Advanced Signal Processing Techniques for Cognitive Radar Systems

Ahmed A. Abouelfadl

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

December 2019

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Contents

Table of Contents i

Abstract v

Sommaire vii

Acknowledgment ix

Contributions of Authors x

List of Figures xi

List of Tables xiv

List of Acronyms xv

1 Introduction 1

1.1 Radar Overview 1
1.2 Cognitive Multiple-Input Multiple-Output Radars 2
1.3 Research Motivations 3
1.4 Objectives and Contributions 6
1.5 Thesis Organization 8

2 Background 9

2.1 Radar System Workflow 9
2.2 Space Time Adaptive Processing (STAP) 11
2.3 MIMO Radars 15

2.4 Cognitive Radars . . . 18
2.5 Radar Detection . . . 19
2.5.1 Scalar CFAR . . . 20
2.5.2 Vector CFAR . . . 21
2.6 Mathematical Background . . . 22
2.6.1 Depth functions . . . 22
2.6.2 Proximal Optimization . . . 25
2.6.3 Dynamic Bayesian graphical models . . . 27
2.6.3.1 State space models . . . 28
2.6.3.2 Hidden Markov models . . . 33

3 Covariance-Free Nonparametric Nonhomogeneity Detector 35
3.1 Introduction . . . 35
3.2 Signal Model . . . 37
3.3 Spherical Invariant Random Process (SIRP) Clutter Model . . . 39
3.4 The non-homogeneity Detector (NHD) . . . 41
3.5 The Proposed NHD . . . 43
3.5.1 Projection depth function . . . 43
3.5.2 Covariance-free reformulation of GIP and NAMF . . . 45
3.5.3 Robust, Covariance-free, and nonparametric NHD . . . 46
3.5.4 Correlated Clutter . . . 49
3.6 Performance Assessment . . . 53
3.6.1 Simulation Parameters . . . 53
3.6.2 Results . . . 54
3.6.2.1 The Low Dimensional Case ($J/L_2 = 0.25$) . . . 54
3.6.2.2 The Higher Dimensional Case ($J/L_2 = 0.5$) . . . 58
3.6.2.3 Number of Projections . . . 59
3.6.2.4 Complexity Analysis . . . 65
3.7 Conclusion . . . 66

4 Design of Power-Efficient Waveforms for Cognitive MIMO Radars 68
4.1 Introduction . . . 68
4.2 MIMO Radar Background . . . 71

4.2.1 MIMO Radar Signal Model . . . 71
4.2.2 Doppler Division Multiple Access in MIMO Radars . . . 73
4.2.3 Cognitive MIMO Waveform Design for Extended Targets . . . 74
4.3 MIMO Antenna Arrays . . . 75
4.3.1 MIMO Virtual Antenna Array . . . 75
4.3.2 Mutual Coupling of Transmitting Antenna Arrays . . . 77
4.3.3 Microwave Techniques for Protection against High Reflection . . . 78
4.4 Proposed waveform design for cognitive MIMO radar . . . 79
4.4.1 Solution Using Lagrange Method with ℓ_2-Norm Regularization . . . 80
4.4.2 Solution Using Proximal Gradient with ℓ_∞ Norm Regularization . . . 81
4.4.2.1 Computation of the Projection onto the unit ℓ_1-Ball . . . 83
4.4.2.2 Convergence of the Proposed Proximal-Based Algorithm . . . 85
4.5 Performance Evaluation . . . 86
4.5.1 Simulation Setup . . . 86
4.5.2 Performance Evaluation with 4-Element Transmitting Antenna Array 88
4.5.3 Performance Evaluation with 8-Element Transmitting Antenna Array 93
4.5.4 Complexity Analysis . . . 100
4.6 Conclusion . . . 102

5 Extended Target Frequency Response Estimation Using Infinite HMM 104
5.1 Introduction . . . 104
5.2 Extended Target Model . . . 106
5.3 TFR Generating Models and Distributions . . . 108
5.3.1 Linear State Space Generating Model . . . 109
5.3.2 Correlated Random Process Model . . . 109
5.3.3 TFR distributions . . . 110
5.4 TRF Estimation: A New Formulation . . . 112
5.4.1 HMM as a Stochastic Finite State Machine . . . 112
5.4.2 TFR Modeling Using HMM . . . 113
5.5 TFR Estimation and Tracking . . . 115
5.5.1 Bayesian Nonparametric (BNP) Models and Dirichlet Process . . . 115
5.5.2 Employing iHMM in TFR Modeling . . . 117
5.5.3 TFR Estimation Using iHMM . . . 118

5.5.3.1 TFR inference . . . 118
5.5.3.2 Inference of hyperparameters . . . 120
5.6 Performance Evaluation . . . 122
5.6.1 Simulation Setup . . . 122
5.6.1.1 Radar and clutter signal models . . . 122
5.6.1.2 Kalman Filter Design for TFR Estimation . . . 123
5.6.1.3 Particle Filter Design for TFR Estimation . . . 123
5.6.1.4 Parameters of the Proposed iHMM for TFR Estimation . . . 124
5.6.2 Results and Discussion . . . 125
5.6.2.1 Linear State Space (LSS) TRF Model . . . 125
5.6.2.2 SIRP TRF Model . . . 130
5.6.2.3 Complexity Analysis . . . 134
5.7 Conclusion . . . 136

6 Conclusion and Future Work **138**

Appendix A Proof of Proposition 3.1 **142**

Appendix B Proof of Proposition 3.2 **144**

Appendix C Proof of Proposition 3.3 **146**

Appendix D Proof of Proposition 4.1 **147**

Appendix E Proof of Lipschitz continuity of $\nabla u(\mathbf{f})$ in Eq. (4.33) **150**

Bibliography **152**

Abstract

This thesis introduces novel signal processing algorithms for cognitive radar systems that consider the constraints of operation under real scenarios as well as hardware limitations. Specifically, the thesis focuses on the analysis and practical solution of three key problems encountered in the design and realization of the signal processing chain within the cognitive new-generation radars.

Firstly, we consider detecting and excluding the non-homogeneous received data from the estimation of the interference covariance matrix. The available non-homogeneity detectors (NHDs) in the literature require estimating the covariance matrix for each examined data cell, leading to exacerbating the NHD complexity, especially with large-dimensional data. Instead, we employ the projection depth functions, inherited from the field of robust statistics, to formulate a new NHD test statistic that avoids estimating the covariance matrix. Moreover, the projection depth function converts the multivariate problem to a scalar one, evading the exponential growth of the computational complexity with the data dimension.

Secondly, we turn our attention to a scarcely but nevertheless important discussed aspect of radar system, namely the waveform design for cognitive multi-input multi-output (MIMO) radars taking into account the reflective properties of the transmitting antenna array. For the first time, we propose a waveform design method using proximal optimization that not only improves the signal-to-interference plus noise ratio (SINR), but also lowers the reflected power from the transmitting antenna array. Consequently, the proposed waveform design method increases the radar system efficiency and protects the amplification unit of the transmitter, while at the same time, significantly improves the SINR.

Finally, we introduce a novel formulation of the target frequency response (TFR) estimation problem, a crucial requirement for cognitive radars. Under the conventional assumption of a linear Gaussian model, the TFR is usually estimated using the Kalman filter. Surprisingly, even though in practice this assumption is often violated and the Kalman filter is no longer an optimal solution, the study of TFR estimation for more general models has not yet been considered. In our proposed formulation, the infinite hidden Markov model (iHMM) is used in TFR estimation without prior knowledge of the channel or the interference. Interestingly, when iterated over multiple pulses and under jamming conditions, the proposed estimation method exhibits superior performance compared to the Kalman and particle filters for different TFR models.

Throughout the thesis, the newly proposed algorithms are evaluated by objective Monte Carlo simulations with different clutter distributions and radar parameters. Under the considered evaluation conditions, the results clearly show that the proposed methods can provide superior performance to existing benchmarks from the literature.

Sommaire

Cette thèse présente de nouveaux algorithmes de traitement du signal pour les systèmes de radar cognitif qui tiennent compte des contraintes de fonctionnement dans des scénarios réels ainsi que des limitations matérielles. Plus précisément, la thèse se concentre sur l'analyse et la solution pratique de trois problèmes clés rencontrés dans la conception et la réalisation de la chaîne de traitement du signal au sein des radars cognitifs de nouvelle génération.

Tout d'abord, nous considérons l'exclusion des données reçues non homogènes de l'estimation de la matrice de covariance d'interférence. Les détecteurs de non-homogénéité (non-homogeneity detectors, NHDs) disponibles dans la littérature nécessitent l'estimation de la matrice de covariance pour chaque cellule de données examinée, ce qui exacerbe la complexité de mise en ouevre, en particulier avec des données de grande dimension. Alternativement, nous utilisons les fonctions de profondeur de projection, héritées du domaine des statistiques robustes, pour formuler une nouvelle statistique de test NHD qui évite d'estimer la matrice de covariance. De plus, la fonction de profondeur de projection convertit le problème multivarié en un problème scalaire, évitant ainsi la croissance exponentielle de la complexité de calcul avec la dimension des données.

Deuxièmement, nous concentrons notre attention sur un aspect peu discuté du système radar, à savoir la question importante de la conception de forme d'onde pour les radars cognitifs multi-entrées multisorties (multi-input multi-output, MIMO), prenant en compte les propriétés de réflexion du réseau d'antennes émettrices. Pour la première fois, nous proposons une méthode de conception de forme d'onde qui améliore non seulement le rapport signal sur brouillage plus bruit (signal-to-interference plus noise ratio, SINR), mais réduit également la puissance réfléchie par le réseau d'antennes d'émission en utilisant l'optimisation proximale. Par conséquent, la méthode de conception de forme d'onde proposée augmente l'efficacité du système radar et protège l'unité d'amplification de l'émetteur, tout en améliorant considérablement le SINR.

Enfin, nous introduisons une nouvelle formulation du problème d'estimation de la réponse en fréquence cible (target frequency response, TFR), une exigence cruciale des radars cognitifs. Dans l'hypothèse conventionnelle d'un modèle linéaire gaussien, le TFR est généralement estimé à l'aide du filtre de Kalman. Alors que, dans la pratique, cette hypothèse est souvent violée et que le filtre de Kalman n'est plus une solution optimale, l'étude de l'estimation de l'ISF pour des modèles plus généraux n'a pas encore été envisagée. Dans notre formulation

proposée, le modèle de Markov caché à l'infini (infinite hidden Markov model, iHMM) est utilisé dans l'estimation du TFR sans connaissance préalable du canal ou du brouillage. Fait intéressant, lorsqu'elle est itérée sur plusieurs impulsions et dans des conditions de brouillage, la méthode d'estimation proposée présente des performances supérieures à celles de Kalman et aux filtres à particules pour différents modèles de TFR.

Les algorithmes proposés sont évalués par des simulations objectives de Monte Carlo avec différentes distributions de fouillis, signaux de brouillage et paramètres radar. Dans les conditions de test considérées les résultats montrent que les méthodes proposées offrent une performance supérieure à leurs homologues de référence dans la littérature.

Acknowledgment

This thesis represents a milestone after the continuous hard work of three years at McGill university. My experience at McGill has been nothing short of amazing.

First and foremost I would like to thank Professors Benoit Champagne and Ioannis Psaromiligkos for giving me the wonderful opportunity to complete my PhD thesis under their supervision, it is truly an honor. Thank you for all the advice, ideas, moral support and patience in guiding me through this research. Without your guidance and constant feedback this PhD would not have been achievable. Thank you also for your enthusiasm for the study of radar signal processing, your wealth of knowledge in the field of signal processing in particular is inspiring. During the most difficult times when writing this thesis, they gave me the moral support and the freedom I needed to move on.

I am also grateful to my Doctoral Advisory Committee—Professors Mark Coates, Hannah Michalska, and Marcello Colombino for their thorough comments and fruitful discussions. I want also to thank Dr. Mohamed Samir Abdel Latif, the supervisor of my MSc thesis, who supported me allover this journey. The thanks also extend to all the staff of the administration of the PhD program in Egypt and special thanks to Professor Moataz Salah for his unlimited support. I express my deep appreciation to professor Yahia Mohasseb for all the support and guidance he provided before and during my joining McGill. I was honored to have been ranked the first among the candidates to the generous PhD scholarship offered by the Egyptian Ministry of Defense, which was a great motivation to conduct my research.

I finish where the most basic source of my life energy resides—my amazing family. The old dream of my mom and dad for me is becoming true now by virtue of their efforts and prayers. Thank you for all you provided, and still provide, to me. Amina, my beloved wife, I owe you a great debt of gratitude for all the support and understanding you showed during these three years. As I was absent most of the time, you took the greater share of responsibility at home. You got my back with all the ups and downs of this journey we went through together. I want also to thank my children, Baraa and Anas, for their patience for missing me when they wanted me around. Lastly, special thanks to my little princesses, Maria and Malika, who offered all the joy in the world that helped me to overcome any difficulty.

Contribution of Authors

The contributions of the authors to the work presented in this thesis is as follows: the first author, Mr. Abouelfadl, developed the ideas, derived and implemented the algorithms, conducted the simulations and wrote a first draft of each chapter. The co-authors, Professors Psaromiligkos and Champagne, supervised the work by providing guidance, validating theoretical developments, and contributing to the editing and writing of the final manuscripts.

List of Figures

1.1 Generic functional diagram of cognitive MIMO radar 4

2.1 Radar Echo 10
2.2 Radar Pulsed Waveform 10
2.3 Radar receiver block diagram. 12
2.4 CPI Cube 13
2.5 MIMO radar. 15
2.6 The virtual array of MIMO radar. 16
2.7 Cognitive radar work flow. 19
2.8 Block diagram of CA-CFAR 20
2.9 Example of a probabilistic graphical model 27
2.10 The different types of state space models 29

3.1 The relative SD of k_f at different number of samples (secondary cells) with $J = 20$ 48
3.2 Detection performance in K-distributed clutter ($\alpha = \ 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$) 55
3.3 Detection performance in Gaussian clutter ($J = 16$, $L_2 = 64$) 55
3.4 Detection performance in Gaussian and K-distributed clutter for PD-NAMF with MAD and SD ($J = 16, L_2 = 64$) 57
3.5 Detection performance in Gaussian clutter for GIP, modified PD-GIP, and PD-NAMF ($J = 16, L_2 = 64$) 58
3.6 Detection performance in K-distributed clutter for GIP, modified PD-GIP, and PD-NAMF ($J = 16, L_2 = 64$) 59

3.7 Detection performance in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}[1,2]$ or $\mathcal{U}(0,1]$, $J = 16$, $L_2 = 64$, $\rho_s = \rho_t = 0.99$ or 0.2) 60
3.8 Detection performance in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 32$) 61
3.9 Detection performance in Gaussian distributed clutter ($J = 16, L_2 = 32$) . . 62
3.10 Interference and the interfering target powers at the output of the PD-NAMF in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$) 62
3.11 Detection performance in K-distributed clutter with different Q values ($\hat{\mathbf{\Psi}}^S$, ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$) 63
3.12 Detection performance in K-distributed clutter with different Q values ($\hat{\mathbf{\Psi}}^K$, $\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$) 63
3.13 Detection performance in Gaussian distributed clutter with different Q ($\hat{\mathbf{\Psi}}^K$, $J = 16, L_2 = 64$) 64
3.14 Detection performance in Gaussian distributed clutter with different Q ($\hat{\mathbf{\Psi}}^S$, $J = 16, L_2 = 64$) 64
3.15 The run times of the PD-NAMF normalized by that of NAMF ($L_2 = 4J$) . . 65

4.1 ARC of the 4 antenna elements (Gaussian TIR). 89
4.2 ARC of the 4 antenna elements (K-distributed TIR). 90
4.3 SINR for the 4-element antenna array. 92
4.4 TARC ECDF for the 4-element antenna array. 93
4.5 ARC of the 8 antenna elements (Gaussian TIR). 96
4.6 ARC of the 8 antenna elements (K-distributed TIR). 98
4.7 SINR for the 8-element antenna array. 99
4.8 TARC ECDF for the 8-element antenna array. 100
4.9 The efficiency of the each element in the 4-element array for Gaussian and K-distributed TIR 101
4.10 The efficiency of the each element in the 8-element array for Gaussian and K-distributed TIR 101
4.11 Execution times of the three considered waveform design methods. 102

5.1 Extended target model. 107
5.2 Hierarchical Polya urn scheme. 119
5.3 Estimation error of Gaussian TFR in Gaussian clutter (LSS model). 125

5.4 Estimation error of Gaussian TFR in jamming and Gaussian clutter ($m = 1$, LSS model). 126
5.5 Estimation error of $K-$distributed TFR and clutter (LSS model). 127
5.6 Estimation error of Gaussian TFR in $K-$distributed clutter (LSS model). . . 127
5.7 Estimation error of Log-Normal TFR in $K-$distributed clutter (LSS model). 128
5.8 Estimation error of Weibull TFR in $K-$distributed clutter (LSS model). . . 129
5.9 Estimation error of Weibull TFR in jamming and K-distributed clutter ($m = 1$, SIRP model). 130
5.10 Estimation error of Gaussian TFR in Gaussian clutter (SIRP model). 131
5.11 Estimation error of $K-$distributed TFR and clutter (SIRP model). 132
5.12 Estimation error of Log-normal TFR and $K-$distributed clutter (SIRP model).133
5.13 Estimation error of Weibull TFR and $K-$distributed clutter (SIRP model). . 134
5.14 Complexity of the proposed iHMM-based method compared to the PF assuming the LSS model. 136
5.15 Complexity of the proposed iHMM-based method compared to the PF assuming the SIRP model. 137

List of Tables

3.1 The performed operations by the proposed and the NAMF tests. 66

4.1 ARC statistical data of 4-element antenna array 94
4.2 ARC statistical data of 8-element antenna array 97

List of Acronyms

ADC	Analog-to-digital converter
AMF	Adaptive matched filter
AWGN	Additive white Gaussian noise
BNP	Bayesian non-parametric
CA-CFAR	Cell average-CFAR
CAOS-CFAR	Cell average OS-CFAR
CDF	Cumulative distribution function
CFAR	Constant false alarm rate
CPI	Coherent pulse interval
CT-HMM	Continuous-time hidden Markov model
CUT	Cell under test
CW	Continuous wave
DAC	Digital-to-analog converter
FSM	Finite state machine
GIP	Generalized inner product
GLRT	Generalized likelihood ratio test

GMTI	Ground moving target identification.
GO-CFAR	Greatest of-CFAR
HMM	Hidden Markov model
iHMM	Infinite hidden Markov model
iid	Independent and identically distributed
LRT	Likelihood ratio test
MAD	Median absolute deviation
MEMS	Micro-electro-mechanical systems
MI	Mutual information
MIMO	Multiple-input multiple-output
ML	Maximum likelihood
MMSE	Minimum mean-square error.
MSE	Mean square error
NAMF	Normalized adaptive matched filter
NHD	Non-homogeneity detector
NSCM	Normalized sample covariance matrix
OS-CFAR	Order statistics-CFAR
PD-GIP	Projection depth-GIP
PD-NAMF	Projection depth-NAMF
PDF	Probability density function
PRF	Pulse repetition frequency
PRI	Pulse repetition interval

PSLR	Peak-to-sidelobes level ratio.
PW	Pulse width
RSD	Relative standard deviation
SCM	Sample covariance matrix
SD	Standard deviation
SFSM	Stochastic FSM
SiGe	Silicon-Germanium
SINR	Signal-to-interference plus noise ratio
SIRP	Spherical invariant random process
SIS	Sequential importance sampling
SMC	Sequential Monte Carlo
STAP	Space time domain adaptive processing
TIR	Target impulse response
TRF	Target frequency response
ZMNL	Zero memory nonlinear

Chapter 1

Introduction

1.1 Radar Overview

RADAR (RADio Detection And Ranging) was first invented by the British scientist Robert Watson-Watt in 1935. In his own words, radar is "the art of detecting by means of radar echoes the presence of objects, determining their direction and ranges, recognizing their character and employing data thus maintained in the performance of military, navy or other operations [1]." Since their invention, radar systems have been tremendously evolved to ensure their robust performance under harsh conditions such as the presence of dense echos from the environment, known as clutter, and electronic counter measures that have been conceived to deny their functionality. Although the radar has been initially invented for military applications, radar systems have been utilized in a many other applications. This includes, geophysics radars used to create soil profiles, anti-collision radars in modern vehicles, and medical radars used as monitoring systems and for diagnostic procedures.

There are different classification criteria for radar systems such as the type of transmitted waveform, the targeted application, or the carrier platform. A simple approach is to classify radar systems on the basis of their functionality, which leads to three main categories, namely: search, tracking, and imaging radars. In this thesis, we are concerned with search radars, whose task is to detect the presence of a target in the presence of noise, clutter, jamming, or any other forms of interference.

The main purpose of modern radar systems is to: improve detection, tracking, and target identification performance and capabilities; decrease the probability of intercept of radar signals; counter electronic warfare (EW) operations and severe environmental conditions;

and increase the abilities of field operation. To achieve these objectives, different advanced radio and signal processing techniques are applied to different radar subsystems.

The evolution of phased array antennas provides modern radar systems with several performance advantages including better spatial resolution, superior scanning via electronic beam steering (as opposed to mechanical steering), and multi-target capability. In fact, the phased array antenna is not only a breakthrough as a hardware development, it also initiated the development of a plethora of signal processing algorithms that in turn have given rise rise to different types of modern radar systems. Since real time operation is one of the distinguishing features of modern radar systems, the development and study of advanced algorithms for on-line processing of phased array antenna signals is henceforth an active area of research within the radar community.

1.2 Cognitive Multiple-Input Multiple-Output Radars

Phased arrays have become the most commonly used antenna type in modern radar systems. The reasons for this are plentiful. First, phased arrays provide high reliability, high bandwidth, and excellent sidelobe control. Second, there are certain applications for which phased arrays are uniquely qualified. For instance, in airborne applications, because they can electronically steer the beam to extreme angles while maintaining a low profile, hence minimizing aircraft drag. They are also ideal for ground radar systems, which in some cases are too large for mechanical rotation. Finally, phased arrays have remarkable electronic beam agility that enables multiple functions to be performed nearly simultaneously by a single radar.

The number of antenna elements in the phased array defines the spatial resolution of the radar system, the number of directions where the interference can be nulled, and the processing gain obtained by the coherent processing of the multiple spatial channels. Therefore, it is apparent that increasing the number of antenna elements improves the performance of the radar. However, besides the high cost of large phased array antennas, the complexity of the signal processing associated with the increase in the number of the spatial channels can be overwhelming.

A multiple-input multiple-output (MIMO) radar transmits from its antenna elements different and usually orthogonal set of waveforms characterized with better spatial resolution. This waveform diversity has motivated a huge body of research with the aim of optimizing

the transmitted waveforms from various perspectives, including: SINR maximization, good ambiguity functions, and optimizing the spatial coverage. To achieve these objectives, besides the requirement of orthogonal waveforms, MIMO radars need to know the covariance matrix of the interference, which complicates the waveform design problem for MIMO radars.

The classic model for radar targets is the point target model, which consists of a single physical reflector. In practice, however, some targets are more complex in nature and this simple model does not apply, as for instance when the target return is composed of several reflections. In this case, the target is more adequately modeled by the target impulse response (TIR), which is in effect the impulse response of a linear system modeling the reflection process.. For this class of targets, the transmitted waveform is preferably adapted to the TIR. This adaptation of the radar transmitter to the TIR, and to the environment in general, is the defining characteristic of the so-called cognitive radars. Therefore, the waveform design problem for cognitive MIMO radars is even more complex than for conventional MIMO radars as it involves the estimation of the TIR, which is conventionally unknown, in addition to improving the SINR and preserving the orthogonality of the transmitted waveforms.

1.3 Research Motivations

We discuss the motivations of the research work presented in this thesis with the aid of Fig. 1.1, which depicts a generic functional diagram of a radar system. The received signals at the N_R receiving antenna elements are each passed to a chain of radio frequency (RF) circuits for filtration, amplification, and down conversion. The output baseband signals then proceed to the analog-to-digital converters (ADCs) and matched filters (MF). The outputs of the matched filters from all receiving channels for multiple pulses are then processed by the non-homogeneity detector (NHD), which finds the outliers within the set of received signal samples to be excluded from covariance estimation. The estimated covariance matrix is then used in the final detection of the target based on space time adaptive processing (STAP). In the case of a MIMO radar, it is also fed to the transmitter for the purpose of waveform design. For cognitive radars, both the covariance matrix and the received signal samples are passed to the transmitter for TIR estimation. In all cases, the transmit waveforms are converted to the analog domain using digital-to-analog converters (DACs) and then up-converted to the required RF and finally amplified for transmission by the N_T antennas. It should be noted that the functional diagram in Fig. 1.1 reduces to that of the conventional

phased array radar by removing the TIR estimation and the waveform design blocks and, consequently, the feedback from the receiver to the transmitter from Fig. 1.1. Therefore, the NHD is common to the conventional phased array radars as well as the more advanced MIMO and cognitive MIMO radars.

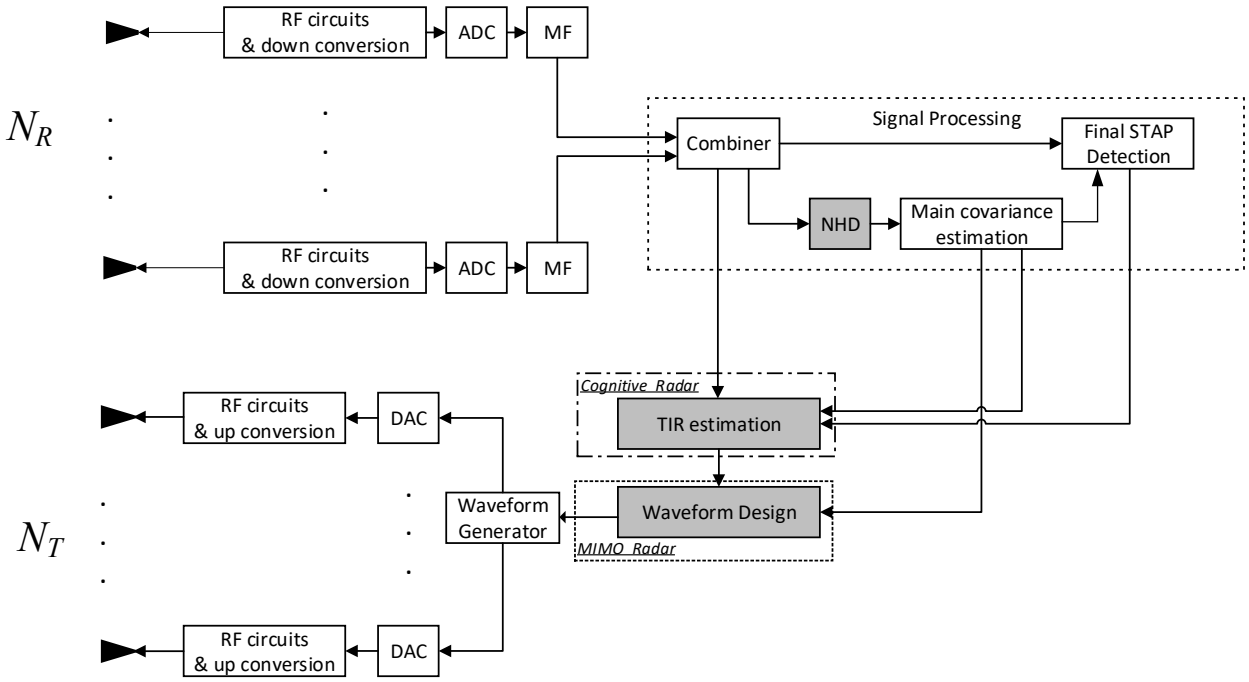


Figure 1.1 Generic functional diagram of cognitive MIMO radar

The focus of the research presented in this thesis targets the three shaded blocks in Fig. 1.1, i.e., NHD, TIR estimation, and waveform design. Following a thorough survey of the literature, our research motivations are summarized as follows:

- **NHD**: A STAP detector applies a test statistic on a given range cell to explore the target presence in this cell. The used test statistics require estimating the covariance matrix of this cell, and this estimation process is denoted as "main covariance estimation", as in Fig. 1.1. Since the covariance matrix of this given cell is unknown in practice, it is estimated from the sample data surrounding this cell, which are nothing but other range cells. These cells may contain non-homogeneous components, including target, jammer and high-power clutter signals, which if included in the sample cells may significantly degrade the estimation accuracy of the covariance matrix. The NHD

tests the sample cells for outliers, so that the non-homogeneous cells are censored from the "main covariance estimation" [2]. However, the NHD is basically a STAP detector that requires estimating the covariance matrix of each sample cell tested for homogeneity, which we denote here as the "scondary covariance estimation". Therefore, to detect a target in a given range cell, we need to perform the "main covariance estimation", which in turn requires performing the "scondary covariance estimation" for all the sample cells. For real-time processing applications, the covariance estimation needed as part of the NHD is quite challenging, especially in the presence of non-Gaussian interference. Existing solutions are not computationally efficient and require considerable resources [3].

- **Waveform Design**: As we mentioned earlier, in addition to the orthogonality among transmitted waveforms, there are several criteria for waveform design in MIMO radars such as achieving ambiguity functions and spatial coverage with desirable properties, as well as matching the waveform spectrum to that of the TIR in the case of extended targets. However, the effect of the transmitted waveforms on the magnitude of the reflected power back from the transmitting antenna array has not been considered in the waveform design algorithms so far [4]. The ratio of the reflected power to the input power of the antenna determines its efficiency and, in turn, the effective radiated power. Moreover, if the reflected power is high, it may cause damage to the amplification stage preceding the transmitting antenna. From a practical perspective, the reflection properties of the antenna should be considered in the design of cognitive MIMO radars for efficient performance and hardware durability [5].

- **TIR Estimation**: The TIR is conventionally assumed to be known or to follow a linear Gaussian model. While the former assumption is merely introduced for the sake of simplification, the latter assumption is made to allow the use of the Kalman filter in TIR estimation [6]. In practice, however, the measured TIR data exhibits non-Gaussian distributions, for which the Kalman filter is no longer optimal [7]. While the particle filter seems to be a rational alternative to the Kalman filter assuming non-Gaussian distributions, the random generating model of the TIR may degrade the accuracy of the particle filter estimation due to model mismatch [8]. In addition, the complexity of the particle filter is another concern, especially in real-time applications. Hence, there is a need for TIR estimation algorithms that can be applied to a wider range

of distributions and generating models, and this with low to moderate complexity of implementation.

1.4 Objectives and Contributions

Considering the the above limitations of existing techniques for the three signal radar signal processing stages identified in Fig. 1.1, the main objectives of this thesis are:

- To develop a covariance-free test statistic for the NHD that provides a robust performance and lower computational cost compared to competing detectors in the literature, and this without prior knowledge about the statistics of the interference.
- To design orthogonal cognitive MIMO radar waveforms, assuming known TIR, that achieve both SINR improvement and low reflected power back from the transmitting antenna array.
- To estimate the TIR without prior information about neither the TIR nor the interference. The estimation accuracy should be also evaluated over multiple pulses and under harsh operating conditions such as jamming.

The main contributions of this thesis in the light of the above-listed objectives are as follows:

1. Towards the first objective, inspired by robust statistics, the projection depth function is employed to derive a novel test statistic for the NHD based on the normalized adaptive matched filter (NAMF), the most well-known robust nonparametric NHD. By exploiting robust nonparametric statistical measures, such as the median, the median absolute deviation, and the Spearman/Kendall correlation matrices, as well as the dimension reduction of the projection depth function, the proposed detector provides robust performance, yet with significant computational savings. The performance of the proposed NHD is evaluated for different clutter distributions and for different radar parameter configurations and compared to that of a robust benchmark in the literature.
2. Regarding the second objective, we propose to add a regularization term to the objective function, which represents the reflected power from the transmitting antenna array in the waveform design problem. We use two different regularization terms that are expressed as functions of the scattering matrix of the transmitting antenna array. The

first regularization term is the ℓ_2-norm of the reflected signal from the transmitting antenna array, which represents the average reflected power from the antenna. The second regularization term is the ℓ_∞-norm of the reflected signal from the antenna, which represents the maximum reflected power from the transmitting antenna among all pulses and antenna elements. The proposed waveform design with the ℓ_∞-norm regularization is shown to provide a robust control over the reflected power from the transmitting antenna. The evaluation of the proposed method relies on the most relevant figures of merit, i.e., the active reflection and the total active reflection coefficients. In this evaluation, we consider different antenna array sizes along with various TIR distributions.

3. To achieve the third objective, we propose a novel formulation for the TIR estimation problem, or equivalently for the target frequency response (TRF) estimation problem, in which the samples of the TRF are modeled as stochastic finite state machine. To overcome the difficulty of obtaining prior knowledge of the number of states taken by the TRF, we adopt the nonparametric Bayesian models, which do not require prior information about the statistics of the TIR nor the interference. Specifically, we employ the infinite hidden Markov model (iHMM) and a modified beam sampling algorithm to infer the TFR. We also apply a new generating model for the TIR that does not suffer from the limitations of the linear Gaussian state-space model employed in the literature. The estimation accuracy of the proposed algorithm is shown to be superior to those of the Kalman and particle filters for different TIR distributions and under the effect of both traditional and smart noise jamming.

These contributions led to the following publications:

Journal papers

- Ahmed A. Abouelfadl, I. Psaromiligkos, and B. Champagne, "Covariance-free nonhomogeneity STAP detector in compound Gaussian clutter based on robust statistics," *IET Radar, Sonar & Navigation*, in press.

Conference papers

- Ahmed A. Abouelfadl, I. Psaromiligkos, and B. Champagne, "A low complexity nonparametric STAP detector," in *IEEE National Aerospace and Electronics Conf. (NAECON)*, (Ohio, USA), pp. 592–596, July 2018.

- Ahmed A Abouelfadl, I. Psaromiligkos, and B. Champagne, "Extended target frequency response estimation using infinite HMM in cognitive radars," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (Ottawa, Canada), Nov. 2019.

Regarding the contributions of the authors in all papers above, the first author, Mr. Abouelfadl, developed the idea, derived and implemented the algorithms, conducted the simulations and wrote a first draft of each manuscript. The co-authors, Professor Psaromiligkos and Champagne, supervised the work by providing guidance, validating theoretical developments, and contributing to the editing and writing of the final manuscripts.

1.5 Thesis Organization

Following this introduction, Chapter 2 provides a review of basic background on radar topics within the scope in thesis. This chapter also provides a very brief overview on the main mathematical tools used throughout the thesis. The first contribution of this thesis is presented in Chapter 3, where we propose a novel NHD based on projection depth function. In Chapter 4, we introduce a new waveform design method for cognitive MIMO radars that improves the SINR and, simultaneously, minimizes the reflected power from the transmitting antenna array based on the gradient proximal optimization method. The formulation of the TIR estimation problem using the iHMM and its performance evaluation are presented in Chapter 5. The conclusion and future work are discussed in Chapter 6.

Chapter 2

Background

In this chapter, we present a brief survey of fundamental radar topics within the scope of the thesis, including: radar system workflow; the basics of space time adaptive signal processing; modern radar systems such as multiple-input multiple-output (MIMO) and cognitive radars; and the principles of radar detection. In addition, we provide a quick review for the mathematical tools used throughout the thesis.

2.1 Radar System Workflow

Radar systems can extract target range, velocity, azimuth and elevation by processing the returned echo from the target as in Fig. 2.1. Range measurement depends on two radar parameters, namely maximum radar transmitted power and pulse repetition interval (PRI). In a pulse Doppler radar, the radar transmits a pulsed waveform as shown in Fig. 2.2. In the case of unambiguous target range, the target return-pulse is received in the time interval between the end of a transmitted pulse and the start of the next one . If range ambiguity happens, a target return pulse may be received after the end of the next pulse. There are many techniques to resolve radar ambiguity, but we do not consider the ambiguous case in this thesis. Target range is determined by measuring the time delay t_d between the leading edges of the transmitted and received radar pulses. Hence, the target range is given by [9]

$$R_t = \frac{c\,t_d}{2} \tag{2.1}$$

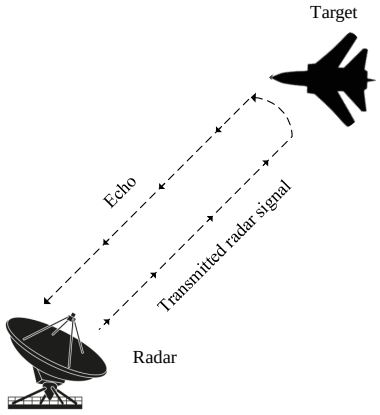


Figure 2.1 Radar Echo

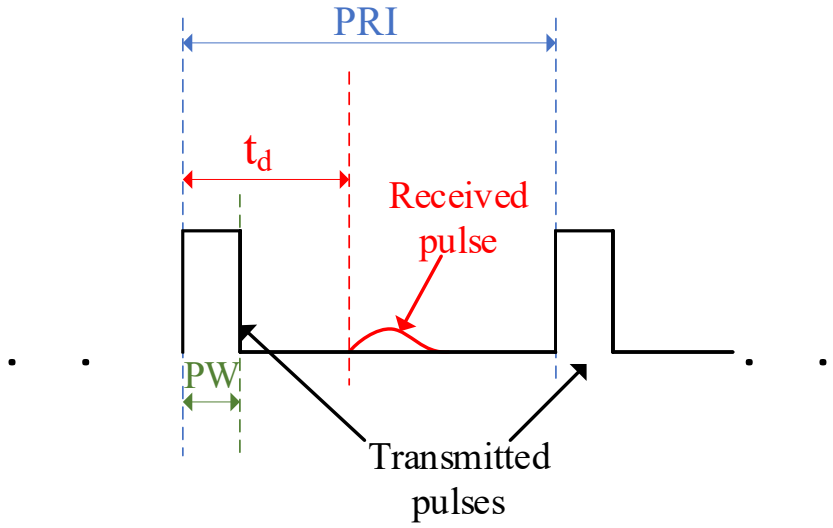


Figure 2.2 Radar Pulsed Waveform

where c is the speed of light. In the case of unmodulated pulse, the radar range resolution $\triangle R$ is determined by its pulse width (PW) using

$$\triangle R = \frac{c}{2}\mathrm{PW} \tag{2.2}$$

For an intra-pulse[1] modulated waveform, the range resolution is related to the radar bandwidth B as

$$\triangle R = \frac{c}{2B} \tag{2.3}$$

To determine a moving target velocity, radar measures the frequency shift f_d between the transmitted and the received pulses, i.e., Doppler shift. Pulse Doppler radars sample the Doppler frequency shift at the pulse repetition frequency (PRF), PRF=1/PRI, which leads to Doppler ambiguity if the PRF is not high enough. Therefore, the maximum and minimum unambiguous Doppler frequency shifts a radar can measure are determined by its PRF using [10]

$$f_{d_{max}} = \frac{\text{PRF}}{2}, \quad f_{d_{min}} = \frac{-\text{PRF}}{2} \tag{2.4}$$

Consider a phased array pulsed Doppler radar using a uniform linear array (ULA) of N antenna elements that are spaced $d = \lambda/2$ apart, where λ is the wavelength at the radar's center frequency. The radar simultaneously transmits from each antenna element a sequence of M coherent pulses with a PRI T, which define the so-called *slow* time domain. The transmitted signal from each antenna element is assumed to be narrowband, that is, its bandwidth B satisfies $B \ll c/Nd$ [11].
Each antenna element is preceded with a radio frequency (RF) processing stage for filtration, amplification, and down conversion. The resulting baseband signals are digitized using an analog-to-digital converter (A/D) and then passed through a matched filter as shown in Fig. 2.3. The matched filter is employed to maximize the signal-to-noise ratio in the presence of additive white Gaussian noise (AWGN). The M received pulses from each antenna element are stacked to form the coherent pulse interval (CPI) using M-tap delay line that are fed to the space time adaptive processor.

2.2 Space Time Adaptive Processing (STAP)

The normalized Doppler frequency shift due to the target velocity is expressed as

$$f_d = \frac{2\nu_r T \cos(\theta_t)}{\lambda} \tag{2.5}$$

[1]Intra-pulse modulation refers to the frequency and/or phase modulation of the carrier signal inside the radar pulse.

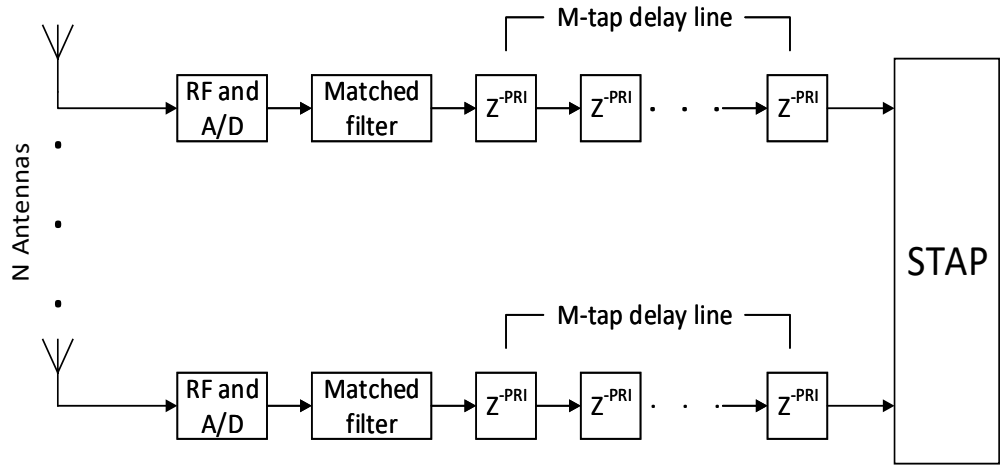


Figure 2.3 Radar receiver block diagram.

where θ_t is the target azimuth angle from the boresight of the radar antenna array; and the Doppler shift is normalized with respect to the pulse repetition frequency $1/T$. Let T_u be the time delay corresponding to the radar maximum unambiguous range, while the time delay corresponding to the radar range resolution is $1/B$. Hence, the total number of range cells in the so called *fast* time domain is

$$L = \lfloor T_u B \rfloor \tag{2.6}$$

where $\lfloor \cdot \rfloor$ denotes the floor integer. The CPI can be visualized as an $L \times M \times N$ data cube as shown in Fig. 2.4. For each range cell the data is an $MN \times 1$ vector that contains the received signal from the target from each PRI and antenna element. The M-dimensional temporal[2] steering vector of the target is given by [11]

$$\mathbf{b}(f_d) = \begin{bmatrix} 1 & \exp(j2\pi f_d) & \exp(j2\pi(2f_d) & \ldots & \exp(j2\pi((M-1)f_d)) \end{bmatrix} \tag{2.7}$$

The N-dimensional azimuth space steering vector[3] is given by

$$\mathbf{a}(\theta_t) = \begin{bmatrix} 1 & \exp(j2\pi \frac{d}{\lambda} \sin(\theta_t)) & \exp(j2\pi \frac{2d}{\lambda} \sin(\theta_t)) & \ldots & \exp(j2\pi \frac{(N-1)d}{\lambda} \sin(\theta_t)) \end{bmatrix} \tag{2.8}$$

[2]By temporal domain it is meant the slow time domain or the Doppler domain.

[3]The discussion here is limited to the azimuth plane; however, the same rules are applied to the elevation plane.

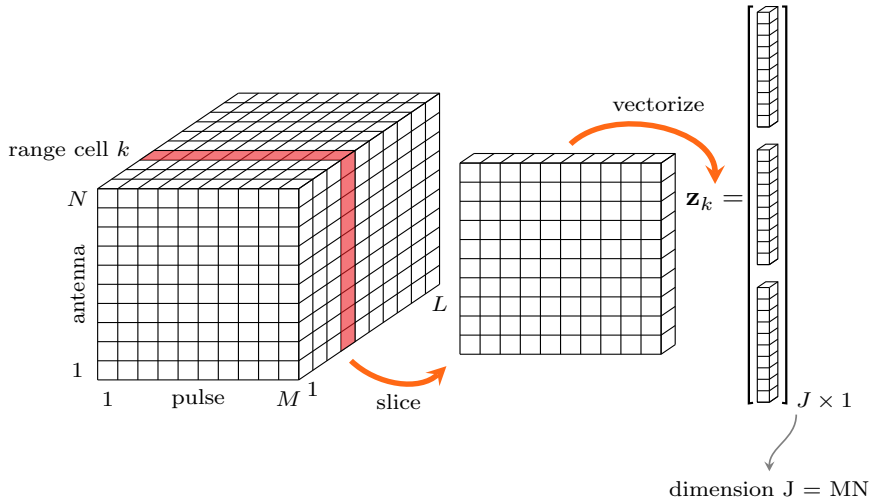


Figure 2.4 CPI Cube

The $MN \times 1$ spatio-temporal steering vector is given by

$$\mathbf{s}(f_d, \theta_t) = \frac{\mathbf{b}(f_d) \otimes \mathbf{a}(\theta_t)}{\|\mathbf{b}(f_d) \otimes \mathbf{a}(\theta_t)\|_2} \tag{2.9}$$

where $\otimes$ represents the Kronecker product. The baseband received signal $\mathbf{x}$ is expressed as

$$\mathbf{x} = a\mathbf{s} \tag{2.10}$$

where a is the complex amplitude of the received signal at the output of the matched filter. The received signal $\mathbf{z}$ can be expressed as

$$\mathbf{z} = \mathbf{x} + \mathbf{c} + \mathbf{n} \tag{2.11}$$

where $\mathbf{x}$ is the received signal from the target defined in Eq. (2.10), $\mathbf{c}$ is the clutter signal, and $\mathbf{n}$ is the additive white Gaussian noise. The covariance matrix of the received signal is an $MN \times MN$ matrix that is given by

$$\mathbf{R} = E[\mathbf{z}\mathbf{z}^H] \tag{2.12}$$

where $(\cdot)^H$ denotes the Hermitian transpose. Due to the presence of correlated clutter and, possibly, jamming, the covariance matrix $\mathbf{R}$ of the received signal is not diagonal. However,

since the white noise is generated in the receiver, $\mathbf{R}$ can be guaranteed to be positive definite and full rank [11]. With this fact in hand, a closed form solution for the optimum weight vector that maximizes the signal to interference noise ratio (SINR) can be reached using the following optimization problem

$$\max_{\mathbf{w}} \quad \frac{|\mathbf{w}^H\mathbf{s}|^2}{\mathbf{w}^H\mathbf{R}\mathbf{w}} \tag{2.13}$$

$$\text{subject to} \quad \mathbf{w}^H\mathbf{w} = 1 \tag{2.14}$$

Solving Eq. (2.13) results in the optimum weight vector

$$\mathbf{w} = k\mathbf{R}^{-1}\mathbf{s} \tag{2.15}$$

where k is a scalar and the optimum SINR is [11–13]

$$\text{SINR}_{\text{opt}} = |a|^2\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{2.16}$$

There is another, normalized, form of the weight vector, that is [14]

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}} \tag{2.17}$$

From Eq. (2.16) one can see that the signal should be adaptively processed in both space and temporal domains. That is why the name "space time domain adaptive processing" (STAP) is used.

As shown in Eq. (2.15), both the covariance matrix and space-time steering vector of the received signal should be known to form the weighting vector which is not a valid assumption in most cases. The spatial steering vector may be known in the case of radars that perform electronic scanning; however, the temporal steering vector is totally unknown to the radar receiver. The steering vector can be estimated by finding the steering vector that maximizes the SINR in Eq. (2.16) through scanning different Doppler shifts and angles [13]. It is assumed that target steering vectors are stationary during the CPI. This assumption is valid as long as the relative motion between the radar and the target does not result in a an angle difference more than orders of $1/100^{th}$ of the beam width, which is the case in most radar situations [15].

Different estimation techniques are used to estimate the covariance matrix. The maximum likelihood (ML) estimator of the covariance matrix $\hat{\mathbf{R}}$ for a Gaussian distributed interference is given by

$$\hat{\mathbf{R}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{z}_k \tag{2.18}$$

where $K \geq 2MN$ is the required number of the secondary cells [16], $\hat{\mathbf{R}}$ in Eq. (2.18) is known as the sample covariance matrix (SCM). It should be emphasized that the SCM is not robust in the case of non Gaussian interference, in which case other estimators should be used as will be shown in the next chapter.

2.3 MIMO Radars

In contrast to phased array radars, MIMO radars transmit independent waveforms from the transmitting antenna elements and observing the target(s) returns by the receiving antenna elements. The operation of the MIMO radar is illustrated in Fig. 2.5.

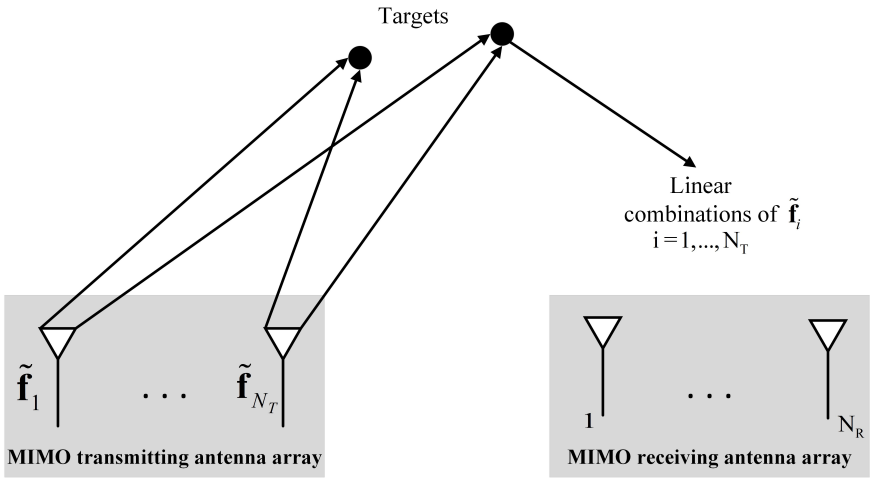
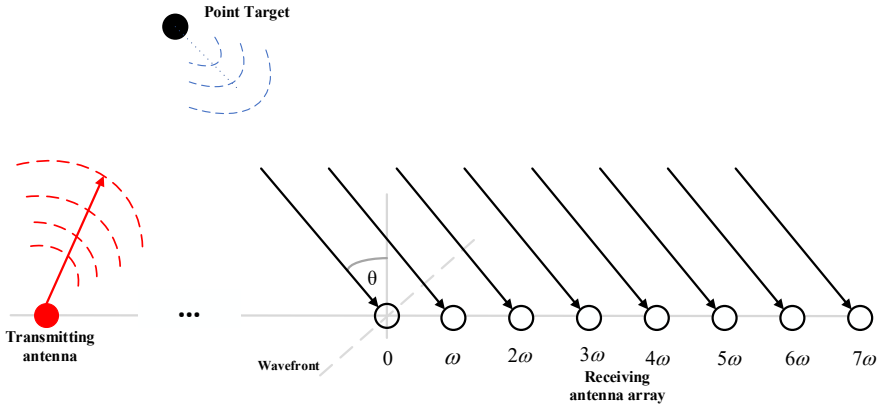


Figure 2.5 MIMO radar.

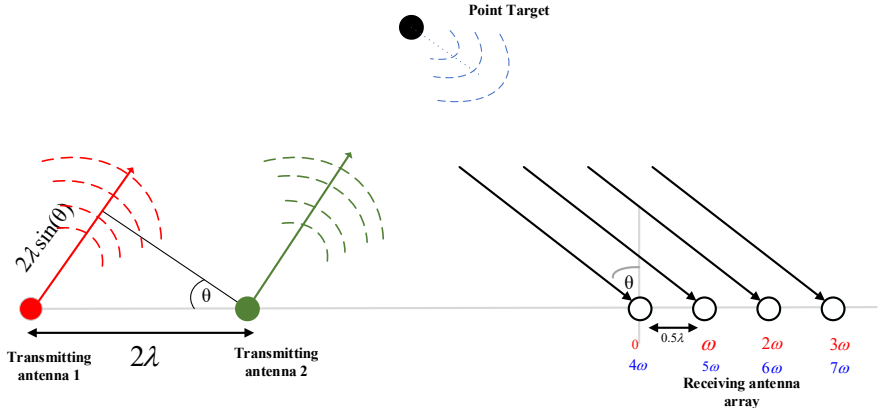
Consider a MIMO radar system with N_T transmitting antennas and N_R receiving antennas. The ith transmitting antenna element radiates a discrete-time baseband waveform $\bar{\mathbf{f}}_i \in \mathbb{C}^{L_s}$, where L_s is the number of samples within the pulse width. The receiving antenna array is a filled ULA[4] and the transmitting antenna array has an inter-element spacing of

[4]A filled phased array has its elements placed with half-wavelength spacing between each consecutive

$N_R\lambda/2$. When the waveforms $\bar{\mathbf{f}}_i, i = 1, \cdots, N_T$ are orthogonal, the MIMO radar has a virtual filled ULA of $N_T N_R$ elements. The idea of the virtual array in MIMO radar is illustrated with the aid of Fig. 2.6.



(a) A Radar with $N_T = 1$ and $N_R = 8$



(b) A MIMO Radar with $N_T = 2$ and $N_R = 4$

Figure 2.6 The virtual array of MIMO radar.

In Fig. 2.6a, we depict a radar system consisting of one transmitting antenna and eight receiving antennas with inter-element spacing of $\lambda/2$. The transmitted signal from the transmitting antenna results in phase shifts of $\omega, \cdots, 7\omega$ at the receiving antennas, with the first antenna element as the reference. Using two transmitting antennas and four receiving antennas, as shown in Fig. 2.6b, results in the same phase shift sequence at the receiving antennas. The signal transmitted from the first transmitter results in phase shifts $0, \omega, 2\omega, 3\omega$. Since the second transmitting antenna is separated from the first one by four times the

elements [17].

separation between the receiving antennas, its transmitted signal arrives the receiver with an additional 4ω phase shift. Therefore, the transmitted signal from the second antenna element results in a phase-shift sequence of $4\omega, 5\omega, 6\omega, 7\omega$. Combining the two phase shifts resulting from the first and second transmitting antenna elements, we obtain the same sequence of phase shifts obtained by the radar configuration in Fig. 2.6a. In general, using the proper placement of N_T transmitting and N_R receiving antennas, a virtual array of $N_T N_R$ antennas is synthesized at the receiver.

The advantages of the MIMO radars over the phased array radars include higher spatial resolution, better parameter identification, improved performance for ground moving target identification (GMTI) [18, Ch.2], and enhanced detection performance due to their spatial diversity [19].

The problem of waveform design of MIMO radars has attracted a wide interest in the last decade, which resulted in the following main trends in the design of MIMO radar waveforms [20]:

1. To consider only the covariance matrix of the waveforms instead of the entire waveform, to control the spatial distribution of the transmitted power. However, this design method covers the spatial domain only.

2. Waveform design based on the optimization of the ambiguity waveform properties such as the autocorrelation peak to sidelobes level ratio (PSLR), the cross-correlation between the waveforms, the Doppler and range resolutions, and Doppler tolerance.

3. In his seminal book [21], Woodward employed, for the first time, the information theory in the design of radar receivers. After three decades, this was followed by the work in [22], where it was shown that the radar performance is enhanced by maximizing the conditional mutual information between the target and the radar reflected signal. This category of waveform design is concerned with the extended target model[5]. In [23], it has been shown that maximizing the MI between the random target impulse response and the reflected radar signal is equivalent to minimizing the value of the minimum mean-square error (MMSE) of the target impulse response estimation.

[5]Extended targets are those targets that occupy more than one range cell. While this abridged definition is sufficient for the purpose of the discussion here, more details about extended targets will be provided in Chapters 4 and 5.

In Chapter 4, we delve more into the problem of waveform design of the third trend in the aforementioned waveform design trends in cognitive MIMO radars.

2.4 Cognitive Radars

Adaptive radars involve adjusting the receiver to improve different aspects of radar performance. This adjustment includes setting the detection threshold automatically according to the environment through employing adaptive detectors and antenna arrays. The latter led to devising the STAP that adaptively filters the signal in both Doppler and spatial domains. The more advanced cognitive radars extend the concept of adaptation to the radar transmitter as well as the receiver [24]. Since the cognitive radar is still in the research and development phase, there is no unique, formal definition on what constitutes a cognitive radar. However, in the following, we describe briefly the distinguishing features of the cognitive radar over conventional radars.

While the concept of cognitive radars can be rooted back to the work on knowledge-based radar in the late nineties [25–27], the first formulation of the cognitive radar framework was introduced by Haykin in [28]. Haykin outlined three main elements of the cognitive radar that distinguish it from the adaptive radar:

1. The transmitter, receiver, and the environment form a dynamic closed-loop system as shown in Fig. 2.7.

2. The radar system continuously *learns* from the environment through the received observations and the obtained information is used to adapt the receiver.

3. The transmitted waveform is also adapted according to the acquired information about the environment and the target parameters.

The knowledge-based radars use prior knowledge of the environment to improve the performance by employing the available environment database to choose the optimum signal processing approach [29]. Therefore, the knowledge-based radars can be seen as employing "inside-out" information, in which the prior knowledge, which can be considered as an integral part of the receiver, is used to improve the radar performance. Conversely, cognitive radars use "outside-in" information, which is gathered online by the radar from the environment [30]. One of the most important information obtained by the cognitive radars is the

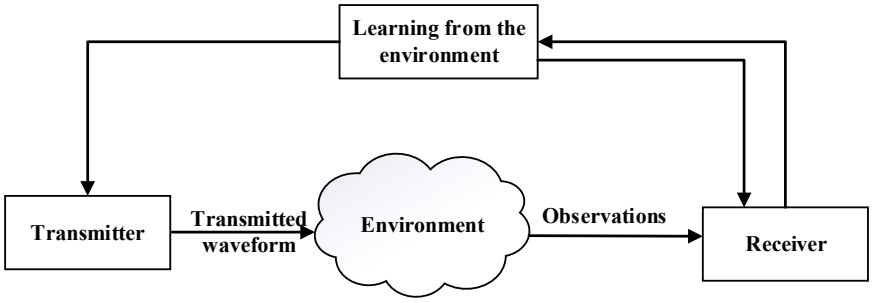


Figure 2.7 Cognitive radar work flow.

target impulse response (TIR), which is used to optimize the radar waveform as we discuss in more details in Chapter 5.

2.5 Radar Detection

The final stage in the signal processing chain of the radar receiver for all the radar types discussed previously including phased array, MIMO, cognitive radars. Radar detection means the ability of its receiver to decide whether a target is present or not in the presence of noise, environment clutter[6], and jamming. Radar detection is a binary hypothesis testing problem with the null hypothesis H_0 representing that no target is present and the alternative hypothesis H_1 corresponding to the target is present. This binary-hypothesis testing problem reduces to the likelihood ratio test

$$\Lambda = \frac{f_{R|H_1}(r|H_1)}{f_{R|H_0}(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{2.19}$$

where r is the observation, γ is the threshold, $f_{R/H_1}(r/H_1)$ and $f_{R/H_0}(r/H_0)$ are the conditional probability density functions (PDF) of r under H_1 and H_0 respectively. The most appropriate criterion to obtain the threshold is the Neyman Pearson criterion that maximizes the probability of detection P_D at a fixed probability of false alarm P_{fa}, which determines the probability that H_1 is decided while H_0 is true. The value of the threshold is calculated to achieve the required P_{fa} at a given level of interference. False alarms are generated due to different sources of interference as clutter, high noise power, or jamming.

[6]Radar clutter is defined as "unwanted echoes, typically from the ground, sea, rain or other precipitation, chaff, birds, insects, or aurora." [31]

To lower P_{fa}, the threshold γ should be raised, but this leads to a lower P_D at lower noise or clutter power levels than those at which P_{fa} has been calculated. To resolve this, an adaptive threshold can be used, to maintain a fixed P_D in different clutter and noise environments [9, 32] among other solutions. The detector that maximizes the probability of detection at a fixed level false alarm rate is the constant false alarm rate (CFAR) detector.

The basic assumption of adaptive threshold detectors is that the PDF of the interference is known except for the variance σ^2 or the covariance matrix $\mathbf{R}$ in the case of vector detectors. The presence of an unknown parameter in the detection problem raises the need for the generalized likelihood ratio test (GLRT) that is formed by estimating the unknown parameter (the variance or the covariance) and substituting this estimate into the likelihood ratio test. If the probability of false alarm does not depend on this estimated parameter, a GLRT is possible [32, 33].

There are mainly two types of CFAR radar detectors based on the dimension of the received radar signal: scalar and vector CFAR detectors, which will be described briefly below.

2.5.1 Scalar CFAR

In the scalar CFAR detectors, as shown in Fig. 2.8, the secondary cells along with the primary cell (the Cell Under Test (CUT) are complex scalars in time or frequency domain. The threshold is calculated from the secondary cells, after excluding the guard cells [34], and then it is compared to the CUT to decide about the target presence. The basic form CFAR detector is the cell average CFAR (CA-CFAR) detector, whose threshold depends on the average of the surrounding cells (reference or secondary cells) of the CUT. The secondary cells are assumed to be independent and identically distributed (iid).

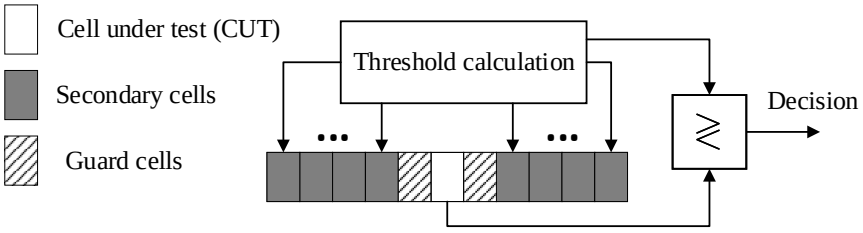


Figure 2.8 Block diagram of CA-CFAR

If the distribution of the secondary cells and CUT is Gaussian, the CA-CFAR detector

is the optimum detector. however, the performance of CA-CFAR detector is degraded when clutter or jamming are present in the secondary cells. That is why other types of CFAR detectors have been proposed such as the Greatest of CFAR (GO-CFAR) [34], order statistic CFAR (OS-CFAR) [35], and censored CFAR [36].

2.5.2 Vector CFAR

While scalar CFAR detectors deal with the received signal that is represented in one dimension only, i.e., frequency (Doppler) or time (range), vector CFAR detectors handle multidimensional signals. For the three-dimensional case, the received signal represents the target in range (fast time), temporal (Doppler or slow time), and spatial (angle) domains where the concept of "data cube" is used. The detector scans the signal in one dimension (fast time) and vectorizes the remaining 2-D matrix into an array.

As indicated before, the received signal matrix is stacked into an $MN \times 1$ single column vector. In light of the well-known RMB procedures[7], Kelly in [37] has formulated the following likelihood ratio test (LRT)

$$\Lambda_{1,0} = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{(\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}[1 + \frac{1}{K}(\mathbf{z}^H \hat{\mathbf{R}}^{-1}\mathbf{z})])} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{2.20}$$

where H_0 and H_1 are the null and alternative hypotheses denoting the target absence or presence, respectively, and η is a threshold that is determined based on the required probability of false alarm P_{fa} according to the Neyman-Pearson criteria [38]. The covariance matrix can be estimated from the range cells surrounding the CUT (also called primary data) under the assumption that the surrounding cells (also called secondary cells) are homogeneous and free of targets. Both Kelly and Reed, at the same time, simplified the LRT in Eq. (2.20) to [39, 40]

$$\Lambda_{1,0} = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{\mathbf{s}^H \hat{R}^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.21}$$

This detector is known as the adaptive matched filter (AMF) detector. To improve the CFAR property of the detector, a normalized version of this detector is the normalized adaptive

[7] RMB are the initials of the authors of [16]

matched filter (NAMF)

$$\Lambda_{\text{NAMF}} = \frac{\left|\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{z}\right|^2}{\left|\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}\right| \left|\mathbf{z}^H \hat{\mathbf{R}}^{-1}\mathbf{z}\right|} \tag{2.22}$$

2.6 Mathematical Background

In this section, we briefly describe the basic mathematical tools used in the thesis.

2.6.1 Depth functions

To identify outliers in a data set associated with a cumulative distribution function (CDF) F defined on $\mathbb{R}$, the data points are compared to a threshold, e.g., one of the quantiles of F, or a function of it. The outliers are identified by those data points whose values exceed the threshold. Given a sample data set, the sample quantiles are obtained by applying linear ordering to the data points, which induces a ranked or ordered data set. However, applying a similar procedure on multidimensional data defined on $\mathbb{R}^d$, with $d > 2$, is cumbersome, since the concept of ranking is not defined for multidimensional data. Alternatively, employing a *center* for the multivariate data using the mean or the median, the concept of *center-outward* ordering can be applied to the multivariate data instead of the linear ordering [41]. Based on this ordering, the *depth* of each point relative to the center is used to identify outliers with the center as the deepest point.

Definition 2.1 (Depth function) [42]: Let the function $D(\mathbf{x}; F_{\mathbf{x}}) : \mathbb{R}^d \to \mathbb{R}$ of a random vector $\mathbf{x} \in \mathbb{R}^d$ and its CDF be $F_{\mathbf{x}}$. If $D(x; F_{\mathbf{x}})$ satisfies the following:

(a) $D(\mathbf{A}\mathbf{x}+\mathbf{b}; F_{\mathbf{A}\mathbf{x}+\mathbf{b}}) = D(\mathbf{x}; F_{\mathbf{x}})$ for a non-singular $d \times d$ matrix $\mathbf{A}$ and any $d-$dimensional vector $\mathbf{b}$. In other words, $D(\mathbf{x}; F_{\mathbf{x}})$ is affine invariant.

(b) $D(\bar{\mathbf{x}}; F_{\mathbf{x}}) = \sup_{\mathbf{x}\in\mathbb{R}^d} D(\mathbf{x}; F_{\mathbf{x}})$, where $\bar{\mathbf{x}}$ is the center of $F_{\mathbf{x}}$.

(c) $D(\mathbf{x}; F_{\mathbf{x}}) \to 0$ as $\|\mathbf{x}\| \to \infty$.

Then $D(\mathbf{x}; F_{\mathbf{x}})$ is a statistical depth function. There are four main approaches in constructing depth functions: weighted mean depth functions, depth functions based on halfspaces, spatial depth function, and distance based depth functions [43, 44].

Weighted mean-based depth functions are defined by the so-called weighted-mean regions that are convex sets, whose support functions are weighted means of order statistics [43]. For the vectors $\mathbf{x}^1, \cdots, \mathbf{x}^L$ and a vector $\mathbf{u} \in \mathbb{R}^d$, a linear ranking can be obtained by projecting the data vectors as follows

$$\mathbf{u}^T\mathbf{x}^{\mathbf{p}(1)} \leq \mathbf{u}^T\mathbf{x}^{\mathbf{p}(2)} < \cdots < \mathbf{u}^T\mathbf{x}^{\mathbf{p}(L)} \tag{2.23}$$

where $\mathbf{p}$ is a permutation of the vectors' indices. Let $w_{i,\alpha}$, $i = 1, \cdots, L$ and $\alpha \in [0, 1]$, be scalar weights, where $\sum_{i=1}^{L} w_{i,\alpha} = 1$, then the weighted-mean (WM) depth is defined as [43]

$$D^{WM}(\mathbf{x}; F_{\mathbf{x}}) = \sum_{i=1}^{L} w_{i,\alpha}\mathbf{u}^T\mathbf{x}^{\mathbf{p}(i)}, \tag{2.24}$$

Different weights result in different notions of data depths. For instance, one of the known statistical depth functions is the zonoid regions, whose weights are given by [43]

$$w_{i,\alpha} = \begin{cases} 0, & \text{if } \; i < L - \lfloor L\alpha \rfloor, \\ \frac{L\alpha - \lfloor L\alpha \rfloor}{L\alpha}, & \text{if } \; i = L - \lfloor L\alpha \rfloor, \\ \frac{1}{L\alpha}, & \text{if } \; i > L - \lfloor L\alpha \rfloor \end{cases} \tag{2.25}$$

However, the Zonoid depth function, which is the most widely used weighted mean depth function, has a higher complexity compared to other depth functions [45].

Depth functions based on halfspaces do not use a metric on $\mathbb{R}^d$; instead they use closed halfspaces. The most famous form of the halfspace depth function is the location depth, also known as Tukey depth, whose population version is defined as [44]

$$D^{Tukey}(\mathbf{x}; F_{\mathbf{x}}) = \inf_{\mathbf{H}} \{F(\mathbf{H}) : \mathbf{H} \text{ is a closed halfspace }, \mathbf{x} \in \mathbf{H}\} \tag{2.26}$$

However, Tukey depth is not informative in the case of high dimensional data, i.e., $d > L$ [46].

Spatial depth functions are based on the spatial quantiles [47]. The spatial median $\breve{\mathbf{x}}$ is

the solution to the following optimization problem[8] [49]

$$\min_{\check{\mathbf{x}}\in\mathbb{R}^d} \quad \sum_{i=1}^{L} \|\check{\mathbf{x}} - \mathbf{x}_i\|_2 \tag{2.27}$$

The above problem is solved by setting the derivative of its objective function with respect to $\check{\mathbf{x}}$ to 0 to obtain

$$\sum_{i=1}^{L} \xi(\check{\mathbf{x}} - \mathbf{x}_i) = 0 \tag{2.28}$$

where

$$\xi(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|_2}, & \mathbf{x} \neq 0 \\ \mathbf{0}, & \mathbf{x} = 0 \end{cases} \tag{2.29}$$

The spatial depth function is given by [50]

$$D^S(\mathbf{x}; F_{\mathbf{x}}) = 1 - \left\| \int \xi(\mathbf{y} - \mathbf{x}) dF(\mathbf{y}) \right\|_2 \tag{2.30}$$

While the spatial depth function has various desirable properties, such as robustness to give an example, its computation depends on the sample size L rather than the data dimension d [50]. In radar applications, L, i.e., the number of secondary cells, is often larger than d, leading to a high computational cost of the spatial depth function.

A distance-based depth function uses the distance from the center as a measure of depth. One of the first and most famous distance-based functions is the Mahalanobis depth, whose sample version is defined as

$$D^{MH}(\mathbf{x}; F_{\mathbf{x}}) = \left(1 + (\mathbf{x} - \bar{\mathbf{x}})^T \hat{\mathbf{\Sigma}}_X^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right)^{-1} \tag{2.31}$$

where $\hat{\mathbf{\Sigma}}$ is the estimated covariance matrix of $\mathbf{x}$ and $\bar{\mathbf{x}}$ is its sample mean. Another distance-

[8] The definition of the spatial median given in Eq. (2.27) is equivalent to that of the median in the univariate case [48].

based depth function is the projection-depth function, which is defined as [43]

$$D^{Proj}(\mathbf{x}; F_{\mathbf{x}}) = \left(1 + \sup_{\|\mathbf{u}\|=1} \frac{\left|\mathbf{u}^T\mathbf{x} - \text{Med}(\mathbf{u}^T\mathbf{X})\right|}{\text{MAD}(\mathbf{u}^T\mathbf{X})}\right)^{-1} \tag{2.32}$$

where $\mathbf{X} \in \mathbb{R}^{d\times L}$ is the sample of $\mathbf{x}$ of size L, $\mathbf{u} \in \mathbb{R}^d$, Med denotes the median, and MAD denotes the median absolute deviation. It is noteworthy that for a one-dimensional data set $X = \{X_1, X_2, \cdots, X_L\}$ the rule $\frac{|X_i - \text{Med}(X)|}{\text{MAD}(X)}, i = 1, \cdots, L$ has been widely used as a robust measure to detect outliers [51, 52]. Among other benefits, the projection depth function requires the simplest computations compared to other types of depth functions [53, 54]. In Chapter 3, we employ the projection depth function in the problem of detecting non-homogeneous secondary cells for a more robust estimation of the covariance matrix.

2.6.2 Proximal Optimization

Consider the following optimization problem [55]

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) + g(\mathbf{x}) \tag{2.33}$$

where $f(\mathbf{x})$ is a smooth function, possibly non-convex, and $g(\mathbf{x})$ is a convex function, possibly non-smooth. The form of the problem in Eq. (2.33) is encountered in many applications of signal processing and machine learning, where $f(\mathbf{x})$ is an objective function that is dependent on some observation and $g(\mathbf{x})$ is a regularization term that imposes some favorable properties on the solution [56]. The difficulty in solving Eq. (2.33) arises from the fact that $g(\mathbf{x})$ can be non-differentiable, which impedes the solution using conventional convex optimization methods. One approach to solve such problems is to split the objective function of Eq. (2.33), which leads to efficient solution algorithms that are known as *proximal algorithms* [57]. Proximal algorithms can solve the problems of the form of Eq. (2.33) with non-smooth $g(\mathbf{x})$ if its proximal operator can be calculated, which explains the name "proximal algorithms".

Definition 2.2 (Proximal operator) [57]: Let $g(\mathbf{x})$ be a convex function and $\mathbf{x} \in \mathbb{R}^d$, $d > 2$. The minimization problem

$$\min_{\mathbf{y}\in\mathbb{R}^d} \quad g(\mathbf{y}) + \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.34}$$

admits a unique solution that is known as the proximal operator of g and denoted as $\text{prox}_g(\mathbf{x})$. Intuitively, $\text{prox}_g(\mathbf{x})$ minimizes Eq. (2.34) with a constraint on the distance from $\mathbf{x}$. The proximal operator is closely related to the Moreau envelope $M_g(\mathbf{x})$, which is defined as [56]

$$M_g(\mathbf{x}) = \inf_{\mathbf{y}} \left\{ g(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 \right\} \tag{2.35}$$

The Moreau envelope can be viewed as a regularized version of g [56]. The proximal operator and the Moreau envelope are related as [58]

$$M_g(\mathbf{x}) = g(\text{prox}_g(\mathbf{x})) + \frac{1}{2}\|\mathbf{x}-\text{prox}_g(\mathbf{x})\|_2^2 \tag{2.36}$$

For a scaled version of $g(\mathbf{x})$, we have

$$\nabla M_{\lambda g}(\mathbf{x}) = \frac{1}{\lambda}(\mathbf{x}-\text{prox}_{\lambda g}(\mathbf{x})) \tag{2.37}$$

where $\lambda > 0$. Eq. (2.37) can be rewritten as

$$\text{prox}_{\lambda g}(\mathbf{x}) = \mathbf{x} - \lambda \nabla M_{\lambda g}(\mathbf{x}) \tag{2.38}$$

Therefore, the proximal operator can be considered as a gradient step to minimize $M_{\lambda g}(\mathbf{x})$, and equivalently $g(\mathbf{x})$, with a step size λ.

The basic optimization algorithm based on the proximal operator is the proximal point optimization. This algorithm solves the minimization of the convex and possibly non-smooth function $\lambda g(\mathbf{x})$ and its solution is the proximal operator itself such that $\mathbf{x}^{k+1} = \text{prox}_{\lambda g}(\mathbf{x}^k)$. For the solution of problems in the form of Eq. (2.33), the proximal gradient method is applied. Using the gradient proximal algorithm, which is an iterative method. The kth iteration is

$$\mathbf{x}^{k+1} = \text{prox}_{\lambda^k g}(\mathbf{x}^k - \lambda^k \nabla f \mathbf{x}^k) \tag{2.39}$$

where $\lambda^k > 0$ is the step size and the solution is obtained as $k \to \infty$. The proximal gradient method reduces to the proximal point method when $f(\mathbf{x}) = 0$. When $g(\mathbf{x}) = 0$, the proximal gradient method is the conventional gradient descent method. In Chapter 5, the proximal gradient algorithm is used to design power-efficient cognitive MIMO radar waveforms.

2.6.3 Dynamic Bayesian graphical models

Graphical models combine the use of both graph and probability theories. From the graph theory, the graphical models inherit the ability of modeling system modularity, i.e., simplifying the system into a number of connected parts. The probability theory is used to define the connections, specifically the probabilistic relations, among those connected parts [59]. In dynamic (or dynamical) Bayesian graphical models the nodes represent random variables, whose dependencies are represented by the arcs between the nodes. As shown in Fig. 2.9, the dependencies among the random variables A, B, C, D are described by the arcs between the pairs of the variables, where the independent variables are not connected with arcs. It should be noted that the use of the word "dynamic" means that the graph models are used to describe dynamic systems[9] and it does not mean that the model changes over time [61].

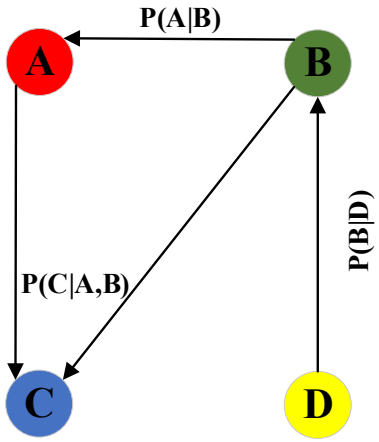


Figure 2.9 Example of a probabilistic graphical model

The main trend in the literature is to differentiate between two main classes of Bayesian graphical models that are used widely in different applications: state space and hidden Markov models. Both models, as will be detailed shortly, embody hidden states from the observer; however, according to some authors [62], the state space model is defined with continuous states while the hidden Markov model (HMM) assume discrete states. However, some authors, like Murphy in [61], considers HMM as a type of state space model. While we adopt the mainstream in the literature, the following descriptions show that the state space models and HMM are strongly related.

[9] A dynamic system is one whose states are changing over time according to a family of transformations that are parameterized by time [60]

2.6.3.1 State space models

In the state space model, the observations and the hidden states are expressed as

$$\mathbf{y}(t) = v(\mathbf{x}(t)) + \mathbf{n}_m(t) \tag{2.40a}$$

$$\mathbf{x}(t) = u(\mathbf{x}(t-1)) + \mathbf{n}_s(t) \tag{2.40b}$$

where $\mathbf{x}(t) \in \mathbb{C}^d$ is the hidden state of the system at time t, $\mathbf{y}(t) \in \mathbb{C}^d$ is the observation vector, $\mathbf{n}_m(t), \mathbf{n}_s(t)$ are the independent observation and state noise vectors, respectively, and $v, u : \mathbb{C}^d \to \mathbb{C}^d$ are linear or nonlinear functions, assumed to be static, i.e., do not change with time. Without loss of generality, it is assumed here that both the observation and state vectors have the same dimension. It is customary to call Eq. (2.40a) as the measurement equation and Eq. (2.40b) as the state or plant equation. As observed in Eq. (2.40b), the current state is assumed to depend only on the previous state which is known as the first-order Markov chain. A widely used model is the linear state space model [62]

$$\mathbf{y}(t) = \mathbf{V}\mathbf{x}(t) + \mathbf{n}^s(t) \tag{2.41a}$$

$$\mathbf{x}(t) = \mathbf{U}\mathbf{x}(t-1) + \mathbf{n}^m(t) \tag{2.41b}$$

where $\mathbf{V} \in \mathbb{R}^{d\times d}$ is the design or observation matrix and $\mathbf{U} \in \mathbb{R}^{d\times d}$ is the transition matrix; both of which are assumed to be constant. Bayesian inference is concerned with finding the posterior probability density function (pdf) of the states given the observation. Specifically, Bayesian filters are used to recursively, i.e., in a sequential manner, estimate the posterior pdf from the observations [63]. Before delving into the concept and application of Bayesian filter, it is important to discuss first the different types of the state-space models. As Fig. 2.10 depicts, there are four main types of the state space model based on the linearity or nonlinearity of the functions v, u and the distribution of $\mathbf{n}^m(t)$ and $\mathbf{n}^s(t)$. The type of Bayesian filter realized to solve a problem of the form of a state-space model is determined based on the properties of this model, which is summarized as follows:

(a) **Linear Gaussian model:** The Bayesian filter is realized exactly through the Kalman filter, which is also, under the linear Gaussian model, the optimal filter based on the mean square error (MSE) criterion.

(b) **Nonlinear Gaussian model:** The posterior pdf of the states is directly and locally

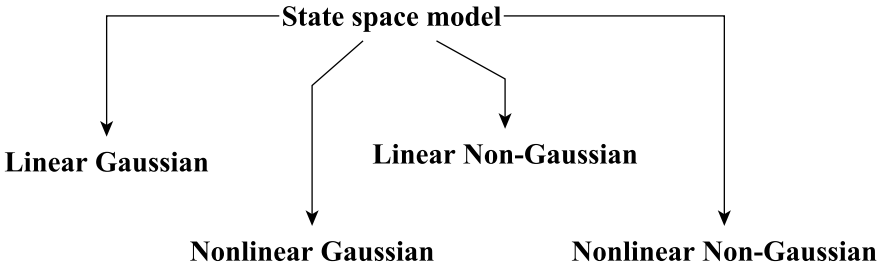


Figure 2.10 The different types of state space models

approximated around the filtered estimate of the tth state given all the observations up to t [63]. Examples of this approach include the extended, cubature, unscented, and decoupled Kalman filters [63]. The other approach of approximating the posterior pdf of the states is by using indirect approximation through sampling the posterior from a set of randomly chosen samples (particles) with associated weights, which is the approach adopted by the particle filter [64].

(c) **Linear non-Gaussian model:** In this case the non-Gaussian distribution of the measurement and/or the state noise vectors are approximated by a Gaussian mixture to form what is known as the mixture Kalman filter [65]. Moreover, the indirect approximation of the posterior, i.e., the particle filter can be used for this model.

(d) **Nonlinear non-Gaussian model:** For this case, the particle filter is the only available approximation for the Bayesian filter [63].

For the sake of succinctness, brief descriptions of the Kalman and particle filter, which are employed as benchmarks for the TIR estimation method proposed in Chapter 5, are provided here.

Kalman filter

In practical applications, the estimation is performed in discrete time at time instants $t_1, \cdots, t_k$, which leads to the discrete Kalman filter. For simplicity of notation, we use $1, \cdots, k$ to denote the time instants $t_1, \cdots, t_k$. Let $\mathbf{x}_k^-$ be the *a priori* estimate of the state at time instant k, while $\hat{\mathbf{x}}_k$ be its *a posteriori* estimate given a measurement $\mathbf{y}_k$. Therefore,

the *a priori* and *a posteriori* error covariance matrices are defined by [66]

$$\mathbf{P}_k^- = E[(\mathbf{x}_k - \mathbf{x}_k^-)(\mathbf{x}_k - \mathbf{x}_k^-)^H] \tag{2.42a}$$
$$\hat{\mathbf{P}}_k = E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^H] \tag{2.42b}$$

The *a posteriori* and *a priori* state estimates are related as

$$\hat{\mathbf{x}}_k = \mathbf{x}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{V}\mathbf{x}_k^-) \tag{2.43}$$

where $\mathbf{K}_k$ is the Kalman gain or blending factor. Eq. (2.43) is known as the measurement innovation, which reflects the discrepancy between the actual and the expected observations. The Kalman filter operates in two main steps: the ***prediction step*** that involves estimating the current state *a priori* and the ***update step*** in which the estimated state is adjusted by the actual measurement. It is assumed that the state noise vector $\mathbf{n}_k^s \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ and observation noise vector $\mathbf{n}_k^m \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, where $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Both $\mathbf{Q}$ and $\mathbf{R}$ are assumed to be known and they do not change with the time index k. The two steps of the Kalman filter are summarized below.

Prediction step:

$$\mathbf{x}_k^- = \mathbf{U}\hat{\mathbf{x}}_{k-1} \tag{2.44a}$$
$$\mathbf{P}_k^- = \mathbf{U}\hat{\mathbf{P}}_k\mathbf{U}^H + \mathbf{Q} \tag{2.44b}$$

Update step:

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{V}^H(\mathbf{V}\mathbf{P}_k^-\mathbf{V}^H + \mathbf{R})^{-1} \tag{2.45a}$$
$$\hat{\mathbf{x}}_k = \mathbf{x}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{V}\mathbf{x}_k^-) \tag{2.45b}$$
$$\hat{\mathbf{P}}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{V})\mathbf{P}_k^- \tag{2.45c}$$

Particle filter

To fully grasp the idea behind the particle filter, we need to look back to Monte Carlo methods, or more specifically, the sequential Monte Carlo (SMC) methods. Suppose we

want to approximate a multivariate pdf $\pi(x_{1:n})$ using Monte Carlo methods, where $x_{1:n} = x_1, \cdots, x_n$ and $n \geq 2$. We sample N independent random variables, $X^i_{1:n} \sim \pi(x_{1:n})$, such that the approximated measure of $\pi(x_{1:n})$ is

$$\hat{\pi}_n(x_{1:n}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_{1:n}}(x_{1:n}) \tag{2.46}$$

where $\delta_{x_0}(x)$ is the Dirac delta function at x_0. The samples, or the "particles", $X^i_{1:n}$ are obtained from the state equation assuming known state noise distribution [67]. In addition, we can also approximate any marginal pdf, say $\pi(x_k)$ as

$$\hat{\pi}_n(x_k) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_k}(x_k) \tag{2.47}$$

However, it is often difficult to sample directly from the target distribution, $\pi(x_{1:n})$. This problem can be solved using importance sampling, also known as weighted sampling, whose purpose is to sample from a distribution that is different from the original distribution due to the computational advantage of sampling from the former over the latter [68]. In particular, let the original pdf be $\pi(x_{1:n})$, also called the target or nominal pdf, and assume that we have $q(x_{1:n}) \propto \pi(x_{1:n})$ such that $q(x_{1:n}) > 0$ whenever $\pi(x_{1:n}) > 0$. Then[10] [70]

$$\hat{\pi}(x_k) = \frac{1}{N} \sum_{i=1}^{N} w(X^i_{1:n}) \tag{2.48}$$

with

$$w(X^i_{1:n}) \propto \frac{\pi(X^i_{1:n})}{q(X^i_{1:n})} \tag{2.49}$$

where $X^i_{1:n}$ are sampled from $q(x_{1:n})$, known as the importance or proposal pdf, instead of $\pi(x_{1:n})$. Eq. (2.49) implies that we should be able to compute $\pi(X^i_{1:n})$; however, in some cases we are only able to compute an unnormalized version $\pi^u(X^i_{1:n}) = c\pi(X^i_{1:n})$, where $c > 0$ is unknown. The same can be applied to $q(x_{1:n})$ and its unnormalized version $q^u(X^i_{1:n}) = bq(X^i_{1:n})$, where $b > 0$ is also unknown. To overcome this difficulty, the new

[10] Some references state that $q(x_{1:n})$ should share the same support with $\pi(x_{1:n})$; however, it is sufficient that the support of $q(x_{1:n})$ includes that of $\pi(x_{1:n})$ [69].

weights $w^u(X^i_{1:n}) = \pi^u(X^i_{1:n})/q^u(X^i_{1:n})$ can be used instead of $w(X^i_{1:n})$ and Eq. (2.48) can be modified to be

$$\hat{\pi}(x_k) = \sum_{i=1}^{N} \frac{w^u(X^i_{1:n})}{\sum_{i=1}^{N} w^u(X^i_{1:n})} = \sum_{i=1}^{N} \tilde{w}(X^i_{1:n}) \tag{2.50}$$

where $\frac{1}{N}\sum_{i=1}^{N} w^u(X^i_{1:n})$ is a normalizing constant, $\tilde{w}(X^i_{1:n})$ are known as the self-normalized importance weights, and the ratio c/b cancels out. It is worthy to emphasize that importance sampling can be used to reduce the variance of the pdf estimation by concentrating the sampled points in the regions that are more *"important"* in the target distribution instead of sampling equally from all the regions [71].

In practice, the sampling is performed sequentially by choosing the importance density such that [64]

$$q(x_{1:n}) = q(x_n)q(x_{1:n-1}) = q(x_n)\prod_{k=2}^{n} q(x_k) \tag{2.51}$$

and the weights are given by [70]

$$w(X^i_{1:n}) = w(X^i_1)\prod_{k=2}^{n} \alpha(X^i_{1:k}) \tag{2.52}$$

where $\alpha(X^i_{1:k}) = \pi(X^i_{1:k})/\pi(X^i_{1:k-1})q(X^i_k)$. After calculating the weights, the estimated state $\hat{x}_{1:n}$ is obtained from the particles using different schemes. The classic approach is to pick the estimated state from the particles according to $P(x_{1:n} = X^i_{1:n}) = w(X^i_{1:n})$ as initially proposed in [67]. Other approaches are also possible, for instance the weighted mean, that is $\hat{x}_{1:n} = \sum_{i=1}^{N} w(X^i_{1:n})X^i_{1:n}$, or the best particle $\hat{x}^i_{1:n} = \text{argmax} \quad w(X^i_{1:n})$ [72].

The aforementioned sampling scheme is known as the sequential importance sampling (SIS). A common problem with the SIS is that after some iterations all the particles except one will have negligible weights, which is known as the *degeneracy problem*. To detect the degeneracy problem, the number of effective particles is calculated as [64]

$$N_{eff} = \frac{1}{\sum_{i=1}^{N}(w(X^i_k))^2} \tag{2.53}$$

A small N_{eff} means degeneracy. When detected, the problem of degeneracy can be solved through the appropriate choice of the importance density and resampling. The latter involves sampling N independent and identically distributed (i.i.d) particles with equal weights, $1/N$ [64].

2.6.3.2 Hidden Markov models

The hidden Markov model (HMM) embeds two stochastic processes: a *"hidden"* stochastic process that is not observed but it can be inferred from the second stochastic process that produces a sequence of observations [73]. The basic building block of the HMM is a Markov chain, which describes the evolution of the "states", each of which can takes values from a discrete set $\mathscr{X} = \{X_k | k \in \mathbb{N}\}$. The main characteristic of the Markov chain is that the current state X_k is independent from all previous states given X_{k-1} [74], i.e.,

$$P(X_k|\mathscr{X}) = P(X_k|X_{k-1}) \tag{2.54}$$

which is known as the "*Markov property*" or Markov assumption. In the HMM, the evolution of the *hidden* states is governed by a Markov chain and an observation is generated depending on the current state. Specifically, the HMM is determined by the following five elements [73]:

1. The set of the model's hidden states $\mathsf{S} = \{S_1, \cdots, S_{N_s}\}$, where N_s is its cardinality.

2. The model's observation set $\mathsf{O} = \{O_1, \cdots, O_{N_s}\}$, where N_o is its cardinality.

3. The transition matrix $\mathbf{A}$, whose (i, j)th element is defined as

$$\mathbf{A} = [a_{ij}] = P(X_k = S_i|X_{k-1} = S_j), \qquad 1 \leq i, j \leq N_s \tag{2.55}$$

   where $\mathbf{A}$ is a stochastic matrix such that

$$a_{ij} \geq 0, \quad \sum_{j=1}^{N_s} a_{ij} = 1 \tag{2.56}$$

4. The observation matrix $\mathbf{B}$, whose (i, j)th element is defined as

$$\mathbf{B} = [b_{ij}] = P(Y_k = O_i|X_k = S_j) \qquad 1 \leq j \leq N_o, 1 \leq i \leq N_s \tag{2.57}$$

where Y_k is the observation at the kth time instant and $\mathbf{B}$ is a stochastic matrix that admits the form of Eq. (2.56).

5. The initial distribution $\boldsymbol{\zeta} = [\zeta_i]$, whose ith element is defined as

$$\zeta_i = P(X_1 = S_i), \qquad 1 \leq i \leq N_s \tag{2.58}$$

It is customary to define the HMM using $\vartheta = (\mathbf{A}, \mathbf{B}, \boldsymbol{\zeta})$. The research on the HMM is concentrated on three problems: (a) the estimation of the observation $Y_{k+1}, \cdots, Y_T$ for a time duration T given ϑ and $Y_1, \cdots, Y_k$, (b) the inference of the states $X_1, \cdots, X_k$ given the observations $Y_1, \cdots, Y_k$ and the model ϑ, and (c) how to adjust the model parameters ϑ to maximize $P(Y_1, \cdots, Y_k|\vartheta)$. Dynamic programming algorithms are employed to solve the first two problems [73], while the adjustment of the model parameters is performed through model training [75]. In Chapter 5 we deal with the HMM from a different stand point, from which we propose a new formulation of the TIR estimation problem based on the HMM assuming an uncountable number of states.

Chapter 3

Covariance-Free Nonparametric Nonhomogeneity Detector

In this chapter, we consider the problem of detecting outliers in the secondary cells used to estimate the covariance matrix of the interference, which is an essential requirement for target detection. In this regard, we propose a novel detector based on robust statistics, which provides both robust performance and fast computations.

3.1 Introduction

Upon reflection of the transmitted pulses by a target, the radar antenna receives distorted versions of these pulses due to other scatterers, clutter, and noise. A space time adaptive processing (STAP) detector discretely scans the range dimension and, for each range bin, arranges the data along the angle and Doppler dimensions into a vector, called a range cell. It then linearly combines the spatio-temporal data in each range cell to form the test statistics. To this end, it needs to compute a set of weight vectors corresponding to the different spatio-temporal "look" directions, which depend on the covariance matrix of the background clutter and noise within the cell under test (CUT), also called the primary cell [11]. However, this covariance matrix is not known in practice and it is commonly estimated from the adjacent range cells, known as the secondary or training cells in this context.

The estimation of the covariance matrix from the secondary cells relies on the assumption that they are homogeneous, i.e., independent and identically distributed (iid). In reality, the homogeneity assumption is hardly met due to the presence of discrete scatterers, in-band

interferers, target-dependent jammers [76, 77], or a combination thereof. In this case, the estimated covariance matrix does not represent accurately the background clutter and noise, and hence, the weight vectors computed from this matrix lead to a performance degradation of the STAP detector. To tackle this problem, the non-homogeneity detector (NHD) was introduced to detect the anomalous secondary cells to be censored from covariance matrix estimation [78].

Conceptually, a secondary cell is considered to be homogeneous to its surrounding secondary cells if it shares with them the same covariance matrix up to a scalar. Since the true covariance matrix of a given secondary cell is unknown, the work in [78] used the generalized inner product (GIP) test to examine the similarity between this unknown covariance matrix and the test covariance matrix estimated from the surrounding secondary cells. Later, the normalized adaptive matched filter test (NAMF) was used as an NHD with Gaussian and non-Gaussian clutter models in [79], where the NAMF detector was shown to be the most robust NHD.

Recent research efforts on NHD have focused on improving the performance of the aforementioned classical detectors or reducing their complexity. For instance, [80] proposed a soft NHD concept, wherein the covariance matrix of the CUT is calculated using the weighted secondary cells assuming Gaussian distributed clutter. In turn, the calculation of a weight for each secondary cell is formulated as a non-linear optimization problem based on the output of a modified version of the adaptive matched filter (AMF). An iterative approximate maximum-likelihood (ML) approach based on the GIP detector was developed in [2] for estimating the subset of non-homogeneous cells. This approach shows a comparable performance to the iterative original GIP test using the ML covariance estimator for the Gaussian interference [81]. In addition, a large body of research has been devoted to reduce the dimensions of the STAP detection problem using different transformation and rank reduction techniques as in [82], which can also be applied to NHD [83]. However, these partially adaptive detectors generally exhibit inferior performance compared to their fully adaptive counterparts. Based on the GIP detector, other NHD procedures for the special cases of spaceborne or side-looking radars were introduced in [84, 85].

The above referenced covariance-based NHDs share the need to estimate the covariance matrix and its inverse (known as the precision matrix) for each secondary cell, which leads to a high computational cost, especially for non-Gaussian clutter. Some covariance estimators need *a priori* knowledge about the clutter distribution [86], which is imperfect in most

cases, while other estimators need to solve non-convex optimization problems with high computational complexity [3]. To avoid such difficulties, a covariance-free NHD with a comparable performance to the GIP was introduced in [87]. However, it is known that the GIP test is not robust, especially in non-Gaussian clutter scenarios [79].

In this chapter, we are concerned with the problem of detecting the non-homogeneous cells out of the secondary cells in the Gaussian and non-Gaussian clutter distributions. We introduce a novel covariance-free and nonparametric NHD based on projection depth function, a well known tool in robust statistics. This detector provides robust performance, does not require any prior assumptions about the clutter distribution, and, most importantly, does not require estimating the covariance or the precision matrices for each cell, and therefore, it reduces the computational burden significantly. The results show that the proposed test maintains the robust performance of the NAMF test, but inherits the nonparametric framework and simple computations of the projection depth functions.

This chapter is organized as follows: The signal model is introduced in Section 3.2. The non-homogeneity detection problem is presented in Section 3.4. The proposed NHD is introduced in Section 3.5 where its approximate equivalence to the NAMF test is proven. The comparative analysis for the detection performance of the proposed detector with the NAMF test is investigated in Section 3.6. Section 3.7 is a brief conclusion of the results.

3.2 Signal Model

As we introduced in Chapter 2, the total received signal $\mathbf{z}$ is expressed as

$$\mathbf{z} = \mathbf{r} + \mathbf{c} + \mathbf{n} \tag{3.1}$$

and

$$\mathbf{r} = a\mathbf{s} \tag{3.2}$$

where a is an unknown deterministic complex amplitude (i.e., Swerling case 0 [88]), $\mathbf{s}$ is the target steering vector, $\mathbf{c}$ is the clutter vector and $\mathbf{n}$ is the noise vector; $\mathbf{n}$ and $\mathbf{c}$ are assumed to be statistically independent. The noise vector $\mathbf{n}$ is drawn from a complex circular symmetric Gaussian distribution $\mathcal{CN}(\mathbf{0}, \varsigma_n^2\mathbf{I}_J)$ with zero mean and covariance matrix $\varsigma_n^2\mathbf{I}_J$ where ς_n^2 is the noise variance. The clutter vector is modeled as [11]

Let $\mathbf{R} = E(\mathbf{z}\mathbf{z}^H)$ be the covariance matrix of the received signal $\mathbf{z}$ in Eq. (2.11). For

each CUT, STAP aims at forming the optimal beamforming (or weight) vector in real time to maximize the received signal-to-interference-plus-noise ratio (SINR) with respect to $\mathbf{s}$. Under the minimum variance distortionless response (MVDR) criterion, the optimal weight vector takes the form [11]

$$\mathbf{w} = g\mathbf{R}^{-1}\mathbf{s} \tag{3.3}$$

where g is a complex scalar.

For the complex vector $\mathbf{z} = \mathbf{z}_{\mathrm{R}} + j\mathbf{z}_{\mathrm{I}}$, where $\mathbf{z}_{\mathrm{R}} = \Re(\mathbf{z})$ and $\mathbf{z}_{\mathrm{I}} = \Im(\mathbf{z})$, the covariance matrix is expressed as [89]

$$\mathbf{R} = \mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}} + \mathbf{R}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}} + j(\mathbf{R}^{T}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}} - \mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}}) \tag{3.4}$$

where $\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}} = E(\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}^{T})$, $\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}} = E(\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}^{T})$, $\mathbf{R}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}} = E(\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}^{T})$, and $\mathbf{z}$ is a proper complex signal, i.e., $\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}} = -\mathbf{R}^{T}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}}$ and $\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}} = \mathbf{R}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}}$, which is common in the radar context. Moreover, it is customary to assume that the in-phase and quadrature components of $\mathbf{z}$ are independent, i.e., $\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{I}}} = \mathbf{0}$, where $\mathbf{0}$ is $J \times J$ zero matrix [90, 91]. Hence,

$$\mathbf{R} = 2\mathbf{R}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}} = 2\mathbf{R}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}} \tag{3.5}$$

In practice, the covariance matrix $\mathbf{R}$ is unknown and different techniques are used to estimate it from the adjacent $L-1$ secondary cells, assuming no guard cells. In the case of Gaussian clutter, the ML estimator is the sample covariance matrix (SCM) given by:

$$\hat{\mathbf{R}}_{\mathrm{SCM}} = \frac{1}{L-1}\sum_{l=1}^{L-1}\mathbf{z}_l\mathbf{z}_l^{H} \tag{3.6}$$

where $\mathbf{z}_l$ denotes the total received signal in the lth secondary cell, and the condition $L-1 \geq 2J$ is needed to ensure robustness. If $\mathbf{c}$ follows a non-Gaussian distribution, the SCM is neither a consistent nor robust estimator and other estimators should be used. More details on these estimators will be presented shortly.

3.3 Spherical Invariant Random Process (SIRP) Clutter Model

The clutter vector is modeled as [11]

$$\mathbf{c} = \sum_{i=0}^{N_c-1} \kappa_i \mathbf{e}_i \tag{3.7}$$

where N_c is the number of clutter patches, κ_i is the complex amplitude of the ith patch and $\mathbf{e}_i \in \mathbb{C}^J$ is the corresponding steering vector, which admits the form of Eq. (2.9). We are concerned with coherent processing of the received signal vector $\mathbf{z}$ in Eq. (2.11), where both the real and imaginary parts (i.e., in-phase and quadrature components) of each vector entry are considered. In this regard, it is essential to employ a probabilistic model of the clutter vector $\mathbf{c}$ in Eq. (3.7) that takes into account the joint statistics of the real and imaginary parts of all its entries. In particular, for proper clutter modeling, both the spatio-temporal correlation properties and probability density function (PDF) of the clutter envelope should comply with experimental data. Under the SIRP model, which meets these requirements [92], the clutter vector is modeled as a product of two independent components, that is: a zero-mean complex Gaussian vector, known as the speckle component, and a positive random variable, known as the texture component and assumed to vary slowly across range cells. Therefore, the clutter vector in Eq. (3.7) can be represented as

$$\mathbf{c} = v\mathbf{y}, \tag{3.8}$$

where $\mathbf{y} \in \mathbb{C}^J$ follows a complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ with zero mean and covariance matrix $\mathbf{\Sigma}$, and v is a positive random variable. By choosing the proper PDF of the texture component v in the SIRP model Eq. (3.8), denoted as $f_V(v)$ in the sequel, we can obtain different non-Gaussian clutter distributions, also known as the compound Gaussian distributions, while the particular choice $v = 1$ (with probability one) yields the Gaussian clutter model. Moreover, through a suitable choice of the covariance matrix $\mathbf{\Sigma}$ of the Gaussian speckle vector $\mathbf{y}$, the desired spatio-temporal correlation properties can be fulfilled. The PDF of $\mathbf{c}$ can be expressed as [79]:

$$f_{\mathbf{c}}(\mathbf{c}) = (2\pi)^{-J} \det(\mathbf{\Sigma})^{-\frac{1}{2}} h_{2J}(\mathbf{c}^H \mathbf{\Sigma}^{-1} \mathbf{c}) \tag{3.9}$$

where the function $h_{2J}(x)$ is defined as

$$h_{2J}(x) = \int_0^\infty v^{-J} \exp(-\frac{x}{v^2}) f_V(v) dv. \tag{3.10}$$

The covariance matrix of the SIRP vector $\mathbf{c}$ is given by $\mathbf{R_c} = E(v^2)\ \mathbf{\Sigma}$.

Other models for compound Gaussian clutter use zero memory nonlinear (ZMNL) transformations. These methods apply nonlinear transformations on sequences of coherent Gaussian samples that result in the desired marginal PDF of the clutter envelope. However, due to the nonlinear transformations, the covariance matrix of the resulting non-Gaussian clutter is related to that of the original Gaussian samples in an intricate manner, which makes it difficult to obtain a desired covariance matrix. Moreover, these methods do not guarantee that the resulting covariance matrix is nonnegative definite [92]. On the contrary, the SIRP model in Eq. (3.9) allows to control both the envelope PDF and the covariance matrix of the generated clutter.

One of the most common clutter distributions is the K-distribution, which provides a good fit to the envelope of the data acquired from different environments. The K-distribution of the clutter envelope is given by [79]

$$f(r) = \frac{2\delta}{\Gamma(\alpha)} \left(\frac{\delta r}{2}\right)^\alpha K_{\alpha-1}(\delta r), \tag{3.11}$$

where $\alpha > 0$ and $\delta > 0$ are the shape and scale parameters, respectively, $\Gamma(\cdot)$ is the Gamma function, and $K_\alpha(\cdot)$ is the modified Bessel function of the second kind of order α. In order to arrive at the K-distribution for the clutter envelope using the SIRP model, the PDF of the texture component $f_V(v)$ should be selected as [79]

$$f_V(v) = \frac{2\delta}{\Gamma(\alpha) 2^\alpha} (\delta v)^{2\alpha-1} \exp(-\delta^2 v^2). \tag{3.12}$$

In this case, the second moment of v is given by $E(v^2) = 2\alpha/\delta^2$. For the detailed simulation procedures to generate coherent K-distribution clutter with the desired covariance matrix, the reader can refer to [93].

3.4 The non-homogeneity Detector (NHD)

To calculate the adaptive weight vector $\mathbf{w}$ in Eq. (3.3) for a given CUT within the available L range cells, one needs to estimate the covariance matrix of this CUT from the adjacent $L_1 = L - 1$ secondary cells $\mathbf{z}_l$, where $l \in \mathcal{L} = \{1, \cdots, L_1\}$, that together form the secondary sample matrix $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_{L_1}] \in \mathbb{C}^{J \times L_1}$. To censor non-homogeneous secondary cells from the estimation, the NHD decides if a secondary cell, say $\mathbf{z}_k$ for $k \in \mathcal{L}$, is non-homogeneous with respect to the remaining $L_2 = L_1 - 1$ secondary cells $\mathbf{z}_l$ for $l \in \mathcal{L} - \{k\}$, which together form a matrix $\mathbf{Z}_k$ (obtained from $\mathbf{Z}$ by removing the column $\mathbf{z}_k$). The NHD is basically a STAP detector that sequentially processes the L_1 secondary cells with one of them, $\mathbf{z}_k$, temporarily considered as the CUT (also termed secondary CUT), while the remaining secondary cells $\mathbf{Z}_k$ are used to estimate the covariance matrix of $\mathbf{z}_k$.

A basic test employs the general inner product (GIP), which is equivalent to the square of the Mahalanobis distance [78], i.e.,

$$\Lambda_{\text{GIP}} = (\mathbf{z}_k - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{R}}^{-1} (\mathbf{z}_k - \hat{\boldsymbol{\mu}}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_1 \tag{3.13}$$

where $\hat{\boldsymbol{\mu}} \in \mathbb{C}^{J \times 1}$ is the sample mean of $\mathbf{z}_k$, $\hat{\mathbf{R}} \in \mathbb{C}^{J \times J}$ is its estimated covariance matrix, and η_1 is a threshold that is determined based on the required probability of false alarm P_{F}. In this test, H_0 is the null hypothesis that $\mathbf{z}_k$ is homogeneous with respect to $\mathbf{Z}_k$, while H_1 is the alternative hypothesis. However, the GIP test is not robust in non-Gaussian clutter environment as reported in [79], where a more robust detector, namely the normalized adaptive matched filter (NAMF), is proposed as

$$\begin{aligned} \Lambda_{\text{NAMF}} &= \frac{|\hat{\mathbf{w}}^H \mathbf{z}_k|^2}{(\hat{\mathbf{w}}^H \hat{\mathbf{R}} \hat{\mathbf{w}})(\mathbf{z}_k^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k)} \\ &= \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k|^2}{(\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s})(\mathbf{z}_k^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_2 \end{aligned} \tag{3.14}$$

where $\hat{\mathbf{w}} = g \hat{\mathbf{R}}^{-1} \mathbf{s}$. For this detector, P_F has been derived in [79] assuming Gaussian clutter, but it is not tractable analytically for non-Gaussian SIRP clutter. In the latter case, Monte Carlo simulations are used to set the threshold.

While the SCM in Eq. (3.6) is the ML estimator in the case of Gaussian clutter, the ML

estimator of the covariance matrix in the case of compound Gaussian clutter cannot generally be obtained in analytical form. Tyler introduced a generalization of the ML estimator for elliptical distributions (which include Gaussian along with other distributions) that can be expressed as the solution to the nonlinear equation [94]

$$\hat{\mathbf{R}} = \frac{J}{L_1} \sum_{\substack{l=1 \\ l \neq k}}^{L_1} \frac{\mathbf{z}_l \mathbf{z}_l^H}{\mathbf{z}_l^H \hat{\mathbf{R}}^{-1} \mathbf{z}_l} \tag{3.15}$$

However, besides the difficulties posed by solving Eq. (3.15) due to the high computational cost, it needs a large number of secondary cells L_1 for estimator accuracy [79]. An approximation to the ML estimator for the covariance matrix of SIRP clutter is given by [79]

$$\hat{\mathbf{R}}_{\text{SIRP}} = \frac{1}{L_1} \sum_{\substack{l=1 \\ l \neq k}}^{L_1} \zeta_l \mathbf{z}_l \mathbf{z}_l^H, \tag{3.16}$$

where

$$\zeta_l = \frac{h_{2J+2}(\mathbf{z}_l^H \hat{\mathbf{R}}_{\text{SIRP}}^{-1} \mathbf{z}_l)}{h_{2J}(\mathbf{z}_l^H \hat{\mathbf{R}}_{\text{SIRP}}^{-1} \mathbf{z}_l)} \tag{3.17}$$

where the function $h_{2J}(\cdot)$ is defined in Eq. (3.10). The scalar ζ_l cannot be expressed in a closed form, since both sides of Eq. (3.16) contain $\hat{\mathbf{R}}_{\text{SIRP}}$, but it can be found by the iterative expectation-maximization (EM) algorithm [95]. However, the EM algorithm converges slowly [79], especially for low values of the shape parameter α introduced in Eq. (3.11), which are common for many clutter environments. Moreover, the estimator in Eq. (3.16) needs *a priori* knowledge of the clutter distribution. Another approximation to the ML covariance estimator in case of non-Gaussian clutter is the iterative normalized sample covariance matrix (NSCM) that is obtained through the following recursive formula [90]

$$\hat{\mathbf{R}}_{\text{NSCM}}^{(t+1)} = \frac{J}{L_1} \sum_{\substack{l=1 \\ l \neq k}}^{L_1} \frac{\Re(\mathbf{z}_l)\Re(\mathbf{z}_l^T)}{\Re(\mathbf{z}_l^T)(\hat{\mathbf{R}}_{\text{NSCM}}^{(t)})^{-1}\Re(\mathbf{z}_l)} \tag{3.18}$$

where t denotes the iteration index. The computation is initialized with the estimator [90]

$$\hat{\mathbf{R}}_{\text{NSCM}}^{(0)} = \frac{J}{L_1} \sum_{\substack{l=1 \\ l \neq k}}^{L_1} \frac{\Re(\mathbf{z}_l)\Re(\mathbf{z}_l^T)}{\Re(\mathbf{z}_l^T)\Re(\mathbf{z}_l)} \tag{3.19}$$

Although it is also based on iterative procedures, its rate of convergence is faster than the EM-based algorithm mentioned above for the solution of Eq. (3.16) and Eq. (3.17), and it has been reported to converge after only four iterations [90]. Moreover, the NSCM shows a detection performance that is very close to that of the EM-based estimator [96]. Henceforth, whenever we use $\hat{\mathbf{R}}$ we mean $\hat{\mathbf{R}}_{\text{NSCM}}$.

3.5 The Proposed NHD

In this section, we first introduce the projection depth function and use it to provide covariance-free interpretations of the GIP and NAMF test statistics. We then introduce a covariance-free NHD that employs a novel nonparametric (distribution-free) test statistic based on the projection depth function and extend it to the case of correlated clutter.

3.5.1 Projection depth function

Let $\mathbf{z} \in \mathbb{C}^J$ be a random vector with joint cumulative distribution function $F(\mathbf{z})$. As we discussed in 2.6, a depth function is a random scalar $D(\mathbf{z}, F) \in [0, 1]$, defined as a function of $\mathbf{z}$ and taking into account the features of its distribution F. Ideally, the value of $D(\mathbf{z}, F)$ provides an inverse measure of "distance" from a central point (such as the median or the mean of the distribution F), which can used for the center-outward ordering of observations of vector $\mathbf{z}$ [42]. Based on this ordering, outliers can be detected when their distance from the center is larger than a certain threshold. Hence, the concepts of depth function and outliers are related. Specifically, we can define a measure of outlyingness as the function [54]

$$O(\mathbf{z}, F) = \frac{1}{D(\mathbf{z}, F)} - 1 \tag{3.20}$$

Let $\mu(\cdot)$ and $\sigma(\cdot)$ be univariate location and scale measures, respectively. Then, the

projection-based outlyingness of $\mathbf{z}$ is [54]

$$O(\mathbf{z}, F) = \sup_{\mathbf{u}\in\mathbb{C}^J, \|\mathbf{u}\|=1} \frac{|\mathbf{u}^H\mathbf{z} - \mu(F_\mathbf{u})|}{\sigma(F_\mathbf{u})} \tag{3.21}$$

where $F_\mathbf{u}$ is the cumulative distribution function (CDF) of $\mathbf{u}^H\mathbf{z}$. In practice, the sample version of Eq. (3.21) is found by replacing $F_\mathbf{u}$ by its empirical version $\hat{F}_\mathbf{u}$.

The projection-based outlyingness has a higher breakdown value in comparison to other types of outlyingness functions [54], which motivates its use in this work. To understand the concept of the breakdown value, consider the estimation of a scalar parameter θ from n observations $X_n = \{x_1, ..., x_n\}$, with $T_\theta(X_n)$ denoting the resulting estimator. Let us assume that out of these observations, m are replaced by arbitrary values (outliers), resulting in the contaminated sample set $X_{n,m}$. The estimator $T_\theta(X_{n,m})$ is calculated for the same parameter θ, but from the contaminated set $X_{n,m}$. The finite sample breakdown value of the estimator $T_\theta(\cdot)$ is the smallest ratio of contamination m/n for which the distance between $T_\theta(X_n)$ and $T_\theta(X_{n,m})$ can become arbitrarily large for certain choices of outliers [97]. For example, the sample mean has a breakdown value of $1/n$, which means that a single outlier in the sample data can affect the sample mean estimator. However, the breakdown value of the sample median is $1/2$, which means that as long as the outliers are less than half of the sample size, the estimator value is still unaffected.

The projection-based outlyingness function in Eq. (3.21) is a robust alternative to the Mahalanobis distance and, hence, to the GIP. A nonparametric GIP NHD detector was introduced in [87] for Gaussian clutter using the outlyingness function in Eq. (3.21) that evades the high computational burden of estimating the covariance matrix and its inverse with increasing dimensions of the range cells. We refer to it as the projection depth GIP (PD-GIP). However, the generation of the projection vectors $\mathbf{u}$ to approximate the supremum operation requires calculating the median of the secondary cells in $\mathbf{Z}_k$ for each $\mathbf{z}_k$; besides, the performance of the original GIP detector in case of non-Gaussian clutter environment is not robust [79]. Below, we propose a covariance-free detector based on Eq. (3.21) that is approximately equivalent to the NAMF detector in its robust performance, while at the same time sharing the nonparametric character of Eq. (3.21) and its lower computational complexity.

3.5.2 Covariance-free reformulation of GIP and NAMF

We begin by stating a proposition about the equivalence of the outlyingness function in Eq. (3.21) to the GIP in Eq. (3.13). This equivalence, which was demonstrated in [98] for the case of real-valued data in image processing applications, is extended here to complex-valued radar observations, as needed to comply with the case of coherent clutter model under consideration in this chapter.

Proposition 3.1. *Let* $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_{L_1}] \in \mathbb{C}^{J \times L_1}$ *be a secondary sample matrix. For any target steering vector* $\mathbf{s}$ *as in Eq.* (2.9) *and an arbitrary secondary cell* $\mathbf{z}_k$*, where* $k \in \{1, \cdots, L_1\}$*, is associated with an estimated covariance matrix* $\hat{\mathbf{R}}$ *and mean vector* $\hat{\boldsymbol{\mu}}$*, we have*

$$\sup_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H \mathbf{z}_k - \hat{\mu}(\mathbf{u}^H \mathbf{Z}_k)|}{\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)} \right)^2 = (\mathbf{z_k} - \hat{\boldsymbol{\mu}})^H \hat{\mathbf{R}}^{-1} (\mathbf{z_k} - \hat{\boldsymbol{\mu}}) \tag{3.22}$$

and

$$\sup_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H \mathbf{s}|}{\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)} \right)^2 = \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} \tag{3.23}$$

where $\hat{\mu}(\mathbf{u}^H \mathbf{Z}_k)$*,* $\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)$ *are the sample mean and standard deviation (SD) of* $\mathbf{u}^H \mathbf{Z}_k$*, respectively, and* $\mathbf{Z}_k$ *denotes the secondary cells after excluding* $\mathbf{z}_k$*.*

In Eq. (3.22) and Eq. (3.23), the supremum operation is taken over all unit-norm vectors $\mathbf{u} \in \mathbb{C}^J$. The proof of Proposition 1 is given in Appendix A. We note that the denominator of the test statistic of the NAMF detector in Eq. (3.14) can be expressed using the outlyingness function in Eq. (3.21) as shown in Appendix A, specifically Eq. (A.4) and Eq. (A.6). However, the numerator of Eq. (3.14), $\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k$, cannot be directly expressed in terms of Eq. (3.21). To circumvent this difficulty, we suggest replacing $\mathbf{z}_k$ in the numerator of Eq. (3.14) by $(\mathbf{s}^H \mathbf{z}_k)\mathbf{s}$ to obtain

$$|\mathbf{s}^H \hat{\mathbf{R}}^{-1} (\mathbf{s}^H \mathbf{z}_k) \mathbf{s}|^2 = |\mathbf{s}^H \mathbf{z}_k|^2 (\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s})^2 \tag{3.24}$$

The following proposition states that, in case of a dominant target, the expression in Eq. (3.24) is approximately equivalent to $|\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k|^2$, which is the numerator of Eq. (3.14).

Proposition 3.2. *Let* $\mathbf{s}$*,* $\mathbf{z}_k$*, and* $\hat{\mathbf{R}}$ *be as defined in Proposition 3.1, then* $(\mathbf{s}^H \mathbf{z}_k)\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}$ *has the same target's signal component as* $\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k$*.*

The proof of this proposition is provided in Appendix B. The next proposition introduces

a modified test statistic, which approximates the original NAMF test statistic in Eq. (3.14) in terms of the projection-based outlyingness in Eq. (3.21).

Proposition 3.3. *Let* $\mathbf{s}$, $\mathbf{z}_k$, $\mathbf{Z}_k$, *and* $\hat{\mathbf{R}}$ *be as defined in Proposition 3.1. Then*

$$\begin{aligned}\Lambda'_{NAMF} \triangleq & \frac{|\mathbf{s}^H\hat{\mathbf{R}}^{-1}(\mathbf{s}^H\mathbf{z}_k)\mathbf{s}|^2}{(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s})(\mathbf{z}_k^H\hat{\mathbf{R}}^{-1}\mathbf{z}_k)} \\ = & \frac{|\mathbf{s}^H\mathbf{z}_k|^2 \sup\limits_{\|\mathbf{u}\|=1}\left(\frac{|\mathbf{u}^H\mathbf{s}|}{\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2}{\sup\limits_{\|\mathbf{u}\|=1}\left(\frac{|\mathbf{u}^H\mathbf{z}_k|}{\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2}\end{aligned} \tag{3.25}$$

The proof of this proposition is provided in Appendix C. As observed from Eq. (3.25), the test statistic Λ'_{NAMF} is covariance-free. Moreover, besides its approximate equivalence to the NAMF test in Eq. (3.14) as shown in Proposition 3.2, it inherits the nonparametric characteristic of the projection-based outlyingness.

3.5.3 Robust, Covariance-free, and nonparametric NHD

Although the projection-based outlyingness in Eq. (3.21) does not dictate a specific scale measure, the median absolute deviation (MAD) has been widely used in robust statistics to detect outliers due to its robustness with respect to heavy-tailed distributions and higher breakdown value compared to the SD [99].

For the real-valued random sample data $X_n = [x_1, \cdots, x_n]$ with order statistics $x_{(1)} \leq \cdots \leq x_{(n)}$, the sample median $\text{med}(X_n)$ and sample median absolute deviation $\text{mad}(X_n)$ are calculated as [100]

$$\text{med}(X_n) = \begin{cases} x_{((n+1)/2)} & n \text{ is odd} \\ \\ 0.5(x_{(n/2)} + x_{((n/2)+1)}) & n \text{ is even} \end{cases} \tag{3.26}$$

and

$$\text{mad}(X_n) = \text{med}(|x_i - \text{med}(X_n)|), \qquad i = 1, \cdots, n \tag{3.27}$$

respectively. The population MAD of the random variable X $\text{MAD}(X)$ is related to its

population SD $\sigma(X)$ as [101]

$$\mathrm{MAD}(X) = k_f \sigma(X) \tag{3.28}$$

where k_f is a positive constant to achieve consistency with the scale parameter and its value depends on the population CDF of X. For the standard normal distribution, $k_f \approx 0.6745$ [99]. In the absence of outliers, the sample versions $\mathrm{mad}(X_n)$ and $\hat{\sigma}(X_n)$ are related, approximately, by the same constant k_f even with a sample size as low as 10 [99].

The breakdown value of the MAD is 0.5 [101], which is the best possible breakdown value, compared to a value of 0 for the SD [97]. Therefore, the MAD is more robust than the SD, especially for heavy-tailed clutter distributions as the K-distribution. Since heavy-tailed distributions tend to have many outliers with very high values, the MAD constitutes a better estimate for the scale parameter than the SD and leads to a lower threshold for the same false alarm rate and, consequently, a better detection.

By employing the $\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)$ as a robust scale measure instead of $\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)$ in Eq. (3.25) we obtain

$$\Lambda'_{\mathrm{NAMF}} \approx \frac{|\mathbf{s}^H\mathbf{Z}_k|^2 \sup\limits_{\|\mathbf{u}\|=1} \left(\frac{k_f|\mathbf{u}^H\mathbf{s}|}{\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2}{\sup\limits_{\|\mathbf{u}\|=1} \left(\frac{k_f|\mathbf{u}^H\mathbf{z}_k|}{\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2} \tag{3.29}$$

where, for the proper complex signal vectors $\mathbf{Z}_k$, we have according to Eq. (3.4)

$$\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k) = 2\mathrm{mad}(\Re(\mathbf{u}^H\mathbf{Z}_k)) = 2\mathrm{mad}(\Im(\mathbf{u}^H\mathbf{Z}_k)) \tag{3.30}$$

Under the SIRP model considered in this chapter, all the CDFs $F_{\mathbf{u}}$ of the projections of a given SIRP vector are the same [91]; this means that the value of k_f, that is determined based on $F_{\mathbf{u}}$, does not depend on the projection vector $\mathbf{u}$. To verify the independence of k_f from the projection vector $\mathbf{u}$ for the considered signal vector, a secondary sample matrix $\mathbf{Z} \in \mathbb{C}^{J\times L}$ of uncorrelated clutter vectors is simulated with the dimension of the secondary cells fixed at $J = 20$, while L/J changes from 2 to 10. For each sample size L, we perform 10^4 Monte Carlo simulation trials. For each trial, a sample of k_f from different 1000 projection vectors $\mathbf{u}$ is calculated using Eq. (3.28), but using the sample median absolute deviation $\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)$ as defined in Eq. (3.27) in place of $\mathrm{MAD}(\mathbf{u}^H\mathbf{Z}_k)$. The average relative standard

deviation (RSD) of the k_f sample is

$$\mathrm{RSD}(k_f) = \frac{\hat{\sigma}(k_f)}{\hat{\mu}(k_f)} \tag{3.31}$$

where $\hat{\sigma}(k_f)$ and $\hat{\mu}(k_f)$ are the sample SD and mean of k_f, respectively, averaged over 10^4 trials. As Fig. 3.1 shows, the value of k_f exhibits a low variation for both of the considered distributions at all considered values of L. This also shows that Eq. (3.28) holds also for $\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)$ even for a low sample size as shown in [99].

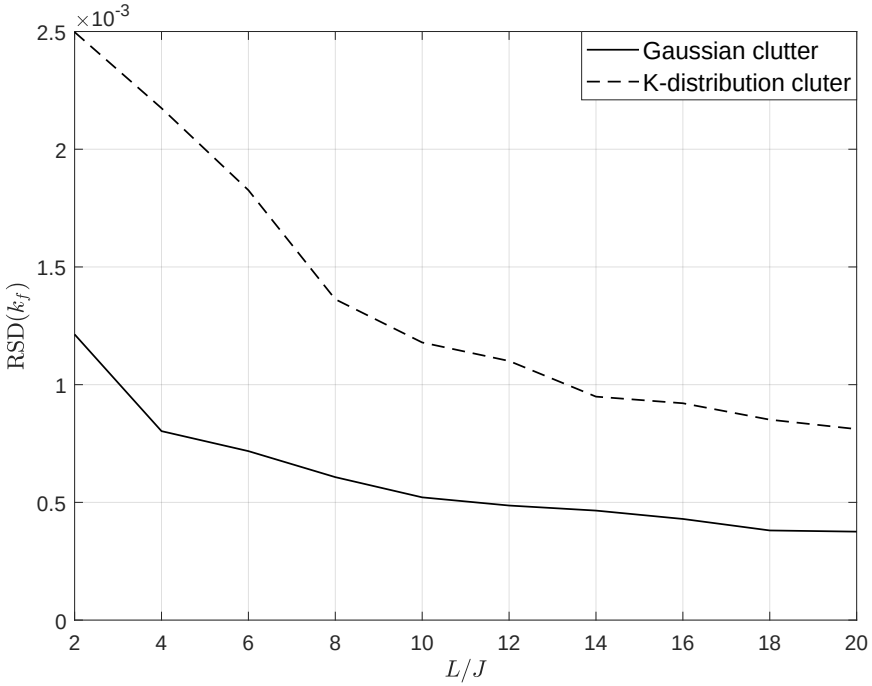


Figure 3.1 The relative SD of k_f at different number of samples (secondary cells) with $J = 20$

With the agreement of the presented simulation results with the theoretical analysis in [91, 99], the constant k_f can be taken out of the supremum in Eq. (3.29). Therefore, the proposed test (Λ_{PD}) based on the NAMF and projection depth (PD) outlyingness is

$$\Lambda_{\text{PD-NAMF}} \triangleq \frac{|\mathbf{s}^H\mathbf{Z}_k|^2 \sup\limits_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H\mathbf{s}|}{\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2}{\sup\limits_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H\mathbf{z}_k|}{\mathrm{mad}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_3 \tag{3.32}$$

We call the detector based on the test statistic in Eq. (3.32) as the PD-NAMF. Theoretically, implementing the supremum requires calculating the projections of an infinite number of vectors that cover the unit hypersphere in J-dimensional space. In practice, as shown in [98, 102] for different applications, the supremum can be approximated by taking the maximum magnitude of a finite number (Q) of projections of $\mathbf{z}_k$ or $\mathbf{s}$ on randomly generated vectors over this hypersphere. As suggested in [103], each of these vectors is obtained by first generating J independent complex Gaussian variates $u_i \sim \mathcal{CN}(0,1), 1 \leq i \leq J$ with zero mean and unit variance to form the vector $\mathbf{u} = [u_1, \cdots, u_J]$, and then normalizing $\mathbf{u}$ with respect to $\|\mathbf{u}\|_2$. As the steps above show, the generation method used in this work is totally independent of the steering vector $\mathbf{s}$ or the CUT $\mathbf{z}_k$, hence it is performed once and the obtained vectors are stored to be used for all range cells and any steering vector $\mathbf{s}$. This off-line method of generation of the projection vectors is different from that used in [87] for the PD-GIP, whose test statistic is on the basic form of the projection-based outlyingness function in Eq. (3.21). In [87], the projection vectors were recomputed for each CUT from the secondary cells $\mathbf{Z}$. The discussion on the choice of Q is left for Section 3.6.

3.5.4 Correlated Clutter

The correlation matrix of the clutter signal vector $\mathbf{c}$ in Eq. (3.7) is a Kronecker product of the temporal (i.e., between pulses) and spatial (i.e., between antenna elements) covariance matrices $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_s$, respectively, that is [104]

$$\mathbf{\Psi} = \mathbf{\Psi}_t \otimes \mathbf{\Psi}_s \tag{3.33}$$

The spatial correlation of the clutter depends on the inter-element spacing of the antenna array as in [105, 106]. It can be approximated as

$$\mathbf{\Psi}_s = \left[\rho_s^{|i-j|}\right], \qquad 1 \leq i, j \leq N \tag{3.34}$$

where ρ_s is the one-lag spatial correlation coefficient. Based on experimental measurements for different clutter environments, e.g., [107, 108], the temporal covariance matrix of the clutter can be expressed similarly as

$$\mathbf{\Psi}_t = \left[\rho_t^{|i-j|}\right], \qquad 1 \leq i, j \leq M \tag{3.35}$$

where ρ_t is the one-lag temporal correlation coefficient.

In practice, the sample version of the projection-based outlyingness function $O(\mathbf{z}, \hat{F})$ is used rather than the population version $O(\mathbf{z}, F)$. The former is obtained by replacing $F_{\mathbf{u}}$ in Eq. (3.21) with its empirical version $\hat{F}_{\mathbf{u}}$. Based on [109, Theorem B.1], we have

$$\sup_{\mathbf{z}} |O(\mathbf{z}, \hat{F}) - O(\mathbf{z}, F)| = o(1) \quad a.s., \tag{3.36}$$

if

$$\sup_{\|\mathbf{u}\|=1} |\mu(\hat{F}_{\mathbf{u}}) - \mu(F_{\mathbf{u}})| = o(1) \quad a.s., \tag{3.37}$$

and

$$\sup_{\|\mathbf{u}\|=1} |\sigma(\hat{F}_{\mathbf{u}}) - \sigma(F_{\mathbf{u}})| = o(1) \quad a.s. \tag{3.38}$$

In the case of (μ, σ) =(med, mad), as assumed in this work, Zuo has proven that equations (3.37) and (3.38) hold for elliptical distributions under the assumption of $\hat{F}_{\mathbf{u}} \rightarrow_d F_{\mathbf{u}}$, where $\rightarrow_d$ denotes convergence in distribution [54, Remark 2.4]. However, this assumption has been made assuming the samples drawn from $F_{\mathbf{u}}$ are iid, which is not true in the case of correlated clutter.

Therefore, we need to discuss the convergence of the empirical CDF to the population CDF for correlated data, which is addressed in [110, Theorem 1]. Let $\{x_i\}_{i=1}^n$, be random univariate samples that follow a joint normal distribution with correlation matrix $\mathbf{\Phi}$. If $\{x_i\}_{i=1}^n$ are not weakly correlated[1], then $E[\hat{G} - G]^2$ does not tend to 0 as $n \rightarrow \infty$. Hence, $O(\mathbf{z}, \hat{F})$ does not converge to $O(\mathbf{z}, \hat{F})$ and, consequently, the test in Eq. (3.32) may deviate from the true test value. Therefore, given the strong correlation shown by the available experimental data for different clutter environments [107, 108], the detection performance of the test in such environments may be degraded.

To handle this problem, we propose decorrelating $\mathbf{Z}$ before applying Eq. (3.32). The

[1] As a rule of thumb [111], the data $\{x_i\}_{i=1}^n$ is said to be weakly-correlated if its correlation coefficient is ≤ 0.4. For a more formal definition of weak correlation, define the average ℓ_1-norm of the correlation matrix $\mathbf{\Phi} \in \mathbb{R}^{J \times J}$ of the data $\{x_i\}_{i=1}^n$ as $\|\mathbf{\Phi}\|_1^{(J)} = \frac{1}{J^2} \sum_{i,j=1}^{J} |\phi_{ij}|$. If $\|\mathbf{\Phi}\|_1^{(J)} \rightarrow 0$, then $\{x_i\}_{i=1}^n$ are weakly correlated. Otherwise it is called strongly correlated [110, Definition 1].

decorrelated secondary cells are

$$\mathbf{Z}_{\mathrm{d}} = \hat{\mathbf{\Psi}}^{-1/2}\mathbf{Z} \tag{3.39}$$

where $\hat{\mathbf{\Psi}}$ is an estimate of the correlation matrix of $\mathbf{Z}$. To keep the nonparametric characteristic of the PD-NAMF, we use a nonparametric correlation estimator.

There are two prevalent nonparametric rank correlation coefficients, namely, the Kendall's and Spearman correlation coefficients. Compared to the Spearman coefficient, the Kendall's has a lower bias, shows better accuracy at lower number of samples, and has a lower mean square error (MSE) for heavily correlated data [112]. However, calculating the Spearman coefficient has a lower computation complexity than that of the Kendall coefficient. Generally, the Kendall correlation matrix estimator of the E-dimensional vector $\mathbf{x}$ calculated from the sample data $\mathbf{X} \in \mathbb{R}^{E\times D}$ is given by [113]

$$\hat{\psi}_{jk}^{K} = \frac{2}{D(D-1)} \sum_{i<i'} \mathrm{sign}(x_{ji} - x_{ji'})\mathrm{sign}(x_{ki} - x_{ki'}) \tag{3.40}$$

where $\hat{\psi}_{jk}^{K}$ is the (j,k)-th entry of $\hat{\mathbf{\Psi}}^{K}$, $1 \leq j,k \leq E$, $1 \leq i \leq D$, and $2 \leq i' \leq D$. The Spearman correlation matrix estimator calculated from the same sample data is [113]

$$\hat{\psi}_{jk}^{S} = \frac{\sum_{i=1}^{D} \left(o(x_{ji}) - \bar{D}\right)\left(o(x_{ki}) - \bar{D}\right)}{\sqrt{\sum_{i=1}^{D} \left(o(x_{ji}) - \bar{D}\right)^2 \sum_{i=1}^{L_1} \left(o(x_{ki}) - \bar{D}\right)^2}} \tag{3.41}$$

where $\bar{D} = \frac{D+1}{2}$ and $o(x_{ki})$ denotes the order of x_{ki} within $x_{k1}, ..., x_{kD}$.

The correlation estimators in Eq. (3.40) and Eq. (3.41) cannot be directly applied to the complex-valued secondary cells assumed in this chapter. Based on Eq. (3.4), the correlation matrix $\hat{\mathbf{\Psi}} \in \mathbb{C}^{MN\times MN}$ is given by

$$\hat{\mathbf{\Psi}} = 2\hat{\mathbf{\Psi}}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}} = 2\hat{\mathbf{\Psi}}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}} \tag{3.42}$$

where $\hat{\mathbf{\Psi}}_{\mathbf{z}_{\mathrm{R}}\mathbf{z}_{\mathrm{R}}}$ and $\hat{\mathbf{\Psi}}_{\mathbf{z}_{\mathrm{I}}\mathbf{z}_{\mathrm{I}}}$ are the estimated autocorrelation of $\Re(\mathbf{z})$ and $\Im(\mathbf{z})$, respectively. Equation Eq. (3.42) is applied to both Kendall and Spearman correlation matrix estimators $\hat{\mathbf{\Psi}}^{K}$ and $\hat{\mathbf{\Psi}}^{S}$, respectively.

Remark 1: It should be emphasized that the estimation of the correlation matrix $\mathbf{\Psi}$ is not the same as estimating the covariance matrix $\mathbf{R}$. The estimation of $\mathbf{\Psi}$ can be seen as a

step to estimate $\mathbf{R}$; a step that should be followed by estimating the SDs of the components of $\mathbf{z}_k$. To illustrate this, the correlation matrix $\mathbf{\Psi} = [\psi_{ij}]$ is related to the covariance matrix $\mathbf{R} = [r_{ij}]$ as

$$r_{ij} = \sigma_i \sigma_j \psi_{ij}, \qquad 1 \leq i, j \leq MN \tag{3.43}$$

where σ_i, σ_j are the SDs of the i-th and j-th components of $\mathbf{z}_k$, respectively. The step of estimating the SDs of $\mathbf{z}_k$'s components is cumbersome for non-Gaussian clutter models in addition to the need to calibrate the resulting covariance matrix by solving multiple optimization problems as shown in [114]. Under the proposed algorithm, it suffices to estimate $\mathbf{\Psi}$; avoiding the complexity of estimating $\mathbf{R}$.

Remark 2: In the case of Gaussian distributed data, both the Spearman and Kendall coefficient are related to the linear Pearson correlation coefficient $\hat{\psi}^P_{jk}$ by [113]

$$\hat{\psi}^P_{jk} = \sin\left(\frac{\pi}{2}\hat{\psi}^K_{jk}\right) = 2\sin\left(\frac{\pi}{6}\hat{\psi}^S_{jk}\right). \tag{3.44}$$

Nonetheless, we do not use the transformed coefficients for two reasons. The first is that they are derived for the Gaussian distributed data, while we do not make any assumptions about the distribution of the received signal vector. The second is that the transformations in Eq. (3.44) do not guarantee the positive semidefiniteness of the estimated matrices [115], in contrast to the estimators in Eq. (3.40) and Eq. (3.41).

The flow of the PD-NAMF with both of Kendall and Spearman decorrelation matrices is shown below.

Algorithm 1 Using Kendall

Input: $\mathbf{Z}$,η_3

Calculate $\hat{\mathbf{\Psi}}^{K}$

$\mathbf{Z} \leftarrow (\hat{\mathbf{\Psi}}^{K})^{-1/2}\mathbf{Z}$

Generate $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_Q]$

for $k = 1$ to L **do**

Perform the proposed test as Eq. (3.32)

end for

Algorithm 2 Using Spearman

Input: $\mathbf{Z}$,η_3

Generate $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_Q]$

for $k = 1$ **to**L **do**

$\mathbf{Z}_k \leftarrow [\mathbf{z}_l], l \in \mathcal{L} - \{k\}$

Calculate $\hat{\mathbf{\Psi}}^{S}$ from $\mathbf{Z}_k$

$\mathbf{Z}_k \leftarrow (\hat{\mathbf{\Psi}}^{S})^{-1/2}\mathbf{Z}_k$

Perform the proposed test as Eq. (3.32)

end for

3.6 Performance Assessment

In this section, the performance of the PD-NAMF in Eq. (3.32) is compared to that of the NAMF detector in Eq. (3.14) using Monte Carlo simulations. To justify the robustness of the PD-NAMF, we evaluate its performance with different clutter distributions and signal configurations. Moreover, we study the different choices for the algorithm parameters, specifically, the type of the correlation estimator used and the minimum required number of projections. Finally, we investigate the complexity and the execution time of the PD-NAMF compared to the NAMF detector for different design parameters.

3.6.1 Simulation Parameters

The simulated radar signal has a fixed dimension $J = 16$ and L_1 is either 65 or 33 cells. The non-homogeneity detector is applied on a sequential basis on each secondary cell where $\hat{\mathbf{R}}$ and mad$(\mathbf{u}^H\mathbf{Z}_k)$ are estimated from the remaining $L_2 = 64$ or 32 cells. An interfering target is injected in a secondary CUT, representing a non-homogeneous cell, with a normalized Doppler frequency $f_d = 0.3$ and azimuth angle $\theta_t = 35°$.

The Kendall correlation matrix $\hat{\mathbf{\Psi}}^{K}$ is estimated once from all the L_1 secondary cells $\mathbf{Z}$ including the secondary CUT $\mathbf{z}_k$, i.e., it is not recalculated for each $\mathbf{z}_k$. However, the Spearman correlation matrix $\hat{\mathbf{\Psi}}^{S}$, due to its lower immunity to outliers, is calculated for each

secondary CUT $\mathbf{z}_k$ from the remaining secondary cells. If the secondary CUT is included in the calculations of $\hat{\mathbf{\Psi}}^{S}$, a self-nulling effect appears at the output of the detector, especially at low number of secondary cells and/or high interfering target's power.

Regarding the clutter vectors, they are generated as proper complex SIRP vectors with independent quadrature and in-phase components. For the clutter's envelope distribution, we consider two extreme cases: K-distributed clutter with $\alpha = 0.1$, which represents heavy-tailed spiky clutter, and Gaussian clutter. For the K-distributed clutter, δ is allowed to be randomly and independently changed from a range cell to another as indicated by [86].

As for the value of δ, [116] gives measured values between $(0, 1]$, which is the range of values considered in most of the relevant works in the literature, where only the shape parameter is considered to have an impact on the detection performance [79]. However, [117] provides experimental data showing that $1 \leq \delta \leq 2$. Therefore, we examine the performance of both the NAMF and the PD-NAMF in K-distributed clutter with the foregoing two cases of δ for each range cell: $\delta \sim \mathcal{U}(0, 1]$, as a default case, and $\delta \sim \mathcal{U}[1, 2]$, where $\mathcal{U}$ denotes the uniform distribution. The average clutter-to-noise ratio (CNR) is assumed to be 20 dB. The one-lag spatial and temporal correlation coefficients of the clutter in equations (3.34) and (3.35) are $\rho_s = \rho_t = 0.99$ [106, 108], unless other values are specified.

The projection vectors are generated randomly over the J-dimensional unit hypersphere as defined in Proposition 3.1. The default number of projections is $Q = 4J$, however, we consider other values for Q later in this section. For all detection performance simulations, P_{F} is set to 0.01. The probability of detection P_{D} is evaluated versus the input signal-to-noise (SNR) of the interfering target using Monte Carlo simulation with 10^5 trials.

3.6.2 Results

3.6.2.1 The Low Dimensional Case ($J/L_2 = 0.25$)

The STAP radar signal is considered low dimensional when $J \leq 0.5L_2$. Fig. 3.2 shows the detection performance of the PD-NAMF and the NAMF detectors in K-distributed clutter with $\alpha = 0.1$ and $\delta \sim \mathcal{U}(0, 1]$. We can observe that the detection performance of the PD-NAMF is comparable to that of the NAMF detector with a maximum loss in P_{D} of 0.051 in the case of $\hat{\mathbf{\Psi}}^{S}$. For SNR > -7dB, it is also observed that using $\hat{\mathbf{\Psi}}^{K}$ provides a relative improvement over $\hat{\mathbf{\Psi}}^{S}$, narrowing the loss in P_{D} relative to the NAMF detector to 0.036. The difference in the performance between the NAMF detector and the PD-NAMF is getting

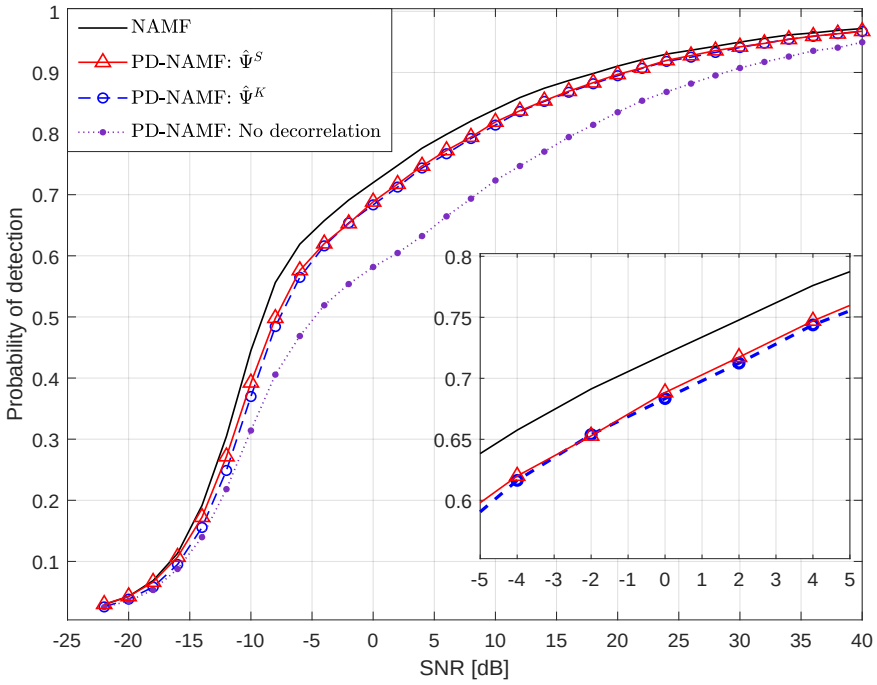


Figure 3.2 Detection performance in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$)

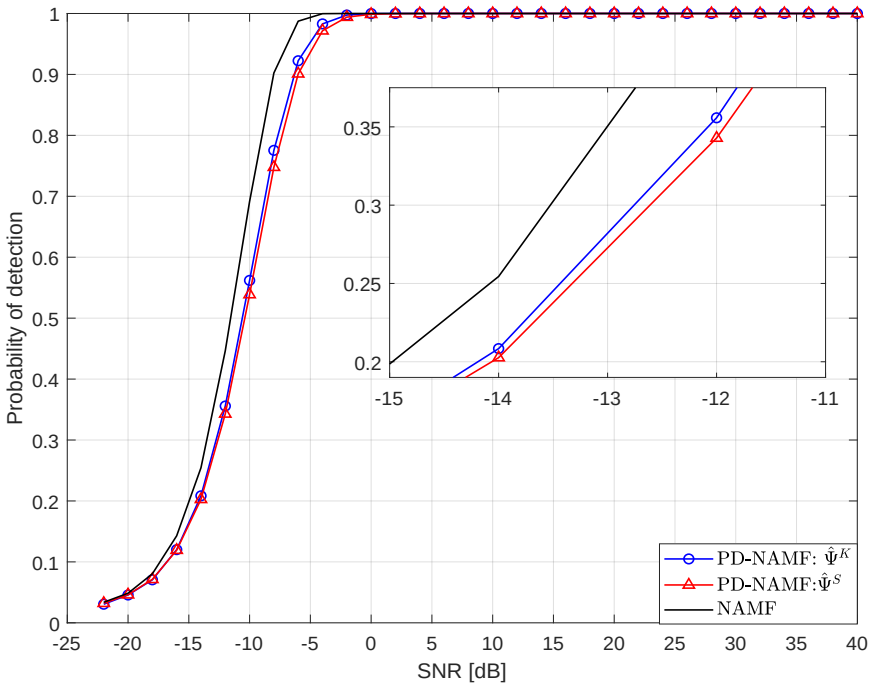


Figure 3.3 Detection performance in Gaussian clutter ($J = 16$, $L_2 = 64$)

narrower for SNR values beyond 5 dB and below -6 dB. While the observed slight advantage of the NAMF over the PD-NAMF is not common among all the cases studied in this chapter, it also comes with the cost of much higher complexity, as we show shortly. In Fig. 3.2, we can also point out that the decorrelation is not only dictated by the theoretical need to achieve the Fisher-consistency of $\text{mad}(\mathbf{u}^H\mathbf{Z}_k)$, but it also has a significant effect on the performance of the PD-NAMF.

The performance in Gaussian distributed clutter is shown in Fig. 3.3. Compared to the NAMF detector, the PD-NAMF has a maximum loss in P_{D} of 0.051 with both $\hat{\mathbf{\Psi}}^{S}$ and $\hat{\mathbf{\Psi}}^{K}$. Moreover, the performance of the PD-NAMF with $\hat{\mathbf{\Psi}}^{S}$ is similar to, or slightly better than that with $\hat{\mathbf{\Psi}}^{K}$.

To validate our claim of the robustness of the PD-NAMF using the MAD, we replaced MAD by SD in Eq. (3.29) and we evaluated the resulting detection performance in both Gaussian and K-distributed clutter. The results are shown in Fig. 3.4. As we can observe, the performance of the PD-NAMF using the SD is almost equivalent to that using the MAD in the case of the Gaussian distribution for both Spearman $\hat{\mathbf{\Psi}}^{S}$ and Kendall $\hat{\mathbf{\Psi}}^{K}$ decorrelation matrices. This is attributed to the equivalence of the SD and MAD, up to a constant k_f, in the case of the Gaussian distribution as we indicate in Eq. (3.28), which is based on [99]. In the case of the K-distribution, however, the performance of the PD-NAMF with both $\hat{\mathbf{\Psi}}^{S}$ and $\hat{\mathbf{\Psi}}^{K}$ degrades when SD is used in place of MAD. This is consistent with the theoretical reasoning provided in Subsection 3.5.3. We can also notice that when using SD, the PD-NAMF with$\hat{\mathbf{\Psi}}^{K}$ is more robust than the one with $\hat{\mathbf{\Psi}}^{K}$, which is ascribed to the higher robustness of the former in the presence of outliers as shown in [118].

It is important to demonstrate the performance of the original GIP and the PD-GIP detectors compared to the PD-NAMF in simulation. Interestingly, to the best of our knowledge, the detection performance of GIP in correlated compound Gaussian clutter has not been investigated in the open literature. Furthermore, the detection performance of the PD-GIP has not yet been investigated. For a fair comparison between the PD-GIP and the proposed PD-NAMF, we use a modified version of PD-GIP that differs from the one originally proposed in [87] in the following ways: the secondary cells are decorrelated using $\hat{\mathbf{\Psi}}^{K}$, and the projection vectors are generated in the same way as in the proposed PD-NAMF. To make this point clear, we refer to this detector as the "modified PD-GIP." Figs 3.5 and 3.6 show the performance of GIP, PD-GIP, modified PD-GIP, and PD-NAMF for both Gaussian and K-distributed clutter, respectively. We first note from the figure that the modifications

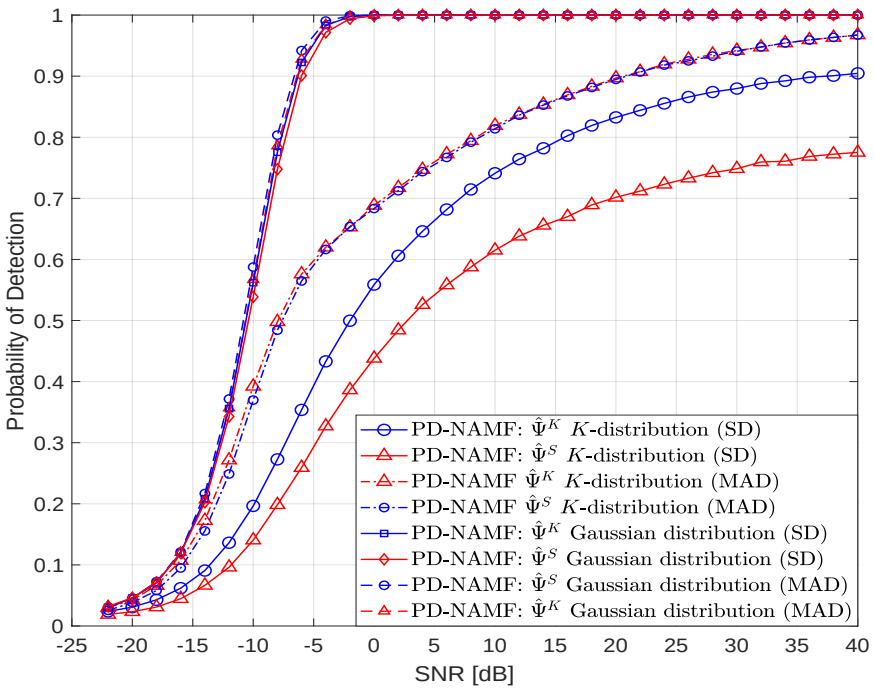


Figure 3.4 Detection performance in Gaussian and K-distributed clutter for PD-NAMF with MAD and SD ($J = 16, L_2 = 64$)

made to the PD-GIP contribute to improve its performance, as observed in the degraded performance of the original PD-GIP. Furthermore, both the GIP and the modified PD-GIP show a performance degradation in the case of Gaussian clutter of approximately 5dB and 7dB compared to the proposed PD-NAMF, respectively. However, this degradation is much greater in the case of K-distributed clutter. These results are consistent with the false alarm results for GIP presented in [79].

The effect of the scale parameter on the detection performance of both detectors can be observed in Fig. 3.7. The PD-NAMF performs approximately the same with both $\hat{\mathbf{\Psi}}^S$ and $\hat{\mathbf{\Psi}}^K$. The maximum detection loss by the PD-NAMF relative to NAMF is 0.034. In general, the detection performance of both the PD-NAMF and NAMF detectors are considerably affected by the change in the scale parameter of the clutter. In addition to the typical choice of 0.99 for both ρ_s and ρ_t, we consider the case of a low value $\rho_s = \rho_t = 0.2$ for these coefficients. The results are also shown in Fig. 3.7. The lower correlation coefficients of the clutter lead to a degraded detection performance which coincides with the results in [119, Fig. 6].

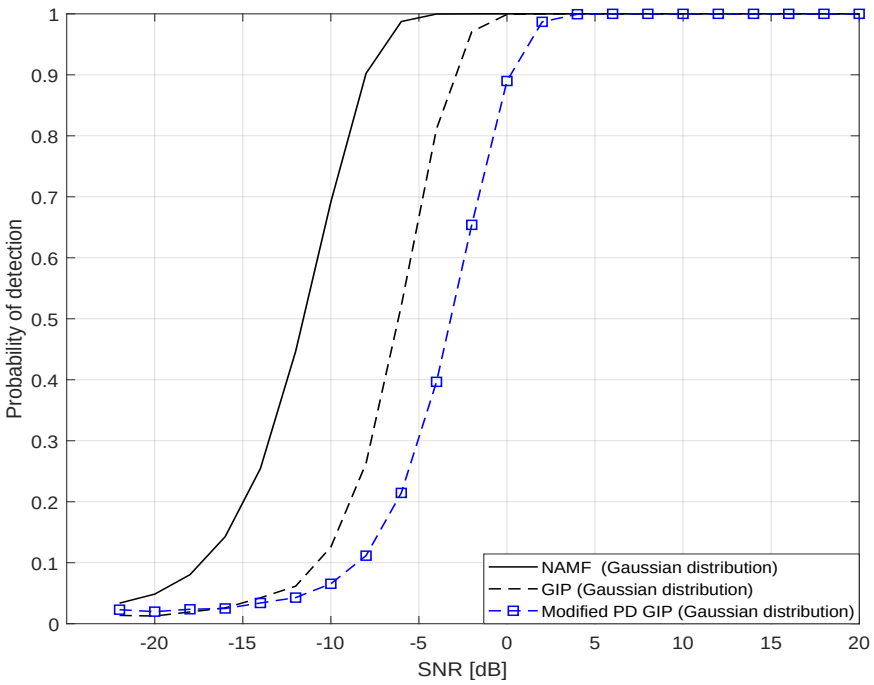


Figure 3.5 Detection performance in Gaussian clutter for GIP, modified PD-GIP, and PD-NAMF ($J = 16, L_2 = 64$)

3.6.2.2 The Higher Dimensional Case ($J/L_2 = 0.5$)

The performance of the PD-NAMF is investigated at a higher dimensional case, where $L_2 = 32$ and $J = 16$. As we notice in Fig. 3.8, in the presence of K-distributed clutter, the PD-NAMF with $\hat{\mathbf{\Psi}}^{K}$ provides a relative advantage over the NAMF at SNR< -6 dB, with a maximum increase in P_{D} of 0.06. Beyond this point, the maximum loss in detection of the PD-NAMF with $\hat{\mathbf{\Psi}}^{K}$ relative to NAMF detector is 0.03. With $\hat{\mathbf{\Psi}}^{S}$ the PD-NAMF shows a maximum loss in P_{D} of 0.061 relative to the NAMF for -10 dB $\leq$ SNR $\leq$ 0 dB. It is noteworthy that the overall performance of both detectors is relatively degraded by lowering L_2/J as we observe by comparing the performance of each detector in Fig. 3.2 with its counterpart in Fig. 3.8.

As Fig. 3.9 depicts for the Gaussian clutter, P_{D} of the PD-NAMF with $\hat{\mathbf{\Psi}}^{K}$ is higher than that of the NAMF with a maximum difference of 0.16. It is noteworthy that this improvement in P_{D} provided by the PD-NAMF is higher than any loss it shows relative to the NAMF in the previous cases. When the PD-NAMF uses $\hat{\mathbf{\Psi}}^{S}$, it shows a maximum improvement of 0.014 over the NAMF. Beyond the crossover point at SNR$= -6$ dB, the PD-NAMF with

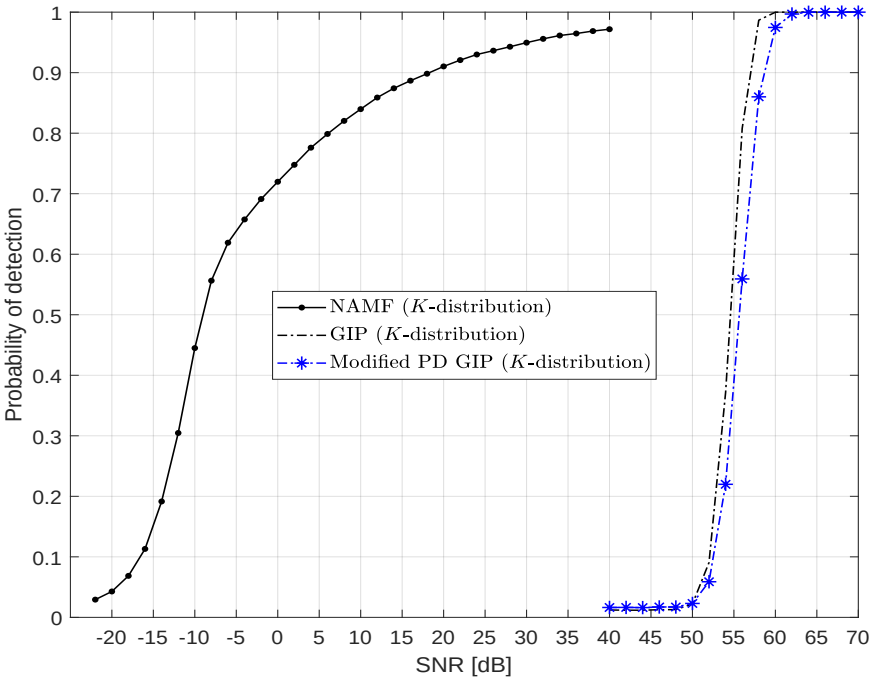


Figure 3.6 Detection performance in K-distributed clutter for GIP, modified PD-GIP, and PD-NAMF ($J = 16, L_2 = 64$)

$\hat{\mathbf{\Psi}}^{S}$ shows a comparable detection performance to the NAMF with a maximum loss in P_{D} of 0.041.

To summarize, using $\hat{\mathbf{\Psi}}^{K}$ with the PD-NAMF improves the detection performance over that of the NAMF detector in case of high dimensional signals at all SNR values in Gaussian clutter and at lower SNR values for K-distributed clutter. This is explained by the robustness of the Kendall's coefficient in small sample conditions as mentioned before.

3.6.2.3 Number of Projections

Theoretically, the higher the number of random projections Q, the more accurate Eq. (3.22) holds [98]. However, in practice, the used number of projections should be as small as possible for fast computations. Unfortunately, there is no analytical method to determine the minimum number of projections required for Eq. (3.22) to hold at a given approximation level, consequently, simulations are used to determine this value as in [98, 102]. While the simulations in these references are concerned with the convergence of Eq. (3.22), the simulation in this chapter is concerned with maximizing P_{D} at a given level of false alarm.

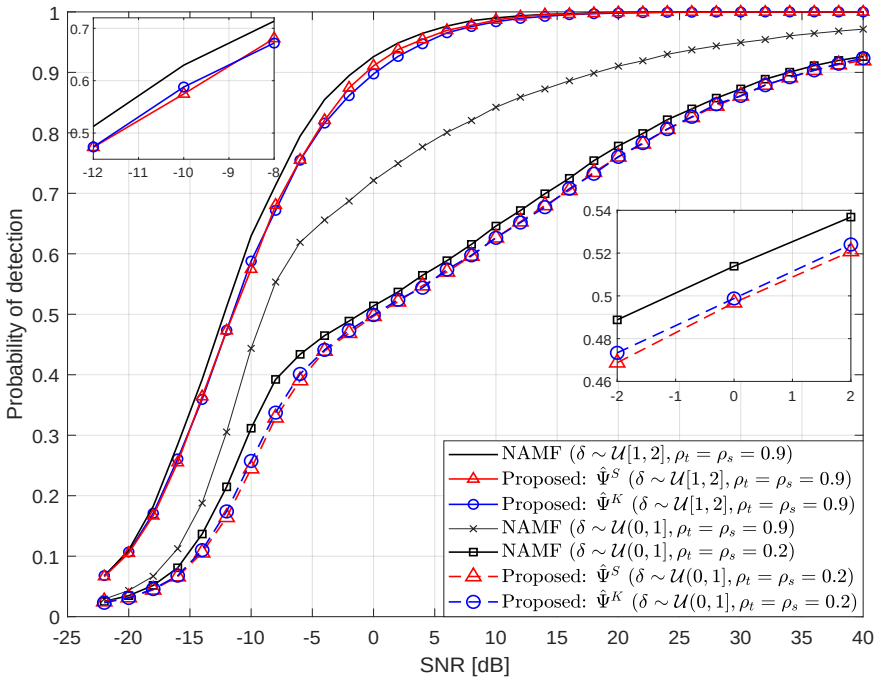


Figure 3.7 Detection performance in K-distributed clutter ($\alpha = 0.1$, $\delta \sim \mathcal{U}[1,2]$ or $\mathcal{U}(0,1]$, $J = 16$, $L_2 = 64$, $\rho_s = \rho_t = 0.99$ or 0.2)

Fig. 3.10 illustrates the power of a secondary CUT with homogeneous interference (clutter and noise) or an interfering target's signal power at the output of the PD-NAMF with $\hat{\mathbf{\Psi}}^K$ and $\hat{\mathbf{\Psi}}^S$, as a function of the ratio Q/J, averaged over 10^5 trials. The powers of both interference and target signals of the PD-NAMF, with both correlation estimators, are normalized with respect to those of the NAMF detectors, respectively. The simulated clutter envelope follows the K-distribution and the radar signal is low-dimensional ($J = L/4$). As shown in Fig. 3.10, using $\hat{\mathbf{\Psi}}^K$ results in a lower interference power level at the output of the PD-NAMF than $\hat{\mathbf{\Psi}}^S$, which explains its superior performance relative to the latter. The interference level of both of the correlation estimators decreases as Q increases up to $Q = 4J$, beyond this point the interference level is almost constant with $\hat{\mathbf{\Psi}}^S$ while it decreases slightly using $\hat{\mathbf{\Psi}}^K$. In general, it is also obvious that the interference power at the output of the PD-NAMF for $\hat{\mathbf{\Psi}}^K$ and $\hat{\mathbf{\Psi}}^S$ is lower than that at the output of the NAMF detector. This demonstrates the validity of the proof in Appendix B. Fig. 3.10 also reveals that Q has a negligible effect on the target signal level at the output of the PD-NAMF for both decorrelators with a relative higher target's signal level for $\hat{\mathbf{\Psi}}^K$ than that of $\hat{\mathbf{\Psi}}^S$.

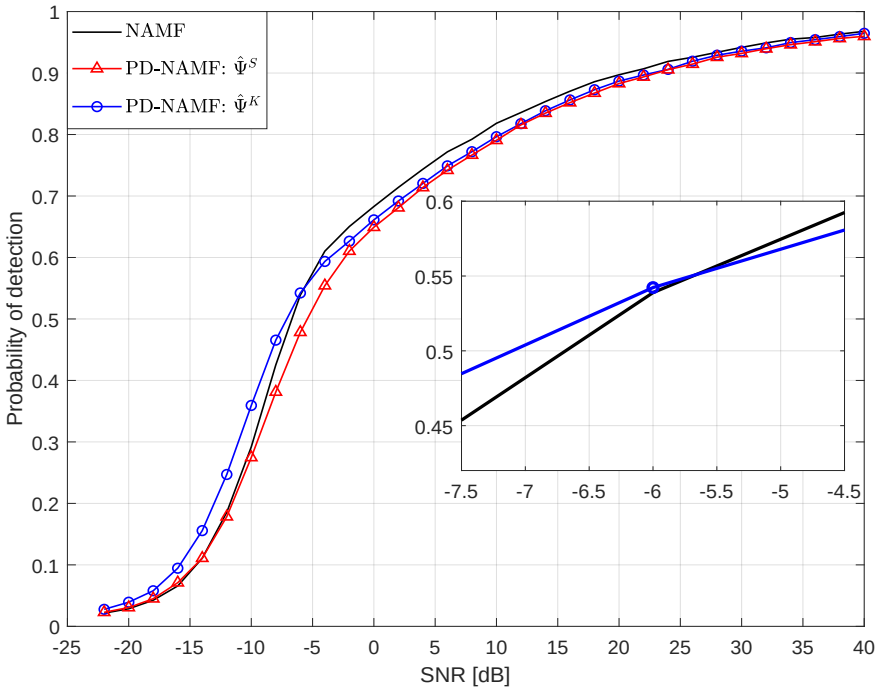


Figure 3.8 Detection performance in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 32$)

The final choice of the minimum required Q is based on the detection performance of the PD-NAMF as depicted in Fig. 3.11, where $\hat{\mathbf{\Psi}}^S$ is used in the presence of K-distributed clutter ($\delta \sim \mathcal{U}(0,1]$) and $J = L_2/4$. We can see that the increase of Q beyond $4J$ has a negligible impact on the detection performance and for most values of SNR there is no difference in the performance. When Q is reduced to $2J$, P_{D} decreases slightly with a maximum loss of 0.022. The same is shown for $\hat{\mathbf{\Psi}}^K$ in Fig. 3.12, but with a slight improvement with $Q = 10J$ at lower SNR values even over the NAMF. To investigate the dependence of Q on the clutter distribution, we performed additional simulations for different values of Q, but in presence of Gaussian clutter. As shown in Figs 3.13 and 3.14, for both $\hat{\mathbf{\Psi}}^K$ and $\hat{\mathbf{\Psi}}^S$, the performances of the PD-NAMF in the Gaussian clutter for different values of Q exhibit the same trend as in the K-distributed clutter shown in Figs 3.11 and 3.12. Therefore, we can conclude that using $Q = 4J$ projections is an appropriate rule of thumb that does not depend on the clutter distribution.

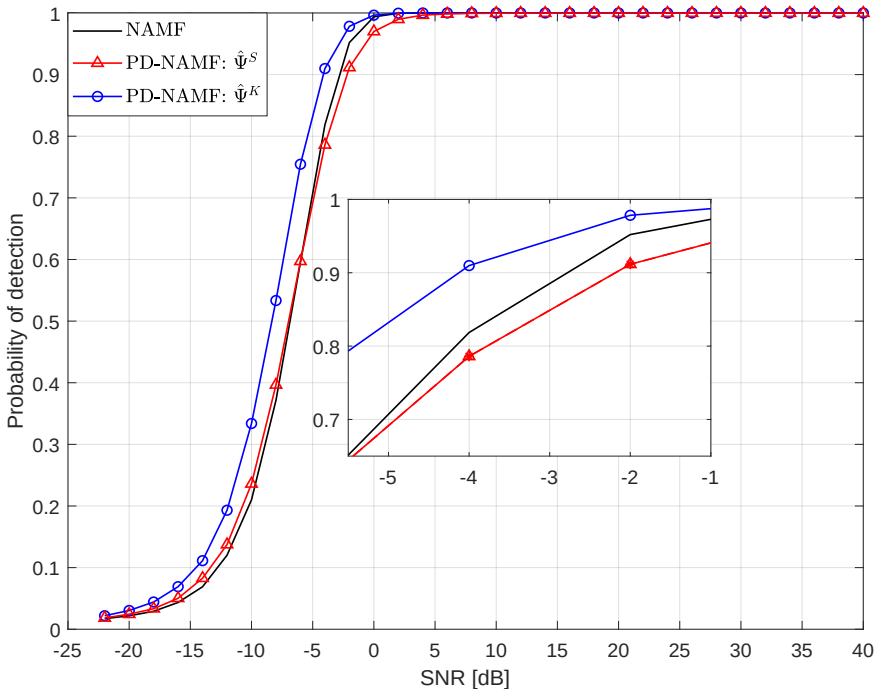


Figure 3.9 Detection performance in Gaussian distributed clutter ($J = 16, L_2 = 32$)

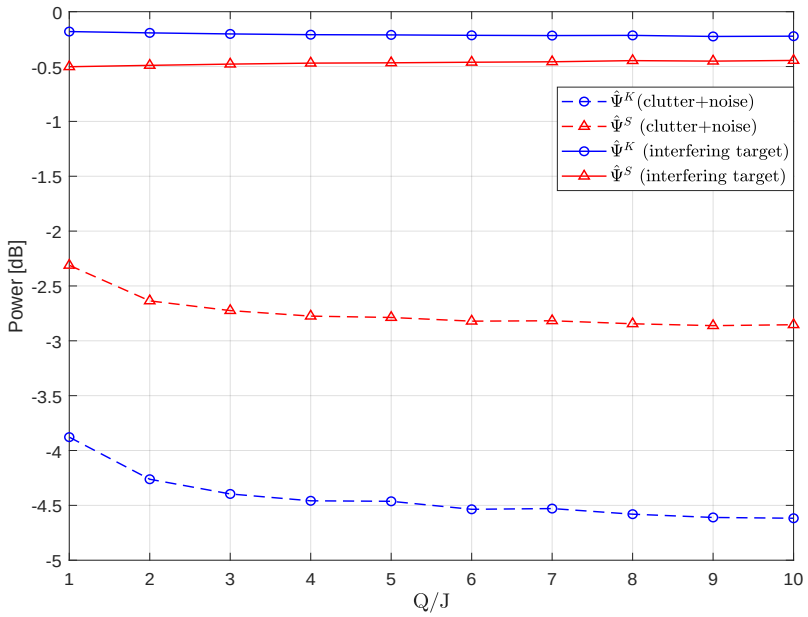


Figure 3.10 Interference and the interfering target powers at the output of the PD-NAMF in K-distributed clutter ($\alpha = 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$)

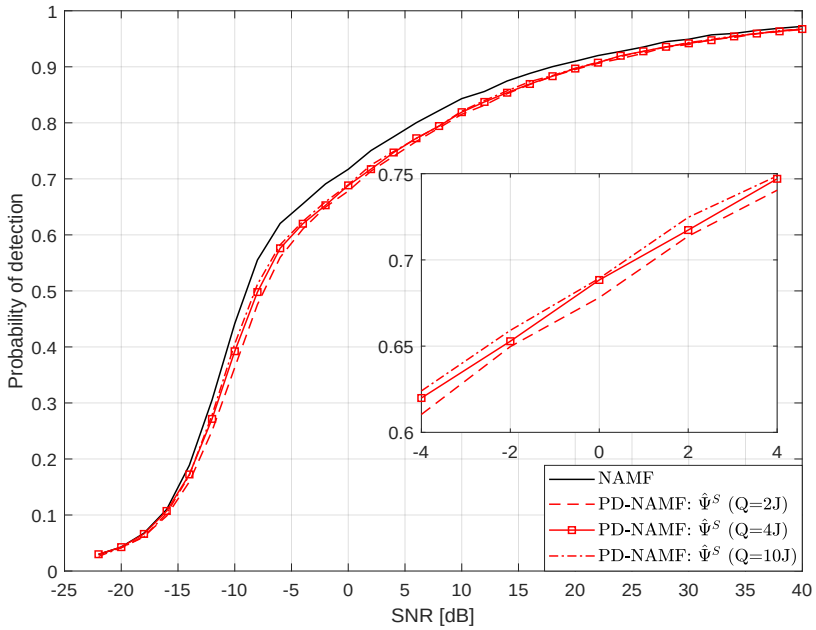


Figure 3.11 Detection performance in K-distributed clutter with different Q values ($\hat{\mathbf{\Psi}}^S$, ($\alpha = \ 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$)

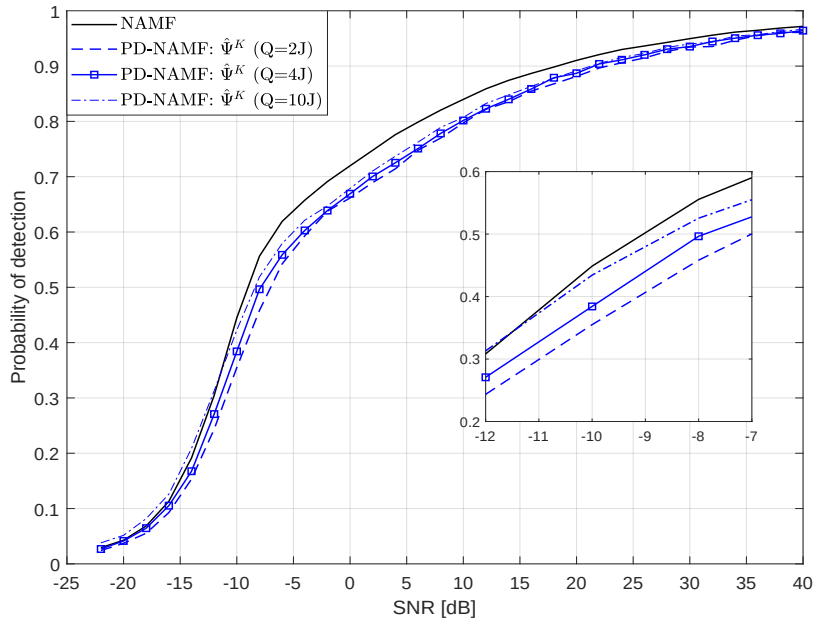


Figure 3.12 Detection performance in K-distributed clutter with different Q values ($\hat{\mathbf{\Psi}}^K$, $\alpha = \ 0.1, \delta \sim \mathcal{U}(0,1], J = 16, L_2 = 64$)

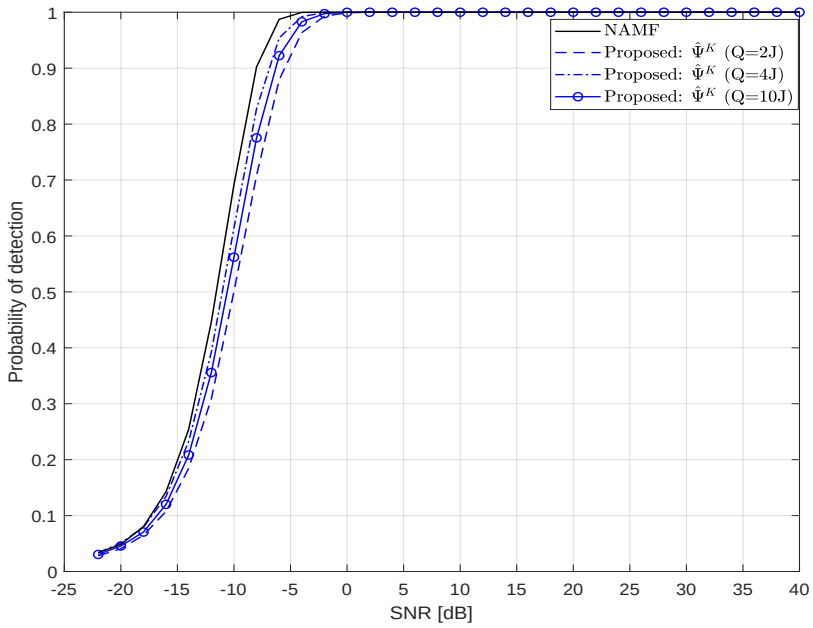


Figure 3.13 Detection performance in Gaussian distributed clutter with different Q ($\hat{\mathbf{\Psi}}^K$, $J = 16, L_2 = 64$)

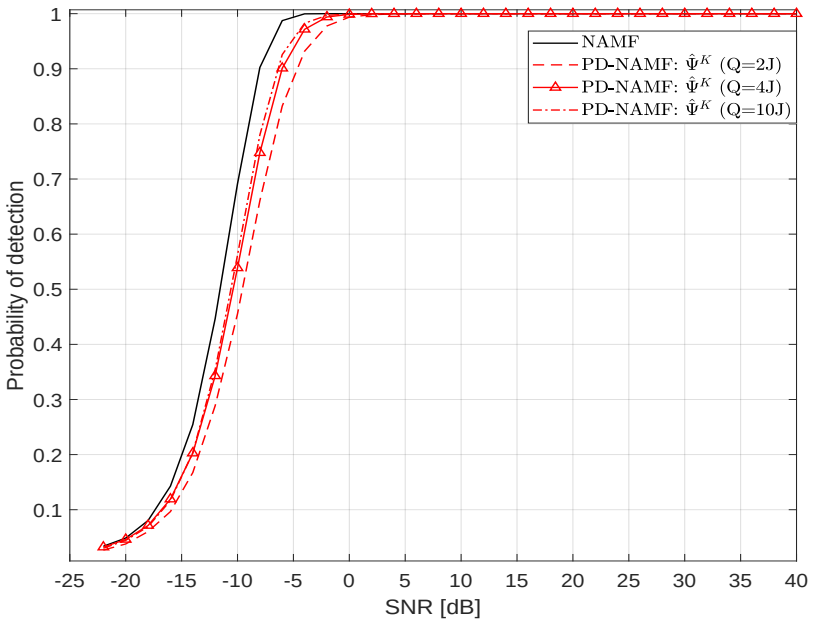


Figure 3.14 Detection performance in Gaussian distributed clutter with different Q ($\hat{\mathbf{\Psi}}^S$, $J = 16, L_2 = 64$)

3.6.2.4 Complexity Analysis

The complexity of the PD-NAMF (with both $\hat{\mathbf{\Psi}}^{K}$ and $\hat{\mathbf{\Psi}}^{S}$) compared to the NAMF is analyzed in terms of the required arithmetic operations and the run time of each detector. table 3.1 summarizes the mathematical operations performed by the PD-NAMF compared to the NAMF for each secondary cell. The reported complexities are based on the Gaussian-Jordan elimination, Schoolbook, and merge sort algorithms for matrix inversion, matrix multiplication, and sorting, respectively [120]. By $1/L_1$ we mean that $\hat{\mathbf{\Psi}}^{K}$ is calculated once for all the L_1 secondary cells and not for each cell in contrast to $\hat{\mathbf{\Psi}}^{S}$, which is estimated for each secondary cell from the remaining L_2 cells.

The computation reduction is more obvious in Fig. 3.15, where the run times of the NAMF detector and the PD-NAMF are computed on the same platform dedicated only for this job. The specifications of the machine used for this simulations are 64-bit Intel® Core™ i7-6700 central processing unit (CPU) @3.4GHz and 16 GB of random access memory (RAM). For both versions of the PD-NAMF with $\hat{\mathbf{\Psi}}^{S}$ and $\hat{\mathbf{\Psi}}^{K}$, the figure shows their average run times normalized by the run time of NAMF with $L_2 = 4J$ secondary cells for different J. It

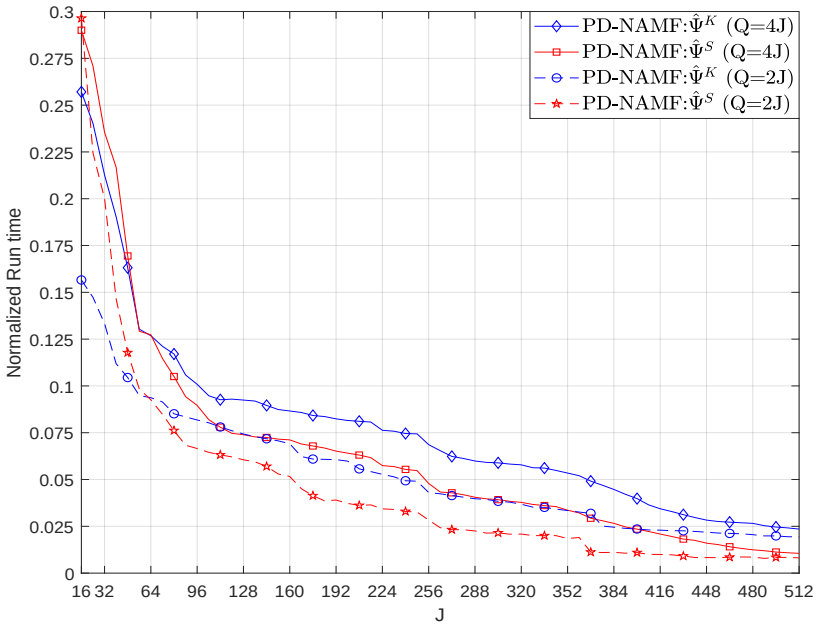


Figure 3.15 The run times of the PD-NAMF normalized by that of NAMF ($L_2 = 4J$)

is conspicuous that the PD-NAMF, either with $\hat{\mathbf{\Psi}}^{S}$ or $\hat{\mathbf{\Psi}}^{K}$, substantially reduces the NHD

Table 3.1 The performed operations by the proposed and the NAMF tests.

	NAMF		Proposed		
	Complexity	Times	Complexity	Times	
				Kendall	Spearman
Multiplication	$\mathcal{O}(J^2)$	$5L_2$	$\mathcal{O}(J^2L_1)$	$1/L_1$	–
	$\mathcal{O}(J)$	L_2	$\mathcal{O}(J^2L_2)$	–	1
	$\mathcal{O}(J^3)$	$4L_2+3$	$\mathcal{O}(QJ)$	2	2
			$\mathcal{O}(QJL_2)$	1	1
Inverse	$\mathcal{O}(J^3)$	5	$\mathcal{O}(J^3)$	$1/L_1$	1
Addition	$\mathcal{O}(J^2)$	$5(L_2-1)$	–	–	–
Sort	–	–	$\mathcal{O}(QL_2\log(L_2))$	4	4
Correlation Estimation			$\mathcal{O}(J^2L_1^2)$	$1/L_1$	–
			$\mathcal{O}(J^2L_2\log(L_2))$	–	1

run time depending on J. The larger the dimension of the cell J (and consequently L_2), the greater the reduction, which is of a great importance for modern radar systems with large antenna arrays.

Remark 2: The complexity of the PD-NAMF can be reduced further using parallel processing, given the independence of the random projections from each other. Moreover, the use of parallel programming on graphical processing units (GPUs) can reduce the complexity of calculating the median, and consequently the MAD, as in [121]. Furthermore, the calculation of $\hat{\mathbf{\Psi}}^{K}$, the most computationally demanding step, can be parallelized as well [122]. Nevertheless, the parallelization is only possible partially in the NAMF due to the iterative nature of the robust covariance estimators.

3.7 Conclusion

In this chapter, we introduced a novel covariance-free, nonparametric NHD detector for correlated clutter environments with Gaussian and non-Gaussian distributions. Based on the

projection depth function, the proposed PD-NAMF avoids the computationally expensive estimation of the covariance matrix. Interestingly, the larger the dimension of the radar signal vector, the higher the computation reduction the PD-NAMF provides relative to the NAMF. This advantage fosters the application of the PD-NAMF in modern radars with large antenna arrays. Further, this significant complexity reduction is not achieved at the expense of a degraded performance. That is, the detection performance of the new detector is shown to be comparable to, and in some cases better than, the full adaptive NAMF detector at different dimensions and clutter distributions. With this robust performance and the considerable reduction in computations, the PD-NAMF is superior to its covariance-based counterparts in the literature for real time applications and it can be a more efficient replacement of the computationally demanding GIP and NAMF detectors in iterative NHD approaches. The feasible utilization of parallel processing and GPUs paves the way for more efficient implementations of the PD-NAMF in the future.

With this robust performance of the proposed NHD and the choice of robust covariance estimators as the one in Eq. (3.18), we can obtain an accurate estimation of the interference covariance matrix. The estimated covariance matrix is employed in the waveform design of cognitive MIMO radars as we will show in Chapter 4.

Chapter 4

Design of Power-Efficient Waveforms for Cognitive MIMO Radars

In this chapter, we address the second contribution of this thesis, in which we are concerned with the high reflected power from the transmitting antenna of cognitive MIMO radars back to the amplification stage. We establish a signal-processing approach that can reduce the reflected power instead of reducing its effects. We show that the proposed approach controls the power level of the antenna reflection, improves the SINR of the target, and exhibits a lower computational burden than the standard method of waveform design of cognitive MIMO radars.

4.1 Introduction

Multi-input multi-output (MIMO) radars are distinguished from the phased array radars by their ability to transmit independent waveforms from the transmitting antenna elements. The advantages of MIMO radars over phased array radars include better spatial resolution, better parameter identification, improved performance for ground moving target identification (GMTI) [18, Ch.2], and enhanced detection performance due to their spatial diversity [19].

The problem of waveform design for MIMO radars has attracted considerable interest in the last decade, which resulted in three main trends in designing MIMO radar waveforms, as we mentioned in Section 2.3 and repeat here for convenience. The first approach is to control the spatial distribution of the transmitted power, through the spatial covariance matrix of the

waveforms. The second approach is to design the waveforms by the optimizing the ambiguity waveform properties such as the autocorrelation peak-to-sidelobe level ratio (PSLR), cross-correlation between the waveforms, Doppler and range resolutions, and Doppler tolerance. The third approach is concerned with the extended target model and aims to maximize the conditional mutual information (MI) between the target impulse response and the radar signal for optimal target detection.

In the aforementioned waveform design methods, the spatial cross-correlation between the different transmitted waveforms can range between perfect coherence, as in the phased array radar, and mutual orthogonality [123]. The spatial orthogonality is customarily realized in the fast-time domain, which is defined by the time samples transmitted from each antenna element within a single pulse [124–130]. This means that the waveform differs from one antenna element to another, but is repeated for each pulse. Another approach to achieve orthogonality is to apply phase coding on the slow-time domain, which is defined by the pulses within the coherent pulse interval (CPI); in this case the orthogonality is realized in the Doppler domain, which is known as Doppler division multiple access (DDMA) [131]. This Doppler division multiple access (DDMA) approach gives the radar designer the flexibility to use the same waveform for all antenna elements, while the required orthogonality is maintained after the Doppler processing in the receiver [132]. For the extended target models, research has focused on matching the transmitted waveforms to the TIR and maximizing the signal-to-interference plus noise ratio (SINR). This is achieved by solving an optimization problem that maximizes the SINR assuming prior information about the TIR [20]; however the orthogonality of the waveforms obtained by solving this optimization problem is not discussed.

For all of the discussed waveform design approaches, the efficiency of the radar transmitter using the obtained MIMO waveforms has not been explicitly considered. By efficiency we mean the effect of the scattering parameters (S-parameters) of the transmitting antenna array and its interaction with the used MIMO waveforms, which results in the reflection of a part of the transmitted power back to the amplification stage. This reflection is not only crucial to the efficiency of the radar system, but also to the durability of the microwave components preceding the antenna array. Specifically, the high reflected power from the transmitting antenna can damage the power module feeding the antenna [133]. The problem of power reflection has earned a lot of attention from the microwave and antenna design perspectives. From the signal processing standpoint, however, this problem has received scant coverage

in the literature. The problem of mutual coupling in MIMO transmitting antennas was first considered in [134] through the electromagnetic analysis of antenna arrays when all the elements of the array are active. This work was followed by [135,136], where electromagnetic simulations have shown the significant increase of antenna reflection in MIMO radars using phase coded signals assuming a point target model. In [137], also assuming a point target and phase coded signals, the transmitting antenna array was divided into groups, i.e., subarrays. Each subarray treated as a single element with a single waveform to reduce the reflection coefficient, or equivalently the voltage standing wave ratio (VSWR) of each antenna element. Later, the significance of the waveform design to the MIMO radar efficiency has been pointed out in [138]. Recently, the mutual coupling between the receiving antenna elements of MIMO radars has been studied in [139]. However, to the best of our knowledge, the mutual coupling between the transmitting antenna elements has not been considered in the design of MIMO waveforms in the literature.

In this chapter, we consider the problem of power reflection in cognitive MIMO radars from the transmitting antenna by proper design of power-efficient waveforms. In particular, we propose a cognitive MIMO radar system in which the transmitted waveforms are adapted to the TIR of the extended target of interest, but are also optimized to minimize the reflected power from the transmitting antenna array. The reflected signal is related to the transmitted signal by the S-parameter matrix. To achieve this, we formulate a new optimization problem with the ordinary objective of minimizing the interference power, but with adding a regularization term that includes the reflected power from the transmitted antenna array. The regularization term takes two forms. The first form uses the Euclidean norm (ℓ_2-norm) of the reflected signal and is solved using the Lagrange method. The second form utilizes the infinity norm (ℓ_∞-norm) of the reflected signal, which is a non-smooth function of the transmitted signal. In the latter case, and due to the non-differentiable regularization term, we propose using the proximal gradient method to solve the formulated optimization problem. To guarantee the orthogonality of the designed waveforms, the DDMA is employed. Monte Carlo simulations are used to evaluate our algorithm with the two proposed solutions using different figures of merit. The results show that the proposed algorithm improves the efficiency of the cognitive MIMO radar and has a lower complexity than the original cognitive waveform design method in [20], yet with acceptable SINR.

The rest of this chapter is organized as follows. Section 4.2 provides a background of MIMO radar that includes the mathematical signal model and the waveform designs of

MIMO and cognitive MIMO radars. The configurations of the MIMO antenna arrays and the problem of their mutual coupling are discussed in Section 4.3. In Section 4.4 we introduce the proposed waveform design approach for cognitive MIMO radars. The performance of the proposed approach is evaluated in Section 4.5. Section 4.6 concludes the chapter.

4.2 MIMO Radar Background

In this section, we provide a brief description of the mathematical model of the MIMO radar signal, the different schemes of the MIMO waveform orthogonality, and the waveform design of the cognitive MIMO radar for extended targets.

4.2.1 MIMO Radar Signal Model

Consider a MIMO radar system with N_T sparsely spaced transmitting antennas and N_R filled receiving antenna array. Assume a point target at azimuth angle θ_t and a uniform linear array (ULA), then the steering vectors at the transmitter and receiver are expressed as

$$\mathbf{a}_T(\theta_t) = \begin{bmatrix} 1 & e^{j2\pi \frac{d_T}{\lambda} \sin(\theta_t)} & \cdots & e^{j2\pi \frac{(N_T-1)d_T}{\lambda} \sin(\theta_t)} \end{bmatrix}^T \tag{4.1}$$

and

$$\mathbf{a}_R(\theta_t) = \begin{bmatrix} 1 & e^{j2\pi \frac{d_R}{\lambda} \sin(\theta_t)} & \cdots & e^{j2\pi \frac{(N_R-1)d_R}{\lambda} \sin(\theta_t)} \end{bmatrix}^T \ , \tag{4.2}$$

respectively, where d_T and d_R are the inter-element antenna spacings of the transmitting and receiving antenna arrays, respectively, and λ is the wavelength corresponding to the radar center frequency. Assuming N_T orthogonal waveforms and denoting $\tilde{\mathbf{f}}_i \in \mathbb{C}^{L_s}$ as the discrete baseband signal within a pulse duration T_p from the ith transmitting antenna, the transmitted pulses from the N_T elements are

$$\mathbf{X}^{(m)} = e^{j2\pi m T_r} \mathrm{diag}(\mathbf{a}_T(\theta_t))\mathbf{F} \tag{4.3}$$

where $\mathbf{F} = [\tilde{\mathbf{f}}_1^T \ \cdots \ \tilde{\mathbf{f}}_{N_T}^T]^T \in \mathbb{C}^{N_T \times L_s}$. Each antenna element receives a group of M coherent pulses resulting in a total of MN_R pulses. The mth received group of N_R pulses at the

receiving antenna array output is

$$\mathbf{Y}^{(m)} = e^{j2\pi(m-1)f_d}\mathbf{T}_p\mathbf{F} \tag{4.4}$$

where $\mathbf{T}_p = \text{diag}(\mathbf{a}_R)\tilde{\mathbf{T}}_p\text{diag}(\mathbf{a}_T)$, $\tilde{\mathbf{T}}_p \in \mathbb{C}^{N_R \times N_T}$ is the TIR whose (i,j)th component is the point target response between the ith receiving and the jth transmitting antenna elements, and f_d is the normalized Doppler frequency. The components of $\tilde{\mathbf{T}}_p$ represent the two-way attenuation and the phase difference between each pair of the receiving and the transmitting antenna elements. The received signal can be expressed in the vector form as

$$\mathbf{y}^{(m)} = e^{j2\pi(m-1)f_d}\text{blkdiag}(\mathbf{T}_p)\mathbf{f} \tag{4.5}$$

where $\text{blkdiag}(\mathbf{T}_p) \in \mathbb{C}^{N_R L_s \times N_T L_s}$ is the block diagonal matrix of $\mathbf{T}_p$ and $\mathbf{f} = \text{vec}(\mathbf{F})$.

In the case of the extended target model, the TIR between each transmitting antenna element and each receiving antenna element has a length L_t and the mth received pulse is expressed as

$$\mathbf{y}^{(m)} = e^{j2\pi(m-1)f_d}\mathbf{T}\mathbf{f} \tag{4.6}$$

where $\mathbf{T} \in \mathbb{C}^{N_R L_R \times N_T L_s}$ is the Toeplitz matrix of the extended TIR defined as

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_e(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{T}_e(1) & \mathbf{T}_e(0) & \ddots & \mathbf{0} \\ \mathbf{T}_e(L_t-1) & \cdots & \ddots & \mathbf{T}_e(0) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_e(L_t-1) \end{bmatrix} \tag{4.7}$$

where $\mathbf{T}_e(l) = \text{diag}(\mathbf{a}_R)\tilde{\mathbf{T}}_e(l)\text{diag}(\mathbf{a}_T)$, and $\tilde{\mathbf{T}}_e(l) \in \mathbb{C}^{N_R \times N_T}$ is the lth tap of the extended TIR[1].

[1]For the sake of generality, we assume here that the TIRs between all the pairs of the transmitting and receiving antenna elements are statistically independent .

4.2.2 Doppler Division Multiple Access in MIMO Radars

A characteristic feature of the MIMO radar is its spatial diversity achieved by utilizing N_T orthogonal waveforms. The orthogonality of the transmitted waveforms can be attained by: (1) time division multiple access (TDMA), (2) frequency division multiple access (FDMA), (3) code division multiple access (CDMA), and (4) Doppler division multiple access (DDMA) [140]. The first three methods are implemented in the fast-time domain, i.e., using the samples of the transmitted pulses. Conversely, the DDMA is implemented in the slow-time domain, i.e., from pulse to pulse within the coherent pulse interval (CPI). The CPI consists of a group of M identical pulses and each pulse is phase-coded, where the waveform of the pulses of each antenna element can be any of the conventional radar waveforms [137]. The orthogonality of the DDMA waveforms is then obtained by the Doppler processing in the receiver. The use of DDMA has the advantages of utilizing the full radar spectrum and transmission time, which are not offered by the FDMA and TDMA, respectively. In contrast to CDMA MIMO radar, the DDMA MIMO radar enjoys a simple hardware design by evading the need for a waveform generator for each transmitting antenna element [138].

Let the pulse repetition interval (PRI) of the radar be T_r and the pulse repetition frequency (PRF) $f_r = 1/T_r$, which is the maximum Doppler frequency of the radar. The full Doppler spectrum of the radar is divided to N_T sub-bands each of width f_c so that $f_c \leq f_r/N_T$. In DDMA, the transmitted waveform at the mth group of pulses in Eq. (4.3) becomes [137] $\mathbf{X}^{(m)} = \text{diag}(\mathbf{b}_T^{(m)})\text{diag}(\mathbf{a}_T(\theta_t))\mathbf{F}$, where $\mathbf{b}_T^{(m)} = [\exp(j2\pi\alpha_1 mT_r), \cdots, \exp(j2\pi\alpha_{N_T} mT_r)]$ and

$$\alpha_n = -\frac{f_c}{2}(N_T - 1 - 2n) \qquad n = 1, \cdots, N_T \tag{4.8}$$

In this manner, the DDMA establishes N_T orthogonal channels that are separated in the receiver using Doppler processing. Since the orthogonality is achieved in the slow-time (Doppler) domain and not the fast-time domain, $\tilde{\mathbf{f}}_i$, $i = 1, \cdots, N_T$, can be identical and chosen to achieve desirable radar signal properties as the linear frequency modulated (LFM) signal.

4.2.3 Cognitive MIMO Waveform Design for Extended Targets

A salient aspect of the cognitive radars is to use the environmental information to adapt the transmitted waveform to match the TIR(s) of the extended target(s) of interest [141]. Assuming known TIR and clutter statistics, a method has been proposed in [20] to jointly design the transmitting waveforms and receive filter impulse response of the cognitive MIMO radar with the aim of maximizing the SINR. The output of the receive filter at the mth pulse is expressed as

$$\mathbf{r}^{(m)} = e^{j2\pi(m-1)f_d}\mathbf{h}^H\mathbf{T}\mathbf{f} + \mathbf{h}^H\mathbf{C}\mathbf{f} + \mathbf{h}^H\mathbf{n} \tag{4.9}$$

where $\mathbf{h} \in \mathbb{C}^{N_R L_R}$ is the impulse response of the receive filter, $\mathbf{C} \in \mathbb{C}^{N_R L_R \times N_T L_s}$ is the clutter impulse response, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_R L_R})$. The clutter impulse response is defined as

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{C}_e(0) & \mathbf{C}_e(-1) & \cdots & \mathbf{C}_e(-L_s+1) \\ \mathbf{C}_e(1) & \mathbf{C}_e(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{C}_e(0) \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{C}_e(L_R-1) & \mathbf{C}_e(L_R-1) & \cdots & \mathbf{C}_e(L_t-1) \end{bmatrix} \tag{4.10}$$

where $\mathbf{C}_e(l) \in \mathbb{C}^{N_R \times N_T}$ is the lth tab of the clutter impulse response between the ith receiving antenna element and the jth transmitting antenna element. The SINR at the filter output is

$$\chi(\mathbf{f}, \mathbf{h}) \triangleq \frac{|\mathbf{h}^H\mathbf{T}\mathbf{f}|^2}{E[|\mathbf{h}^H\mathbf{C}\mathbf{f}|^2] + E[|\mathbf{h}^H\mathbf{n}|^2]} \tag{4.11}$$

Both $\mathbf{f}$ and $\mathbf{h}$ are jointly optimized to maximize the $\chi(\mathbf{f}, \mathbf{h})$. The resulting optimization problem is generally nonconvex and it is solved iteratively by solving $\mathbf{h}$ in terms of $\mathbf{f}$ and vice versa. This problem can be recast as two minimum variance distortionless response (MVDR) problems for both $\mathbf{h}$ and $\mathbf{f}$. The first MVDR problem to solve for $\mathbf{f}$ is

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{T}\mathbf{f} = 1 \end{aligned} \tag{4.12}$$

The solution for $\mathbf{f}$ is

$$\mathbf{f} = \alpha_f(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})^{-1}\mathbf{T}^H\mathbf{h} \tag{4.13}$$

where α_f is a scalar, $\mathbf{R}_n \triangleq E[\mathbf{n}\mathbf{n}^H]$, and $\mathbf{R}_{c,h} \triangleq E[\mathbf{C}^H\mathbf{h}\mathbf{h}^H\mathbf{C}]$, which is related to the clutter covariance matrix $\mathbf{R}_c(m) = E[\text{vec}(\mathbf{C}(n))\text{vec}(\mathbf{C}(n-m))^H]$, $\infty < n, m < \infty$, as in [142, Lemma 2]. The solution in Eq. (4.13) is then normalized with respect to $\|\mathbf{f}\|_2$, so that the scalar α_f can be neglected. The MVDR problem for $\mathbf{h}$ can be expressed as

$$\begin{aligned} \min_{\mathbf{h}} \quad & \mathbf{h}^H(\mathbf{R}_{c,f} + \mathbf{R}_n)\mathbf{h} \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{T}\mathbf{f} = 1 \end{aligned} \tag{4.14}$$

from which, the following solution is obtained

$$\mathbf{h} = \alpha_h(\mathbf{R}_{c,f} + \mathbf{R}_n)^{-1}\mathbf{T}\mathbf{f} \tag{4.15}$$

where α_h is a scalar that satisfies the equality constraint in Eq. (4.14), $\mathbf{R}_{c,f} \triangleq E[\mathbf{C}\mathbf{f}\mathbf{f}^H\mathbf{C}^H]$, and $\mathbf{R}_{c,f}$ is related to $\mathbf{R}_c(m)$ as shown in [142, Lemma 1]. Note that the scalar α_h does not affect the objective function in Eq. (4.11), hence it can be neglected.

4.3 MIMO Antenna Arrays

In this section, we discuss the different configurations of MIMO antenna arrays and give a brief review of the mutual coupling of antenna arrays and the different reflection coefficients.

4.3.1 MIMO Virtual Antenna Array

A key characteristic of the MIMO radar is to form a virtual array that improves the spatial resolution of the MIMO radar system. By transmitting orthogonal waveforms from the transmitting antenna array, the MIMO radar can form a virtual array, whose elements' locations result from the convolution between the locations of the elements of the transmitting and receiving antenna elements. The resulting virtual array is larger than the total number of receiving and transmitting antenna elements [143]. Conventionally, if the transmitting antenna elements in a ULA are spaced by $d_T = N_R d_R$, the resulting virtual antenna array

has effectively $N_T N_R$ elements with $N_T N_R - 1$ normalized[2] aperture length, which improves radar spatial resolution. [144]. However, this configuration requires large N_T and N_R to achieve high aperture lengths.

One approach to achieve high spatial resolution using lower number of antennas is the minimum redundancy array (MRA) [144]. Let the locations of the elements of the transmitting and the receiving antenna arrays be $d_T^{(m)}$ and $d_R^{(n)}$, respectively, where $m = 1, \cdots, N_T$ and $n = 1, \cdots, N_R$. The concept of the MRA is to use the minimum number of antenna elements as long as the spacings between the pairs of antenna elements take all the integer values between 0 and N_V, where N_V is the required normalized aperture length of the virtual array [144].

The majority of the research efforts in implementing the MRA were focused on employing optimization techniques to optimally configure the receiving and transmitting antenna arrays as in [144–146]. Distinctively, the work in [147] proposed employing the difference basis and simple perfect cyclic difference sets (SPCDS) to find the MRA for the MIMO radars. As the number theory constitutes, a difference basis for the segment $[0, P]$, $P \in \mathbb{N}$, is a set of K integers ($K < P$) such that all the integers $0, \cdots, P$ can be expressed using this set [148]. An SPCDS $W(V, K)$ is a set of K integers chosen from $\{0, 1, \cdots, V\}$, where $K(K-1) = V - 1$, such that any one of the ordered differences between the elements of $W(V, K)$ is not repeated [149]. As indicated in [147], the locations of both the transmitting and receiving antenna elements are determined as follows:

1. The locations of the transmitting antenna elements $d_T^{(m)}$, $m = 1, 2, \cdots, N_T$, are the elements of an SPCDS with parameters V, M with the first element is 0.

2. The locations of the receiving antenna elements $d_R^{(n)}$, $n = 1, 2, \cdots, N_R$, are the elements of a set $\{p_n \odot V\}$, where $\odot$ denotes set multiplication, $\{p_n\}$ is a difference basis for a segment $[0, P]$, and P is chosen such that the resulting difference basis has N_R elements.

3. The resulting virtual minimum redundancy array (VMRA), $d^{(k)}$, is expressed as

$$\{d^{(k)}\} = \{d_T^{(m)}\} \oplus \{d_R^{(n)}\} \tag{4.16}$$

where $k = 1, \cdots, N_T N_R$; $m = 1, \cdots, N_T$; $n = 1, \cdots, N_R$; and $\oplus$ denotes the addition of sets.

[2]Normalized with respect to $\lambda/2$.

As has been proven in [150], the resulting VMRA is a difference basis for $[0, N_V]$, where

$$N_V = V(P+1) - d_T^{(M)} - 1 \tag{4.17}$$

As shown in [147, 150], the resulting virtual MRA has a larger normalized aperture length than the conventional length $N_T N_R$, yet with a lower number of antenna elements.

To illustrate the idea of MRA, consider $d_T^{(m)} = \{0, 1, 3\}$ and $d_R^{(m)} = \{0, 6, 13, 40, 60\}$. The resulting virtual array is $\{0, 1, 3, 6, 7, 9, 13, 14, 16, 40, 41, 43, 60, 61, 63\}$, which has a normalized effective aperture length of 63 using a total of 8 antenna elements. To achieve the same normalized aperture length using the conventional ULA, we need a total of 16 antenna elements. However, as the MRA reduces the inter-element spacing in the transmitting antenna array, the mutual coupling between each pair of the N_T antenna elements increases compared to the ULA.

4.3.2 Mutual Coupling of Transmitting Antenna Arrays

The input-output relations of an N_T-element antenna array are described using S-parameters, which are represented by the matrix $\mathbf{S} \in \mathbb{C}^{N_T \times N_T}$. In particular, (i, j)th element of $\mathbf{S}$, $1 \leq i, j \leq N_T$, represents the power reflected from element i to element j. The elements of $\mathbf{S}$ are given by [151]

$$S_{pn} = \left.\frac{b_p}{a_n}\right|_{a_i=0} \qquad 1 \leq p, n, i \leq N_T, i \neq n \tag{4.18}$$

where a_n is the input voltage to element n and b_p is the reflected voltage from element n to element p, with all the inputs to all ports except n are inactive. The reflected signal from an N_T-element antenna array is related to the input signal as $\mathbf{F}_{ref} = \mathbf{SF}$. The scattering parameters are related to the reflection coefficients of the antenna elements by

$$\Gamma_i = S_{ii} \qquad 1 \leq i \leq N_T \tag{4.19}$$

where $0 < \Gamma_i < 1$. However, the assumption of passive ports except one in the measurement of Γ is not adequate in MIMO operation, where all the elements could be simultaneously active. In this case, the mutual coupling between antenna elements makes the active properties of the array different from those measured under passive conditions [152]. The active

reflection coefficient (ARC) of the ith antenna element is given by [153]

$$\Gamma_i^a = \frac{b_i}{a_i} = \frac{\sum_{j=1}^{N_T} S_{ij} a_j}{a_i} \tag{4.20}$$

As reported in [154], the ARC can exceed 1 for some antenna elements as the reflected power is formed by coupling the reflected power from different elements. To characterize the performance of the whole antenna array, the total active reflection coefficient (TARC) is used [155, Eq. (12)]

$$\Gamma^t = \frac{\sqrt{\sum_{i=1}^{N_T} |b_i|^2}}{\sqrt{\sum_{i=1}^{N_T} |a_i|^2}} \tag{4.21}$$

where $0 < \Gamma^t < 1$. The TARC has been used as a figure of merit for MIMO antenna arrays as in [156, 157].

4.3.3 Microwave Techniques for Protection against High Reflection

Before introducing the proposed method to reduce the reflected power from the transmitting antenna arrays to the preceding amplification stage, it is important to discuss, briefly, the microwave techniques developed to protect the amplification stage from this reflected power. Radio frequency power amplifiers data sheets show that a reflection coefficient of 0.8 can be damaging to the output stage of the amplifier [158]. Even if the reflection is not high enough to damage the output stage, it can cause a reduction in the output power [159]. Conventionally, isolators have been employed between the antenna and the power amplifier; however, they have a large space and weight requirements, which may make them not suitable for many radar applications, besides their insertion loss. Additionally, conventional protection methods also use clipping diodes, but they introduce parasitic capacitance that make them not appropriate for RF applications [160].

The main challenge to microwave techniques to provide protection to the power amplifier in the presence of high reflection is to achieve a balance between the required protection and acceptable insertion loss [161]. Another challenge is to achieve fast response time in the case of varying reflection coefficients. In the case of cognitive MIMO radars, the reflection coefficients change on a pulse-to-pulse basis, which means they can change in less than

a fraction of a millisecond. To face these challenges, closed- and open-loop circuits have been proposed to sense the output of the power amplifier and change its operating point[3] accordingly [163]. This is implemented by controlling the gain of the amplifier or the driving input power if the reflected power exceeds predefined limits. Nevertheless, these methods require complex design of the power amplifier and its driving stage to maintain the required stability at different input and supply levels [163]. Moreover, sensing the output of the amplifier, which is implemented using directional couplers, leads to losses of the output power and, consequently, a lower transmitter efficiency. In addition, closed-loop techniques require delicate design of the loop gain to cope with the nonlinear operation of the amplifier and open-loop techniques need higher cost [164]. Finally, class-A amplifiers show robust performance against high reflection; however, this comes at the cost of bulky modules, high cost, and low efficiency that may drop down to %50 [165]. Generally, the methods reported in the literature on the microwave techniques of the protection from high reflections, with the disadvantages we have just reported, can protect the amplifier in the case of reflection coefficient less than 1, but not with an ARC greater than unity as may be encountered in cognitive MIMO radars, as we will show shortly.

4.4 Proposed waveform design for cognitive MIMO radar

In this section, we propose a novel approach to design power-efficient waveforms for cognitive MIMO radars. This new approach takes into account the reflection characteristics of the radar antenna array as well as the TIR of the target of interest. This approach reduces the reflected power from the transmitting antenna rather than reducing its effects, as offered by the microwave techniques. In this context, we formulate two optimization problems with two different regularization terms and provide the solutions to both of them.

To maximize the SINR, the problem in Eq. (4.12) minimizes the interference term of Eq. (4.11). To simultaneously minimize the reflected signal from the antenna array, we propose adding a regularization term to the objective function of Eq. (4.12) as follows

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \gamma_1 g(\mathbf{S}_d\mathbf{f}) \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{T}\mathbf{f} = 1 \end{aligned} \tag{4.22}$$

[3] An operating point of the amplifier is the intersection between the load line of the amplifier and an output characteristic at certain biasing conditions [162].

where γ_1 is a regularization parameter, $\mathbf{S}_d = \text{blkdiag}(\mathbf{S})$, $\mathbf{S}_d \in \mathbb{C}^{N_T L_s \times N_T L_s}$, and $g(\mathbf{S}_d\mathbf{f})$ is a function of the reflected signal from the N_T antenna elements at all samples of each pulse. In the following, we propose using two different choices for $g(\mathbf{S}_d\mathbf{f})$, which result in two different optimization problems.

4.4.1 Solution Using Lagrange Method with ℓ_2-Norm Regularization

By using the squared ℓ_2-norm as the regularization function $g(\mathbf{S}_d\mathbf{f})$, Eq. (4.22) becomes

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \gamma_1\|\mathbf{S}_d\mathbf{f}\|_2^2 \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{T}\mathbf{f} = 1 \end{aligned} \tag{4.23}$$

The problem in Eq. (4.23) is in the form of MVDR problem, which can be solved using Lagrange method. The Lagrangian function is defined as

$$\mathfrak{L}(\mathbf{f}, \lambda) \triangleq \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \gamma_1\|\mathbf{S}_d\mathbf{f}\|_2^2 + \lambda_1(\mathbf{h}^H\mathbf{T}\mathbf{f} - 1) \tag{4.24}$$

where λ_1 is the Lagrangian multiplier. By differentiating $\mathfrak{L}(\mathbf{f}, \lambda)$ with respect to $\mathbf{f}$ and equating it to zero we obtain

$$\mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s}) + \gamma_1\mathbf{f}^H\mathbf{S}_d^H\mathbf{S}_d + \lambda_1\mathbf{h}^H\mathbf{T} = 0 \tag{4.25}$$

Therefore, the solution for the waveform $\mathbf{f}$ is

$$\mathbf{f} = -\lambda_1 \left((\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s} + \gamma_1\mathbf{S}_d^H\mathbf{S}_d)^{-1}\right)^H \mathbf{T}^H\mathbf{h} \tag{4.26}$$

The scalar λ_1 can be neglected as $\mathbf{f}$ in Eq. (4.26) is then normalized with respect to $\|\mathbf{f}\|_2$ as in [20]. The detailed steps of the proposed waveform design using the Lagrange method are shown in Algorithm 3. For faster convergence, γ_1 is set to grow with each iteration as discussed later in Section 4.5.

Algorithm 3 Waveform design using Lagrange method

Input: $\mathbf{R}_{c,f}, \mathbf{R}_v, \mathbf{T}, \mathbf{S}_d, \gamma_1^{(1)}$
Initialize: $\mathbf{f}^{(1)}, \mathbf{h}^{(1)}$, SINR
Output: $\mathbf{f}^{(k)}, \mathbf{h}^{(k)}$
while $\epsilon < 1$ **do**
Calculate $\mathbf{h}^{(k)}$ according to Eq. (4.15)
Calculate $\mathbf{R}_{c,h}^{(k)}$
Calculate $\mathbf{f}^{(k)}$ according to Eq. (4.26)
$\mathbf{f}^{(k)} \leftarrow \mathbf{f}^{(k)}/\|\mathbf{f}^{(k)}\|_2$
Calculate $\mathbf{R}_{c,f}^{(k)}$
Calculate SINR according to Eq. (4.11)
$\epsilon \leftarrow$ Improvement in SINR
$k \leftarrow k+1$
Calculate $\gamma_1^{(k)}$
end while

4.4.2 Solution Using Proximal Gradient with ℓ_∞ Norm Regularization

Using the ℓ_∞-norm (maximum norm) as the regularization term $g(\mathbf{S}_d\mathbf{f})$ in Eq. (4.22), we obtain

$$\begin{aligned} \min_{\mathbf{f}} \quad & \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{T}\mathbf{f} = 1 \end{aligned} \tag{4.27}$$

While minimizing the ℓ_2-norm reduces the average reflected power from the antenna array, minimizing the ℓ_∞-norm reduces the reflected power of each single sample reflected from each antenna element. This means a higher reduction to the reflected power at all pulses and all antenna elements, which motivates the use of the ℓ_∞-norm as a regularization term. The objective function in Eq. (4.27) contains two terms, which are convex. However, the regularization term $\gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty$ is nonsmooth, which impedes the solution of Eq. (4.27) using Lagrange or conventional gradient methods. To solve this problem we propose using the proximal method [58].

As mentioned in Section 2.6, for a convex function $g(\mathbf{f})$, the proximal operator $\text{prox}_g(\mathbf{f})$

is defined by [56]

$$\text{prox}_g(\mathbf{f}) \coloneqq \underset{\mathbf{v}}{\text{argmin}} \left(g(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}-\mathbf{f}\|_2^2 \right) \qquad \forall \mathbf{v} \in \mathbb{C}^{N_T L_s} \tag{4.28}$$

For a scaled function $\gamma_2 g(\mathbf{f})$, the proximal operator is defined as

$$\text{prox}_g(\mathbf{v}) \coloneqq \underset{\mathbf{v}}{\text{argmin}} \left(g(\mathbf{f}) + \frac{1}{2\gamma_2}\|\mathbf{f}-\mathbf{v}\|_2^2 \right) \tag{4.29}$$

The proximal operator, as perceived from Eq. (4.28), minimizes g, but at the same time considers the Euclidean distance of the solution from $\mathbf{f}$.

First we convert the constrained optimization problem in Eq. (4.27) to an unconstrained optimization problem. To achieve this, we propose using the quadratic penalty method [166], that is

$$\min_{\mathbf{f}} \quad \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty + \frac{\mu}{2}(\mathbf{h}^H\mathbf{T}\mathbf{f}-1)^2 \tag{4.30}$$

where $\mu > 0$ is the penalty parameter. The objective function in Eq. (4.30) can be written as

$$\min_{\mathbf{f}} \quad u(\mathbf{f}) + w(\mathbf{f}) \tag{4.31}$$

where $u(\mathbf{f}) = \mathbf{f}^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_T L_s})\mathbf{f} + \frac{\mu}{2}(\mathbf{h}^H\mathbf{T}\mathbf{f}-1)^2$ and $w(\mathbf{f}) = \gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty$. Discernibly, $u(\mathbf{f})$ is convex and smooth, while $w(\mathbf{f})$ is convex but non-smooth. For solving this problem, we propose applying the proximal gradient method. This method can iteratively find the minimizer of $u(\mathbf{f}) + w(\mathbf{f})$ by calculating the so-called proximal operator of $w(\mathbf{f})$ as a function of $\nabla u(\mathbf{f})$, thus tackling the problem of a non-smooth $w(\mathbf{f})$. Using the proximal gradient method, the minimizer of Eq. (4.30) at the $(k+1)$th iteration is [58, (4.6)]

$$\mathbf{f}^{(k+1)} \coloneqq \text{prox}_{\beta w}(\mathbf{f}^{(k)} - \beta^{(k)}\nabla u(\mathbf{f}^{(k)})) \tag{4.32}$$

where k is the iteration number, $\beta^{(k)} > 0$ is the step size at the kth iteration, and $\text{prox}_{\beta w}$ is the proximal operator of $\beta^{(k)} w(\mathbf{f})$. The solution of Eq. (4.31), which includes the proximal operator of $\gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty$, is given by Proposition 4.1, whose proof is provided in Appendix D.

Proposition 4.1. *Consider the following optimization problem*

$$\min_{\mathbf{f}} \quad u(\mathbf{f}) + w(\mathbf{f}) \tag{4.33}$$

where $u(\mathbf{f}) = \mathbf{f}^H(\mathbf{R}_{c,h}+\mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_TL_s})\mathbf{f}+\frac{\mu}{2}(\mathbf{h}^H\mathbf{T}\mathbf{f}-1)^2$ *and* $w(\mathbf{f}) = \gamma_2\|\mathbf{S}_d\mathbf{f}\|_\infty$. *The waveform* $\mathbf{f}$ *in Eq.* (4.33) *at the* $(k+1)$*th iteration is*

$$\mathbf{f}^{(k+1)} = \mathbf{d}^{(k)} - (\gamma_2\mathbf{S}_d^{-1})\mathrm{Proj}_{\|\cdot\|_1<1}\left(\frac{\mathbf{S}_d\mathbf{d}^{(k)}}{\gamma_2}\right) \tag{4.34}$$

where $\mathbf{d}^{(k)} = \mathbf{f}^{(k)} - \beta^{(k)}\nabla u(\mathbf{f})$, $\nabla u(\mathbf{f}) = (\mathbf{f}^{(k)})^H(\mathbf{R}_{c,h} + \mathbf{h}^H\mathbf{R}_n\mathbf{h}\mathbf{I}_{N_TL_s}) + \mu(\mathbf{h}^H\mathbf{T}\mathbf{f}^{(k)} - 1)\mathbf{h}^H\mathbf{T}$, *and* $\mathrm{Proj}_{\|\cdot\|_1<1}(\cdot)$ *is the projection operator of the argument on the unit* ℓ_1*-norm ball.*

As observed in Proposition 4.1, the computation of $\mathrm{Proj}_{\|\cdot\|_1<1}\left(\frac{\mathbf{S}_d\mathbf{d}^{(k)}}{\gamma_2}\right)$ is crucial to the proposed waveform design. In the following, we discuss how to perform this computation.

4.4.2.1 Computation of the Projection onto the unit ℓ_1-Ball

This problem has been discussed for real-valued variables in the context of the matrix estimation problem [167], and several algorithms have been proposed [168–170]. In this chapter, we adopt a simple algorithm that has a straightforward geometrical interpretation.

As shown in [171], the projection onto the unit ℓ_1-ball is similar to the projection onto a simplex[4] of the same dimension. For the simplex projection of a vector $\mathbf{p} \in \mathbb{R}^Q$, $\mathrm{Proj}_\Delta(\mathbf{p})$, there exists a unique scalar $\tau \in \mathbb{R}$ such that [170, Proposition 2.2]

$$\mathrm{Proj}_\Delta(\mathbf{p}(i)) = \max\{\mathbf{p}(i) - \tau, 0\} \qquad \forall i = 1, \cdots, Q \tag{4.35}$$

For the projection onto the unit ball, we seek finding τ such that $\sum_{i=1}^{Q}\max\{\mathbf{p}(i)-\tau, 0\} = 1$. Let $\mathscr{I}$ be the set of indices i, for which $\mathrm{Proj}_\Delta(\mathbf{p}(i)) > 0$. Then we have

$$\sum_{i\in\mathscr{I}}(\mathbf{p}(i) - \tau) = 1 \tag{4.36}$$

[4]The simplex is the generalization of triangles and tetrahedra to any dimension. A more formal definition of the simplex is that it is the convex hull of its vertices [172]

Hence [170],

$$\tau = \sum_{i \in \mathscr{I}} (\mathbf{p}(i) - 1)/\|\mathscr{I}\| \tag{4.37}$$

where $\|\cdot\|$ denotes the cardinality of the set. Based on equations (4.35) and (4.37), a simple algorithm has been proposed in [173] for real-valued projections. For complex-valued projections, as the problem at hand, the projection is computed for the vector of the components' modula of the original complex vector onto the real unit ℓ_1-ball, then the complex soft thresholding operation is applied on the real projection [174]. Denoting $\mathbf{p}^{\perp}$ as the real $\text{Proj}_{\|\cdot\|_1<1}(\mathbf{p})$, then, for $\underline{\mathbf{p}} \in \mathbb{C}^Q$, the complex projection $\underline{\mathbf{p}}^{\perp} = \text{Proj}_{\|\cdot\|_1<1}\left(\underline{\mathbf{p}}\right)$ is given by

$$\underline{\mathbf{p}}^{\perp}(i) = \text{csgn}(\mathbf{p}(i))\mathbf{p}^{\perp}(i) \qquad \forall i = 1, \cdots, Q \tag{4.38}$$

where $\text{csgn}(\cdot)$ is the complex signum function. In this way, the phases of the input vector components are preserved, while the main computations are performed on the vector of components' modula of the original complex vector. The steps of computing $\text{Proj}_{\|\cdot\|_1<1}(\mathbf{p})$ are indicated in Algorithm 4. The complexity of the algorithm is determined by the complexity of the sort operation. If merge sort algorithm is used to sort the components of $|\underline{\mathbf{p}}|$, the worst case complexity is $\mathcal{O}(Q\log(Q))$ [120].

Algorithm 4 Calculation of $\text{Proj}_{\|\cdot\|_1<1}\left(\underline{\mathbf{p}}\right)$

Input: $\underline{\mathbf{p}} \in \mathbb{C}^Q$
Output: $\underline{\mathbf{p}}^{\perp}$
  Sort $|\underline{\mathbf{p}}|$ to $\mathbf{u} : u_1 \geq \cdots \geq u_Q$
  for $i = 1$ **to** Q **do**
    $\mathbf{r}(i) = \sum_{j=1}^{i}(u_j - 1)/i$
  end for
  Find $K = \underset{i}{\text{argmax}}\, \mathbf{r}(i) < \mathbf{u}(i)$
  $\tau = \sum_{k=1}^{K} u_k - 1/K$
  for $i = 1$ **to** Q **do**
    $\mathbf{p}^{\perp}(i) = \max\{\mathbf{p}(i) - \tau, 0\}$
    $\underline{\mathbf{p}}^{\perp}(i) = \text{csgn}(\underline{\mathbf{p}}(i))\mathbf{p}^{\perp}(i)$
  end for

The proposed waveform design method is detailed in Algorithm 5. By comparing the two

proposed waveform designs using Lagrange and proximal gradient methods in Algorithms 3 and 5, respectively, we can highlight two main points. First, the proximal gradient method establishes a separate inner loop for the waveform design that is concerned with minimizing the interference, the reflection, and the deviation from the constraint. The outer loop is responsible for improving the SINR with $\mathbf{h}$ is taken into account. This separation of loops offers a better control over the reflected power than the Lagrange method, besides the advantage offered by the ℓ_∞-norm over the ℓ_2-norm. Second, the proximal gradient method avoids the evaluation of $\mathbf{R}_{c,s}$ for each iteration k, which is expected to reduce the complexity of the design using the proximal gradient method. More details about the complexity of both algorithms are provided in Section 4.5.

Algorithm 5 Waveform design using proximal gradient

Input: $\mathbf{R}_{c,f}, \mathbf{R}_v, \mathbf{T}, \mathbf{S}_d, \gamma_1, \beta^{(1)}$
Initialize: $\mathbf{f}, \mathbf{h}$, SINR, $k = 1$
Output: $\mathbf{f}^{(k)}, \mathbf{h}^{(k)}$
  while $\epsilon < 1$ **do**
    Calculate $\mathbf{h}$ according to Eq. (4.15)
    Calculate $\mathbf{R}_{c,h}$
    while $\zeta < 1$ **do**
      Calculate $\nabla u(\mathbf{f}^{(k)}), \mathbf{d}^{(k)}$ according to Proposition 4.1.
      Calculate $\text{Proj}_{\|\cdot\|_1<1}\left(\frac{\mathbf{S}_d\mathbf{d}^{(k)}}{\beta^{(k)}\gamma_2}\right)$ as in Algorithm 4.
      Calculate $\mathbf{f}^{(k+1)}$ as in Eq. (4.34)
      $\zeta \leftarrow$ reduction of the objective function in Eq. (4.30).
      $k \leftarrow k + 1$.
      Calculate $\beta^{(k)}$.
    end while
    $\epsilon \leftarrow$ Improvement in SINR.
  end while

4.4.2.2 Convergence of the Proposed Proximal-Based Algorithm

It is well-known that if $u(\mathbf{f})$ is convex and $\nabla u(\mathbf{f})$ is Lipschitz continuous with constant q, the convergence rate of the proximal gradient method in solving Eq. (4.33) is $\mathcal{O}(1/k)$ if $\beta^{(k)} = \beta \in (0, 1/q]$ [175, Theorem 1 (a)] [58, 176]. Obviously, $u(\mathbf{f})$ is convex; we also prove its Lipschitz continuity in Appendix E. Therefore, the convergence rate of the proposed algorithm is

linear. The step size can also be found using line search with the same convergence rate as indicated in [177], where

$$\beta^k = \eta\beta^{k-1} \qquad \eta \in (0, 1] \tag{4.39}$$

with $k \geq 1$ and $\beta^{(1)} \in (0, 1/q]$.

4.5 Performance Evaluation

In this section, we evaluate the performance of the two proposed method using Monte Carlo simulations and we compare it to two benchmarks. The first is the standard waveform design method for extended targets in cognitive radars introduced in [20] and briefed in Section 4.2, which does not consider the mutual coupling of the transmitting antenna. Since, as far as we know, there is no waveform design in the literature that takes into account the reflection coefficients of the antenna elements, we opt to comparing the proposed methods with the subarray solution used in [137]. It should be noted that the subarray configuration has been used in [137] with point targets and with standard radar waveforms without any adaptation done to the transmitted waveform. However, we will use the subarray configuration with extended targets and the transmitted waveforms that are adapted to the TIR.

4.5.1 Simulation Setup

We consider two sizes of the transmitting antenna array: 4 elements and 8 elements. The transmitting antenna is a linear MRA, as discribed in Subsection 4.3.1, where the locations of the two antenna arrays are $\{0, 1, 3, 7\}$ and $\{0, 9, 10, 12, 16, 27, 35, 40\}$. It should be noted that there are different configurations for each assumed antenna size using SPCDS. However, the configurations we use in this work have the minimum spatial dimension among other configurations. The antenna elements are assumed to be half-wave dipoles, and the impedance values of the resulting array elements are calculated as [178]

$$Z_{ii} = 30[\ln(\xi 2\pi) - C_i(2\pi) + jS_i(2\pi)] \qquad 1 \leq i \leq N_T \tag{4.40}$$

where ξ is the Euler constant and C_i and S_i are the cosine and sine integral functions. The coupling impedance values $Z_{ij} = R_{ij} + jX_{ij}$, $i \neq j$, are given by

$$R_{ij} = 30[2C_i(\mu_0) - C_i(\mu_1) - C_i(\mu_2)] \qquad 1 \leq i \leq N_T, \quad i \neq j \tag{4.41a}$$

$$X_{ij} = -30[2S_i(\mu_0) - S_i(\mu_1) - S_i(\mu_2)] \qquad 1 \leq i \leq N_T, \quad i \neq j \tag{4.41b}$$

where

$$\begin{aligned} &\mu_0 = 2\pi d, \mu_1 = 2\pi\sqrt{d^2 + 0.25} + 0.5 \\ &\mu_2 = 2\pi\sqrt{d^2 + 0.25} - 0.5 \end{aligned} \tag{4.42}$$

where d is the spacing between the ith and jth antenna elements normalized with respect to the wavelength corresponding to the center frequency. The S-parameters matrix $\mathbf{S}$ is obtained from the impedance matrix $\mathbf{Z}$ by [151]

$$\mathbf{S} = \frac{\mathbf{Z} - Z_0\mathbf{I}_{N_T}}{\mathbf{Z} + Z_0\mathbf{I}_{N_T}} \tag{4.43}$$

where Z_0 is the matched load impedance at which there is no reflection and it is customarily taken as $Z_0 = 50\Omega$.

Regarding the TIR, we consider two different distributions. The first distribution is the Gaussian model, which is considered widely in the literature [179, 180]. The Gaussian TIR between each pair of the transmitting and receiving antenna elements is generated as a random vector distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{L_t})$. The second distribution is the K-distribution [181], which has not been explored in the literature on the cognitive radar applications, neither any other non-Gaussian distribution. The K-distributed TIR between each pair of the transmitting and receiving antenna elements is generated as a spherical invariant random vector (SIRV), as shown in Section 3.3 and [182, 183]. For both distributions, the TIR is conventionally assumed to be known by the radar system [20, 22, 23, 180]. If the TIR is unknown, it is estimated from the received data[5] as in [179, 184]. For a comprehensive evaluation of the proposed method, we used 1000 Monte Carlo simulation trials each with a

[5]In Chapter 5, we discuss in details the problem of TIR estimation.

different TIR.

The initial radar waveform is assumed to be LFM signal, as has been also assumed in [20, 179]. The signal bandwidth is 250 KHz and sampled at 0.5 MHz. The pulse width is $10\mu s$ and the PRI is $500\mu s$, which corresponds to a maximum Doppler of 2 KHz. For the proposed Lagrange-based method, the regularization parameter is taken as $\gamma_1^{(k)} = 3\gamma_1^{(k-1)}$, with $\gamma_1^{(1)} = 10$. For proximal-based method, the step size is $\beta^{(k)} = 0.99\beta^{(k-1)}$ with $\beta^{(1)} = 10$, the quadratic penalty parameter $\mu = 0.001$, and the regularization parameter $\gamma_2 = 0.5$. These values have been chosen using simulations to achieve the best balance between the SINR improvement and the reflection reduction. However, it should be emphasized that while these are the recommended values we found through the simulations, the proposed algorithms are not sensitive to these values. This means that the deviation from these values does not cause a significant effect on performance.

4.5.2 Performance Evaluation with 4-Element Transmitting Antenna Array

We first consider the case of 4-element transmitting antenna array in the case of Gaussian TIR. Fig. 4.1 shows the worst case scenario among the 1000 simulated Gaussian TIRs for the four considered algorithms. The figure shows the values of Γ^a for each antenna element with some elements having $\Gamma^a > 1$ at some pulses for the standard waveform design method. The subarray configuration offers lower Γ^a when used with point targets as reported in [137], i.e., standard DDMA with standard radar signals [137]. However, when it is used in cognitive radars with extended targets it can be greater than the full MIMO configurations at some pulses and some antenna elements. Regarding the proposed algorithms, we can observe that the Lagrange-based algorithm does not guarantee a lower Γ^a for all pulses and all antenna elements. This can be attributed to the criterion of lowering the average reflected power, which does not guarantee the reduction of the reflection at each pulse and each antenna element. Moreover, the Lagrange method works on two objectives simultaneously, the SINR and Γ^a, which limits its ability to obtain a waveform that satisfies both objectives.

On the other hand, the proximal gradient method guarantees lower Γ^a at all elements and all pulses, such that Γ^a barely exceeds 0.6 at maximum. The relative reduction achieved by the proximal-based proposed method is approximately between 27.24% and 84.64%. This superior performance of the proximal-based method can be explained by the opposite of the reasons behind the unsatisfactory performance of the Lagrange-based method. That is,

the proximal-based method in Algorithm 5 deals with lowering the reflected power from the antenna array as a separate objective in an inner loop, while dealing with the SINR separately in the outer loop. This gives the proximal-based method more flexibility in minimizing Γ^a. In addition, the ℓ_∞-norm metric used to measure the reflected power in this algorithm guarantees lowering Γ^a, iteratively, for each pulse and each antenna element.

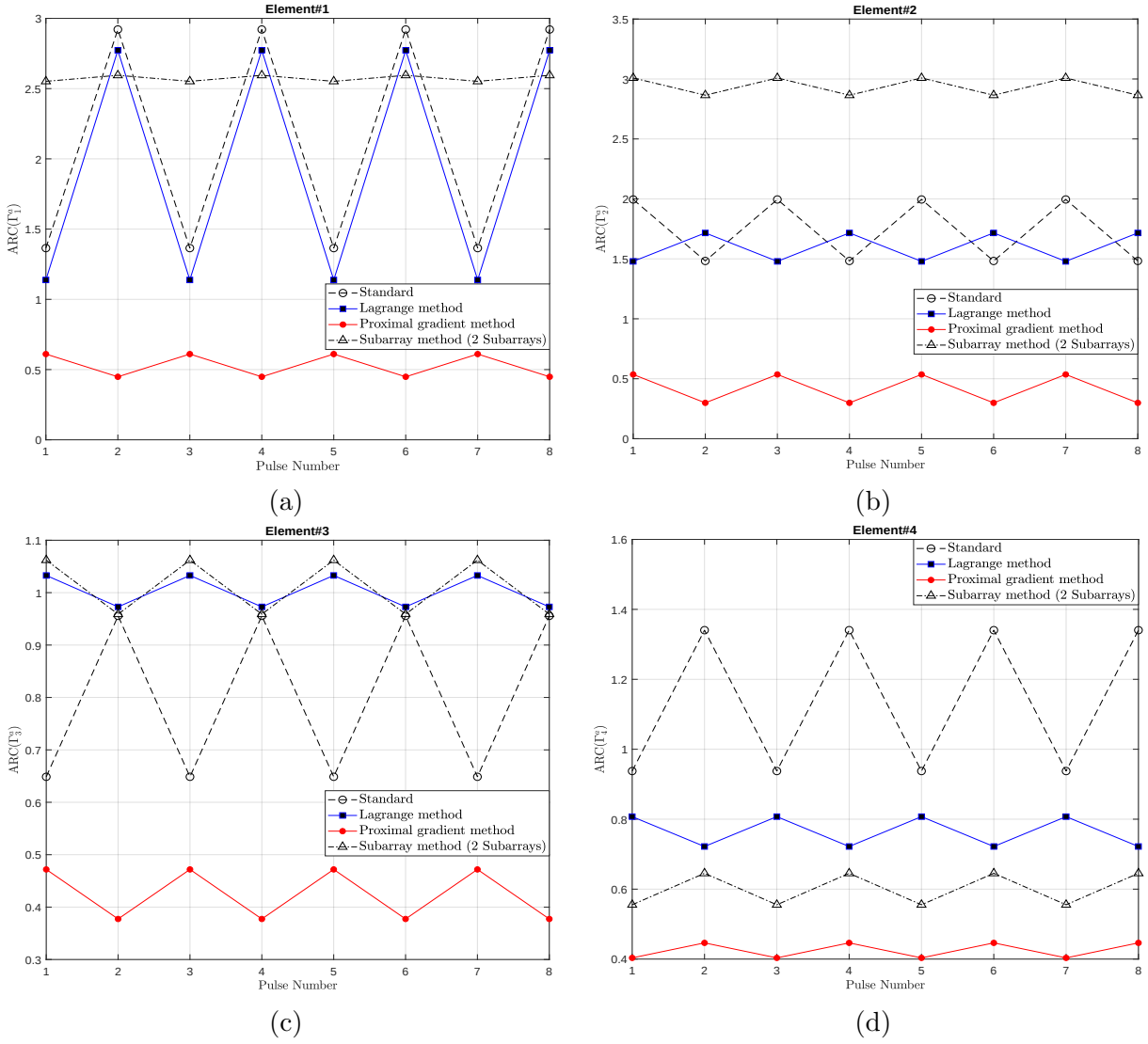


Figure 4.1 ARC of the 4 antenna elements (Gaussian TIR).

Fig. 4.2 shows Γ^a, but with K-distributed TIR, where we can draw the same discussion provided for the Gaussian TIR, noting that Γ^a is larger for most antenna elements.

The proximal-based method provides a minimum relative reduction in Γ^a of 52.28% and maximum reduction of 82.26%. To investigate the effect of each algorithm on the SINR

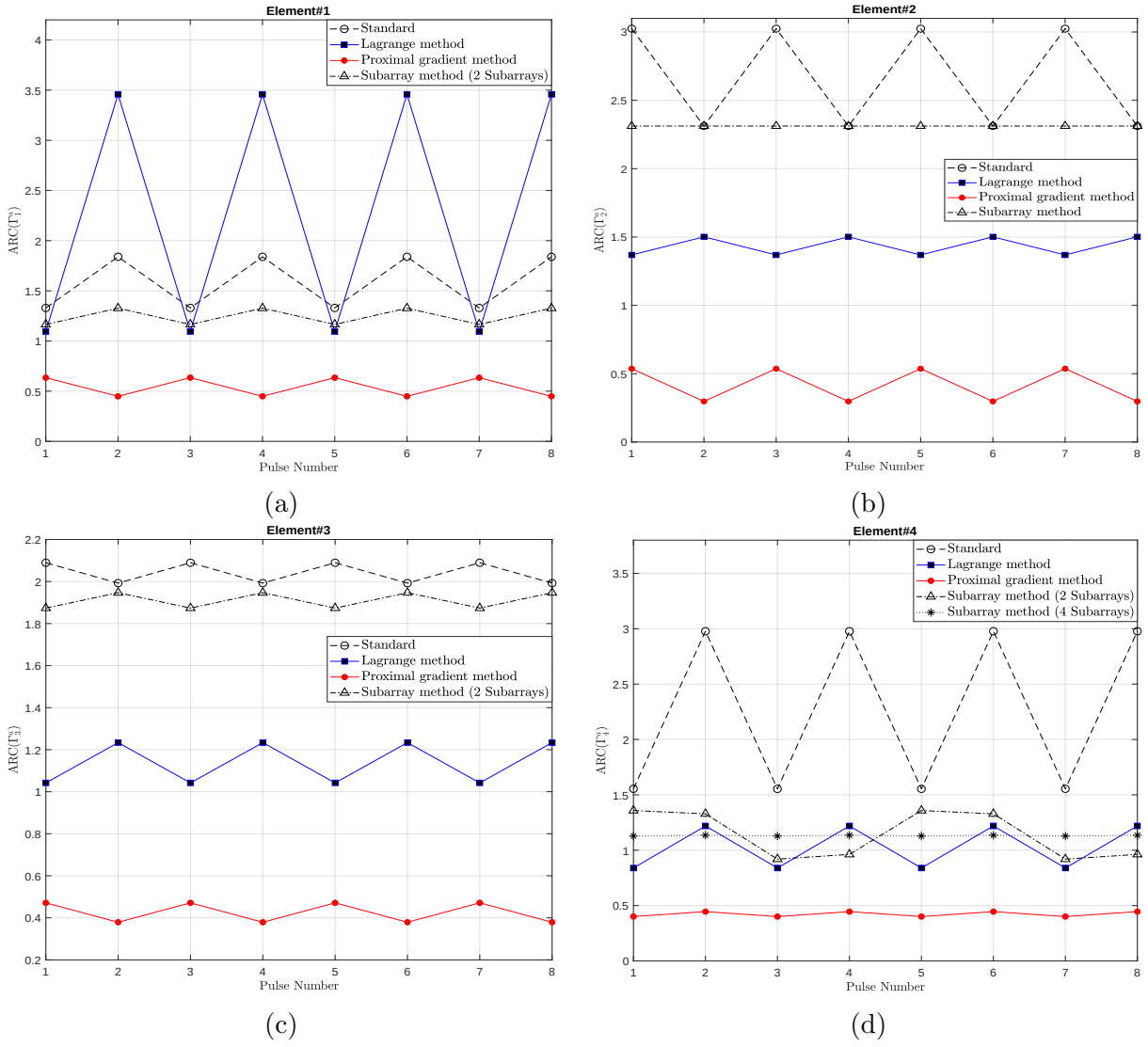
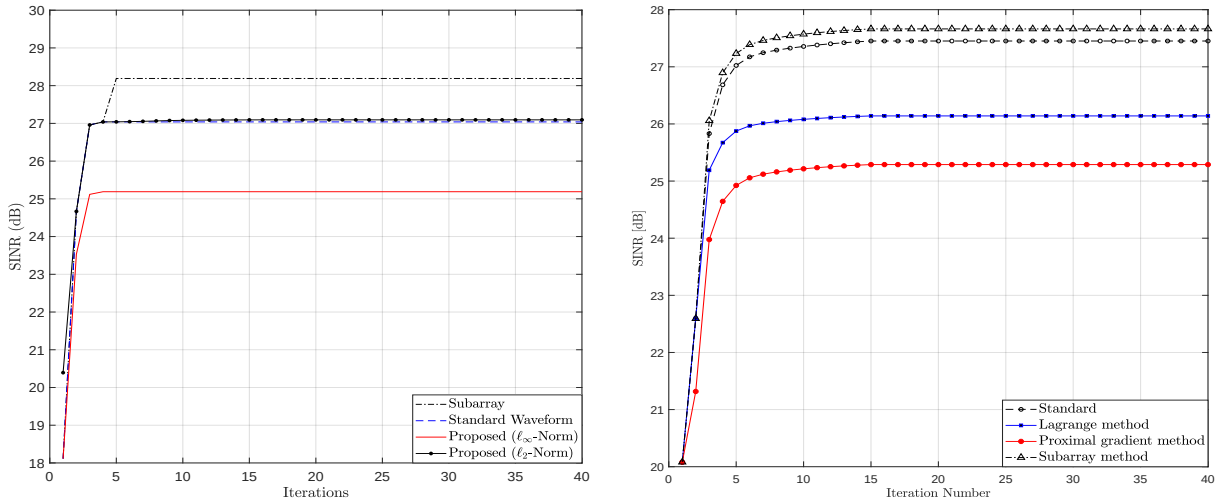


Figure 4.2 ARC of the 4 antenna elements (K-distributed TIR).

improvement, Fig. 4.3 shows the average SINR versus the iteration number over the 1000 simulated TIRs for both considered distributions. It should be noted that the iteration number reported for the proximal-based method in the figure is that of the outer loop. The first notice is that, on average, all the considered algorithms converge to a steady-state SINR value after approximately the same number of iterations. For both distributions, the subar-

ray configuration provides larger improvement of the SINR. The larger improvement of the subarray configuration over the full MIMO array is ascribed to the formation of two phased array radar within the MIMO radar, which enhances the processing gain of the receiver as reported in [185]. For the Gaussian TIR shown in Fig. 4.3a, the Lagrange-based proposed method provides the same SINR improvement as the standard algorithm, while the superior performance of the proximal-based proposed method comes on the expense of a lower improvement of the SINR than the standard method. That is to say, while the proximal-based method provides a superior Γ^a compared to all other considered algorithms, it still improves the SINR by 7 dB approximately compared to the 9 dB improvement provided by the standard and the proposed Lagrange-based methods. As shown in Fig. 4.3b for the K-distributed TIR, Lagrange-based method cannot provide the same SINR improvement as the standard method even without providing Γ^a at all pulses and antenna elements, while the proximal-based method still provides the same SINR as for the Gaussian TIR along with guaranteed superior Γ^a at all pulses and antenna elements. It should be emphasized that the reduction in SINR improvement is inevitable if Γ^a is to be minimized by proper design of the transmitted waveforms. The basic idea of maximizing the SINR is to minimize the interference power while in the same time maximizing the target signal power by matching the transmitted waveform to the TIR of the target of interest. However, constraining this matching by a low reflected power form the transmitting antenna array has the effect of limiting the maximization of the target power and, in turn, the SINR. This limitation of the SINR improvement should be weighed in the light of the achieved power efficiency of the transmitter and the protection of its power module. We give the results of the efficiency under the two considered TIR distributions with the used antenna sizes at the end of this section.

The TARC has been conventionally used as a figure of merit to evaluate the design of MIMO antenna arrays as mentioned in Subsection 4.3.2. For both TIR distributions, the TARC is calculated over all the array elements and all the transmitted pulses for each TIR. Fig. 4.4 shows the empirical cumulative distribution function (ECDF) of the TARC over 1000 TIR realizations for both Gaussian and K-distributions. For both of these distributions, it is depicted that the TARC for all algorithms does not exceed 0.5 with the standard, Lagrange, and proximal gradient methods. Surprisingly, the ECDF of the TARC gives the impression that the standard and the Lagrange-based design methods have better reflection properties than the proximal based method. However, we emphasize that the TARC cannot



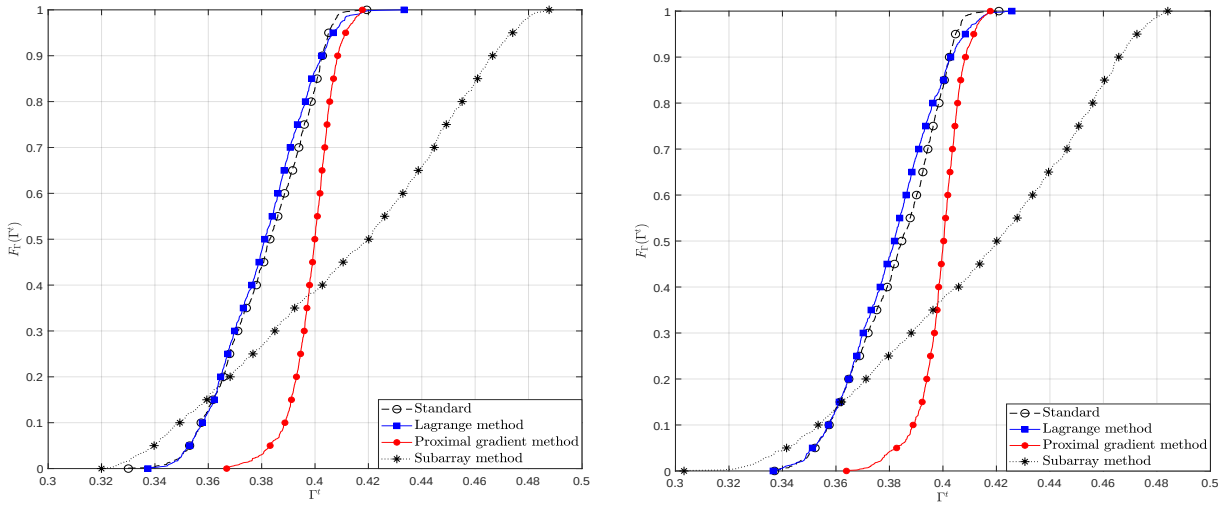
(a) Gaussian TIR. (b) K-distributed TIR.

Figure 4.3 SINR for the 4-element antenna array.

be used as a unique characteristic for the antenna design evaluation. While the TARC provides an *average* measure of the whole antenna array, it does not reflect the actual reflection properties encountered at each antenna element and each transmitted pulse as we can obviously conclude by comparing Fig. 4.4 with Fig. 4.1 and Fig. 4.2.

To have a more comprehensive evaluation of the ARC, we present in Table 4.1 some statistical measures of the ARC for the simulated TIR from both Gaussian and K-distributions denoted as "G" and "K", respectively. We can see that both sample mean, $\hat{\mu}_\Gamma$, and median, $\tilde{\Gamma}^a$, are approximately equal for all the considered algorithms and both TIR distributions. This, again, asserts the importance of not depending solely on the average reflection properties of the MIMO antenna array. The sample standard deviation, $\hat{\sigma}_\Gamma$, shows very similar values for the standard, Lagrange, and subarray methods with a relatively larger values for the K-distributed TIR than that for the Gaussian TIR. However, the proximal gradient method exhibits $\hat{\sigma}_\Gamma$ that is down to half of that of the standard method in the case of the K-distributed TIR. This means a lower scattering of Γ^a values around the mean.

To have a closer look on the trend of the scattering of Γ^a values around the mean, we calculated the skewness of Γ^a, denoted as $\hat{\varsigma}_\Gamma$. Except for the proximal gradient method, all the considered methods have $\hat{\varsigma}_\Gamma > 4$, which means that the Γ^a are scattered more to the right of the mean, i.e., greater than the mean. It should be noted that symmetrically distributed data has $\hat{\varsigma}_\Gamma = 0$ [186]. We can also notice that $\hat{\varsigma}_\Gamma$ is larger in the case of K-distributed TIR



(a) Gaussian TIR. (b) K-distributed TIR.

Figure 4.4 TARC ECDF for the 4-element antenna array.

than the Gaussian TIR. Only the proximal gradient method has a negative skewness, which means more scattered values of Γ^a are lower than the mean. Further, we seek to have more information about the magnitude of the scattered values of Γ^a, which can be perceived from measuring the tailedness of Γ^a values. We use the sample kurtosis, $\hat{\kappa}_\Gamma$, to measure the tail heaviness of Γ^a values. The proximal gradient method provides the lowest $\hat{\kappa}_\Gamma$ for both TIR distributions, even lower than that of Gaussian distributed data for which $\hat{\kappa}_\Gamma = 3$ [186]. For all the other methods, the values of $\hat{\kappa}_\Gamma$ are much larger than that of the proximal method with the K-distributed TIRs have larger $\hat{\kappa}_\Gamma$ than that of the Gaussian TIR. This means a much higher tendency of the standard, subarray, and Lagrange methods to have high-valued outliers than that of the proximal-gradient proposed method. This explains the difference between the average and worst-case evaluations of Γ^a shown in Fig. 4.4 compared to Figs 4.1 and 4.2, respectively. Moreover, this statistical analysis emphasizes our claim of the false conclusions that can be drawn depending on the TARC.

4.5.3 Performance Evaluation with 8-Element Transmitting Antenna Array

In this subsection, we consider a larger transmitting antenna array with $N_T = 8$ assuming Gaussian and K-distributed TIR. In this size of antenna array, we consider the two possible configurations of the subarray MIMO. The first configuration is two subarrays with each subarray having four antenna elements, while the second configuration is four subarrays

Method	Standard		Lagrange		Proximal		Subarray	
Dist.	G	K	G	K	G	K	G	K
$\hat{\mu}_\Gamma$	0.4025	0.4084	0.4028	0.4029	0.4002	0.4004	0.4044	0.4116
$\tilde{\Gamma}^a$	0.3840	0.3857	0.3836	0.3838	0.4175	0.4176	0.3825	0.3873
$\hat{\sigma}_\Gamma$	0.1185	0.1429	0.1219	0.1213	0.0703	0.0704	0.1297	0.1562
$\hat{\varsigma}_\Gamma$	4.1582	5.3415	4.3319	4.7418	-0.2620	-0.2612	4.0734	5.7962
$\hat{\kappa}_\Gamma$	48.67	60.54	46.30	69.10	2.08	2.09	43.09	70.13

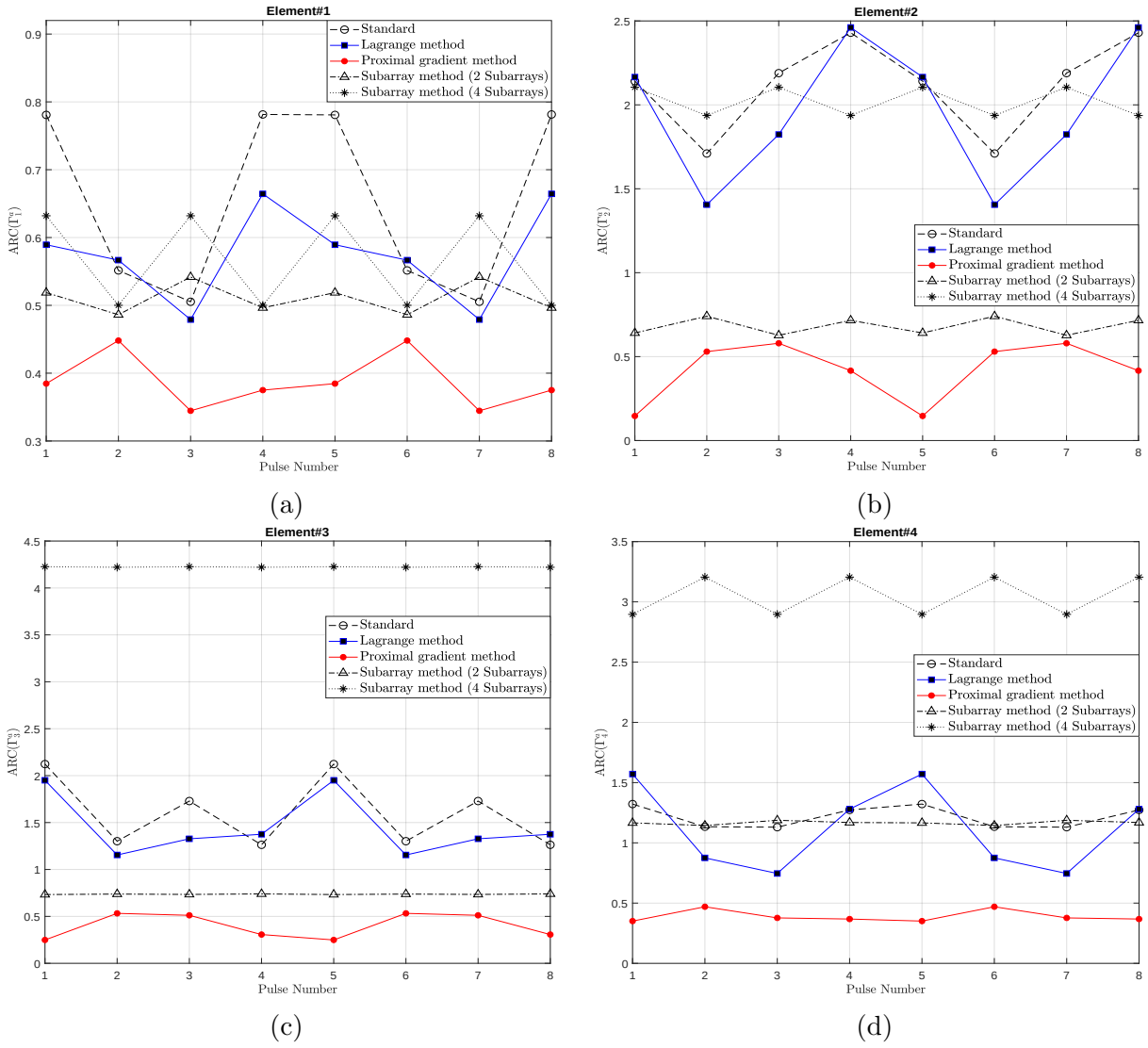
Table 4.1 ARC statistical data of 4-element antenna array

with two elements in each subarray. In Fig. 4.5, Γ^a for each antenna element in the array is shown with Gaussian TIR. The results shown here in Fig. 4.5 and in Fig. 4.1 are similar. That is, with both of the subarray configurations, Γ^a can exceed that of the full MIMO radar. We can also see that the proposed proximal-based method has the smallest Γ^a among all other waveform design techniques, while the Lagrange-based waveform design method cannot guarantee a lower Γ^a at all pulses and antenna elements.

The same can be shown for the K-distributed TIR in Fig. 4.6. We can see that the proposed proximal-based method provides the best control over Γ^a along all the pulses and antenna elements as in the case of Gaussian TIR. Conversely, both the standard and the proposed Lagrange-based methods show high Γ^a that exceeds 1 at some antenna elements and pulses. As a general remark, from Figs 4.1, 4.2, 4.5 and 4.6, we can see how Γ^a changes from pulse to pulse and from one antenna element to another. The fast varying Γ^a challenges the implementation of impedance matching circuits [187] and highlights the relevance of the signal-processing handling of the reflection problem proposed in this work.

The effect of all the considered waveform design techniques on the SINR is shown in Fig. 4.7. The improvement in the SINR with the proposed proximal-based method is lower than that of the standard method with less than 3 dB as a cost for its superior reflection performance. This reduction in SINR improvement is slightly greater than what the proposed proximal-based method exhibits for the 4-element array. While the Lagrange-based method does not guarantee lowering Γ^a at all pulses or antenna elements, it limits the SINR improvement by more than 1 dB.

Fig. 4.8 shows the ECDF of Γ^t for the 8-element array. Just as we have shown for the 4-element array, the TARC ECDFs for the standard, Lagrange-based, and proximal-based methods do not show a significant difference from those of the Gaussian TIR, despite the relative reduction of probability of higher Γ^t in the proximal-based method relative to the



other two methods. We can also observe that the subarray method has slightly higher TARC values than the latter three methods. This emphasizes our previous conclusion about the insufficiency of the average reflection performance represented by Γ^t that does not trace the reflected power at each antenna element or each pulse.

The statistical data for the three waveform design methods and the subarray configuration (2 subarrays), shown in Table 4.2, gives the same deduction as in Table 4.1. The proposed proximal-based method provides the lowest scattering of the Γ^a around its mean value and the lowest tendency to exhibit high-valued outliers among the 1000 simulated TIR of both

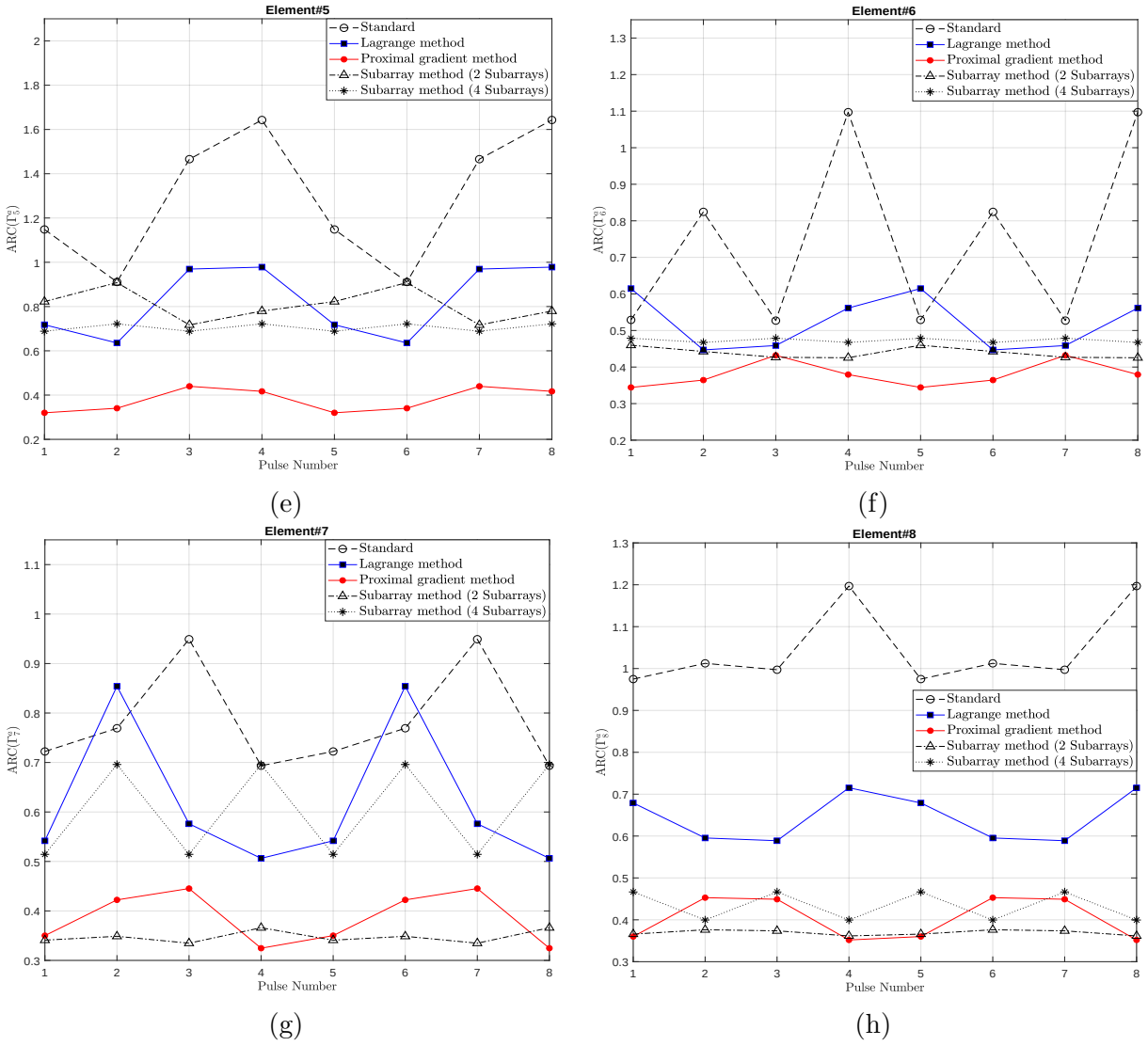
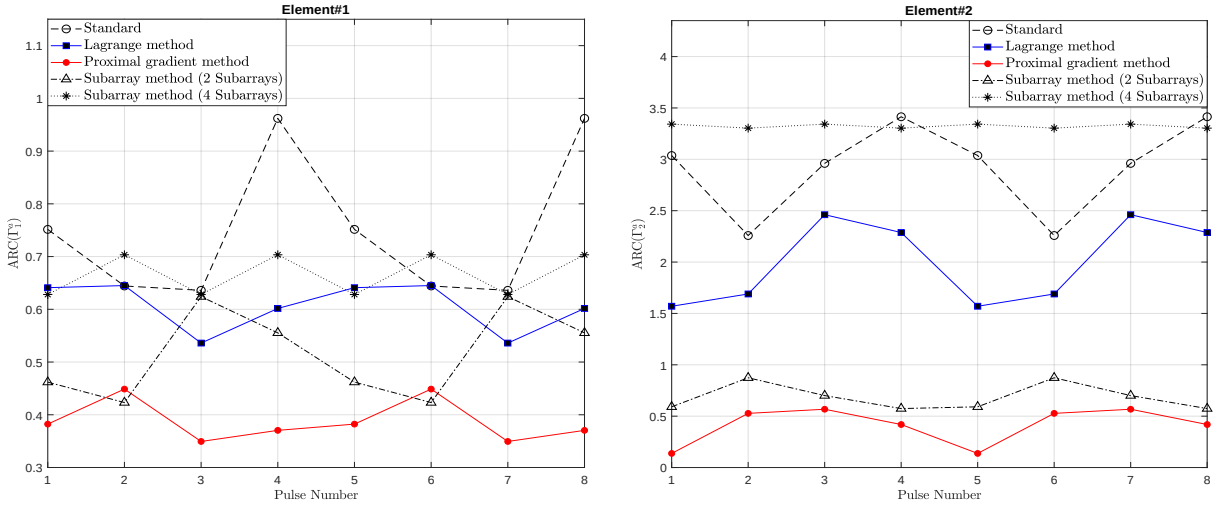


Figure 4.5 ARC of the 8 antenna elements (Gaussian TIR).

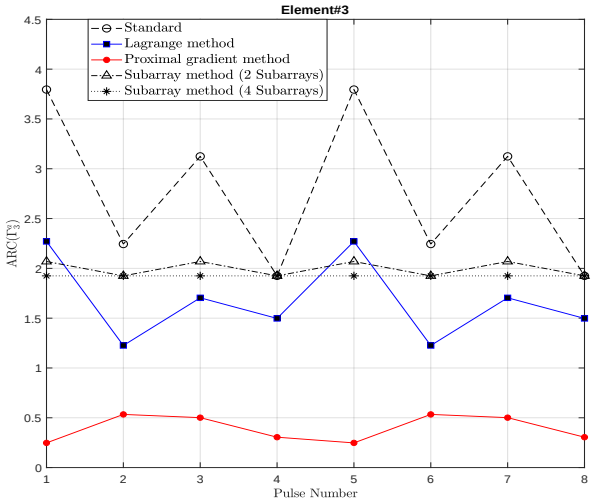
distributions considered in this work. The data in Table 4.2 also corroborates the closeness of the mean and median values of Γ^a for all the methods. We can also notice that $\hat{\kappa}_\Gamma$ is greater in the case of $N_T = 8$ than in the case of $N_T = 4$ for all considered algorithms. In addition, $\hat{\varsigma}_\Gamma$ in the case of $N_T = 8$ is greater than the case of $N_T = 4$ for all algorithms, with that of the proposed proximal-based has larger negative values, which means more compactness around the mean.

To evaluate the effect of the proposed algorithm using the ℓ_∞-norm on the efficiency of

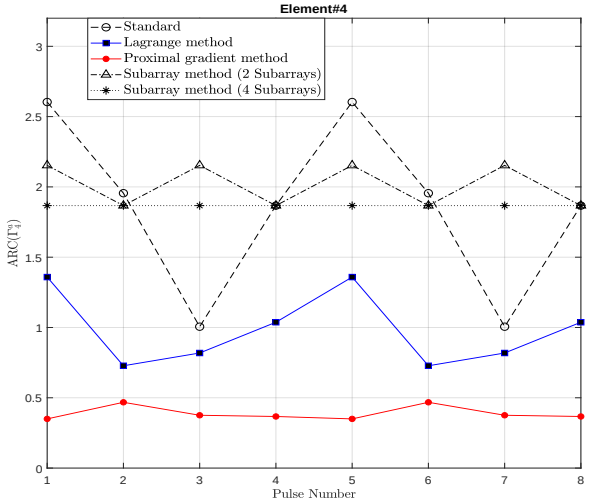


(a)

(b)



(c)



(d)

Method	Standard		Lagrange		Proximal		Subarray	
Dist.	G	K	G	K	G	K	G	K
$\hat{\mu}_\Gamma$	0.3915	0.3924	0.3876	0.3881	0.3700	0.3699	0.4191	0.4211
$\widetilde{\Gamma}^a$	0.3766	0.3769	0.3755	0.3756	0.3617	0.3616	0.3983	0.3991
$\hat{\sigma}_\Gamma$	0.1130	0.1207	0.0955	0.0973	0.0807	0.0806	0.1105	0.1172
$\hat{\varsigma}_\Gamma$	6.4826	7.2141	4.2844	4.1433	-0.9419	-0.9476	6.2984	6.7264
$\hat{\kappa}_\Gamma$	107.52	119.3	49.47	69.1	5.26	5.29	84.39	97.58

Table 4.2 ARC statistical data of 8-element antenna array

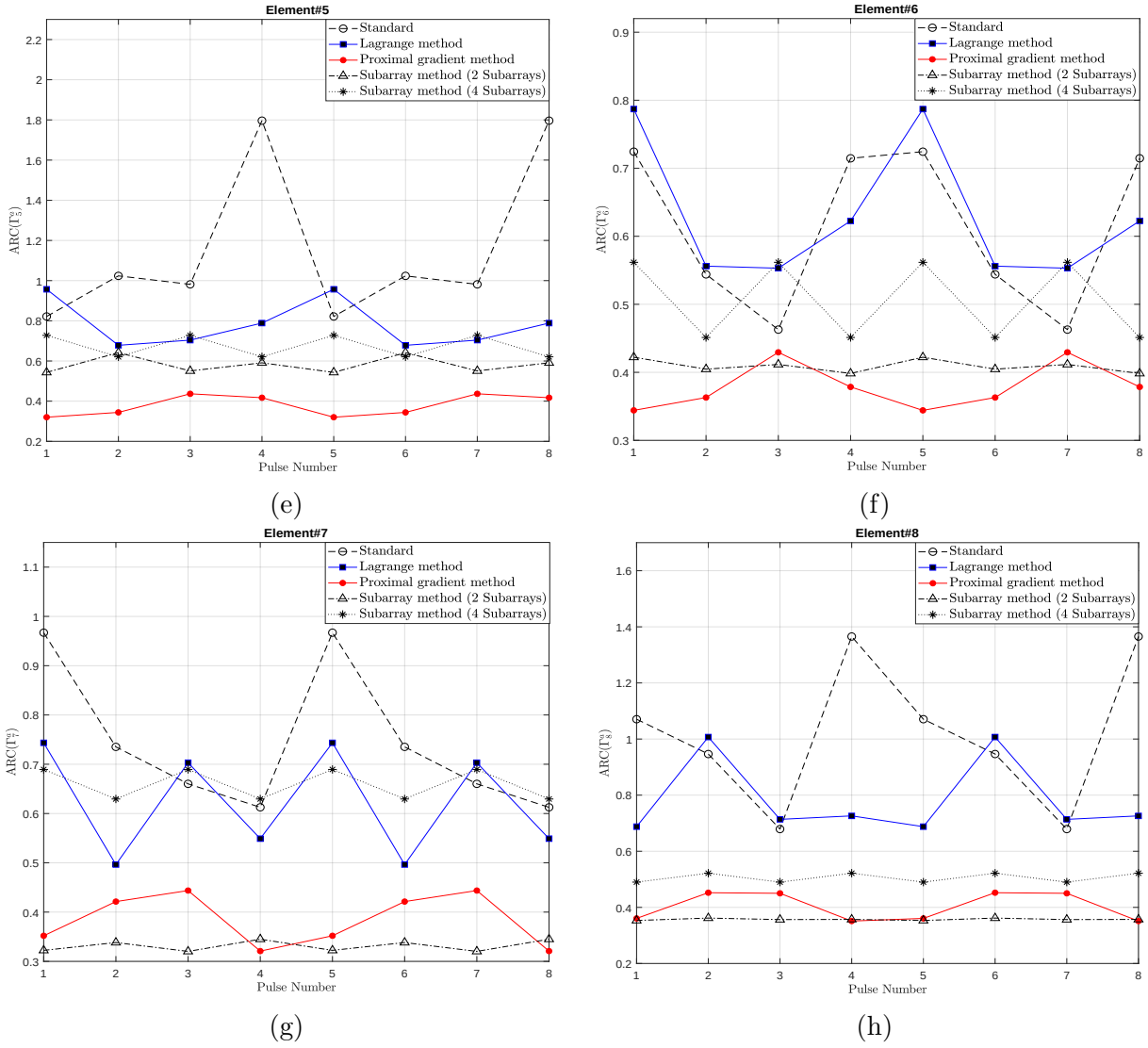
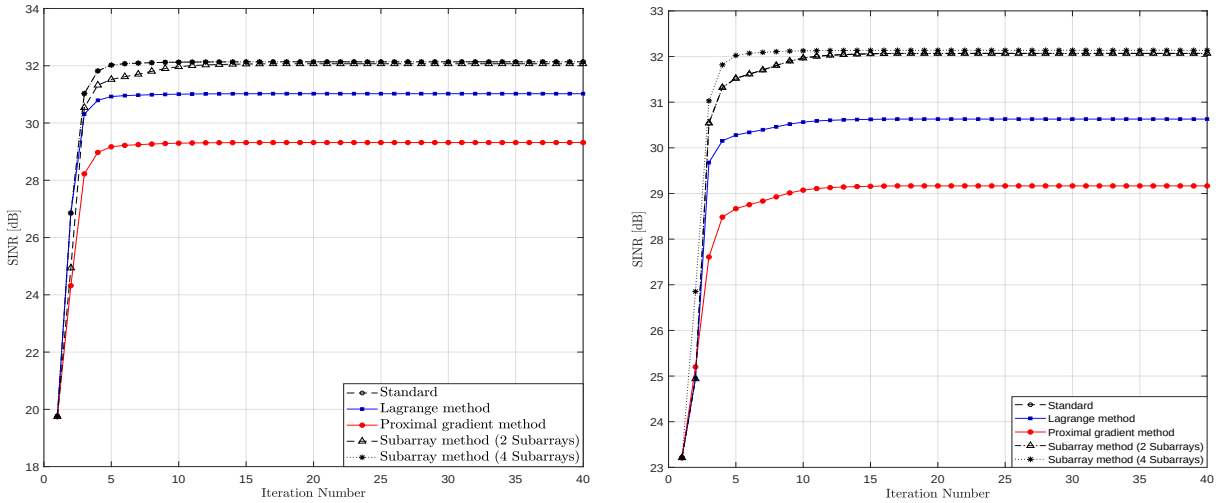


Figure 4.6 ARC of the 8 antenna elements (K-distributed TIR).

the transmitting antenna, we need to calculate the ratio between the effective transmitted power from each antenna element to its input power. The total reflected power from the jth port is given by

$$P_j^{ref} = \sum_{k=1}^{L_T} \sum_{i=1}^{N_T} |S_{ij} F_{jk}|^2 \tag{4.44}$$

where S_{pn} and F_{pn} are the components of the (p, n)th element of $\mathbf{S}$ and $\mathbf{F}$, respectively. On



(a) Gaussian TIR. (b) K-distributed TIR.

Figure 4.7 SINR for the 8-element antenna array.

the other hand, the input power fed to the jth antenna element is

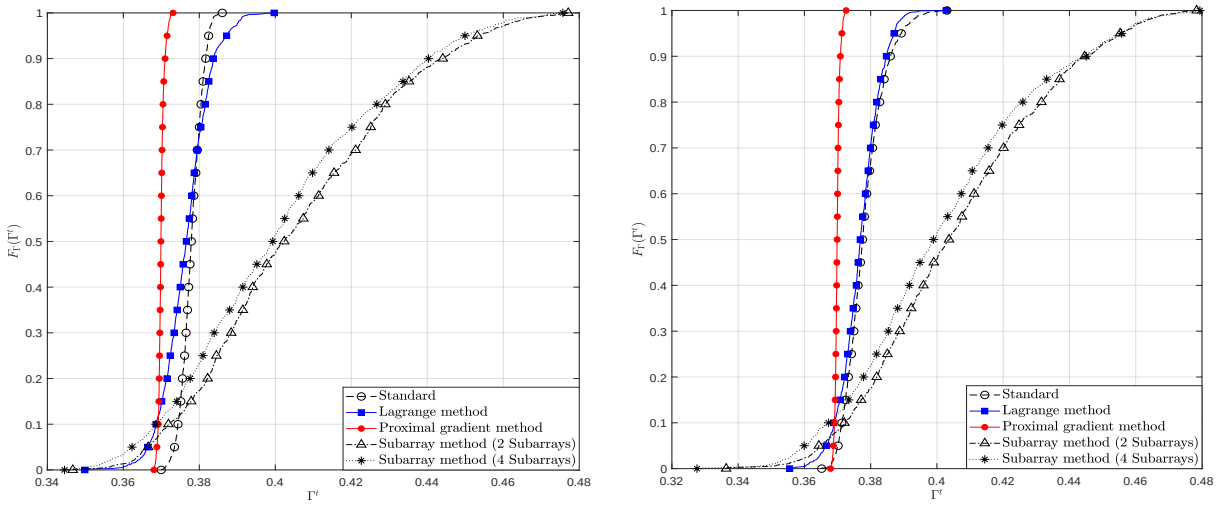
$$P_j^{in} = \sum_{k=1}^{L_T} |F_{jk}|^2 \tag{4.45}$$

Therefore, the efficiency factor of the jth port can be expressed as

$$\eta_j = 10 \log \left(1 - \frac{\sum_{k=1}^{L_T} \sum_{i=1}^{N_T} |S_{ij} F_{jk}|^2}{\sum_{k=1}^{L_T} |F_{jk}|^2} \right) \tag{4.46}$$

The efficiency factors of all antenna elements of the two considered arrays over 1000 simulated TIRs from both Gaussian and K-distributions are shown in Figs 4.9 and 4.10. As observed in these figures, the proposed algorithm with ℓ_∞-norm regularization guarantees the efficiency factor to be better than -1 dB for the two considered antenna sizes and TIR distributions. Conversely, we observe that the standard cognitive MIMO waveform design method can have the efficiency factor as low as -5.5 dB for some antenna elements. It is also obvious that the proposed method offers a stable efficiency factor at all antenna elements for the two considered cases for both the TIR and antenna arrays.

It should be emphasized that the data depicted in Figs 4.9 and 4.10 is not related to that illustrated in Figs 4.1, 4.5, 4.2 and 4.6 for the ARC. While the ARC expresses the total power



(a) Gaussian TIR. (b) K-distributed TIR.

Figure 4.8 TARC ECDF for the 8-element antenna array.

reflected *to* an antenna element compared to the input power to this element, the efficiency factor relates the effective transmitted power from an antenna element to the input power to this element. Of course, the effective transmitted power from an antenna element is closely related to the total power reflected *from* this element. Therefore, there is no contradiction to have some elements with a slightly higher efficiency, with 0.5 dB approximately, using the standard method than the proposed method.

4.5.4 Complexity Analysis

The complexity of the ℓ_∞-based proposed method compared to the standard method is evaluated through the calculation of the execution times of both algorithms using a machine dedicated for this task with 64-bit Intel® Core™ i7-6700 CPU @3.4GHz and 16 GB RAM. Fig. 4.11 shows the execution time of the two proposed waveform design methods compared to the standard algorithm in logarithmic scale. It is manifest that the proposed proximal-based method provides the lowest execution time among the considered algorithms at all array sizes. This result is not surprising since the proximal-based algorithm has two major reductions in calculations compared to the standard and proposed Lagrange-based methods. The first reduction is eliminating the calculation of $\mathbf{R}_{c,f}$ from $\mathbf{R}_c$, which requires scanning all the elements of $\mathbf{R}_c$ leading to a complexity of $\mathcal{O}((N_T L_s)^2)$. The second reduction is replacing the calculation of the inverse of $(\mathbf{R}_{c,h} + \mathbf{h}^H \mathbf{R}_n \mathbf{h} \mathbf{I}_{N_T L_s})$ by calculating the inverse

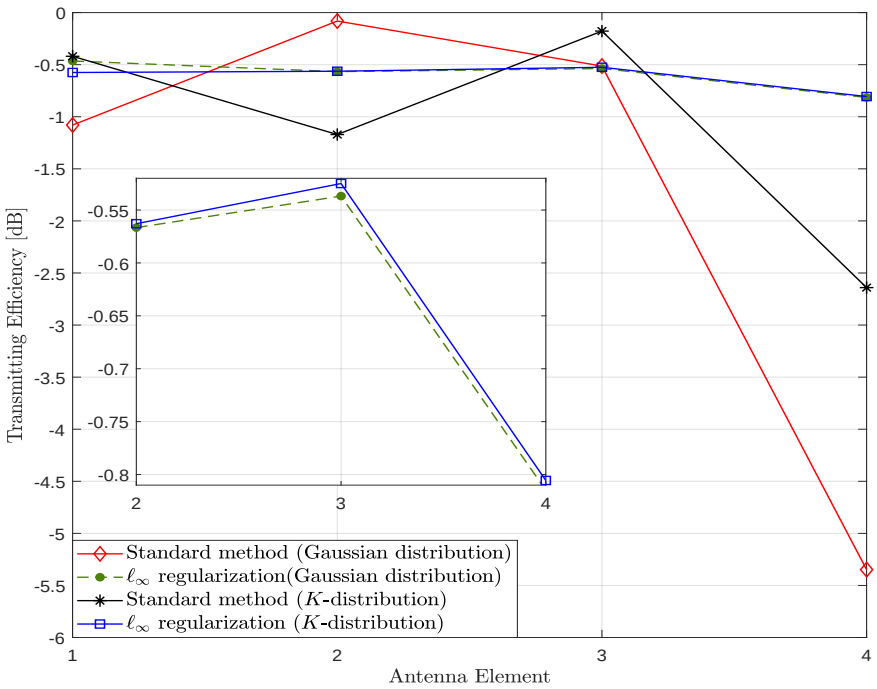


Figure 4.9 The efficiency of the each element in the 4-element array for Gaussian and K-distributed TIR

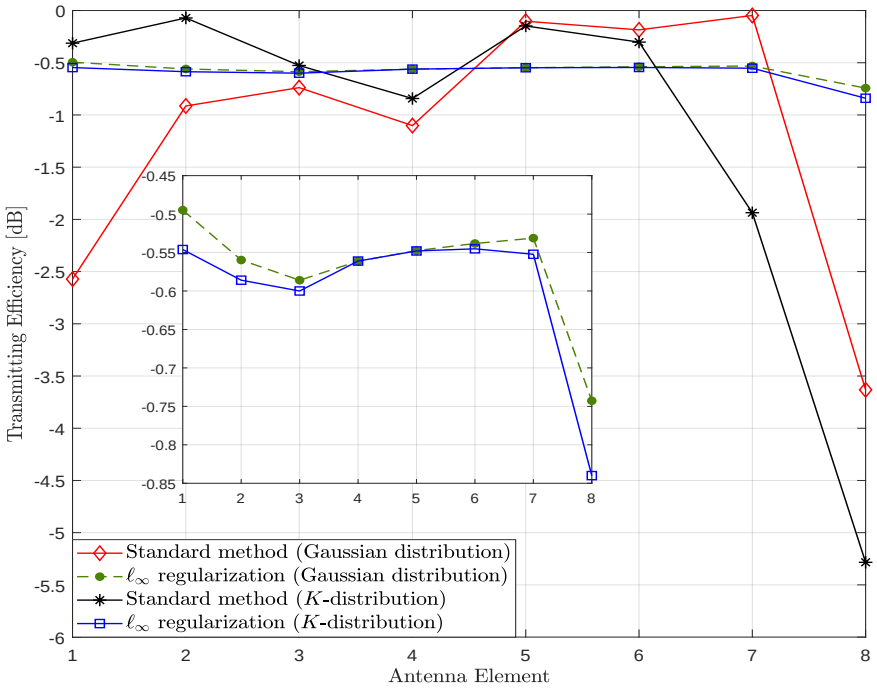


Figure 4.10 The efficiency of the each element in the 8-element array for Gaussian and K-distributed TIR

of $\mathbf{S}_d$. Obviously, in contrast to $\mathbf{R}_{c,h}$, $\mathbf{S}_d$ is fixed and does not change with the proximal algorithm iterations, so that its inverse is calculated once and stored to be used for all the iterations[6]. With a complexity of $\mathcal{O}((N_T L_s)^3)$ of the inverse operation and $\mathcal{O}((N_T L_s)^2)$ of the search operation compared to $\mathcal{O}(N_T L_s \log(N_T L_s))$ for finding the proximal operator of the ℓ_∞-norm, the low complexity of the proposed proximal-based algorithm is justified.

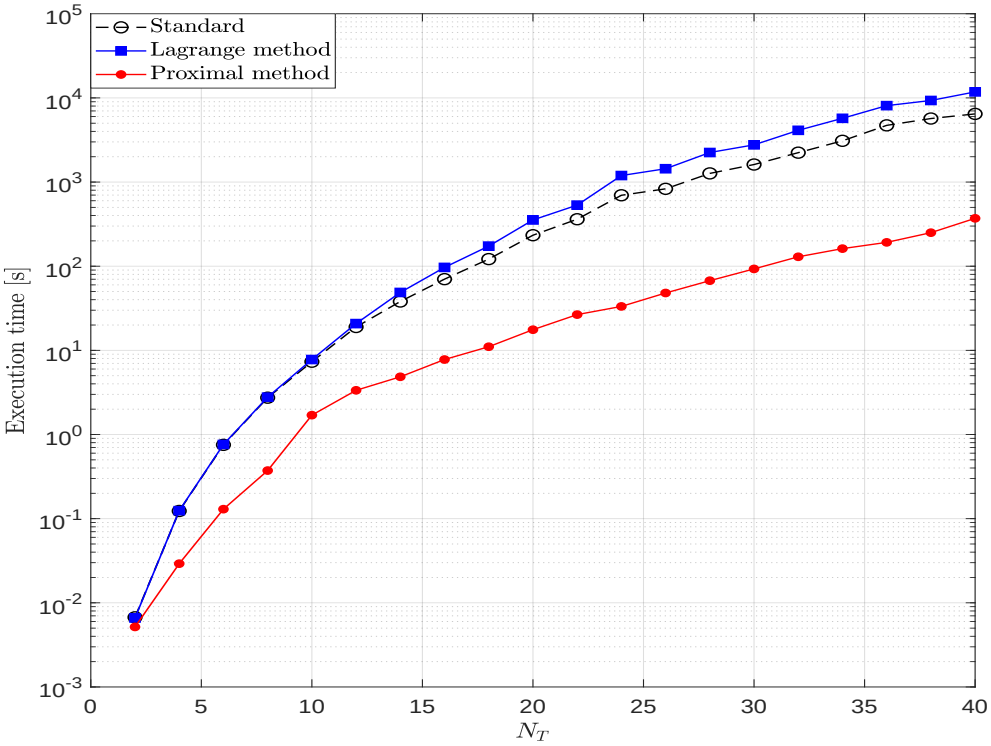


Figure 4.11 Execution times of the three considered waveform design methods.

4.6 Conclusion

In this chapter, we delved into the problem of power-efficient waveform design for cognitive MIMO radars. We proposed two novel algorithms that maximize the SINR for a certain target of interest, while at the same time minimizing the reflected power from the transmitting antenna array. The first proposed method adds an ℓ_2-norm regularization to the original ob-

[6]For lower memory storage in practice, the inverse of the matrix is calculated using lower-upper (LU) factorization or using the block recursive inversion [188].

jective function of the standard waveform design method and the resulting problem is solved using Lagrange method. The second proposed method employs the ℓ_∞-norm regularization, which results in a non-smooth term in the objective function. This non-smoothness necessitates the use of the proximal methods, from which the proximal gradient method is chosen to find the optimal waveform. The performances of the two proposed methods are compared to the standard waveform design method and to the subarray configuration, the only solution introduced in the literature to reduce the reflected power, yet with point targets and less degrees of freedom for the MIMO radar. Through extensive Monte Carlo simulations, the proposed proximal-based method has demonstrated outstanding performance from the reflection standpoint compared to the other considered benchmark methods in addition to the lowest complexity. These merits come with the cost of a reduced improvement of the SINR. This cost is considered low, especially when compared to the offered efficiency for the radar transmitter and the guaranteed protection for its microwave components. The work in this chapter lays the foundation for new methods of power-efficient waveform design.

In this chapter, we assumed that the TIR is known by the radar receiver *apriori* and we also pointed out that the TIR is estimated if this knowledge is not available. In Chapter 5, we deal with the problem of estimating the TIR under different distributions and generating models.

Chapter 5

Extended Target Frequency Response Estimation Using Infinite HMM

In this chapter, we delve into the estimation problem of the target impulse response (TIR), which has been assumed to be known *apriori* in the previous chapter. We introduce a new estimation method of the TIR based on Bayesian nonparametric models. Moreover, we introduce a new generating model and explore non-Gaussian distributions of the TIR. Through extensive Monte Carlo simulations, we show the robustness of the proposed method and its computational efficiency compared to the benchmark methods.

5.1 Introduction

As we discussed before in Section 2.4, cognitive radar systems are distinguished by their dynamic adaptation of their transmitter and receiver operations through continuous learning from the environment [141]. One goal of the transmitter adaptation is to optimize its waveform relative to the target of interest. The radar may encounter two types of targets: point targets, with dimension less than the radar range cell, and extended targets, occupying more than one range cell. An extended target can be viewed as a combination of multiple point targets and is modeled as a linear time-invariant or time-variant system and characterized by its TIR or, equivalently, by its target frequency response (TFR) [88, 189]. Since the TIR is band-limited in practice [181], the spectrum of the transmitted waveform can be matched to that of the TIR to improve radar detection [179].

In Chapter 4, we assumed that the TIR is known *apriori* by the radar. In practice,

however, the TFR is unknown and is conventionally estimated as the hidden state of a state-space model using a Bayesian filter. For a linear Gaussian state-space model, a Bayesian filter is realized exactly by the Kalman filter (KF). For nonlinear Gaussian models, the KF can be approximated using extended, decoupled, unscented, or cubature KF [141]. As indicated in Section 2.6, if both the Gaussian and linear assumptions are not met, the particle filter (PF) is the best possible approximation for the Bayesian filter [63]. Previous works reported in the literature have focused on using the KF assuming Gaussian TFR and interference (noise and clutter) with known statistics [6, 179, 180, 190], which are not always available. That is, the clutter signal can deviate from the Gaussian distribution and the TIR may not admit a linear or Gaussian model. In this case, the PF is the most viable option for TFR estimation. Moreover, the estimation performance of KF was evaluated within a single pulse, but its tracking performance over multiple pulses was not considered [179]. The estimation accuracy of the TRF is vital for the operation of cognitive radars. That is, it has been proven in [23] that minimizing the minimum mean square error (MMSE) of the TRF estimation is equivalent to maximizing the mutual information between the target return and the transmitted signal. This leads to a better target detection and optimal target information extraction [22, 191].

In contrast to KF, the hidden Markov model (HMM) is not limited to linear Gaussian models. While in the KF the state transitions follow a continuous Gaussian linear model, the HMM assumes discrete states whose transitions follow a Markov chain. Interestingly, the discrete states assumption is well-suited to modern digital radar receivers, where the amplitudes of the processed signals are quantized to a finite number of values [192]. These observations motivate our investigation of applying the HMM to the TFR estimation problem. However, in a similar manner to KF and PF, to apply HMM to TFR estimation the model structure (e.g., transition probabilities) must be known, which is rarely the case. A promising solution to this difficulty is the nonparametric Bayesian framework, which when applied to the HMM results in the infinite HMM (iHMM).

In this chapter, we provide a new formulation for the TFR estimation problem that makes it amenable to iHMM-based solutions. Further, we propose a new iHMM-based TFR estimation method that inherits all the desirable properties of nonparametric Bayesian approaches that is, it does not require any prior knowledge about the statistical properties of the TFR or the interference. Monte Carlo simulations are performed to compare the proposed method with the KF assuming Gaussian TFR and interference. We take the performance

analysis a step further than the literature by considering the tracking performance over multiple pulses rather than the estimation performance at a single pulse. Moreover, we extend the analysis to the non-Gaussian TFR or clutter cases, for which we develop the PF and use it as a benchmark. Furthermore, we put forward a new generating model for the TIR instead of the linear state-space model considered in the literature so far. Finally, we consider severe operating conditions such as smart noise jamming, which has not been considered before in TFR estimation context. Our simulations show that the proposed method outperforms KF and PF in terms of tracking error in all considered scenarios. Since there is no benchmark in the literature for the non-Gaussian case, we applied the PF to the TFR estimation as a benchmark. Regarding the complexity of the proposed method, the latter shows lower complexity compared to the PF at all TRF distributions and generating models.

The remainder of the chapter is organized as follows. Section 5.2 is a background on the extended target model. The TFR generating models and distributions are discussed in Section 5.3, in which we propose a new generating model for the TRF. In Section 5.4 we introduce our new formulation to the TRF estimation problem. Section 5.5 provides the details of the proposed method to estimate the TRF equipped with the new formulation introduced in the previous section. The performance of the proposed algorithm is compared to that of Kalman and PFs in Section 5.6. Section 5.7 is a summary.

5.2 Extended Target Model

Let $\tilde{\mathbf{g}} \in \mathbb{C}^{L_s}$ be the discrete-time transmitted radar waveform, which is fixed for M pulses, where L_s is the number of samples. A target with a range span larger than the radar's range cell can be divided into multiple, say L_t, discrete scattering centers, as shown in Fig. 5.1. Hence, after a time delay corresponding to its range from the radar, the reflected signal from the ith scattering center is [189, Ch. 9] [193]

$$\mathbf{x}_i = \tilde{\mathbf{g}} * (\sqrt{p_t} \sum_{j=0}^{P} b_{ij}) \sum_{i=1}^{L_t} \delta(t - \tau_i) \tag{5.1}$$

where * denotes convolution, τ_i is the time delay corresponding to the ith scattering center, $b_{ij} \in \mathbb{C}$ is the reflection coefficient of the jth reflecting surface within the ith scattering center,

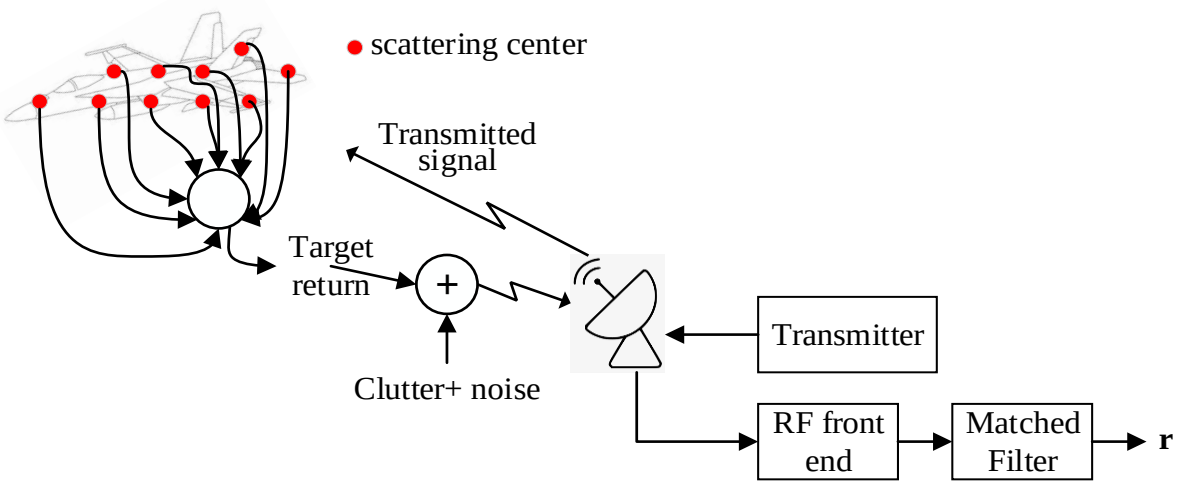


Figure 5.1 Extended target model.

p_t is the average transmitted power of the radar, and P is the total number of reflecting surfaces within the ith scattering center. It should be noted that the radar antenna gain[1], path loss, and the radar cross sectional area of a reflecting surface are all absorbed in b_{ij}. Hence, the TIR can be expressed as [193]

$$\tilde{\mathbf{h}} = \sum_{i=1}^{L_t} A_i \delta(t - \tau_i) \tag{5.2}$$

where $A_i = \sum_{j=0}^{P} b_{ij}$ is a random scalar representing the amplitude of the ith scattering center. The model in Eq. (5.2) is known as the scattering center model. Therefore, the extended target can be seen as a composition of L_t point targets.

When the P reflections are approximately equal, A_i can be modeled as a Gaussian random variable, where the central limit theorem applies [2]. In this case, the total received signal at the mth pulse $\tilde{\mathbf{r}}^{(m)} \in \mathbb{C}^{L_s+L_t-1}$ is

$$\tilde{\mathbf{r}}^{(m)} = \tilde{\mathbf{g}} * \tilde{\mathbf{h}}^{(m)} + \tilde{\mathbf{c}}^{(m)} + \tilde{\mathbf{n}}^{(m)} \tag{5.3}$$

$\tilde{\mathbf{h}}^{(m)} \in \mathbb{C}^{L_t}$ is the target impulse response, which changes on a pulse-to-pulse basis, $\tilde{\mathbf{c}}^{(m)}$ is the clutter vector and $\tilde{\mathbf{n}}^{(m)}$ is the noise vector. In this work, $\tilde{\mathbf{n}}^{(m)}$ and $\tilde{\mathbf{c}}^{(m)}$ are modeled as

[1] Monostatic radar configuration is assumed in this chapter.

[2] It was reported in [189, Ch. 9] that the Gaussian approximation is valid while P can be as low as 10 .

independent random vectors with $\tilde{\mathbf{n}}^{(m)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_n)$, while the distributions of $\tilde{\mathbf{h}}^{(m)}$ and $\tilde{\mathbf{c}}^{(m)}$ are discussed later.

After passing through an analog-to-digital converter (ADC), the baseband received signal $\tilde{\mathbf{r}}^{(m)}$ is passed through a receive filter of length L_s. The filter output of length $L_r = 2L_s + L_t - 2$ is then transformed to the frequency domain via an L_r-point discrete Fourier transform (DFT). Denoting $\mathfrak{T}\{\cdot\}$ as the combined effect of the L_r-point DFT and the receive filter operations, which are both linear, the frequency-domain received signal is

$$\mathbf{r}^{(m)} = \mathbf{G}\mathbf{h}^{(m)} + \mathbf{c}^{(m)} + \mathbf{n}^{(m)} \tag{5.4}$$

where $\mathbf{r}^{(m)} = \mathfrak{T}\{\tilde{\mathbf{r}}^{(m)}\}$, $\mathbf{G} = \text{diag}(\mathfrak{T}\{\tilde{\mathbf{g}}\})$ is the diagonal matrix of $\mathfrak{T}\{\tilde{\mathbf{g}}\}$, $\mathbf{h}^{(m)}$ is the target frequency response (TFR), $\mathbf{c}^{(m)} = \mathfrak{T}\{\tilde{\mathbf{c}}^{(m)}\}$, and $\mathbf{n}^{(m)} = \mathfrak{T}\{\tilde{\mathbf{n}}^{(m)}\}$.

The TFR is usually assumed to be linear and Gaussian distributed, an assumption that is validated if there are at least 10 scattering centers with approximately equal reflections, as the central limit theorem applies [88,189]. In this case, the Bayesian filter for TFR estimation is realized using the KF. However, the TIR may deviate from the Gaussian distribution, as we will show shortly. Moreover, the clutter signals in many radar environments are non-Gaussian [194]. In such nonlinear and non-Gaussian models, the PF can be used [63], which is not limited to linear or Gaussian assumptions. The PF approximates the Bayesian filter by using sequential importance sampling, which represents the posterior distribution of the states by weighted particles [64] as shown in Section 2.6.

The lack of prior knowledge about the TFR distribution hinders the choice of the right approach for its estimation and tracking (over pulse index m) the TFR. Even if this information is available, knowledge of the distribution parameters would be necessary for the proper design of the Bayesian filter or any of its approximations. In this chapter, we are concerned with estimating and tracking $\mathbf{h}^{(m)}$ from the received signal $\mathbf{r}^{(m)}$ without any prior information about $\mathbf{h}^{(m)}$ or interference terms $\mathbf{c}^{(m)}$ and $\mathbf{n}^{(m)}$.

5.3 TFR Generating Models and Distributions

Before delving into the TRF estimation problem, we discuss the TRF, or equivalently the TIR, generating models. This discussion is important for two reasons. The first reason is that employing the most realistic generating model is essential to the reliability of the assessment

of different estimation approaches. The second reason is that all the related previous work in the literature has assumed the linear state space model as the exclusive generating model of the TIR. This assumption, along with the Gaussian assumption, led to the prevalence of the KF in TFR estimation as the best candidate for estimating the TRF. In this section, we investigate this assumption and whether it is based on experimental data or not. Moreover, we discuss the different distributions of the TIR according to the available measured data published in the literature.

5.3.1 Linear State Space Generating Model

The TFR is conventionally modeled as a state-space model [179],

$$\mathbf{h}^{(m)} = e^{-T_r/\zeta}\mathbf{h}^{(m-1)} + \mathbf{v}^{(m-1)} \tag{5.5}$$

where ζ is the decay time constant, T_r is the radar pulse repetition interval, and $\mathbf{v}^{(m-1)}$ is the white Gaussian state noise vector in frequency domain. Both $\mathbf{v}^{(m-1)}$ and $\mathbf{h}^{(m-1)}$ are assumed to be independent.

Comparing Eq. (5.5) to Eq. (2.41) in Subsection 2.6.3, we can say that the model assumed for the TIR is a very special case of the linear state space model. That is, in Eq. (5.5) the transition matrix is not only assumed to be time-invariant, but it is also reduced to a scalar. Besides the constraints this model imposes on the generation of the TIR, there is a crucial question about the validity of these assumptions in real scenarios.

5.3.2 Correlated Random Process Model

We noticed that the linear state space model in Eq. (5.5) has been used recently in the literature, but it is not based on real measured data. We traced the use of Eq. (5.5) in the literature, and we found out that the first time it has been used was in [195] based on the fact that "*the TIRs of different time in a short interval are correlated, and the correlation coefficient decreases with increasing time interval*". However, this statement, which is based on measured data reported in [196], does not dictate or suggest a certain generating model as the one in Eq. (5.5), which has been later followed in [179, 190, 197].

Nevertheless, the measured data from real targets suggests that the correlation model of target return follows a first-order Markov process [198]. This conclusion also agrees with

Edrington's findings in [199], also based on measured data, that the target returns are exponentially correlated. Therefore, the only available information about the generating model of the TIR, based on experimental data, is about the correlation of this model. The correlation matrix of the TIR is expressed as [200]

$$\mathbf{\Psi}_h = \left[\rho_h^{|i-j|}\right], \qquad 1 \leq i, j \leq L_t \tag{5.6}$$

As noticed from Eq. (5.6), the correlation properties of the TIR admits the same form as that of the clutter in equations (3.35) and (3.34) as we have shown in Chapter 3. With this available information about the generating model of the TIR and the distributions to which the TIR measured data has been fitted, it is more convenient to model the TIR as a random vector generated from a correlated random process. In this regard, we propose to model the TIR using an SIRP, which has been shown before in Section 3.3, as a less restrictive model than the linear state space model conventionally assumed in the literature.

5.3.3 TFR distributions

As we have noticed in the literature, both the Gaussian distribution and the linear state space model have been jointly assumed for the TIR. These assumptions facilitate the task of estimating the TIR and make it viable for the KF to be employed. However, as we disputed the validity of the linear state space model to the TIR in the previous section, we bring the generalization of the Gaussian TIR assumption into question.

The conditions for the Gaussian assumption are not always met in real scenarios. As Swerling outlined in [201] for point targets, and the same is applied for extended targets under the scattering-center model, the random amplitude A_i in Eq. (5.2) deviates from the Gaussian assumption when a scattering center is formed of a dominant reflection and other roughly equal small reflections. Even without physical justification, several non-Gaussian distributions fit the measured reflections from different target types [181, 202].

Measured data for different target types suggests different non-Gaussian distributions, among them the Log-normal, Weibull, and K-distributions, which are common to different targets types [203]. The Log-normal distribution can be easily generated by applying the exponential function to a complex multivariate Gaussian distribution and the generation of the complex multivariate K-distribution has been discussed under the SIRP in Section 3.3. Unfortunately, the generation of the complex multivariate Weibull distribution is not

straight forward as the complex multivaraite K-distribution. As mentioned in Eq. (3.8) in Section 3.3, and revisited here for convenience, the SIRV is on the form

$$\tilde{\mathbf{h}} = v\mathbf{y}, \tag{5.7}$$

where $\mathbf{y} \in \mathbb{C}^{L_t}$ follows a complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ with zero mean and covariance matrix $\mathbf{\Sigma}$, and v is a positive random variable. The pdf of v for the complex multivariate Weibull distribution is given in terms of infinite summations [204]. To approximate these summations, [92] suggested using an algorithm based on the Rejection Method, which can be computationally unattainable for low values of shape parameters and high dimensions. Furthermore, in [205] the ZMNL method has been employed to generate the mutlivariate complex Weibull distribution. However, as we have also mentioned in Section 3.3 for multivariate complex distributions, the ZMNL method suffers from the difficulty of controlling the pdf and the correlation matrix of the generated vectors at the same time. A better approximation for the multivariate complex Weibull distribution has been recently proposed in [206], which also uses the Rejection Method. The envelope r of the complex multivariate Weibull distribution for an L_t-dimensional vector $\tilde{\mathbf{h}}$ is given by

$$f_H(r) = \frac{2(-1)^{L_t}}{\Gamma(L_t)} \sum_{k=1}^{L_t} C_k \frac{a^k}{k!} r^{kb-1} \exp(-ar^b) \tag{5.8}$$

where $a > 0$ and $0 < b \leq 2$ are the scale and shape parameters and

$$C_k = \sum_{m=1}^{k} (-1)^m \begin{pmatrix} k \\ m \end{pmatrix} \frac{\Gamma(1 + \frac{mb}{2})}{\Gamma(1 + \frac{mb}{2} - L_t)} \tag{5.9}$$

To generate the pdf of v, the variable w, where $v = \sqrt{w}$, is first generated and its pdf is given by

$$f_W(w) = ab(2\pi)^{b/2-1} \sum_{n=0}^{\infty} \frac{(-a(\sqrt{2w})^b)^n}{n!\,\Gamma(1 - \frac{b}{2}(n+1))}. \tag{5.10}$$

Performing one-to-one transformation from w to $x = (2a'w)^{b/2}$, where $a' = a^{2/b}$, the pdf of

x is

$$\sum_{n=0}^{\infty} \frac{(-x)^n}{n!\,\Gamma(1-\frac{b}{2}(n+1))} \tag{5.11}$$

This infinite summation can be upper-bounded by the exponential pdf defined for a variable z by

$$f_Z(z) = \lambda \exp(-\lambda z) \tag{5.12}$$

and $\lambda = 1/\Gamma(1-\frac{b}{2})$. The Rejection Method can be used to generate v with the upper bound in Eq. (5.12) and reversing the performed one-to-one transformation.

5.4 TRF Estimation: A New Formulation

In this section, we formulate of TFR estimation as a nonparametric Bayesian iHMM estimation problem and provide its solution.

5.4.1 HMM as a Stochastic Finite State Machine

A finite state machine (FSM), also known as finite state automaton, consists of a set of states, a set of outputs (observations), an optional set of inputs, a transition function that controls the evolution of the states, and an emission function that describes the evolution of the observations [207]. Stochastic FSM (SFSM) is an FSM in which the transition and/or the emission functions are probabilistic. The simplest form of the SFSM is the Markov model, where the observations are associated to the states. The HMM is an extension to the simple Markov model in which the states are hidden from the observer, which suits many applications as channel modeling [208,209], speech signal processing [73,210], and many other applications. As we have briefly discussed in Chapter 2, the HMM has been considered in the literature as the cousin of state space models under the umbrella of graphical Bayesian models. However, HMM does not dictate generating models for the states or the observations, as the state space models. Conversely, the HMM probabilistically describes the dependencies among the states and observations regardless of their generating models. While the state-space model are susceptible to the generating model mismatch, the HMM does not assume a generating model at all.

As we discussed in Chapter 2, the main trend in the literature is to differentiate between KF, or the Bayesian filter in general, and the HMM based on the continuity of the states and the outputs. However, we argue that this is not quite accurate. As a matter of fact, the signal processing community has employed discrete time KF, as in [211] to give an example. Moreover, the continuous-time HMM (CT-HMM) has been used recently in different applications and different fields [212–214]. While the CT-HMM still assumes finite space of states, the iHMM can relax this assumption, as we will show shortly. Therefore, we believe that the distinguishing feature of the HMM over the Bayesian filters is that the former is not restricted to a generating model and is concerned only with the probabilistic model of the states and observations. Therefore, viewing the HMM as an SFSM is a more accurate and generic approach than viewing it as a variant of the state space model as claimed by Murphy in [61]. In addition, it is also more accurate than distinguishing the HMM from the state-space model based on the continuity of the states. In the following, we introduce a new scope of the TRF estimation based on the HMM as an SFSM.

5.4.2 TFR Modeling Using HMM

In related previous works, $\mathbf{h}^{(m)}$ is considered as a random vector with known distribution and generating model. This vector is estimated recursively over the pulse index m using a Bayesian filter [179]. Alternatively, we propose estimating $\mathbf{h}^{(m)} = [h_1^{(m)}, \cdots, h_l^{(m)}, \cdots, h_{L_r}^{(m)}]^T$, where $h_l^{(m)}$ denotes the lth frequency sample, by considering a recursion over the frequency index l within each pulse. The fact that the observations $\mathbf{r}^{(m)} = [r_1^{(m)}, \cdots, r_l^{(m)} \cdots, r_{L_r}^{(m)}]^T$ are quantized to a finite number of quantization levels allows us to assume a discrete model for the amplitudes of the TFR samples. Regardless of the TFR generating model or distribution, the finite set of values taken by $h_l^{(m)}$ can be seen as the possible states in a scalar SFSM. In this SFSM, the sample amplitude at frequency l can transit from a given state to any other state at frequency $l+1$ according to a certain probability distribution. Specifically, the TFR samples for each pulse can be modeled as an HMM, in which the output value (i.e. $r_l^{(m)}$) associated to each state is also stochastic and the states are hidden from the observer. The proposed formulation differs from the Bayesian filter not only in the employed tool but also in the scope. The Bayesian filter considers the TRF as a random vector and track its evolution from pulse to pulse based on their statistics, assumed to be known. The proposed formulation, however, views the samples of the TRF of each pulse as the random variables

of an SFSM, specifically HMM, based on a learned transition model that is updated each pulse.

To apply the HMM to the TFR at the mth pulse, the components of $\mathbf{h}^{(m)}$ are considered as the hidden state sequence, while the components of $\mathbf{r}^{(m)}$ are the observations. Each sample of $\mathbf{h}^{(m)}$ or $\mathbf{r}^{(m)}$ can take any value from the discrete level sets, $\mathbf{Q} = \{q_1, \cdots, q_{N_s}\}$ or $\mathbf{O} = \{o_1, \cdots, o_{N_o}\}$, respectively. Without loss of generality, it is assumed that $N_s = N_o \equiv N$, where N denotes the quantization levels of the used ADC. Within the mth pulse, we assume that the components of $\mathbf{h}^{(m)}$ form a Markov chain of first order, that is

$$\Pr(h_l^{(m)} = q_i | h_{l-1}^{(m)} = q_j, \cdots, h_1^{(m)} = q_k) = \Pr(h_l^{(m)} = q_i | h_{l-1}^{(m)} = q_k) \tag{5.13}$$

where $1 \leq i, j, k \leq N$ and $1 \leq l \leq L_R$. We also assume a homogeneous HMM within the same pulse, but not from pulse to pulse. That is, the probabilities in Eq. (5.13) do not depend on l but may change with m. To simplify the notation, we temporarily drop the index m noting that the following steps are applied to each pulse.

To estimate the hidden states, the HMM structure should be specified *a priori*. For $1 \leq l \leq L_r$, this structure is defined by:

1. The discrete sets of states $\mathbf{Q}$ and observations $\mathbf{O}$.

2. The state transition matrix $\mathbf{A}$ defined as

$$\mathbf{A} = [a_{ij}] = \Pr(h_l = q_j | h_{l-1} = q_i), \qquad 1 \leq i, j \leq N \tag{5.14}$$

3. The emission matrix $\mathbf{B}$ is

$$\mathbf{B} = [b_{ij}] = \Pr(r_l = o_j | h_l = q_i), \qquad 1 \leq i, j \leq N \tag{5.15}$$

4. The prior distribution of the states $\boldsymbol{\pi}$.

Both $\mathbf{A}$ and $\mathbf{B}$ do not depend on l based on the homogeneity assumption. To achieve both low quantization noise and high dynamic range, the number of bits of the ADC can be as high as 14 bits or more [215]. This implies transition and emission matrices of very high dimensions, let alone the difficulty of obtaining prior knowledge about them. To solve these

problems, nonparametric Bayesian models allow the parameters to grow or shrink according to the observed data rather than assume a fixed number of parameters [216].

5.5 TFR Estimation and Tracking

In this section, we provide a discussion about the Bayesian non-parametric (BNP) models[3], which may be unfamiliar to the radar signal processing community. From the BNP, we set the stage for the iHMM to be employed in the TRF estimation problem.

5.5.1 Bayesian Nonparametric (BNP) Models and Dirichlet Process

Before delving into the BNP models and how they differ from conventional Bayesian parametric models, we need first to clarify what is meant by "non-parameteric". The importance of this clarification stems from two reasons. The first is that there is no agreement on a unique definition of the term between the statisticians [217] and the second, and the most important, is that we use the term in Chapter 3 in a way that may seem different from the way we are going to use in this chapter.

Conventional Bayesian models consists of a set of parameters and a prior distribution. The prior distribution is then updated to the posterior distribution using the observations [218]. As Jacob Wolfowitz[4] outlined, the parametric statistical analysis is the one where the distribution is determined by a finite set of parameters. He denoted the non-parametric case as the one where functional forms of those distributions, in terms of their parameters, are unknown [220]. Therefore, the non-parametric statistical analysis tries to use the minimum number of assumptions about the underlying distribution. From this standpoint, the non-parametric statistical methods are commonly denoted as the distribution-free methods. However, we emphasize that using the two terms, non-parametric and distribution-free, as synonymous terms is not quite accurate. That is to say, while the non-parametric analysis involves one or more unknown parameters of the distribution, distribution-free tools do not make any assumptions about the form of the entire distribution [221]. In Chapter 3, we used the term "non-parametric" in the context of the proposed NHD to refer to a distribution-free detector.

[3]We use the term "statistical model" to denote a probabilistic measure on the sample or observations.

[4]Jacob Wolfowitz is a mathematician who first coined the term "non-parametric" in a seminal paper in 1942 [219].

Recently, "non-parametric" has been used in a way that can be seen as a generalization to Wolfowitz's definition, in which the model is determined by an *infinite* set of parameters. This allows the model parameters to grow with the observed data without specifying the cardinality of the parameters' set. In other words, the word "*non-parametric*" in the context of BNP does not mean the model does not have any parameters, but, in fact, it has infinite set of parameters. This recent definition of non-parametric models is the one used in this chapter.

The BNP models are probability models with infinite number of parameters. Therefore, following the Bayesian framework, we need to use a prior distribution on an infinite dimensional space with the following desirable properties: (1) a large support and (2) simple posterior inference [222]. The Dirichlet process has been introduced in [223] to fulfill these requirements. A Dirichlet process is a random probability measure that generates a distribution F for any measurable partitions $A_1, \cdots, A_K$ of the sample space. The vector of random probabilities $F(A_i)$ follows a Dirichlet distribution that is defined by [222]

$$[F(A_1), \cdots, F(A_K)] \sim \text{Dir}(\alpha F_0(A1), \cdots, \alpha F_0(A_K)) \tag{5.16}$$

where $\alpha > 0$ known as the concentration parameter and F_0 is a distribution known as the base measure. The Dirichlet distribution, denoted also as $\mathfrak{D}(\alpha, F_0)$, can be better understood with the aid of the stick-breaking process.

Definition (Stick-breaking process) Let $c > 1$ be an integer and β_c be independent, the stick length $\boldsymbol{\iota}$ is

$$\iota_1 = \beta_1 \tag{5.17a}$$

$$\iota_c = \beta_c \prod_{1<c'<c} (1 - \beta_{c'}) \tag{5.17b}$$

The process starts with a stick of unit length from which a proportion β_c is broken at the step c and the remaining is kept for the subsequent proportions. The proof of the equivalence between the stick-breaking process and the Dirichlet distribution can be found in [224].

As we mentioned above, one of the attractive properties of the Dirichlet distribution is the simple form of its posterior distribution. Assume $x_1, \cdots, x_n \sim F$ and $\delta_x(.)$ is the delta Dirac function at x, then the posterior distribution $F|x_1, \cdots, x_n$ is proportional to a Dirichlet distribution $\mathfrak{D}(\alpha + n, F_0 + \sum_{i=1}^{n} \delta_{x_i})$. This means that the posterior distribution of the

Dirichlet prior is also a Dirichlet distribution. The extension of the Dirichlet distributions to the infinite dimension is straight forward by letting $K, c \to \infty$ in equations (5.16) and (5.17).

5.5.2 Employing iHMM in TFR Modeling

Inspired by the nonparametric Bayesian models and the new formulation we introduced for the TRF estimation problem using HMM, the problem of determining $\mathbf{A}$ and $\mathbf{B}$ can be avoided using an iHMM with unbounded number of states [225]. In iHMM, only a finite number of states, say K, are invoked initially at $m = 1$; K may grow or shrink for $m > 1$ depending on $\mathbf{r}^{(m)}$. Each row of $\mathbf{A}$ or $\mathbf{B}$ is modeled using a Dirichlet process (DP), also known as stick-breaking process, which does not need prior knowledge of the number of states that is possibly unbounded (in our application $K \leq N$). As both $\mathbf{r}$ and $\mathbf{h}$ are for the same target, the DP of the rows of $\mathbf{A}$ as well as those of the rows of $\mathbf{B}$ should be linked. To model this relationship, we propose using the hierarchical DP (HDP) [226].

For any $\gamma > 0$, we define the infinite length random vector $\boldsymbol{\beta} = [\beta_i]_{i=1}^{\infty} \triangleq \text{Stick}(\gamma)$ as [226]

$$\beta_i = \hat{\beta}_i \Pi_{k=1}^{i-1}(1 - \hat{\beta}_k) \tag{5.18a}$$

$$\hat{\beta}_k \overset{iid}{\sim} \text{Beta}(1, \gamma) \tag{5.18b}$$

where $\text{Beta}(1, \gamma)$ denotes the Beta distribution with shape parameters 1 and γ. The ith row of $\mathbf{A}$ of length K is

$$\mathbf{a}_i = [a_{ij}, \cdots, a_{iK}] \sim \text{Dir}(\alpha\beta_1, \cdots, \alpha\beta_K) \tag{5.19}$$

where α is a non-negative scalar as before. The emission matrix $\mathbf{B}$ is generated in the same way as in equations (5.18) and (5.19), but using $\boldsymbol{\beta}^e$, γ^e, and α^e. Therefore, using HDP, the rows of $\mathbf{A}$ are linked through the common vector $\boldsymbol{\beta}$. Similarly, the rows of $\mathbf{B}$ are linked through the common vector $\boldsymbol{\beta}^e$. The prior probabilities of the states are generated as a random vector, that is

$$\boldsymbol{\pi} \sim \text{Dir}(\beta_K^i, \cdots, \beta_K^i) \tag{5.20}$$

where the parameter $\boldsymbol{\beta}^i$ is generated as a random vector distributed uniformly over a K-

dimensional hypersphere. It is should be noted that the proposed algorithm does not depend on the choice of the prior-probabilities vector $\boldsymbol{\pi}$.

Therefore, using only four hyperparameters $\alpha, \alpha^e, \gamma, \gamma^e$, the iHMM model is fully specified and controlled. Specifically, choosing higher values of hyperparameters the model is more biased to explore new states, while using lower values of the hyperparameters makes the states more concentrated around lower number of atoms in $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^e$.

For a constructive definition and better understanding of the HDP, we opt for the hierarchical *Polya urn* scheme shown in Fig. 5.2. Polya urn schemes are used to represent discrete probability distributions through filling colored urns with colored balls. For hierarchical Polya urns, we have an additional urn denoted as *oracle urn* [227]. We denote the number of balls of color j in a Polya urn of color i as n_{ij}. We also record the color of the last drawn ball. As Fig. 5.2 depicts, we choose a ball of color i to be put in an urn of color j with a probability proportional to the number of balls of color i in the urn of color j. With a probability α, we query the oracle urn. We choose a ball with a certain color according to the number of balls of this color in the oracle urn, otherwise we choose a new color with a probability γ. The number of balls of color j in the oracle urn is denoted as n^o_j. In iHMM, n_{ij} corresponds to the probability of moving from state i to state j, while n^o_j and α are common to all states.

From Fig. 5.2 we can grasp the idea behind the fact that the larger the values of α and γ with respect to n_{ij} and n^o_{ij}, respectively, the higher the tendency of the iHMM algorithm to explore new states. The same is applied to the hyperparameters α^e and γ^e.

5.5.3 TFR Estimation Using iHMM

In the following, we show how to infer the state sequence $\mathbf{h}$ and the iHMM hyperparameters $\alpha, \alpha^e, \gamma, \gamma^e$ for each pulse.

5.5.3.1 TFR inference

The first step in estimating $\mathbf{h}$ is to estimate the posterior probability density function (pdf) $f(h_l|r_{1:l})$ of the lth sample within the L_r samples, where $r_{1:l} = [r_1, \cdots, r_l]$. The canonical state inference algorithm is the Gibbs sampler, however, its convergence is slow, especially with correlated data. Moreover, the posterior and the prior pdfs of $\mathbf{h}$ should be conjugate[5]

[5]Conjugate distributions are members of the same distribution family. An example of these families is the exponential family that includes Dirichlet, normal, exponential, and Gamma distributions.

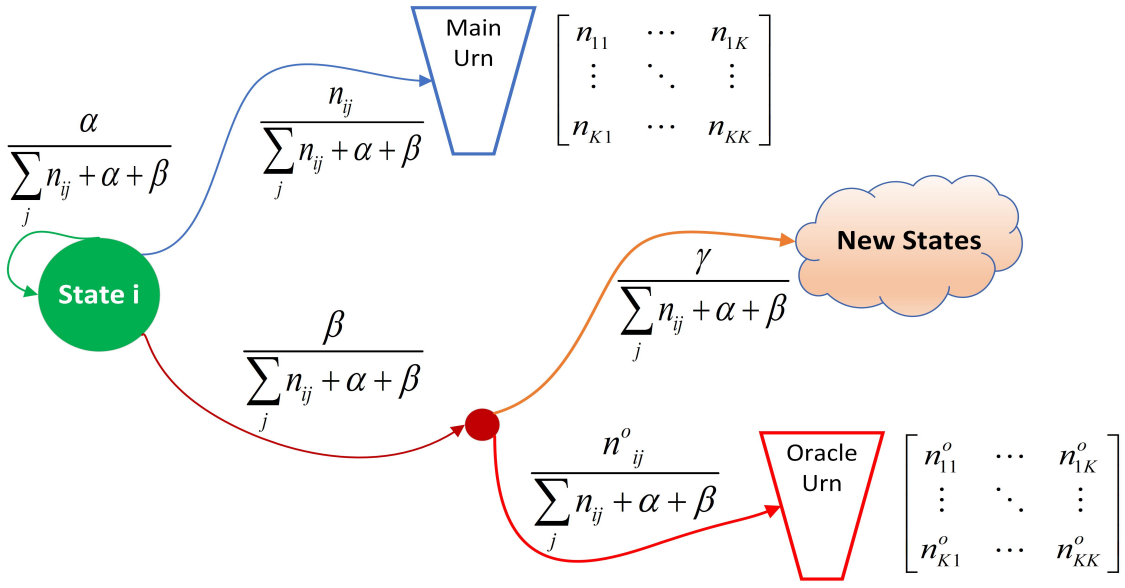


Figure 5.2 Hierarchical Polya urn scheme.

[228]. However, most of the measured data for radar clutter are correlated and the conjugate posterior needs *a priori* information about the prior probability distribution that is not usually available. To avoid these drawbacks, we adopt another inference algorithm, the beam sampling [228].

The beam sampler utilizes auxiliary variables to reduce the states of $\mathbf{A}$ and $\mathbf{B}$ at each l resulting in a finite number of states. Consequently, dynamic programming algorithms can be used to estimate the posterior pdf of the states as in the conventional HMM. Using the auxiliary variables $u_{1:L_r} = [u_1, \cdots, u_{L_r}]$, the posterior pdf can be estimated as [228]

$$f(h_l|r_{1:L_r}, u_{1:L_r}) \propto f(r_l|h_l) \sum_{u_l < \Pr(h_l|h_{l-1})} f(h_{l-1}|r_{1:l-1}, u_{1:l-1}) \tag{5.21}$$

In Eq. (5.21), the sum at each l is evaluated only over a limited number of states, say K_u, out of the invoked states K, whose transition probabilities exceed a threshold u_l. The choice of u_l is important. On one hand, a large u_l may result in underestimating the actual number of states. On the other hand, a small u_l may result in a higher number of states that increases the complexity of the model and the resulting error. The threshold is conventionally taken

as [225]

$$u_l \sim \mathcal{U}(0, \Pr(h_{L_r}|h_{l-1})) \tag{5.22}$$

with $\mathcal{U}(a, b)$ denoting the uniform distribution in the interval $[a, b]$. It can also be generated as [227]

$$u_l \sim \Pr(h_{L_r}|h_{l-1})\text{Beta}(w, z) \tag{5.23}$$

with $w, z > 0$. In the latter case, the appropriate choices of w and z, which have not been specified in the literature, should force u_l to be either close to 0 or $\Pr(h_{L_r}|h_{l-1})$.

In this work, we propose adjusting u_l depending on the pulse number m. Since there is no prior information about the true states of $\mathbf{h}^{(m)}$, the model is initialized at $m = 0$ with a low K. As m advances, the number of the invoked states K grows and, consequently, their relative transition probabilities tends to be lower. Therefore, u_l needs to be decreased as m increases, otherwise the number of surviving states K_u will be too low for an accurate estimation of $\mathbf{h}^{(m)}$. The details of the choice iHMM parameters are provided in Section 5.4.

Finally, to estimate the state sequence $\mathbf{h}$, h_{L_r} is first sampled using $f(h_{L_r}|r_{1:L_r}, u_{1:L_r})$, then the backward induction is used to estimate the remaining states from the posterior pdf as [228]

$$f(h_l|h_{l+1}, r_{1:L_r}, u_{1:L_r}) \propto f(h_{l+1}|h_l, u_{l+1}) f(h_l|r_{1:l}, u_{1:l}) \tag{5.24}$$

In this work, we employ the forward-backward algorithm [73] to infer the state at each time step as shown in Algorithm 6, where the lth column of the matrix $\mathbf{X}$ is denoted as $\text{col}_l\mathbf{X}$ and $\odot$ denotes the Hadamard multiplication. The process of estimating $\mathbf{h}^{(m)}$ is performed for each received signal $\mathbf{r}^{(m)}$, where the hyperparameters inferred based on $\mathbf{h}^{(m-1)}$ are used in generating $\mathbf{A}$ and $\mathbf{B}$ to estimate $\mathbf{h}^{(m)}$.

5.5.3.2 Inference of hyperparameters

After inferring the states, we infer the model hyperparameters α, γ, α^e, and γ^e as a second step. For $\mathbf{A}$, at $m = 0$ the two hyperparameters α and γ are initialized as [228]

$$\alpha^{(0)} \sim \text{Gamma}(a_\alpha, b_\alpha), \quad \gamma^{(0)} \sim \text{Gamma}(a_\gamma, b_\gamma), \tag{5.25}$$

Algorithm 6 Forward-backward Algorithm

Input: $\mathbf{A}$,$\mathbf{B}$,$\boldsymbol{\pi}$
Initialize: $\boldsymbol{\varepsilon}_1 = \boldsymbol{\pi} \odot \mathrm{col}_1(\mathbf{B})$,$\boldsymbol{\tau}_{L_r} = \mathbf{1}$
$\boldsymbol{\varepsilon}_1, \boldsymbol{\tau}_{L_r} \in \mathbb{R}^K$
for $l = 2$ **to** L_r **do**
$\boldsymbol{\varepsilon}_l = \mathbf{A}\boldsymbol{\varepsilon}_{l-1} \odot \mathrm{col}_l(\mathbf{B})$
end for
for $l = L_r - 1$ **to** 1 **do**
$\boldsymbol{\tau}_l = \mathbf{A}\mathrm{col}_{l+1}(\mathbf{B}) \odot \boldsymbol{\tau}_{l+1}$
$\boldsymbol{\varphi}_l = \frac{\boldsymbol{\tau}_l \odot \boldsymbol{\varepsilon}_l}{\boldsymbol{\varepsilon}_l^T \boldsymbol{\tau}_l}$
$\hat{h}_l = \mathrm{argmax}_{1 \geq i \geq K}\, \boldsymbol{\varphi}_l(i)$
end for

where $\mathrm{Gamma}(a, b)$ denotes the Gamma distribution with shape parameter a and inverse scale parameter b, where $a_\alpha, b_\alpha, a_\gamma, b_\gamma > 0$. At the mth pulse, $\alpha^{(m)}$ is generated as [229, eq. (47)]

$$\alpha^{(m)} \sim \mathrm{Gamma}(a_\alpha + E - \sum_{k=1}^{K_u} e_k, b_\alpha - \sum_{k=1}^{K_u} \log q_k), \tag{5.26}$$

where e_k is a binary variable that randomly takes a value of 0 or 1, E is the number of inferred states within $\mathbf{h}^{(m)}$ obtained after solving the dynamic program in the TFR inference step, and $q_k \sim \mathrm{Beta}(\alpha^{(m-1)} + 1, n)$ with n being the number of times each state of E is visited. Moreover, $\gamma^{(m)}$ is obtained by [226, eq. (13)]

$$\gamma^{(m)} \sim \pi_\eta \mathrm{Gamma}(a_\gamma + E, b_\gamma - \log(\eta)) + (1 - \pi_\eta)\mathrm{Gamma}(a_\gamma + E - 1, b_\gamma - \log(\eta)) \tag{5.27}$$

where

$$\pi_\eta = (a_\gamma + E - 1)/(E(b_\gamma - \log(\eta))) \tag{5.28}$$

and

$$\eta \sim \mathrm{Beta}(\gamma^{(m-1)} + 1, E) \tag{5.29}$$

Equations equations (5.25), (5.26) and (5.27) can be applied to infer α^e, γ^e given h_l using $a_{\alpha^e}, b_{\gamma^e}$ as initialization parameters, K_u^e as the number of surviving output values, n^e as the number of times an output value is visited, and E^e as the total number of visited observations.

The steps of the proposed method based on beam sampling for the inference of the states and the hyperparameters are shown in Algorithm 7.

Algorithm 7 Beam sampling of TFR iHMM

Input: $\mathbf{A}$,$\mathbf{B}$,$\boldsymbol{\pi}$
Initialize: $h_{1:L_r}^{(0)}$, $\alpha, \alpha^e, \gamma, \gamma^e, \mathbf{A}, \mathbf{B}$
for $m = 1$ **to** M **do**
Generate $u_{1:L_r}$
Apply Algorithm 6 $\forall$ $\Pr(h_l|h_{l-1}) > u_l$
Sample $h_{T-1:1}$ from $\Pr(h_l|h_{t+1}, y_{1:L_r})$
Sample $\alpha, \alpha^e, \gamma, \gamma^e, \mathbf{A}, \mathbf{B}$ using equations (5.25), (5.26), (5.27) and (5.19)
end for

5.6 Performance Evaluation

In this section we evaluate the TFR estimation accuracy of the proposed iHMM-based method and compare it to the particle and KF through Monte Carlo simulations.

5.6.1 Simulation Setup

Here, we briefly give the details of the simulation parameters and assumptions for the radar signal, clutter, TFR, the proposed iHMM-based estimator, Kalman, and particle filters.

5.6.1.1 Radar and clutter signal models

The radar transmitted waveform $\tilde{\mathbf{g}}$ is simulated as a linear frequency modulated signal with 1 MHz bandwidth and 4 μs pulse width sampled at 2.5 MHz and the ADC has $N = 2^{14}$ quantization levels. The clutter vector $\tilde{\mathbf{c}}$ is generated in the same way as $\mathbf{h}$ for both Gaussian and K-distributions with a covariance matrix

$$\boldsymbol{\Sigma}_c = [0.9^{|i-j|}], \qquad 1 \leq i, j \leq L_s \tag{5.30}$$

The noise vector $\tilde{\mathbf{c}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{L_s})$, where $\mathbf{I}_{L_s}$ is the identity matrix with dimension L_s.

5.6.1.2 Kalman Filter Design for TFR Estimation

The performance of the proposed method is compared to the KF in the case of Gaussian $\mathbf{h}^{(m)}$ and $\mathbf{c}^{(m)}$ assuming a linear state space model. The minimum mean square error (MMSE) estimator is used to initialize the KF, which is given by [23, 230]

$$\mathbf{h}^{(0)} = \left(\mathbf{G}^H\mathbf{G} + \mathbf{D}\mathbf{R}_i\mathbf{D}^H\mathbf{R}_h^{-1}\right)^{-1}\mathbf{G}^H\mathbf{r}^{(0)} \tag{5.31}$$

where $\mathbf{D}$ is the diagonal matrix of the receive filter in the frequency domain, $\mathbf{R}_i$ is the total interference covariance matrix, clutter plus noise, in the frequency domain, and $\mathbf{R}_h$ is the covariance matrix of the TRF. The final estimate is taken after 50 filter iterations. The recursion of this KF is based on equations (2.44) and (2.45) and is shown below

Prediction step

$$\hat{\mathbf{h}}^{(m)^-} = \exp(-T_r/\zeta)\hat{\mathbf{h}}^{(m-1)} \tag{5.32a}$$

$$\mathbf{P}^{(m)^-} = \exp(-2T_r/\zeta)\mathbf{P}^{(m-1)} + (1 - \exp(-2T_r/\zeta)\mathbf{R}_h \tag{5.32b}$$

Update step

$$\mathbf{K}^{(m)} = \mathbf{P}^{(m)^-}\mathbf{G}^H\left(\mathbf{G}\mathbf{P}^{(m)^-}\mathbf{G}^H + \mathbf{D}\mathbf{R}_i\mathbf{D}^H\right) \tag{5.33a}$$

$$\hat{\mathbf{h}}^{(m)} = \hat{\mathbf{h}}^{(m)^-} + \mathbf{K}^{(m)}(\mathbf{r}^{(m)} - \mathbf{G}\mathbf{h}^{(m)^-}) \tag{5.33b}$$

$$\mathbf{P}^{(m)} = \mathbf{P}^{(m)^-} - \mathbf{K}^{(m)}\mathbf{G}\mathbf{P}^{(m)^-} \tag{5.33c}$$

where $\hat{\mathbf{h}}^{(m)^-}$ and $\hat{\mathbf{h}}^{(m)}$ are the *a priori* and *a poteriori* estimates of the TIR, respectively, $\mathbf{P}^{(m)^-}$ and $\mathbf{P}^{(m)}$ are the *a priori* and *a poteriori* error covariance matrices, respectively, and $\mathbf{K}^{(m)}$ is the Kalman gain; all at the mth pulse.

5.6.1.3 Particle Filter Design for TFR Estimation

In the case of K-distributed $\mathbf{h}^{(m)}$ and/or $\mathbf{c}^{(m)}$, we apply the particle filter as we explained in Section 2.6. The number of particles is $N_p = 50$ as recommended in [231] for the considered signal dimension, with the ith particle at the m-th pulse having a weight $w_i^{(m)}$ calculated

as [64, eq. (63)]

$$w_i^{(m)} \propto w_i^{(m-1)} f(\mathbf{r}^{(m)}|\mathbf{h}^{(m)}) \qquad i = 1, \cdots, N_p \tag{5.34}$$

where $f(\mathbf{r}^{(m)}|\mathbf{h}^{(m)})$ is the likelihood pdf of $\mathbf{r}^{(m)}$, which is as deduced for the a mixture of K-distributed clutter and Gaussian noise in [232] as

$$f_I(\mathbf{i}) = \int_0^\infty f_V(v) \frac{1}{(2\pi)^{L_s} |v\boldsymbol{\Sigma}_c + \mathbf{I}_{L_s}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{i}^H (v\boldsymbol{\Sigma}_c + \mathbf{I}_{L_s})^{-1}\mathbf{i}\right) dv \tag{5.35}$$

where $\mathbf{i} = \tilde{\mathbf{c}}^{(m)} + \tilde{\mathbf{n}}^{(m)}$ and Eq. (5.35) is solved numerically. To overcome the degeneracy problem, the particles with negligible weights are removed and N_p particles are resampled, each with a weight $1/N_p$ [64].

For the PF and KF, we assume three cases. In Case I, the filters have complete prior knowledge about the statistical parameters of $\mathbf{h}^{(m)}$, $\mathbf{c}^{(m)}$, and $\mathbf{n}^{(m)}$. Case II assumes partial knowledge of the three vectors, that is, the distributions are known, but not their parameters. In Case III, the filters have no prior knowledge about the three vectors.

5.6.1.4 Parameters of the Proposed iHMM for TFR Estimation

In the iHMM, we found through simulations that $u_l = 10^{-\kappa}\mathcal{U}(0, \Pr(h_l|h_{l-1}))$, with $\kappa = 0.15m$, is an appropriate value. There are no recommended values for a_α, a_γ and b_α, b_γ in the literature, but we found $a_\alpha, a_\gamma \sim \mathcal{U}(1, 10)$ and $b_\alpha, b_\gamma \sim \mathcal{U}(0.1, 1)$ to be appropriate. Again, these choices are independent on the distribution or the generating model of $\mathbf{h}^{(m)}$.

The estimation error for all algorithms is calculated over $M = 16$ pulses using the normalized root mean square error (NRMSE) defined as

$$\epsilon(m) = \|\mathbf{h}^{(m)} - \hat{\mathbf{h}}^{(m)}\|_2 / \|\bar{\mathbf{h}}\|_2 \tag{5.36}$$

where $\mathbf{h}^{(m)}$ and $\hat{\mathbf{h}}^{(m)}$ are the true and estimated TFR at the mth pulse, respectively, $\bar{\mathbf{h}}$ is the sample mean of the true TFR over the 16 pulses, and $\|\cdot\|_2$ denotes the ℓ_2-norm. The NRMSE is averaged over 250 trials.

5.6.2 Results and Discussion

In this subsection, we present the results obtained for the linear state-space model of the TFR, used in the literature, but we also assume non-Gaussian distributions. Moreover, we consider the new model proposed in this thesis for the TFR based on the SIRP.

5.6.2.1 Linear State Space (LSS) TRF Model

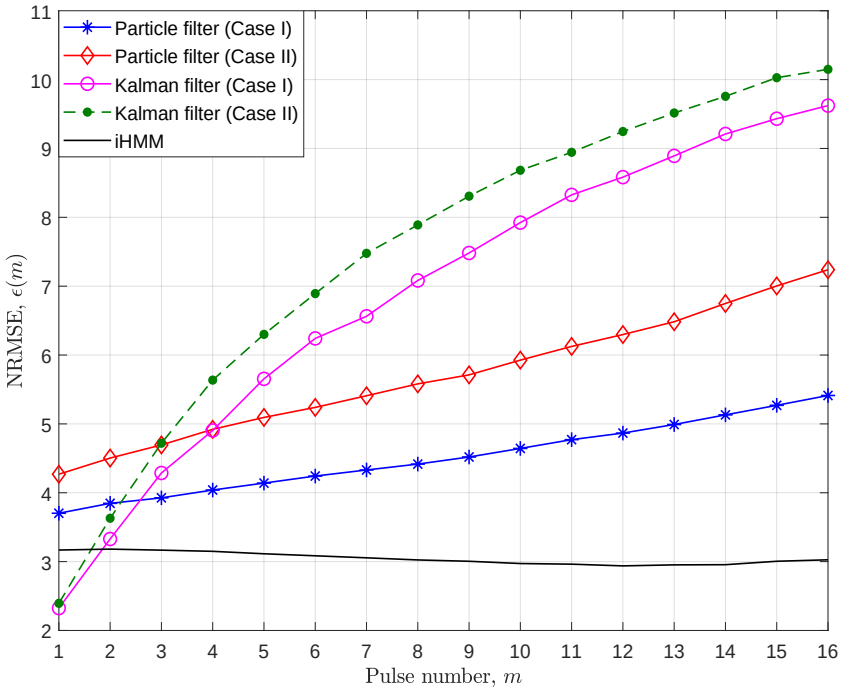


Figure 5.3 Estimation error of Gaussian TFR in Gaussian clutter (LSS model).

Assuming Gaussian $\mathbf{h}^{(m)}$, $\mathbf{c}^{(m)}$ and linear $\mathbf{h}^{(m)}$ model, the KF can be used to estimate $\mathbf{h}^{(m)}$. Fig. 5.3 shows that the KF provides the lowest error at $m = 1$ compared to the PF and iHMM. However, beyond $m = 1$, its error is proportional to m at a higher rate than that of the PF and above the iHMM. This lower error is ascribed to the fact that the KF is the exact Bayesian filter for the TFR estimation problem under the assumption of Gaussian $\mathbf{h}^{(m)}$ and $\mathbf{c}^{(m)}$. Nevertheless, as $\mathbf{h}^{(m)}$ itself is a random vector, the model mismatch increases with the accumulation of noise as m increases.

However, even this relative lower error of the KF at $m = 1$ is not guaranteed in all

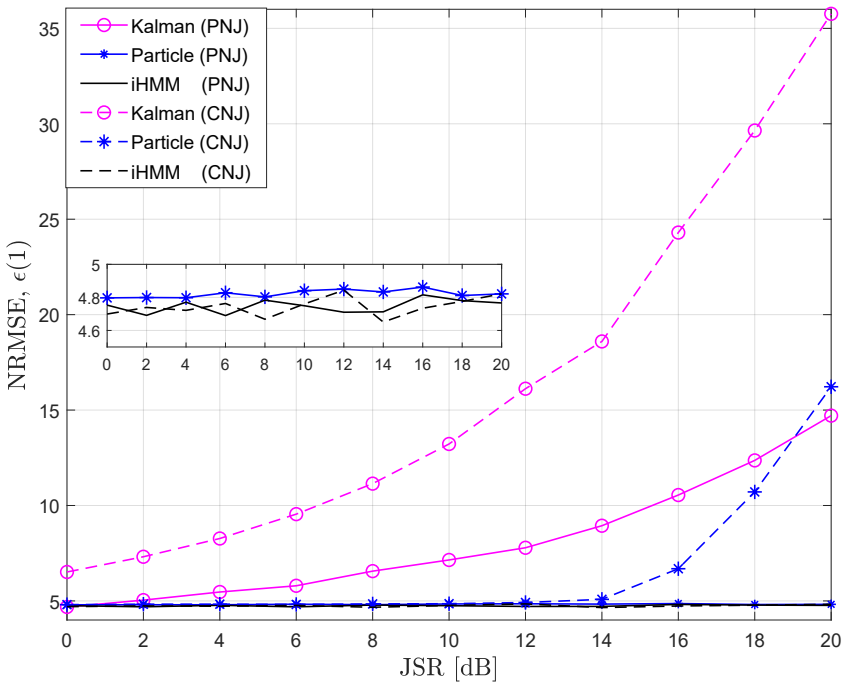


Figure 5.4 Estimation error of Gaussian TFR in jamming and Gaussian clutter ($m = 1$, LSS model).

operating conditions. Specifically, when the radar system is under the effect of jamming. We use the pulsed noise jamming (PNJ) and the convolution noise jamming (CNJ) to evaluate the performance of the three methods. Both are generated by a repeater jammer matched to the radar pulse width and repetition interval [77]. CNJ is the result of the convolution between the intercepted radar pulse and a noise pulse generated by the jammer. Fig. 5.4 shows ϵ at $m = 1$ under the effect of the PNJ and CNJ versus different jamming-to-signal ratios (JSRs). Compared to the jamming-free case, the KF is vulnerable to both jamming techniques, especially to the CNJ, with a higher ϵ that increases with the (JSR). For the PF, ϵ is higher at all JSR values in the case of PNJ, while ϵ increases significantly at higher JSRs under the effect of the CNJ. Except for the small increase in ϵ at all JSRs, the iHMM shows better stability in terms of ϵ in the presence of both PNJ and CNG.

Fig. 5.5 shows ϵ of the $K-$distributed $\mathbf{h}^{(m)}$ and $\mathbf{c}^{(m)}$ for both the iHMM and PF. Even when the Case I of the PF is considered, ϵ increases with m. In Cases II and III, ϵ is higher and increases in a faster rate. This is true whether the PF has prior partial knowledge or it has no information about $\mathbf{h}^{(m)}$ and $\mathbf{c}^{(m)}$. On the contrary, the iHMM shows a lower ϵ at pulses

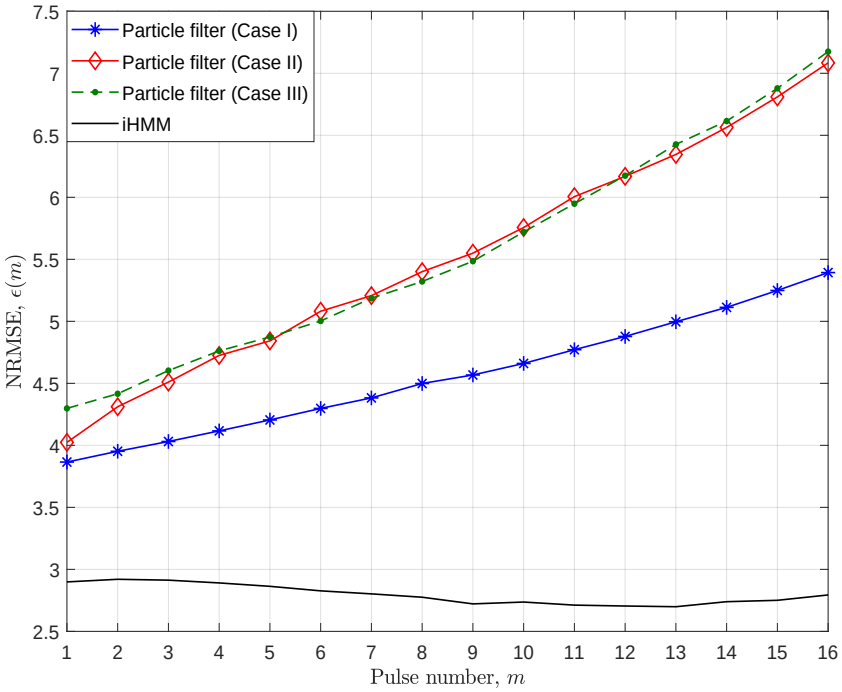


Figure 5.5 Estimation error of $K-$distributed TFR and clutter (LSS model).

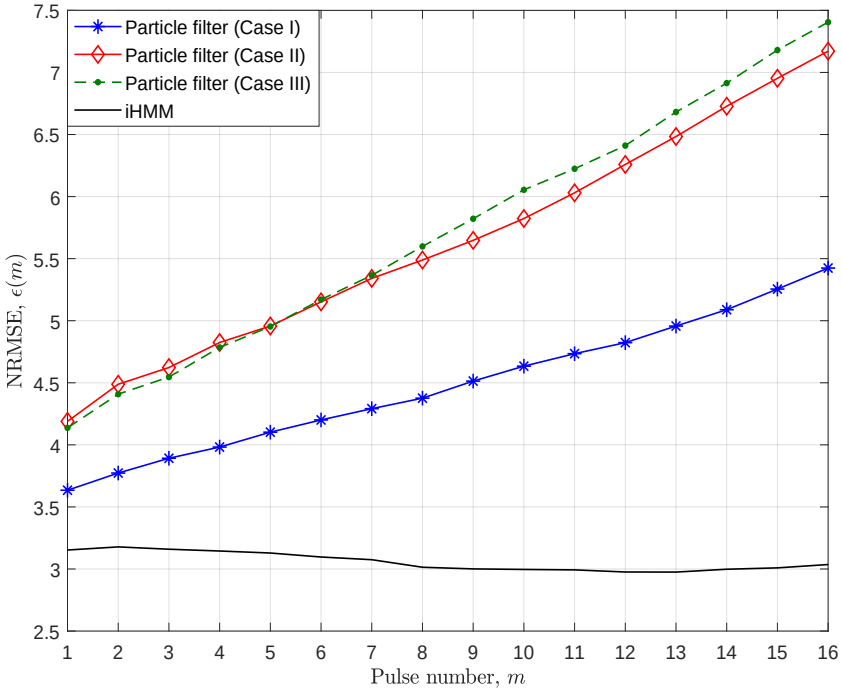


Figure 5.6 Estimation error of Gaussian TFR in $K-$distributed clutter (LSS model).

and it does not increase with m; it even slightly decreases. Fig. 5.6 shows the performance for the Gaussian $\mathbf{h}^{(m)}$ with $K-$distributed $\mathbf{c}^{(m)}$. It is observed that the iHMM shows the same error trend as for the $K-$distributed $\mathbf{h}^{(m)}$, but with a slight increase in ϵ at all m. While the PF introduces a lower ϵ at $m = 1$ relative to its error with the $K-$distributed $\mathbf{h}^{(m)}$, it reaches approximately the same error at the $m = 16$ for the three cases considered. It should be emphasized that the iHMM performance is obtained with no prior knowledge about the $\mathbf{h}$, $\mathbf{c}$, or $\mathbf{n}$. Fig. 5.7 shows the NRMSE of the TRF under the Log-normal distribution. It

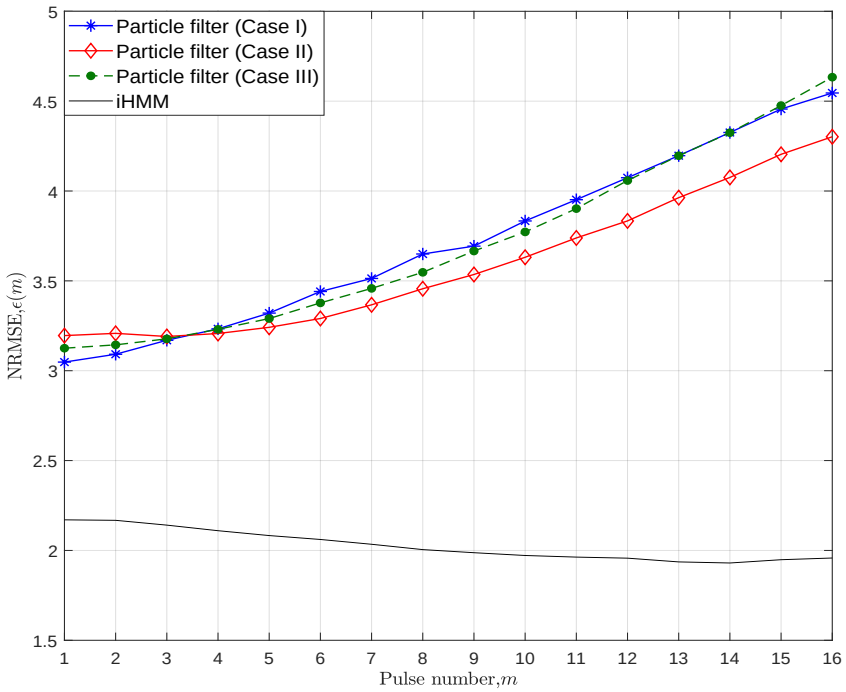


Figure 5.7 Estimation error of Log-Normal TFR in $K-$distributed clutter (LSS model).

is obvious that the estimation NRMSE of the Log-normal TRF is lower for both the PF and the proposed iHMM-based algorithm. Other than that, the PF continues to show the same estimation error divergence while the proposed algorithm keeps the error below that of the PF and lower than the initial value at different m. It is observed also that there is no substantial performance advantage obtained by providing the PF partial knowledge about the TRF over not having this knowledge at all.

The estimation performance of the PF under the Weibull distribution of the TRF is different from the other distributions in the first few pulses. As observed in Fig. 5.8, the

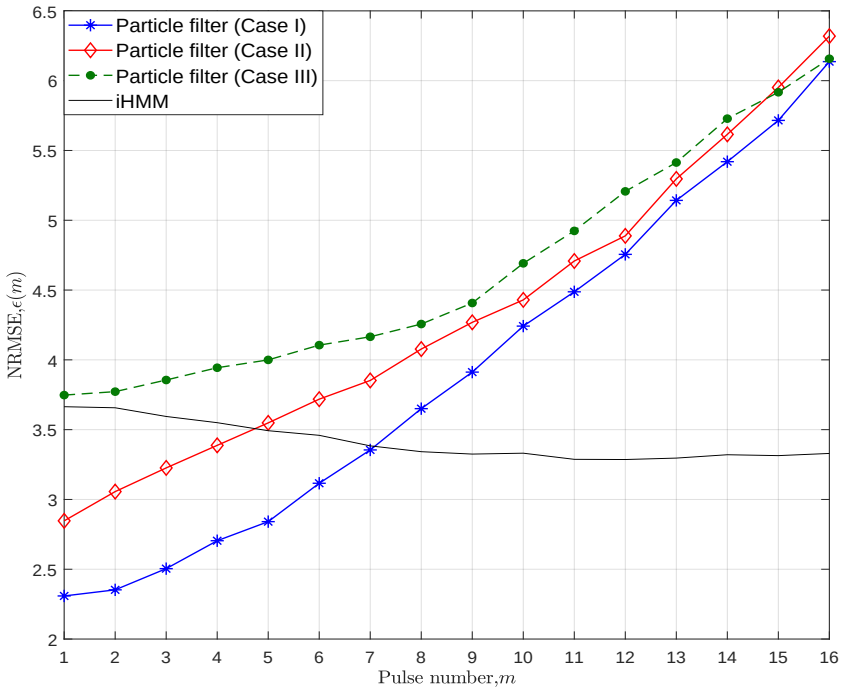


Figure 5.8 Estimation error of Weibull TFR in $K-$distributed clutter (LSS model).

PF shows the lowest NRMSE across all its three cases, especially in Case I. However, this relative advantage does not last for more than 4 pulses for Case II and III and 7 pulses for Case I, out of the considered 16 pulses. Beyond that, the proposed algorithm offers the lowest NRMSE that is approximately half of that of the PF with all its scenarios.

However, the relative advantage of the PF at the first few pulses given a Weibull TRF is not obtainable under jamming conditions. As we have shown in Fig. 5.4 for the Gaussian TRF, Fig. 5.9 depicts the performance of the proposed algorithm compared to the PF in the presence of both PNJ and CNJ. The PF is assumed in Case I. As can be observed in the figure, the NRMSE of the PF is increased under PNJ and CNJ at the first pulses to be higher than that of the proposed method. The proposed method maintains the same NRMSE level for both jamming techniques, while the PF shows more degradation under CNJ.

Based on the presented results under the LSS model, the KF and PF suffer from tracking-error divergence in all considered cases of TIR and clutter, attributed to their sensitivity to model mismatch. Moreover, when the jamming scenario is considered, both filters are vulnerable to smart noise jamming as CNJ while the KF is also susceptible to the classic

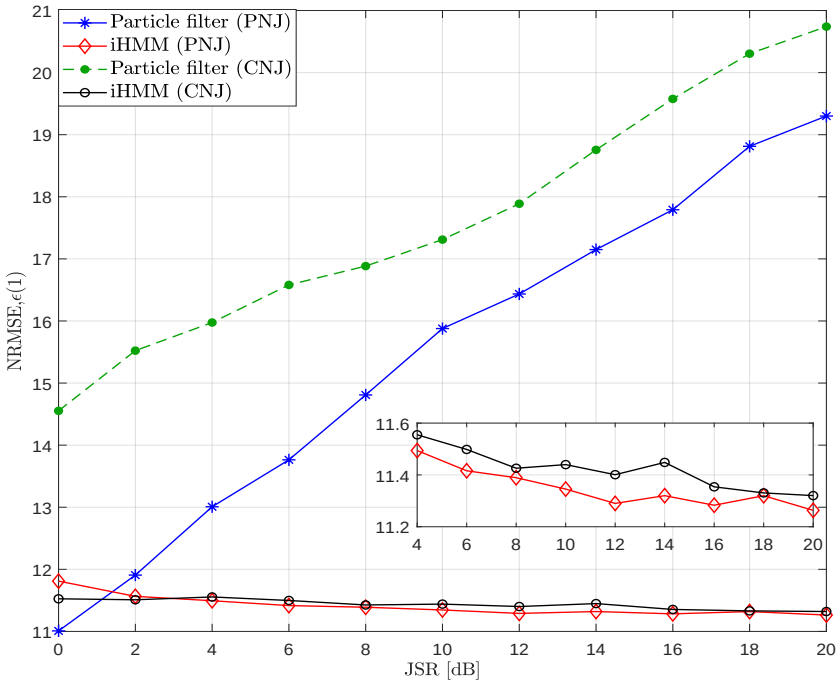


Figure 5.9 Estimation error of Weibull TFR in jamming and K-distributed clutter ($m = 1$, SIRP model).

PNJ. On the contrary, the proposed iHMM shows lower and steady state estimation error of TIR and better resistance to both the classic and smart noise jamming.

5.6.2.2 SIRP TRF Model

In the following, we explore the TRF estimation accuracy using the SIRP model for the proposed iHMM-based algorithm compared to the PF. It should be noted that the use of the KF under the SIRP model is not optimal because of the violation of the KF assumptions of linearity and/or Gaussian distribution. The correlation coefficient of the TIR under the SIRP model, ρ_h in Eq. (5.6), is assumed to be 0.5.

In Fig. 5.10 we can see the effect of using the SIRP TRF model on the estimation accuracy of the considered algorithms. For the PF, there are two main observations. The first observation from Fig. 5.10 is the estimation NRMSE, which is approximately fourfold that under the LSS TRF model for the three cases of the PF—complete, partial, and no knowledge. The second observation regarding the PF is the trend of the NMRSE with the pulse number m, which is shown to change randomly around a mean value for the three

knowledge-based scenarios of the PF. This can be explained by the fact that the PF does not assume a linear model anymore, which causes the accumulation of the error from pulse to pulse. Instead, the PF samples from a random process, SIRP, at each pulse without using the particles of the previous pulse as in the linear model, except for the weights as shown in Eq. (5.34).

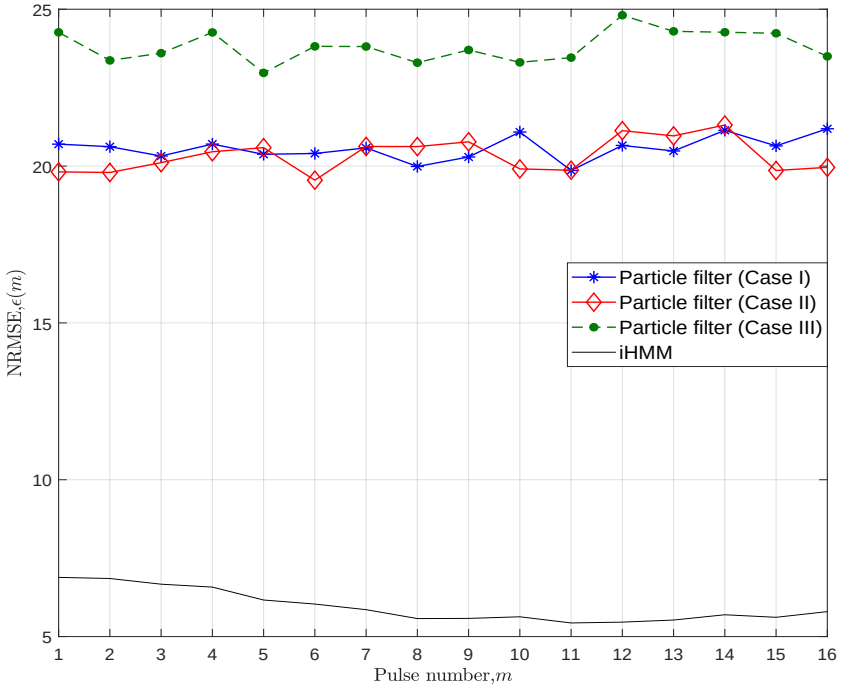


Figure 5.10 Estimation error of Gaussian TFR in Gaussian clutter (SIRP model).

For the proposed iHMM-based algorithm we can notice that the estimation error compared to the LSS TRF model for the Gaussian distribution is approximately doubled, but it preserves its trend in lowering the estimation error from the initial one. It should be emphasized that the parameters of the proposed algorithm are fixed for the two assumed TRF models, which means there is still a possibility of parameter adaptation in the case of the availability of training data even without prior information about the exact generating model or the TRF distribution.

The same can be concluded for the K-distribution from Fig. 5.11, but with the estimation error of the PF being higher than twice that of LSS TRF for the same distribution. For the iHMM-based TRF model, it keeps the same trend of the NRMSE with m, but it is

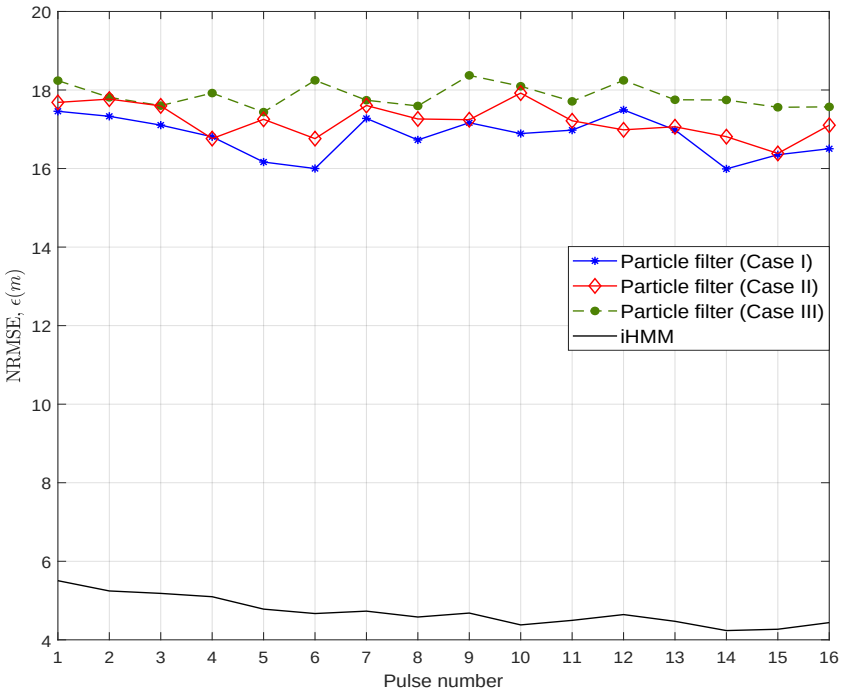


Figure 5.11 Estimation error of $K-$distributed TFR and clutter (SIRP model).

approximately below one third of that of the PF under the same conditions. We can also notice that there is no significant effect on the PF accuracy under K-distributed TRF with the filter having the prior knowledge of the TRF distribution, which was not the case under the Gaussian TRF.

For the Log-normal distribution, we can see that it still shows the easiest challenge for the TRF estimation among the other considered distributions. We can also notice that the NRMSE introduced by the particle filter is at least three times that introduced by the proposed iHMM-based algorithm. While lacking the prior knowledge about the TRF distribution increases the NRMSE by about 15% compared to having this knowledge, there is no noticeable difference between Case II and the Case I of the PF.

The Weibull distributed TFR under the SIRP shows higher NRMSE for the two considered algorithms as depicted in Fig. 5.13. However, while both the PF and iHMM-based algorithm start with the same value of NRMSE, the iHMM-based algorithm can lower the NRMSE with about 30% over the 16 pulses. Moreover, without complete prior knowledge about the TRF, the NRMSE of the PF is noticeably increased to be more than 150% that

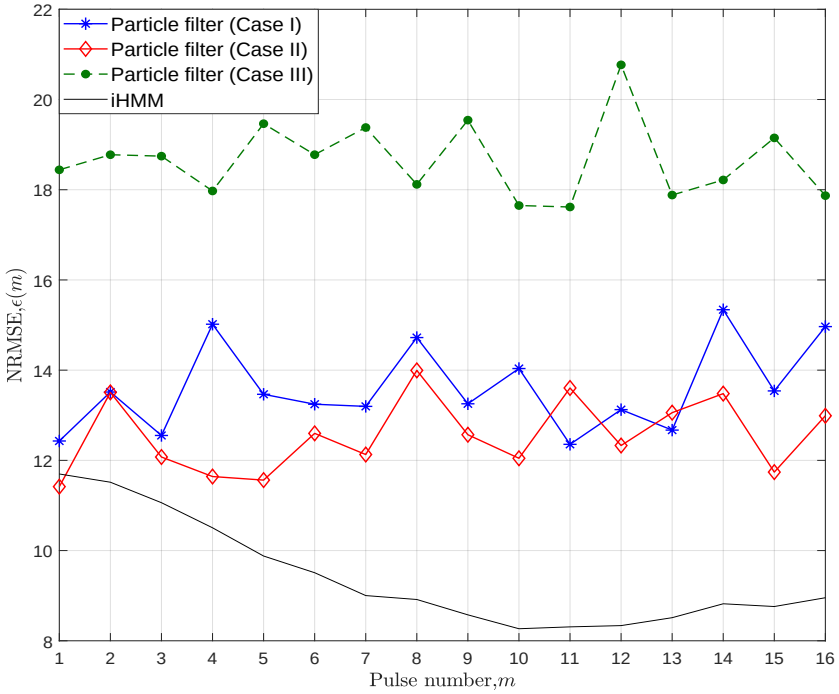


Figure 5.12 Estimation error of Log-normal TFR and $K-$distributed clutter (SIRP model).

of the iHMM-based algorithm.

In general, we can clearly see that the SIRP TRF model is more challenging than the simple LSS model assumed in the literature heretofore. This highlights our motivation behind proposing the SIRP model for the TRF not only for its generality over that of the LSS, but also for the challenge it poses for the TRF estimation. While the proposed algorithm does not adapt its parameters for the SIRP TRF model, it can keep its NRMSE trend as that of the LSS model but at different values, with the Weibull TRF having the highest bias. As shown before for the LSS TRF model, the proposed iHMM-based algorithm guarantees the lowest NRMSE at all pulses compared to the simulated PF.

In general, for both the LSS and the SIRP models of the TRF we can see how the particle and Kalman filters suffer from the model mismatch due to the fact that the TRF is a random process without any deterministic model for the dynamic state evolution. As we mentioned in Section 2.6, the particle filter needs a deterministic state model, to generate particles that describe the actual system. In the absence of this deterministic model, the particle filter draws its particles from a random process, whose realizations differ in actual values from the

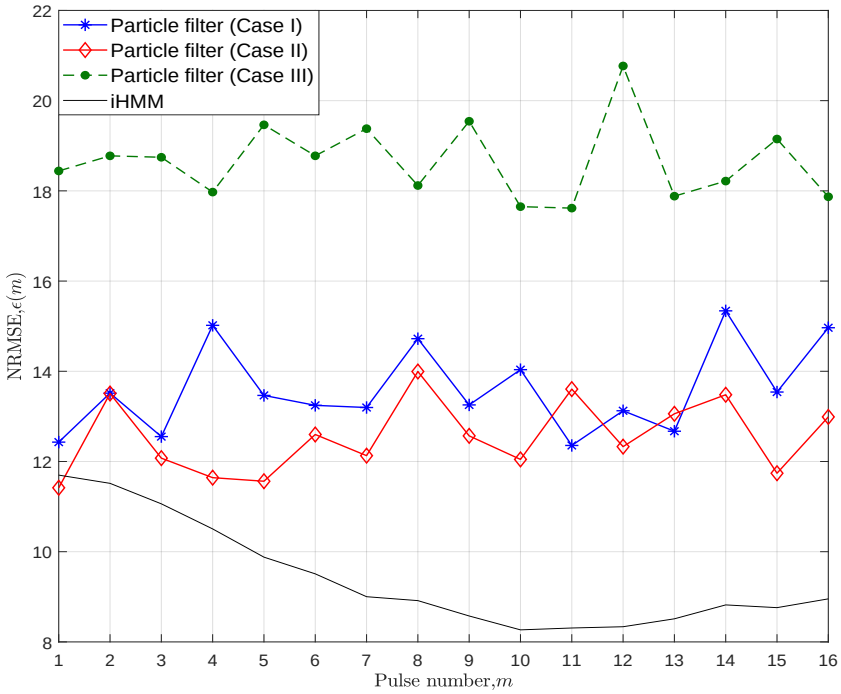


Figure 5.13 Estimation error of Weibull TFR and $K-$distributed clutter (SIRP model).

true TRF.

5.6.2.3 Complexity Analysis

After the comparative evaluation of the proposed iHMM estimation of the TRF, we evaluate the complexity of the proposed method compared to the PF for the two considered generating models and four distributions. As we observed in the performance assessment introduced before, the KF is not optimal to be applied except for the special case of the Gaussian LSS model only, therefore we exclude it from the complexity analysis. For this purpose, we dedicated a machine with 64 bit Intel® Core™ i7-8850H CPU @2.6GHz and 32 GB RAM. In the following simulations, we assumed that the PF in Case I. Fig. 5.14 depicts the execution time of the proposed algorithm compared to that of the PF in the case of the LSS with different distributions. As we can observe in the figure, the time complexity of the PF does not depend on the pulse number m; however, it strongly depends on the distribution of the TRF the PF samples from. As we may anticipate, both the Gaussian and the Log-normal distributions have approximately the same execution time as the generation

operations of the TRF using both distributions are similar to each other. Nevertheless, in the case of the K-distributed TRF we can see that the execution time increases compared to the Gaussian and Log-normal distributions. For the Weibull distributed TRF, it is obvious that it has the highest execution time among the other distributions. This is attributed to the Rejection Method used in generating the modulating scalar random variable of the Weibull SIRV compared to that of the K-distribution.

For the proposed iHMM-based method, the execution time is independent from the distribution, that is why we display a single curve for the proposed method. This is consistent with its nonparametric nature that does not make any assumptions about the TRF or interference. However, the execution time of the proposed method depends on m, which can be explained by the formula we proposed for the auxiliary variable u_l. Recalling that $u_l = 10^{-\kappa}\mathcal{U}(0, \Pr(h_t|h_{t-1}))$, with $\kappa = 0.15m$, the auxiliary variable for all the frequency bins is inversely proportional to m. This setting of u_l allows the proposed model to start with a low number of states and then consider more states as m increases according to the parameters learned from the environment. Therefore, as m increases, the number of considered states also gets higher, which explains the increase of the execution time. Nevertheless, it is obvious that the average execution time of the proposed algorithm is lower than that of the PF for all distributions. Moreover, we assumed the cognitive cycle[6] of the radar is 16 pulses, which is relatively long, to investigate the cumulative estimation error of the proposed algorithm compared to the benchmark estimators at the worst case scenario. In real applications, it is preferred to have shorter cognitive cycles, which means the proposed algorithm will maintain the lowest execution time and error at the same time for all m.

Approximately, the same performance is obtained under the correlated SIRP model shown in Fig. 5.15. We can see that the PF execution time is also sensitive to the generating model of the TRF it samples from and the correlated SIRP model increases the execution time. In addition, we can also observe that the execution time required for the Weibull distributed TRF has been substantially increased compared to the same TRF distribution under the LSS model. This is attributed to the fact that the PF is now forced to sample from the Weibull distribution for each pulse for all the used 50 particles. The proposed method under the correlated SIRP generating model of the TRF has the same execution time as the LSS model and it appears to be lower than that of the PF even at $m = 16$ for the Weibull and

[6]The cognitive cycle of the radar is the time required to update the transmitter waveform through learning from the environment.

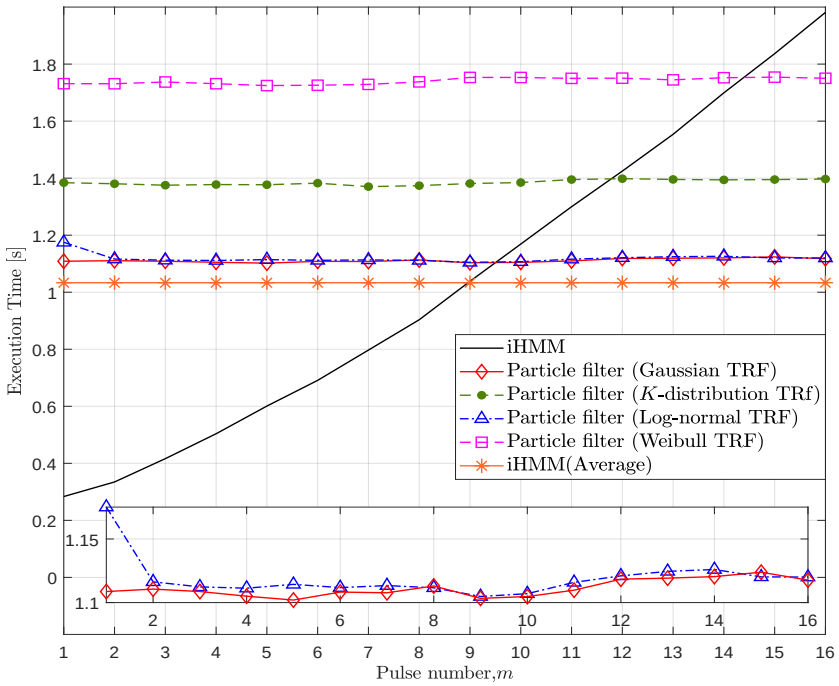


Figure 5.14 Complexity of the proposed iHMM-based method compared to the PF assuming the LSS model.

K-distributions and slightly higher than that of the Gaussian and Log-normal distributions. However, on average, the proposed method has a lower execution time than the PF at all distributions.

5.7 Conclusion

In this chapter we delved into the problem of Gaussian and non-Gaussian TFR estimation under the assumption of Gaussian and non-Gaussian clutter. We have also proposed a new generating model for the TRF based on the SIRP, which is more general and challenging than the LSS assumed in the literature heretofore. In this context, the performance of the nonparametric Bayesian framework represented by the iHMM is compared to the classic Bayesian frameworks of the KF and PF. Compared to the latter filters, the iHMM improves the tracking accuracy of the TFR without any prior knowledge about its statistics or that of the interference, even if the filters know completely the statistical parameters of the TFR and interference. This robust performance is provided with a lower complexity compare to

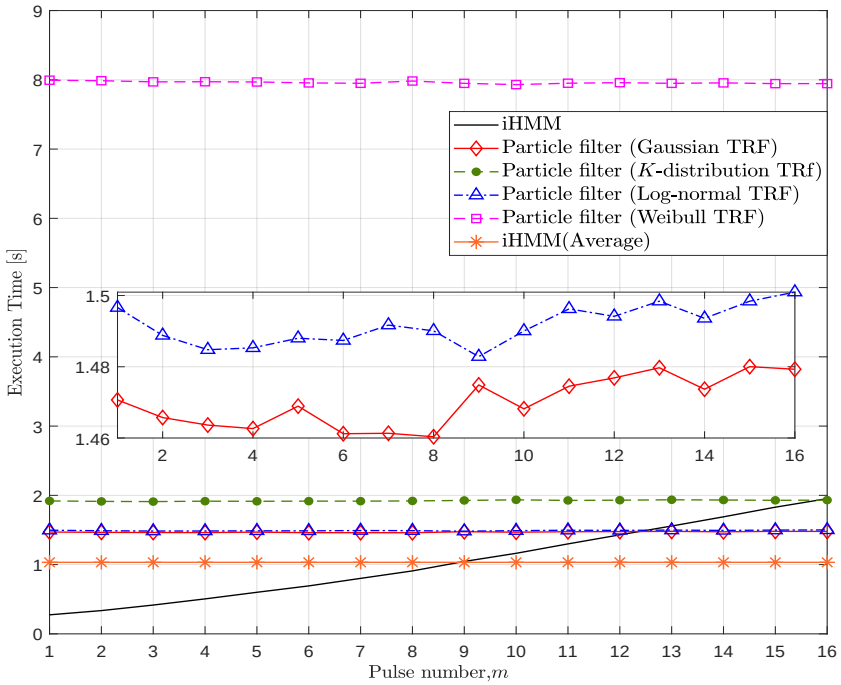


Figure 5.15 Complexity of the proposed iHMM-based method compared to the PF assuming the SIRP model.

the PF. These promising results encourage further research in employing the nonparametric Bayesian methods in cognitive radar applications.

Chapter 6

Conclusion and Future Work

In this thesis, we introduced advanced radar signal processing algorithms for enhancing the performance of phased array, MIMO, and cognitive MIMO radars considering the the real field operation constraints and the limitations of radar systems hardware.

In Chapter 3, we discussed the NHD problem for robust covariance matrix estimation. In this regard, we derived a new test statistic for the NHD based on the projection depth function, termed PD-NAMF. To ensure the convergence of sample version of the test of the proposed detector to its population version, we studied the convergence of the empirical CDF to the population CDF. Our analysis showed that before applying the PD-NAMF to the secondary cells, they need first to be decorrelated in the case of correlated clutter. To preserve the nonparametric feature of the PD-NAMF, we proposed employing the nonparametric Spearman and Kendall correlation matrices in the decorrelation operation. The performance of the proposed PD-NAMF has been shown to be comparable to that of the NAMF, but at much lower complexity. This performance has been demonstrated in different clutter distributions and radar parameters.

In Chapter 4, we considered a signal processing approach to overcome the problem of high reflection from the transmitting antenna back to the preceding amplification stage. Specifically, we developed a waveform design algorithm for cognitive MIMO radars, which maximizes the SINR and reduces the reflected power from the transmitting antenna array. While the TARC has been conventionally used as a measure of the reflection performance of antenna arrays of MIMO systems, we found that it is not sufficient to guarantee a low reflection coefficient for each antenna element and each pulse. Therefore, we used the ARC

as a more robust figure of merit to evaluate the performance of the proposed design method. We employed the ℓ_∞-norm of the reflected signal as a regulation term, which is a non-smooth function that hinders the use of the conventional gradient methods. To circumvent this issue, we used the proximal gradient method. Since there is no similar benchmark in the literature, we employed a competing method that handles the reflection from the transmitting antenna array through grouping the antenna elements into subarrays. Monte Carlo simulations showed that the proposed method guarantees low ARCs for all pulses and antenna elements and simultaneously improves the SINR to a comparable level to the original waveform design techniques for cognitive MIMO radars, yet with lower complexity.

It is assumed in Chapter 4 that the TIR is known *apriori*, which is not the case in practice. In Chapter 5, we discussed the problem of estimating the TFR from the target received signal. We introduced a new formulation to the estimation of the TRF using the iHMM and beam sampling inference, which does not require prior information about the distribution of the TRF or the interference. We extended the analysis further than the literature by assuming non-Gaussian distributions of the TRF and by investigating the estimation accuracy over multiple pulses. In addition, we proposed employing a new generating model for the TRF based on the SIRP that is more generic than the linear Gaussian state model. While there is no benchmark for the estimation of the non-Gaussian TRF in the literature, we used the particle filter as our benchmark in this case. The results showed that the proposed method exhibits superior estimation accuracy compared to both Kalman and particle filters at all TRF distributions and models under different clutter distributions and jamming signals. Equally important, the proposed algorithm requires a lower execution time compared to the particle filter at all distributions and generating models.

The research results described in this thesis emphasized the importance of considering the real-time operation of radar systems in devising radar signal processing techniques. The complexity and the execution times of the used algorithms should be reduced in addition to providing desirable theoretical properties and performance merits. The findings of this thesis also highlighted the need of using nonparametric statistical techniques, such as the depth functions and iHMM, in radar signal processing. Specifically, in most of radar scenarios encountered in practice, the prior information about the distribution of the underlying radar data or its parameters is not available. We have shown in Chapter 3 that the nonparametric techniques can achieve a comparable performance as the parametric techniques. Furthermore, nonparametric methods can be more efficient than their parametric counter-

parts when the model mismatches are severe, as we have shown in Chapter 5. A substantial outcome of this thesis is the significance of contemplating the RF hardware limitations of the radar system in signal processing algorithms. For instance, in Chapter 4 we have discussed how the microwave design techniques are used to reduce the effect of the reflection from the transmitting antenna back to the power amplifier. However, properly designed cognitive MIMO radar waveforms can reduce the reflected power itself, not just its effect, to safe levels, yet without compromising performance. Guerci has noticed the lack of interest within the signal processing community in considering the characteristics of the RF components. He attributed this scarce coverage to the difference in the backgrounds between the signal processing and RF hardware researchers. One year ago [4], he anticipated this situation to be changed in the future, we hope that we took the first step to bridge the gap between the two research communities.

Finally, we discuss some key avenues for future research. For the NHD, future research may consider the parallel implementation of the PD-NAMF, specifically the median and Kendall correlation matrix, which can lower the complexity of the proposed detector further. As we pointed out in Chapter 3, employing GPUs can provide more efficient implementations for different ordered-statistics measures. Regarding the power-efficient design of cognitive MIMO radar waveforms, we found the basis for a new signal-processing approach that deals with problem of the reflection from the transmitting antenna array. However, this paves the way for more optimization methods that may provide the same good ARCs, yet with a better SINR performance. Moreover, the research should also consider the experimental evaluation of the ARC for cognitive MIMO radar waveforms, keeping in mind that the available network analyzers measure the reflection coefficient using the standard continuous wave (CW) signal only. With the joint work between researchers from both the microwave and signal processing communities can lead to more efficient cognitive MIMO radars that achieve, at the same time, the desirable performance quality. It is also important to consider the performance of the proposed method under uncertainty of the TIR estimation. Regarding the third contribution of this thesis, the new formulation of the TRF estimation using the iHMM opens the door for more inference algorithms that may offer better estimation accuracy, especially with high-dimensional signals. A promising and growing area of research in the field of Bayesian inference is the variational inference [233]. Variational inference embodies different algorithms, for instance, stochastic variational inference [234] and variational Hamiltonian Monte Carlo inference [235]. Lastly, in this thesis, we introduced three contributions in three

subsystems in the cognitive MIMO radar and analyzed the performance of each technique separately. It is still an open question, and an interesting one, to explore the closed-loop performance of the whole system using the proposed techniques compared to using the conventional one.

Appendix A

Proof of Proposition 3.1

Assuming that the covariance matrix $\hat{\mathbf{R}}$ is positive definite, it is invertible and admits of a square-root $\hat{\mathbf{R}}^{1/2}$, which is also invertible. Hence, applying the Cauchy-Schwartz inequality

$$\begin{aligned} |\mathbf{u}^H\mathbf{z}_k|^2 = \left|\mathbf{u}^H\hat{\mathbf{R}}^{\frac{1}{2}}\hat{\mathbf{R}}^{-\frac{1}{2}}\mathbf{z}_k\right|^2 &\leq \left\|\mathbf{u}^H\hat{\mathbf{R}}^{\frac{1}{2}}\right\|_2^2\left\|\hat{\mathbf{R}}^{-\frac{1}{2}}\mathbf{z}_k\right\|_2^2 \\ &\leq (\mathbf{u}^H\hat{\mathbf{R}}\mathbf{u})(\mathbf{z}_k^H\mathbf{R}^{-1}\mathbf{z}_k) \end{aligned} \tag{A.1}$$

Consequently, for any non-zero vector $\mathbf{u} \in \mathbb{C}^J$, we have

$$\frac{|\mathbf{u}^H\mathbf{z}_k|^2}{\mathbf{u}^H\hat{\mathbf{R}}\mathbf{u}} \leq \mathbf{z}_k^H\hat{\mathbf{R}}^{-1}\mathbf{z}_k \tag{A.2}$$

Hence,

$$\sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}^H\mathbf{z}_k|^2}{\mathbf{u}^H\hat{\mathbf{R}}\mathbf{u}} = \mathbf{z}_k^H\hat{\mathbf{R}}^{-1}\mathbf{z}_k \tag{A.3}$$

Among the possible scale measures of the scalar random variable $\mathbf{u}^H\mathbf{z}_k$, let us consider the sample variance $\hat{\sigma}^2(\mathbf{u}^H\mathbf{Z}_k) = \mathbf{u}^H\hat{\mathbf{R}}\mathbf{u}$. Then according to Eq. (A.3)

$$\sup_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H\mathbf{z}_k|}{\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2 = \mathbf{z_k}^H\hat{\mathbf{R}}^{-1}\mathbf{z_k} \tag{A.4}$$

Based on Eq. (3.5), $\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)$ is given by

$$\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k) = 2\hat{\sigma}(\Re(\mathbf{u}^H\mathbf{Z}_k)) = 2\hat{\sigma}(\Im(\mathbf{u}^H\mathbf{Z}_k)) \tag{A.5}$$

Repeating the same steps from Eq. (A.1) to Eq. (A.4), but with replacing $\mathbf{z}_k$ by $\mathbf{s}$, a similar relation can be proven for the steering vector $\mathbf{s}$, that is

$$\sup_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H\mathbf{s}|}{\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2 = \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \tag{A.6}$$

For the centralized vector $\mathbf{z}_k - \hat{\boldsymbol{\mu}}$, we can write

$$\sup_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H\mathbf{z}_k - \hat{\mu}|}{\hat{\sigma}(\mathbf{u}^H\mathbf{Z}_k)}\right)^2 = (\mathbf{z_k} - \hat{\boldsymbol{\mu}})^H\hat{\mathbf{R}}^{-1}(\mathbf{z_k} - \hat{\boldsymbol{\mu}}) \tag{A.7}$$

Appendix B

Proof of Proposition 3.2

Considering the k-th secondary cell $\mathbf{z}_k = a\mathbf{s} + \mathbf{c} + \mathbf{n}$, where for convenience we let $\mathbf{s} = [s_1, \cdots, s_J]^T$, $\mathbf{c} = [c_1, \cdots, c_J]^T$, and $\mathbf{n} = [n_1, \cdots, n_J]^T$, then we have

$$\mathbf{s}^H \mathbf{z}_k \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} = \left(a \sum_{i=1}^{J} s_i s_i^* + \sum_{i=1}^{J} s_i^* \left(c_i + n_i \right) \right) \cdot \sum_{i=1}^{J} \left(s_i^* \sum_{j=1}^{J} p_{ij} s_j \right) \tag{B.1}$$

where $\hat{\mathbf{R}}^{-1} = [p_{ij}]$. From Eq. (2.9), $\sum_{i=1}^{J} s_i s_i^* = 1$, hence

$$\mathbf{s}^H \mathbf{z}_k \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} = a \sum_{i=1}^{J} s_i^* \sum_{j=1}^{J} p_{ij} s_j + \sum_{j=1}^{J} \left(s_j^* \sum_{i=1}^{J} \left(s_i^* \sum_{k=1}^{J} p_{ik} s_k \right) \left(c_j + n_j \right) \right) \tag{B.2}$$

$$\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{z}_k = a \sum_{i=1}^{J} s_i^* \sum_{j=1}^{J} p_{ij} s_j + \sum_{i=1}^{J} s_i^* \sum_{j=1}^{J} p_{ij} \left(c_j + n_j \right) \tag{B.3}$$

From Eq. (B.2) and Eq. (B.3) we can observe that the two tests have the same target signal component (i.e., the first term in each equation, which is equivalent to $a(\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s})$), but they differ in the interference component. There is no analytical way to compare the

interference components in the two tests due to the different random weights of each term, hence, we rely on simulation to compare them. Nevertheless, as Eq. (B.2) considers only the interference component in the spatio-temporal direction of the target, its average interference power is anticipated to be lower than that of Eq. (B.3), which is confirmed by simulation in Section 3.6.

Appendix C

Proof of Proposition 3.3

Using the term in Eq. (3.24), we obtain the modified NAMF test statistic

$$\Lambda'_{\text{NAMF}} = \frac{|\mathbf{s}^H \hat{\mathbf{R}}^{-1}(\mathbf{s}^H \mathbf{z}_k)\mathbf{s}|^2}{(\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s})(\mathbf{z}_k^H \hat{\mathbf{R}}^{-1}\mathbf{z}_k)} \tag{C.1}$$

Based on Eq. (A.6) we have

$$\left|(\mathbf{s}^H \mathbf{z}_k)\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}\right|^2 = \left|(\mathbf{s}^H \mathbf{z}_k)\Big(\sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}^H \mathbf{s}|}{\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)}\Big)^2\right|^2 \tag{C.2}$$

By substitution of Eq. (A.4), Eq. (A.6), and Eq. (C.2) into Eq. (C.1) and after simple manipulations, Λ'_{NAMF} reduces to

$$\Lambda'_{\text{NAMF}} = \frac{|\mathbf{s}^H \mathbf{z}_k|^2 \sup\limits_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H \mathbf{s}|}{\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)}\right)^2}{\sup\limits_{\|\mathbf{u}\|=1} \left(\frac{|\mathbf{u}^H \mathbf{z}_k|}{\hat{\sigma}(\mathbf{u}^H \mathbf{Z}_k)}\right)^2} \tag{C.3}$$

Appendix D

Proof of Proposition 4.1

Proof. It is known that the proximal operator of the $\|\mathbf{f}\|_\infty$ is the projection of $\mathbf{f}$ onto the unit ℓ_1 ball [58]. Since the penalty term in Eq. (4.30), $\|\mathbf{S}_d\mathbf{f}\|_\infty$, is a function of the linear transformation of $\mathbf{f}$, we seek to find the proximal operator of $w(\mathbf{f}) = y(\mathbf{S}_d\mathbf{f})$ with respect to $y(\mathbf{f})$. In this case, prox_w is defined as the solution of the following minimization problem

$$\begin{aligned} &\min_{\mathbf{u}\in\mathbb{C}^{N_T L_s}} \quad \{w(\mathbf{u}) + \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2\} \\ &\min_{\mathbf{u}\in\mathbb{C}^{N_T L_s}} \quad \{y(\mathbf{S}_d\mathbf{u}) + \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2\} \end{aligned} \tag{D.1}$$

The problem in Eq. (D.1) can be formulated as

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{z}\in\mathbb{C}^{N_T L_s}} \quad \{y(\mathbf{z}) + \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2\} \\ &\text{subject to} \quad \mathbf{z} = \mathbf{S}_d\mathbf{u} \end{aligned} \tag{D.2}$$

Let the solution for the above problem is $(\tilde{\mathbf{u}}, \tilde{\mathbf{z}})$. Fixing $\mathbf{z} = \tilde{\mathbf{z}}$ and solving for $\mathbf{u}$, we have

$$\begin{aligned} &\min_{\mathbf{u}\in\mathbb{C}^{N_T L_s}} \quad \frac{1}{2}\|\mathbf{u}-\mathbf{f}\|_2^2 \\ &\text{subject to} \quad \tilde{\mathbf{z}} = \mathbf{S}_d\mathbf{u} \end{aligned} \tag{D.3}$$

The Lagrange function for this problem is given by

$$\mathfrak{L}(\mathbf{u}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{u} - \mathbf{f}\|_2^2 + (\mathbf{S}_d\mathbf{u} - \tilde{\mathbf{z}})^T\boldsymbol{\lambda} \tag{D.4}$$

where $\boldsymbol{\lambda}$ is the complex Lagrange multiplier vector. From the properties of complex Wirtinger derivatives and considering the real valued Lagrangian function $\mathfrak{L}(\mathbf{u}, \boldsymbol{\lambda})$, we have [236, Result A2.3]

$$\nabla_{\mathbf{u}}\mathfrak{L}(\mathbf{u}, \boldsymbol{\lambda}) = 0 \Leftrightarrow \frac{\partial\mathfrak{L}(\mathbf{u}, \boldsymbol{\lambda})}{\partial\mathbf{u}} = 0 \tag{D.5}$$

Therefore, by setting $\frac{\partial\mathfrak{L}(\mathbf{u},\boldsymbol{\lambda})}{\partial\mathbf{u}} = 0$, we obtain

$$\tilde{\mathbf{u}} = \mathbf{f} - \mathbf{S}_d^T\boldsymbol{\lambda} \tag{D.6}$$

Substituting Eq. (D.6) in the constraint of Eq. (D.2), we get

$$\tilde{\mathbf{z}} = \mathbf{S}_d(\mathbf{f} - \mathbf{S}_d^T\boldsymbol{\lambda}) \tag{D.7}$$

Then, we have

$$\boldsymbol{\lambda} = (\mathbf{S}_d\mathbf{S}_d^T)^{-1}(\mathbf{S}_d\mathbf{f} - \tilde{\mathbf{z}}) \tag{D.8}$$

Substituting Eq. (D.8) in Eq. (D.6) we have

$$\begin{aligned}\tilde{\mathbf{u}} &= \mathbf{f} - \mathbf{S}_d^T(\mathbf{S}_d\mathbf{S}_d^T)^{-1}(\mathbf{S}_d\mathbf{f} - \tilde{\mathbf{z}}) \\ &= \mathbf{S}_d^{T^{-1}}\tilde{\mathbf{z}}\end{aligned} \tag{D.9}$$

Now, solving for $\tilde{\mathbf{z}}$

$$\tilde{\mathbf{z}} = \underset{\mathbf{z}\in\mathbb{C}^{N_T L_s}}{\operatorname{argmin}}\{y(\mathbf{z}) + \frac{1}{2}\|\mathbf{S}_d^{T^{-1}}\mathbf{z} - \mathbf{f}\|_2^2\} \tag{D.10}$$

The problem in Eq. (D.10) can be equivalently expressed as

$$\tilde{\mathbf{z}} = \underset{\mathbf{z}\in\mathbb{C}^{N_T L_s}}{\operatorname{argmin}}\{y(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{S}_d\mathbf{f}\|_2^2\} \tag{D.11}$$

Comparing Eq. (D.11) with the definition of the proximal operator in Eq. (4.28), we have

$$\tilde{\mathbf{z}} = \text{prox}_y(\mathbf{S}_d\mathbf{f}) \tag{D.12}$$

Since $\tilde{\mathbf{u}}$ is the solution for Eq. (D.1), we can write

$$\text{prox}_w(\mathbf{f}) = \mathbf{S}_d^{T-1}\text{prox}_y(\mathbf{S}_d\mathbf{f}) \tag{D.13}$$

When $y(\mathbf{f})$ is defined as the ℓ_∞-norm of its argument, then $\text{prox}_y(\mathbf{f})$ is given by

$$\text{prox}_y(\mathbf{f}) = \mathbf{f} - \text{Proj}_{\|\cdot\|_1<1}(\mathbf{f}) \tag{D.14}$$

Therefore, using the proximal operator property in Eq. (4.29) the proximal operator of $\gamma_2 y(\mathbf{f})$ is [20]

$$\text{prox}_{\gamma_2 y}(\mathbf{f}) = \mathbf{f} - \gamma_2\text{Proj}_{\|\cdot\|_1<1}(\mathbf{f}/\gamma_2) \tag{D.15}$$

Substituting with Eq. (D.15) in Eq. (4.30), we obtain Eq. (4.34) and the proof completes. □

Appendix E

Proof of Lipschitz continuity of $\nabla u(\mathbf{f})$ in Eq. (4.33)

Proof. We begin by the definition of the Lipschitz continuity for a complex function $\mathbf{r} : \mathbb{C}^{N_T L_s} \to \mathbb{C}^{N_T L_s}$ that is given by [237]

$$\|\mathbf{r}(\mathbf{f}) - \mathbf{r}(\mathbf{g})\|_2 \leq \Lambda \|\mathbf{f} - \mathbf{g}\|_2 \qquad \forall \mathbf{f}, \mathbf{g} \in \mathbb{C}^{N_T L_s} \tag{E.1}$$

where $\Lambda \in \mathbb{R}^+$ with $\mathbb{R}^+$ denoting the set of non-negative real numbers. Defining $\mathbf{r}(\mathbf{f})$ and $\mathbf{r}(\mathbf{g})$ as $\nabla_{\mathbf{f}} u(\mathbf{f})$ and $\nabla_{\mathbf{g}} u(\mathbf{g})$, respectively, and defining $\mathbf{R} \triangleq \mathbf{R}_{c,h} + \mathbf{h}^H \mathbf{R}_v \mathbf{h} \mathbf{I}_{N_T L_s}$ then we have

$$\|\mathbf{r}(\mathbf{f}) - \mathbf{r}(\mathbf{g})\|_2 = \|(\mathbf{f}^H - \mathbf{g}^H)\mathbf{R} + \mu \mathbf{h}^H \mathbf{T}(\mathbf{f} - \mathbf{g})\mathbf{h}^H \mathbf{T}\|_2 \tag{E.2}$$

Hence

$$\|\mathbf{r}(\mathbf{f}) - \mathbf{r}(\mathbf{g})\|_2 \leq \|(\mathbf{f}^H - \mathbf{g}^H)\mathbf{R}\|_2 + \mu \|\mathbf{h}^H \mathbf{T}(\mathbf{f} - \mathbf{g})\mathbf{h}^H \mathbf{T}\|_2 \tag{E.3}$$

Equivalently, we can write that

$$\|\mathbf{r}(\mathbf{f}) - \mathbf{r}(\mathbf{g})\|_2 \leq \|(\mathbf{f}^H - \mathbf{g}^H)\|_2 \|\mathbf{R}\|_F + \mu \|\mathbf{h}^H \mathbf{T}\|_2^2 \|(\mathbf{f} - \mathbf{g})\|_2 \tag{E.4}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Therefore

$$\|\mathbf{r}(\mathbf{f}) - \mathbf{r}(\mathbf{g})\|_2 \leq (\|\mathbf{R}\|_F + \mu\|\mathbf{h}^H\mathbf{T}\|_2^2)\|(\mathbf{f} - \mathbf{g})\|_2 \tag{E.5}$$

where $(\|\mathbf{R}\|_F + \mu\|\mathbf{h}^H\mathbf{T}\|_2^2) \geq 0$. Comparing Eq. (E.5) with Eq. (E.1) we can conclude that $\nabla u(\mathbf{f})$ is Lipschitz continuous with constant $\Lambda = \|\mathbf{R}\|_F + \mu\|\mathbf{h}^H\mathbf{T}\|_2^2$. In real scenarios, the values of $\|\mathbf{R}\|_F$ and $\|\mathbf{h}^H\mathbf{T}\|_2^2$ is much smaller than unity. Therefore, with the value assigned to μ in this work, $\mu = 0.001$, it is easy to find $\beta, \beta^1 \in (0, 1/\Lambda]$ for fast convergence. □

Bibliography

[1] O. Blumtritt, H. Petzold, and W. Aspray, *Tracking the history of radar*. New Jersey, NJ: IEEE-Rutgers Center for the History of Electrical Engineering, 1994.

[2] S. Han, A. De Maio, V. Carotenuto, L. Pallotta, and X. Huang, "Censoring outliers in radar data: An approximate ml approach and its analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, pp. 534–546, April 2019.

[3] J. Fan, Y. Liao, and H. Liu, "An overview of the estimation of large covariance and precision matrices," *The Econometrics J.*, vol. 19, no. 1, pp. C1–C32, 2016.

[4] J. Bergin and J. R. Guerci, "Summary and future work," in *MIMO radar: Theory and applications*, London, UK: Artech House, 2018.

[5] M. van Heijningen, G. van der Bent, E. H. van der Houwen, A. Chowdhary, and F. E. van Vliet, "L-band AlGaN/GaN power amplifier with protection against load mismatch," in *European Microwave Conference*, (Nuremberg, Germany), pp. 1379–1382, Oct 2013.

[6] Y. Yao, J. Zhao, and L. Wu, "Waveform optimization for target estimation by cognitive radar with multiple antennas," *Sensors*, vol. 18, no. 6, p. 1743, 2018.

[7] A. Aubry, J. Carretero-Moya, A. D. Maio, A. Pauciullo, J. Gismero-Menoyo, and A. Asensio-Lopez, "Detection of extended target in compound-Gaussian clutter," in *Modern Radar Detection Theory*, Radar, Sonar & Navigation, pp. 333–374, Institution of Engineering and Technology (IET), 2015.

[8] Ö. D. Akyildiz and J. Míguez, "Nudging the particle filter," *Statistics and Computing*, pp. 1–26, Jul. 2019.

[9] M. I. Skolnik, ed., *Radar Handbook*. London, UK: McGraw-Hill, 2008.

[10] M. Budge and S. German, *Basic Radar Analysis*. Artech House Radar Library, London, UK: Artech House, 2015.

[11] J. Guerci, *Space-time Adaptive Processing for Radar*. Artech House Radar Library, London, UK: Artech House, 2014.

[12] J. S. Goldstein and I. S. Reed, "Theory of partially adaptive radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 1309–1325, Oct. 1997.

[13] I. S. Reed, Y. L. Gau, and T. K. Truong, "CFAR detection and estimation for STAP radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 722–735, 1998.

[14] W. L. Melvin, "Space-time adaptive processing for radar," in *Commun. and Radar Signal Process.* (R. C. Nicholas D. Sidiropoulos, Fulvio Gini and S. Theodoridis, eds.), vol. 2 of *Academic Press Library in Signal Process.*, pp. 595 – 665, Amsterdam, Netherlands: Elsevier, 2014.

[15] R. Klemm, *Principles of Space-Time Adaptive Processing*. Electromagnetics and Radar Series, London, UK: Institution of Engineering and Technology, 2006.

[16] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, 1974.

[17] D. G. Leeper, "Isophoric arrays-massively thinned phased arrays with well-controlled sidelobes," *IEEE Trans. Antennas Propag.*, vol. 47, pp. 1825–1835, Dec. 1999.

[18] F. Gini, A. De Maio, and L. Patton, *Waveform Design and Diversity for Advanced Radar Systems*. London, UK: Institution of Engineering and Technology, 2012.

[19] M. Park and H. S. Lee, "Adaptive selection method for generalized likelihood ratio test," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, pp. 2615–2626, Oct. 2015.

[20] C. Chen and P. P. Vaidyanathan, "MIMO radar waveform optimization with prior information of the extended target and clutter," *IEEE Trans. Signal Process.*, vol. 57, pp. 3533–3544, Sep. 2009.

[21] P. Woodward, *Probability and Information Theory: With Applications to Radar*. Amesrdam, Netherlands: Elsevier Science & Technology, 1953.

[22] M. R. Bell, "Information theory and radar waveform design," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1578–1597, 1993.

[23] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, pp. 330–343, January 2007.

[24] H. Griffiths, "Introduction to waveform diversity and cognitive radar," in *Waveform Diversity and Cognitive Radar, and Target Tracking and Data Fusion* (Klemm, Richard, Griffiths, Hugh, Koch, and Wolfgang, eds.), London, UK: Institution of Engineering and Technology, 2018.

[25] W. Melvin, M. Wicks, P. Antonik, Y. Salama, Ping Li, and H. Schuman, "Knowledge-based space-time adaptive processing for airborne early warning radar," *IEEE Aerosp. El. Sys. Mag.*, vol. 13, pp. 37–42, April 1998.

[26] P. Antonik, H. Schuman, P. Li, W. Melvin, and M. Wicks, "Knowledge-based space-time adaptive processing," in *Proceedings of the IEEE National Radar Conference*, (New York, NY), pp. 372–377, May 1997.

[27] P. Antonik, H. K. Schuman, W. L. Melvin, and M. C. Wicks, "Implementation of knowledge-based control for space-time adaptive processing," in *Proceedings of Radar Systems*, (Edinburgh, UK), pp. 478–482, Oct 1997.

[28] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Process. Mag.*, vol. 23, pp. 30–40, Jan. 2006.

[29] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, "Space-time adaptive processing: a knowledge-based perspective for airborne radar," *IEEE Signal Process. Mag.*, vol. 23, pp. 51–65, Jan 2006.

[30] F. Gini and M. Rangaswamy, *Knowledge Based Radar Detection, Tracking and Classification (Adaptive and Learning Systems for Signal Processing, Communications and Control Series)*. New York, NY, USA: Wiley-Interscience, 2008.

[31] S. A. L. David K. Barton, ed., *Radar Technology Encyclopedia*. London, UK: Artech House, 1998.

[32] H. Poor, *An Introduction to Signal Detection and Estimation*. Springer Texts in Electrical Engineering, New York, NY: Springer, 2013.

[33] H. Van Trees, *"Detection, Estimation, and Linear Modulation Theory" in Detection, Estimation, and Modulation Theory*. New York, NY: John Wiley & Sons, 2004.

[34] V. G. Hansen and J. H. Sawyers, "Detectability loss due to greatest of selection in a cell-averaging CFAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, pp. 115–118, Jan. 1980.

[35] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, pp. 608–621, July 1983.

[36] J. T. Rickard and G. M. Dillard, "Adaptive detection algorithms for multiple-target situations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-13, pp. 338–343, July 1977.

[37] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, 1986.

[38] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference: Part I," *Biometrika*, vol. 20A, no. 1/2, pp. 175–240, 1928.

[39] W.-S. Chen and I. S. Reed, "A new CFAR detection test for radar," *Digital Signal Process.*, vol. 1, no. 4, pp. 198–214, 1991.

[40] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[41] R. Serfling, "Depth functions in nonparametric multivariate inference," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 72, Nov. 2006.

[42] Zuo, Yijun and Serfling, Robert, "General notions of statistical depth function," *Ann. Statist.*, vol. 28, pp. 461–482, Apr. 2000.

[43] K. Mosler, *Depth Statistics*, pp. 17–34. Berlin, Heidelberg: Springer, 2013.

[44] X. Dang and R. Serfling, "Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties," *Journal of Statistical Planning and Inference*, vol. 140, no. 1, pp. 198 – 213, 2010.

[45] K. Mosler and R. Hoberg, "Data analysis and classification with the zonoid depth," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 72, p. 49, 2006.

[46] S. Dutta, A. K. Ghosh, and P. Chaudhuri, "Some intriguing properties of tukey's half-space depth," *Bernoulli*, vol. 17, pp. 1420–1434, Nov. 2011.

[47] P. Chaudhuri, "On a geometric notion of quantiles for multivariate data," *J. Am. Stat. Assoc.*, vol. 91, no. 434, pp. 862–872, 1996.

[48] J. B. S. Haldane, "Note on the median of a multivariate distribution," *Biometrika*, vol. 35, pp. 414–417, 12 1948.

[49] T. Kärkkäinen and S. Äyrämö, "On computation of spatial median for robust data mining," in *6th Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems*, (Munich, Germany), Sep. 2005.

[50] Yixin Chen, Xin Dang, Hanxiang Peng, H. L. Bart, and H. L. Bart, "Outlier detection with the Kernelized spatial depth function," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 288–305, Feb. 2009.

[51] S. Seo, *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*. PhD thesis, University of Pittsburgh, Aug. 2006.

[52] F. Mosteller and J. W. Tukey, *Data analysis and regression: a second course in statistics.* Addison-Wesley Series in Behavioral Science: Quantitative Methods, London, UK: Pearson, 1977.

[53] Huber, Peter J., "Projection pursuit," *Ann. Statist.*, vol. 13, pp. 435–475, June 1985.

[54] Y. Zuo, "Projection-based depth functions and associated medians," *Ann. Statist.*, vol. 31, pp. 1460–1490, 10 2003.

[55] S. Chen, S. Ma, A. M.-C. So, and T. Zhang, "Proximal gradient method for manifold optimization," *arXiv preprint arXiv:1811.00980*, 2018.

[56] N. G. Polson, J. G. Scott, B. T. Willard, *et al.*, "Proximal algorithms in statistics and machine learning," *Statistical Science*, vol. 30, no. 4, pp. 559–581, 2015.

[57] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds.), pp. 185–212, New York, NY: Springer, 2011.

[58] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, pp. 127–239, Jan. 2014.

[59] M. I. Jordan, ed., *Learning in Graphical Models*. Cambridge, MA: MIT Press, 1999.

[60] Y. A. Kuznetsov, "Introduction to dynamical systems," in *Elements of Applied Bifurcation Theory*, pp. 1–37, New York, NY: Springer, 2004.

[61] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.

[62] L. Fahrmeir and G. Tutz, *State Space and Hidden Markov Models*, pp. 331–383. New York, NY: Springer, 2001.

[63] S. Haykin, *Cognitive Dynamic Systems: Perception-action Cycle, Radar and Radio*. Cambridge University Press, 2012.

[64] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, Feb. 2002.

[65] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.

[66] M. S. Grewal and A. P. Andrews, "Linear optimal filters and predictors," in *Kalman Filtering: Theory and Practice Using MATLAB*, New York, NY: John Wiley & Sons, Ltd, 2008.

[67] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE proceedings F (radar and signal processing)*, vol. 140, pp. 107–113, IET, 1993.

[68] J. M. Hammersley and D. C. Handscomb, "The general nature of Monte Carlo methods," in *Monte Carlo Methods*, pp. 1–9, Dordrecht, Netherlands: Springer, 2013.

[69] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, pp. 3–14, New York, NY: Springer, 2001.

[70] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *The Oxford Handbook of Nonlinear Filtering*, vol. 12, pp. 656–704, New York, NY: Oxford University Press, 2009.

[71] J. M. Hammersley and D. C. Handscomb, "General principles of the Monte Carlo method," in *Monte Carlo Methods*, pp. 50–45, Dordrecht, Netherlands: Springer, 2013.

[72] I. Rekleitis, "A particle filter tutorial for mobile robot localization," Tech. Rep. TR-CIM-04-02, Centre for Intelligent Machines, McGill University, Montreal, Québec, Canada, Jan. 2004.

[73] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.

[74] R. Douc, E. Moulines, P. Priouret, and P. Soulier, "Markov chains: Basic definitions," in *Markov Chains*, pp. 3–25, Cham, Switzerland: Springer, 2018.

[75] Xiaolin Li, M. Parizeau, and R. Plamondon, "Training hidden Markov models with multiple observations-a combinatorial method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 371–377, April 2000.

[76] Ahmed A. Abouelfadl, F. M. Ahmed, and M. S. Soliman, "A novel noise-free jamming technique against LFM-PC search radar," in *Int. Japan-Egypt Conf. on Electronics, Commun. and Computers (JEC-ECC)*, (Cairo, Egypt), pp. 127–130, May 2016.

[77] Ahmed A. Abouelfadl, A. M. Samir, F. M. Ahmed, and A. H. Asseesy, "Performance analysis of LFM pulse compression radar under effect of convolution noise jamming," in *National Radio Science Conf. (NRSC)*, (Aswan, Egypt), pp. 282–289, Feb. 2016.

[78] W. L. Melvin and M. C. Wicks, "Improving practical space-time adaptive radar," in *Proc. of the IEEE National Radar Conf.*, (New York, USA), pp. 48–53, May 1997.

[79] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, pp. 2101–2111, June 2005.

[80] L. Jiang and T. Wang, "Robust non-homogeneity detector based on reweighted adaptive power residue," *IET Radar, Sonar & Navigation*, vol. 10, pp. 1367–1374, April 2016.

[81] S. Han, L. Pallotta, V. Carotenuto, A. De Maio, and X. Huang, "An approximate regularized ML approach to censor outliers in Gaussian radar data," *IEEE˙O˙ACC*, vol. 7, pp. 66263–66274, 2019.

[82] P. Setlur and M. Rangaswamy, "A family of random and random type projections for radar STAP," in *IEEE Radar Conf.*, (Oklahoma, USA), pp. 0856–0861, April 2018.

[83] Z. Li, H. Liu, Y. Zhang, and Y. Guo, "Robust nonhomogeneous training samples detection method for space-time adaptive processing radar using sparse-recovery with knowledge-aided," *J. Appl. Remote Sens.*, vol. 11(4), pp. 045013–1–045013–13, Oct. 2017.

[84] W. Wang, L. Zou, X. Wang, and Y. Yang, "Deterministic-aided single dataset stap method based on sparse recovery in heterogeneous clutter environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, p. 24, Apr. 2018.

[85] Z. Bowen, D. Zhiguang, and X. Lan, "Secondary non-homogeneity detector for STAP," in *IEEE 3rd International Conf. on Signal and Image Processing (ICSIP)*, (Shenzhen, China), pp. 382–384, July 2018.

[86] J. H. Michels, M. Rangaswamy, and B. Himed, "Performance of parametric and covariance based STAP tests in compound-Gaussian clutter," *Digital Signal Process.*, vol. 12, no. 2, pp. 307 – 328, 2002.

[87] G. N. Schoenig, M. L. Picciolo, and L. Mili, "Improved detection of strong nonhomogeneities for STAP via projection statistics," in *IEEE Int. Radar Conf.*, (Virginia, USA), pp. 720–725, May 2005.

[88] M. Richards, W. Holm, W. Melvin, J. Scheer, and J. Scheer, *Principles of Modern Radar: Basic Principles*. North Carolina: Institution of Engineering and Technology, 2012.

[89] T. Adali and P. J. Schreier, "Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation," *IEEE Signal Process. Mag.*, vol. 31, pp. 112–128, Sep. 2014.

[90] E. Conte, A. D. Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, pp. 1908–1915, Aug. 2002.

[91] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proceedings F - Communications, Radar and Signal Processing*, vol. 134, pp. 191–197, April 1987.

[92] M. Rangaswamy and D. Weiner and A. Ozturk, "Computer generation of correlated non-Gaussian radar clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, pp. 106–116, Jan. 1995.

[93] M. Rangaswamy, *Spherically Invariant Random Processes for Radar Clutter Modeling, Simulation and Distribution Identification*. PhD thesis, Syracuse University, 1992.

[94] I. Soloveychik and A. Wiesel, "Tyler's covariance matrix estimator in elliptical models with convex structure," *IEEE Trans. Signal Process.*, vol. 62, pp. 5251–5259, 2014.

[95] Pulsone, N.B. and Raghavan, R.S., "Analysis of an adaptive CFAR detector in non-Gaussian interference," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 903–916, 1999.

[96] Y. He, T. Jian, F. Su, C. Qu, and D. Ping, "CFAR assessment of covariance matrix estimators for non-Gaussian clutter," *Science China Information Sciences*, vol. 53, pp. 2343–2351, Nov. 2010.

[97] J. W. McKean and T. P. Hettmansperger, "Rank-based analysis of linear models and beyond: A review," in *Robust Rank-Based and Nonparametric Methods* (R. Y. Liu and J. W. McKean, eds.), pp. 1–24, Michigan: Springer International Publishing, 2016.

[98] S. Velasco-Forero and J. Angulo, "Random projection depth for multivariate mathematical morphology," *IEEE J. Sel. Topics Signal Process.*, vol. 6, pp. 753–763, Nov. 2012.

[99] K. Hayes, "Finite-sample bias-correction factors for the median absolute deviation," *Communications in Statistics - Simulation and Computation*, vol. 43, no. 10, pp. 2205–2212, 2014.

[100] D. J. Olive, "Sufficient statisitics," in *Statistical Theory and Inference*, pp. 215–256, Michigan: Springer International Publishing, 2014.

[101] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. Am. Stat. Assoc.*, vol. 88, no. 424, pp. 1273–1283, 1993.

[102] J. Cuesta-Albertos and A. Nieto-Reyes, "The random Tukey depth," *Computational Statistics & Data Analysis*, vol. 52, no. 11, pp. 4979 – 4988, 2008.

[103] M. E. Muller, "A note on a method for generating points uniformly on N-dimensional spheres," *Commun. ACM*, vol. 2, pp. 19–20, Apr. 1959.

[104] J. Ward, *Space-time Adaptive Processing for Airborne Radar.* Technical report, Massachusetts Institute of Technology, Lincoln Laboratory, 1994.

[105] M. Younis, C. Laux, A. Loinger, G. Adamiuk, M. Ludwig, D. Geudtner, and G. Krieger, "Concept and performance of internal instrument calibration for multi-channel SAR," in *12th European Conference on Synthetic Aperture Radar (EUSAR)*, (Aachen, Germany), pp. 1–5, June 2018.

[106] E. Attia and B. Steinberg, "Self-cohering large antenna arrays using the spatial correlation properties of radar clutter," *IEEE Trans. Antennas Propag.*, vol. 37, no. 1, pp. 30–38, 1989.

[107] A. Aubry, A. D. Maio, and L. Pallotta, "A geometric approach to covariance matrix estimation and its applications to radar problems," *IEEE Trans. Signal Process.*, vol. 66, pp. 907–922, Feb. 2018.

[108] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847 – 1859, 2002.

[109] Y. Zuo and R. Serfling, "Structural properties and convergence results for contours of sample statistical depth functions," *Ann. Statist.*, vol. 28, pp. 483–499, 04 2000.

[110] D. Azriel and A. Schwartzman, "The empirical distribution of a large number of correlated Normal variables," *J. Am. Stat. Assoc.*, vol. 110, no. 511, pp. 1217–1228, 2015.

[111] J. D. Evans, *Straightforward statistics for the behavioral sciences.* CA, US: Thomson Brooks/Cole Publishing Co, 1996.

[112] W. Xu, Y. Hou, Y. S. Hung, and Y. Zou, "A comparative analysis of Spearman's Rho and Kendall's Tau in Normal and contaminated Normal models," *Signal Process.*, vol. 93, no. 1, pp. 261–276, 2013.

[113] P.-L. Loh and X. L. Tan, "High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination," *Electron. J. Statist.*, vol. 12, no. 1, pp. 1429–1467, 2018.

[114] T. Zhao and H. Liu, "Calibrated precision matrix estimation for high-dimensional elliptical distributions," *IEEE Trans. on Information Theory / Professional Technical Group on Information Theory*, vol. 60, no. 12, pp. 7874–7887, 2014.

[115] T. Zhao, K. Roeder, and H. Liu, "Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation," *J. of Computational and Graphical Statistics*, vol. 23, no. 4, pp. 895–922, 2014.

[116] A. Melebari, A. K. Mishra, and M. Y. A. Gaffar, "Statistical analysis of measured high resolution land clutter at X-band and clutter simulation," in *European Radar Conf. (EuRAD)*, (Paris, France), pp. 105–108, Sept. 2015.

[117] I. Antipov, *Analysis of Sea Clutter Data*. DSTO Electronic and Surveillance Research Laboratory Salisbury, Australia, 1998.

[118] C. Croux and C. Dehon, "Influence functions of the Spearman and Kendall correlation measures," *Statistical Methods & Applications*, vol. 19, pp. 497–515, Nov. 2010.

[119] S. D. Himonas and M. Barkat, "Adaptive CFAR detection in partially correlated clutter," *IEE Proceedings F - Radar and Signal Processing*, vol. 137, pp. 387–394, Oct 1990.

[120] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA: The MIT Press, 3rd ed., 2009.

[121] J. D. Blanchard, E. Opavsky, and E. Uysaler, "Selecting multiple order statistics with a graphics processing unit," *ACM Trans. Parallel Comput.*, vol. 3, pp. 1–23, July 2016.

[122] Y. Liu, T. Pan, O. Green, and S. Aluru, "Parallelized Kendall's Tau coefficient computation via SIMD vectorized sorting on many-integrated-core processors," *ArXiv e-prints*, p. arXiv:1704.03767, Apr. 2017.

[123] D. R. Fuhrmann and G. San Antonio, "Transmit beamforming for MIMO radar systems using partial signal correlation," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 1, (California, CA), pp. 295–299 Vol.1, Nov. 2004.

[124] H. Deng, Z. Geng, and B. Himed, "MIMO radar waveform design for transmit beamforming and orthogonality," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1421–1433, 2016.

[125] C. Gao, K. C. Teh, A. Liu, and H. Sun, "Piecewise LFM waveform for MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 2, pp. 590–602, 2016.

[126] C. Gao, K. C. Teh, and A. Liu, "Piecewise nonlinear frequency modulation waveform for MIMO radar," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 2, pp. 379–390, 2017.

[127] H. Li, Y. Zhao, Z. Cheng, and D. Feng, "OFDM chirp waveform diversity design with correlation interference suppression for MIMO radar," *IEEE Geosci. Remote Sens. Lett.*, vol. PP, no. 99, pp. 1–5, 2017.

[128] H. Li, Y. Zhao, Z. Cheng, and D. Z. Feng, "Orthogonal frequency division multiplexing linear frequency modulation signal design with optimised pulse compression property of spatial synthesised signals," *IET Radar, Sonar& Navigation*, vol. 10, no. 7, pp. 1319–1326, 2016.

[129] K. Alhujaili, V. Monga, and M. Rangaswamy, "Transmit MIMO radar beampattern design via optimization on the complex circle manifold," *IEEE Trans. Signal Process.*, vol. 67, pp. 3561–3575, July 2019.

[130] E. H. Kim and K. H. Kim, "Random phase code for automotive MIMO radars using combined frequency shift keying-linear FMCW waveform," *IET Radar, Sonar & Navigation*, vol. 12, pp. 1090–1095, April 2018.

[131] V. F. Mecca, D. Ramakrishnan, and J. L. Krolik, "MIMO radar space-time adaptive processing for multipath clutter mitigation," in *Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, (Massachusetts, MA), pp. 249–253, July 2006.

[132] K. W. Forsythe and D. W. Bliss, "MIMO radar waveform constraints for GMTI," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 21–32, Feb. 2010.

[133] J. de Mingo, A. Valdovinos, A. Crespo, D. Navarro, and P. Garcia, "An RF electronically controlled impedance tuning network design and its application to an antenna input impedance automatic matching system," *IEEE Trans. Microw. Theory Techn.*, vol. 52, pp. 489–497, Feb. 2004.

[134] D. M. Pozar, "A relation between the active input impedance and the active element pattern of a phased array," *IEEE Trans. Antennas Propag.*, vol. 51, pp. 2486–2489, Sep. 2003.

[135] L. Savy and M. Lesturgie, "Coupling effects in MIMO phased array," in *IEEE Radar Conference (RadarConf)*, (Philadelphia, PA), pp. 1–6, May 2016.

[136] N. Colon-Diaz, J. G. Metcalf, D. Janning, and B. Himed, "Mutual coupling analysis for colocated MIMO radar applications using CEM modeling," in *IEEE Radar Conference (RadarConf)*, (Washington, WA), pp. 0441–0446, May 2017.

[137] V. F. Mecca, J. L. Krolik, F. C. Robey, and D. Ramakrishnan, "Slow-time MIMO spacetime adaptive processing," in *MIMO Radar Signal Processing*, ch. 7, pp. 283–318, John Wiley & Sons, Ltd, 2008.

[138] J. Bergin and J. R. Guerci, "Introduction to MIMO radar," in *MIMO Radar: Theory and Applications*, London, UK: Artech House, 2018.

[139] B. T. Arnold and M. A. Jensen, "The effect of antenna mutual coupling on MIMO radar system performance," *IEEE Trans. Antennas Propag.*, vol. 67, pp. 1410–1416, March 2019.

[140] H. Sun, F. Brigui, and M. Lesturgie, "Analysis and comparison of MIMO radar waveforms," in *International Radar Conference*, (Lille, France), pp. 1–6, Oct. 2014.

[141] S. Haykin, Y. Xue, and P. Setoodeh, "Cognitive radar: Step toward bridging the gap between neuroscience and engineering," *Proc. IEEE*, vol. 100, pp. 3102–3130, Nov. 2012.

[142] S. Wang, Q. He, and Z. He, "LFM-based waveform design for cognitive MIMO radar with constrained bandwidth," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, pp. 1–9, 2014.

[143] J. Li and P. Stoica, "MIMO radar: Concepts, performance enhancements, and applications," in *MIMO Radar Signal Processing*, Wiley-IEEE Press, 2009.

[144] Chun-Yang Chen and P. P. Vaidyanathan, "Minimum redundancy MIMO radars," in *IEEE International Symposium on Circuits and Systems*, (Washnigton, WA), pp. 45–48, May 2008.

[145] K. Han, M. Jiang, H. Wang, Y. Chen, and G. Liu, "Sparse linear antenna array design in compact MIMO radar using genetic algorithm," in *4th International Conference on Information Science and Control Engineering (ICISCE)*, (Changsha, China), pp. 1717–1720, July 2017.

[146] K. Rezer, W. Gropengießer, and A. F. Jacob, "Particle swarm optimization of minimum-redundancy MIMO arrays," in *German Microwave Conference*, (Darmstadt, Germany), pp. 1–4, March 2011.

[147] J. Dong, Q. Li, and W. Guo, "A combinatorial method for antenna array design in minimum redundancy MIMO radars," *IEEE Antennas Wireless Propag. Lett.*, vol. 8, pp. 1150–1153, 2009.

[148] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, "Difference bases and sparse sensor arrays," *IEEE Trans. Inf. Theory*, vol. 39, pp. 716–721, March 1993.

[149] E. J. Weldon, "Difference-set cyclic codes," *The Bell System Technical Journal*, vol. 45, pp. 1045–1055, Sep. 1966.

[150] J. Leech, "On the representation of $1, 2, \cdots, n$ by differences," *Journal of the London Mathematical Society*, vol. s1-31, no. 2, pp. 160–169, 1956.

[151] S. Orfanidis, "S-parameter," in *Electromagnetic Waves and Antenna*, Rutgers University, 2002.

[152] N. Colon-Diaz, D. Janning, T. Corigliano, L. Wang, and J. Aberle, "Measurement of active reflection coefficient for co-located MIMO radar using dual directional couplers," in *Antenna Measurement Techniques Association Symposium (AMTA) Proceedings*, (Virginia, VA), pp. 1–4, Nov. 2018.

[153] C. Zhang, Q. Lai, and C. Gao, "Measurement of active S-parameters on array antenna using directional couplers," in *IEEE Asia Pacific Microwave Conference (APMC)*, (Kuala Lumpur, Malaysia), pp. 1167–1170, Nov. 2017.

[154] J. Whitaker, "Radar and radionavigation," in *The Electronics Handbook*, Electrical Engineering Handbook, Florida, FL: CRC Press, 2018.

[155] S. H. Chae, S. Oh, and S. Park, "Analysis of mutual coupling, correlations, and TARC in WiBro MIMO array antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 122–125, 2007.

[156] A. K. Gautam, S. Yadav, and K. Rambabu, "Design of ultra-compact UWB antenna with band-notched characteristics for MIMO applications," *IET Microwaves, Antennas Propagation*, vol. 12, no. 12, pp. 1895–1900, 2018.

[157] G. Das, A. Sharma, and R. K. Gangwar, "Dielectric resonator-based two-element MIMO antenna system with dual band characteristics," *IET Microwaves, Antennas Propagation*, vol. 12, no. 5, pp. 734–741, 2018.

[158] A. Scuderi, L. La Paglia, A. Scuderi, F. Carrara, and G. Palmisano, "A VSWR-protected silicon bipolar RF power amplifier with soft-slope power control," *IEEE J. Solid-State Circuits*, vol. 40, pp. 611–621, March 2005.

[159] A. Engineering, "The importance of load tolerance in specifying RF power amplifiers for immunity / susceptibility testing applications," tech. rep., AR RF/Microwave Instrumentation, Philadelphia, PA, 2015.

[160] S. Preis, J. Ferretti, N. Wolff, A. Wiens, R. Jakoby, W. Heinrich, and O. Bengtsson, "Response time of VSWR protection for GaN HEMT based power amplifiers," in *46th European Microwave Conference (EuMC)*, (London, UK), pp. 401–404, Oct. 2016.

[161] Y. Li and Z. Tan, "Simulation and design of RF front-end electromagnetic protection module based on VHF communication," in *International Conference on Electronics Technology (ICET)*, (Chengdu, China), pp. 142–146, May 2018.

[162] D. Dr. R. K. Agarwal, *Solid State Devices and Electronics*. Meerut, Uttar Pradesh, India: Krishna Prakashan, second ed., 2007.

[163] S. Sinha, C. V. N. Rao, and D. Pujara, "Balanced power amplifier protection against load mismatch," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, pp. 165–167, Feb. 2018.

[164] A. Scuderi, L. La Paglia, A. Scuderi, F. Carrara, and G. Palmisano, "A VSWR-protected silicon bipolar RF power amplifier with soft-slope power control," *IEEE J. Solid-State Circuits*, vol. 40, pp. 611–621, March 2005.

[165] N. O. Sokal, "The origin and development of the high efficiency class-E RF power amplifier: A 2011 address accepting the honorary doctorate," *IEEE Microw. Mag.*, vol. 19, pp. 22–25, July 2018.

[166] Jorge Nocedal, Stephen J. Wright, "Penalty and augmented Lagrangian methods," in *Numerical Optimization*, pp. 497–528, New York: Springer, 2006.

[167] J. Shen, H. Xu, and P. Li, "Online optimization for max-norm regularization," *Machine Learning*, vol. 106, pp. 419–457, Mar 2017.

[168] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *IEEE Trans. Signal Process.*, vol. 59, pp. 936–952, March 2011.

[169] S. Sra, "Fast projections onto $\ell_{1,q}$-norm balls for grouped feature selection," in *Machine Learning and Knowledge Discovery in Databases* (D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, eds.), (Berlin, Heidelberg), pp. 305–317, Springer Publishing, 2011.

[170] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Program.*, vol. 158, pp. 575–585, Jul 2016.

[171] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions," in *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, (New York, NY), pp. 272–279, ACM, 2008.

[172] K. Devriendt and P. Van Mieghem, "The simplex geometry of graphs," *J. Complex Networks*, vol. 7, pp. 469–490, Jan. 2019.

[173] M. Held, P. Wolfe, and H. P. Crowder, "Validation of subgradient optimization," *Math. Program.*, vol. 6, pp. 62–88, Dec. 1974.

[174] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[175] P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Math. Program.*, vol. 125, pp. 263–295, Oct 2010.

[176] B. Wen, X. Chen, and T. Pong, "Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems," *SIAM J. Optim.*, vol. 27, no. 1, pp. 124–145, 2017.

[177] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal-recovery problems," in *Convex Optimization in Signal Processing and Communications* (D. P. Palomar and Y. C. E. Eldar, eds.), p. 42–88, Cambridge, UK: Cambridge University Press, 2009.

[178] L. Petit, L. Dussopt, and J. . Laheurte, "MEMS-switched parasitic-antenna array for radiation pattern diversity," *IEEE Trans. Antennas Propag.*, vol. 54, pp. 2624–2631, Sep. 2006.

[179] X. Zhang, K. Wang, and X. Liu, "Joint optimisation of transmit waveform and receive filter for cognitive radar," *IET Radar, Sonar & Navigation*, vol. 12, pp. 11–20, Jan. 2018.

[180] B. Tang, J. Tang, and Y. Peng, "MIMO radar waveform design in colored noise based on information theory," *IEEE Trans. Signal Process.*, vol. 58, pp. 4684–4697, Sep. 2010.

[181] G. A. Clark, "Non-Gaussian correlated process sampling for Bayesian cognitive target classification," *The J. of the Acoustical Society of America*, vol. 137, 2015.

[182] Ahmed A. Abouelfadl, I. Psaromiligkos, and B. Champagne, "A low-complexity non-parametric STAP detector," in *IEEE National Aerospace and Electronics Conf. (NAECON)*, (Ohio, USA), pp. 592–596, July 2018.

[183] Ahmed A. Abouelfadl, I. Psaromiligkos, and B. Champagne, "Covariance-free nonhomogeneity STAP detector in compound Gaussian clutter based on robust statistics," *IET Radar, Sonar & Navigation*, accepted, to appear in 2020.

[184] Ahmed A Abouelfadl, I. Psaromiligkos, and B. Champagne, "Extended target frequency response estimation using infinite HMM in cognitive radars," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, (Ottawa, Canada), Nov. 2019.

[185] A. Hassanien and S. A. Vorobyov, "Phased-MIMO radar: A tradeoff between phased-array and MIMO radars," *IEEE Trans. Signal Process.*, vol. 58, pp. 3137–3151, June 2010.

[186] D. Dagan, E. N. Doric, V. Jevremovic, and J. Malisic, "On measuring skewness and kurtosis," *Quality and Quantity*, vol. 43, pp. 481–493, May 2009.

[187] A. Grebennikov, "Control circuits," in *RF and Microwave Transmitter Design*, ch. 13, pp. 717–757, New York, NY: John Wiley & Sons, Ltd, 2011.

[188] I. C. Cosme, I. F. Fernandes, J. L. de Carvalho, and S. Xavier-de Souza, "Memory-usage advantageous block recursive matrix inverse," *Applied Mathematics and Computation*, vol. 328, pp. 125–136, 2018.

[189] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. New York, NY: Krieger Publishing Co., Inc., 1992.

[190] P. Chen and L. Wu, "Waveform design for multiple extended targets in temporally correlated cognitive radar system," *IET Radar, Sonar & Navigation*, vol. 10, pp. 398–410, Feb. 2016.

[191] M. Bell, C. I. o. T. D. o. Engineering, and A. Science, *Information Theory and Radar: Mutual Information and the Design and Analysis of Radar Waveforms and Systems*. PhD thesis, California Institute of Technology, 1988.

[192] D. C. Jenn, P. E. Pace, and R. A. Romero, "An antenna for a mast-mounted low probability of intercept continuous wave radar: Improving performance with digital architecture," *IEEE Antennas Propag. Mag.*, vol. 61, pp. 63–70, April 2019.

[193] C. R. Smith and P. M. Goggans, "Radar target identification," *IEEE Antennas Propag. Mag.*, vol. 35, pp. 27–38, April 1993.

[194] A. Mezache, F. Soltani, M. Sahed, and I. Chalabi, "Model for non-Rayleigh clutter amplitudes using compound inverse Gaussian distribution: an experimental analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, pp. 142–153, Jan. 2015.

[195] F. Z. Dai, H. W. Liu, P. H. Wang, and S. Z. Xia, "Adaptive waveform design for range-spread target tracking," *Electron. Lett.*, vol. 46, pp. 793–794, May 2010.

[196] G. Y. Delisle and Haiqing Wu, "Moving target imaging and trajectory computation using ISAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, pp. 887–899, July 1994.

[197] X. Zhang and C. Cui, "Range-spread target detecting for cognitive radar based on track-before-detect," *Int. J. Electron.*, vol. 101, no. 1, pp. 74–87, 2014.

[198] I. Kanter, "Exact detection probability for partially correlated Rayleigh targets," tech. rep., The Defense Technical Information Center (DTIC), 1985.

[199] T. S. Edrington, "The amplitude statistics of aircraft radar echoes," *IEEE Trans. Mil. Electron.*, vol. 9, pp. 10–16, Jan. 1965.

[200] I. Kanter, "Exact detection probability for partially correlated Rayleigh targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, pp. 184–196, 1986.

[201] P. Swerling, "Probability of detection for fluctuating targets," *IRE Transactions on Information Theory*, vol. 6, pp. 269–308, April 1960.

[202] W. Weinstock, *Target Cross Section Models for Radar Systems Analysis*. PhD thesis, University of Pennsylvania, 1982.

[203] A. De Maio, A. Farina, and G. Foglia, "Target fluctuation models and their application to radar performance prediction," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 151, pp. 261–269, Oct. 2004.

[204] K. J. Sangston and K. R. Gerlach, "Non-Gaussian noise models and coherent detection of radar targets," tech. rep., Naval Research Lab, Washington DC, 1992.

[205] D. W. Matolak, I. Sen, and W. Xiong, "Generation of multivariate Weibull random variates," *IET Commun.*, vol. 2, pp. 523–527, April 2008.

[206] J. G. Metcalf, K. J. Sangston, M. Rangaswamy, S. D. Blunt, and B. Himed, "A new method of generating multivariate Weibull distributed data," in *IEEE Radar Conference (Radar Conf)*, (Philadelphia, PA), pp. 1–6, May 2016.

[207] H. Bourlard and S. Bengio, "Hidden Markov models," in *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA: MIT Press, second ed., 2002.

[208] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels - a survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, pp. 57–80, Sep. 2008.

[209] J. Lopez-Salamanca, L. O. Seman, M. D. Berejuck, and E. A. Bezerra, "Finite-state markov chains channel model for cubesats communication uplink," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–1, accepted, to appear in 2020.

[210] M. Gales, S. Young, *et al.*, "The application of hidden Markov models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2008.

[211] S. Sangsuk-Iam and T. E. Bullock, "Analysis of discrete-time Kalman filtering under incorrect noise covariances," *IEEE Trans. Autom. Control*, vol. 35, no. 12, pp. 1304–1309, 1990.

[212] T. Inamura, H. Tanie, and Y. Nakamura, "Keyframe compression and decompression for time series data based on the continuous hidden Markov model," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, (Nevada, NV), pp. 1487–1492 vol.2, Oct 2003.

[213] Y.-Y. Liu, S. Li, F. Li, L. Song, and J. M. Rehg, "Efficient learning of continuous-time hidden Markov models for disease progression," in *Advances in Neural Information Processing Systems*, pp. 3600–3608, 2015.

[214] T. Bo, T. Xiaobin, and Y. Baoqun, "Continuous-time hidden Markov models in network simulation," in *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, (Wuhan, China), pp. 667–670, IEEE, 2008.

[215] S. Turso and T. Bertuch, "Zero-IF radar signal processing," in *European Radar Conf. (EuRAD)*, (Paris, France), pp. 513–516, Sep. 2015.

[216] P. Orbanz and Y. W. Teh, "Bayesian nonparametric models," in *Encyclopedia of Machine Learning*, pp. 81–89, New York, NY: Springer, 2011.

[217] L. Wasserman, "Introduction," in *All of Nonparametric Statistics*, pp. 1–11, New York, NY: Springer, 2006.

[218] J. K. Ghosh and R. V. Ramamoorthi, "Preliminaries and the finite dimensional case," in *Bayesian Nonparametrics*, pp. 9–55, New York, NY: Springer, 2003.

[219] J. Wolfowitz, "Additive partition functions and a class of statistical hypotheses," *Ann. Math. Statist.*, vol. 13, pp. 247–279, 09 1942.

[220] P. H. Kvam and B. Vidakovic, *Nonparametric statistics with applications to science and engineering*, vol. 653. John Wiley & Sons, 2007.

[221] J. Bradley, *Distribution-free statistical tests*. Prentice-Hall, 1968.

[222] P. Müller and F. A. Quintana, "Nonparametric Bayesian data analysis," *Statistical Science*, vol. 19, no. 1, pp. 95–110, 2004.

[223] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, vol. 1, pp. 209–230, 03 1973.

[224] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, pp. 639–650, 1994.

[225] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden Markov model," in *Proc. of the 14th Int. Conf. on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, (Vancouver, British Columbia, Canada), pp. 577–584, MIT Press, 2001.

[226] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *J. Am. Stat. Assoc.*, vol. 90, no. 430, pp. 577–588, 1995.

[227] J. V. Gael and Z. Ghahramani, "Nonparametric hidden Markov models," in *Bayesian Time Series Models* (D. Barber, A. T. Cemgil, and S. Chiappa, eds.), p. 317–340, Cambridge, UK: Cambridge University Press, 2011.

[228] J. Van Gael, Y. Saatci, Y. W. Teh, and Z. Ghahramani, "Beam sampling for the infinite hidden Markov model," in *Proc.of the International Conf. on Machine Learning*, (Helsinki, Finland), pp. 1088–1095, ACM, 2008.

[229] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1566–1581, 2006.

[230] Y. Yang and R. S. Blum, "Minimax robust MIMO radar waveform design," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 147–155, June 2007.

[231] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Mon. Weather Rev.*, vol. 136, no. 12, pp. 4629–4640, 2008.

[232] A. Farina and P. Lombardo, "Modelling of a mixture of K-distributed and Gaussian clutter for coherent radar detection," *Electron. Lett.*, vol. 30, pp. 520–521, March 1994.

[233] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Am. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[234] A. Zhang, S. Gultekin, and J. Paisley, "Stochastic variational inference for the HDP-HMM," in *Artificial Intelligence and Statistics*, pp. 800–808, 2016.

[235] C. Zhang, B. Shahbaba, and H. Zhao, "Variational Hamiltonian Monte Carlo via score matching," *Bayesian Analysis*, vol. 13, no. 2, pp. 485–506, 2018.

[236] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, UK: Cambridge university press, 2010.

[237] A. Edalat, "A derivative for complex Lipschitz maps with generalised Cauchy–Riemann equations," *Theoretical Computer Science*, vol. 564, pp. 89 – 106, 2015.