519	0.571	0.792	1.039

Bibliography

- [1] Picciolo M.L., *Robust Adaptive Processors*, Ph.D. thesis, Virginia Tech, May 2003.
- [2] Stahel W. A., *Robuste Schatzungen: Infinitesimale Optimalitat und Schatzungen von Kovarianzmatrizen*, Ph.D. thesis, E.T.H. Zurich, Switzerland, 1981.
- [3] D.L. Donoho , Qualifying Paper, 1982.
- [4] Gasko M. and Donoho D., “Influential observation in data analysis,” American Statistical Association, 1982, Proc. of the Business and Economic Statistics Section, pp. 104–110.
- [5] Donoho D. and Gasko M., “Breakdown properties of location estimates based on halfspace depth and projected outlyingness,” *The Annals of Statistics*, vol. 20, no. 4, pp. 1803–1827, 1992.
- [6] Rousseeuw P.J. and Van Zomeren B.C. , ”Robust Distances: Simulations and Cutoff Values”, in *Directions in Robust Statistics and Diagnostics, Part II*, edited by W. Stahel and S. Weisburg, 1991.
- [7] Rousseeuw P.J. and Van Zomeren B.C., “Unmasking multivariate outliers and leverage points,” *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 633–651, 1990.

- [8] Mili L., Cheniae M., Vichare N., and Rousseeuw P., “Robust state estimation based on projection statistics,” *IEEE Trans. on Power Syst.*, vol. 11, no. 2, pp. 1118–1127, May 1996.
- [9] Mili L. and Coakley C.W., “Robust estimation in structured linear regression,” *The Annals of Statistics*, vol. 24, no. 6, pp. 2593–2607, 1996.
- [10] Gerlach K.R., “Outlier resistant adaptive matched filtering,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 885–901, 2002.
- [11] Wang Y.-L., Chen J.-W., Bao Z., and Peng Y.-N., “Robust space-time adaptive processing for airborne radar in nonhomogeneous clutter environments,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 70–81, 2003.
- [12] Michels J.H., Himed B., and Rangaswamy M., “Robust stap detection in a dense signal airborne radar environment,” *Signal Processing*, vol. 84, no. 9, pp. 1625–1636, 2004.
- [13] Roman J., Rangaswamy M., Davis D., Zhang Q., Himed B., and Michels J., “Parametric adaptive matched filter for airborne radar applications,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 677–692, 2000.
- [14] Blunt S., Gerlach K., and Rangaswamy M., “The enhanced fracta algorithm with knowledge aided covariance estimation,” Barcelona, Spain, 2004, Third IEEE Workshop on Sensor Array Multichannel Processing.
- [15] Chen P., Melvin W.L., and Wicks M.C., “Screening among multivariate normal data,” *Journal of Multivariate Analysis*, vol. 69, no. 1, pp. 10–29, 1999.
- [16] Rangaswamy M., Himed B., and Michels J.H., “Statistical analysis of the nonhomogeneity detector,” 2000, vol. 2 of *Proc. Asilomar Conference on Signals, Systems and Computers*, pp. 1117–1121.

- [17] Rangaswamy M., Lin F., and Gerlach K., “Robust adaptive signal processing methods for heterogeneous radar clutter scenarios,” *Signal Processing*, vol. 84, no. 9, pp. 1653–1665, 2004.
- [18] Rangaswamy M., “Statistical analysis of the non-homogeneity detector for non-gaussian interference backgrounds,” *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2101–2111, 2005.
- [19] Rangaswamy M., Michels J.H., and Himed B., “Statistical analysis of the nonhomogeneity detector for STAP applications,” *Digital Signal Processing*, vol. 14, no. 3, pp. 253–267, 2004.
- [20] Melvin W.L., Wicks M.C., and Chen P. United States Patent 5,706,013, 1999.
- [21] Hampel F.R., Ronchetti E.M., Rousseeuw P.J., and Stahel W.A., *Robust Statistics: The Approach based on Influence Functions*, Wiley, New York, 1986.
- [22] Rousseeuw P.J. and Leroy A.M., *Robust Regression and Outlier Detection*, John Wiley and Sons, New York, 1987.
- [23] Schoenig G.N., Picciolo M.L., and Mili L., “Improved detection of strong nonhomogeneities for STAP via projection statistics,” Crystal City, VA, 2005, IEEE Intl. Radar Conference.
- [24] Schoenig G.N., Picciolo M.L., Gerlach K., and Mili L., “Adaptive processor convergence improvement using reiterative projection statistics,” Verona, NY, 2006, IEEE Intl. Radar Conference.
- [25] Kleiner B. and Martin R.D., “Robust estimation of power spectra,” *Journal of the Royal Statistical Society*, vol. 41, no. 3, pp. 313–351, 1973.

- [26] Efron A.J., Swaszek P.F., and Tufts D.W., “Insight into detection of deterministic and gaussian signals in correlated plus impulsive noise environments,” *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 603–611, 1991.
- [27] Efron A.J., Swaszek P.F., and Tufts D.W., “Detection in correlated gaussian plus impulsive noise,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 4, pp. 932–943, 1992.
- [28] Efron A.J. and Jeon H., “Pre-whitening for detection in correlated plus impulsive noise,” 1992, vol. II of *Proceedings of IEEE ICASSP*, pp. 469–472.
- [29] Efron A.J. and Jeon H., “Detection in impulsive noise based on robust whitening,” *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1572–1576, 1994.
- [30] Mollah N.H., Eguchi S., and Minami M. Draft Chapter, Kluwer Academic Publishers, Netherlands, 2006.
- [31] Mallows C.L. Unpublished Memorandum, Bell Telephone Laboratories, Murray Hill, NJ, 1975.
- [32] Schweppe F. C., Wildes J., and Rom D.B., “Power system staticstate estimation,” *IEEE Trans on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–135, 1970.
- [33] Picciolo M.L., Schoenig G.N., Gerlach K., and Mili L., “Robust cascaded canceller using projection statistics for adaptive radar,” Big Sky, MT, 2005, IEEE Aerospace Conference.
- [34] des Rosiers A., Schoenig G.N., and Mili L., “Robust space-time adaptive processing using projection statistics,” Toulouse, France, 2004, IEEE Intl. Radar Conference.
- [35] Ollila E. and Koivunen V., “Influence functions for array covariance matrix estimators,” 2003, Statistical Signal Processing, IEEE Workshop on, pp. 462–465.

- [36] Ollila E. and Koivunen V., “Robust antenna array processing using m-estimators of pseudo-covariance,” Beijing, China, 2003, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.
- [37] Hayes M. H., *Statistical Digital Signal Processing and Modeling*, John Wiley and Sons, New York, 1996.
- [38] Haykin S., *Adaptive Filter Theory*, Prentice Hall, & Sons, Upper Saddle River, NJ, 1996.
- [39] Papoulis A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.
- [40] Fisher R.A., “Theory of statistical estimation,” *Proc. Cambridge Philos. Soc.*, vol. 22, 1925.
- [41] Wiener N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley and Sons, New York, 1949.
- [42] Van Trees H. L., *Detection, Estimation, and Modulation Theory*, vol. 1, John Wiley & Sons, New York, NY, 1968.
- [43] Scharf L.L., *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, USA, 1991.
- [44] Goldstein J.S., *Optimal Reduced-Rank Statistical Signal Processing, Detection and Estimation Theory*, Ph.D. thesis, University of Southern California, December 1997.
- [45] Poor H.V., “On robust wiener filtering,” *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 531–536, June 1980.

- [46] Poor H.V., “Signal detection in the presence of weakly dependent noise – part ii: Robust detection,” *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 744–752, September 1982.
- [47] Moustakides G. and Kassam S.A., “Robust wiener filters for random signals in correlated noise,” *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 614–619, July 1983.
- [48] Chen C. and Kassam S.A., “Robust wiener filtering for multiple inputs with channel distortion,” *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 674–677, July 1984.
- [49] Widrow B. and McCool J., “A comparison of adaptive algorithms based on the methods of steepest descent and random search,” *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 5, pp. 615–637, September 1976.
- [50] Kailath T., Sayed A.H., and Hassibi B., *Linear Estimation*, Prentice Hall, Upper Saddle River, NJ, 2000.
- [51] Moon T.K. and Stirling W.C., *Mathematical Methods and Algorithms*, Prentice Hall, Upper Saddle River, N.J., 2000.
- [52] Goldstein J.S. , Private communication, 2004.
- [53] Brennan L.E. and Reed I.S., “Theory of adaptive radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-9, pp. 237–252, 1973.
- [54] Reed I.S., Mallett J.D., and Brennan L.E., “Rapid convergence rate in adaptive arrays,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, no. 6, pp. 853–863, November 1974.

- [55] Ward J., “Space-time adaptive processing for airborne radar,” Technical Report TR-1015, Lincoln Laboratory, MIT, Lexington, MA, December 1994.
- [56] Rangaswamy M., “An overview of space-time adaptive processing for radar,” Huntsville, AL, USA, 2003, IEEE Intl. Radar Conference.
- [57] Wicks M. C., Rangaswamy M., Adve R., and Hale T. B., “A step overview,” *Space-Time Adaptive Processing: A Knowledge-based Perspective for Airborne Radar*, pp. 51–65, January 2006.
- [58] Klemm R., *Principles of Space-Time Adaptive Processing*, IEE Press, Bodmin, UK, 2002.
- [59] Melvin W. L., “A step overview,” *IEEE A&E Systems Magazine*, vol. 19, no. 1, pp. 19–35, January 2004.
- [60] Schoenig G.N. , Ph.D. Preliminary Examination, 2005.
- [61] Ogle W.C., Witzgall H.E., Tinston M.A., Goldstein J.S., and P.A. Zulch, “Independent sample mean squared error for adaptive detection statistics,” Big Sky, MT, 2005, IEEE Aerospace Conference.
- [62] Picciolo M.L. and Gerlach K., “A median cascaded canceller for robust adaptive array processing,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 3, pp. 883–900, July 2003.
- [63] Kay S., *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, 1998.
- [64] Mahalanobis P.C., “On the generalised distance in statistics,” *Proceedings of the National Institute of Science of India*, vol. 12, pp. 49–55, 1936.

- [65] Melvin W.L. and Wicks M.C., “Improving practical space-time adaptive radar,” Syracuse, NY, USA, 1997, Proceedings of the IEEE Radar Conference, pp. 48–53.
- [66] Guerci J.R., *Space-Time Adaptive Processing for Radar*, Artech House, Boston, 2003.
- [67] Teixeira C. M., Bergin J. S., and Techau P. M., “Adaptive thresholding of the gip statistic to remove ground target returns from the training data for STAP applications,” Lexington, MA, USA, 2004, MIT Lincoln Laboratory, 12th Adaptive Sensor Array Processing Workshop.
- [68] Morgan D.R., “Partially adaptive array techniques,” *IEEE Transactions on Antennas and Propagation*, vol. 26, no. 6, pp. 823–833, November 1978.
- [69] Van Veen B.D. and Roberts R.A., “Partially adaptive beamformer design via output power minimization,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 11, pp. 1524–1532, November 1987.
- [70] Van Veen B.D., “Eigenstructure-based partially adaptive array design,” *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 3, pp. 357–362, March 1988.
- [71] Goldstein J.S. and Reed I.S., “Theory of partially adaptive radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1309–1325, October 1997.
- [72] Goldstein J.S. and Reed I.S., “Reduced rank adaptive filtering,” *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 492–496, February 1997.
- [73] Goldstein J.S. and Reed I.S., “Subspace selection for partially adaptive sensor array processing,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 539–544, April 1997.

- [74] Goldstein J.S., Ingram M.A., Holder E.J., and Smith R.N., “Adaptive subspace selection using subband decompositions for sensor array processing,” 1994, vol. 4 of *Proceedings of IEEE ICASSP*, pp. 281–284.
- [75] Goldstein J.S., Kogon S.M., Reed I.S., Williams D.B., and Holder E.J., “Partially adaptive radar signal processing: The cross-spectral approach,” 1995, vol. 2 of *Asilomar Conference on Signals, Systems, and Computers*, pp. 1383–1387.
- [76] Goldstein J.S., Reed I.S., and Tague J.A., “Reduced complexity, robust, CFAR detectors for large sensor arrays,” 1996, vol. 2 of *Asilomar Conference on Signals, Systems, and Computers*, pp. 1268–1272.
- [77] Tinston M.A., Ogle W.C., Picciolo M.L., Goldstein J.S., Zulch P.A., and Wicks M.C., “Classification of training data with reduced-rank generalized inner product,” Philadelphia, PA, USA, 2004, Proceedings of the IEEE Radar Conference.
- [78] Huber P.J., *Robust Statistics*, John Wiley and Sons, Hoboken, NJ, 2004.
- [79] R.A. Maronna, R.D. Martin, and V.J. Yohai, *Robust Statistics: Theory and Applications*, Wiley, West Sussex, England, 2006.
- [80] Barnett V. and Lewis T., *Outliers in Statistical Data*, John Wiley and Sons, Chichester, 1978.
- [81] Atkinson A.C., “Regression diagnostics, transformations and constructed variables,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 1, pp. 1–36, 1982.
- [82] Atkinson A.C., “Diagnostic regression analysis and shifted power transformations,” *Technometrics*, vol. 25, no. 1, pp. 23–33, Feb 1983.

- [83] Atkinson A.C., “Diagnostic tests for transformations,” *Technometrics*, vol. 28, no. 1, pp. 29–37, Feb 1986.
- [84] Atkinson A.C., “Transformations unmasked,” *Technometrics*, vol. 30, no. 3, pp. 311–318, Aug 1988.
- [85] Mili L. Robust Estimation and Filtering, Class Notes, 2004.
- [86] Mili L., Phaniraj V., and Rousseeuw P.J. in book *Control and Dynamic Systems*, ed. C.T. Leondes, 1990.
- [87] Huber P.J., “Robust estimation of a location parameter,” *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.
- [88] Tukey J., “A survey of sampling from contaminated distributions,” in *Contributions to Probability and Statistics*, in I. Olkin, Ed. Stanford University Press, 1960, IEEE Intl. Symp. on Phased Array Systems and Tech., pp. 448–485.
- [89] Hampel F.R., *Contributions to the theory of robust estimation*, Ph.D. thesis, University of California, Berkeley, 1968.
- [90] Hampel F.R., “A general qualitative definition of robustness,” *Ann. Math. Statist.*, vol. 42, pp. 1887–1896, 1971.
- [91] Hampel F.R., “The influence curve and its role in robust estimation,” *Ann. Math. Statist.*, vol. 69, pp. 383–393, 1974.
- [92] Ronchetti E., *Robustheitseigenschaften von Tests*, Ph.D. thesis, E.T.H. Zurich, Switzerland, 1979.
- [93] Rousseeuw P.J., “Optimally robust procedures in the infinitesimal sense,” 1979, 42nd Session of the ISI, pp. 467–470.

- [94] Rousseeuw P.J. and Ronchetti E., “Influence curves for general statistics,” *Journal of Comput. Appl. Math.*, vol. 7, pp. 161–166, 1981.
- [95] Stigler S.M., “Simon Newcomb, Percy Daniell, and the history of robust estimation 1885 - 1920,” *Journal of the American Statistical Association*, vol. 68, no. 344, pp. 872–879, December 1973.
- [96] Krasker W.S. and Welsch R.E., “Efficient bounded influence regression estimation,” *Journal of the American Statistical Association*, vol. 77, pp. 595–604, 1982.
- [97] Maronna R.A., Bustos O.H., and Yohai V.J. in book *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757 eds. T. Gasser and J.M. Rossenblat, 1979.
- [98] Thomas L. and Mili L., “A robust GM-Estimator for the automated detection of external defects on barked hardwood logs and stems,” *IEEE Transaction on Signal Processing*, pp. Submitted,, 2007.
- [99] Rousseeuw P.J. and Van Driessen K., “A fast algorithm for the minimum covariance determinant estimator,” *Technometrics*, vol. 41, pp. 212–223, 1999.
- [100] Miller K.S., *Complex Stochastic Processes*, Addison-Wesley Publishing Company, Inc., London, 1974.
- [101] Rousseeuw P. and Croux C., “Alternatives to the median absolute deviation,” *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, December 1993.
- [102] Babu B.N.S., Torres J.A., and Melvin W.L., “Processing and evaluation of multichannel airborne radar measurements (MCARM) measured data,” 1996, IEEE Intl. Symp. on Phased Array Systems and Tech., pp. 395–399.

- [103] Rohling H., “Radar cfar thresholding in clutter and multiple target situations,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 19, no. 4, pp. 608–621, 1983.
- [104] Robey F.C., Fuhrman D.R., Kelly E.J., and Nitzberg R., “A cfar adaptive matched filter detector,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, January 1992.
- [105] Chen W.S. and Reed I.S., “A new CFAR detection test for radar,” *Digital Signal Proc.*, vol. 4, pp. 198–214, October 1991.
- [106] Meng H., Wang X., Zhang H., and Peng Y., “An approach based on influence function to evaluate robustness and detection performance of CFAR detectors,” 2004, vol. 2 of *Proceedings of IEEE ICASSP*, pp. 1053–1056.
- [107] Meng H., Wang X., Zhang H., and Peng Y., “A method using influence function for evaluating robustness of cfar detectors,” Philadelphia, PA, USA, 2004, Proceedings of the IEEE Radar Conference, pp. 620–623.
- [108] Bergin J. S., Teixeira C. M., Techau P. M., and Guerri J. R., “Space-time beamforming with knowledge-aided constraints,” 2002, Asilomar Conference on Signals, Systems, and Computers, pp. 881–884.
- [109] Bergin J.S., Teixeira C.M., Techau P.M., and Guerri J.R., “STAP with knowledge-aided data pre-whitening,” Philadelphia, PA, USA, 2004, Proceedings of the IEEE Radar Conference, pp. 289–294.

Vita

Gregory Neumann Schoenig was born on August 3, 1977 in Philadelphia, PA, USA. He earned the B.S. in Electrical Engineering, M.S. in Electrical Engineering, and M.S. in Computer Science from Johns Hopkins University in 1998, 2000, and 2001, respectively. He received an M.B.A. from The George Washington University in 2003. Mr. Schoenig is currently an Assistant Vice President and Senior Systems Engineer at Science Applications International Corporation (SAIC) in Chantilly, VA, responsible for the conception and development of new signal processing capabilities for the Department of Defense.