OPTIMAL AND SUBOPTIMAL
SIGNAL DETECTION–ON THE RELATIONSHIP BETWEEN
ESTIMATION AND DETECTION THEORY

by

ALI GHOBADZADEH

A thesis submitted to the
Department of Electrical and Computer Engineering
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
October 2015

Copyright © Ali Ghobadzadeh, 2015
Abstract

In this thesis, we consider a general binary and $M$-ary hypothesis testing problems with unknown parameters. In a hypothesis testing problem, we assume that the probability density function (pdf) of observation is given while some unknown parameters exit in the structure of the pdf and the set of unknowns under each hypothesis is given.

In the first part of this study, for a binary composite testing, we prove that the Minimum Variance and Unbiased Estimator of a Separating Function (SF) serves as the optimal decision statistic for the Uniformly Most Powerful (UMP) unbiased test. In many problems, the UMP test does not exist. For such cases, we introduce new suboptimal SF-Estimator Tests (SFETs) which are easy to derive for many problems.

In the second part, we study the relationship between Constant False Alarm Rate (CFAR) and invariant tests. We generally show that for a family of distributions, the unknown parameters are eliminated from the distribution of the maximal invariant statistic under the Minimal Invariant Group (MIG) while the maximum information of the observed signal is preserved. We prove that any invariant test with respect to the MIG is CFAR. Then, we introduce the UMP-CFAR test as the optimal CFAR bound among all CFAR tests.

In the third part, the asymptotical optimality of the CFAR tests is studied after
reduction using MIG. We show that the CFAR tests obtained after MIG reduction using the Wald test is SFET and the Generalized Likelihood Ratio Test and the Rao test are asymptotically optimal. To find an improved test, we maximize the asymptotic probability of detection of the SFET using the Maximum Likelihood Estimation (MLE). We propose a systematic method allowing to derive the Asymptotically Optimal SFET. Finally, we extend the concept of SFET to the $M$-ary hypothesis testing. Defining $M$ different SFs for an $M$-ary problem, we show that the optimal minimum error test is achieved using the MLE of SFs. Moreover, in the case that the optimal minimum error test does not exist, the error probability of the proposed SFET tends to zero when the number of independent observations tends to infinity.
Co-Authorship

List of publications as a result of the contributions of this thesis:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 5:

Section 7.2:
A. Ghobadzadeh and S. Gazor, “Hybrid estimation and detection for random sparse

Section 7.1:


Section 7.1:


Section 7.3:

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Gazor for the continuous support of my Ph.D study and related research, for his patience, motivation, and knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. I had the honor of working with Dr. Gazor over the past few years. Through his help and guidance, we addressed many challenging topics. I would like to thank him for not just being a supervisor but also for being a great friend and mentor.

I was privileged to have courses with Dr. Blostein, Dr. Taban, Dr. Tadaion and Dr. Moshtaghion. I would like to thank Dr. Taban, Dr. Tadaion, Dr. Moshtaghion, Dr. M. Gazor and M. Naderpour who helped me in reviewing the theorems and simulations. Moreover, I wish to express my gratitude to Dr. Adve, Dr. Ibnkahla, Dr. McAuley, Dr. Afsahi, Dr. Thomson and Dr. Jain for accepting to serve as the examiners for my defense and Comprehensive exams. Absolutely their valuable comments helped to enhance the quality of the thesis.

I would also like to thank my wife and my little girl Rosanna who are the true source of LOVE in my life. Certainly, without my wives support I would not be able to finish this thesis. Her support, encouragement, quiet patience and unwavering love
were undeniably the bedrock upon which the past fourteen years of my life have been built. Her tolerance of my occasional vulgar moods is a testament in itself of her unyielding devotion and love. I thank my parents, for their faith in me and allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on. I wish to thank my sisters, Parastoo and Pardis, my brother Wahid and my mother-in-law for supporting me all the time.

I want to thank all my friends, who made the past few years a blast, especially, Sohail, Mohammad, Hossein, Saeed, Toritseju, Alex and all my labmates. Finally, I would like to extend my gratitude to the Natural Sciences and Engineering Research Counsel (NSERC), Ontario Graduate Scholarship (OGS), and Queens Graduate Awards for funding my graduate studies during the past three years.
Contents

Abstract i

Co-Authorship iii

Acknowledgments v

Contents vii

List of Figures x

List of Abbreviations xiv

Chapter 1: Introduction 1
  1.1 Thesis Organization ........................................... 4

Chapter 2: Literature Review 7
  2.1 Constant False Alarm Rate tests ............................. 10
  2.2 Preliminary Definitions and Terminology .................. 13

Chapter 3: Separating Function Estimation Tests 17
  3.1 The Relationship Between MVUE Statistic and the UMP test . 18
    3.1.1 One-Sided Hypothesis Test: 1D Parameter Space .......... 18
    3.1.2 One-Dimensional Parameter Space: The General Form ..... 21
    3.1.3 Multi-Dimensional Parameter Space ..................... 22
  3.2 A Perspective from Estimation Theory ....................... 23
    3.2.1 Suboptimal Decision Statistics: One-Sided-1D Case .... 24
    3.2.2 Suboptimal SFETs ........................................ 26
  3.3 How to find an SF? ......................................... 26
  3.4 Examples .................................................. 32

Chapter 4: Invariance and Optimality of CFAR tests 43
  4.1 Reduction Using MIG ........................................ 45
4.1.1 Some Gaussian Families . . . . . . . . . . . . . . . . . . . . . 47
4.2 Relationship of Invariant and CFAR Tests . . . . . . . . . . . . . . 51
4.3 Optimal and Suboptimal CFAR Tests . . . . . . . . . . . . . . . . 56
4.3.1 SFET-CFAR . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
4.3.2 Average Likelihood Ratio of Maximal invariant CFAR . . . . 59
4.3.3 GLRT-CFAR . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
4.4 Signal Processing Examples . . . . . . . . . . . . . . . . . . . . . . 62

Chapter 5: Asymptotically Optimal CFAR Test 72
5.1 GLRT-CFAR, Wald-CFAR and Rao-CFAR . . . . . . . . . . . . . . . 74
5.2 Asymptotically Optimal SFET (AOSFET) . . . . . . . . . . . . . . . 77
5.2.1 A tight Lower Bound for Detection Probability . . . . . . . . . 77
5.2.2 Asymptotically Optimal SF . . . . . . . . . . . . . . . . . . . . 80
5.3 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

Chapter 6: Multi Class classification using SFET 96
6.1 Relationship between UMET and MLE . . . . . . . . . . . . . . . . 97
6.2 Non-Detectable Problems using SFET . . . . . . . . . . . . . . . . . 103
6.2.1 Minimizing the Probability of Error Using NP criterion . . . . 103
6.2.2 Relaxing the Hypothesis . . . . . . . . . . . . . . . . . . . . . . 105
6.2.3 Exponentially Embedded Family . . . . . . . . . . . . . . . . . 110
6.3 SFET for Random Unknowns . . . . . . . . . . . . . . . . . . . . . . 113

Chapter 7: Practical Examples 115
7.1 MIMO Radar Target Detection . . . . . . . . . . . . . . . . . . . . . 115
7.1.1 UMPI Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
7.1.2 SFET based on SIR . . . . . . . . . . . . . . . . . . . . . . . . 119
7.1.3 Optimal Signal Design . . . . . . . . . . . . . . . . . . . . . . . 120
7.2 Sparse Signal Detection and VAD . . . . . . . . . . . . . . . . . . . . 122
7.2.1 Separating Function . . . . . . . . . . . . . . . . . . . . . . . . 124
7.2.2 Parametric Estimations . . . . . . . . . . . . . . . . . . . . . . 127
7.2.3 Hybrid SFETs . . . . . . . . . . . . . . . . . . . . . . . . . . . . 130
7.2.4 CFAR Property . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
7.2.5 Extension to Matrix Observation . . . . . . . . . . . . . . . . . 137
7.2.6 Simulation Result . . . . . . . . . . . . . . . . . . . . . . . . . . 139
7.3 Signal Detection in Uniform Linear Antennas . . . . . . . . . . . . . 147
7.3.1 Signal Model and Invariant Groups . . . . . . . . . . . . . . . . 148
7.3.2 Asymptotically Optimal SF . . . . . . . . . . . . . . . . . . . . 151
7.3.3 Simulations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 155

Chapter 8: Conclusion and Future Work 158
## Bibliography

### Appendix A: Appendix to Chapter 3

A.1 Proof of Theorem 2 ........................................... 181
A.2 Proof of Theorem 3 ........................................... 183
A.3 Proof of Lemma 1 ............................................. 186
A.4 Proof of Theorem 4 ........................................... 188
A.5 Proof of Theorem 6 ........................................... 189

### Appendix B: Appendix to Chapter 4

B.1 Proof of Proposition 2 ........................................ 193
B.2 Proof of Theorem 9 ........................................... 194
B.3 Proof of Theorem 12 ......................................... 195
B.4 Derivation of (4.7) .......................................... 197

### Appendix C: Appendix to Chapter 5

C.1 Proof of Proposition 3 ........................................ 199
C.2 Proof of Proposition 4 ........................................ 199
C.3 Proof of Proposition 6 ........................................ 200
C.4 Derivation of (5.24) ......................................... 201

### Appendix D: Appendix to Section 7.2

D.1 Derivation of $\hat{\alpha}_{ML_{k+1}}$ and $\hat{\sigma}^2_{ML_{k+1}}$ when $\hat{\alpha}_{ML_k}$ is zero ......... 203
D.2 Proof of Proposition 7 ....................................... 204
D.3 Proof of (D.3c) .............................................. 206
List of Figures

1.1 The general structure of a detection problem. ............................... 2

3.1 The general block diagram for deriving the SFET. ........................... 27

3.2 Detection probability versus false alarm probability for SFET\textsubscript{1}, SFET\textsubscript{2} and Optimal Neyman-Pearson bound in Example 4 for $L = 1, 2, 5, 10$ .......................................................... 40

3.3 Detection probability of $\psi_m$ in Example 5 for $m = 1, 2, 3$ and 10, (a) versus ENR$=\frac{N|A|^2}{\sigma^2}$ for $P_{fa} = 0.2$, (b) versus the probability of false alarm for ENR$= 9$ dB. ................................................................. 40

3.4 The $(\epsilon, p)$-accuracy of $g(\hat{\theta}_m)$ as an estimate of $\frac{|A|^2}{\sigma^2}$, i.e., $\Pr_\theta(|g(\hat{\theta}_m) - \frac{|A|^2}{\sigma^2}| < \epsilon)$ versus $\epsilon$. ................................................................. 41

3.5 Relationship between the performance two estimators for $f$ (MLE and zero crossing estimator) and the corresponding SEFTs in Example 5 (a) Mean Square Error of these estimators, (b) Probability of detection versus SNR for these SEFTs. ................................................................. 41

3.6 Comparison of the performance of the SFET and the GLRT in Example 6, (a) ROC curve, (b) $\hat{\rho}_k$ versus the number of iterations .................................................. 42

3.7 Missed detection probability of v.s. probability of false alarm for SFET\textsubscript{1}, SFET\textsubscript{2}, and SFET\textsubscript{3}. and comparing their performance by Optimal Neyman-Pearson bound for $M = 3$, $L = 5, 10, 15$ in Example 7. .............................................. 42
4.1 The general block diagram for deriving the SFET-CFAR. 60
4.2 Performance comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson for Problem (4.4), $\rho = 1.5 + j1.5$ and $N = 5$. 66
4.3 Performance comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson versus the number of observations for Problem (4.4), $\rho = 1.5 + j1.5$ and $P_{fa} = 10^{-3}$. 68
4.4 Probability of detection comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson, versus signal to noise ratio (in dB) for Problem (4.4) for $P_{fa} = 10^{-3}$, $N = 5$, and phase of $\rho$ is $\pi/4$. 68
4.5 Probability of detection versus probability of false alarm for MP-CFAR, CFAR SFET\textsubscript{1}, CFAR SFET\textsubscript{2} and CFAR GLRT for Example 18 for $\rho = 2$, $N = 100$ and $M = 30$. 71
5.1 Probability of miss detection versus the false alarm probability for Example 22, for the AOSFET, the GLRT and the NP test when $\sigma^2$ is known and equals to 1, and $\mu = 0.5$, and $N = 100$. 94
5.2 Probability of missed detection versus the false alarm probability for Example 22, for the AOSFET, the GLRT and the NP test when $\sigma^2$ and $\mu$ are unknown. In this simulation, we consider $\sigma^2 = 1$, $\mu = 0.5$, and $N = 200$. 94
5.3 Probability of detection versus false alarm probability for AOSFET\textsubscript{1}, AOSFET\textsubscript{2}, the GLRT and MP-CFAR, for $\theta_0 = 1_5$ versus $\theta_1 = [1 1.5 1.5 2 2]^T$ and $N = 20$. 95
5.4 Probability of detection versus false alarm probability for AOSFET₁, AOSFET₂, the GLRT and the MP-CFAR, for $\mathbf{\theta}_0 = 1_5$ versus $\mathbf{\theta}_1 = [1 1 1 1 0.5]^T$ and $N = 20$. .............................. 95

6.1 Probability of error versus the threshold, for (6.10), when $N = 10$, $L = 800$ and SNR varies from $-40$dB to $-8$dB. ...................... 110

6.2 Probability of error versus SNR, for (6.10) and (6.20), when $N = 10$, $L = 800$. ................................................................. 113

7.1 $P_d$ v.s. $P_{fa}$ curves of SFET in comparison with MPI bound for orthogonal narrowbound signals and proposed set of signals. .............. 122

7.2 Hybrid estimation of unknown parameters and SFET. ............... 127

7.3 a) Performance comparison of the SFETs and MP-CFAR versus $P_{fa}$, for $\alpha = 0.1, \sigma^2 = 1, N = 100$. b) Probability of miss detection of the SFETs and MP-CFAR versus $P_{fa}$, for $\beta = 1, \alpha = 0.5, \sigma^2 = 1, N = 100$. 143

7.4 Detection performance of the SFETs and MP-CFAR versus SNR, for $P_{fa} = 0.01, N = 100$ and $\beta = 1$. .............................. 143

7.5 Sample mean of the estimation of $\alpha$ using (7.21) and (7.24) versus the number of iterations for $\alpha = 0.5$ and $\alpha = 1$. ...................... 144

7.6 The objective function $\hat{\text{Obj}}(s)$ and the $l_\beta$ norm of sparse signal estimator using (7.26) and (7.27) versus the number of iterations, a. for $\alpha = 0.2, \beta = 0.7$ and $N = 100$ and, b. for $\alpha = 1, \beta = 0.9$ and $N = 100$. 144

7.7 a. Performance evaluation of the MP-CFAR test and SFET_{MA} versus $\beta$ for $\alpha = 0.1$. b. Probability of detection versus the number of observations for $P_{fa} = 0.01, \alpha = 1$ and $\beta = 1.5$. ...................... 145
7.8 Probability of miss detection versus the probability of false alarm for SFET_{MA} for detecting a sparse vector generated with a 40 point FFT of a received target signal when the number of targets are 2, 5 and 10, number of cells is 20.

7.9 Hard and Soft detection process for $\beta = 1$, sampling frequency is 48000, (a) input clean signal, (b) noisy signal at SNR = $-5.5$dB, (c) noisy signal at SNR = 3.46dB, (d) decision statistic $y_i$ out put of (7.29) for SNR = $-5.5$dB (e) $y_i$ out put of (7.29) for SNR = 3.46dB, (f,g) soft decision (7.30), (h,i) enhanced speech signal $\hat{s}_{i,MA}$.

7.10 Probability of detection versus probability of false alarm for AOSFET_{MLE}, AOSFEET_{OP} and the GLRT. The simulation parameters are $\rho = 0.01$, $\phi = \frac{\pi}{4}$ and $\omega T_s = \frac{\pi}{3}$.

7.11 Probability of missed-detection versus the number of snapshots for AOSFET_{MLE}, AOSFEET_{OP} and the GLRT. The simulation parameters are $\rho = 0.04$, $\phi = \frac{\pi}{4}$, $\omega T_s = \frac{\pi}{3}$ and $P_{fa} = 0.01$.

7.12 MSE versus the number of snapshots for AOSFET_{MLE}, AOSFEET_{OP} and the GLRT. The simulation parameters are $\rho = 0.04$, $\phi = \frac{\pi}{4}$ and $\omega T_s = \frac{\pi}{3}$.
List of Abbreviations

1D        one-Dimensional
AWGN      Additive White Gaussian Noise
AOSFET    AOSF Estimation Test
AOSF      Asymptotically Optimal Separating Function
ALM       Average Likelihood Ratio of Maximal invariant
ALRT      Average Likelihood Ratio Test
BPSK      Binary phase-shift keying
CA-CAFR   Cell Averaging Constant False Alarm Rate
CUT       Cell Under Test
CLA       Co-Located Antennas
CFAR      Constant False Alarm Rate
CM-SFET   Continuous M-ary-Separating Function Estimator Tests
CRLB      Cramer-Rao Lower Bound
DCT       Discrete Cosine Transform
DoA       Direction-of-Arrival
DFT       Discrete Fourier Transformation
EM        Expectation Maximization
EEF       Exponentially Embedded Family
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIM</td>
<td>Fisher Information Matrix</td>
</tr>
<tr>
<td>GLRT</td>
<td>Generalized Likelihood Ratio Test</td>
</tr>
<tr>
<td>IFD</td>
<td>Invariant Family of Distributions</td>
</tr>
<tr>
<td>KKT</td>
<td>Karush-Kuhn-Tucker</td>
</tr>
<tr>
<td>LR</td>
<td>Likelihood Ratio</td>
</tr>
<tr>
<td>LMPI</td>
<td>Locally Most Powerful Invariant</td>
</tr>
<tr>
<td>DM-SF</td>
<td>M-ary Separating Function</td>
</tr>
<tr>
<td>MAPE</td>
<td>Maximum A Posteriori Estimation</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum A posteriori Probability</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>MIG</td>
<td>Minimal Invariant Group</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>MPI</td>
<td>Most Powerful Invariant</td>
</tr>
<tr>
<td>MVUE</td>
<td>Minimum Variance and Unbiased Estimator</td>
</tr>
<tr>
<td>MD</td>
<td>Multi-Dimensional</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multi-Input Multi-Output</td>
</tr>
<tr>
<td>NP</td>
<td>Neyman-Pearson</td>
</tr>
<tr>
<td>NDP</td>
<td>Non-detectable Problems</td>
</tr>
<tr>
<td>OP-MLE</td>
<td>Outlier Protector Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>pdf</td>
<td>probability density function</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>RV</td>
<td>Random Variables</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>SF</td>
<td>Separating Function</td>
</tr>
<tr>
<td>SFET</td>
<td>Separating Function Estimator Test</td>
</tr>
<tr>
<td>SIR</td>
<td>Signal to Interference Ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>UMPI</td>
<td>Uniformly Most Powerful Invariant</td>
</tr>
<tr>
<td>UMPU</td>
<td>Uniformly Most Powerful unbiased</td>
</tr>
<tr>
<td>UMET</td>
<td>Uniformly Minimum Error test</td>
</tr>
<tr>
<td>UMP</td>
<td>Uniformly Most Powerful</td>
</tr>
<tr>
<td>VAD</td>
<td>Voice Activity Detection</td>
</tr>
<tr>
<td>WSA</td>
<td>Widely Separated Antennas</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Detection theory or signal detection theory is a means to quantify the ability to distinguish between defined patterns or hypotheses. Each pattern or hypothesis is defined by the statistical features and information. A detector or test is a function to determine each observation belongs to which hypothesis. Figure 1.1 shows a general structure of a hypothesis testing problem and a detector.

Hypothesis testing problems are encountered in many fields such as communication, radar, sonar, speech, image, biomedical and astronomy signal processing [1–8]. The Neyman-Pearson (NP) lemma states that the detection probability is maximized by comparing the Likelihood Ratio (LR) with a threshold where the threshold satisfies a given false alarm rate [9]. A binary composite test is a hypothesis problem where there exist some unknown parameters in the involved probability density functions (pdf) of observations [1]. For the composite problems, the NP test is referred to as the Uniformly Most Powerful (UMP) test if the rejection and acceptance regions of the NP test do not depend on the unknown parameters [9]. Since the UMP test does not exist in many problems, many suboptimal tests are sought in the literature such as
Generalized Likelihood Ratio Test (GLRT), Rao and Wald tests [10–13]. The GLRT statistic is the ratio of the supremums of the two families of pdfs over their unknown parameters, i.e., the GLRT statistic is given by the ratio of the pdfs evaluated at the Maximum Likelihood Estimation (MLE) of the unknown parameters. Similarly, the Rao test and Wald test are given by replacing the MLE of some of unknown parameters into two approximations of the LR respectively [1]. In the Rao test the MLE of the nuisance parameters is applied under the null hypothesis while in the Wald test the MLE of unknown parameters is exploited under the alternative hypothesis. It is shown that the three tests are asymptotically equivalent under some conditions [1].

The GLRT, the Rao and the Wald tests are applied in many signal processing problems (for example see [14–17]). The signal detection problem in Gaussian interference is studied using the concept of invariance property of the hypothesis in [16] and [17], where the Constant False Alarm rate (CFAR) property is also embedded in the proposed tests. A maximal invariant is proposed in [15] for signal detection in Gaussian noise with unknown covariance matrix and it is shown that the Adaptive Matched
Filter Test and the GLRT depend on observations only trough the presented maximal invariant. The problem of adaptive detection in Gaussian interference is studied in [14] assuming unknown covariance matrix. A reduction using a bi-dimensional maximal invariant statistic is proposed in [14] such that the achieved GLRT, Rao and Wald tests are all CFAR. This reduction eliminates the unknown parameters under null hypothesis and makes the reduced problem one dimensional, one-sided hypothesis testing problem. The relationship between UMP Invariant (UMPI) test and the CFAR property is studied in [18], where the signal detection problem is considered in the interference with a general statistical characterization including spherically invariant random vector. Derivation of CFAR tests in radar application is important, where several creative approaches have been proposed [19–22]. For example, a high resolution radar target detection is considered in [22] where the transmitting signal is a linear frequency modulation. The proposed detector in [22] is derived in three steps, first the cross S-method is found from the S-method, then the cross S-method is decomposed by the signal synthesis method based on the singular value decomposition and finally the time-frequency decomposition feature is proposed by the ratio of the sum of biggest singular values to the median of the rest.

In general, most of works, as mentioned above, are based on the estimating unknown parameters and replacing them into the LR or an approximation of LR, e.g., the GLRT, the Rao and the Wald tests. The main idea behind these approaches is the NP lemma. Although the NP test is optimal when the all parameters are known, in general the optimality of replacing the estimation of unknowns into LR is not proven. The main object of this thesis are finding answers for the following questions.

- To find a suboptimal test, should all unknown parameters be estimated? What
is the optimal set of unknowns that should be estimated?

- Should the estimation of unknown parameters be replaced into LR or its approximation?

- Is there any transformation such that eliminates some of unknown parameters and preserves the maximum information needed for detection?

In this thesis, we provide answers for those questions. We assume that the unknown parameters are deterministic and the set of values under each hypothesis is given variation. We exploit a tight relationship between estimation and detection to propose new tests. Furthermore, the invariance property of hypothesis provides powerful tools to deal with the unknowns.

The rest of this chapter is organized as follows. Chapter 2 summarizes the previous works related to the relationship between estimation and detection. In Section 2.1, the CFAR property of a test is defined and the related works are briefly studied. In Section 2.2 the invariance property of hypothesis tests and family of distribution are explained. The notations and all necessary definitions are collected in this section. Finally in Section 1.1, we scheme, the general structure of the thesis and the list of main achievements.

1.1 Thesis Organization

In Chapter 3, we derive some relationships between the detection and estimation theories for a binary composite hypothesis test. We show that for a 1D space problem and one-sided hypothesis problems, the UMP test is given by comparing the Minimum Variance and Unbiased Estimator (MVUE) with a threshold. Then, we extend out
1.1. THESIS ORGANIZATION

results into more general cases. In general cases, we show that the UMP test is given by comparing the MVUE of a Separating Function (SF). In the case that the optimal test or optimal estimator does not exist, a suboptimal test based in estimating the SF is proposed. In this case, it is proven that, if one detector between two detectors has a better probability of detection, then we can estimate the SF more $\epsilon$-accurately. These results motivate us to introduce new suboptimal Separating Function Estimator Tests (SFETs) which are easy to derive for many problems.

In Chapter 4, we study the relationship between CFAR and invariant tests using the concept of Minimal Invariant Group (MIG). We show that in a hypothesis testing problem, the unknown parameters under null hypothesis are eliminated from the distribution of the maximal invariant statistic of MIG while the maximum information of the observed signal is preserved. We prove that any invariant test with respect to MIG is CFAR and conversely, for any CFAR test an invariant statistic exists with respect to an MIG under some mild conditions. Moreover we propose a method to improve the performance of a CFAR test using the related MIG of the CFAR test. The UMP-CFAR test as the optimal CFAR bound among all CFAR tests is proposed in this chapter. In the case that the UMP-CFAR test depends on the unknown parameters, three suboptimal novel CFAR tests is proposed.

Chapter 4, we show that any test after reduction using MIG is CFAR. In Chapter 5, we study the performance of the GLRT, the Rao, the Wald tests and SFET after reduction. We show that the CFAR tests obtained after MIG reduction using the Wald test is SFET and GLRT and the Rao test are asymptotically SFET using MLE under some mild conditions. Thus, they are asymptotical optimal. Moreover, an improved SFET is proposed to maximize the probability of detection when the
number of observations tends to infinity. The proposed SFET called Asymptotically Optimal SFET (AOSFET) which is an extension of the Wald test after reduction.

In Chapter 6, we extend the concept of SFET to the $M$-ary hypothesis problems. The extension of SF is defined by an $M$ different SFs. We show that the optimal minimum error test is achieved using the MLE of SFs. Moreover, in the case that the optimal minimum error test does not exist, the proposed suboptimal test is asymptotically optimal. In some problems that the Lebesgue measure of set of unknowns is zero, the proposed SFET does not have a reliable performance for finite observations. Applying the concept of mixture modelling and exponentially embedded distributions, the performance of the proposed $M$-ary SFET is improved.

Finally, we provide some practical examples to evaluate the proposed tests. In Chapter 7, three practical problems are considered and the proposed tests are applied on the problems. A radar target detection, a voice activity detection and finally a narrowband signal detection using a linear array are studied in Section 7.1, 7.2 and 7.3.
Chapter 2

Literature Review

Consider two companion problems:

1) A hypothesis test described by

\[
\begin{align*}
\mathcal{H}_0 : \theta &\in \Theta_0, \\
\mathcal{H}_1 : \theta &\in \Theta_1,
\end{align*}
\]

where \(\Theta\) is the unknown parameter vector of the probability density function (pdf) of observation vector \(x \in \mathbb{R}^N\). The pdf of \(x\) is denoted by \(f(x; \theta)\) with the unknown deterministic parameter \(\theta \in \Theta \triangleq \Theta_0 \cup \Theta_1 \subseteq \mathbb{R}^M\).

2) An estimation problem described by

\[
x \sim f(x; \theta), \quad \theta \in \Theta \triangleq \Theta_0 \cup \Theta_1.
\]

These related problems are important in statistical signal processing and are comprehensively studied [1, 9, 23–25], separately. However, they are rarely studied together in the literature [26–30]. Our main objective is to establish a relationship between
the results related to these problems in order to gain a new perspective. For estimation of unknown parameters, there are two approaches in literature. In the classical approach, the unknown parameters are considered deterministic, whereas in the Bayesian approach, the unknown parameters are considered as Random Variables (RVs). As examples, the MVUE and the MLE are the most important ones obtained in the classical approach, whereas the Maximum A posteriori Probability (MAP) is obtained from the Bayesian approach [23].

In hypothesis testing, the optimum test based on the NP criterion is obtained by employing the NP lemma [10]. However, the NP test may not be used in composite hypothesis tests, because of unknown parameters and then the UMP test does not exist for many problems [1, 9]. There are two alternative approaches similar to the estimation theory. In the classical approach, the unknown parameters are considered as deterministic and are estimated via a classical estimation method, e.g., in GLRT. In contrast, in the Bayesian approach, the unknown parameters are assumed as RVs with a given distribution and the NP test is obtained after averaging over the distribution of the unknowns, e.g., in Average Likelihood Ratio Test (ALRT). In this thesis, we focus on the classical approach.

There are some works on the relationship between the Bayesian estimation and detection. In [26], it is shown that for a simple hypothesis (in which there is no unknown parameters), the NP detection is nothing other than the comparing of the Minimum Mean Square Error (MMSE) estimation of a discrete binary RV with a threshold. This RV takes the values 0 and 1 with prior probabilities Pr(\(H_1\)) and Pr(\(H_0\)), respectively. Kailath showed that for the problem of detecting a random signal in Additive White
Gaussian Noise (AWGN), the ALRT statistic is the estimator-correlator [27]. The authors in [28] showed that a detector based on the Bayesian approach with a quadratic cost function is equivalent to maximizing a weighted sum of least square estimators of signals in each hypothesis. In [29], it is shown that the LR is a linear combination of MVUE of unknown parameters. In some other works, without implicit study of such relationships, they have benefited from such a relationship to improve the detection performance, e.g., [31–35]. For example in [31], authors proposed an estimator for cumulants of non-Gaussian processes in the presence of unknown deterministic and Gaussian signals and employed this statistic to study the detection in colored non-Gaussian noise. In [33] a detector and an estimator are derived based on the least square solutions to detect and estimate unknown parameters for a second-order model. In [34] two lower bounds are derived for deterministic parameters conditioned by a binary hypothesis testing problem in a joint detection-estimation problem, while in [35], the authors proposed a derivation of the Cramér-Rao Lower Bound (CRLB) for the deterministic signal model based on an energy detector. A well known application of estimation in detection theory is within the GLRT which is frequently used particularly when the UMP test does not exist. To derive the GLRT, the unknown vectors under each hypothesis are replaced by their MLEs. The GLRT is widely used in signal processing (e.g., see [36–41]) and some of its statistical properties are studied in the literature, e.g., see [42–45]. In addition, there are many other suboptimal detectors proposed employing estimates of the unknown parameters, where the relationship between their performance and the employed estimators is not clear yet.

In this thesis we highlight some of these relationships. These relationships, despite their intrinsic importance, may lead to suboptimal detectors with acceptable
performance. We show that for a one-sided hypothesis testing problem the UMP test may be performed by comparing the MVUE of the unknown parameter with a threshold. Furthermore for a more general case (where $\Theta_0$ and $\Theta_1$ are the union of some subintervals), we show that the UMP test may be performed by comparing the MVUE of an SF with a threshold. This SF $g(\theta)$ is a continuous differentiable function, is positive for $\Theta_1$ and is negative (or zero) for $\Theta_0$. In many practical problems the MVUE does not exist. For such cases, we also use a similar suboptimal detection approach and show that an improved estimator for the SF results in a better decision statistics for our test problem, where the estimation criterion is the probability that estimator be in the $\epsilon$–neighborhood of the true parameter, i.e., $\Pr(|\hat{g}(\theta) - g(\theta)| < \epsilon)$.

2.1 Constant False Alarm Rate tests

The CFAR tests are employed in many signal processing applications [1, 46–49] to maintain the probability of false alarm $P_{fa}$ constant in spite of the uncertainties and the variation of the unknown parameters [1]. The tests which are not CFAR are impractical for many applications where the false alarm requirement must be guaranteed. The adaptive thresholding methods are often applied to maintain the CFAR property of tests in classical coherent and incoherent radar target detection [19, 20, 50, 51]. For example for the incoherent radar target detection in [19], the Cell Averaging CFAR (CA-CAFR) test is optimal among all CFAR tests for slowly fluctuating targets with Swerling-I model in homogenous background interference. This test compares the energy of the Cell Under Test (CUT) with an adaptive threshold [52]. The Order Statistic CFAR test is proposed in [53] as an adaptive thresholding method for the non-homogenous background and the modified OS-CFAR tests are suggested
2.1. CONSTANT FALSE ALARM RATE TESTS

in [54, 55].

Various statistical methods are used to derive CFAR tests usually using the coherent signal model [56–59]. For example, a CFAR test is derived in [60] using the waveform entropy and the Rao test proposed in [61] is CFAR. For some problems, the GLRT turns out to be CFAR, e.g., the GLRT in [62] for a multiple input multiple output radar target detection. In [63], an adaptive CFAR test has been derived for signal detection with unknown complex amplitude and uncertain steering vector.

In this thesis, we present new definitions and theorems in order to better understand and describe the CFAR property of tests which are widely used in signal processing. The invariant tests have been proposed for numerous practical signal detection problems (e.g., see [36, 44, 48, 64–68] and references therein). Many properties of the invariant tests and their applications have been investigated in statistical literature (see e.g. [10, 24, 67–69]). Recently, some new works have attempted and developed some invariance concepts. For example, some invariant tests are proposed in [65] for detecting a phase shift keying signal in Gaussian noise. In [70], invariant tests are proposed for a radar detection problem. In [44] for a signal detection problem, it has been shown that the performance of the GLRT tends to that of the UMPI test as $P_{fa}$ tends to zero, or as the signal-to-noise-ratio tends to infinity. The invariance property of the GLRT have been clarified in [64]. In [66], some invariance properties and the relationship between the Rao test, the Wald test, the GLRT and the UMPI test have been investigated. In [71], some transformations have been suggested allowing to simplify the derivations of GLRT and UMPI test. In [14], the idea of deriving Rao test and GLRT using the maximal invariant statistic of the observation with respect to the total group of invariant transformations is proposed for
detection of a signal in noise with unknown covariance matrix. Moreover in [72], it is shown that the UMPI test and GLRT are CFAR for signal detection in a wide class of noise distributions. Nevertheless, the important and practical properties of CFAR are not fully and efficiently investigated in the literature.

In this thesis, we introduce the minimally Invariant Family of Distributions (IFD) and the MIG of transformations. We show that the pdf of any maximal invariant with respect to an MIG does not depend on the unknown parameters, while preserves the maximum information of observations. Moreover, we prove that the necessary condition to eliminate the unknown parameters in the pdf of any invariant statistic is that the family of distributions be minimally invariant. This result allows us to prove that any invariant test with respect to an MIG is CFAR. In addition, any CFAR test is also invariant with respect to an MIG if the support of the pdf of observations does not depend on the unknown parameters. This result confirms the CFAR property reported for the GLRT proposed for signal processing problems, e.g., [15, 36, 42, 48, 73]. For any given CFAR test $\psi(\cdot)$, we show that there exists a function of observations $z(\cdot)$ such that the family of distributions of $z(\cdot)$ is minimally invariant and $\psi(\cdot)$ depends on the observations only through $z(\cdot)$. Moreover, we prove that the LR test using the maximal invariant of $z(\cdot)$ with respect to the MIG outperforms or performs the same as $\psi(\cdot)$. For the minimally invariant problems, we show that the optimal CFAR performance bound is given by the LR of the maximal invariant of an MIG. We derive this upper performance bound for a wide class of signal detection in Gaussian noise (for which the optimal CFAR test is not realizable) and propose three novel suboptimal CFAR tests (SFET-CFAR, average LR of maximal invariant Test and GLRT-CFAR) using the maximal invariant of MIG where the
SFET-CFAR is asymptotically optimal.

2.2 Preliminary Definitions and Terminology

Throughout this paper bold-face upper case letters (e.g. \(X\)) denote matrices, bold-face lower case letters (e.g. \(x\)) denote vectors, light-face upper case letters (e.g. \(Q\) and \(X\)) denote sets or groups, and light-face lower case letters (e.g. \(x\) and \(q\)) denote scalars or transformations (deterministic or random). The pdf of a random vector \(x\) is denoted by \(f_x(x; \theta)\), where \(\theta\) is the unknown parameter vector \(f_x(x; \theta, \mathcal{H})\) is the pdf under hypothesis \(\mathcal{H}\). The probability and expectation are defined as \(\Pr_{\theta}(A) = \int_A f_x(x; \theta) dx\) and \(E_{\theta}(g(x)) = \int_{\mathbb{R}^N} g(x) f_x(x; \theta) dx\), respectively. A vector is considered as \(x = [x_1, \cdots, x_N]^T\) where \(x_n = [x]_n\) is the \(n^{th}\) element of \(x\) and \((\cdot)^T\) shows the transpose operation. Moreover \([X]_{n,m}\) denotes the \((n,m)^{th}\) element of matrix \(X\). The \(N \times N\) identity matrix and \(N\) dimensional vector of ones are denoted by \(I_N\) and \(1_N\), respectively. For a real valued function \(g\) from a vector space, the derivative with respect to \(x\) is denoted by \(\frac{\partial g}{\partial x}\), where \(\frac{\partial g}{\partial x_n}\) is its \(n^{th}\) element. The derivative of a vector function \(g\) from a vector space is the matrix \(\frac{\partial g}{\partial x}\) where \([\frac{\partial g}{\partial x}]_{m,n} = \frac{\partial |g|_{m,n}}{\partial |x|_{m,n}}\). The \(L_2\) norm of \((\cdot)\) is denoted by \(\| (\cdot) \|\).

Consider a family of distributions as \(\mathcal{P} = \{f_x(x; \theta)|x \in \mathcal{X} \subseteq \mathbb{C}^N, \theta \in \Theta \subseteq \mathbb{R}^M\}\), where \(x\) is the observation vector, \(\theta\) is the parameter vector, \(\mathcal{X}\) is the sample space and \(\Theta\) is the parameter space of this family. We denote the support of a real valued function by \(\text{supp}(f(\cdot, \theta)) = \text{closure}(\{x|f_x(x; \theta) \neq 0\})\), where \(\text{closure}(A)\) is the closure of set \(A\) [74].

**Definition 1 (Ch. 6 in [10])** We say \(\mathcal{P}\) is an Invariant Family of Distributions (IFD) under the group of transformations \(Q\), if for any \(q \in Q\) there exists a \(\overline{q}(\cdot) \in \overline{Q}\)
such that \( q(\theta) \in \Theta \) and

\[
\Pr_{\theta}\{q(x) \in A\} = \Pr_{\bar{q}(\theta)}\{x \in A\},
\]

(2.3)

where \( \Pr_{\theta}\{x \in A\} = \int_A f(x; \theta) dm_N(x) \) is the probability of an arbitrary Borel set \( A \in \mathcal{X} \) and \( m_N \) denotes the Lebesgue measure in \( \mathbb{C}^N \), \( q \) is a one-to-one and onto transformation on \( \mathcal{X} \) and the set of functions \( \bar{q} : \Theta \mapsto \Theta \) is a group referred to as the induced group denoted by \( \bar{Q} \).

In other words, \( \mathcal{P} \) is invariant with respect to \( Q \), if for any \( q \in Q \), the pdf of \( y = q(x) \) belongs to \( \mathcal{P} \) and is given by \( f_x(x; \bar{q}(\theta)) \).

Consider (2.1), given \( \mathcal{P} \) and the class of tests defined by

\[
\psi(x) = \begin{cases} 
0, & x \in \Gamma_0, \\
\gamma, & x \in \partial \Gamma_1, \\
1, & x \in \Gamma_1,
\end{cases}
\]

(2.4)

where \( \Gamma_1 \) and \( \Gamma_0 \) are the rejection and acceptance regions, respectively, \( 0 < \gamma < 1 \), and \( \partial \Gamma_1 \) is the boundary of \( \Gamma_1 \). In other words, the test \( \psi(\cdot) \) uses the observed signal to decide.

**Definition 2 ([10, Ch. 6])** The hypothesis testing problem (2.1) is invariant under the group \( Q \) if distribution families \( \mathcal{P}_i = \{f_x(x; \theta) | \theta \in \Theta_i\} \) for \( i = 0, 1 \) are both invariant under \( Q \).

The invariant groups of transformations \( Q \) depend on the sets \( \Theta_0 \) and \( \Theta_1 \) as well as on the family of distributions in each hypothesis. Unfortunately, there is no systematic and universal way for finding these groups for all detection problems. However for
many solved problems in signal processing, these groups have been determined which
give hints about possible subgroups that could be verified for each given problem (e.g.,
see [36,63,65] and references therein).

**Definition 3** (Ch. 6 in [10]) A test (or function) $\psi(\cdot)$ is invariant under $Q$ if
$\psi(q(x)) = \psi(x)$, for all $x \in \mathcal{X} \subseteq \mathbb{C}^N$ and for all $q \in Q$.

**Definition 4** For a group $Q$, $m(x)$ is called a maximal invariant if

1. for all $x$, $m(q(x)) = m(x)$,

2. for all $x_1, x_2$, if $m(x_1) = m(x_2)$, there exists a $q(\cdot) \in Q$ such that $x_2 = q(x_1)$.

The pdf of $m(x)$ depends on $\theta$ only through the induced maximal invariant, which is
the maximal invariant of $\theta \in \Theta$ with respect to $\overline{Q}$, denoted by $\rho(\theta)$ [10].

An invariant test (or function) depends on $x$ only through a maximal invariant
of $x$ [10]. It is shown that the pdf of $m(x)$ under $\mathcal{H}_i$ depends on $\theta$ through $\rho_i(\theta) \triangleq \rho(\theta), \theta \in \Theta_i$ for $i = 0, 1$ [10]. Let $\overline{Q}_i \triangleq \{\overline{q}_i, \overline{q}_i(\theta) = \overline{q}(\theta), \overline{q} \in \overline{Q}, \theta \in \Theta_i\}$, then we
have $\rho_i(\theta)$ as the maximal invariant of $\theta \in \Theta_i$ with respect to $\overline{Q}_i$. In [75], the authors
showed that the detection probability $P_d(\theta) = \int_{\mathcal{C}^N} \psi(x)f_x(x; \theta)dm_N(x)$, $\theta \in \Theta_1$ and
the false alarm probability $P_a(\theta) = \int_{\mathcal{C}^N} \psi(x)f_x(x; \theta)dm_N(x)$, $\theta \in \Theta_0$ of any invariant
test $\psi(\cdot)$ are both monotonic functions of $\rho_1(\cdot)$ and $\rho_0(\cdot)$, respectively.

**Definition 5** We call the test $\psi(\cdot)$ in (2.4) as CFAR if

1. $\Gamma_0$ and $\Gamma_1$ are Lebesgue measurable sets in $\mathcal{X}$,

2. the set $\Gamma_0 \cap (\Gamma_1 \cup \partial \Gamma_1)$ has a zero Lebesgue measure.
3. its false alarm probability is constant versus the unknown parameters under $H_0$, i.e., $P_{fa}(\theta_1) = P_{fa}(\theta_2)$, $\forall \theta_1, \theta_2 \in \Theta_0$.

**Remark 1** Lemma 1 of [75] implies that there exist a function $h : \mathcal{X} \rightarrow \mathbb{R}$ and $\eta \in \mathbb{R}$ such that $\psi(x) = 1$ if $h(x) > \eta$ and $\psi(x) = 0$ if $h(x) < \eta$. In other words, our above CFAR definition is restricted to the class of tests for which the pair $(h, \eta)$ exists.

**Example 1** Let $\psi(x)$ be a binary random variable independent of $x$ with $P[\psi(x) = 1] = P_{fa}$. This trivial test is not CFAR as the second condition of Definition 5 is not satisfied.

**Definition 6** For two groups $(Q, \circ)$ and $(\overline{Q}, \overline{\circ})$, a group homomorphism is a function $h : Q \rightarrow \overline{Q}$ such that for all $q_1, q_2 \in Q$ we have $h(q_1 \circ q_2) = h(q_1) \overline{\circ} h(q_2)$, where $\circ$ and $\overline{\circ}$ are the group laws of $Q$ and $\overline{Q}$, respectively.

The relationship between $Q$ and $\overline{Q}$ is a homomorphism for any IFD under the group $Q$ and the induced group of $\overline{Q}$. For an IFD $\mathcal{P}$ with respect to a group of transformations $Q$, we define a function $h : Q \rightarrow \overline{Q}$ such that $h(q) = \overline{q}$. The groups $Q$ and $\overline{Q}$ are homomorphism, since for all $q_1, q_2 \in Q$ we have $h(q_1 \circ q_2) = \overline{q_1 \circ q_2} = \overline{q_1} \overline{\circ} \overline{q_2} = h(q_1) \overline{\circ} h(q_2)$.

**Definition 7** A group monomorphism is a one-to-one group homomorphism.

In this thesis, the group laws are only the composition of functions, hence we refer to groups only by their sets.
Chapter 3

Separating Function Estimation Tests

The result of this chapter is published in [75]. In this chapter, we highlight some of these relationships. These relationships, despite their intrinsic importance, may lead to suboptimal detectors with acceptable performance. We show that for a one-sided hypothesis testing problem the UMP test may be performed by comparing the MVUE of the unknown parameter with a threshold. Furthermore for a more general case (where $\Theta_0$ and $\Theta_1$ are the union of some subintervals), we show that the UMP test may be performed by comparing the MVUE of an SF with a threshold. This SF $g(\theta)$ is a continuous differentiable function, is positive for $\Theta_1$ and is negative (or zero) for $\Theta_0$. In many practical problems the MVUE does not exist. For such cases, we also use a similar suboptimal detection approach and show that an improved estimator for the SF results in a better decision statistics for our test problem, where the estimation criterion is the probability that estimator be in the $\epsilon$–neighborhood of the true parameter, i.e., $\Pr(\|\hat{g}(\theta) - g(\theta)\| < \epsilon)$. 
3.1. THE RELATIONSHIP BETWEEN MVUE STATISTIC AND THE UMP TEST

The chapter is organized as follows. Section 3.1 deals with the relationship between MVUE and UMP. In Section 3.2 for the cases where the MVUE of the SF does not exist, we show that an improved SF estimator results in an improved detector. The existence conditions and properties of SF are described in Section 3.3. Section 3.4 presents examples and simulations to justify our theorems in Sections 3.1 and 3.2.

3.1 The Relationship Between MVUE Statistic and the UMP test

In the following, we only refer to the efficient MVUEs which attain the CRLB. In this Section, we study the relationship between the UMP test for the problem in (2.1) and the MVUE of some function of the unknown parameters in (2.2). We start with the special one-sided hypothesis tests in one-dimensional (1D) parameter space; then generalize the results in this 1D space. Finally, we extend the results to the multi-dimensional (MD) parameter space.

3.1.1 One-Sided Hypothesis Test: 1D Parameter Space

We first study a one-sided to right hypothesis test where \( \Theta_1 \triangleq \{ \theta | \theta > \theta_b \} \subseteq \mathbb{R} \) and \( \Theta_0 \triangleq \{ \theta | \theta \leq \theta_b \} \) in (2.1) for a known \( \theta_b \) [10]. The following theorem establishes a link between the existence of the UMP test for (2.1) and the MVUE for \( \theta \) in (2.2).

**Theorem 1 (The MVUE and the UMP test)** If \( T(x) \) is the MVUE of \( \theta \) in (2.2), then for the one-sided binary hypothesis test in (2.1) with \( \Theta_1 = \{ \theta | \theta > \theta_b \} \) and \( \Theta_0 = \{ \theta | \theta \leq \theta_b \} \), the size-\( \alpha \) UMP test will be \( \psi(x) = u(T(x) - \eta) \) where \( u(\cdot) \) is the unit step function (i.e., \( u(x) = 1 \) for \( x \geq 0 \) and \( u(x) = 0 \) for \( x < 0 \)) and \( \eta \) is obtained from \( E_{\theta_b}(\psi(x)) = \int_{\mathbb{R}^N} \psi(x)f(x; \theta_b)dm_N = \alpha \) to satisfy a pre-determined value for the false alarm probability, where \( m_N \) is the appropriate Lebesgue measure in \( \mathbb{R}^N \) [74].
3.1. THE RELATIONSHIP BETWEEN MVUE STATISTIC AND THE UMP TEST

Proof 1 For two arbitrary parameters \( \theta_1 \in \Theta_1 \) and \( \theta_0 \in \Theta_0 \), the log-LR is \( \ln(LR(x)) = \ln(f(x; \theta_1)) - \ln(f(x; \theta_0)) \). According to Theorem 3.2 of [23], there exists a \( T(x) \) such that \( \frac{\partial \ln(f(x; \theta))}{\partial \theta} = I(\theta)(T(x) - \theta) \) where \( I(\theta) = \frac{1}{\text{var}(T(x))} \) is the Fisher information function and \( \text{var}(\cdot) \) is the variance of \( (\cdot) \). The indefinite integral of \( \frac{\partial \ln(f(x; \theta))}{\partial \theta} \) with respect to \( \theta \in \Theta_0 \cup \Theta_1 \) gives

\[
\ln(f(x; \theta)) = C(x) + \int I(\theta)(T(x) - \theta)d\theta, = C(x) + I(\theta)T(x) - \theta I(\theta), \quad (3.1)
\]

where \( C(x) \) is a function of \( x \). \( I(\cdot) \) and \( \theta I(\cdot) \) are the anti-derivative of \( I(\theta) \) and \( \theta I(\theta) \), respectively \( (\frac{dI(\theta)}{d\theta} = I(\theta); \frac{d\theta I(\theta)}{d\theta} = \theta I(\theta)) \). Substituting (3.1) into \( \ln(LR(x)) \), we obtain

\[
\ln(LR(x)) = (I(\theta_1) - I(\theta_0))T(x) - (\theta I(\theta_1) - \theta I(\theta_0)). \quad (3.2)
\]

Since the function \( I(\theta) \) is the inverse of the minimum variance attained by the MVUE, we have \( I(\theta) > 0 \) for all \( \theta \in \Theta_0 \cup \Theta_1 \). Therefore, the function \( I(\theta) \) is an increasing function of \( \theta \). Thus from \( \theta_1 > \theta_0 \), we have \( I(\theta_1) - I(\theta_0) > 0 \). This means that the log-LR \( \ln(LR(x)) \) in (3.2) is an increasing function in \( T(x) \). Theorem [10, Theorem 3.4] states that the UMP test exists if the likelihood ratio statistic is increasing in a statistic. In this case, since the likelihood ratio is increasing in \( T(x) \), then the UMP is given by \( \psi(x) = u(T(x) - \eta) \), where \( \eta \) is determined by \( E_{\theta_b}(\psi(x)) = \alpha \).

Remark 2 This theorem allows finding the UMP using the MVUE. However, we have examples in which the UMP test exists whereas the MVUE does not exist for the corresponding unknown parameter. As a counter example, consider a uniform distribution \( f(x, \theta) = U(0, \theta) \), and a hypothesis testing problem \( H_0 : \theta = \theta_b \) versus \( H_1 : \theta > \theta_b \). The size-\( \alpha \) UMP test for this problem is \( \psi(x) = 1 \) if \( x > \theta_b \), \( \psi(x) = \alpha \) if
3.1. THE RELATIONSHIP BETWEEN MVUE STATISTIC AND THE UMP TEST

\[ x \leq \theta_b \] where, the MVUE of \( \theta \) does not exist\(^1\) [23].

**Remark 3 (The MVUE and the UMP test to left)** Here we prove a similar theorem for 1D one-sided to left hypothesis tests, where \( \Theta_1 = \{ \theta | \theta < \theta_b \} \) and \( \Theta_0 = \{ \theta | \theta \geq \theta_b \} \). Any one-sided to left hypothesis test with parameter \( \theta \) can be reformulated as a one-sided to right hypothesis test by defining \( \theta' = -\theta \in \Theta' = \Theta'_0 \cup \Theta'_1 \), \( \Theta'_1 = \{ \theta' | \theta' > -\theta_b \} \) and \( \Theta'_0 = \{ \theta' | \theta' \leq -\theta_b \} \). Let \( -T(x) \) be the MVUE for \( \theta' \), where \( x \sim f(x; -\theta') \). Theorem 1 indicates that the UMP test for \( \Theta'_0 \) versus \( \Theta'_1 \) is given by

\[
\psi(x) = u(\eta - T(x)),
\]

where \( -\eta \) is set such that

\[
E_{\psi}(\psi(x)) = \int_{R^N} \psi(x)f(x; \theta_b)dm_N = \alpha.
\]

Switching back to the original one-sided to left hypothesis test with \( \theta = -\theta' \), the UMP decision rule rejects \( H_0 \) if \( T(x) > \eta \), where \( T(x) \) is the MVUE of \( \theta \).

**Remark 4** We deduce that for the one-sided hypothesis tests, the UMP test follows the structure of the hypotheses; i.e., for the one-sided to right test the UMP test rejects \( H_0 \) if \( T(x) > \eta \) and the one-sided to left test rejects \( H_0 \) if \( T(x) < \eta \). The MVUE of the unknown parameter also gives the UMP decision statistic which should be compared with the threshold in the similar direction in the hypothesis.

**Remark 5** We must note that the MLE is asymptotically efficient, i.e., as the number of observations tends to infinity, the MLE becomes equivalent to the MVUE. In most practical situations, the MVUE of \( \theta \) does not exist; hence, we can use MLE as an asymptotically optimal test, i.e., the performance of \( u(\hat{\theta}_{ML} - \eta) \) tends to the performance of UMP test when \( N \to \infty \), in one-sided to right problem. The result is similar for the test \( 1 - u(\hat{\theta}_{ML} - \eta) \) in one-sided to left problem.

\(^1\)The notation \( \psi(x) = \alpha \) means that \( \psi(x) = 1 \) with probability \( \alpha \) and \( \psi(x) = 0 \) with probability \( 1 - \alpha \) [9].
3.1. THE RELATIONSHIP BETWEEN MVUE STATISTIC AND THE UMP TEST

3.1.2 One-Dimensional Parameter Space: The General Form

For a class of problems in which a UMP test does not exist, there exist a UMP unbiased (UMPU) test. A test is called unbiased if its detection probability $P_d$ is greater than its false alarm probability $P_{fa}$ for all possible unknown parameters [10]. Hence, the UMPU test is the optimal test based on Neyman-Pearson criterion among all unbiased tests. It is known that the UMP test (if exists) is unbiased [10]. However, a size-$\alpha$ UMPU test $\psi(\cdot)$ may be obtained by comparing the LR with a threshold, provided that the threshold satisfies $\sup_{\theta \in \Theta_0} E_\theta(\psi) = \alpha$. In the following theorem, we extend Theorem 1 for a more general case where $\Theta_0$ is the union of some intervals.

Theorem 2 (The MVUE, the UMPU and UMP tests for 1D case) In the hypothesis test (2.1), let $\Theta_0 = \bigcup_{i=0}^{n-1}[\theta_{2i+1}, \theta_{2i+2}]$, where $\theta_1 < \theta_2 < \cdots < \theta_{2n}$, $\Theta_1 = \mathbb{R} - \Theta_0$ be the complement set of $\Theta_0$ and $g(\theta) : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $\Theta_1 = g^{-1}((0, \infty))$ and $\Theta_0 = g^{-1}((\infty, 0])$, where $g^{-1}(A) \triangleq \{\forall x | g(x) \in A\}$. If $T_g(x)$ is the MVUE of $g(\theta)$ in (2.2), then $\psi(x) = u(T_g(x) - \eta)$ is a size-$\alpha$ UMPU test for the hypothesis test in (2.1) where $\eta$ is set such that $E_{\theta_1}(\psi(x)) = \cdots = E_{\theta_{2n}}(\psi(x)) = \alpha$ to satisfy the false alarm probability, where $E_{\theta}(\psi(x)) = \int_{\mathbb{R}^N} \psi(x) f(x; \theta) dm_N$. In addition if UMP exists, is given by $\psi(x)$.

Proof 2 See Appendix A.1.

This theorem states that maximizing the detection probability imposes that the false alarm probability at all boundary points of $\Theta_0$ be equal. This is because of the constraint on the false alarm probability.

For many problems, the MVUE does not exist. In contrast, it is usually easy to find and compute the MLE which is not necessarily efficient. This is why in practice,
the MLE is the most common estimator. The following remark extends the statement of Remark 4 and allows to use the MLE in this more general case for a large number of observations.

**Remark 6** As it is known that \( \hat{g}(\hat{\theta}_{ML}) = g(\hat{\theta}_{ML}) \) [23, Theorem 7.2], the Remarks 4 and 5 can be generalized, i.e., the test \( u(g(\hat{\theta}_{ML}) - \eta) \) is asymptotically UMPU. This test rejects \( \mathcal{H}_0 \) if \( g(\hat{\theta}_{ML}) > \eta \). By solving \( g(\hat{\theta}_{ML}) > \eta \) with respect to \( \hat{\theta}_{ML} \), the rejection region for this test would be \( \hat{\theta}_{ML} < \eta_1, \eta_2 < \hat{\theta}_{ML} < \eta_3, \ldots, \eta_{2N-2} < \hat{\theta}_{ML} < \eta_{2N-1}, \ldots \), where \( \eta_k \) is the \( k^{th} \) sorted solution of \( g(\eta_k) = \eta \), i.e., \( \eta_1 < \eta_2 < \cdots \). So we deduce that the asymptotical UMPU test follows the structure of the parameter space.

Theorems 1 and 2 clarify the relationship between the UMP and the UMPU tests in (2.1) and the MVUE of unknown parameter in (2.2) where the unknown parameter is a real number. They allow us to find the UMPU and the UMP tests (if they exist) from the MVUE of the unknown parameter in (2.2).

### 3.1.3 Multi-Dimensional Parameter Space

In this section we investigate on similar extended relationships where there are multiple parameters involved in (2.1) and (2.2), i.e., \( \Theta_0, \Theta_1 \subseteq \mathbb{R}^M \). The following theorem generalizes Theorem 2 for the case of MD parameter space.

**Definition 8 (Separating Function)** Let \( \Theta_0 \) and \( \Theta_1 \) be disjoint subsets of \( \mathbb{R}^M \). Then, the function \( g : \mathbb{R}^M \to \mathbb{R} \) is called an SF for \( \Theta_0 \) and \( \Theta_1 \) if it continuously maps the parameter sets \( \Theta_0 \) and \( \Theta_1 \) into two separated real intervals, i.e., \( \Theta_0 \subseteq g^{-1}((-\infty,0]) \) and \( \Theta_1 \subseteq g^{-1}((0,\infty)) \).
The threshold in this definition for SF can be an arbitrary value; however, we choose a zero threshold to simplify the definition.

**Theorem 3** Let \( g(\cdot) \) be an SF for two disjoint subsets \( \Theta_0 \) and \( \Theta_1 \) of \( \mathbb{R}^M \) in (2.1). Then, if the MVUE of \( g(\theta) \) denoted by \( T_g(x) \) for (2.2) exists, then the size-\( \alpha \) UMP test for (2.1) is given by 

\[
\psi(x) = u(T_g(x) - \eta),
\]

where \( \eta \) satisfies \( E_{\theta_g}(\psi(x)) = \alpha \), for all \( \theta_g \in \{\theta | g(\theta) = 0\} \). And if UMP test exists, it is given by \( \psi(x) \).

**Proof 3** See Appendix A.2.

The highlighted relationship between the MVUE and the UMP test allows us to find the UMP test from the MVUE in many examples (if it exists). In some practical problems, the UMP test for hypothesis testing model or the MVUE for a \( g(\theta) \) does not exist. As a new approach (even for these cases) finding the SF \( g(\theta) \) allows us to seek for optimal or alternative suboptimal estimators of \( g(\theta) \) in order to find optimal or suboptimal decision rules. To elaborate on this new approach, we aim to establish a closer relation between the performance of an estimator of the SF, \( \hat{g}(\theta) \) for (2.2) and that of its corresponding test \( \psi(x) = u(\hat{g}(\theta) - \eta) \) for (2.1).

### 3.2 A Perspective from Estimation Theory

In the previous section, we studied the relationship between the UMP and UMPU tests and the MVUE of SF. Unfortunately for many practical problems, the MVUE does not exist or cannot be found in order to lead us to the UMP decision statistic. In the following, we develop a theorem which assists us to find suboptimal decision statistic employing suboptimal estimators.
3.2. A PERSPECTIVE FROM ESTIMATION THEORY

3.2.1 Suboptimal Decision Statistics: One-Sided-1D Case

Definition 9 Let \( \hat{\theta} \) be an estimator for the unknown parameter \( \theta \) in (2.2). We define \( \hat{\theta} \) as \((\epsilon,p)\)-accurate at the true value \( \theta \) provided that \( \Pr(|\hat{\theta} - \theta| \leq \epsilon) = p \) for \( \epsilon > 0 \). In addition, comparing two estimators \( T_1(x) \) and \( T_2(x) \), we say \( T_1(x) \) is more \( \epsilon \)-accurate at \( \theta \), if \( \Pr(|T_1(x) - \theta| \leq \epsilon) > \Pr(|T_2(x) - \theta| \leq \epsilon) \).

Definition 10 For a one-sided to right hypothesis test, we call \( \psi(\cdot) \) as a reliable detector, if

1. \( E_\theta(\psi) \) is monotonically increasing in \( \theta \in \Theta_1 \cup \Theta_0 \),

2. and the ROC curve is continuous with bounded inflection points\(^2\) for all \( \theta \).

Similarly for a one-sided to left hypothesis test, the \( E_\theta(\psi) \) should be decreasing in \( \theta \in \Theta_1 \cup \Theta_0 \).

Let \( \Gamma_1 \) and \( \Gamma_0 \) denote the rejection and acceptance regions of the test \( \psi(\cdot) \). In Section 3.3, we show that, there exists a continuous function \( h(x) : \mathbb{R}^N \to \mathbb{R} \) and a threshold \( \eta \in \mathbb{R} \) such that, \( \Gamma_1 = h^{-1}((\eta, \infty)) \) and \( \Gamma_0 = h^{-1}((-\infty, \eta]) \). Hence, by the decision rule we mean that \( \psi(x) = 1 \) for \( x \in \Gamma_1 \) and \( \psi(x) = 0 \) if \( x \in \Gamma_0 \), where \( x \) is the observation vector in \( \mathbb{R}^N \). We must note that here \( h(x) : \mathbb{R}^N \to \mathbb{R} \) is a real valued function from the observation space.

The following theorem allows us to compare two reliable detectors and determine which estimator results in a better detector according to \((\epsilon,p)\)-accurate criterion.

Theorem 4 Consider a one-sided to right hypothesis test in 1D, i.e. \( \Theta_0 = \{\theta|\theta \leq \theta_b\} \) and \( \Theta_1 = \{\theta|\theta > \theta_b\} \) and its corresponding \( \theta \)-estimation problem \( x \sim f(x;\theta), \theta \in \Theta = \)

\(^2\)An inflection point is where the convexity of a curve is exchanged.
\( \Theta_0 \cup \Theta_1 \). For two reliable detectors \( \psi_1(x) \) and \( \psi_2(x) \), if \( P_{d1} = E_{\theta \in \Theta_1} (\psi_1) > P_{d2} = E_{\theta \in \Theta_1} (\psi_2) \) for all \( P_{fa} \in [0, 1] \), then there exist two estimators for \( \theta \) (i.e., two functions) \( h_1(x) \), \( h_2(x) \) and correspondingly \( c_1 \) and \( c_2 \) such that \( \psi_1(x) = u(h_1(x) - c_1) \) and \( \psi_2(x) = u(h_2(x) - c_2) \). Furthermore there exists \( \epsilon_{\text{max}} > 0 \) such that for all \( \epsilon \in (0, \epsilon_{\text{max}}] \) and \( \theta > \theta_0 \), \( h_1(\cdot) \) is more \( \epsilon \)-accurate than \( h_2(\cdot) \).

**Proof 4** See Appendix A.4.

The results of this theorem can be easily extended to the one-sided to left tests similar to our previous discussion, by converting the one-sided to left hypothesis test to a one-sided to right one, through a simple variable change of the parameter, \( \theta' = -\theta \).

After such a transformation, provided that the conditions of Theorem 4 are satisfied, the most \( \epsilon \)-accurate estimator of \( \theta' \) that we could find can serve as the best decision statistic for the test problem.

We now generalize Theorem 4 for the MD case using the same approach as in Section 3.3. In particular, we attempt a continuous function \( g(\theta) : \Theta \rightarrow \mathbb{R} \) which separates \( \Theta_0 \subset \mathbb{R}^M \) and \( \Theta_1 \subset \mathbb{R}^M \). First, we update the definition of a reliable detector for MD case.

**Definition 11** For (2.1) where an SF \( g(\theta) : \Theta \rightarrow \mathbb{R} \) exists for \( \Theta_0, \Theta_1 \subset \mathbb{R}^M \), we call a test \( \psi(x) \) as reliable if \( E_{\theta}(\psi(x)) \) is an increasing function in \( g(\theta) \) and its ROC curve is continuous with bounded inflection points for all \( \theta \).

The following theorem extends Theorem 4 for this case.

**Theorem 5** Consider a general form of the hypothesis test in MD, \( \mathcal{H}_0 : \theta \in \Theta_0 \) versus \( \mathcal{H}_1 : \theta \in \Theta_1 \) where the pdf of observation \( x \) is \( f_x(x; \theta) \) by its corresponding SF \( g(\theta) \). Also consider the \( \theta \)-estimation problem \( x \sim f_x(x; \theta), \theta \in \Theta = \Theta_0 \cup \Theta_1 \).
For two reliable detectors $\psi_1(x)$ and $\psi_2(x)$, if $P_{d1} = E_{\theta \in \Theta_1}(\psi_1) > P_{d2} = E_{\theta \in \Theta_1}(\psi_2)$ for all $P_{fa} \in [0, 1]$, then there exist two estimators for $g(\theta)$ (i.e., two functions) $h_1(x)$, $h_2(x)$ and correspondingly $c_1$ and $c_2$ such that $\psi_1(x) = u(h_1(x) - c_1)$ and $\psi_2(x) = u(h_2(x) - c_2)$. Furthermore there exists $\epsilon_{max} > 0$ such that for all $\epsilon \in (0, \epsilon_{max}]$ and $\theta \in \Theta_1$, $h_1(\cdot)$ is more $\epsilon$-accurate than $h_2(\cdot)$ with respect to $g(\theta)$.

**Proof 5** The proof is similar to the proof of Theorem 4. Just substitute $\theta$ by $g(\theta)$.

### 3.2.2 Suboptimal SFETs

Our results lead us to propose the following approach for employing a suboptimal estimator as a decision statistic for testing problems:

- Given $\Theta_0$ and $\Theta_1$ find an SF $g(\cdot)$ which satisfies Definition 8. Note that Lemmas 1 and 2 guarantee the existence of $g(\cdot)$.

- Find an estimator for $g(\theta)$ and denote it by $\hat{g}(\theta)$, or use an estimator of the unknown parameters $\hat{\theta}$ (such as MLE) and define $\hat{g}(\theta) \triangleq g(\hat{\theta})$.

- Use $\hat{g}(\theta)$ as the decision statistic, i.e., $\psi_{g-ET}(x) = u(\hat{g}(\theta) - \eta)$, where $\eta$ satisfies $\epsilon_{max} = \sup_{\theta \in \Theta_0} E_{\theta}(\psi(x)) = \alpha$ and $\alpha$ is the test size.

For the SF $g(\cdot)$, hereafter, we sometime refer to such an SFET as the $g$-Estimator Test. Figure 3.1 shows the block diagram for deriving the SFET.

### 3.3 How to find an SF?

In some signal processing applications (e.g., see some of examples in Section 3.4), we intuitively expect that the detection probability should improve as the Signal to
3.3. HOW TO FIND AN SF?

Noise Ratio (SNR) increases. In addition Definition 11 and Theorem 5 reveal that the detection probability for a reliable test must be increasing with respect to SF. Therefore in these cases, the SNR may be a valid SF (this shall be verified for each case) and an SNR estimator gives a decision statistic.

In this section, three methods for finding an SF are proposed in (3.3), (3.4), and (3.5). The first two are based on the Euclidian distance between the unknown vector and the boundary of $\Theta_0$ and $\Theta_1$. The third method is only applicable to invariant problems (e.g., see [65, 67, 68, 76–78]) and is based on maximal invariants.
In the first two methods all unknown parameters are involved in the SF (3.3) and (3.4), whereas the third method, only takes into account the effective parameters. The following lemma guarantees the existence of the SF and is indeed a measure theoretical extension of Urysohn’s lemma.

**Lemma 1** Let $\Theta_1$ and $\Theta_0$ be two arbitrary Lebesgue measurable sets in $\mathbb{R}^M$ such that the set $\Theta_0 \cap \overline{\Theta_1}$ has a zero Lebesgue measure, where $\overline{A}$ denotes the topological closure of $A$. Then, there exists a continuous function $g : \Theta = \Theta_1 \cup \Theta_0 \rightarrow \mathbb{R}$ and a threshold $c \in \mathbb{R}$ such that $\Theta_1 - \partial \Theta_1 = \{ \theta \in \Theta | g(\theta) > c \}$ and $\partial \Theta_1 = \{ \theta \in \Theta | g(\theta) = c \}$, where $\partial \Theta_1$ denotes the boundary set of $\Theta_1$. Furthermore, $\Theta_0 = \{ \theta \in \Theta | g(\theta) < c \}$ holds almost everywhere with respect to the Lebesgue measure $m_M$.

**Proof 6** See Appendix A.3.

The function given by Lemma 1 may fail to separate the sets for a zero Lebesgue measurable subset of $\Theta_0 \cap \overline{\Theta_1}$. For most applications, the set $\Theta_0 \cap \overline{\Theta_1}$ is empty and there is no need to involve the Lebesgue measure. The condition $m_M(\Theta_0 \cap \overline{\Theta_1}) = 0$ is not necessary also for the case $\Theta_1$ and $\Theta_0$ are disjoint and $\Theta_1$ is open. This is because given $\Theta_1 \cap \partial \Theta_1 = \emptyset$, the proof of Lemma 1 is satisfied with some minor changes and a continuous SF exists.

**Remark 7** The proof of Lemma 1 reveals that the SF is not unique. In addition, the condition that $\Theta_0$ must be disjoint from closure of $\Theta_1$ cannot be omitted from Lemma 1 when $\Theta_1$ is not open. As counter example, there exist disjoint subsets $\Theta_0, \Theta_1 \subset I = [0,1]$ such that for any interval $[a, b] \subset [0,1]$ we have $m_1([a, b] \cap \Theta_0) \neq 0$ and $m_1([a, b] \cap \Theta_1) \neq 0$, see [79]. Suppose for a function $g(\cdot)$ we have $g(\theta_0) \leq c$ for almost all $\theta_0 \in \Theta_0$, and $g(\theta_1) > c$ for almost all $\theta_1 \in \Theta_1$. Then, $g(\cdot)$ cannot be
3.3. HOW TO FIND AN SF?

continuous as \( \Theta_0 \) and \( \Theta_1 \) are dense in \([0, 1]\). This is a contradiction.

Consider an SF \( g(\cdot) : \Theta = \Theta_0 \cup \Theta_1 \to \mathbb{R} \) by

\[
g(\theta) = \begin{cases} 
- \inf_{\vartheta \in \partial \Theta_1} \| \vartheta - \theta \|^2 , & \theta \in \Theta_0, \\
\inf_{\vartheta \in \partial \Theta_0} \| \vartheta - \theta \|^2 , & \theta \in \Theta_1,
\end{cases}
\]

where \( \| \cdot \| \) denotes the Euclidian norm. The function \( \inf_{\vartheta \in \partial \Theta_0} \| \vartheta - \theta \|^2 \) is continuous since \( \| \cdot \|^2 \) and \( \inf(\cdot) \) are both continuous functions. Also for all \( \theta \in \partial \Theta_0 \cap \partial \Theta_1 \) we have \( \inf_{\vartheta \in \partial \Theta_0} \| \vartheta - \theta \|^2 = 0 \). Then the Pasting Lemma, [79, Theorem 3.7] garanties the continuous property of \( g(\cdot) \) in (3.3). For example, \( g(\theta) = \theta^2 - (a + b)\theta - ab \) is an SF function for \( \Theta_0 = (a, b) \) and \( \Theta_1 = \mathbb{R} - \Theta_0 \).

The following lemma ensures the existence of a differentiable function \( g(\cdot) \) under some conditions that are satisfied for many practical signal processing problems.

**Lemma 2** Let \( \Theta_1 \) and \( \Theta_0 \) be disjoint subsets of \( \mathbb{R}^M \) and \( \Theta_0 \) be a compact convex set with a \( C^1 \)-smooth boundary\(^3\). Then a \( C^1 \)-smooth function \( g : \mathbb{R}^M \to \mathbb{R} \) exists such that \( \Theta_0 \subseteq g^{-1}(\{0\}) \) and \( \Theta_1 \subseteq g^{-1}((0, \infty)) \).

**Proof 7** Since \( \Theta_0 \) has a connected and compact \( C^1 \)-smooth boundary, the boundary of \( \Theta_0 \) can be written as \( \partial \Theta_0 = \{ \sigma(t) | t \in [0, 1]^{M-1} \} \), where \( \sigma : [0, 1]^{M-1} \to \mathbb{R}^{M-1} \to \mathbb{R}^M \) is a \( C^1 \)-smooth function. Define \( g : \mathbb{R}^M \to \mathbb{R} \) by

\[
g(\theta) \triangleq \begin{cases} 
\inf_{t \in [0, 1]^{M-1}} \| \theta - \sigma(t) \|^2 , & \theta \in \Theta_0^c, \\
0 , & \theta \in \Theta_0,
\end{cases}
\]

\(^3\)A differentiable function with a continuous derivative is called \( C^1 \)-smooth.
where the superscript $^c$ is the complement operator. It is obvious that $g(\cdot)$ is a $C^1$-smooth function, $\Theta_0 \subseteq g^{-1}(\{0\})$ and $\Theta_1 \subseteq g^{-1}((0, \infty))$.

Many problems in signal processing are invariant under some group of transformations (e.g., see [36,41,65,67,68,76–78,80,81]). We define the family of distribution $\{f_x(x; \theta) | \theta \in \Theta\}$ as invariant under a group of transformations $Q = \{q|q(x) : \mathbb{R}^N \to \mathbb{R}^N\}$ if for any $q \in Q$ the pdf of $y = q(x)$ denoted by $f(y; q(\theta))$ be within the same family, where $q(\theta) \in \Theta$ is referred to as the induced transformation [10, 36]. The induced group of transformations is denoted by $\mathcal{Q} = \{q|q(\theta) : \mathbb{R}^M \to \mathbb{R}^M\}$.

A hypothesis test is called invariant under $Q$ if both families $\{f_x(x; \theta) | \theta \in \Theta_i\}$ for $i = 0, 1$ are invariant under the group of transformation (see [81] for more details on how to construct the UMP Invariant (UMPI) test which is optimal under the defined transformation using the maximal invariant and the maximal invariant of the induced groups). A test $\psi(x)$ is called invariant under $Q$ if for all $q \in Q$, $\psi(q(x)) = \psi(x)$. It is proven that an invariant test depends on $x$ only through $m(x)$ which is referred to as the maximal invariant with respect to $Q$ [10]. The maximal invariant $m(x)$ with respect to $Q$ satisfies two conditions, 1) for all $q \in Q$, $m(q(x)) = m(x)$ and 2) for all $x_1$ and $x_2$, if $m(x_1) = m(x_2)$ then there exists a $q \in Q$ such that $x_1 = q(x_2)$. It is shown that the pdf of maximal invariant depends on $\theta$ only through the maximal invariant of $\theta$ with respect to $Q$ denoted by $\rho(\theta)$ [10,81].

The following theorem provides a method for finding an appropriate SF.

**Theorem 6** Assume that the problem (2.1) is invariant under the group of transformations $Q = \{q|q(x) : \mathbb{R}^N \to \mathbb{R}^N\}$ and $\overline{Q} = \{q|q(\theta) : \Theta \to \Theta\}$ denotes the induced group of transformations under the unknown parameters $\Theta = \Theta_0 \cup \Theta_1$. Then a maximal invariant of $\theta \in \Theta$ with respect to $\overline{Q}$ denoted by $\rho(\theta) = [\rho_1(\theta), \cdots, \rho_L(\theta)]^T$ gives
3.3. HOW TO FIND AN SF?

an SF for (2.1) by

\[ \rho_{\text{tot}}(\theta) \overset{\Delta}{=} \sum_{l=1}^{L} f_l(\rho_l(\theta)), \quad g(\theta) \overset{\Delta}{=} \rho_{\text{tot}}(\theta) - \sup_{\theta \in \Theta_0} \rho_{\text{tot}}(\theta), \quad (3.5) \]

where \( f_l(\cdot) \) is any monotonic decreasing or increasing function such that \( E_{\theta}(\psi) \) is increasing in \( f_l(\rho_l(\theta)) \).

**Proof 8** See Appendix A.5.

This SF gives flexibility to combine different components of the maximal invariant, since various functions \( f_l(\cdot) \) can be used in (3.5).

**Theorem 7** All SFETs using (3.5) and the MLE of the unknown parameters in \( \Theta = \Theta_0 \cup \Theta_1 \) are invariant tests with respect to \( Q \).

**Proof 9** The definition of MLE implies that \( f(x; \hat{\theta}_{\text{ML}}(x)) \geq f(x; \theta), \forall \theta \in \Theta_1 \cup \Theta_0 \). Thus for all \( q \in Q \) and \( \theta \in \Theta_1 \cup \Theta_0 \), we have \( f(q(x); \hat{\theta}_{\text{ML}}(q(x))) \geq f(q(x); \theta) \). It is known that \( \overline{q}(\cdot) \) is an onto function [10]. Hence for all \( \overline{q}(\theta) \in \Theta_1 \cup \Theta_0 \) we have

\[ f(q(x); \hat{\theta}_{\text{ML}}(q(x))) = f(q(x); \overline{q}(\overline{\theta}_{\text{ML}}(q(x)))) \]

\[ \geq f(q(x); \overline{q}(\theta)). \quad (3.6) \]

In [64], it is shown that, \( f_x(x; \theta) = p(q(x); \overline{q}(\theta)) \left| \frac{\partial q(x)}{\partial x} \right| \), therefore multiplying (3.6) by \( \left| \frac{\partial q(x)}{\partial x} \right| \) we get \( f(x; \overline{q}^{-1}(\hat{\theta}_{\text{ML}}(q(x)))) \geq f_x(x; \theta) \) for all \( \theta \in \Theta_1 \cup \Theta_0 \). Since \( \overline{q} \) is an onto function the later equation deals with \( \overline{q}^{-1}(\hat{\theta}_{\text{ML}}(q(x))) = \hat{\theta}_{\text{ML}}(x) \). Applying the induced transformation and \( \rho(\cdot) \) in both sides of this equation respectively, we have \( \hat{\theta}_{\text{ML}}(q(x)) = \overline{q}(\hat{\theta}_{\text{ML}}(x)) \), and using the maximal invariant property of \( \rho(\cdot) \) we get
\[ \rho(\hat{\theta}_{ML}(q)) = \rho(\bar{q}(\hat{\theta}_{ML}(x))) = \rho(\hat{\theta}_{ML}(x)). \]

This theorem shows that the SFET using the decision statistic \( g(\hat{\theta}_{ML}) \) obtained from the SF in (3.5) and the MLE is an invariant test and has full robustness against the variation of unknown parameters [10,37,38].

**Theorem 8** Using a set of independent observations \( \{x_i\}_{i=1}^L \), the performance of any SFET using MLE asymptotically tends to the performance of UMPU as \( L \to \infty \).

**Proof 10** According to [23, Theorem 7.1] as the population size \( L \) grows, the pdf of \( g(\hat{\theta}_{ML}) \) tends to a Gaussian distribution, its bias tends to zero, and its variance tends to the CRLB (i.e., \( N(g(\theta), \frac{1}{I_g(\theta)}) \)). In addition the MVUE of \( g(\theta) \) denoted here by \( \hat{g}(\theta) \) is obtained from \( \sum_{i=1}^L \frac{\partial}{\partial g(\theta)} \ln(f(x_i; \theta)) = I_g(\theta)(\hat{g}(\theta) - g(\theta)). \) The law of large numbers guarantees that the pdf of \( \sum_{i=1}^L \frac{\partial}{\partial g(\theta)} \ln(f(x_i; \theta)) \) and \( \hat{g}(\theta) \) tend to Gaussian distribution. Moreover, the mean and variance of \( \hat{g}(\theta) \) are \( g(\theta) \) and \( \frac{1}{I_g(\theta)} \) respectively. Hence the pdf of \( g(\hat{\theta}_{ML}) \) tends to that of \( \hat{g}(\theta) \) for large \( L \). Therefore, \( E_\theta(u(g(\hat{\theta}_{ML}) - \eta)) \) approaches \( E_\theta(u(g(\theta) - \eta)) \) as \( L \) grows. The proof is complete as \( E_\theta(u(g(\hat{\theta}_{ML}) - \eta)) \) and \( E_\theta(u(g(\theta) - \eta)) \) for \( \theta \in \Theta_0 \) and \( \theta \in \Theta_1 \) represent the probabilities of false alarm and detection using SFET and UMPU, respectively.

### 3.4 Examples

**Example 2 (relation between UMP and SFET)** Consider the hypothesis test \( \mathcal{H}_0 : x = n \) versus \( \mathcal{H}_1 : x = \theta s + n \), where \( \theta > 0 \) is the unknown scale parameter, \( n \) is a zero mean Gaussian random vector with covariance matrix \( \Sigma \), i.e. \( n \sim \mathcal{N}(0, \Sigma) \) and \( s \) is a known target signal vector with \( N \) samples. The rejection region of the UMP test in this one-sided detection problem is \( s^T \Sigma^{-1} x > \eta_{UMP} \), which is the generalized
3.4. EXAMPLES

Matched-Filter [9] and $\eta_{UMP}$ is set to satisfy the predetermined $P_{fa}$. The MVUE for $\theta$ is given by $\hat{\theta}_{MVU} = \frac{s^T \Sigma^{-1} x}{s^T \Sigma^{-1} s}$, in which the denominator $s^H \Sigma^{-1} s > 0$ is a known positive constant, since $\Sigma$ is positive-definite [23]. Therefore, the UMP decision statistic in this detection problem is nothing but the MVUE of $\theta$, as it is proved in Lemma 1.

Example 3 (relation between UMP and SFET) Let $x_n$'s, $n = 0, \ldots, N - 1$ are independent and identically exponential distributed by the pdf $\theta \exp(-\theta x n) u(x_n)$, where $\theta > 0$ is the parameter of the pdf. Then for the hypothesis test $\mathcal{H}_0 : \theta \in (0, \theta_b]$ versus $\mathcal{H}_1 : \theta \in (\theta_b, \infty)$, we have $\Theta_0 = (0, \theta_b]$ and $\Theta_1 = (\theta_b, \infty)$ in which $\theta_b$ is known. The LR for a given $\theta_1 \in \Theta_1$ and $\theta_0 \in \Theta_0$ is $L(x) = \frac{\theta_1^N \exp(-\theta_1 \sum_{n=0}^{N-1} x_n)}{\theta_0^N \exp(-\theta_0 \sum_{n=0}^{N-1} x_n)}$, which is monotonically decreasing in $\frac{1}{N} \sum_{i=1}^{N} x_i$. Using Karlin-Rubin theorem [9, 10], the UMP test rejects $\mathcal{H}_0$ if $\frac{1}{N} \sum_{n=0}^{N-1} x_n < \eta_{UMP}$.

Now consider the separating function $g : \Theta = \Theta_0 \cup \Theta_1 \rightarrow \mathbb{R}$ by definition $g(\theta) = \frac{1}{\theta_b} - \frac{1}{\theta}$, then we have $\Theta_0 = g^{-1}((-\infty, 0])$ and $\Theta_1 = g^{-1}((0, \infty))$. The MVUE of $g(\theta)$ is $\frac{1}{\theta_b} - \frac{1}{N} \sum_{n=0}^{N-1} x_n$ hence Theorem 2 guaranties that the UMPU test rejects $\mathcal{H}_0$ if $\frac{1}{\theta_b} - \frac{1}{N} \sum_{n=0}^{N-1} x_n > \eta$ or equally $\sum_{n=0}^{N-1} x_n < \eta_{UMP}$. As the hypothesis testing problem has UMP then UMPU and UMP are equivalent. Again, this confirms that the UMP test is obtained by comparing the MVUE with a threshold.

Example 4 (relation between GLRT and SFET) A hypotheses test as in (2.1) is formulated in [82, 83] for distributed spectrum sensing using non-coherent sensors in which the complex observed data $X = [x_1, \ldots, x_M] \in \mathbb{C}^{L \times M}$ has a pdf given by $f(x, \theta) = \pi^{-LM} \prod_{i=1}^{M} \theta_i^{-L} \exp\left(- \sum_{i=1}^{M} \frac{||x_i||^2}{\theta_i}\right)$, where $\theta = [\theta_1, \ldots, \theta_M]^T$ and $\theta_i > 0$. In this problem, the unknown parameters are all equal only under $\mathcal{H}_0$, i.e., $\Theta_0 = \{\theta_1, \ldots, 1^T | \theta > 0\}$. The GLRT for this problem is derived in [82] as $\psi_{GLR}(x) = u \left( \log \left( \sum_{i=1}^{M} ||x_i||^2 \right) - \frac{1}{M} \sum_{i=1}^{M} \log \left( ||x_i||^2 - \eta \right) \right)$. This hypothesis is invariant under
the scale group, \( Q = \{ q(\mathbf{X}) = c\mathbf{X} | c \in \mathbb{C} - \{0\} \} \) and then the induced group of transformations is given by \( \overline{Q} = \{ q(\mathbf{\theta}) = [|c|^2 \theta_1, \cdots, |c|^2 \theta_M]^T | c \in \mathbb{C} - \{0\} \} \). A maximal invariant of \( \mathbf{\theta} \) under \( \overline{Q} \) is \( \rho(\mathbf{\theta}) = [\frac{\theta_1}{(\prod_{m=1}^{M} \theta_m)^{1/M}}, \cdots, \frac{\theta_M}{(\prod_{m=1}^{M} \theta_m)^{1/M}}]^T \). Using the identical function for \( f_1(\cdot) \) in (3.5), an SF based on induced maximal invariant is

\[
g(\mathbf{\theta}) = \frac{\sum_{m=1}^{M} \theta_m}{(\prod_{m=1}^{M} \theta_m)^{1/M}} - M.
\]

In addition, \( \hat{\theta}_m = \frac{1}{L} \|x_m\|^2 \) is the ML estimate of \( \theta_m \). Thus, the SFET using the logarithm of this function and the ML estimates becomes identical with this GLRT (denoted as SFET \(_1\) in Figure 3.2). This property proves that this GLRT converges to the optimal detector as \( L \to \infty \). Interestingly, the MVUE of \( \theta_i \) is also proportional to \( \|x_i\|^2 \). Hence, the SFET obtained using MVUEs also becomes identical with this GLRT.

To better understand, consider

\[
g(\mathbf{\theta}) = \log(\frac{\max_i \{\theta_i\}}{\min_i \{\theta_i\}})
\]

which is another separating functions for this problem and leads to an alternative SFET. Using ML estimates results, we obtain another simpler SFET \( \psi_{\text{SFET}_2}(\mathbf{x}) = u \left( \log(\frac{\max_i \{\|x_i\|^2\}}{\min_i \{\|x_i\|^2\}}) - \eta \right) \) which is not a GLRT. Figure 3.2 shows the performance of SFET \(_1\), SFET \(_2\) and the optimal Neyman-Pearson bound (UMP Bound) which is obtained assuming the true value of parameters are known under both hypothesis. The performance of these SFETs converge to the optimal bound as \( L \to \infty \).

### Example 5 (relation between SFET and GLR)

Consider the problem of detecting a complex sinusoid with \( N \) noisy samples with the pdf of

\[
f(\mathbf{x}; \mathbf{\theta}) = \frac{1}{\pi^N \sigma^{2N}} \prod_{n=1}^{N} \exp\left( -\frac{1}{\sigma^2} |x_n - Ae^{2\pi fn}|^2 \right),
\]

where the unknown vector is \( \mathbf{\theta} = [A, f, \sigma^2]^T \), \( f \in [0, 1) \) is the frequency and the complex amplitude \( A \) is zero only under \( \mathcal{H}_0 \). The GLRT statistic for this problem is given
3.4. EXAMPLES

by $T_{GLR}(\mathbf{x}) = \frac{I(\hat{f}_{ML})}{\frac{N}{||\mathbf{x}||^2}}$, where $I(f) = \frac{1}{N} |\sum_{n=0}^{N-1} x_n \exp(-j2\pi fn)|^2$, $\hat{f}_{ML} = \arg \max_{f \in [0,1]} I(f)$ and $\mathbf{x} = [x_0, \cdots, x_{N-1}]$ is the observation vector [36]. The GLRT rejects $\mathcal{H}_0$, if $T_{GLR}(\mathbf{x}) > \eta_{GLR}$ where $\eta_{GLR}$ is set such that the $P_{fa}$ requirement is satisfied. Now consider the following quantized implementations of the GLRT statistic, $T_m(\mathbf{x}) = \frac{I(\hat{f}_m)}{N ||\mathbf{x}||^2}$, $m \in \mathbb{N}$, where $\hat{f}_m = \arg \max_{f \in \{k/N\}_{k=0}^{mN-1}} I(f)$. Obviously as $m$ increases, $T_m(\cdot)$ tends to $T_{GLR}(\cdot)$. Correspondingly let $\psi_m(\mathbf{x}) = u(T_m(\mathbf{x}) - \eta_m)$ denote the test using $T_m(\cdot)$ instead of $T_{GLR}(\cdot)$, where $\eta_m$ is set to satisfy $P_{fa}$ requirement. Figure 3.3(a) shows the detection probability $P_d$ of $\psi_m(\cdot)$ versus energy to noise ratio (ENR=$\frac{N|A|^2}{\sigma^2}$) for $N = 10$, $m = 1, 2, 3, 10$ and a constant false alarm of $P_{fa} = 0.2$. The ROC curve of $\psi_m(\cdot)$ is plotted in Figure 3.3(b) for $m = 1, 2, 3, 10$ and ENR = 9 dB. The performance improvement in detection is justified by Theorem 5 as $T_m$ gives a better estimate for the SNR as $m$ increases. To better understand this justification, consider $g(\theta) = \frac{\Delta}{\sigma^2}$ which is an SF and converts this test into $\mathcal{H}_0 : g(\theta) = 0$ versus $\mathcal{H}_1 : g(\theta) > 0$. Using $\hat{\sigma}^2_{ML} = \frac{1}{N} \sum_{n=0}^{N-1} |x_n - \hat{A}_{ML} e^{j2\pi \hat{f}_{ML}n}|^2$, $\hat{f}_{ML}$ and $\hat{A}_{ML} = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-j2\pi \hat{f}_{ML}n}$, the SFET proposed in Subsection 3.2.2 can be calculated as

$$g(\hat{\theta}_{ML}) = \frac{T_{GLR}(\mathbf{x})}{1 - T_{GLR}(\mathbf{x})} \begin{array}{c} H_1 \\ \geq \end{array} \eta_{SFET}, \quad (3.7)$$

where $\eta_{SFET}$ satisfies $P_{fa}$ requirement. Since $g(\hat{\theta}_{ML})$ is a monotonic function of $T_{GLR}(\cdot)$ for $0 \leq T_{GLR}(\cdot) \leq 1$, the SFET using the MLE is equivalent to the GLRT. In a similar way $\psi(\mathbf{x}) = u\left(g(\hat{\theta}_m) - \eta'_m\right)$, since we have $g(\hat{\theta}_m) = \frac{T_m(\mathbf{x})}{1 - T_m(\mathbf{x})} \geq H_1$ $\eta_{SFET}$. The $(\epsilon, p)$-accuracy of $g(\hat{\theta}_m)$ (as an estimate of $\frac{|A|^2}{\sigma^2}$) is evaluated in Figure 3.4 versus $\epsilon$ for $N = 10$ and two values of $\frac{|A|^2}{\sigma^2}$. As expected from Theorem 5 in Figure 3.4, we observe that an $\epsilon_{max}$ exists such that for all $\epsilon < \epsilon_{max}$, $g(\hat{\theta}_{10})$ is more $\epsilon$-accurate than $g(\hat{\theta}_3)$, $g(\hat{\theta}_2)$, and $g(\hat{\theta}_1)$ and so on. Note that $\hat{f}_{ML}$ maximizes $I(f)$ over $[0,1]$ a $\hat{f}_m$.
does the same over quantized finite subset. This quantization impacts $I(\hat{f}_m)$ and $\psi_m$ depending on the true value $f$. In particular, if the true value $f$ is close to $\frac{k}{mN}$ for some integer $k$ between 0 and $mN - 1$, then $\psi_m$ outperforms $\psi_{GLR}$ at high SNRs as the quantization eliminates small estimation errors of $f$. However at lower SNRs, the performance improves as $m$ increases.

Now, we consider an additional SFET using the zero crossing estimate $\hat{f}_{zc}$ of $f$ (e.g., see [84]), the scale estimates $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x_n \exp(-j2\pi \hat{f}_{zc} n)$ and $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} |x_n - \hat{A} \exp(j2\pi \hat{f}_{zc})|^2$. We observe that the SEFT statistic $|\hat{A}|^2 \hat{\sigma}^2$ is an increasing function of $\frac{1}{\|x\|^2} |\sum_{n=0}^{N-1} x_n \exp(-j2\pi \hat{f}_{zc} n)|^2$. Figure 3.5(a) shows the Mean Square Error (MSE) of $\hat{f}_{zc}$ and $\hat{f}_{ML}$ as estimates of $f$. Only at low SNRs, the MSE of $\hat{f}_{zc}$ is lower than that of $\hat{f}_{ML}$. Therefore from theorems, it is expected that the SFET using the zero crossing outperforms the SFET using MLE only in low SNRs, which is illustrated in Figure 3.5(b) for different $P_{fa}$.

**Example 6 (comparison between SFET and GLRT)** Consider a particular multi-antenna radar target detection problem [85] with $f_x(x; \theta) = \exp \left( -x^H (\theta_2 R_s + \theta_1 I_N)^{-1} x \right) \pi^N \det (\theta_2 R_s + \theta_1 I_N)$, where $R_s$ is a known $N \times N$ signal covariance matrix, $I_N$ is the identity matrix, $\theta_1 > 0$ is the unknown noise variance and $\theta_2$ is zero under $H_0$ and is positive under $H_1$. It is obvious that the SNR defined by $\rho = \frac{\theta_2}{\theta_1}$ is an SF for this problem. To find the MLE of $[\theta_1, \rho]^T$, we substitute the eigenvalue decomposition for $R_s = U \text{diag} [\lambda_1, \cdots, \lambda_N] U^H$ in this pdf and define $[y_1, \cdots, y_N]^T = U^H x$. The MLE of $[\theta_1, \rho]^T$ is the solution of two nonlinear equations $\hat{\theta}_1 = \frac{1}{N} \sum_{n=1}^{N} \frac{|y_n|^2}{\hat{\rho} \lambda_n + 1}$, and

$$
\hat{\rho} = \frac{\sum_{n=1}^{N} \frac{|y_n|^2 \lambda_n}{\hat{\rho} \lambda_n + 1}^2}{\left( \hat{\theta}_1 \sum_{n=1}^{N} \frac{1}{\hat{\rho} \lambda_n + 1} \right)^2},
$$
which are obtained by setting the partial derivatives of \( \log(f_x(x; \theta)) \) with respect to \( \theta_1 \) and \( \rho \). We propose to solve

\[
\hat{\rho} = \frac{N \sum_{n=1}^{N} \left| y_n \right|^2 \lambda_n}{\left( \sum_{n=1}^{N} \frac{\left| y_n \right|^2}{\hat{\rho} \lambda_n + 1} \right) \sum_{n=1}^{N} \frac{1}{\hat{\rho} \lambda_n + 1}}
\]

iteratively by initializing \( \hat{\rho}_0 = 0 \), and updating \( \hat{\rho}_{k+1} = \xi(\hat{\rho}_k) \), for \( k = 0, 1, 2, \cdots \). The simulation result shows that this iterative algorithm converges rapidly after small number of iterations and the result are accurate (e.g., see \( \hat{\rho}_k \) versus the number of iterations in Figure 3.6(b), for \( N = 10 \) and \( k = 5 \) for \( \rho = 0.4 \) and 4). Therefore the proposed SFET is given by \( \psi_k(x) = u(\hat{\rho}_k - \eta) \) for some \( k \). However the GLRT is easily derived by substituting the MLE of unknown parameters under each hypothesis into the LR as follows

\[
\psi_{\text{GLR}}(x) = u(N \log ||x||^2 - N \log \hat{\theta}_1 - \log |\hat{\rho}R_s + I_N| - \eta_{\text{GLR}}).
\]

This GLRT does not seem to be an SFET. Interestingly, as shown in Figure 3.6(a), our numerical results reveal that the proposed SFET is superior to this GLRT for all \( P_{fa} \).

Example 7 (relation between SFET and GLRT) Let

\[
f(X; \theta) = \frac{1}{|\pi \theta|^L} \exp(-\text{trace}(X^H \theta^{-1}X)),
\]

where \( X \in \mathbb{C}^{M \times L} \), \( \theta \) is the unknown covariance matrix and \( | \cdot | \) is the determinant. We first consider the detection of a zero-mean white complex Gaussian noise
\( \Theta_0 = \{ \theta | \theta = \sigma^2 I_M, \sigma^2 \geq 0 \} \) from a colored one, i.e., the elements of \( \Theta_1 \) are positive definite, Hermitian matrices that are not in \( \Theta_0 \). The GLRT for this problem is obtained in [82] as \( \psi_{\text{GLRT}}(X) = u \left( M \log (\text{trace}(\frac{1}{L}XX^H)) - \log (|\frac{1}{L}XX^H|) - \eta \right) \). The problem is invariant under \( Q = \{ q(X) = cUX | c \in \mathbb{C} - \{0\}, U \in \mathbb{U}^{M \times M} \} \), where \( \mathbb{U}^{M \times M} \) is the set of all \( M \times M \) unitary matrices. The group of induced transformations is given by \( Q_0 = \{ q(\theta) = |c|^2 U \theta U^H | c \in \mathbb{C} - \{0\}, U \in \mathbb{U}^{M \times M} \} \), where \( \lambda_m(\theta) \geq 0 \) refers to the \( m \)th eigenvalue of \( \theta \). According to Theorem 6, using \( f_1(x) = x \) and \( f_2(x) = x^2 \), we can define \( g_1(\theta) = \frac{\sum_{m=1}^{M} \lambda_m(\theta)}{\prod_{m=1}^{M} (\lambda_m(\theta))^{1/M}} - M \) and \( g_2(\theta) = \frac{\sum_{m=1}^{M} (\lambda_m(\theta))^2}{\prod_{m=1}^{M} (\lambda_m(\theta))^{2/M}} - M \), respectively. Replacing the MLE of unknown parameters, \( \hat{\theta} = \frac{1}{L}XX^H \) into \( g_1(\cdot) \) and \( g_2(\cdot) \) the SFET\(_1\) and SFET\(_2\) are given. As Logarithm function is a monotonic increasing function the SFET\(_1\) and GLRT are equivalent.

Also we can consider \( g_3(\theta) = \log (\max_m \lambda_m(\theta)) - \log (\min_m \lambda_m(\theta)) \) as another eigenvalue based SF for this problem. According to Theorem 7, SFET\(_1\), SFET\(_2\), and SFET\(_3\) are invariant tests because \( g_1(\cdot), g_2(\cdot), \) and \( g_3(\cdot) \) depend on \( \theta \) only via \( \rho(\theta) \). Each is a superior detector in some specific scenarios depending on collective distribution of other unknown eigenvalues. In general case it is not easy to prove that a GLRT is asymptotically optimal as \( L \to +\infty \). According to Theorem 8 any SFET using the MLE is asymptotically optimal. Thus, this GLRT is asymptotically optimal.

Figure 3.7 depicts the performance of SFET\(_1\), SFET\(_2\), and SFET\(_3\) for \( M = 3, L = 5, 10 \) and \( L = 15 \). In this simulation we consider the covariance matrix under \( \mathcal{H}_0 \) by \( 0.5 I_M \) and the \((i,j)^{th}\) element of covariance matrix under \( \mathcal{H}_1 \) by \([0.9]_{i,j} \). This simulation shows that the SFETs converge to the optimal bound by increasing \( L \).
3.4. EXAMPLES

Example 8 (relation between SFET and GLRT) Using the notations of the previous example, let us consider the detection of a single source in zero-mean white noise complex Gaussian where \( \Theta = \{ \theta | \theta = \sigma^2 I_M + vv^H, \sigma^2 \geq 0, v \in \mathbb{C}^M \} \), and \( \Theta_0 = \{ \theta | v = 0 \} \) versus \( \Theta_1 = \{ \theta | v \neq 0 \} \). Here, the GLRT is \( \psi_{\text{GLR}}(X) = u \left( \log(\lambda_{\max}(XX^H)) - \log(\text{trace}[XX^H]) - \eta \right) \). Similar to Example 7, this problem is also invariant under \( Q = \{ q(X) = cUX | c \in \mathbb{C} - \{0\}, U \in U_{M \times M} \} \), therefore the maximal invariant depends on the eigenvalues of \( \theta \), in this problem the eigenvalues of \( \theta \) is given by \( [\lambda_{\max}(\theta), \lambda_{\min}(\theta), \ldots, \lambda_{\min}(\theta)]^T \). It can be shown that a maximal invariant of \( \theta \in \Theta \) with respect to the induced group is given by \( \rho(\theta) = \frac{\lambda_{\max}(\theta)}{\lambda_{\min}(\theta)} \). According to Theorem 6, this function is an SF for this problem. The MLE of \( \lambda_{\max}(\theta) \) is \( \hat{\lambda}_{\max}(\theta) = \max_m \lambda_m(\frac{1}{L}XX^H) \) and the MLE of \( \lambda_{\min}(\theta) \) is \( \frac{1}{M-1} \left( \sum_{m=1}^M \lambda_m(\frac{1}{L}XX^H) - \max_m \lambda_m(\frac{1}{L}XX^H) \right) \). Then the SFET based on induced maximal invariant is

\[
\frac{\max_m \lambda_m(\frac{1}{L}XX^H)}{\frac{1}{M-1} \left( \sum_{m=1}^M \lambda_m(\frac{1}{L}XX^H) - \max_m \lambda_m(\frac{1}{L}XX^H) \right)} \overset{\mathcal{H}_1}{\gtrsim} \eta_{\text{SFET}}
\]

. The statistic of SFET is increasing in the statistic of the GLRT, hence this GLRT is also an SFET and therefore is asymptotically optimal.
3.4. EXAMPLES

Figure 3.2: Detection probability versus false alarm probability for SFET1, SFET2 and Optimal Neyman-Pearson bound in Example 4 for $L = 1, 2, 5, 10$

Figure 3.3: Detection probability of $\psi_m$ in Example 5 for $m = 1, 2, 3$ and 10, (a) versus ENR=$\frac{N|A|^2}{\sigma^2}$ for $P_{fa} = 0.2$, (b) versus the probability of false alarm for ENR= 9 dB.
3.4. EXAMPLES

Figure 3.4: The $(\epsilon, p)$-accuracy of $g(\hat{\theta}_m)$ as an estimate of $|A|^2/\sigma^2$, i.e., $\Pr_\theta(|g(\hat{\theta}_m) - |A|^2/\sigma^2| < \epsilon)$ versus $\epsilon$.

Figure 3.5: Relationship between the performance two estimators for $f$ (MLE and zero crossing estimator) and the corresponding SEFTs in Example 5 (a) Mean Square Error of these estimators, (b) Probability of detection versus SNR for these SEFTs.
3.4. EXAMPLES

Figure 3.6: Comparison of the performance of the SFET and the GLRT in Example 6, (a) ROC curve, (b) $\hat{\rho}_k$ versus the number of iterations

Figure 3.7: Missed detection probability of v.s. probability of false alarm for $\text{SFET}_1$, $\text{SFET}_2$, and $\text{SFET}_3$, and comparing their performance by Optimal Neyman-Pearson bound for $M = 3$, $L = 5, 10, 15$ in Example 7.
Chapter 4

Invariance and Optimality of CFAR tests

The result of this chapter is published in [86]. In this chapter, we introduce the minimally IFD and the MIG of transformations. We show that the probability density function (pdf) of any maximal invariant with respect to an MIG does not depend on the unknown parameters, while preserves the maximum information of observations. Moreover, we prove that the necessary condition to eliminate the unknown parameters in the pdf of any invariant statistic is that the family of distributions be minimally invariant. This result allows us to prove that any invariant test with respect to an MIG is CFAR. In addition, any CFAR test is also invariant with respect to an MIG if the support of the pdf of observations does not depend on the unknown parameters. This result confirms the CFAR property reported for the GLRT proposed for signal processing problems, e.g., [15, 36, 42, 48, 73]. For any given CFAR test $\psi(\cdot)$, we show that there exists a function of observations $z(\cdot)$ such that the family of distributions of $z(\cdot)$ is minimally invariant and $\psi(\cdot)$ depends on the observations only through $z(\cdot)$. 
Moreover, we prove that the LR test using the maximal invariant of $z(\cdot)$ with respect to the MIG outperforms or performs the same as $\psi(\cdot)$. For the minimally invariant problems, we show that the optimal CFAR performance bound is given by the LR of the maximal invariant of an MIG. We derive this upper performance bound for a wide class of signal detection in Gaussian noise (for which the optimal CFAR test is not realizable) and propose three novel suboptimal CFAR tests (CFAR-SFET, average LR of maximal invariant Test and CFAR-GLRT) using the maximal invariant of MIG where the CFAR-SFET is asymptotically optimal.

A critical point in this chapter is the concept invariant hypothesis testing problem. The first step of any procedure in this chapter is finding the group of transformations that a problem is invariant with respect to that groups. Finding the invariant groups are proposed in [36,37,44,65,67,70,73] for some practical problems.

This chapter is organized as follow. In Section 4.1, the concept of minimally invariant family of distributions and MIG are developed and it is shown that we can eliminate the unknown parameters using the pdf of an MIG, while preserving the maximum information of observations. Using these concepts, the optimal statistic for some important Gaussian problems are proposed. In Section 4.2, the relationship between the CFAR tests and minimally invariant property is investigated. Several novel tests are proposed in Section 4.3 including the Uniformly Most Powerful CFAR (UMP-CFAR) test as an optimal CFAR test and three suboptimal CFAR tests: i) CFAR SFET, ii) average LR of maximal invariant CFAR test and iii) CFAR-GLRT.
4.1 Reduction Using MIG

The results in this section can be used for the development of novel estimators and detectors such as in CFAR tests using the maximal invariant statistic. We first define a *minimally IFD* using the induced group of transformations. Then an MIG via the induced group is defined for the minimally IFD. We show that the distribution of the maximal invariant with respect to the MIG does not depend on the unknown parameters, while this maximal invariant preserves the maximum information. We also show that the existence of an MIG is a necessary and sufficient condition allowing to eliminate the unknown parameters while preserving all relevant information.

**Definition 12 (minimally invariant family of distributions)** Under a group $Q$ and the induced group $\overline{Q}$, an IFD $\mathcal{P} = \{f_x(x; \theta) | x \in \mathcal{X} \subseteq \mathbb{C}^N, \theta \in \Theta \subseteq \mathbb{R}^M\}$ is called minimally invariant if for all $\theta, \theta' \in \Theta$, there exists a $\bar{q} \in \overline{Q}$ such that $\bar{q}(\theta) = \theta'$.

**Definition 13 (minimal invariant group)** For a minimally IFD, we refer to a subgroup $Q_m$ of $Q$ as an MIG if $h : Q_m \mapsto \overline{Q}$ is a group monomorphism, where $h(q_m) = \overline{q_m}$ for all $q_m \in Q_m$.

**Proposition 1** The pdf of any invariant function $\phi(x)$ with respect to the MIG $Q_m$ for any minimally IFD does not depend on the unknown parameters. Conversely, the condition for the pdf of a maximal invariant $m(x)$ be constant versus the unknown parameters is that the IFD be a minimally IFD.

**Proof 11** To show that the pdf of $\phi(x)$ does not depend on the unknown parameters, we prove $\Pr_{\theta}(\phi(x) \in A) = \Pr_{\theta'}(\phi(x) \in A)$ for any arbitrary measurable set $A \in \phi(\mathcal{X})$ [74, Definition 1.3] and arbitrary $\theta, \theta' \in \Theta$. This is proven since we
4.1. REDUCTION USING MIG

have \( \Pr_{\theta}(\phi(x) \in A) = \Pr_{\theta}(\phi(q_m(x)) \in A) = \Pr_{\theta}(q_m(x) \in \phi^{-1}(A)) = \Pr_{q_m(\theta)}(x \in \phi^{-1}(A)) = \Pr_{\theta}(x \in \phi^{-1}(A)) = \Pr_{\theta}(\phi(x) \in A) \). Conversely, using [10, Theorem 6.2.1] any invariant function \( \phi(x) \) with respect to \( Q_m \) depends on \( x \) only through the maximal invariant \( m(x) \) with respect to \( Q_m \). Moreover, the pdf of \( m(x) \) is a function of unknown parameters only through \( \rho(\theta) \) as in Definition 4 [10, Theorem 6.3.2]. Assuming that \( \rho(\theta) \) is constant (i.e., \( \rho(\theta) = \rho(\theta') \) for all \( \theta, \theta' \in \Theta \)), the second property in Definition 4 for the induced maximal invariant \( \rho(\theta) \) implies that there exists a \( \varphi \in \overline{Q} \) such that \( \theta = \varphi(\theta') \). Hence, the family \( \mathcal{P} \) must be a minimally invariant.

Proposition 1 implies that the necessary and sufficient condition to eliminate the unknown parameters using maximal invariant is that the IFD be a minimally invariant family as any invariant function is a function of the maximal invariant.

The maximal invariant with respect to \( Q \) is an invariant function with respect to \( Q_m \). Hence, the MIG \( Q_m \) in Definition 13 is the smallest set (it is not unique) which preserves the impact of any transformation \( q \in Q \) on \( \theta \). In other words, the maximal invariant with respect to \( Q_m \) preserves the maximum information from the observation while its distribution has no unknown parameters. Moreover, all MIGs are monomorphism; because for two MIGs, we can define the monomorphism \( \zeta : Q_{m1} \mapsto Q_{m2} \) by \( q_{m2} = \zeta(q_{m1}) \triangleq h_2^{-1}(h_1(q_{m1})) \), where \( h_1 : Q_{m1} \mapsto \overline{Q}, h_1(q_{m1}) = \overline{q_{m1}} \) and \( h_2 : Q_{m2} \mapsto \overline{Q}, h_2(q_{m2}) = \overline{q_{m2}} \) are the monomorphism transformations from \( Q_{mi} \) to \( \overline{Q} \) for \( i = 1, 2 \) respectively.

**Proposition 2** For a minimally IFD \( \mathcal{P} \), let \( m_{m1}(x) \) and \( m_{m2}(x) \) be the maximal invariants with respect to two MIGs \( Q_{m1} \) and \( Q_{m2} \), respectively. Then, there exists a one-to-one function \( \zeta : \mathcal{X} \mapsto \mathcal{X} \), such that \( m_{m2}(x) = m_{m1}(\zeta(x)) \).
4.1. REDUCTION USING MIG

Proof 12 See Appendix B.1.

The above proposition states that the maximal invariants of any two MIGs of an IFD have a one-to-one relationship. This means that the maximal invariants obtained employing different MIGs contain the same information. In other words, finding an MIG is enough to eliminate the unknown parameters while its maximal invariant preserves the relevant information. Hence in a given problem, the uniqueness of MIG is not important and the unknown parameter problem can be dealt with by finding any MIG and employing its maximal invariant using Proposition 1. We refer to a maximal invariant \( m_m(x) \) of the MIG \( Q_m \) as a maximin invariant, since it maintains the maximum information from the observations, while its distribution does not depend on the unknown parameters as for MIG, \( h : Q_m \mapsto \overline{Q} \) is a group monomorphism, then there is no smaller invariant group that provides this property.

In some signal processing problems, we need to eliminate a subset of unknown parameters while preserve as much information as possible (see Example 11). Let \( \theta = [\theta_r^T, \theta_s^T]^T \) where \(^T\) is the transpose operator and \( \theta_s \) is the vector of nuisance parameters which must be eliminated [66]. This can be achieved by treating \( \theta_r \) as known and applying Proposition 1 to the family of distributions \( \mathcal{P}_{\theta_r} \triangleq \{ f(x; \theta_r, \theta_s) | \theta_s \in \Theta_s, x \in \mathcal{X} \} \), where \( \Theta_s \) is the set of possible values of \( \theta_s \).

4.1.1 Some Gaussian Families

Example 9 Consider the family of zero mean complex white Gaussian distributions \( \mathcal{P} = \{ \exp(-||x||^2/(\pi\theta^N)) | \theta > 0, x \in \mathbb{C}^N \} \). This family is invariant with respect to \( Q = \{ q \mid q(x) = cUx, \forall c \in \mathbb{C}, \forall U \in \mathbb{C}^{N \times N}, UU^H = I \} \), where \( x \) is the observation vector. The induced group for this problem is given by \( \overline{Q} = \{ \bar{q} | \bar{q}(\theta) = |c|^2\theta \} \). This
4.1. REDUCTION USING MIG

IFD is a minimally invariant, as for any \( \theta, \theta' > 0 \), a \( |c|^2 \) exists such that \( \theta = |c|^2 \theta' \).

The group \( Q_m = \{ q_m | q_m(x) = (x, \xi > 0) \} \) is an MIG for \( \mathcal{P} \), because \( \overline{Q_m} = \overline{Q} \), and the function \( h : Q_m \mapsto \overline{Q} \), \( h(q_m) = \overline{q_m} \) is one-to-one and onto. A maximin invariant is given by \( m_m(x) = [x_1, \ldots, x_{N-1}, x_N] \) which has a generalized Cauchy-Lorentz distribution as follows \( f(m_m) = \frac{(N-1)!}{\pi|m_m|^2} \times \frac{\delta(m_m^{-1})}{2\pi} \), where \( \delta(\cdot) \) is the Dirac delta function and \( m_{m_N} \) is the \( N^{th} \) element of \( m_m \). It is known that \( m = [\frac{x_1}{x_N}, \ldots, \frac{x_{N-1}}{x_N}, 1]^T \) is a maximal invariant with respect to \( Q \). Obviously \( m_m \) preserves more information as \( m \) can be calculated from \( m_m \) and not the other way around. In this example, the unitary transformation does not appear in \( Q_m \), i.e., the smaller group, the more information is preserved in the corresponding maximal invariant. For a problem the maximin may not be unique, but there is a one-to-one relationship between two existing maximins (see Proposition 2). In this problem, \( [\frac{x_1}{\sum_{n=1}^N x_n}, \ldots, \frac{x_{N-1}}{\sum_{n=1}^N x_n}, \frac{x_N}{\sum_{n=1}^N x_n}]^T \) is another maximin.

Example 10 For \( X = [x_1, \ldots, x_M] \), consider the family of complex normal distributions \( \mathcal{P} \) defined by \( f(X; \theta) = \exp(-\sum_{i=1}^M (x_i - \mu)^H \Sigma^{-1} (x_i - \mu)) \) where we have the independent and identically distributed \( x_i \in \mathbb{C}^N \) and \( \theta = (\mu, \Sigma) \) is the pair of unknown parameters. The set of all possible unknown parameters is given by \( \Theta = \{(\mu, \Sigma) | \mu \in \mathbb{C}^N, \Sigma = \Sigma^H > 0\} \) where \( \Sigma > 0 \) means that \( \Sigma \) is a positive definite covariance matrix. This family is invariant under the composition of the affine transformations \( Q_A = \{q_A | q_A(X) = AX + a1^T, A \in \mathbb{C}^{N \times N}, a \in \mathbb{C}^N \} \) and \( Q_p = \{q_p | q_p(X) = X \cdot P, P \in \mathbb{C}^{M \times M} \} \) where \( 1 = [1, \ldots, 1]^T \) is an \( M \)-dimensional vector. The induced group \( \overline{Q} \) is obviously given by \( \overline{Q} = \{\overline{q} | \overline{q}(\theta) = (A\mu + a, A\Sigma A^H)\} \). For \( \mathcal{P} \) to be a minimally IFD, a solution \( q \in Q \) must exist for \( \overline{q}(\theta') = \theta \), for all \( \theta', \theta \in \Theta \). A solution for \( \overline{q}(\theta') = \theta \), i.e., \( (A\mu' + a, A\Sigma'A^H) = (\mu, \Sigma) \) is given by \( A = U\Lambda^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}U^H \)
and \( a = \mu - A\mu' \) where \( \Sigma' = U'\Lambda'U'^H \) and \( \Sigma = U\Lambda U^H \) are eigenvalue decompositions of the covariance matrices. The maximal invariant with respect to \( Q \) under \( \mu = 0 \) is the vector of non-zero eigenvalues of \( (X^HXX^H)^{-1}X \) in ascending order, denoted by \( \lambda(X^HXX^H)^{-1}X \) [15].

The group \( Q \) is not minimal for this IFD as the solution of \( (A\mu' + a, A\Sigma' A^H) = (\mu, \Sigma) \) is not unique. We have found a minimal invariant subgroup as \( Q_m \Delta = \{ q_m(X) = \Upsilon X + \varrho 1^T, \varrho \in \mathbb{C}^N, \Upsilon \in L \} \subset Q \) where \( L \) is the set of all \( N \times N \) positive definite lower triangular matrices. This group is minimal because \( \overline{q_m}(\theta') = \theta \), i.e., \( \Upsilon\Sigma'\Upsilon^H = \Sigma \) and \( \Upsilon\mu' + \varrho = \mu \) has a unique solution in \( Q_m \) given by \( q_m(X) = \Upsilon X + \varrho \) where \( \Upsilon \overset{\Delta}{=} L'^{-1}L \) is calculated using the LU decomposition of \( \Sigma = LL^H \) and \( \Sigma' = L' L'^H \), \( \varrho \overset{\Delta}{=} \mu - \Upsilon\mu' \). A maximal invariant with respect to \( Q_m \) is given by \( M_m(X) = \Pi(X)^{-1}(X - x_M1^T) \), where \( \Pi(X) \in L \) is obtained from the LU decomposition of \( XX^H = \Pi(X)\Pi(X)^H \). The statistic \( M_m(X) \) is a maximin invariant because:

1. For all \( q_m \in Q_m \), \( M_m(q_m(X)) = \Pi(\Upsilon X + \varrho 1^T - (\Upsilon x_M + \varrho)1^T) = (\Upsilon \Pi(X))^{-1}(\Upsilon X + \varrho 1^T - (\Upsilon x_M + \varrho)1^T) = \Pi(X)^{-1}(X - x_M1^T) = M(X) \).

2. For \( M_m(X') = M_m(X) \) (i.e., \( \Pi(X')^{-1}(X' - x'_M1^T) = \Pi(X)^{-1}(X - x_M1^T) \)) there exist a \( q_m \in Q_m \) such that

\[
q_m(X) = \Pi(X')\Pi(X)^{-1}X + x'_M1^T - \Pi(X')\Pi(X)^{-1}x_M1^T = X'.
\]

**Example 11** Consider the estimation of the eigenvalues of the covariance matrix for...
the family of complex normal distributions $\mathcal{P}$ defined by $f(X; \theta) = \frac{\exp(-\text{tr}(XX^H\Sigma^{-1}))}{\pi^{MN} |\Sigma|^{M}}$

where $X \in \mathbb{C}^{N \times M}$ and the unknown parameters $U$ and $\Lambda$ are defined by the eigenvalue decomposition of $\Sigma = U\Lambda U^H$. Since we aim to extract the eigenvalues $\Lambda$, we use $\theta_r = \Lambda$ and define $\mathcal{P}_\Lambda \overset{\Delta}{=} \{f(X; U) = \exp(-\text{tr}(XX^H U \Lambda^{-1} U^H)) |, UU^H = I\}$, where $\Lambda$ is assumed to be known and $\text{tr}(\cdot)$ is the trace of a matrix. The family $\mathcal{P}_\Lambda$ is invariant with respect to $Q = \{q|q(X) = VX, PP^H = I, VV^H = I\}$. The maximal invariant with respect to $Q$ is $m(X) = \lambda(XX^H)$. An MIG for $\mathcal{P}_\Lambda$ is $Q_m = \{q_m|q_m(X) = VX, VV^H = I\}$. A maximin invariant for $Q_m$ is the pair $M_m(X) = (\Delta, \Psi)$ calculated from the Singular Value Decomposition (SVD) of $X = \Xi \Delta \Psi$, as

1. for any unitary matrix $U$, we have $M_m(UX) = (\Psi, \Delta)$

2. a unitary matrix $\Xi \Xi' \Sigma$ exists such that if $M_m(X') = M_m(X)$ then $X = \Xi \Delta \Psi = \Xi \Xi' \Xi' \Delta' \Psi' = \Xi \Xi' \Sigma X'$, where $X = \Xi \Delta \Psi$ and $X' = \Xi' \Delta' \Psi'$ are the SVDs of $X$ and $X'$ respectively.

Interestingly, the joint distribution of $\Delta$ and $\Psi$ is constant with respect to $\Psi$. Thus for any optimal estimation of $\Lambda$ we can use only $\Delta$ \[87, 88\].

**Example 12** This example presents an IFD which is not minimally invariant. Let $\mathcal{P} = \{f_x(x; \theta) = \frac{\exp(-x^H(\theta_1 R + \theta_2 I)^{-1} x)}{\pi^N |\theta_1 R + \theta_2 I|} | \theta = [\theta_1, \theta_2]^T, \theta_1 > 0, \theta_2 > 0, x \in \mathbb{C}^N\}$ and $R$ is an $N \times N$ known (given) positive definite matrix. This problem is invariant under $Q = \{q|q(x) = cx, c \in \mathbb{C}\}$, then we have $Q = \{q|q([\theta_1, \theta_2]^T) = [|c|^2 \theta_1, |c|^2 \theta_2]^T, c \in \mathbb{C}\}$. This IFD is not minimally invariant because, there is no solution $c \in \mathbb{C}$ for $[|c|^2 \theta_1', |c|^2 \theta_2']^T = [\theta_1, \theta_2]^T$ for arbitrary $[\theta_1, \theta_2]^T$ and $[\theta_1', \theta_2']^T$. 
4.2 Relationship of Invariant and CFAR Tests

In this section, we generalize the concept of MIG to the hypothesis testing problems. Then, we present Theorems 9 and 10 in order to establish some relationships between CFAR tests and the invariant tests in the invariant hypothesis testing problem with respect to a MIG. We show that any invariant test with respect to a MIG is CFAR if the hypothesis testing problem is invariant with respect to the MIG. The converse of this phrase is proven under some mild conditions.

Finally, Theorem 11 shows that if a reliable CFAR test exits for a hypothesis testing problem, then there exists a function of observations which is invariant with respect to an MIG. In the next section, these theorems allow us to define the optimal CFAR test, to give a novel approach to find the optimal CFAR test and to propose some suboptimal CFAR tests for the class of hypothesis testing problems which have a provided MIG.

**Definition 14** The hypothesis testing problem (2.1) is called invariant with respect to an MIG if (2.1) is invariant with respect to group $Q$ and a subgroup $Q_m \subseteq Q$ exists such that $Q_m$ is an MIG of $P_0 \triangleq \{f_x(x;\theta) | \theta \in \Theta_0\}$.

**Theorem 9** If (2.1) is invariant under the MIG $Q_m$, then any CFAR test $\psi(\cdot)$ must satisfy $\psi(q_m(x)) = \psi(x)$, $\forall q_m \in Q_m$ $\forall \theta \in \Theta_0$, $\forall x \in \text{supp}(f(\cdot;\theta))$. Moreover, if $\text{supp}(f_x(\cdot;\theta))$ does not depend on $\theta$, then any CFAR test $\psi(\cdot)$ for (2.1) is invariant under $G_m$.

**Proof 13** See Appendix B.2.
Example 13 Consider the pdf

\[ f_x(x; \theta) = \pi^{-N} \theta_2^{-M} \theta_3^{-N} \exp\left(-\frac{|x_1 - \theta_1|^2 + \sum_{i=2}^{M} |x_i|^2}{\theta_2} - \frac{\sum_{i=M+1}^{N} |x_i|^2}{\theta_3}\right) \]

for \( x = [x_1, x_2, \ldots, x_N]^T \) with the unknown parameter vector \( \theta = [\theta_1, \theta_2, \theta_3]^T \) where \( N > M \geq 3 \) and \( \theta_1 = 0 \) under \( H_0 \) and \( \theta_1 \neq 0 \) under \( H_1 \). This is an invariant problem (with respect to the scale group) and is not invariant with respect to an MIG. The test rejecting \( H_0 \) where \( \frac{|x_1|^2}{\sum_{i=2}^{M} |x_i|^2} > \eta \) is a CFAR test; i.e., this is an example for the existence of a CFAR test for a none-minimally invariant problem.

Interestingly, most signal processing problems are invariant (for example see [70] [72] [16] [36]) for which there exists an MIG. For such cases, this theorem proves that the invariant tests GLRT, UMPI, and Locally Most Powerful Invariant (LMPI) are all CFAR. Moreover for such problems, we can find the optimal CFAR bound as a benchmark to evaluate any suboptimal CFAR test. This theorem also guarantees that any test using the maximal invariant with respect to MIG is CFAR. Hence, the suboptimal tests derived after reduction using MIG result in CFAR test which are anticipated to have good performance since the maximal relevant information is preserved after reduction. Motivated by Theorem 9 we present three suboptimal methods using MIG reduction in Subsections 4.3.1, 4.3.3 and 4.3.2.

The following example shows that a reliable CFAR test may be found even for a non-invariant problem, i.e., the converse of this theorem is not correct in general. Theorem 11 presents a converse form of this theorem under some specific condition. Moreover, the following theorem proves that any invariant test is CFAR under a mild sufficient condition.
Theorem 10 All invariant tests are CFAR if (2.1) is invariant under the MIG, $Q_m$.

Proof 14 Consider an arbitrary invariant test $\psi(\cdot)$ defined in (2.4). The probability of false alarm is given by $P_{fa}(\theta) = Pr_{\theta}\{x \in \Gamma_1\}$ for $\theta \in \Theta_0$. To prove the CFAR property, we must show $P_{fa}(\theta) = P_{fa}(\theta')$ for all $\theta, \theta' \in \Theta_0$. Since the problem is invariant with respect to the MIG $Q_m$, using Definition 14, there exists a $q_m \in Q_m$, where $Q_m$ is the induced maximal invariant of $Q_m$, such that $\theta' = q_m(\theta)$ and consequently $P_{fa}(\theta') = P_{fa}(q_m(\theta))$. Since $q_m$ is a one-to-one and onto transformation then we have, $P_{fa}(\theta) = Pr_{\theta}\{x \in \Gamma_1\} = Pr_{\theta}\{q_m(x) \in q_m(\Gamma_1)\}$. Thus from (2.3), we have $P_{fa}(\theta) = Pr_{q_m(\theta)}\{x \in q_m(\Gamma_1)\}$. In the next paragraph, we show that $\Gamma_1 = q_m(\Gamma_1)$, which implies that the last equation becomes $P_{fa}(\theta) = Pr_{q_m(\theta)}\{x \in q_m(\Gamma_1)\} = Pr_{q_m(\theta)}\{x \in \Gamma_1\} = P_{fa}(\theta')$. To complete the proof, we show $\Gamma_1 = q_m(\Gamma_1)$, i.e., $q_m(\Gamma_1) \subseteq \Gamma_1$ and $\Gamma_1 \subseteq q_m(\Gamma_1)$. For the first one, for all $x \in \Gamma_1$, we have $\psi(q_m(x)) = \psi(x) = 1$; thus $q_m(x) \in \Gamma_1$. Therefore $q_m(\Gamma_1) \subseteq \Gamma_1$. The second one $\Gamma_1 \subseteq q_m(\Gamma_1)$ can be concluded by using $q_m^{-1}$ instead of $q_m$. Note that since $Q_m$ is a group, $q_m \in Q_m$ results in $q_m^{-1} \in Q_m$.

Remark 8 Theorem 10 gives a sufficient condition on an invariant test to be CFAR.

We here prove the minimally invariant property for $H_0$ is necessary for an invariant test to be CFAR. For a CFAR invariant test, $\rho_0(\theta)$ is constant for all $\theta \in \Theta_0$, as the false alarm probability is a function of $\rho_0(\theta)$ as in Definition 4. Moreover, the second property in Definition 4 for the induced maximal invariant $\rho_0(\theta)$ implies that for all $\theta, \theta' \in \Theta_0$ a $\bar{q} \in \bar{Q}_0$ exists such that $\theta = \bar{q}(\theta')$. Thus according to Definition 12, $H_0$ is a minimally IFD.

The GLRT is widely used as a suboptimal detector for such problems and some of
4.2. RELATIONSHIP OF INVARIANT AND CFAR TESTS

its properties have been studied in [44,63,64]. In [64], it is shown that the GLRT is
an invariant test for an invariant problem. However, the GLRT is not a CFAR test.
Interestingly, Theorem 10 proves that the GLRT is also a CFAR test for all minimally
invariant problems.

Theorem 10 shows that, any invariant test is CFAR for any invariant problem
with respect to an MIG. Conversely, Theorem 9 provides a converse theorem for
Theorem 10 which shows that under some mild conditions, any CFAR test is invariant
for an invariant hypothesis testing problem with respect to an MIG. In the following,
we establish that this condition is required only for some function of observations.

Theorem 11 If a CFAR test exists for (2.1), an invariant degenerate function of x
under some MIG exists under $H_0$.

Proof 15 Consider a CFAR test $\psi(\cdot)$ with the decision statistic $h(x)$ and the thresh-
old $\eta$ as stated in Remark 1. The third property in Definition 5 implies $Pr_{\theta}(h(x) \geq
\eta) = Pr_{\theta'}(h(x) \geq \eta)$ for all $\eta$ and $\theta', \theta \in \Theta_0$. This means that the cumulative dis-
tribution function (cdf) of $h(x)$ does not depend on $\theta$. It is obvious that the collection of
sets $h^{-1}(\{\eta\})$ for $\eta \in h(\mathcal{X})$ forms a partition for $\mathcal{X}$ and $\mathcal{X} = \cup_{\eta \in h(\mathcal{X})} h^{-1}(\{\eta\})$. Now
let $Q_\eta$ denote the group of all permutations on $h^{-1}(\{\eta\})$, (any one-to-one and onto
transformation from $h^{-1}(\{\eta\})$ to $h^{-1}(\{\eta\})$ is referred to as a permutation). Since
$\eta \in h(\mathcal{X})$, then $h^{-1}(\{\eta\})$ is not empty, so $q_\eta$ exists. Thus for all $q_\eta \in Q_\eta$, we have
$h(q_\eta(x)) = h(x)$. Therefore by defining $Q = \{q|q(x) \triangleq q_\eta(x), \text{ if } x \in h^{-1}(\{\eta\}), q_\eta \in
Q_\eta\}$, for such $q \in Q$, we have $h(q(x)) = h(x)$. Then for any arbitrary set $A \in \mathbb{R}$, we
have $Pr_{\theta}(h(x) \in A) = Pr_{\theta'}(h(x) \in A) = Pr_{\theta'}(h(q(x)) \in A)$. This shows that $h(x)$ is
invariant with respect to $Q$. It means that for any $\theta$ and $\theta'$, we have found a $q$ such
that $Pr_{\theta}(h(x) \in A) = Pr_{\theta'}(h(q(x)) \in A)$. The induced group for $Q$ depends on the
structure of $q_\eta$. Moreover, $Q$ is an MIG since this relationship is correct for all $\theta, \theta'$.

This theorem shows that from an existing CFAR test, we can find a Group $Q$ such that a function of observations is invariant with respect to $Q$. Hence, the CFAR test is a function of maximal invariant of $Q$. The maximal invariant statistic (or an invariant statistic) of $Q$ can be directly extracted from $h^{-1}(\{\eta\})$ such that the existing CFAR test depends on $x$ only through this maximal invariant (or the invariant statistic). Obviously, the LR test for this maximal invariant (or the invariant statistic) is CFAR and outperforms the existing one. Interestingly, this method directly finds the maximal invariant (or the invariant statistic) from the statistic of the existing CFAR test without requiring to determine $Q_\eta$. The following two examples illustrate the application of Theorem 11.

Example 14 The GLRT for the problem in [73, eq. (1)] is given by $\|x\|_2^2 \frac{H_1}{H_0} \sim \tau_{\text{GLR}}$ where $\tau_{\text{GLR}}$ is set to satisfy the false alarm requirement, $x$ is the observation vector with length of $N$ and $\|x\|_1$ and $\|x\|_2$ are the $L_1$-norm and $L_2$-norm of $x$. In [73], it is shown that this GLRT is CFAR. Based on proof of Theorem 11, we have $h(x) = \frac{\|x\|^2_1}{\|x\|^2_2}$ and consider the orbits of $h(x)$ for a fixed threshold $\eta$ as $h^{-1}(\{\eta\}) = \{x \mid \sqrt{\eta} = \frac{\|x\|_1}{\|x\|_2} = \frac{\|x\|_2}{\|x\|_1}\}$. We define $m_n \triangleq \frac{|x_n|}{\|x\|_2}$ for all $n = 1, \cdots, N$. Thus, the orbit $h^{-1}(\{\eta\})$ is $\{x \mid \sum_{n=1}^N m_n = \sqrt{\eta}\}$. The proof of Theorem 11 shows that $h(x)$ depends on $x$, only via an invariant statistic. Since all orbits depends on $x$ only via $m$, then $m$ is an invariant statistic of a function of observations with respect to an MIG. The LR of $m$ provides the UMPI test as given in [73] and outperforms this GLRT. This example shows that we can directly find an enhanced invariant statistic from an existing CFAR test without finding $Q_\eta$ and its maximal invariant allows to find an improved test.
Example 15 As an application of Theorem 11, consider the LMPI test for the problem in [14, eq. (1)] given by

\[ h(X) = L\tilde{\rho}(X) + (L - 1)\tilde{\rho}(X)\tilde{\eta}(X) \geq \tau_{\text{LMPI}}, \]

where \( \tau_{\text{LMPI}} \) is set to satisfy the false alarm requirement and \( \tilde{\rho}(X) \) and \( \tilde{\eta}(X) \) are defined in [14, eq. (3)]. Theorem 11 states that the orbits depend on the observation via a maximal invariant (or an invariant statistic) of a function of \( X \). Setting \( \tilde{\eta}(X) = 0 \), we have \( h^{-1}(\{\eta\}) \) depends on \( \tilde{\rho}(X) \). Theorem 11 shows that \( \tilde{\rho}(X) \) is an invariant statistic. Since \( [h(X), \tilde{\rho}(X)] \) is an invariant statistic of a function of observations, the second term of \( h(X) \), i.e., \( \tilde{\rho}(X)\tilde{\eta}(X) \) is also invariant. As \( \tilde{\rho}(X) \) and \( \tilde{\rho}(X)\tilde{\eta}(X) \) are invariant, \( \tilde{\eta}(X) \) will be invariant too. Thus \( [\tilde{\rho}(X), \tilde{\eta}(X)]^T \) is an invariant vector, then the LR of \( [\tilde{\rho}(X), \tilde{\eta}(X)]^T \) provides an enhanced CFAR detector. Interestingly, this result is confirmed in [14] as this LR test is the Most Powerful Invariant (MPI) test for [14, eq. (1)] and gives a tight upper bound performance for the LMPI.

4.3 Optimal and Suboptimal CFAR Tests

In this section, we define the optimal CFAR test and discuss about the optimality of various CFAR tests. We first show how we can employ an existing CFAR test \( \psi(x) \) which rejects \( H_0 \) for \( h(x) > \eta \) in order to derive an improved CFAR test. The proof of Theorem 9 provides a group \( Q \) for which the family of the distributions of \( h(x) \) is invariant. For any such group, there exists a maximal invariant \( m(x) \) where \( h(x) \) is a function of \( m(x) \), i.e., a function \( t(\cdot) \) exists such that \( h(x) = t(m(x)) \). Thus by virtue of the NP lemma, the LR test of \( m(x) \) outperforms \( \psi(\cdot) \) and is CFAR.

Definition 15 (UMP-CFAR Test) A CFAR test which maximizes the detection
probability for all $\theta \in \Theta_1$ is called the UMP-CFAR test.

The following theorem gives a systematic method to derive the UMP-CFAR for a wide class of problems. For the cases where the UMP-CFAR test does not exist, we propose sub-optimal CFAR tests in Subsections 4.3.1, 4.3.2 and 4.3.3.

**Theorem 12** Assume that the problem in (2.1) is invariant under the group $Q$, and $H_0$ is minimally invariant with respect to $Q_m \subseteq Q$. If the UMP-CFAR test exists, it rejects $H_0$ if

$$LR(m_x) = f(m_x; \rho(\theta)) \geq \eta$$

where $m_x$ is the maximal invariant of $x$ with respect to $Q_m$, $\rho(\theta)$ is the maximal invariant of the induced MIG under $H_1$ for $\theta \in \Theta_1$ and $\eta$ is set to satisfy the false alarm probability requirement.

**Proof 16** See Appendix B.3.

We say the UMP-CFAR does not exist and call it as Most Powerful-CFAR (MP-CFAR) test if the obtained optimal test statistic depends on the unknown parameters (i.e., if any increasing function of $LR(m_x)$ depends on the unknown parameters). The MP-CFAR gives a tight upper-bound performance for all CFAR tests even if the UMP-CFAR test is not realizable. The proof of Theorem 12 in Appendix B.3 reveals that even if $LR(m_x)$ depends on the unknown parameters, the test $u(LR(m_x) - \eta)$ provides an upper-bound performance for all CFAR tests where $u(\cdot)$ is the unit step function. We use the performance of this test as a benchmark to evaluate all suboptimal CFAR tests.

The UMPI test is derived for the composition of all invariant groups $Q$. The decision statistic of the UMPI test is the LR of the maximal invariant $m(x)$ with respect to $Q$. In contrast, the UMP-CFAR test is obtained using the MIG $Q_m \subseteq Q$ as in Theorem 12. In the following we briefly show that UMP-CFAR outperforms
UMPI \cite{10}. From the first condition of Definition 4 and \( Q_m \subseteq Q \), we conclude that \( m(x) \) is invariant with respect to \( Q_m \). However, the second condition of Definition 4 is not always satisfied for \( m(x) \) and \( Q_m \). The UMPI is invariant with respect to \( Q_m \) since the UMPI depends on \( x \) only through \( m(x) \). According to Theorem 12, the UMP-CFAR is given by the LR of \( m_m(x) \), which is the optimal test among all invariant tests with respect to \( Q_m \). The UMP-CFAR outperforms the UMPI since UMPI is also invariant with respect to \( Q_m \). These two tests become identical if \( Q_m = Q \), i.e., if the problem is invariant only under an MIG. Otherwise, the UMP-CFAR test outperforms the UMPI test.

4.3.1 SFET-CFAR

The SFET is a suboptimal test and is given by estimating a SF under both hypotheses. The SF is a function on the unknown parameters to real values that separate both hypotheses by comparing it with a threshold. The following procedure allows to systematically find a novel SFET-CFAR assuming that the distribution under \( H_0 \) forms a minimally IFD with the MIG \( Q_m \):

- Using \( Q_m \), determine the induced MIG \( \overline{Q}_m \) for \( \Theta = \Theta_1 \cup \Theta_0 \) as in Section 4.1.

- Find the induced maximal invariant of \( \theta \) with respect to \( \overline{Q}_m \) denoted by \( \rho_m(\theta) \) as described after Definition 4. Note: the induced maximal invariants with respect to \( Q_m \) under \( H_0 \) and \( H_1 \) are \( \rho_m(\theta) = \rho_m(\theta) \) for \( \theta \in \Theta_1 \) and \( \rho_m(\theta) = \rho_m(\theta) \) for \( \theta \in \Theta_0 \), respectively. The reduced hypothesis testing problem after reduction
is given by

\[
\begin{cases}
    \mathcal{H}_0 : \rho = \rho_0, \\
    \mathcal{H}_1 : \rho \neq \rho_0,
\end{cases}
\]  

(4.1)

where \( \rho_0 \overset{\Delta}{=} \rho_{m_0}(\theta) \) and \( \rho \overset{\Delta}{=} \rho_m(\theta) \)

- Use SF(\( \theta \)) = \( \sum_{l=1}^{L} f_l(\rho_{m,l}(\theta)) \) as a SF where \( f_l(\cdot) \) are arbitrary increasing functions [75].

- Define the SFET-CFAR as \( \psi_{\text{SFET}}(x) = u(\text{SF}(\hat{\theta}) - \eta) \) where \( \hat{\theta} \) is the maximum likelihood estimation (MLE) of \( \theta \in \Theta_0 \cup \Theta_1 \) and the detection threshold \( \eta \) is set to satisfy the false alarm requirement.

This test is invariant with respect to \( Q_m \), because we have \( \theta(\hat{q}_m(x)) = \hat{q}_m(\hat{\theta}(x)) \) [10], thus \( \rho_m(\hat{q}_m(\hat{\theta}(x))) = \rho_m(\hat{\theta}(x)) \), where \( \hat{\theta}(x) \) is the MLE of \( \theta \) as a function of observations. As a result by virtue of Theorem 10, \( \psi_{\text{SFET}}(x) \) is a CFAR test. This SFET is practical because of its asymptotical optimality [75]. Figure 4.1 shows the block diagram for deriving the SFET-CFAR.

### 4.3.2 Average Likelihood Ratio of Maximal invariant CFAR

For some practical problems the UMP-CFAR decision statistic depends on the unknown parameters (for such cases, we say the UMP-CFAR test does not exists). In these cases, the parameters may be eliminated by averaging this decision statistic over some distribution of the unknown parameters. Provided the MIG \( Q_m \), Theorem 10 and Remark 8 guarantee that the distribution of the maximal invariant
4.3. OPTIMAL AND SUBOPTIMAL CFAR TESTS

Figure 4.1: The general block diagram for deriving the SFET-CFAR.

\[ m_m(x), f(m_m; H_0), \] does not depend on the unknown parameters under \( H_0 \). However, the distribution \( f(m_m; \rho_{m_1}, H_1) \) depends on \( \rho_{m_1}(\theta) \) under \( H_1 \). We propose the Average Likelihood Ratio of Maximal invariant-CFAR (ALM-CFAR) test as

\[
\int_{\rho_{m_1}(\theta_1)} \frac{f(m_m; \rho_{m_1}, H_1)dF(\rho_{m_1})}{f(m_m; H_0)} \geq \eta_{ALM}.
\]  \hspace{1cm} (4.2)

where \( F(\rho_{m_1}) \) is an a-priori distribution for the residual unknown parameters under \( H_1 \) and \( \eta_{ALM} \) is set to false alarm requirement. The well known ALR test is not
necessarily CFAR. In contrast, the ALM-CFAR test is always CFAR since it depends on \( \mathbf{x} \) only through \( m_m(\mathbf{x}) \) and is invariant with respect to an MIG. Moreover, this test requires less prior information about the distribution of \( \theta \) because the averaging is performed over distribution of reduced parameter \( \rho_{m_1}(\theta) \) instead of that of \( \theta \). In other words, we propose to first eliminate and reduce some of unknown parameters by using the maximal invariant of the MIG and then eliminate the residual unknowns by averaging. As a result this detector requires less prior information of unknowns and is more robust to mismatch of such information.

4.3.3 GLRT-CFAR

An alternative method to eliminate the unknown parameters is to use the GLRT. In [14], De Maio and Conte proposed GLRT and Rao test after applying a reduction using the maximal invariant with respect to \( Q \) for a specific problem of adaptive detection in homogeneous Gaussian clutter with unknown covariance matrix. Using the concept of the maximal invariant, they have first reduced the unknown parameters and then have dealt with the residual unknowns.

In Propositions 1 and 2, Theorem 10 and Remark 8, we proved that the maximal invariant with respect to the MIG \( Q_m \) reduces the unknown parameters and preserves the necessary and sufficient information for a wide class of problems. Note that for some problems, using \( Q \) instead of \( Q_m \) may cause loss of useful information. For a minimally invariant problem, we propose to reduce the data via the maximin invariant \( m_m(\mathbf{x}) \) of the MIG \( Q_m \) to achieve a CFAR test and preserve the maximum information of observations.

The pdf of \( m_m(\mathbf{x}) \) under \( H_0 \) is given by \( f(m_m; H_0) \), which is independent on the
unknown parameters. However the pdf under $\mathcal{H}_1$ is given by $f(m; \rho_m, \mathcal{H}_1)$ where the residual parameter $\rho_m$ may still depend on $\theta$. In this case, we can use the GLRT to deal with the residual unknowns $\rho_m(\theta)$. Theorem 12 guarantees that the GLRT for such a problem is CFAR. However, we propose to apply the GLRT approach on the distribution of the reduced data $m(x)$ as follows

$$
\sup_{\rho_m} \frac{f(m; \rho_m, \mathcal{H}_1)}{f(m; \mathcal{H}_0)} \geq \frac{f(m; \hat{\rho}_m, \mathcal{H}_1)}{f(m; \mathcal{H}_0)} \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\geq}} \eta_{G\text{-CFAR}}.
$$

where $\hat{\rho}_m$ is the MLE of $\rho_m(\theta)$ under $\mathcal{H}_1$ and $\eta_{G\text{-CFAR}}$ is set to satisfy the false alarm requirement. The invariance of MLE yields $\hat{\rho}_m = \rho_m(\hat{\theta})$ where $\hat{\theta}$ is the MLE of the unknown parameters under $\mathcal{H}_1$. From [75, Theorem 7], we have $\rho_m(\hat{\theta}(g_m(x))) = \rho_m(\hat{\theta}(x))) = \rho_m(\hat{\theta}(x)))$ which reveals that the GLRT-CFAR is invariant with respect to $G_m$, i.e., this test is CFAR. Comparing the GLRT-CFAR and GLRT, the maximal invariant reduction with respect to an MIG eliminates all unknown parameters under $\mathcal{H}_0$ and converts all unknown parameters under $\mathcal{H}_1$ into $\rho_m(\theta)$ where the size of $\rho_m(\theta)$ is smaller than that of $\theta$, i.e., we only use the estimate of $\rho_m(\theta)$ whereas in GLRT we use the estimate of $\theta$.

### 4.4 Signal Processing Examples

We here present three illustrative examples for the application of the proposed methods. Example 16 is a fundamental signal detection problem in spectrum sensing and target detection. Example 17 presents a coherent radar target detection problem.
Example 16 Consider the signal detection problem

\begin{equation}
\begin{cases}
    \mathcal{H}_1 : x = as + w, \\
    \mathcal{H}_0 : x = w,
\end{cases}
\end{equation}

where \( s \in \mathbb{C}^N \) is a complex known signal with \( \|s\|^2 = 1 \), \( a \in \mathbb{C} \) is an unknown complex amplitude and \( w \in \mathbb{C}^N \) is a zero-mean white complex Gaussian noise with unknown variance \( \sigma^2 \). Rearranging this problem in the standard form of (2.1), we have

\[ \Theta_1 = \{ \theta = [a_R, a_I, \sigma^2]^T | a_R \neq 0 \text{ or } a_I \neq 0, \sigma^2 > 0 \} \quad \text{and} \quad \Theta_0 = \{ [0, 0, \sigma^2]^T | \sigma^2 > 0, \} \],

where \( a_R = \text{Re}(a) \) and \( a_I = \text{Im}(a) \), are the real and Image part of \( a \).

Let \( Q \) denote the invariant group of transformations of (4.4). This problem is invariant under the positive scale group \( Q_m = \{ q_m(\mathbf{x}) = \xi \mathbf{x}, \xi > 0 \} \subseteq Q \). The induced groups under both hypotheses are given by

\begin{align*}
    Q_m|_{\mathcal{H}_0} &= \{ q_m([0, 0, \sigma^2]^T) = [0, \xi \sigma^2]^T \}, \quad (4.5) \\
    Q_m|_{\mathcal{H}_1} &= \{ q_m([a_R, a_I, \sigma^2]^T) = [\xi a_R, \xi a_I, \xi \sigma^2]^T \}. \quad (4.6)
\end{align*}

Definition 14 implies that \( G_m \) is an MIG for (4.4), because (4.4) is invariant with respect to \( Q_m \) and the family of distributions is equal to the family of Gaussian noise with unknown variance under \( \mathcal{H}_0 \). Example 9 gives an MIG \( Q_m \) for the problem in (4.4) (this is because for all \( \sigma^2 > 0, \sigma^{2'} > 0 \), there exists a unique \( \xi > 0 \) such that \( \xi^2 \sigma^2 = \sigma^{2'} \)). Thus Theorem 10 guarantees that any invariant test such as the GLRT, SFET-CFAR and the UMPI test are CFAR. In the following, the UMPI, MP-CFAR, GLRT, GLRT-CFAR, and three SFET-CFARs are derived.

The induced maximal invariant with respect to \( Q_m \) is \( \boldsymbol{\rho} = [\frac{a_R}{\sigma}, \frac{a_I}{\sigma}]^T \). Based on [75],
we need to define the induced group of transformation under the union of both hypotheses \( \mathcal{H}_1 \) and \( \mathcal{H}_0 \), i.e., \( \overline{Q}_m \equiv Q_m|_{\mathcal{H}_0} \cup Q_m|_{\mathcal{H}_1} \) for the union of families of distributions under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), where we have \( \overline{q}_m([a_R, a_I, \sigma^2]^T) = [\xi a_R, \xi a_I, \xi^2 \sigma^2]^T \) for the amplitude and noise variance parameters. A maximal invariant of \([a_R, a_I, \sigma^2]^T\) with respect to \( \overline{Q}_m \) is given by \( \rho([a_R, a_I, \sigma^2]^T) = \left[ \frac{a_R}{\sigma}, \frac{a_I}{\sigma} \right]^T \) as it satisfies two conditions of maximal invariant: 1) we have \( \rho(\overline{q}_m([a_R, a_I, \sigma^2]^T)) = \rho([\xi a_R, \xi a_I, \xi^2 \sigma^2]^T) = \left[ \frac{a_R}{\sigma}, \frac{a_I}{\sigma} \right]^T = \rho([a_R, a_I, \sigma^2]^T); \) 2) from \( \left[ \frac{a_R}{\sigma}, \frac{a_I}{\sigma} \right]^T = \rho([a_R, a_I, \sigma^2]^T) = \rho([a'_R, a'_I, \sigma^2]^T) = \left[ \frac{a'_R}{\sigma}, \frac{a'_I}{\sigma} \right]^T \) we get \( [a_R, a_I, \sigma^2]^T = [\xi a'_R, \xi a'_I, \xi^2 \sigma^2]^T \) for \( \xi = \frac{a}{\sigma} \). Hereafter, we define \( \rho_R = \frac{a_R}{\sigma} \) and \( \rho_I = \frac{a_I}{\sigma} \) and \( \rho = \rho_R + j \rho_I \) and so \( \rho(\theta) = [\rho_R, \rho_I]^T \).

To find the optimal CFAR test, first, we find a maximal invariant with respect to \( G_m \), then its LR provides the optimal CFAR bound. A maximin invariant for this problem is \( m_m(x) = \left[ \frac{x_0}{|x_{N-1}|}, \cdots, \frac{x_{N-1}}{|x_{N-1}|} \right]^T \), where \( x_n \)'s, \( n = 0, 1, \cdots, N - 1 \) are the elements of \( x \). The pdf of \( m_m(x) \) under \( \mathcal{H}_1 \) is derived in Appendix B.4 as follows

\[
    f(m_m; \rho, \mathcal{H}_1) = \frac{v_\rho}{\pi \| m_m \|^{2N}},
\]

where \( v_\rho \equiv e^{-\frac{\Re(\rho m_m^H s)}{\| m_m \|^2} \sum_{i=0}^{2N-1} (2N-1)_i \Re(\rho m_m^H s)^i \tau(2N - i, \Re(\rho m_m^H s))} \) and \( \tau(\cdot, \cdot) \) is defined in Appendix B.4. The pdf of \( m_m(x) \) under \( \mathcal{H}_0 \) is obtained by substituting \( \rho = 0 \) in (4.7). Thus, using Theorem 12 the MP-CFAR test is given by

\[
    v_\rho \geq \eta_1 \mid_{\mathcal{H}_0},
\]

where \( \eta_1 \) is set to satisfy the false alarm requirement.

To find the UMPI test, we must consider all invariant groups of transformations.

For this problem, this group is \( Q = \{ q(q(x) = cx; c \in \mathbb{C} \setminus \{0\} \). Similarly to find
the UMPI test first a maximal invariant with respect to \( Q \) is derived then its LR is considered as the UMPI statistic. The maximal invariant with respect to \( Q \) is \( m = \left[ \frac{x_0}{x_{N-1}}, \cdots, \frac{x_{N-2}}{x_{N-1}} \right]^T \) which has the following pdf under \( H_1 \)

\[
f(m; \rho, H_1) = \frac{(N-1)!}{\pi^{N-1} \| m \|^{2N}} \exp \left( \frac{|m^H s|^2}{\| m \|^2} - |\rho|^2 \right) \sum_{k=0}^{N-1} \frac{1}{\begin{pmatrix} N-1 \\ k \end{pmatrix}} \frac{|\rho m^H s|^{2k}}{\| m \|^{2k}}.
\] (4.9)

The pdf under \( H_0 \) is given by replacing \( \rho = 0 \) in (4.9). Thus, the LR of \( m(x) \) is increasing in \( |x^H s| \| x \| \), i.e., the UMPI test is \( \frac{|x^H s|^2}{\| x \|^2} \gtrsim \eta_2 \), where \( \eta_2 \) is set to guarantee the \( P_{fa} \) requirement.

To derive the GLRT, the unknown parameters are substituted by their MLE under each hypothesis in the likelihood ratio. In this case, the MLEs of \( \sigma^2 \) under \( H_0 \) and \( H_1 \) are \( \hat{\sigma}_0^2 = \frac{1}{N} \| x \|^2 \), and \( \hat{\sigma}_1^2 = \frac{1}{N} \sum_{n=0}^{N-1} |x_n - \hat{a}_1 s_n|^2 \), respectively, where \( \hat{a}_1 = s^H x \) is the MLE of \( a \) under \( H_1 \). By substituting these MLEs into the LR, we find the GLRT statistic as \( T_{GLR}(x) = \frac{\| x \|^2}{\| x \|^2 - |x^H s|^2} \), that is increasing in \( \frac{|x^H s|^2}{\| x \|^2} \), i.e., the GLRT and the UMPI are identical.

Now, we derive three SFET-CFARs for this problem using \( \rho(\theta) \). After reduction using \( G_m \), the problem becomes to test

\[
\begin{align*}
\mathcal{H}_0 : \rho_R + j \rho_I &= 0 \\
\mathcal{H}_1 : \rho_R + j \rho_I &\neq 0.
\end{align*}
\] (4.10)

We propose three SFs for this problem as \( SF_1 \overset{\Delta}{=} \rho_R^2 + \rho_I^2 \), \( SF_2 \overset{\Delta}{=} \epsilon |\rho_R + \rho_I|^2 + (1 - \epsilon) |\rho_R - \rho_I|^2 \) and \( SF_3 \overset{\Delta}{=} \max \{ \epsilon |\rho_R + \rho_I| + (1 - \epsilon) |\rho_R - \rho_I|, (1 - \epsilon) |\rho_R + \rho_I| + \epsilon |\rho_R - \rho_I| \} \), where \( \epsilon \in (0, 1) \). These SFs depend on the unknown parameters only through \( \rho \), hence based on Subsection 4.3.1, using the MLE of unknown parameters,
Figure 4.2: Performance comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson for Problem (4.4), \( \rho = 1.5 + j1.5 \) and \( N = 5 \).

the SFETs related to SF\(_1\), SF\(_2\), SF\(_3\) are CFAR. The MLE must be derived from the signal model \( \mathcal{H}_1 \cup \mathcal{H}_0 \) (see [75]), then the family of distributions is the set of all complex Gaussian distributions with mean \( a \) and the covariance matrix \( \sigma^2 I \), where \( a \in \mathbb{C} \) and \( \sigma^2 > 0 \) are unknown. The MLE of \( a \) and \( \sigma \) are given by \( \hat{a} = s^H x \) and \( \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} |x_n - \hat{a}s_n|^2} \), respectively. Thus SFET\(_1\) statistic is given by replacing the MLEs into SF\(_1\) is \( T_{SFET_1}(x) = \frac{|s^H x|^2}{\frac{1}{N} \sum_{n=0}^{N-1} |x_n - \hat{a}s_n|^2} \). This statistic is increasing in \( \frac{|s^H x|^2}{\|x\|^2} \), thus, SFET\(_1\) rejects \( \mathcal{H}_0 \) if \( \frac{|s^H x|^2}{\|x\|^2} > \eta_2 \). In this case, SFET\(_1\) and GLRT are equivalent. Moreover for known \( s \), SFET\(_1\) is equivalent to the UMPI test. Similarly SFET\(_2\) and SFET\(_3\) are obtained by replacing \( \rho_R \) and \( \rho_I \) respectively by the real and imaginary parts of \( \hat{a}/\hat{\sigma} \) in SF\(_2\) and SF\(_3\).

The GLRT-CFAR for this problem is given by replacing the MLE of \( \rho_R \) and \( \rho_I \) under \( \mathcal{H}_1 \) using (4.7) in the LR with respect to \( m_m \). Thus the GLRT-CFAR is given
4.4. SIGNAL PROCESSING EXAMPLES

by $\max_{\rho} v_{\rho} \overset{\mathcal{H}_1}{\gtrless} \eta_3$. This maximization gives the MLE of $\rho$ using the pdf of $m_m$. In our simulations, we calculate this MLE using a numerical search as it has no closed form expression. Alternatively, we suggest a heuristic estimator of $\rho$ using $m_m$ as $\hat{\rho}_m = s^H m_m$. Substituting this estimator in SF$_1$, SF$_2$, and SF$_3$, we obtain SFET$_4$, SFET$_5$, and SFET$_6$. These tests are all CFAR.

In Figure 7.11 for $\rho = 1.5 + j1.5$, $N = 5$ and $\epsilon = 0.1$, we compare the performance of the proposed tests with the Neyman-Pearson (NP) test that rejects $\mathcal{H}_0$ if $\Re(s^H x) > \eta_4$, where $\eta_4$ is set to satisfy the $P_{fa}$ requirement [1]. As discussed, the SFET$_1$, GLRT, and MPI test are equivalent in this case. The NP and the MP-CFAR tests depend on the unknown parameters and provide two upper bounds for all tests and all CFAR tests, respectively. We see that SFET$_5$ outperforms SFET$_1$ (MPI and GLRT) and SFET$_2$ are SFET$_6$ are close to SFET$_1$. We observe that no CFAR test outperforms the MP-CFAR bound as proven earlier. Figures 4.3 shows the performance of detectors versus the number of observations for $\rho = 1.5 + j1.5$. It is seen that SFET$_5$ outperforms SFET$_1$ when $N \leq 7$, while for $N \geq 8$, SFET$_1$ outperforms SFET$_5$. Although for $N \geq 8$, the probabilities of detection of SFET$_1$ and SFET$_5$ are so close to one. In Figures 7.12, the performance of detectors versus the signal-to-noise Ratio (SNR) in dB is depicted for $N = 5$. It is shown that SFET$_5$ and SFET$_6$ outperform SFET$_1$. Also SFET$_2$ outperforms SFET$_1$ for SNR less than 5 dB.

Example 17 (MP-CFAR for coherent radar detection) Here, we apply the results of the problem (4.4) for the target detection problem of a coherent radar in a homogenous clutter field [52]. In a classical radar detection, the distribution of a number of reference vectors is assumed to be the same as the noise from the CUT.
4.4. SIGNAL PROCESSING EXAMPLES

Figure 4.3: Performance comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson versus the number of observations for Problem (4.4), \( \rho = 1.5 + j1.5 \) and \( P_{fa} = 10^{-3} \).

Figure 4.4: Probability of detection comparison of the MP-CFAR, GLRT-CFAR, SFETs and Neyman-Pearson, versus signal to noise ratio (in dB) for Problem (4.4) for \( P_{fa} = 10^{-3} \), \( N = 5 \), and phase of \( \rho \) is \( \pi/4 \).
4.4. SIGNAL PROCESSING EXAMPLES

The reference vectors are signals within the same pulse repetition interval corresponding to the cells around the CUT. Consider the I-Q output of the matched filter of any cell as a complex random variable \( x_n \), where \( n \in \{1, \ldots, N\} \) and \( N \) is the number of cells. The pdf of \( x_n \) if there is no target in the \( n \)th cell, is zero mean Gaussian with unknown variance \( \sigma^2 \). On the other hand, the pdf of \( x_n \) with a presented target in the \( n \)th cell is complex Gaussian with unknown variance \( \sigma^2 \) and nonzero unknown mean of \( a \). Then we can rewrite the target problem as 

\[
H_1 : x = ax + w \quad \text{and} \quad H_0 : x = w,
\]

where \( s \triangleq [0, 0, \cdots, 0, 1, 0, \cdots, 0]^T \) and \( w \) is a white Gaussian clutter with unknown variance \( \sigma^2 \). Note that, in \( s \), 1 is replaced in the location at the \( c \)th place as CUT.

This problem is invariant under the scale group. Similarly the UMPI test rejects \( H_0 \) if 

\[
\frac{|s^T x|^2}{\|x\|^2} = \frac{|x_c|^2}{\|x\|^2} > \eta_{\text{UMPI}},
\]

where \( \eta_{\text{UMPI}} \) is set to satisfy the false alarm requirement. This UMPI test is equivalent to comparing 

\[
\frac{|x_c|^2}{\|x\|^2 - |x_c|^2} \quad \text{with} \quad \frac{\eta_{\text{UMPI}}}{1 - \eta_{\text{UMPI}}} = \eta_{\text{CA-CFAR}},
\]

which is known as the CA-CFAR test. However, our previous results reveal that the optimal MP-CFAR test is given by (4.8).

**Example 18** Consider a non-homogeneity field detection applied in radar imaging.

The observation vector is as \( x = [x_1, \cdots, x_M, \cdots, x_N]^T \), where \( x_n \)'s are exponentially distributed with parameter \( \theta_1 \) if \( n \leq M \) and \( \theta_2 \) if \( n > M \), i.e., the pdf of observations is 

\[
f(x; \theta) = \exp\left(-\frac{1}{\theta_1} \sum_{n=1}^{M} x_n - \frac{1}{\theta_2} \sum_{n=M+1}^{N} x_n\right)/(\theta_1^M \theta_2^{N-M}),
\]

where the unknown vector is defined by \( \theta = [\theta_1, \theta_2]^T \) and \( M \) is known. Consider a non-homogeneity field detection problem as 

\[
\begin{cases}
H_0 : \theta_1 = \theta_2 \\
H_1 : \theta_1 \neq \theta_2.
\end{cases}
\quad (4.11)
\]

The sets of unknown parameters under \( H_0 \) and \( H_1 \) are \( \Theta_0 = \{[\theta_1, \theta_2]^T | \theta_1 = \theta_2\} \) and
\[ \Theta_1 = \{ [\theta_1, \theta_2]^T | \theta_1 \neq \theta_2 \} \], respectively. Here, we show that a minimal group for this problem is \( G_m = \{ g_m(x) = bx, b > 0 \} \). The induced group of \( G_m \) under \( \mathcal{H}_0 \) is \( \overline{Q}_m|\mathcal{H}_0 = \{ \overline{q}_m(\theta) = [b\theta_1, b\theta_2]^T, \theta_1 = \theta_2 \} \); thus for any \( \theta > 0 \) and \( \theta' > 0 \), there exists a unique \( b \) such that \( b\theta = \theta' \). A maximal invariant with respect to \( G_m \) is \( \mathbf{m}_m = \begin{bmatrix} x_1 \\ x_N \\ \vdots \\ x_{N-M+1} \end{bmatrix}^T \). Using the Jacobian method, the pdf of \( \mathbf{m}_m \) is \( f(\mathbf{m}; \rho) = \frac{\Gamma(N)}{(\rho \frac{M}{N} \sum_{n=1}^{M} m_n + \rho \frac{N}{M} \sum_{n=M+1}^{N} m_n)^N} \) where \( m_N = 1 \) and \( \rho = \frac{\theta_2}{\theta_1} \). Therefore, the hypothesis problem after reduction becomes

\[
\begin{cases}
\mathcal{H}_0 : \rho = 1 \\
\mathcal{H}_0 : \rho \neq 1.
\end{cases}
\tag{4.12}
\]

Using the pdfs of \( \mathbf{m}_m \), the MP-CFAR and the GLRT-CFAR are

\[
\frac{\sum_{n=1}^{N} m_n}{\rho^{1-\frac{M}{N}} \sum_{n=1}^{M} m_n + \rho \frac{M}{N} \sum_{n=M+1}^{N} m_n} \overset{\mathcal{H}_1}{\gtrless} \eta_{\text{MP-CFAR}}.
\]

where \( \eta_{\text{MP-CFAR}} \) and \( \eta_{\text{GLR}} \) are the adjusting thresholds given by the probability of false alarm. The MLE of \( \rho \geq 0 \) under the union of hypotheses is \( \hat{\rho} = \frac{M \sum_{n=1}^{N} m_n}{(N-M) \sum_{n=1}^{M} m_n} \).

Defining three SFs as \( SF_1 = (\rho - 1)^2 \), \( SF_2 = |\ln(\rho)| \) and \( SF_3 = (\rho^{M/N} + \frac{M \rho^{N-M+1}}{N-M}) - (1 + \frac{M}{N-M}) \), we have three SFET-CFARs as, \( (\hat{\rho} - 1)^2 \overset{\mathcal{H}_0}{\gtrless} \eta_{SF_1} \), \( |\ln(\hat{\rho})| \overset{\mathcal{H}_0}{\gtrless} \eta_{SF_2} \) and \( SF_3(\hat{\rho}) \overset{\mathcal{H}_0}{\gtrless} \eta_{\text{GLR}} \) where \( \eta_{SF_1}, \eta_{SF_2} \) and \( \eta_{SF_3} \) are the thresholds for satisfying the required false alarm. Interestingly, it turns out that SFET-CFAR_3 is the GLRT-CFAR. Figure 4.5 depicts the probability of detection versus the probability of false alarm for \( \rho = 2, N = 100 \) and \( M = 30 \). It is seen that, the performance of CFAR SFET_1 is so close the optimal CFAR bound of MP-CFAR. We conclude that by using the SFET after reducing the number of unknown parameters from three to one, the performance
Figure 4.5: Probability of detection versus probability of false alarm for MP-CFAR, CFAR SFET$_1$ CFAR SFET$_2$ and CFAR GLRT for Example 18 for $\rho = 2$, $N = 100$ and $M = 30$.

of detectors (e.g., GLRT in this example ) is improved significantly while the CFAR property is preserved.
Chapter 5

Asymptotically Optimal CFAR Test

The results of this chapter are published in [89, 90]. The main contributions of this chapter are as follows:

We consider the maximal invariant with respect to MIG to remove the nuisance parameters. The resulting reduced problem has unknown parameters $\rho$ only under $\mathcal{H}_1$. We first show that the MLE of $\rho$ is invariant with respect to the MIG. As a consequence, any SFET using the MLE of $\rho$ is CFAR. Then, we propose GLRT-CFAR, Wald-CFAR, and Rao-CFAR as the suboptimal CFAR tests by applying the concept of the GLRT, the Wald and the Rao tests on the reduced problem and study their asymptotical behaviour using SFET concept.

Motivated by invariance of the MLE, we consider the class of SFETs using the MLE of $\rho$ and show that, the Wald-CFAR is asymptotically optimal for identically independent distribution (iid) observations. Moreover, we prove that the Rao-CFAR and the GLRT-CFAR are asymptotically SFET under some mild conditions which are satisfied in most signal processing problems. This means that the GLRT-CFAR, Wald-CFAR and Rao-CFAR tests are asymptotically optimal under those conditions.
By the performance gap amongst these AOSFETs, we are motivated to find the most powerful SFET using MLE which we refer to it as AOSFET. To derive an AOSFET, we express the asymptotical Probability of Detection ($P_d$) in terms of the SF $g(\rho)$, $\frac{\partial g(\rho)}{\partial \rho}$ and the FIM $I(\rho)$ of the family of distributions after reduction. We propose a novel systematic method to find the optimal SF by maximizing this $P_d$. This maximization results in an SF which is the Euclidean distance of a transformation of unknown parameters under two hypotheses. This gradient of this transformation is the Cholesky decomposition of the FIM $I(\rho)$. This AOSFET asymptotically outperforms any other SFET. As a confirmation, we also maximize a tight lower bound expression of $P_d$ which results in the same AOSFET.

Interestingly, this AOSFET simplifies to the Wald-CFAR wherever the FIM does not depend on the unknown parameters. Moreover, the pdf of $m$ is not required in the derivation of the AOSFET. The simulation results reveal that the proposed AOSFET outperforms the Wald-CFAR. Moreover, our results are applicable without the MIG reduction to any composite problem without unknown parameters under the null hypothesis. For such problems, the AOSFET approach is directly applicable without the MIG reduction and the GLRT-CFAR, Wald-CFAR and Rao-CFAR tests are the regular the GLRT, the Wald and the Rao tests respectively.

This chapter is organized as follows. In Section 5.1, the asymptotic performance of the GLRT-CFAR, Wald-CFAR and Rao-CFAR test are discussed and it is proven that they are asymptotically SFET. Section 5.2 introduces a lower bound for the $P_d$ of SFETs. As an example, this bound leads to a novel lower bound for the $P_d$ the NP in simple problems. In Section 5.2, a method is proposed to derive the optimal SF which maximizes the asymptotic $P_d$. We propose to use the MLE estimate of this optimal
SF as AOSFET in Section 5.2 which asymptotically outperforms all SFETs using MLE. Section 5.3 provides four examples to evaluate and compare the performance of these tests with the optimal MP-CFAR bound.

5.1 GLRT-CFAR, Wald-CFAR and Rao-CFAR

The following proposition shows that the MLE of unknown induced/reduced unknowns $\rho(\theta)$ gives us an AOSFET.

**Proposition 3** For any $q_m \in Q_m$, we have $\rho(\hat{\theta}(q_m(x))) = \rho(\hat{\theta}(x))$ where $\hat{\theta}(x)$ is the MLE of unknown parameters and $Q_m$ is the MIG the family of distributions of $x$.

**Proof 17** See Appendix C.1

This means that the MLE of $\rho$ using $x$ is invariant with respect to the MIG. Clearly the MLE of $\rho$ using $m$ is also invariant because $m$ is invariant. Thus hereafter, we aim to find a SF using $\rho$ which provides a SFET-CFAR using MLE.

Although the asymptotically optimality of SFETs is proven, no similar theorem exists for the well-known suboptimal detectors Rao, Wald and GLRT under such a mild conditions. In the following, we consider these tests for the reduced hypothesis (4.1) and prove that under some conditions the Wald test is an SFET; thereby is asymptotically optimal. Then, similar property for Rao and GLRT is proven.
The Rao-CFAR, the Wald-CFAR and the GLRT-CFAR are defined by

\[
\begin{align*}
\text{Rao} & \quad \frac{\partial \ln(f_m(m, \rho))}{\partial \rho}\bigg|_{\rho = \rho_0}^T \left[ \mathbf{I}^{-1}(\rho_0) \frac{\partial \ln(f_m(m, \rho))}{\partial \rho}\bigg|_{\rho = \rho_0} \right] \gtrless \eta_R, \\
\text{Wald} & \quad (\hat{\rho}_1 - \rho_0)^T \mathbf{I}(\hat{\rho}_1)(\hat{\rho}_1 - \rho_0) \gtrless \eta_W, \\
\text{GLRT} & \quad \frac{\sup_{\rho \in \Omega} f_m(m, \rho)}{f_m(m, \rho_0)} \gtrless \frac{\eta_1}{\eta_0} \eta_G,
\end{align*}
\]  

where \(T\) is the transposition, \(\mathbf{I}(\rho)\) is the Fisher Information Matrix (FIM) with respect to \(\rho\), \(\hat{\rho}_1 = \arg\max_{\rho \in \Omega} f_m(m; \rho)\) is the MLE of \(\rho\) under \(\mathcal{H}_1\) and the thresholds \(\eta_R, \eta_W\) and \(\eta_G\) are set to satisfy the required probability of false alarm. These tests are CFAR as they are invariant with respect to MIG \(Q_m\), since their decision statistics depend only on \(m\).

Following proposition proves that the Wald-CFAR test is an SFET under some mild sufficient conditions (which are satisfied in most signal processing problems).

**Proposition 4** The Wald-CFAR test in (5.1b) for (2.1) is an SFET using the MLE, if \(\mathbf{I}(\rho)(\rho - \rho_0) = 0\) has a unique solution at \(\rho = \rho_0\) and \(\hat{\rho}\) is a continuous random vector, where \(\hat{\rho} = \arg\max_{\rho \in \Omega \cup \{\rho_0\}} f_m(m; \rho)\).

**Proof** See Appendix C.2

Under these mild conditions, the Wald-CFAR is SFET using MLE; thus it is asymptotically optimal by virtue of Theorem 7 in [75].

The relationship between GLRT, Rao and Wald tests are studied in [42] and interesting necessary and sufficient conditions are proposed for them to be equivalent for problems without nuisance parameters, i.e., for (4.1). The following proposition combines Theorem 5 of [42] with Proposition 4 and shows that under the following
conditions the tests in (5.1) are all SFETs using MLE.

**Proposition 5** The Rao and GLRT are SFETs using MLE under the conditions in Proposition 4, where

\[ f_m(m; \rho) = \exp \left( (\rho - \rho_0)^T t(m) - k(\rho - \rho_0) + \ln(f_m(m; \rho_0)) \right) \]

\[ t(m) = I(\rho)(\hat{\rho}_1 - \rho_0) \quad \text{and} \quad k(\rho) = \frac{1}{2} \rho^T I(\rho) \rho, \text{ for all } \rho \in \Omega. \]

The following remarks their asymptotic optimality.

**Remark 9** The SFET property of the Rao-CFAR test and the GLRT-CFAR under the conditions of Proposition 5 guarantees their asymptotic optimality. Moreover Proposition 4 easily shows that the Rao-CFAR test and the GLRT-CFAR under these conditions are asymptotically optimal. This is because of these tests are asymptotically equivalent, i.e., for (4.1), the Rao-CFAR and the Wald-CFAR tests tend to

\[ 2 \ln(T_{GLR}(m)), \text{ where } T_{GLR}(m) \text{ is the GLRT-CFAR statistic given by the left hand side of (5.1c). Under the assumptions of Proposition 3, the Wald-CFAR test is asymptotically optimal and since Rao-CFAR, Wald-CFAR and GLRT-CFAR are asymptotically SFET. Thus, they are all asymptotically optimal.} \]

We have so far shown that these tests converge to a similar SFET using MLE, under some mild conditions. We know that for any SF the SFET using MVUE performs as the optimal NP test [75]. The MLE is a reliable replacement for the MVUE for many signal processing problems in which the MVUE of unknown parameters does not exist. Proposition 4, Proposition 5 and Remark 9 state that the Wald-CFAR is an SFET. Moreover, Rao-CFAR and GLRT-CFAR asymptotically perform similar to the Wald-CFAR which is an SFET. Hence finding an optimal SF and applying SFET using MLE provides a detector which outperforms Wald-CFAR test and asymptotically outperforms the Rao-CFAR the and GLRT-CFAR.
5.2 Asymptotically Optimal SFET (AOSFET)

We aim to find an asymptotically optimal SF by maximizing the $P_d$ of the resulting SFET using MLE. Unfortunately, the $P_d$ does not have a closed form expression in terms of SF. Thus we maximize the asymptotic expression of the $P_d$ with respect to SF to find an (asymptotically) optimal SF. This leads to generalize the Wald-CFAR test and improve the $P_d$ while the CFAR property is preserved.

In this section, we derive a lower bound for $P_d$ of an SFET in terms of the mean and the variance of SFET statistic $g(\hat{\rho})$. For a simple hypothesis testing problem, this gives a new lower bound for NP test. Finally, we propose a method to maximize the $P_d$ (or the lower bound) with respect to the SF when the number of observations tends to infinity. Using the MLE of this AOSF, the AOSFET is achieved which is CFAR as guaranteed by Proposition 3. We show that this AOSFET generalizes the Wald test, and for the case that FIM does not depend on the unknown parameters is equal to the Wald test.

5.2.1 A tight Lower Bound for Detection Probability

Consider an SFET for (4.1) given by $g(\hat{\rho}) \gtrless_{\mathcal{H}_1}^{H_0} \eta$, where $\eta$ is set to satisfy the probability of false alarm. Using Chernoff inequality, a lower bound for the $P_d$ is $P_d(\rho) = \Pr_{\rho \neq \rho_0}(g(\hat{\rho}) > \eta) \geq 1 - e^{s\eta}E_{\rho \neq \rho_0}(e^{-sg(\hat{\rho})})$, where $\Pr_{\rho \neq \rho_0}(\cdot)$ and $E_{\rho \neq \rho_0}(\cdot)$ are the probability and the expectation operators under $\mathcal{H}_1$ and $s > 0$. From Theorem 7.2 of [23], $g(\hat{\rho})$ is the MLE of $g(\rho)$, then $E_{\rho \neq \rho_0}(e^{-sg(\hat{\rho})})$ asymptotically tends to $\exp(-sg(\rho) + \frac{s^2}{2}I_{g(\rho)}^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho} - (g(\rho) - \eta)s)$ at
5.2. ASYMPTOTICALLY OPTIMAL SFET (AOSFET)

\[ s = \frac{g(\rho) - \eta}{\frac{\partial g(\rho)}{\partial \rho}} \frac{1}{I^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho}}, \] for all \( \eta < g(\rho) \) gives

\[ P_d(\rho) \geq 1 - \exp\left( \frac{-(g(\rho) - \eta)^2}{2 \frac{\partial g(\rho)}{\partial \rho} I^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho}} \right), \] (5.2)

The bound in (5.2) is expressed in terms of the SF and the inverse of FIM. We must also note that for the SFETs using the MLE, \( g(\hat{\rho}) \) is a consistent estimator and asymptotically tends to normal as the number of independent observations increases.

Thus, the false alarm probability \( P_{fa} \) and the \( P_d \) asymptotically tend to \( P_{fa} = Q(\frac{\eta - g(\rho_0)}{\sqrt{\frac{\partial g(\rho_0)}{\partial \rho_0} I^{-1}(\rho_0) \frac{\partial g(\rho_0)}{\partial \rho_0}}}) \), and \( P_d(\rho) = Q(\frac{\eta - g(\rho)}{\sqrt{\frac{\partial g(\rho)}{\partial \rho} I^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho}}}) \), respectively, where \( Q(\cdot) \) is the \( Q \) function. Note that \( g(\rho) \leq \eta \) results in \( P_d \geq \frac{1}{2} \). The bound in 5.2 is derived for any composite hypothesis testing problem and for any SFET including Wald, Rao and GLRT tests (see Remark 9 and Proposition 4). Since the Wald, Rao and GLRT tests are asymptotically SFETs. In Appendix C.2, the SF for the Wald test is found to be \( g_W(\rho) \triangleq (\rho - \rho_0)^T I(\rho)(\rho - \rho_0) \). Thus, the \( P_d \) lower bound for the Wald test is \( 1 - \exp\left( -\frac{(g_W(\rho) - \eta)^2}{2g_W(\rho)} \right) \).

The lower bound (5.2) is given for a composite hypothesis testing problem. In Example 19, we use (5.2) and derive a new bound for the \( P_d \) of the NP test for simple hypothesis testing problems using Theorem 1 of [75].

**Example 19 (Simple Hypothesis Testing Problem)** Consider a simple hypothesis testing problem [9]

\[
\begin{align*}
\mathcal{H}_1 : x &\sim f_x(x|\mathcal{H}_1), \\
\mathcal{H}_0 : x &\sim f_x(x|\mathcal{H}_0),
\end{align*}
\] (5.3)
where \( f_x(x|H_1) \) and \( f_x(x|H_0) \) are the pdf of observations \( x \) with no unknown parameter, under \( H_1 \) and \( H_0 \) respectively. Consider the Exponentially Embedded Family (EEF) proposed in [91, 92] generated by \( f_x(x|H_1) \) and \( f_x(x|H_0) \) as follows

\[
P_{\gamma} \triangleq \left\{ f_x(x; \gamma) = \exp \left( \gamma \ln(f_x(x|H_1)) + (1 - \gamma) \ln(f_x(x|H_0)) - k(\gamma) \right) \bigg| 0 \leq \gamma \leq 1 \right\}, \tag{5.4}
\]

where \( k(\gamma) = \ln(\int_{\mathbb{R}^N} f_x(x|H_1)^\gamma f_x(x|H_0)^{1-\gamma} dx) \), \( dx \) is the lebesgue measure of \( \mathbb{R}^N \) and \( N \) is the dimension of observations. A composit problem can be defined as

\[
\begin{align*}
\tilde{H}_0 : & \gamma \leq \frac{1}{2}, \\
\tilde{H}_1 : & \gamma > \frac{1}{2}.
\end{align*}
\tag{5.5}
\]

The simple problem (5.3) is a sub-problem of (5.5), since \( f_x(x; \gamma) \) is equal to \( f_x(x|H_0) \) and \( f_x(x|H_1) \), respectively at \( \gamma = 0 \) and at \( \gamma = 1 \). We first prove that the NP test for (5.3) is an SFET for (5.5). Then, we use (5.2) to provide a new lower bound for the \( P_d \) of the NP test.

**Proposition 6** The NP test of (5.3), i.e., \( \frac{f_x(x|H_1)}{f_x(x|H_0)} \sim_{H_0} \eta \) is an SFET for (5.5) using \( g_{NP}(\gamma) = \gamma - \frac{1}{2} \) as the SF and the MLE as the estimator.

**Proof 19** See Appendix C.3.

The FIM of \( \gamma \) is \( \frac{\partial^2 k(\gamma)}{\partial \gamma^2} \) which in Appendix C.3 is obtained to be the variance of the log-likelihood function under \( f_x(x; \gamma) \). Hence using (5.2), the lower bound for the NP test using SF \( \gamma \) is given by \( P_d(\gamma) \geq 1 - \exp\left(\frac{-(\gamma - \eta)^2}{2\frac{\partial^2 k(\gamma)}{\partial \gamma^2}}\right) \). Replacing \( \gamma = 1 \), the new
5.2. ASYMPTOTICALLY OPTIMAL SFET (AOSFET)

Asymptotic lower bound of the $P_d$ is

$$P_d = Q\left(\frac{\eta' - 1}{\sqrt{\text{var}_1(\alpha)}}\right) \geq 1 - \exp\left(-\frac{(1 - \eta')^2}{2\text{var}_1(\alpha)}\right)$$  \hspace{1cm} (5.6)$$

where $\eta'$ is the solution of $\exp(\eta) = \frac{\partial k(\eta')}{\partial \eta}$, $\alpha = \ln \frac{f_x(x|\mathcal{H}_1)}{f_x(x|\mathcal{H}_0)}$ and

$$\text{var}_1(\alpha) = \int_{\mathbb{R}^N} \alpha^2 f_x(x|\mathcal{H}_1)dx - \left(\int_{\mathbb{R}^N} \alpha f_x(x|\mathcal{H}_1)dx\right)^2$$

is the variance of the log-LR under $\mathcal{H}_1$.

5.2.2 Asymptotically Optimal SF

Here, we intend to find the SF which asymptotically maximizes the $P_d$. The following remark shows that for any SFET using the MLE, maximizing the proposed bound in (5.6) also maximizes the asymptotic $P_d$.

**Remark 10** The invariance property of the MLE [23] implies that $g(\hat{\rho}) = \hat{g}(\tilde{\rho})$. Moreover, $P_d(\rho) = \Pr_\rho(g(\hat{\rho}) > \eta)$ as the probability of detection of an SF using the MLE asymptotically tends to $Q\left(\frac{(\eta - g(\rho))^2}{I_\alpha^{-1}(\rho)}\right)$, as the number of observations increase. This is because the MLE of $\rho$ asymptotically tends to a Gaussian random variable with mean $g(\rho)$ and variance $I_\alpha^{-1}(\rho)$. Hence to maximize $P_d(\rho)$, we need to minimize $\frac{(\eta - g(\rho))^2}{I_\alpha^{-1}(\rho)}$ with respect to $g$. Maximizing the bound in (5.2) with respect to $g$ is equivalent to maximize $\frac{(g(\rho) - \eta)^2}{I_\alpha^{-1}(\rho)}$. Thus, under the necessary condition $g(\theta) > \eta$ in deriving (5.2), these maximizing are equivalent.

Thus we refers to the SF obtained by maximizing of (5.2) with respect to $g$ as the AOSF and the resulting SFET is referred to as AOSFET.
We show that the Wald test and the AOSFET are equivalent if the FIM does not depend on the unknown parameters, i.e., the AOSFET extends the Wald test.

We have $g(\rho_0) = 0$ since $g$ is an SF for (4.1). Hence we need to optimize $g$ for $\rho \neq \rho_0$. The following constrained optimization problem leads to the AOSFET

$$\max_g 1 - \exp\left(\frac{-(g(\rho) - \eta)^2}{2 \frac{\partial g(\rho)}{\partial \rho}^T \mathbf{I}^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho}}\right), \text{ s.t., } g(\rho) > \eta > 0, \quad g(\rho_0) = 0, \quad \text{for all } \rho \neq \rho_0. \quad (5.7)$$

In this problem the first constraint is given by the derivation of the lower bound in (5.2). The rest of constrains provide the SF property of $g$. To avoid the constraints, we maximize asymptotical expression of $P_d$ which is equivalent to

$$\min_g \sqrt{\frac{\partial g(\rho)}{\partial \rho}^T \mathbf{I}^{-1}(\rho) \frac{\partial g(\rho)}{\partial \rho}} \frac{\partial g(\rho)}{\partial \rho}, \quad \text{s.t., } g(\rho) > 0, \quad g(\rho_0) = 0. \quad (5.8)$$

Any increasing function of an optimal solution of (5.7) is also an optimal solution of (5.7) as their contours remain the same. Thus, the optimal solution is not unique, however, all optimal solutions have identical contours. In other words, we need to find only one of solutions.

To solve (5.8) for the general case, we first solve it for the simple scenario where the FIM is the identity matrix. We will then extend this solution in two steps. First, we extend it to the case where the FIM is any arbitrary constant non-negative semi-definite matrix. In the second step, we extend the results to the case where $\mathbf{I}(\rho)$ is any arbitrary FIM.
First case; \( I(\rho) = \mathbb{I} \)

Consider the simple case where the FIM is an identity matrix \( I(\rho) = \mathbb{I} \). In this case the problem simplifies to

\[
\min_{g} \frac{\| \frac{\partial g(\rho)}{\partial \rho} \|}{\eta - g(\rho)}, \quad \text{s.t., } g(\rho) > 0, \quad g(\rho_0) = 0,
\]

(5.9)

where \( \| \cdot \| \) is the \( L_2 \) norm. To find the contours of \( g(\rho) \), we consider \( \rho \) such that \( g(\rho) \) is constant and evaluate \( g(\rho) \) at \( \varrho = U(\rho - \rho_0) \) where \( U \) is an arbitrary unitary transformation, i.e., \( U^T U = \mathbb{I} \). For the transformed induced maximal invariant \( \varrho \) we have \( \| \frac{\partial g(\varrho)}{\partial \varrho} \| = \| \frac{\partial g(\rho)}{\partial \rho} \| \) and \( \frac{\partial g(\rho)}{\partial \rho} = U^T \frac{\partial g(\varrho)}{\partial \varrho} \). This means that the problem in (5.9) is invariant with respect to the group of unitary transformations, i.e., \( g(\varrho) = g(U(\rho - \rho_0)) = g(\rho) \) for all unitary transformations \( U \). Thus \( g(\rho) \) depends on \( \rho \) only via the maximal invariant of the group which is \( \| \rho - \rho_0 \|^2 \), i.e., \( g(\rho) \) is a function of \( \| \rho - \rho_0 \|^2 \). Therefore the SF \( g(\rho) = \| \rho - \rho_0 \|^2 \) is an optimal solution of (5.9), i.e., the contours of the solution of (5.9) are symmetrical hyperspheres and are identical with those of \( \| \rho - \rho_0 \|^2 \).

Second case; \( I(\rho) \) is constant

In this part, we solve (5.8) using the result of the previous case in more general case where the FIM constant \( I(\rho) = F \) with respect to \( \rho \) where \( F \) is any arbitrary constant non-negative semi-definite matrix. The idea is simply to apply a linear transform on \( \rho \) such that the problem is converted to the canonical form of (5.9). This transformation is defined by \( \varrho \triangleq F^{\frac{1}{2}}(\rho - \rho_0) \), where \( F^{\frac{1}{2}} = U \Gamma U^T \) is a symmetric square root of \( F \) where \( F = U \Gamma U^T \) is the eigenvalue decomposition, such that \( \Gamma \) is a diagonal
matrix and \( U \) is unitary. The chain rule and the symmetric property of \( F^{\frac{1}{2}} \) lead to 
\[
\frac{\partial g(\rho)}{\partial \rho} = F^{\frac{1}{2}} \frac{\partial g(\rho)}{\partial \varrho}.
\]
Replacing this result in (5.8), the problem is converted to the form in (5.9) with respect to \( \varrho \) instead of \( \rho \). Thus, the results of the previous case imply that the optimal solution is 
\[
g(\rho) = \|\varrho\|_2 = (\rho - \rho_0)\Gamma_0 \eta_{AO}.
\]
Hence for this case AOSFET is given by 
\[
(\hat{\rho} - \rho_0)^T F(\hat{\rho} - \rho_0) \geq \eta_{AO},
\]
where \( \eta_{AO} \) is set to satisfy the false alarm probability. Interestingly, in this case \( I(\rho) \) does not depend on \( \rho \) the AOSFET simplifies to the Wald test.

**General form of FIM**

Similar to the previous case, the approach to solve (5.8) is to apply a similar transform on \( \rho \) such that the problem is converted to the canonical form of (5.9). Unlike the previous part this transformation is not constant and shall be a function of \( \rho \). Let \( \varrho(\rho) \) denote this transformation which must be differentiable with respect to \( \rho \). The function \( \varrho(\rho) \) is selected such that the problem is converted to the canonical form of (5.9) in terms of \( \varrho \) which is already solved in the first case, i.e., where \( I(\varrho) = \tilde{I}_M \).

Thus, we derive a set of equations to find \( \varrho(\rho) \). We have 
\[
\frac{\partial g(\rho)}{\partial \rho} = \frac{\partial \varrho(\rho)}{\partial \varrho} \frac{\partial g(\rho)}{\partial \varrho},
\]
where \( \frac{\partial g(\rho)}{\partial \rho} \) and \( \frac{\partial \varrho(\rho)}{\partial \varrho} \) are \( \tilde{M} \times 1 \) vectors, the \( m \)th element of them are \( \left[\frac{\partial g(\rho)}{\partial \rho}\right]_m = \frac{\partial g(\rho)}{\partial \varrho_m} \) and \( \left[\frac{\partial \varrho(\rho)}{\partial \varrho}\right]_m = \frac{\partial \varrho(\rho)}{\partial \varrho_m} \), respectively where \( \varrho_m \) is the \( m \)th element of \( \varrho \) and \( \frac{\partial \varrho(\rho)}{\partial \varrho} \) is an \( \tilde{M} \times \tilde{M} \) matrix with \( \left[\frac{\partial \varrho(\rho)}{\partial \varrho}\right]_{m,m'} = \frac{\partial \varrho(\rho)}{\partial \varrho_{m'}} \), where \( \varrho(\rho) \) is the \( m \)th element of \( \varrho(\rho) \). Thus, to convert (5.8) to (5.9) in terms of \( \varrho \), \( \varrho(\rho) \) must satisfy

\[
(\frac{\partial \varrho(\rho)}{\partial \rho})^T I^{-1}(\rho) \frac{\partial \varrho(\rho)}{\partial \rho} = \tilde{I}_M.
\] 

(5.10)

The partial differential equation (5.10) with respect to \( \rho \) must be solved to find \( \varrho(\rho) \). Since \( I^{-1}(\rho) \) is a symmetric positive definite matrix it has a the Cholesky
decomposition as \( I(\rho) = L(\rho)L(\rho)^T \), where \( L(\rho) \) is a lower triangular matrix. Thus we can rewrite (5.10) as \( (\frac{\partial \varrho(\rho)}{\partial \rho})^T L^{-T}(\rho)L^{-1}(\rho)\frac{\partial \varrho(\rho)}{\partial \rho} = \mathbb{I} \). Therefore, the solution to the following partial differential equation converts our problem to the canonical form

\[
\frac{\partial g(\rho)}{\partial \rho} = L(\rho). \tag{5.11}
\]

Other decompositions of \( I^{-1}(\rho) \) may lead to possibly different solutions. For example any constant unitary transformation of the solution of (5.11) is also a solution of (5.10). However all these different solutions have identical contours and thereby they result in either identical or equivalent AOSEFTs. The significance of the proposed approach is that (5.11) can be systematically driven and solved for most signal processing applications allowing to find the AOSF as

\[
g(\rho) = \| \varrho - \varrho_0 \|^2 = \| \varrho(\rho) - \varrho(\rho_0) \|^2. \tag{5.12}
\]

This optimal SF becomes the SF of the Wald wherever \( I(\rho) \) is constant with respect to \( \rho \). In such case, the Wald test asymptotically outperform other SFETs. In general case, the above AOSFET can be viewed as an extension of the Wald outperform all SFETs using the MLE.

One way to calculate \( I(\rho) \) is to apply reduction using the MIG and employing the distribution of \( m \) given the induced maximal invariant unknown \( \rho \). However, finding and employing this distribution is not always an easy task. Alternatively, \( I(\rho) \) can be calculated using [23, Section 2.8]

\[
I(\rho) = J^T I_\theta(\theta) J, \tag{5.13}
\]
5.3. Examples

where $I_\theta$ is the FIM with respect to $\theta$ which can be directly calculated from the original family of distributions and the elements of $J$ are given by $[J]_{i,j} = \frac{\partial \theta_i}{\partial \rho_j}$ and $\theta_i$ and $\rho_j$ are the $i$th and $j$th elements of $\theta$ and $\rho$ respectively. Note that in one dimensional problem, when the dimension of unknown vector or induced maximal invariant is one, we need to replace Fisher information function instead of FIM in all relations. Hence in such cases, we denote Fisher information function of unknown parameter and induced maximal invariant as $I_\theta$ and $I$ and similarly $\theta$, $\rho$ and $\varrho$ as unknown parameter, induced maximal invariant and the transformed induced maximal invariant, respectively.

5.3 Examples

This section illustrates our theoretical results in four examples and compare the AOS-FET with other suboptimal tests.

Example 20 In this example, a signal detection in Gaussian noise is considered. Assume that the observation vector is $x = av + w$, where $v \in \mathbb{C}^L$ is assumed to be a known complex vector and $w$ is a zero mean circularly symmetrical Gaussian noise with known covariance matrix $\sigma^2 I_2$, where $\mathbb{C}$ is the set of complex numbers. Let denote the unknown vector by $\theta = [a_R, a_I]^T$ where $a_R$ and $a_I$ are the real and the imaginary part of $a$, respectively. Hence the signal detection problem is modeled by (2.1), where $\Theta_0 = \{[0, 0]^T\}$ and $\Theta_1 = \mathbb{R}^2 - \Theta_0$, and $\mathbb{R}$ is the set of real numbers. Since the pdf under $\mathcal{H}_0$ does not have any unknown parameter, we apply the results directly to the pdf of $x$. The FIM for $\theta$ is given by $I = \frac{\|v\|^2}{\sigma^2} I_2$ which does not depend on the unknown parameters. Thus, the result of Subsection 5.2.2 implies that the AOSF is given by $\theta^T I \theta = \frac{\|v\|^2}{\sigma^2}$. Thus the AOSFET is given by $\|v^H x\|^2 \geq_{\mathcal{H}_0} H^1 \eta$, where $\eta$ is set
to satisfy the false alarm probability. In this example the FIM is constant and the results match the results of Subsection 5.2.2 as the GLRT, the Wald test and the Rao test are equivalent with the AOSFET.

**Example 21 (Detection in Unknown Noise Variance)** Consider the previous example problem, while in this problem \( w \) is assumed to be a white noise with unknown noise variance \( \sigma^2 \). Then the unknown parameter vector is \( \theta = [a_R, a_I, \sigma]^T \). The MIG for this problem is \( Q_m = \{ q_m | q_m(x) = \xi x, \xi > 0 \} \). Hence the induced MIG is given by \( \overline{Q_m} = \{ q_m(\theta) = \xi \theta, \xi > 0 \} \), so the induced maximal invariant with respect to MIG is given by \( m = [x_1 | x_N, \ldots, x_N | x_N]^T \), where \( x_n \)'s are the elements of \( x \). Using (5.13), we have \( I(\rho) = (2\|v\|^2)_2 \). Since \( I(\rho) \) does not depend on the unknown parameters and \( \rho_0 = [0, 0]^T \), using Section 5.2.2, the AOSF is given by \( \rho^T I(\rho) \rho = 2\|v\|^2(\rho_1^2 + \rho_2^2) \). Since \( 2\|v\|^2 \) is positive and does not depend on the unknown parameters, the AOSFET is given by \( \hat{\rho}_1^2 + \hat{\rho}_2^2 \geq_{H_0} \eta \), where \( \eta \) is set to the false alarm requirement. The MLE of \( \rho \) is \( [\text{Re}(\frac{v^H m}{|m|}), \text{Im}(\frac{v^H m}{|m|})]^T \), where \( \text{Re}(\cdot) \) and \( \text{Im}(\cdot) \) return the real and imaginary parts respectively. Thus the AOSFET is \( \frac{|v^H m|^2}{|m|^2} \geq_{H_0} \frac{\eta_1}{\eta} \). Since FIM of \( \rho \) does not depend on the unknown parameters the result of Subsection 5.2.2 guarantees that the AOSFET and the Wald test are equivalent. Moreover based on the results of [42], in this problem the Wald test, the Rao test and the GLRT are equivalent.

**Example 22** Consider the pdf of a real vector \( x \) as

\[
f_x(x; \theta, \sigma^2) = \frac{\exp(-\|x - \theta 1_N\|^2/(2(\sigma^2 + \theta^2)))}{(2\pi(\sigma^2 + \theta^2))^{N/2}}.
\]

We first consider the problem \( H_0 : \theta = 0 \), versus \( H_1 : \theta \neq 0 \) assuming known \( \sigma^2 \).
Hence, we have $\Theta_1 = \mathbb{R} - \{0\}$ and $\Theta_0 = \{0\}$. To find the GLRT, we replace the MLE of $\theta$ in to the LR. Setting the derivative of (5.14) with respect to $\theta$ to zero, the MLE of $\theta$ is one of the solutions of

$$\theta^3 + \frac{1}{N}1_N^T x\theta^2 + (2\sigma^2 - \frac{1}{N} \|x\|^2)\theta - \frac{1}{N}1_N^T x\sigma^2 = 0.$$  \hspace{1cm} (5.15)

The MLE of $\theta$ denoted by $\hat{\theta}$, is the real solution of (5.15) for which (5.14) is maximum. Thus the GLRT is given by

$$\left(\frac{\sigma^2}{\sigma^2 + \hat{\theta}^2}\right)^{N/2} \exp \left(\frac{\|x - \hat{\theta}1_N\|^2}{2(\sigma^2 + \hat{\theta}^2)} - \frac{\|x\|^2}{2\sigma^2}\right) \frac{\eta_{GLRT}}{\eta_{H_0}},$$  \hspace{1cm} (5.16)

where $\eta_{GLRT}$ is set to satisfy the false alarm requirement. In the following the AOSFET for this problem is proposed.

This problem has no unknown parameters under $H_0$, hence any test for this problem is CFAR. Thus, we do not need to apply a maximal invariant of MIG for this problem. Thus using the result of Subsection 5.2.2, $\theta$ plays the role of $\rho$ in this problem. The fisher information function with respect to $\theta$ is $I(\theta) = \frac{N}{\sigma^4 + \theta^4} + \frac{4N\theta^2}{(\sigma^2 + \theta^2)^6}$. Hence, using (5.11), we have

$$\frac{\partial \theta}{\partial \theta} = \sqrt{I(\theta)} = \sqrt{\frac{N}{\sigma^4 + \theta^4} + \frac{4N\theta^2}{(\sigma^2 + \theta^2)^6}}.$$  \hspace{1cm} (5.17)

To find AOSFET using (5.12), it is not easy to find the integral of (5.17). However, we prove that $g(\theta) = \|\varphi(\theta) - \varphi(0)\|^2$ is an increasing function of $|\theta|$ and therefore $|\hat{\theta}|$ gives the decision statistic for AOSFET. Since $\frac{\partial \varphi}{\partial \theta}$ even and positive, $\varphi(\theta) - \varphi(0)$
is continuous, odd and increasing in \( \theta \). Thus \( g(\theta) = \| g(\theta) - g(0) \|^2 \) is even and increasing in \( \| \theta \| \). From (5.12), the AOSFET rejects \( \mathcal{H}_0 \) if \( g(\hat{\theta}) = \| g(\hat{\theta}) - g(0) \|^2 > \eta \), or equivalently if

\[
\left| \hat{\theta} \right|_{\mathcal{H}_0} \geq \eta'.
\]  

(5.18)

The optimal bound for (5.16) and (5.18) is the NP test which is given by comparing the log-LR as follows

\[
\frac{\| x \|^2}{2\theta^2} - \frac{\| x - \theta_1 N \|^2}{2(\sigma^2 + \theta^2)} + \frac{N}{2} \ln \frac{\sigma^2}{\sigma^2 + \theta^2} \begin{cases} \frac{\| x \|^2}{\mathcal{H}_0} & \text{or} \eta'_{\text{NP}}, \\ \frac{\| x \|^2}{\mathcal{H}_1} & \text{or} \eta'_{\text{NP}}, 
\end{cases}
\]  

(5.19)

where \( \eta'_{\text{NP}} \) is given by the false alarm requirement. Since \( \theta \) is unknown this test is not practical. The NP test provides an upper performance bound in our simulations using the true value of \( \theta \). Figure 5.1 compares the probability of miss detection of the GLRT, the AOSFET and the NP test versus the false alarm for \( N = 100, \theta = 0.5 \) and \( \sigma^2 = 1 \). This simulation shows that the AOSFET outperforms the GLRT.

Now, consider the case where \( \sigma^2 \) is also unknown. The GLRT is given by replacing the MLEs of \( [\theta, \sigma^2]^T \) under \( \mathcal{H}_1 \) and \( \mathcal{H}_0 \) into the LR. Setting the derivatives of pdfs at zero, the MLE of \( \theta \) under \( \mathcal{H}_1 \) is one of three solutions of

\[
3N^2\theta^3 - 3N \theta N^T x \theta^2 + 2(\theta N^T x)^2 \theta - \| x \|^2 \theta N^T x = 0.
\]  

(5.20)

Denoting this solution by \( \hat{\theta}_1 \), the MLE of \( \sigma^2 \) under \( \mathcal{H}_1 \) is \( \hat{\sigma}_1^2 = \max(0, \frac{1}{N} \| x - \hat{\theta}_1 N \|^2 - \hat{\theta}_1^2) \). The MLEs are selected between three possible pair of \((\hat{\theta}_1, \hat{\sigma}_1^2)\)s for which the pdf is maximum. The MLE of \( \sigma^2 \) under \( \mathcal{H}_0 \) denoted \( \hat{\sigma}_0^2 \) is \( \frac{1}{N} \| x \|^2 \). Hence the GLRT in
this case is given by
\[
\left( \frac{\hat{\sigma}_0^2}{\sigma_1^2 + \theta_1^2} \right)^{N/2} \frac{H_1}{H_0} \geq \eta_{\text{GLRT}},
\]
(5.21)
where \( \eta_{\text{GLRT}} \) provides the desired false alarm probability.

To derive the AOSFET for the case of unknown \( \sigma^2 \), we calculated the FIM as:
\[
I(\theta) = \frac{N}{(\sigma^2 + \theta^2)^2} \begin{bmatrix} \frac{3\theta^2 + \sigma^2}{\theta} & \theta^2 \\ \theta & 1/2 \end{bmatrix}
\] (5.22)
where \( \theta = [\theta, \sigma^2]^T \). The MIG for this problem is the scale group \( Q = \{ q|x(\xi x, \xi > 0) \}; \) thus the induced group of transformations under the union of hypotheses is \( \overline{Q} = \{ q|q([\theta, \xi^2 \sigma^2], \xi > 0) \}. \) From \( \overline{Q} \), we propose \( \rho = \frac{\theta}{\sigma} \) as the induced maximal invariant for the unknown parameters. Using (5.13) the FIM of \( \rho \) is \( I(\rho) = \frac{1}{6N\rho^6 + \frac{13N}{4}\rho^4 + \frac{13N}{4}\rho^2 + N} \) and the induced maximal invariant under \( H_0 \) denoted by \( \rho_0 \) is zero because of \( \mu = 0 \) under \( H_0 \). From (5.11), we have
\[
\frac{\partial \varphi(\rho)}{\partial \rho} = \frac{1}{\sqrt{6N\rho^6 + \frac{13N}{4}\rho^4 + \frac{13N}{4}\rho^2 + N}}.
\] (5.23)
The right hand side is positive and even thus \( \varphi(\rho) - \varphi(0) \) is odd and increasing in \( \rho \), hence \( g(\rho) = \| \varphi(\rho) - \varphi(0) \|^2 \) is even. The AOSF using (5.12) is \( g(\rho) = \| \varphi(\rho) - \varphi(0) \|^2 \) which is an even function. From increasing property of \( \varphi(\rho) - \varphi(0) \), \( g(\rho) = \| \varphi(\rho) - \varphi(0) \|^2 \) is increasing in \( \rho \) for \( \rho > 0 \), and for \( \rho < 0 \) is increasing in \(-\rho\), i.e., \( g(\rho) = \| \varphi(\rho) - \varphi(0) \|^2 \) is an increasing function of \( |\rho| \). Thus, the AOSFET is given by \( |\hat{\rho}| \geq \frac{H_1}{H_0} \eta \). The MLE of \( \rho \) is the ratio of the MLEs of \( \theta \) and \( \sigma^2 \) under the union of
5.3. EXAMPLES

hypotheses, although, the MLE of $\rho$ is $\hat{\rho} \approx \hat{\theta} \sqrt{\hat{\sigma}^2}$.

Now, we derive the MP-CFAR test as the optimal bound of all CFAR tests using the LR of maximal invariant of MIG. A maximal invariant of MIG is $m = x_{|x_N|}$, where $x_N$ is the $N^{th}$ element of $x$. In Appendix C.4, the pdfs of maximal invariant under both hypothesis are derived which using $1^T m = 1^T x$ the following gives the MP-CFAR as the LR of maximal invariant

$$\exp \left( \frac{(\rho' \frac{1^T x}{||x||})^2}{\Gamma(N/2)} \sum_{i=0}^{N-1} \binom{N-1}{i} \left( \frac{\rho' \frac{1^T x}{||x||}}{||x||} \right)^i \tau(N-i, \frac{\rho' \frac{1^T x}{||x||}}{\sqrt{2} ||x||}) \right) \overset{H_1}{\geq} \overset{H_0}{\eta_{MP}}, \quad (5.24)$$

where $\Gamma(\cdot)$ is Gamma function, and $\rho' = \frac{|\rho|}{\sqrt{1+\rho^2}}$, $\tau(n, x) = \frac{1}{2} \Gamma(n/2, x^2)$ for $x < 0$, $\tau(n, x) = \frac{1}{2} \Gamma(n/2, 0) + \frac{(-1)^{n-1}}{2} (\Gamma(n/2, 0) - \Gamma(n/2, x^2))$ for $x \geq 0$, and $\eta_{MP}$ is the threshold achieving by the false alarm probability. Since this test depends on the unknown parameters, then provides only a most powerful test instead of uniformly most powerful test. In simulation by replacing the true values of unknown parameters in the MP-CFAR tests statistic, an upper bound for evaluating the performance of CFAR tests are given. In fact this bound is a benchmark for evaluating the other CFAR tests and is not practical. Figure 5.2 shows probability of miss detection which is $1 - P_d$ versus false alarm probability for $\theta = 0.5$, $\sigma^2 = 1$ and $N = 200$. Here, the AOSFET also outperforms the GLRT.

Example 23 In this example, we consider a fusion decision making in a spectrum sensing application and find different AOSFETs using different induced maximal invariants $\rho$. Consider $M$ sensors observing independent complex Gaussian random
vectors such that the pdf of received measurements is

\[ f_X(X; \theta) = \pi^{-NM} \prod_{m=1}^{M} \theta_m^{-1} \exp\left(\frac{- \sum_{m=1}^{M} \|x_m\|^2}{\theta_m}\right), \]

where \(x_m\) is an \(N\) dimension received vector from the \(m^{th}\) sensor. The set of possible values for unknown parameters is \(\Theta = \{\theta = [\theta_1, \cdots, \theta_M]^T | \theta_m \geq 0, \forall m \in \{1,2,\cdots,M\}\}.\) Let the detection problem be to test if \(\theta_m\)s are equal or not, i.e., we have \(\Theta_0 = \{\theta = \theta 1_M | \theta \geq 0\}\) and \(\Theta_1 = \Theta - \Theta_0.\)

The MIG in this problem is \(Q = \{q(X) = \xi X | \xi > 0\},\) where \(X\) is the matrix of observations defined as \(X = [x_1, \cdots, x_K].\) The induced maximal invariant for this MIG is \(Q = \{q(\theta) = \xi^2 \theta | \xi > 0\},\) An induced maximal invariants of \(\theta\) with respect to \(Q\) is \(\rho = [\theta_2, \cdots, \theta_M]^T\) which allows us to derive AOSFET\(_1\) as follows.}

The problem after reduction using MIG reduces to (4.1), where \(\rho_0 = 1_{M-1}.\) The FIM with respect to \(\theta\) for this problem is diagonal with \([I(\theta)]_{m,m} = \frac{N}{\theta_m^2}\) as its \(m^{th}\) diagonal element. Using (5.11), the FIM of \(\rho\) is diagonal with \([I(\rho)]_{m,m} = \frac{N}{\rho_m^2}\) as its \(m^{th}\) diagonal element. Since the FIM with respect to \(\rho\) depends on \(\rho,\) the AOSF is found from (5.11), where \(L(\rho)\) is a diagonal matrix given by Cholesky decomposition of \(I(\rho),\) i.e., \([L(\rho)]_{m,m} = \sqrt{\frac{N}{\rho_m}}.\) Thus we have \(\frac{\partial \rho_m}{\partial \rho_m} = \frac{\sqrt{N}}{\rho_m}, \ \frac{\partial \rho_m}{\partial \rho_{m'}} = 0, \ \forall m \neq m',\) where \(c_m\) is the \(m^{th}\) element of \(g(\rho).\) The above leads to \(g_m(\rho) = \sqrt{N} \ln(\rho_m) + c_m,\) where \(c_m\) is constant with respect to \(\rho_m.\) Using (5.12), we find the AOSF using \(\rho\) as \(g(\rho) = \|g(\rho) - g(\rho_0)\|^2 = N \sum_{m=1}^{M-1} (\ln(\rho_m))^2.\) The MLE of \(\rho_m\) is \(\hat{\rho}_m = \frac{\partial \rho_m}{\partial \theta} = \frac{\|x_m\|^2}{\|x_1\|^2}.\) Thus using Proposition 3, the MLE of \(g(\rho)\) gives the AOSFET\(_1\) as follows

\[
N \sum_{m=1}^{M} \left( \ln\left(\frac{\|x_m\|^2}{\|x_1\|^2}\right) \right)^2 \frac{\mathcal{H}_1}{R_0} \geq \eta_{\text{AO}_1},
\]

(5.25)
where $\eta_{AO_1}$ is set to satisfy the false alarm requirement.

Alternatively, we can find $AOSFET_2$ by first transforming $\theta$ to $\tilde{\theta} = \sqrt{N}[\ln(\theta_1), \ldots, \ln(\theta_M)]^T$ which leads to $I_{\tilde{\theta}}(\tilde{\theta})$ to become an identity matrix $I_M$. In this case, it obvious that $g(\tilde{\rho}(\tilde{\theta})) = \|\tilde{\theta} - \tilde{\theta}_0\|$ is an AOSF where $\tilde{\theta}_0$ is the unknown parameter under $H_0$. Hence the AOSFET is given by the MLE of $g(\tilde{\rho}(\tilde{\theta}))$ is as follows

$$M \sum_{m=1}^{M} \left( \ln \left( \frac{M\|x_m\|^2}{\sum_{i=1}^{M} \|x_i\|^2} \right) \right)^2 \frac{H_1}{H_0} \geq \eta_{AO_2},$$

where $\eta_{AO_2}$ is set to satisfy the false alarm requirement.

In [75], it is shown that for this problem the GLRT is an SFET using MLE, where its SF is $\sum_{m=1}^{M} \theta_m M(\|x_m\|)$, where $X_{1,1}$ is the $(1\times1)^{th}$ element of $X$. That is because of $M(X)$ satisfies two conditions of maximal invariant as follows, 1) $M(q(X)) = M(X)$ for any $q \in Q$ and 2) if for any arbitrary $X, X'$, we have $M(X) = M(X')$, then $\frac{X}{\|X_{1,1}\|} = \frac{X'}{\|X'_{1,1}\|}$, we have $X = \frac{\|X_{1,1}\|}{\|X'_{1,1}\|} X'$, defining $\xi \triangleq \frac{\|X_{1,1}\|}{\|X'_{1,1}\|}$, we have $X = \xi X'$, i.e., there exists a group such that $X'$ after transformation is $X$.

The pdf of $M(X)$ given $[X]_{1,1}$ under $H_1$ is zero mean Gaussian as

$$f_M(M|X_{1,1}; \theta, H_1) = \frac{\exp(-r^2 \sum_{m=1}^{M} \frac{\|m_m\|^2}{\sigma_m^2} \prod_{m=1}^{M} \theta_m)}{\pi^{MN} \|m_m\| \prod_{m=1}^{M} \theta_m} \times \delta(\text{Im}([M]_{1,1}) - \sin(\phi))\delta(\text{Re}([M]_{1,1}) - \cos(\phi)), \quad (5.27)$$

where $\delta(\cdot)$ is the Dirac delta function and $[X]_{1,1} \overset{\Delta}{=} r \exp(j\phi)$ and $r \geq 0$ and $\phi \in [0, 2\pi]$ and $m_m$ is the $m^{th}$ column of $M$. Since the joint pdf of $r$ and $\phi$ is $f_{r,\phi}(r, \phi; \theta_1) = \frac{\exp(-r^2 \sum_{m=1}^{M} \frac{\|m_m\|^2}{\sigma_m^2} \prod_{m=1}^{M} \theta_m)}{\pi^{MN} \|m_m\| \prod_{m=1}^{M} \theta_m} \times \delta(\text{Im}([M]_{1,1}) - \sin(\phi))\delta(\text{Re}([M]_{1,1}) - \cos(\phi)),$
\( \frac{r}{\pi \theta_1} \exp(-r^2/\theta_1) \), the pdf of maximal invariant is given by

\[
f_M(M; \rho, H_1) = \int_0^{2\pi} \int_0^\infty f_{r, \phi}(r, \phi; \theta_1) drd\phi
\]

\[
= \frac{(MN-2)!}{2\pi^MN(1+\|m_1\|^2+\sum_{m=1}^{M-1} \frac{\|m_{m+1}\|^2}{\rho_m})^{MN} \prod_{m=1}^M \rho_m^{N_m}}
\]

Similarly the pdf of \( M \) under \( H_0 \) is \( f_M(M; H_0) = \frac{(MN-2)!}{2\pi^MN(1+\sum_{m=1}^{M} \|m_m\|^2)^{NL}} \). Using these pdfs, the MP-CFAR is given by

\[
f_M(M; \rho, H_1)/f_M(M; H_0) = \frac{(1 + \sum_{m=1}^{M} \|m_m\|^2)^{MN}}{(1 + \|m_1\|^2 + \sum_{m=1}^{M-1} \frac{\|m_{m+1}\|^2}{\rho_m})^{MN} \prod_{m=1}^M \rho_m^{N_m}} \frac{\eta_{MP}}{\eta_{H_0}}, \quad (5.28)
\]

where \( \eta_{MP} \) is given by the probability of false alarm. Since the test depends on \( \rho_m \)'s which are unknown, the achieved test is not a practical test and only provides an upper performance for CFAR tests [86]. Hence, we can consider the MP-CFAR tests as a benchmark for comparing the performance of the suboptimal CFAR tests. Figures 5.3 and 5.4 show the \( P_d \) versus the false alarm probability for the MP-CFAR bound, AOS-FETs and the GLRT. In Figure 5.3, we consider \( \theta_0 = 1.5 \) versus \( \theta_1 = [1 1.5 1.5 2 2]^T \) and \( N = 20 \) where AOSFET_1 outperforms AOSFET_2 and the GLRT. Moreover, we consider \( \theta_0 = 1.5 \) versus \( \theta_1 = [1 1.5 1.5 2 2]^T \) and \( N = 20 \) in Figure 5.3. In this simulation AOSFET_2 outperforms AOSFET_1 and the GLRT.
Figure 5.1: Probability of miss detection versus the false alarm probability for Example 22, for the AOSFET, the GLRT and the NP test when \( \sigma^2 \) is known and equals to 1, and \( \mu = 0.5 \), and \( N = 100 \).

Figure 5.2: Probability of missed detection versus the false alarm probability for Example 22, for the AOSFET, the GLRT and the NP test when \( \sigma^2 \) and \( \mu \) are unknown. In this simulation, we consider \( \sigma^2 = 1 \), \( \mu = 0.5 \), and \( N = 200 \).
5.3. EXAMPLES

Figure 5.3: Probability of detection versus false alarm probability for AOSFET₁, AOSFET₂, the GLRT and MP-CFAR, for \( \theta_0 = 1_5 \) versus \( \theta_1 = [1 \ 1.5 \ 1.5 \ 2 \ 2]^T \) and \( N = 20 \).

Figure 5.4: Probability of detection versus false alarm probability for AOSFET₁, AOSFET₂, the GLRT and the MP-CFAR, for \( \theta_0 = 1_5 \) versus \( \theta_1 = [1 \ 1 \ 1 \ 1 \ 0.5]^T \) and \( N = 20 \).
Chapter 6

Multi Class classification using SFET

In this chapter, we study the relationship between estimation and detection for M-ary hypothesis testing problems. The concept of SFET is extended to the M-ary problems by defining a vector of SFs. It is shown that the optimal test in MAP criteria and the proposed SFET are equivalent and in the case that the optimal MAP test does not exist, the proposed SFET provides an asymptotically optimal test. We show that the proposed SFET is also asymptotically optimal as the number of observations increases. Unlike the previous chapters, in this chapter the criterion is MAP for finding a test. Moreover, both parametric and Bayesian estimators are studied, i.e., in the last section of this chapter, we consider a known pdf for the unknown parameters. The rest of this chapter is organized as follows; In Section 6.1, a new definition for SF is provided for an M-ary problem. Then it is shown that if Uniformly Minimum Error test (UMET) exists, the SFET using the MLE and the UMET are equivalent. In the case that UMET does not exist, under some conditions, the error probability of the
6.1. RELATIONSHIP BETWEEN UMET AND MLE

The proposed SFET tends to zero by increasing the number of independent observations. For some problems that the Lebesgue measure of one or more set of unknowns is zero, the proposed test does not have a reliable performance. Hence three methods are proposed to improve the performance of the SFET. The improved SFET are proposed in Section 6.2. In Section 6.3, we consider the case that the unknown parameters are random with known pdf. We show that the optimal test is given by comparing the MMSE of a discrete SF.

### 6.1 Relationship between UMET and MLE

Consider the observation vector $x \in \mathbb{R}^N$, with pdf $f_x(x; \theta)$, where $\theta$ is the unknown vector and $\Theta \subseteq \mathbb{R}^M$ is the set of all possible $\theta$ and an M-ary hypothesis testing as

$$
\begin{align*}
\mathcal{H}_0 : & \quad \theta \in \Theta_0, \\
\mathcal{H}_1 : & \quad \theta \in \Theta_1, \\
& \vdots \\
\mathcal{H}_{M-1} : & \quad \theta \in \Theta_{M-1},
\end{align*}
$$

where $\Theta = \Theta_0 \cup_{m=0}^{M-1} \Theta_m$. We assume that the probability of each hypothesis is $\frac{1}{M}$, since the unknown parameters are deterministic. The optimal test to reach the minimum error probability is given by

$$
\psi_{\text{ME}}(x) = \arg\max_{m \in \mathcal{M}} f_x(x; \theta_m),
$$

where $\mathcal{M} \triangleq \{0, \cdots, M-1\}$, $\theta_m \in \Theta_m$ if the unknown parameters are given [9]. In many problems, $\psi_{\text{ME}}(x)$ depends on the unknown parameters, hence in such problems,
the optimal minimum error test does not exist, i.e., the decision regions depend on the unknown parameters. However, the UMET exists and equals to $\psi_{ME}(x)$ if the decision regions of $\psi_{ME}(x)$ do not depend on the unknown parameter. The UMET provides the minimum probability of error among all other tests. In the following, we show that the UMET is achieved by replacing the MLE of unknown parameters into an extension of the SF. The SF for a binary hypothesis testing problem is defined in Chapter 3. The following definition is an extension of SF to the M-ary problem followed by a theorem to show the relationship between estimation and detection in M-ary problems.

**Definition 1 (Continuous M-ary SF (CM-SF))** A vector of $M$ functions $g = [g_0, \cdots, g_{M-1}]^T$ is called CM-SF for problem (6.1), if $g_m$ is an SF for separating $\Theta_m$ and $\Theta - \Theta_m$, i.e., $g_m(\theta) > 0$ if $\theta \in \Theta_m$ and $g_m(\theta) \leq 0$ if $\theta \in \Theta - \Theta_m$.

Based on the definition of CM-SF, we have $\theta \in \Theta_m$ if and only if $g_m(\theta) > g_{m'}(\theta)$ for all $m \neq m'$. The existence of CM-SF using Lemma 1 is guaranteed.

**Theorem 13** If the UMET exists for (6.1), then it is given by

$$\psi_{CM-SFET}(x) = \arg\max_{m \in M} \hat{g}_m(\hat{\theta})$$

where $\psi_{CM-SFET}(x)$ is the Continuous M-ary-SFET (CM-SFET), $\hat{g}_m(\hat{\theta})$ is the MLE of $g_m(\theta)$ under the union of hypotheses, $f_x(x; \theta)$, $\theta \in \Theta$.

**Proof 20** From definition (6.3) and invariance property of MLE, we have $\psi_{CM-SFET}(x) = m$, if and only if, $\hat{\theta} \in \Theta_m$. Thus, $f_x(x; \hat{\theta}) \geq f_x(x; \theta)$ for all $\theta \in \Theta$. From definition of UMET, for the given $x$, since $f_x(x; \hat{\theta}) \geq f_x(x; \theta)$ and $\hat{\theta} \in \Theta_m$, then $\psi_{ME}(x) = m$.  


Remark 11 The UMET does not exist in many practical problems, while \( \psi_{\text{CM-SFET}}(\cdot) \) always exists. Hence, we can apply \( \psi_{\text{CM-SFET}}(\cdot) \) as a suboptimal test in the case that the UMET does not exist. Moreover even if UMET exists, it equals to the CM-SFET using MLE. Generally, the CM-SFET is a suboptimal test which is given by 
\[
\psi_{\text{CM-SFET}}(x) = \arg\max_{m \in \{0, \ldots, M-1\}} \hat{g}_m(\theta),
\]
where \( \hat{g}_m(\theta) \) is the estimation of \( g_m(\theta) \) under the union of hypotheses. In general, we can apply other types of estimation.

Remark 12 Based on the proof of Theorem 13, unlike the SFET of NP class, the performance of CM-SFET using MLE does not depend on selecting the CM-SF. Hence, the CM-SFET using MLE can be written as 
\[
\psi_{\text{CM-SFET}}(x) = \begin{cases} 
0 & \text{if } \min(x_n) \geq 0 \text{ and } \max(x_n) \leq \lambda_0, \\
m & \text{if } \min(x_n) > \lambda_{m-1} \text{ and } \max(x_n) \leq \lambda_m
\end{cases},
\]
which is equal to UMET.

Example 24 Consider a random vector \( x \) with iid elements uniformly distributed as \( U(\theta, \theta + \Delta) \), where \( \Delta \) is known, \( \theta \) is unknown and \( U(a, b) \) shows a 1D uniform distribution in interval \([a, b]\). Assume that \( \Theta_0 = [0, \lambda_0 - \Delta], \Theta_1 = (\lambda_0, \lambda_1 - \Delta], \ldots, \Theta_{M-1} = (\lambda_{M-2}, \lambda_{M-1} - \Delta] \). The UMET using maximizing the pdf under each hypothesis is given by 
\[
\psi_{\text{UME}}(x) = \begin{cases} 
0 & \text{if } x \in [0, \lambda_0]^N, \\
m & \text{if } x \in (\lambda_{m-1}, \lambda_m]^N,
\end{cases}
\]
where \( N \) is the number of observations. Now we show that the CM-SFET using MLE provides the same test. Consider the joint pdf \( x \) in terms of \( \theta \) as follows
\[
f_x(x; \theta) = \begin{cases} 
\frac{1}{\Delta^N}, & \theta \in [\max(x_n) - \Delta, \min(x_n)], \\
0, & \text{otherwise}.
\end{cases}
\] (6.4)

Thus the MLE of \( \theta \) is any value in \([\max(x_n) - \Delta, \min(x_n)]\). From Remark 12, the CM-SFET is given by 
\[
\psi_{\text{CM-SFET}}(x) = \begin{cases} 
0 & \text{if } \min(x_n) \geq 0 \text{ and } \max(x_n) \leq \lambda_0, \\
m & \text{if } \min(x_n) > \lambda_{m-1} \text{ and } \max(x_n) \leq \lambda_m
\end{cases}.
\]

In the following, we study the asymptotic performance of CM-SFET using MLE, if the observations are iid.

**Theorem 14** If \( \mathbf{x} \) is a vector of observations with iid elements and the pdf of each element is differentiable with respect to \( \mathbf{\theta} \) and \( \Theta_m \) is a Borel set for all \( m \in \mathcal{M} \), then the probability of error tends to zero for CM-SFET using MLE when the length of \( \mathbf{x} \) tends to infinity, almost everywhere in Lebesgue measure of \( \mathbb{R}^M \).

**Proof 21** From Theorem 3.7 of [25], since the pdf of each element of \( \mathbf{x} \) is differentiable, we have \( \hat{\mathbf{\theta}} \xrightarrow{P} \mathbf{\theta} \), where \( \mathbf{\theta} \) is the true value of unknown vector and the symbol \( \xrightarrow{P} \) shows the convergence in probability. Then the conditional correct detection probability given \( \mathcal{H}_m \) called \( \Pr(\mathbf{c}; \mathcal{H}_m) \) is given by \( \Pr(\mathbf{c}; \mathcal{H}_m) = \Pr(\hat{\mathbf{\theta}} \in \Theta_m) \), where \( \Theta_m \) is the true value belongs to \( \Theta_m \). Then, for any \( \mathbf{\theta} \in \Theta'_m \), where \( \Theta'_m \) is the set of all interior points of \( \Theta_m \), there exists \( \epsilon > 0 \) such that 1) the set \( \{ \vartheta : ||\vartheta - \mathbf{\theta}|| < \epsilon \} \subset \Theta \) and 2) \( \Pr(\hat{\mathbf{\theta}} \in \Theta_m) \) tends to one. Thus \( \Pr(\hat{\mathbf{\theta}} \in \Theta_m) \) tends to one. Since \( \Theta_m \) is a Borel set, then the measure of the set of all non-interior points is zero.

Theorem 13 and the proof of Theorem 14 shows that the CM-SFET is a reliable test when the UMET does not exist and the Lebesgue measure of \( \Theta_m \) is not zero. As mentioned in Remark 12, regardless of selecting the CM-SF, the performance of all CM-SFETs are equivalent if the MLE is applied as the estimator. However, in general, the MLE does not have a closed form or other estimators may provide more acceptable performance, hence selecting the CM-SF impacts on the performance of the CM-SFET for such cases. In the following, we show that under some mild conditions the better estimator in sense of \( \epsilon \)-accurate, provides a detector with less probability of error.
6.1. RELATIONSHIP BETWEEN UMET AND MLE

Theorem 15  Let \( \Theta_m \) be a Borel set with positive Lebesgue measure in (6.1) for all \( m \in M \) and \( \psi_{CM_1}(\cdot) \) and \( \psi_{CM_2}(\cdot) \) be two CM-SFETs defined as \( \psi_{CM_1}(x) = m \), if \( \hat{\theta}_1 \in \Theta_m \) and \( \psi_{CM_2}(x) = m \), if \( \hat{\theta}_2 \in \Theta_m \), for \( m \in M \) such that \( \hat{\theta}_1 \) is more \( \epsilon \)-accurate than \( \hat{\theta}_2 \) for all \( \theta \in \Theta = \bigcup_{m=0}^{M-1} \Theta_m \) and \( \epsilon > 0 \), then the error probability of \( \psi_{CM_1}(\cdot) \) is less than \( \psi_{CM_2}(\cdot) \).

Proof 22  First, we prove the following claim,

Claim: For any open set, there exists a set of disjoint countable open balls such that their union equals the origin open set.

Proof of claim: Consider an open set \( O \), and also consider \( x_0 \in O \), such that \( B(x_0, r_0) \subseteq O \) and \( r_0 \) is the greatest possible radius between all possible open balls in \( O \), where \( B(x_0, r_0) \) is the open ball with radius \( r_0 \) at point \( x_0 \). Now, we define \( x_1 \in O - \overline{B(x_0, r_0)} \), where \( \overline{B(x_0, r_0)} \) is the closure of \( B(x_0, r_0) \), as the point with greatest radius in \( O - \overline{B(x_0, r_0)} \) and similarly \( x_i \in O - \bigcup_{k=0}^{i-1} B(x_k, r_k) \) such that \( B(x_i, r_i) \) provides the greatest radius in \( O - \bigcup_{k=0}^{i-1} B(x_k, r_k) \). So we have \( O = \bigcup_{k=0}^{\infty} B(x_k, r_k) \).

This is because, if the latest equality is not valid, then there exists an open ball in \( O - \bigcup_{k=0}^{\infty} B(x_k, r_k) \) hence another open ball with greatest radius will be added to \( \bigcup_{k=0}^{\infty} B(x_k, r_k) \), which has a contradiction with the definition of \( \bigcup_{k=0}^{\infty} B(x_k, r_k) \). The claim is proven at this point.

In the following, we show that, the correct detection probability of \( \psi_{CM_1}(\cdot) \) is greater than \( \psi_{CM_2}(\cdot) \). Defining \( \Theta'_m \) as the set of all interior points of \( \Theta_m \), based on the claim, there exists a union of disjoint open balls named \( \bigcup_{k=0}^{\infty} B(x_k, r_k) \). Since \( \hat{\theta}_1 \) is more \( \epsilon \)-accurate than \( \hat{\theta}_2 \), then \( \Pr_{\theta}\{\hat{\theta}_1 \in B(x_k, r_k)\} \geq \Pr_{\theta}\{\hat{\theta}_2 \in B(x_k, r_k)\} \), where \( \theta \in \Theta_m \). So from the presented claim in this proof, we have \( \Pr_{\theta}\{\hat{\theta}_1 \in \Theta'_m\} \geq \Pr_{\theta}\{\hat{\theta}_2 \in \Theta'_m\} \). The conditional correct detection probability given \( H_m \) is \( \Pr_{\theta}\{\hat{\theta}_1 \in \Theta'_m\} \).
6.1. RELATIONSHIP BETWEEN UMET AND MLE

\[ \Theta_m \} \text{ for } l = 1, 2. \] Moreover, \( \Pr_\theta \{ \hat{\theta}_l \in \Theta_m \} = \Pr_\theta \{ \hat{\theta}_l \in \Theta_m' \} + \Pr_\theta \{ \hat{\theta}_l \in \Theta_m - \Theta_m' \} \) and since \( \Pr_\theta \{ \hat{\theta}_l \in \Theta_m - \Theta_m' \} \) is less than or equal to the Lebesgue measure of \( \Theta_m - \Theta_m' \) (from the Cauchy-Schwarz inequality) which is zero then \( \Pr_\theta \{ \hat{\theta}_l \in \Theta_m - \Theta_m' \} = 0 \).

Hence the conditional correct detection probability given \( \mathcal{H}_m \) named \( P_l(c; \mathcal{H}_m) \) equals to \( P_l(c; \mathcal{H}_m) = \Pr_\theta \{ \hat{\theta}_l \in \Theta_m \} = \Pr_\theta \{ \hat{\theta}_l \in \Theta_m' \} \) for \( \psi_{CM}(\cdot) \) for \( l = 1, 2 \). Thus \( P_1(c; \mathcal{H}_m) \geq P_2(c; \mathcal{H}_m) \). Since the total correct detection probability is given by \( P_l(c) = \frac{1}{M} \sum_{m=0}^{M-1} P_l(c; \mathcal{H}_m) \) which is shown that \( P_1(c) \geq P_2(c) \).

The condition of having non-zero Lebesgue measure for \( \Theta_m \) is not removable in this theorem. Moreover this condition has a very significant role in the CM-SFET performance. For example consider a very simple hypothesis as follows

**Example 25** Assume \( \mathcal{H}_0 : \theta = 0 \) versus \( \mathcal{H}_1 : \theta \neq 0 \), where the observation is a Gaussian vector with mean \( \theta \) and variance one. This example is a type of two sided binary hypothesis problem. The UMET for this test does not exist, because the LR depends on \( \theta \). So, we apply the CM-SFET for this problem as follows. The estimation of \( \theta \) under the union of both hypothesis is \( \hat{\theta} = \frac{1}{N} \sum_{n=1}^{N} x_n \), where \( x_n \) is the \( n^{th} \) element of observations and \( N \) is the number of total observations. Hence, the CM-SFET is given by \( \psi_{CM-SFET}(\mathbf{x}) = 0 \), if \( \frac{1}{N} \sum_{n=1}^{N} x_n = 0 \) and \( \psi_{CM-SFET}(\mathbf{x}) = 1 \), if \( \frac{1}{N} \sum_{n=1}^{N} x_n \neq 0 \), where \( \mathbf{x} = [x_1, \cdots, x_N]^T \). Since the pdf of \( \hat{\theta} \) under \( \mathcal{H}_0 \) is a continuous pdf, then the probability of correct detection is always zero, except when the number of observation mathematically is infinity.

In this example, although the error probability of CM-SFET is zero when the number of observations is infinity, this test is not a practical test. We call such hypothesis testing problems Non-Detectable Problems (NDPs) using SFET. In many class of
signal processing, we cannot find a reliable test for the problems with zero-Lebesgue
measure unknown sets, if the joint pdf of estimations is continuous. In the following,
three alternative solutions are proposed to solve the NDP using SFET.

6.2 Non-Detectable Problems using SFET

In this section three modifies CM-SFET methods are proposed to deal with NDP
using SFET. First method is based on minimizing the probability of error using an
SFET given by NP criterion. This method is only applicable on the binary testing.
The second and third methods are based on relaxing the hypothesis testing, such that
the relaxed problem does not have the zero-Lebesgue measure sets while the original
problem is a sub-problem of the relaxed problem. The second and third methods are
applied on general M-ary NDPs.

6.2.1 Minimizing the Probability of Error Using NP criterion

In this subsection, we propose a method to find a CM-SFET for binary hypothesis
testing problem using the SFET achieved from NP criterion. For a CM-SFET, if the
joint pdf of estimation of unknown parameters is a continuous pdf and at least one
of unknown parameter sets, \( \Theta_m \), has a zero Lebesgue measure, then the probability
of correct detection under that hypothesis is zero for finite observations. The proof
of this claim is very straightforward. Using NP criterion, we do not face with this
problem, because, we accept a false alarm probability, and the acceptance of that false
alarm allows us to pick a hypothesis if even the corresponding Lebesgue measure
is zero. For example, consider Example 25 again, the SFET using MLE based on
NP criterion is given by

\[
\frac{1}{N} \sum_{n=1}^{N} x_n |_{H_1} \geq \frac{\eta}{H_0}
\]
6.2. NON-DETECTABLE PROBLEMS USING SFET

requirement. In this test, the hypothesis $\mathcal{H}_0$ is picked when $|\frac{1}{N} \sum_{n=1}^{N} x_n|^2 < \eta$ and absolutely the probability of correct detection is not zero. However, in MAP criterion, we need to minimize the error probability. So we can adjust the threshold $\eta$ to minimize the probability of error. Following remark shows that, from a ROC curve we can find the point that probability of error is minimum for a test which is given by NP criterion.

**Remark 13** Consider an SFET for a binary hypothesis test, achieving by the NP criterion. Then the threshold corresponding to $P_{fa}$ that the slope of ROC curve is one, provides the minimum error for the SFET.

Proof of Remark 13 is straightforward. Consider $P_e = \frac{1}{2} P_{fa} + \frac{1}{2} (1 - P_d)$. Then tacking the derivative on $P_e$ with respect to $P_{fa}$ and setting to zero, we have $\frac{\partial P}{\partial P_{fa}} = 1$. Hence, to solve a binary hypothesis using MAP criterion based on the SFET concept (when at least one of hypotheses has a zero Lebesgue measure) an alternative solution to deal with the zero-Lebesgue measure sets is that, first, we find an SFET using NP criterion, then we consider the threshold of the achieved SFET by the point with slope one in the ROC curve. Although this method provides the minimum error probability using SFET, there are two major problems in the implementations. First, this method is only applicable on the binary problems and second, unfortunately, the point on the ROC curve with slope one depends on the unknown parameters (usually SNR in signal detection problems). However, in the problem that the corresponding threshold does depend on the unknown parameters and the problem is binary, this method is applicable.
6.2. NON-DETECTABLE PROBLEMS USING SFET

6.2.2 Relaxing the Hypothesis

Consider the M-ary hypothesis in (6.1), where for any \( k \in \{0, \cdots, K - 1\} \), the Lebesgue measure of \( \Theta_k \) is zero and \( k \leq M - 1 \) while the Lebesgue measure of \( \Theta_m \) is positive for \( m \in \{K + 1, \cdots, M - 1\} \). Now consider an extension of (6.1) as follows

\[
\begin{align*}
\tilde{H}_0 : & \quad \theta \in \tilde{\Theta}_0, \\
\vdots & \quad \vdots \\
\tilde{H}_{K-1} : & \quad \theta \in \tilde{\Theta}_{K-1}, \\
H_K : & \quad \theta \in \Theta_K, \\
\vdots & \quad \vdots \\
H_{M-1} : & \quad \theta \in \Theta_{M-1},
\end{align*}
\]  

(6.5)

where the Lebesgue measure of \( \tilde{\Theta}_k \) is positive for all \( k \in \{0, \cdots, K - 1\} \) and sets are disjoint. The original problem is a subproblem of the new hypothesis testing problem (6.5) called relaxed problem. Based on Theorem 4 of [93] for two disjoint closed sets two open disjoint covering sets exist, the relaxed hypothesis exists if the zero Lebegue sets of unknowns are given by a finite union of sets with zero Lebesgue measure.

**Remark 14** From Theorem 14, the CM-SFET using the MLE is asymptotically optimal for the relaxed problem, if the pdf of observations is differentiable with respect to all \( \theta \in \cup_{k=0}^{K-1} \tilde{\Theta}_k \cup_{m=K}^{M-1} \Theta_m \). Based on the proof of Theorem 14, if \( \theta \in \Theta_k \) is an interior point of \( \tilde{\Theta}_k \), then the conditional error probability given \( H_k \) for CM-SFET for the NDP tends to zero. Hence a sufficient conditions for asymptotically optimality of CM-SFET using MLE and relaxed problem is that any point of \( \Theta_k \) is an interior point of \( \tilde{\Theta}_k \). In Example 26, we show that, this condition is not necessary.
Finding a relaxed problem is not unique, hence each relaxation provides a corresponding CM-SFET. However, finding the optimal relaxation, or equivalently, finding the optimal relaxed set of unknowns depends on the pdf of estimation of unknown parameters. To find the optimal \( \hat{\Theta}_k \), we need to solve the following maximization problem for \( k \in \{0, \cdots, K-1\} \)

\[
\max_{\Theta_k} \sum_{k=0}^{K-1} \Pr_{\theta \in \hat{\Theta}_k} \{ \hat{\theta} \in \hat{\Theta}_k \}.
\] (6.6)

In many problems, since the pdf of the estimation of unknown parameters depends on the unknown parameters, hence the optimal relaxation does not exist. Nevertheless, any CM-SFET is asymptotically optimal under some mild conditions provided in Remark 14.

**Remark 15** Theorem 15 is valid for problem (6.5), because \( \hat{\Theta}_k \) is a Borel set for all \( k \in \{0, \cdots, K-1\} \). Hence, in the case of existing zero-Lebesgue measure set, a more \( \epsilon \)-accurate estimator for unknown parameters in \( \cup_{k=0}^{K-1} \hat{\Theta}_k \cup_{m=K}^{M-1} \Theta_m \) provides a test with lower error probability.

In the following two methods are proposed to relax (6.1) when the problem is NDP using SFET.

**Mixture Modeling**

Consider problem (6.1), defining a mixture pdf of observations by

\[
f_x(x; \vartheta) = \sum_{m=0}^{M-1} \pi_m f_x(x; \mathcal{H}_m, \theta),
\]
where $\sum_{m=0}^{M-1} \pi_m = 1$, $\pi_m \in [0,1]$ and $f_x(x; H_m, \theta) = f_x(x; \theta)$ if $\theta \in \Theta_m$ and new unknown vector is defined as $\theta = [\pi_0, \cdots, \pi_{M-1}, \theta]^T$. In the following, we show that if the UMET exists, it is given by estimation of $\pi_m$'s.

**Theorem 16** If UMET exists for (6.1), then the UMET is given by

$$\psi_{CM-SFET}(x) = \arg\max_{m \in \mathcal{M}} \hat{\pi}_m, \quad (6.7)$$

where $\hat{\pi}_m$ is the MLE of $\pi_m$ using the relaxed pdf $f_x(x; \theta) = \sum_{m=0}^{M-1} \pi_m f_x(x; H_m, \theta)$ and $\pi_m$'s are the mixture coefficients.

**Proof 23** The MLE of $\pi_m$ is given by maximizing $f_x(x; \theta) = \sum_{m=0}^{M-1} \pi_m f_x(x; H_m, \theta)$. Hence, we have

$$\arg\max_{\pi_m, m \in \{0, \cdots, M-1\}} \sum_{m=0}^{M-1} \pi_m f_x(x; H_m, \theta), \quad (6.8)$$

subject to $\sum_{m=0}^{M-1} \pi_m = 1$, $0 \leq \pi_m \leq 1$.

We consider the Karush-Kuhn-Tucker (KKT) conditions [94] for (6.8) as follows

$$f + \lambda 1_M + \sum_{m=0}^{M-1} \mu_m p_m - \sum_{m=0}^{M-1} \mu'_m p_m = 0_M, \quad (6.9)$$

$$\sum_{m=0}^{M-1} \pi_m = 1$$

$$\mu'_m \geq 0, \mu_m \geq 0,$$

$$\mu_m(\pi_m - 1) = 0, \mu'_m \pi_m = 0, m \in \{0, \cdots, M-1\};$$

where $f = [f_x(x; H_0, \theta), \cdots, f_x(x; H_{M-1}, \theta)]^T$, $1_M$ is the vector with length $M$ such
that all elements are one, \( p_m \) is an \( M \) dimension vector such that all elements are zero except the \( m \)th element which is one and \( \mu'_m \) and \( \mu_m \) are the Lagrange multipliers for \( m \in \{0, \ldots, M-1\} \). The set of equations (6.9) are satisfied if and only if \( \pi = p_m \), where \( \pi = [\pi_0, \ldots, \pi_{M-1}]^T \), for all \( m \in \{0, \ldots, M-1\} \). Applying the achieved \( \pi \) from KKT conditions, the objective function in (6.8) turns to \( \arg\max_{\pi_m, m \in \{0, \ldots, M-1\}} f_x(x; \mathcal{H}_m, \theta) \), which is the UMET.

**Remark 16** In the case that the UMET does not exists, the CM-SFET using (6.7) provides a suboptimal test. Based on the proof of Theorem 16, in the case that UMET does not exist, the test is given by \( \psi_{CM-\text{SFET}}(x) = \arg\max_{\pi_m, m \in \{0, \ldots, M-1\}} f_x(x; \mathcal{H}_m, \hat{\theta}) \), where \( \hat{\theta} \) is estimated based on the mixture pdf model \( f_x(x; \mathcal{H}_m, \theta) = \sum_{m=0}^{M-1} \pi_m f_x(x; \mathcal{H}_m, \theta) \).

A method to estimate the unknown parameters using the mixture model \( f_x(x; \theta) \) is Expectation Maximization (EM) converging to MLE while the new problem is not NDP. Moreover, since the EM of unknowns tends to MLE, the proposed CM-SFET is asymptotically optimal from Theorem 14.

**Example 26** Consider the classification of a Binary phase-shift keying (BPSK) signal as null hypothesis \( \mathcal{H}_0 \) versus Quadrature Phase Shift Keying (QPSK) signal as alternative hypothesis \( \mathcal{H}_1 \), with unknown shaping signal and noise variance. Assume that we have a matrix of observations \( X = [x_1, x_2, \ldots, x_L] \), where \( x_l \) is the \( l \)th received symbol and for all \( l \neq l' \), \( x_l \) and \( x_{l'} \) are independent. The pdf of \( x_l \) for \( l \in \{1, \ldots, L\} \), under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) are given by 
\[
\frac{1}{2\pi^N \sigma^2 N} \exp\left(-\frac{||x_l-s||^2}{\sigma^2}\right) + \frac{1}{4\pi^N \sigma^2 N} \exp\left(-\frac{||x_l+s||^2}{\sigma^2}\right)
\]
and 
\[
\frac{1}{2\pi^N \sigma^2 N} \exp\left(-\frac{||x_l-j||^2}{\sigma^2}\right) + \frac{1}{4\pi^N \sigma^2 N} \exp\left(-\frac{||x_l+j||^2}{\sigma^2}\right)
\]
respectively. So the set of unknown parameters under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) are \( \Theta_0 = \{[\pi_1, \pi_2, \pi_3, \pi_4, s_R, s_I, \sigma^2]^T | \pi_1 = \frac{1}{2}, \pi_2 = \frac{1}{2}, \pi_3 = 0, \pi_4 = 0, s_R, s_I \in \mathbb{R}^N, \sigma^2 > 0 \} \) and \( \Theta_1 = \{[\pi_1, \pi_2, \pi_3, \pi_4, s_R^T, s_I^T, \sigma^2]^T | \pi_1 = \frac{1}{4}, \pi_2 = \frac{1}{4}, \pi_3 = 0, \pi_4 = \frac{1}{2}, s_R, s_I \in \mathbb{R}^N, \sigma^2 > 0 \} \).
\[ \frac{1}{4}, \pi_3 = \frac{1}{4}, \pi_4 = \frac{1}{4}, s_R, s_I \in \mathbb{R}^N, \sigma^2 > 0 \}. \] The dimension of any vector in \( \Theta_0 \) and \( \Theta_1 \) is \( 2N+5 \) and the Lebesgue measure of \( \Theta_0 \) and \( \Theta_1 \) in \( \mathbb{R}^{2N+5} \) is zero, where \( N \) is the length of \( x_l \). Hence the hypothesis testing problem is NDP using SFET. In order to find a relaxed hypothesis testing problem, consider \( \tilde{\mathcal{H}}_0 : \theta \in \tilde{\Theta}_0 \) versus \( \tilde{\mathcal{H}}_1 : \theta \in \tilde{\Theta}_1 \), where \( \tilde{\Theta}_0 = \{ [\pi_1, \pi_2, \pi_3, \pi_4, s_R^T, s_I^T, \sigma^2] | \pi_1, \pi_2 \in [0, 1], \pi_3 + \pi_4 < \eta, s_R, s_I \in \mathbb{R}^N, \sigma^2 > 0 \} \) and \( \tilde{\Theta}_1 = \{ [\pi_1, \pi_2, \pi_3, \pi_4, s_R^T, s_I^T, \sigma^2] | \pi_1, \pi_2 \in [0, 1], \pi_3 + \pi_4 \geq \eta, s_R, s_I \in \mathbb{R}^N, \sigma^2 > 0 \} \), where \( \eta \in (0, \frac{1}{2}) \) determines the boundary of the relaxed set of unknowns \( \tilde{\Theta}_0 \) and \( \tilde{\Theta}_1 \).

From the new relaxed problem, the original problem is a sub-problem of the relaxed problem, i.e., \( \Theta_0 \subset \tilde{\Theta}_0 \) and \( \Theta_1 \subset \tilde{\Theta}_1 \). To find the optimal value of \( \eta \), from (6.6), the probability of error is minimized or equivalently the probability of correct detection is maximized. If the solution of (6.6) depends on the unknown parameters the optimal thresholding in this structure does not exist. Based on the relaxed problem, the CM-SFET is given by

\[ \hat{\pi}_3 + \frac{\eta_1}{\eta_0} \geq \eta, \] (6.10)

where \( \hat{\pi}_2 \) and \( \hat{\pi}_3 \) are the estimations of \( \pi_2 \) and \( \pi_3 \), using the relaxed pdf

\[ f_X(X; \theta) = \prod_{l=1}^{L} \frac{\pi_1}{\pi_N \sigma^2 N} \exp\left( -\frac{\|x_l - s\|^2}{\sigma^2} \right) + \frac{\pi_2}{\pi_N \sigma^2 N} \exp\left( -\frac{\|x_l + s\|^2}{\sigma^2} \right) + \frac{\pi_3}{\pi_N \sigma^2 N} \exp\left( -\frac{\|x_l - js\|^2}{\sigma^2} \right) + \frac{\pi_4}{\pi_N \sigma^2 N} \exp\left( -\frac{\|x_l + js\|^2}{\sigma^2} \right), \] (6.11)

respectively. To estimate \( \pi_1, \pi_2, \pi_3 \) and \( \pi_4 \), an EM method is proposed. In [95] it is shown that for the Gaussian mixture modeling the EM converges to MLE. Figure 7.10 shows the probability of error versus \( \eta \) for Signal to Noise Ratio (SNR). In this simulation the variation of SNR is from \(-40\text{dB}\) to \(-8\text{dB}\), where \( N = 10 \) and \( L = 800 \)
and the SNR is defined by $\text{SNR} = \frac{\|s\|^2}{N L \sigma^2}$. In each SNR, the optimal $\eta$ is given by the point that minimizes the probability of error. The simulation shows that the optimal $\eta$ depends on SNR. Since the SNR is unknown, the optimal $\eta$ does not exist. In such case, we can consider $\eta = \frac{1}{4}$ which is the middle of the interval of all possible $\eta$’s.

### 6.2.3 Exponentially Embedded Family

Another method to find a relaxed problem is Exponentially Embedded Family (EEF). The EEF of the pdfs $f_x(x; H_m, \theta)$, for $m \in M$ is given by [92]

$$f_x(x; \lambda, \theta) = \frac{\prod_{m=0}^{M-1} f_{x}^{\lambda_{m}}(x; H_m, \theta)}{\int_{\mathbb{R}^N} \prod_{m=0}^{M-1} f_{x}^{\lambda_{m}}(x; H_m, \theta) \, dx},$$

(6.12)

where $\lambda = [\lambda_0, \cdots, \lambda_{M-1}]^T$, $\sum_{m=0}^{M-1} \lambda_m = 1$ and $\lambda_m \in [0, 1]$. Thus the CM-SFET is

$$\psi_{\text{CM-SFET}}(x) = \arg\max_{m \in M} \lambda_{m},$$

(6.13)
where $\hat{\lambda}_m$ is the MLE of $\lambda_m$ using $f_x(x; \lambda, \theta)$ which is given by setting the derivative of $f_x(x; \lambda, \theta)$ with respect to $\lambda_m$ equal to zero. Replacing $\lambda_{m'} = 1 - \sum_{m=0, m \neq m'}^{M-1} \lambda_m$, we have

$$f_x(x; \lambda, \theta) = \exp \left( f_x(x; H_{m'}, \theta) - k_{m'}(\lambda) + \sum_{m=0, m \neq m'}^{M-1} \lambda_ml_{m,m'}(x; \theta) \right), \quad (6.14)$$

where $l_{m,m'}(x; \theta) = \ln \left( \frac{f_x(x; H_{m'}, \theta)}{f_x(x; H_{m}, \theta)} \right)$ and

$$k_{m'}(\lambda) = \ln \left( \int_{\mathbb{R}^N} \exp \left( \sum_{m=0, m \neq m'}^{M-1} \lambda_ml_{m,m'}(x; \theta) \right) f_x(x; H_{m'}, \theta) dx \right). \quad (6.15)$$

In the following we show that the MLE of $\lambda_m$ is an increasing function in $f(H_m|x; \theta)$, where $\hat{\theta}$ shows the MLE of $\theta$. Hence in the case of existing the UMET, (6.13) provides the optimal test. To find the MLE of $\lambda_m$, setting $\frac{\partial f_x(x; \lambda, \theta)}{\partial \lambda_m} = 0$, leads us to solve

$$\left. \frac{\partial k_{m'}(\lambda)}{\partial \lambda_m} \right|_{\lambda_m = \hat{\lambda}_m} = l_{m,m'}(x; \hat{\theta}), \quad \frac{\partial k_{m'}(\lambda)}{\partial \lambda_m} = \int_{\mathbb{R}^N} l_{m,m'}(x; \theta) \exp \left( \sum_{i=0, i \neq m'}^{M-1} \lambda_il_{i,m'}(x; \theta) \right) f_x(x; H_{m'}, \theta) dx \bigg|_{\alpha} = \int_{\mathbb{R}} \alpha \exp(\lambda_m \alpha) d\mu_{m,m'}(\alpha),$$

where $d\mu_{m,m'}(\alpha) \triangleq \sum_{i=0, i \neq m'}^{M-1} \lambda_il_{i,m'}(x; \theta) f_x(x; H_{m'}, \theta) dx$ and $\alpha \triangleq l_{i,m'}(x; \theta)$. In the following, we show that $\frac{\partial k_{m'}(\lambda)}{\partial \lambda_m}$ is increasing in $\lambda_m$, which shows that $\hat{\lambda}_m$ is increasing in $l_{m,m'}(x; \hat{\theta})$. To prove the increasing property of $\frac{\partial k_{m'}(\lambda)}{\partial \lambda_m}$, we show that the $\frac{\partial^2 k_{m'}(\lambda)}{\partial \lambda_m^2} \geq 0$. From (6.16), we have

$$\frac{\partial^2 k_{m'}(\lambda)}{\partial \lambda_m^2} = \int_{\mathbb{R}} \alpha^2 dF_{m,m'}(\alpha) - \left( \int_{\mathbb{R}} \alpha dF_{m,m'}(\alpha) \right)^2 \geq 0 \quad (6.17)$$
where \( dF_{m,m'}(\alpha) \triangleq \frac{\exp(\lambda m \alpha) d\mu_{m,m'}(\alpha)}{\int \exp(\lambda m \alpha) d\mu_{m,m'}(\alpha)}. \) The later inequality is given by Jensens inequality theorem. Since \( \frac{\partial k_{m,m'}(\lambda)}{\partial \lambda m} \) is increasing in \( \lambda m \), from (6.16) \( \widehat{\lambda}_m \) is increasing in \( l_{m,m'}(x; \widehat{\theta}) \). Similarly, for any \( m' \neq m \), \( \widehat{\lambda}_m \) is increasing in \( l_{m,m'}(x; \widehat{\theta}) \), then \( \widehat{\lambda}_m \) is also decreasing in \( \sum_{m'=0, m' \neq m}^{M-1} \exp(-l_{m,m'}(x; \widehat{\theta})) = \frac{1-f(\mathcal{H}_m|x; \theta)}{f(\mathcal{H}_m|x; \theta)} \), then \( \widehat{\lambda}_m \) is increasing in \( f(\mathcal{H}_m|x; \theta). \)

**Example 27** In this example the CM-SFET using EEF is applied on the problem of Example (26). The EEF for this problem is given by

\[
fx(X, \lambda, \widehat{\theta}) = \frac{f^1_X(X; \mathcal{H}_1, \widehat{\theta}_1) f^{1-\lambda}_X(X; \mathcal{H}_0, \widehat{\theta}_0)}{\int_{\mathbb{R}^N \times L} f^1_X(X; \mathcal{H}_1, \widehat{\theta}_1) f^{1-\lambda}_X(X; \mathcal{H}_0, \widehat{\theta}_0) dX} \tag{6.18}
\]

where \( \widehat{\theta}_1 \) and \( \widehat{\theta}_0 \) are the EM estimation of \( \theta \) under \( \mathcal{H}_1 \) and \( \mathcal{H}_0 \) respectively. Then the estimation of \( \lambda \) satisfies

\[
\frac{1}{L} \sum_{l=1}^L \ln \left( \frac{f_{x_l}(x_l; \mathcal{H}_1, \widehat{\theta}_1)}{f_{x_l}(x_l; \mathcal{H}_0, \widehat{\theta}_0)} \right) = \frac{\int_{\mathbb{R}^N} \ln \left( \frac{\int_{\mathbb{R}^N} f_{x_l}(y; \mathcal{H}_1, \widehat{\theta}_1)}{\int_{\mathbb{R}^N} f_{x_l}(y; \mathcal{H}_0, \widehat{\theta}_0)} \right)^{\lambda} f_{x_l}(y; \mathcal{H}_1, \widehat{\theta}_1) dy}{\int_{\mathbb{R}^N} \left( \frac{\int_{\mathbb{R}^N} f_{x_l}(y; \mathcal{H}_1, \widehat{\theta}_1)}{\int_{\mathbb{R}^N} f_{x_l}(y; \mathcal{H}_0, \widehat{\theta}_0)} \right)^{\lambda} f_{x_l}(y; \mathcal{H}_0, \widehat{\theta}_0) dy} \tag{6.19}
\]

To find \( \widehat{\lambda} \) from (6.19), an exhaustive search over \( \widehat{\lambda} \in [0, 1] \) is applied. So, the test is

\[
\widehat{\lambda} \overset{\mathcal{H}_1}{\gtrless} \frac{1}{2}. \tag{6.20}
\]

Figure 6.2 shows the probability of error for (6.10) and (6.20), when \( \eta = \frac{1}{4} \). It is seen that for low SNR the CM-SFET using estimating the EEF factor \( \lambda \) outperforms the CM-SFET using mixture modeling, while by increasing the SNR, CM-SFET using mixture modeling provides better performance.
6.3 SFET for Random Unknowns

Figure 6.2: Probability of error versus SNR, for (6.10) and (6.20), when $N = 10$, $L = 800$.

6.3 SFET for Random Unknowns

In this section, we consider the case that the unknown vector $\theta$ has a known pdf, hence the a posteriori probability exists for each hypothesis. Let $f_\theta(\theta)$ be the distribution of $\theta \in \bigcup_{m=0}^{M-1} \Theta_m$. The following theorem shows the relationship between estimation and detection for (6.1) when the pdf of unknowns is available. In this section, we define discrete SF as follows;

**Definition 2 (Discrete M-ary SF)** For (6.1), the Discrete M-ary SF (DM-SF) is defined as vector of functions $[g_0, \cdots, g_{M-1}]^T$ such that

$$g_m(\theta) = \begin{cases} 1 & \theta \in \Theta_m, \\ -1 & \theta \in \Theta - \Theta_m, \end{cases}, \quad \Theta = \bigcup_{m=0}^{M-1} \Theta_m, m \in \mathcal{M}. \quad (6.21)$$

**Theorem 17** The minimum error test for (6.1) is given by

$$\psi_M(x) = \arg\max_{m \in \mathcal{M}} \left( g_m(\hat{\theta}) \right)_{MS}, \quad (6.22)$$
where $\widehat{g_m(\theta)}_{\text{MS}}$ is the MMSE of $g_m(\theta)$ under assumption of having the pdf of $\theta$.

**Proof 24** Since the MMSE of $g_m(\theta)$ is $\widehat{g_m(\theta)}_{\text{MS}} = \int_{\Theta} g_m(\theta) f_\theta(\theta|x) d\theta$, then the right hand side of (6.22) is given by

$$
\arg \max_{m \in M} \widehat{g_m(\theta)}_{\text{MS}} = \arg \max_{m \in M} \int_{\Theta} g_m(\theta) f_\theta(\theta|x) d\theta \\
= \int_{\Theta_m} f_\theta(\theta|x) d\theta - \int_{\Theta - \Theta_m} f_\theta(\theta|x) d\theta \quad (6.23)
$$

Since $\Pr(\mathcal{H}_m|x) = \int_{\Theta_m} f_\theta(\theta|x) d\theta$, then $1 - \Pr(\mathcal{H}_m|x) = \int_{\Theta - \Theta_m} f_\theta(\theta|x) d\theta$, so we can rewrite (6.23) as

$$
\arg \max_{m \in M} \widehat{g_m(\theta)}_{\text{MS}} = \arg \max_{m \in M} 2\Pr(\mathcal{H}_m|x) - 1.
$$

Then $\arg \max_{m \in M} 2\Pr(\mathcal{H}_m|x) - 1 = \arg \max_{m \in M} \Pr(\mathcal{H}_m|x)$, which is the minimum error test for (6.1).
Chapter 7

Practical Examples

In this section, we study the performance of SFET, SFET-CFAR, GLRT and optimal performance bounds for three important practical problems. In the first Section a MIMO radar problem is considered. In the second section, we solve a sparse signal detection problem in Gaussian noise and apply the results for Voice Activity Detection (VAD). Finally, a narrowband signal detection using linear array with unknown parameters is studied.

The result of Sections 7.1 is published in [96,97], Section 7.2 is submitted in [98] and the result of Section 7.3 is submitted in [99].

7.1 MIMO Radar Target Detection

MIMO signal processing has various applications involving radar, wireless communication, and cognitive radio [49,85]. MIMO radar is a type of multi-static radars, that employs multiple transmit antennas with different waveforms and jointly processes signals for detection of target and estimation of its parameters [85]. In general, the
MIMO radar systems are classified into the Widely Separated Antennas (WSA) and Co-Located Antennas (CLA) \cite{85}. In WSA MIMO radar, the antennas are spatially separated at the transmitters and the receivers, such that the target signals received at the receivers are independent \cite{3, 85}. The focus of this section is on the WSA MIMO radar target detection and signal design.

In this chapter, we derive the UMPI test for target detection problem in clutter with unknown covariance matrix using WSA MIMO radar. It is shown that the UMPI test depends on the scatter and Signal to Interference Ratio (SIR). Hence, for a given SIR, this test provides the MPI bound. Since in this problem the UMPI test does not exist, we utilize the SFET using SIR as a suboptimal invariant test. Based on the eigenvalues expansion of SIR, a set of signals is proposed such that maximizes the MPI bound.

Consider a MIMO radar with $K$ transmit and $L$ receive antennas. It is shown that the received signal at the $l$th receiver from a target located at the $i$th cell, is

$$r_i^l(n) = \sqrt{\frac{E_s}{K}} \sum_{k=1}^{K} h_{ik}^i s_k(nT_s) + w_i^l(nT_s),$$

where $n = 0, \ldots, N-1$, $w_i^l(nT_s)$ is Gaussian interference term, $E_s$ is the total energy of transmit signals, $s_k(nT_s)$ denotes the transmitted signal from the $k$th transmitter, and $h_{ik}^i$ denotes the channel coefficients. The samples of the interference are zero mean Gaussian with unknown covariance matrix $\Sigma_{ww}$ \cite{3}. It is shown that $h_{ik}^i$'s are iid Gaussian random variables which depend on the field scattering \cite{3}.

Assuming $r_i^l \overset{\Delta}{=} [r_i^l(0), \ldots, r_i^l(N-1)]^T$ and $s_k \overset{\Delta}{=} [s_k(0), \ldots, s_k(N-1)]^T$, the observation $r_i^l$ is a zero mean, Gaussian random vector with covariance matrix $\Sigma_i^l = \sigma^2 R_{ss} + \Sigma_{ww}$, where $\Sigma_{ww}$ is the interference covariance matrix, $\sigma^2$ is variance of $h_{ik}^i$. 
\( \mathbf{R}_{ss} = \sum_{k=1}^{K} \mathbf{s}_k \mathbf{s}_k^H \), where the superscript \(^H\) denotes the Hermitian operator. In this section, we assume that all channel coefficients have the same variance denoted by \( \sigma^2 \). Hence, we can rewrite the MIMO target detection problem using \( C \) cells by the following hypothesis test:

\[
\begin{aligned}
\mathcal{H}_0 & : \mathbf{r}^i_l \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ww}), \\
\mathcal{H}_1 & : \begin{cases} \\
\mathbf{r}^i_l \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}_{ss} + \mathbf{\Sigma}_{ww}), & i = c, \\
\mathbf{r}^i_l \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ww}), & i \neq c,
\end{cases}
\end{aligned}
\] (7.1)

where \( l = 1, \ldots, L, i = 1, \ldots, C \), and \( c \) is the number of cell under test (CUT).

### 7.1.1 UMPI Test

The decision problem remains unchanged by the set of transformations \( Q_{qu} = \{ q_{qu} : q_{qu}(\mathbf{r}^i_l) = \mathbf{Qr}^i_l \} \) and \( Q_s = \{ q_s : q_s(\mathbf{r}^i_l) = \xi \mathbf{r}^i_l, \xi \neq 0 \} \), where the matrix \( \mathbf{Q} \) is a \( N \times N \) quasi-unitary matrix with respect to \( \mathbf{R}_{ss} \). This matrix must satisfy \( \mathbf{Q} \mathbf{R}_{ss} \mathbf{Q}^H = \mathbf{R}_{ss} \), where \( \mathbf{Q} = \mathbf{R}_{ss}^{1/2} \mathbf{U} \mathbf{R}_{ss}^{-1/2} \); and \( \mathbf{R}_{ss}^{1/2} \) and \( \mathbf{U} \) are the symmetric square root of \( \mathbf{R} \) and a unitary matrix, respectively. Since \( Q_{qu} \) and \( Q_s \) are two linear transformation groups, the distribution of their each element is Gaussian and also the induced parameter transformation groups maintain the parameters space under each hypothesis as follows. The induced parameter transformation groups under each hypothesis are written as follows:

\[
\begin{aligned}
\overline{Q}_{qu}|_{\mathcal{H}_1} & = \{ \overline{q}_{qu}|_{\overline{q}_{qu,1}}([\sigma^2, \mathbf{\Sigma}_{ww}]) = [\sigma^2, \mathbf{Q}\mathbf{\Sigma}_{ww}\mathbf{Q}^H] \}, \quad \overline{Q}_{qu}|_{\mathcal{H}_0} = \{ \overline{q}_{qu}|_{\overline{q}_{qu,0}}([0, \mathbf{\Sigma}_{ww}]) = [0, \mathbf{Q}\mathbf{\Sigma}_{ww}\mathbf{Q}^H] \}, \\
\overline{Q}_s|_{\mathcal{H}_1} & = \{ \overline{q}_s|_{\overline{q}_{s,1}}([\sigma^2, \mathbf{\Sigma}_{ww}]) = [\xi^2\sigma^2, \xi^2\mathbf{\Sigma}_{ww}] \}, \overline{Q}_s|_{\mathcal{H}_0} = \{ \overline{q}_s|_{\overline{q}_{s,0}}([0, \mathbf{\Sigma}_{ww}]) = [0, \xi^2\mathbf{\Sigma}_{ww}] \}.
\end{aligned}
\]

So the hypothesis is invariant under \( Q \) which is given by the combination of two groups \( Q_{qu} \) and \( Q_s \). To derive a MPI bound, we must determine a maximal invariant statistic for \( G \). Hence first we must
7.1. MIMO RADAR TARGET DETECTION

\[
\sum_{\kappa_1=1}^{N} \cdots \sum_{\kappa_L=1}^{N} \left( \rho_{\kappa_1} + 1 \right)^{-1} D_{\kappa_1} \cdots \left( \rho_{\kappa_L} + 1 \right)^{-1} D_{\kappa_L} \left( \frac{\| R_{ss}^{-1/2} r_{i1} \|^2}{\rho_{\kappa_1} \| R_{ss}^{-1/2} r_{i1} \|^2 + \| R_{ss}^{-1/2} r_{iL} \|^2 + 1} \right) - L \\
\sum_{\kappa_1=1}^{N} \cdots \sum_{\kappa_L=1}^{N} \left( \rho_{\kappa_1} \left( \rho_{\kappa_L} \right)^{-1} D_{\kappa_1} \cdots \left( \rho_{\kappa_L}^{-1} \right)^{-1} D_{\kappa_L} \right) \left( \frac{\| R_{ss}^{-1/2} r_{i1} \|^2}{\rho_{\kappa_1} \| R_{ss}^{-1/2} r_{i1} \|^2 + \| R_{ss}^{-1/2} r_{iL} \|^2 + 1} \right) - L 
\]

\[> \eta_{\text{UMPI}} (7.2)\]

determine a maximal invariant statistic for \( Q_{qu} \). A maximal invariant statistic for \( Q_{qu} \) is given by \( m_{qu} (r_t) = \| R_{ss}^{-1/2} r_t \|^2 \). Using the composition lemma \([10, \text{ch.6 th.2}]\), the maximal invariant statistic with respect to the composition of two groups \( Q_{qu} \) and \( Q_s \), for the \( i \)th cell, is given by \( m^i = \left[ \frac{\| R_{ss}^{-1/2} r_{i1} \|^2}{\| R_{ss}^{-1/2} r_{i1} \|^2}, \cdots, \frac{\| R_{ss}^{-1/2} r_{iL} \|^2}{\| R_{ss}^{-1/2} r_{iL} \|^2} \right]^T \), and then the maximal invariant statistic for \( Q \) is given by \( m^T = [m_1^T, m_2^T, \cdots, m^c_T]^T \). Since \( r_t \)’s are independent, the probability density function (pdf) of \( m \) is directly given by the pdf of \( m^i \)’s. It can be shown that the pdf of \( m^i \) under each hypothesis is given by

\[
f_{m^i}(m^i | H_0) = \Gamma(L) \sum_{\kappa_1=1}^{N} \cdots \sum_{\kappa_L=1}^{N} \left( \frac{B_{\kappa_1, \xi} \cdots B_{\kappa_L, \xi}}{\lambda_{\kappa_1, \xi} \cdots \lambda_{\kappa_L, \xi}} \right) \left( \frac{m^i_1}{\lambda_{\kappa_1, 1}} \cdots + \frac{m^i_{L-1}}{\lambda_{\kappa_L, L-1}} + \frac{1}{\lambda_{\kappa_L, i}} \right)^{-L},
\]

where \( m^i_l \) is the \( l \)th element of \( m^i \), \( \zeta = 0, 1 \) and \( \xi = 1 \) if \( \zeta = 1 \) and \( i = c \) and \( \xi = 0 \) for the other cases. \( \Gamma(\cdot) \) is the Gamma function, \( \lambda_{\kappa, \xi} \) is the \( \kappa \)th eigenvalue of matrix \( R_{ss}^{-1/2} (\xi \sigma^2 R_{ss} + \Sigma_{ww}) R_{ss}^{-1/2} \) and \( B_{\kappa, \xi} = \prod_{p=1, p \neq \kappa}^{N} \frac{1}{1 - \lambda_{\kappa, p}} \). The UMPI test statistic is obtained by constructing the LR of \( m \). Constructing the LR of maximal invariant, the UMPI test rejects \( H_0 \) if the condition (7.2) is satisfied. In this equation \( \eta_{\text{UMPI}} \) is set to \( P_{\text{fa}} \) requirement. Note that \( \lambda_{\kappa, 1} = \sigma^2 + \lambda_{\kappa, 0} \), \( \rho_{\kappa} = \frac{\lambda_{\kappa, 0}}{\sigma^2} \) and \( D_{\kappa} = \prod_{j=1, j \neq \kappa}^{N} \frac{1}{1 - \rho_{\kappa - \rho_j}} \). In fact, \( \rho_{\kappa} \) deals with the ratio of interference in the \( \kappa \)th dimension of interference to scatter and signal effects. Since the UMPI test given by (7.2), depends on \( \rho_{\kappa} \)’s, the UMPI test does not exist. Of course, by assuming \( \rho_{\kappa} \) to be known, the decision
rule in (7.2) gives an MPI bound for evaluating the other invariant detectors. In next subsection, we consider the SIN as the SF and moreover we design a new set of signals based on maximizing the proposed SIN given by the induced maximal invariant.

### 7.1.2 SFET based on SIR

According to the results in the previous subsection, the SIR vector for this problem is given by

\[
\text{SIR} = \left[ \frac{1}{\rho_1}, \ldots, \frac{1}{\rho_N} \right]^T = \left[ \frac{\sigma^2}{P_1(R_{ss}^{-1/2} \Sigma_{uw} R_{ss}^{-1/2})}, \ldots, \frac{\sigma^2}{P_N(R_{ss}^{-1/2} \Sigma_{uw} R_{ss}^{-1/2})} \right]^T,
\]

where \( P_n(\cdot) \) is the \( n \)th eigenvalue of \( (\cdot) \). We define the total SIR by

\[
\text{SIR}_{\text{tot}} = \sum_{n=1}^{N} \frac{1}{\rho_n}.
\]

This function of unknown parameters is zero when the parameters belong to \( \mathcal{H}_0 \) and is positive if they belong to \( \mathcal{H}_1 \); hence this function is an SF for this problem. An SFET is given by comparing the estimate of SF by a threshold, hence the SFET rejects \( \mathcal{H}_0 \) if \( \hat{\text{SIR}}_{\text{tot}} > \eta_{\text{SFET}} \), where \( \eta_{\text{SFET}} \) is set to false alarm requirement. In the following, we provide an estimation of the unknown parameters to derive the SFET statistic.

Consider \( \Sigma'_{uw} \triangleq \Sigma_{uw} \sigma^2 \), then the pdf of observation under the union of \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) is given by

\[
f(r^c; \sigma^2, \Sigma'_{uw}) = \frac{\exp \left( -\frac{1}{\sigma^2} \sum_{l=1}^{L} r_{l}^{cH} (R_{ss} + \Sigma'_{uw})^{-1} r_{l}^{c} \right)}{\pi^{NL} \sigma^{2NL} \det(R_{ss} + \Sigma'_{uw})}. \tag{7.3}
\]

Solving \( \frac{\partial f(r^c; \sigma^2, \Sigma'_{uw})}{\partial \sigma^2} = 0 \) and \( \frac{\partial f(r^c; \sigma^2, \Sigma'_{uw})}{\partial \Sigma_{uw}} = 0 \), we have

\[
\hat{\Sigma}_{uw} = \frac{N}{L} \sum_{l=1}^{L} r_{l}^{cH} (R_{ss} + \Sigma'_{uw})^{-1} r_{l}^{c} - R_{ss}, \tag{7.4}
\]

\[
\hat{\sigma}^2 = \frac{1}{NL} \sum_{l=1}^{L} r_{l}^{cH} (R_{ss} + \Sigma'_{uw})^{-1} r_{l}^{c}, \tag{7.5}
\]
7.1. MIMO RADAR TARGET DETECTION

We cannot find $\hat{\Sigma}_{ww}$ and $\hat{\sigma}^2$ directly, so we propose a recursive method to find a proper solving for (7.4). Consider the following recursive method for $\Sigma_{ww}'$ and $\sigma^2$ by

$$
\hat{\Sigma}_{ww}'^{(k+1)} = \frac{N}{L} \sum_{l=1}^{L} r_l r_l^H \left( \tilde{R}_{ss} + \Sigma_{ww}'^{(k)} \right)^{-1} r_l^c - \tilde{R}_{ss},
$$

(7.6)

$$
\hat{\sigma}^2^{(k)} = \frac{1}{NL} \sum_{l=1}^{L} r_l^c r_l^c r_l^H \left( \tilde{R}_{ss} + \Sigma_{ww}'^{(k)} \right)^{-1} r_l^c
$$

(7.7)

where, $\hat{\Sigma}_{ww}'^{(0)} = I$. Hence, the statistic of SFET is given by replacing the estimations of $\Sigma_{ww} = \sigma^2 \Sigma_{ww}'$ and $\sigma^2$ into the SF.

7.1.3 Optimal Signal Design

We define $A \triangleq R_{ss}^{-1/2} \Sigma_{ww} R_{ss}^{-1/2} / \sigma^2$, the eigenvalues of $A$ provide the SIR vector. So the total SIR is given by $\text{SIR}_{\text{tot}} = \frac{1}{\sigma^2} \text{trace} \{ R_{ss}^{1/2} \Sigma_{ww}^{-1} R_{ss}^{1/2} \} = \frac{1}{\sigma^2} \text{trace} \{ R_{ss}^{-1} \Sigma_{ww}^{-1} \}$. We assume that the transmitting energy of each transmitter is limited to a constant $E$.

Hence, we can describe an optimization problem for signal design as bellow:

$$
\max_{s_1, \ldots, s_K} \text{SIR}_{\text{tot}} \quad \text{s.t.} \quad s_k^H s_k = \begin{cases} 
0 & , \quad k' \neq k, \\
E & , \quad k' = k.
\end{cases}
$$

(7.8)

Based on this maximization problem, the desired signals maximize the total SIR and then maximize the MPI bound. To find the optimal solution (7.8), first we relax the problem by removing the orthogonality condition between signals, then finally we add that condition. A typical strategy for solving this problem is based on the Lagrange
coefficients. Thus, we must maximize

\[ L(s_k) = \text{SIR}_{\text{tot}} + \sum_{k=1}^{K} \mu_k(s_k^H s_k - E) = \frac{1}{\sigma^2} \text{trace}(R_{ss} \Sigma_{ww}^{-1}) + \sum_{k=1}^{K} \mu_k(s_k^H s_k - E), \]  

(7.9)

where \( \mu_k \) is the \( k \)th coefficient of Lagrange method. So the optimal signals will be given by solving \( \frac{\partial L(s_k)}{\partial s_k} = 0 \), i.e.,

\[ \frac{\partial L(s_k)}{\partial s_k} = \frac{1}{\sigma^2} \frac{\partial \text{trace}(R_{ss} \Sigma_{ww}^{-1})}{\partial s_k} + 2\mu_k s_k = 0 \Rightarrow \frac{1}{\sigma^2} 2 \Sigma_{ww}^{-1} s_k + 2\mu_k s_k = 0. \]  

(7.10)

According to the last equation, \( \Sigma_{ww}^{-1} s_k = -\sigma^2 \mu_k s_k \), that means \( s_k \) must be a eigenvector of \( \Sigma_{ww}^{-1} \) i.e., \( s_k = \sqrt{E} u_k \), where \( u_k \) is the \( k \)th eigenvector of \( \Sigma_{ww}^{-1} \). Note that this set of signals is also orthogonal. Now consider the total SIR by

\[ \text{SIR}_{\text{tot}} = \frac{1}{\sigma^2} \text{trace}(R_{ss} \Sigma_{ww}^{-1}) = \frac{E}{\sigma^2} \sum_{k=1}^{K} u_k^H \Sigma_{ww}^{-1} u_k = \frac{E}{\sigma^2} \sum_{k=1}^{K} \delta_k, \]

where \( \delta_k \)'s are the eigenvalues of \( \Sigma_{ww}^{-1} \). Therefore by choosing the first \( K \) \( \delta_k \) (from maximum to minimum of eigenvalues of \( \Sigma_{ww}^{-1} \)) and their corresponding \( u_k \)'s, the optimal set of signals for the MIMO radar is derived such that it maximizes the invariant detection bound. Note that the covariance matrix is an unknown matrix and so it must be estimated by reference cells \( (i \neq c) \), \( \hat{\Sigma}_{ww} = \frac{1}{C-1} \sum_{i=1, r \neq c}^{C} r_i r_i^H \).

In the following we simulate the performance of the tests when the transmitted signals are complex exponential and compare the results when the proposed set of signals is used. In the first simulation, we consider \( s_k(n) = e^{j\omega_k n} \) for \( n = 0, \cdots, N-1 \) and \( k = 1, \cdots, K \), where \( \omega_k = (k-1)\Delta \omega \) and \( \Delta \omega \) is the two by two increment between the transmitter carriers, which assumed \( \Delta \omega = 2\pi/K \) and in another experience.
Figure 7.1: $P_d$ v.s. $P_{fa}$ curves of SFET in comparison with MPI bound for orthogonal narrowband signals and proposed set of signals.

consider the same MIMO system with the proposed set of signals. In this simulation we assume that $\Sigma_{ww}$ is unknown, therefore we have calculated $\hat{\Sigma}_{ww}$ using 120 reference cells, in derivation of the proposed set of signal. In all simulations, we consider $K = N = 10$ and $L = 15$. Fig.7.1 depicts $P_d$ versus $P_{fa}$ curves for the MPI test and the SFET. It is seen that the set of proposed signals has a superior performance in comparison with the orthogonal narrowband MIMO radar.

### 7.2 Sparse Signal Detection and VAD

Detecting a sparse signal in noise is an essential problem in statistical signal processing which has many applications such as in image processing, speech processing, hyperspectral imaging, radar target detection, remote sensing and communication [100–104]. In [105–109] some voice activity detectors exploit the speech signal statistical model as a heavy tailed distribution random process. Moreover, in radar
application, some detectors are proposed for the sparse heavy tailed distributed interference, e.g., [110,111], or for detecting an sparse signal of interest, e.g., [112–116]. In [106, 107], it is shown that, speech signal components have heavy tailed probability density function (pdf) which is often modeled by the Laplacian distribution or generalized Gaussian distribution. Modeling signal components with a sparse heavy tailed distributed is described by rapid changes in short time durations [117] which has many applications such as in the genomic [118], radar imaging [119] and hyperspectral imaging [120–125].

In this section, we consider the problem of detecting a sparse signal modeled by generalized Gaussian distribution with unknown scale parameter in additive white Gaussian noise with unknown noise variance. To deal with unknown parameters, we propose five robust detectors which are CFAR and asymptotically optimal and also we evaluate the optimal performance bound for CFAR detectors. We aim to detect the presence of a sparse signal and jointly estimate the signal and its parameters.

The hypothesis testing and signal model are formulated for the sparse signal detection problem in Subsection 7.2.1. We find the MIG in Subsection 7.2.1 and show that the resulting induced maximal invariant of the proposed MIG is an SF for this problem. To find the SFET, we develop two parametric and three hybrid estimators for SF in Subsections 7.2.2 and Subsection 7.2.3, respectively. The parametric estimators are obtained using the moment method and the MLE approach. To further enhance these estimators, we propose three hybrid estimators which are obtained by combination of the Bayesian and the parametric estimators. The hybrid methods estimate the unknowns iteratively. As illustrated in Figure 7.2, the motivation behind the hybrid estimators is to iteratively use the estimate of unknown parameters
to estimate the unknown random sparse signal and visa-versa. The iterative methods are initialised by the moment estimates of unknown parameters. Thereafter, we re-estimate unknown parameters employing an MLE approach and using the previous estimate of the unknown sparse signal. The sparse signal is estimated using MMSE or MAPE. The SF hybrid estimations are presented in Subsection 7.2.3. We show that, the proposed MLE of SF gives the same result as the EM of the SF, which guarantees the convergence of the iteration. In Subsection 7.2.4, we show that all proposed SFETs are CFAR and the optimal performance bound between all CFAR tests is achieved. In Subsection 7.2.5, the problem is extended to a matrix form of sparse signal and the MIG for this matrix form is derived. Correspondingly the SFETs for this problem are proposed using an extension of SFETs for detecting the vector sparse signal. The simulation results in Section 7.2.6 show that the SFETs perform very close to the optimal bound. Our results for two applications in VAD and multi radar target detection illustrate the efficiency of the proposed methods in practice.

### 7.2.1 Separating Function

In this section, we assume that the sparse signal is modeled by a generalized Gaussian distribution, where its shape parameter is assumed to be less than one. Consider a sparse random vector \( s \) in the presence of a zero mean additive white Gaussian noise \( w \) with unknown variance \( \sigma^2 \) as \( x = s + w = [x_1, \cdots, x_N]^T \), where \( N \) is the length of \( x \), \( x_n \) is the \( n^{th} \) element of \( x \) and \( ^T \) represents the transpose operator. The probability of density functions (pdf) of \( w \) and \( s \) are considered as 

\[
    f_w(w; \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{\|w\|_2^2}{2\sigma^2}\right)
\]

and 

\[
    f_s(s; \alpha) = \left(\frac{\beta}{2\alpha^2(1/\beta)}\right)^N \exp\left(-\frac{\|s\|_2^2}{\alpha^2}\right),
\]

respectively, where \( \| \cdot \|_l \) is the \( l \)-norm operator,
$\beta \in (0, 1]$ is known, $\sigma^2, \alpha > 0$ are unknowns and $s_n$ is the $n^{th}$ element of $s$. We denote the vector of unknowns by $\boldsymbol{\theta} \triangleq [\alpha, \sigma^2]^T$, where $\alpha$ is the scale parameter of $s$ and $\sigma^2$ is the variance of $w$. From the independence of $s$ and $w$, we can write the pdf of $x$ as the following convolution

$$f_x(x; \boldsymbol{\theta}) = \frac{\beta^N}{(a \Gamma(\frac{1}{\beta}) \sqrt{8\pi \sigma^2})^N} \int_{\mathbb{R}^N} e^{-\frac{||x-s||^2}{2\sigma^2}} e^{-\frac{||s||^\beta}{\alpha^\beta}} ds.$$  

(7.11)

In this section, we consider a signal activity detection hypothesis testing problem $H_1 : s \neq 0$ versus $H_0 : s = 0$. Clearly the pdf $f_x(x; \boldsymbol{\theta})$ under $H_0$, is obtained by letting $\alpha \to 0$ in (7.11). Since $f_s(s; 0) = \lim_{\alpha \to 0} f_s(s; \alpha) = \prod_n \delta(s_n)$ where $\delta(\cdot)$ is the Dirac function, $f_x(x; \boldsymbol{\theta})$ under $H_0$ is given by a zero mean Gaussian pdf with variance $\sigma^2$. Hence, the set of unknown parameters under $H_1$ and $H_0$ are defined as $\Theta_1 \triangleq \{\boldsymbol{\theta} = [\alpha, \sigma^2]^T | \alpha > 0 \}$ and $\Theta_0 \triangleq \{\boldsymbol{\theta} = [0, \sigma^2]^T | \sigma^2 > 0 \}$, respectively. Thus, the general family of distributions under the union of $H_1$ and $H_0$ is $\mathcal{P} = \{f_x(x; \boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta \}$, where $f_x(x; \boldsymbol{\theta})$ is given by (7.11) and $\Theta = \Theta_0 \cup \Theta_1$. Based on $\mathcal{P}$, the signal activity detection problem can be written as

$$\begin{cases} 
H_1 : & \boldsymbol{\theta} \in \Theta_1, \\
H_0 : & \boldsymbol{\theta} \in \Theta_0.
\end{cases}$$

(7.12)

This problem is invariant with respect to group $Q = \{q(x) = cP x | c \in \mathbb{R} - \{0\}, P \in \mathcal{P}\}$, where $\mathbb{R}$ is the set of real numbers, and $\mathcal{P}$ is the set of all $N \times N$ permutation matrices. This is because the pdf of $y = q(x)$ belongs to $\mathcal{P}$ and the unknown parameters of the pdf of $y$ remain in $\Theta_0$ and in $\Theta_1$, under $H_0$ and $H_1$, respectively, i.e., the pdf of $y = g(x) = cP x$ is $f_x(y; ||c|\alpha, |c|^2\sigma^2|^T)$, since for all $c \in \mathbb{R} - \{0\}$ and
\( P \in \mathbb{P}, \left[ |c|\alpha, |c|^2\sigma^2 \right]^T \in \Theta, \) the pdf of \( y \) belongs to \( P \) with the transformed unknown parameters \( \overline{q}(\theta) \triangleq \left[ |c|\alpha, |c|^2\sigma^2 \right]^T, \) where \( \overline{q}(\theta) \) is the induced transformation. Clearly \( |c|\alpha \) is zero under \( H_0 \) and \( |c|\alpha > 0 \) under \( H_1 \). Thus \( g(\cdot) \) preserves \( H_0 \) and \( H_1 \), i.e., \( \overline{q}(\theta) \in \Theta_0 \) for all \( \theta \in \Theta_0 \) and \( \overline{q}(\theta) \in \Theta_1 \) for all \( \theta \in \Theta_1 \). Therefore, the induced group of transformations is given by \( \overline{Q} = \{ \overline{q}(\theta) = \left[ |c|\alpha, |c|^2\sigma^2 \right]^T | \theta = [\alpha, \sigma^2]^T \in \Theta, c \in \mathbb{R} - \{0\} \} \).

Now, we show that the induce maximal invariant denoted by \( \rho(\theta) \) is a SF for this problem. An induced maximal invariant of \( \overline{Q} \) is \( \rho(\theta) = \frac{\alpha^2}{\sigma^2} \) because it satisfies two conditions: 1) \( \rho(\overline{q}(\theta)) = \rho(\theta) \), and 2) for all \( \theta_1, \theta_2 \in \Theta, \rho(\theta_1) = \rho(\theta_2) \) yields that there exists a \( \overline{q}(\cdot) \in \overline{Q} \) such that \( \theta_1 = \overline{q}(\theta_2) \). Moreover \( \rho(\theta) \) is an SF, because \( \rho(\theta) = 0 \) for all \( \theta \in \Theta_0 \) and \( \rho(\theta) > 0 \) for all \( \theta \in \Theta_1 \).

In this section, we propose several SFETs using different estimators of \( \rho(\theta) \) under \( H_0 \cup H_1 \). An SFET is a suboptimal detector for (7.12), which is given by comparing the estimation of SF under \( H_0 \cup H_1 \) with a threshold, where the threshold is given by \( P_{fa} \). In Section 7.2.2, we propose two detectors using parametric estimates of \( \alpha \) and \( \sigma^2 \). Since \( s \) is unknown and random, in Section 7.2.3, we propose three hybrid Bayesian and parametric estimation methods to jointly estimate \( s, \alpha \) and \( \sigma^2 \). Correspondingly, these estimators result efficient SFETs where the moment method provides a closed form expression and has lower computational cost, whereas the rest need some numerical iterations and result in improved performance. Figure 7.2 shows the block diagram of the hybrid SFETs. Using an iterative estimation, sparse signal \( s \) and parametric unknowns \( \theta \) are estimated and then the estimations provide an estimation of SF to achieve the SFET in the Comparator block.
7.2. SPARSE SIGNAL DETECTION AND VAD

7.2.2 Parametric Estimations

To develop the SFET, the unknown parameters must be estimated only under $H_1 \cup H_0$. Here, we propose two parametric estimators for deterministic unknown vector $\theta = [\alpha, \sigma^2]^T$ via moment method and the MLE where for the MLE, we show that the MLE and the EM algorithm are equivalent.

**SFET via Method of Moments**

It is known that the estimators using moments are consistent [1]. Thus Theorem 3 of [75] implies that the SFET using such estimates is asymptotically optimal.

All odd moments are zero as the pdfs in $P$ are even with respect to $x_n$. Thus, we need to use at least two even moments of $x_n$ to find the estimation of $\alpha$ and $\sigma^2$. Consider $\widehat{E}(x_n^2) = \frac{1}{N} \|x\|_2^2$ and $\widehat{E}(x_n^4) = \frac{1}{N} \|x\|_4^4$ as the estimates of the second and the fourth moments of $x_n$, where $E(\cdot)$ is the expectation operator. From the pdf of $x$, we have

$$E(x_n^2) = \sigma^2 + a\alpha^2, \quad E(x_n^4) = 3\sigma^4 + 6a\sigma^2\alpha^2 + b\alpha^4,$$  \hspace{1cm} (7.13)

where $a = \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}$ and $b = \frac{\Gamma(5/\beta)}{\Gamma(1/\beta)}$. From (7.13), we obtain $\alpha = \sqrt{\frac{E(x_n^4)-3(E(x_n^2))^2}{b-3a^2}}$ and  

![Figure 7.2: Hybrid estimation of unknown parameters and SFET.](image-url)
\[\sigma^2 = E(x_n^2) - a\alpha^2,\] Replacing \(\frac{1}{N}\|x\|_4^4\) and \(\frac{1}{N}\|x\|_2^2\) for \(E(x_n^4)\) and \(E(x_n^2)\), respectively, we have the following estimators

\[
\hat{\alpha}_M = \sqrt{\left(\frac{1}{N}\|x\|_4^4 - \frac{3}{N^2}\|x\|_2^2\right)^+}, \quad \hat{\sigma}_M^2 = \left(\frac{1}{N}\|x\|_2^2 - a\alpha^2\right)^+,
\]

where \((x)^+ = \max(x, 0)\) for \(x \in \mathbb{R}\). Note that \((\cdot)^+\) is used in the above to ensure that these estimates are non-negative.

The SFET using (7.14) named SFET\(_M\) rejects \(H_0\) if \(T_M(x) \overset{\Delta}{=} \hat{\rho}_M = \frac{\hat{\alpha}_M^2}{\hat{\sigma}_M^2} > \eta_M\), where \(\eta_M\) is set by the probability of false alarm \(P_{fa}\).

**SFET via EM and MLE**

First we apply the EM estimation to estimate unknown parameters and show that the EM is an iterative algorithm leading to the MLE. To derive the EM, we consider \(s\) as the hidden observation vector, hence the expectation and maximization steps are given as follows

**Expectation:** \[Q(\theta; \hat{\theta}_{EMk}) \overset{\Delta}{=} E(\ln(f_{x,s}(x, s; \theta)) | x; \hat{\theta}_{EMk})\]

\[= \int_s \ln(f_{x,s}(x, s; \theta)) f_s(s|x; \hat{\theta}_{EMk}) ds, \quad (7.15)\]

**Maximization:** \[\hat{\theta}_{EMk+1} \overset{\Delta}{=} \arg\max_{\theta \in \Theta} Q(\theta; \hat{\theta}_{EMk}). \quad (7.16)\]
We propose to initialize the EM algorithm by conditional pdf of \( f_k \) where \( \hat{x} \). Applying the Bayes’ rule, the joint pdf of \( f_s(x; \hat{\theta}_{EM}) \) are given by

\[
f_{x,s}(x, s; \theta) = \frac{\beta^N \exp\left(\frac{-||x-s||^2}{2\sigma^2} - \frac{||s||^2}{\alpha^2}\right)}{(2\sqrt{2\pi\sigma^2\alpha\Gamma(1/\beta)})^N},
\]

\[
f_s(s|x; \hat{\theta}_{EM}) = \frac{\exp\left(\frac{-||x-s||^2}{2\sigma^2_{EM}} - \frac{||s||^2}{\alpha^2_{EM}}\right)}{\int_{R^N} \exp\left(\frac{-||x-s||^2}{2\sigma^2_{EM}} - \frac{||s||^2}{\alpha^2_{EM}}\right) ds},
\]

where \( \hat{\theta}_{EM} = [\hat{\alpha}_{EM}, \hat{\sigma}^2_{EM}]^T \), \( \hat{\alpha}_{EM} \) and \( \hat{\sigma}^2_{EM} \) are the EM estimates of \( \theta \), \( \alpha \) and \( \sigma^2 \) at the \( k \)th iteration, respectively. Substituting (7.17) and (7.18) into (7.15), \( Q(\theta; \hat{\theta}_{EM}) \) is given as

\[
Q(\theta; \hat{\theta}_{EM}) = N \ln\left(\frac{\beta}{2\sqrt{2\pi\sigma^2\alpha\Gamma(1/\beta)}}\right) - \int_{R^N} \left( \frac{-||x-s||^2}{2\sigma^2} - \frac{||s||^2}{\alpha^2} \right) e^{-\frac{-||x-s||^2}{2\sigma^2_{EM}} - \frac{||s||^2}{\alpha^2_{EM}}} ds
\]

\[
\int_{R^N} \exp\left(\frac{-||x-s||^2}{2\sigma^2_{EM}} - \frac{||s||^2}{\alpha^2_{EM}}\right) ds.
\]

For (7.16), setting \( \frac{\partial Q(\theta; \hat{\theta}_{EM})}{\partial \alpha} = 0 \), \( \frac{\partial Q(\theta; \hat{\theta}_{EM})}{\partial \sigma^2} = 0 \) provides

\[
\hat{\alpha}_{EMk+1} = \left( \frac{\beta}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} \left[ \frac{(x_n - s_n)^2}{2\sigma^2_{EMk}} - \frac{|s_n|^2}{\alpha^2_{EMk}} \right] e^{-\frac{(x_n - s_n)^2}{2\sigma^2_{EMk}} - \frac{|s_n|^2}{\alpha^2_{EMk}}} ds_n \right)^{1/3},
\]

\[
\hat{\sigma}^2_{EMk+1} = \sum_{n=1}^{N} \left[ \frac{1}{N} \int_{-\infty}^{\infty} \frac{(x_n - s_n)^2}{2\sigma^2_{EMk}} - \frac{|s_n|^2}{\alpha^2_{EMk}} ds_n \right].
\]

We propose to initialize the EM algorithm by \( \hat{\alpha}_{EM0} = \hat{\alpha}_M \) and \( \hat{\sigma}^2_{EM0} = \hat{\sigma}^2_M \) and the halt condition as \( |Q(\hat{\theta}_{EMk+1}; \hat{\theta}_{EM}) - Q(\theta_{EMk}; \theta_{EMk-1})| < \epsilon \), where \( \epsilon \) is a small positive number.

In the following, we show that the MLE of \( \alpha \) and \( \sigma^2 \) involve to solve similar
equations as in (7.20). The MLE of \([\alpha, \sigma^2]\) is obtained by the maximization of the pdf of \(x\) over \(\Theta\):

\[
\arg\max_{[\alpha, \sigma^2]^T \in \Theta} \frac{\beta^N \int_{R^N} \exp\left(\sum_{n=1}^{N} -\frac{1}{2\sigma^2} (x_n - s_n)^2 - \frac{1}{\alpha^2} |s_n|^\beta\right)ds}{(2\alpha \Gamma(1/\beta) \sqrt{2\pi \sigma^2})^N}.
\]

Applying the first order derivative condition, we have

\[
\hat{\alpha}_{\text{ML}}^\beta = \frac{\beta}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} \frac{|s_n|^\beta \exp\left(-\frac{(x_n - s_n)^2}{2\hat{\sigma}_{\text{ML}}^2}\right)}{\frac{|s_n|^\beta}{\hat{\sigma}_{\text{ML}}^\beta}} ds_n,
\]

\[
\hat{\sigma}_{\text{ML}}^2 = \sum_{n=1}^{N} \frac{\int_{-\infty}^{\infty} \frac{(x_n - s_n)^2 \exp\left(-\frac{(x_n - s_n)^2}{2\hat{\sigma}_{\text{ML}}^2}\right)}{\frac{(x_n - s_n)^2}{\hat{\sigma}_{\text{ML}}^2}} ds_n}{N \int_{-\infty}^{\infty} \exp\left(-\frac{(x_n - s_n)^2}{2\hat{\sigma}_{\text{ML}}^2}\right) \frac{|s_n|^\beta}{\hat{\sigma}_{\text{ML}}^\beta} ds_n}.
\]

There is no closed form solutions for (7.21). However, the above expressions are identical with (7.20). Since the EM algorithm for this exponential family of distributions is guaranteed to converge, it shall converge to the fixed point of (7.21), i.e., the EM algorithm converges to the MLE.

The estimation of SF in each iteration is given by \(\hat{\rho}_{\text{ML}}^k = \frac{\hat{\alpha}_{\text{ML}}^2}{\hat{\sigma}_{\text{ML}}^2}\). We can show that if \(\hat{\alpha}_{\text{ML}}^k = 0\) then \(\hat{\alpha}_{\text{ML}}^{k+1} = 0\). In such case no more iteration is required as the EM converges to \(\hat{\alpha}_{\text{ML}} = 0\) and \(\hat{\sigma}_{\text{ML}}^2 = \frac{1}{N} \sum_{n=1}^{N} |x_n|^2\).

We name the resulting test using the MLE (or EM) by SFET_{ML}. This test rejects \(H_0\) if \(T_{ML}(x) \overset{\Delta}{=} \frac{\hat{\alpha}_{\text{ML}}^2}{\hat{\sigma}_{\text{ML}}^2} > \eta_{\text{ML}}\), where \(\eta_{\text{ML}}\) is set to satisfy the false alarm requirement and \(\hat{\alpha}_{\text{ML}}\) and \(\hat{\sigma}_{\text{ML}}^2\) are the converged values in (7.20).

### 7.2.3 Hybrid SFETs

To further improve the estimation of \(\rho(\theta) = \frac{\alpha^2}{\sigma^2}\), we exploit two Bayesian estimators (MMSE and MAPE) of \(s\).
SFET via MMSE

We propose to estimate $\alpha$ and $\sigma^2$ given $\hat{s}_{\text{MS}}$ as the MMSE of $s$. According to the signal model $x = s + w$ and motivated by the expression of the MLE of $\sigma^2$ and $\alpha$, we propose to use the following hybrid estimator

$$\hat{\sigma}^2_{\text{HMS}} = \frac{1}{N} \|x - \hat{s}_{\text{MS}}\|_2^2, \quad \hat{\alpha}_{\text{HMS}} = \left( \frac{\beta}{N} \|\hat{s}_{\text{MS}}\|_2^2 \right)^{\frac{1}{\beta}},$$  \tag{7.22}

where $\hat{\sigma}^2_{\text{HMS}}$ and $\hat{\alpha}_{\text{HMS}}$ denote the hybrid estimation of $\sigma^2$ and $\alpha$ using MMSE, respectively.

The MMSE of $s$ is the conditional mean of $s$ given $x$, i.e., $E(s|x)$ [1]. From the Bayes’ rule, we have

$$f(s|x) = \frac{\exp(\sum_{i=1}^{N} \frac{-(x_i - s)^2}{2\sigma^2} - |s_i|^\beta)}{\int_{\mathbb{R}^N} \exp(\sum_{i=1}^{N} \frac{-(x_i - s_i)^2}{2\sigma^2} - |s_i|^\beta) ds}.$$  \tag{7.23}

Thus, the $n^{th}$ element of $\hat{s}_{\text{MS}}$ is given by

$$\hat{s}_{n\text{MS}} = E(s_n|x) = \frac{\int_{-\infty}^{\infty} s_n \exp(\frac{-(x_n - s_n)^2}{2\sigma^2} - |s_n|^\beta) ds_n}{\int_{-\infty}^{\infty} \exp(\frac{-(x_n - s_n)^2}{2\sigma^2} - |s_n|^\beta) ds_n}.$$  \tag{7.24a}

Note that $\hat{s}_{n\text{MS}}$ depends on $\alpha$ and $\sigma^2$ which are unknown. Therefore, we propose the following iterative method to find $\hat{\alpha}_{\text{HMS}}$ and $\hat{\sigma}^2_{\text{HMS}}$,

$$\hat{s}_{n\text{MS}k+1} = \frac{\int_{-\infty}^{\infty} s_n \exp(\frac{-(x_n - s_n)^2}{2\sigma^2_{\text{HMS}k}} - |s_n|^\beta_{\text{HMS}k}) ds_n}{\int_{-\infty}^{\infty} \exp(\frac{-(x_n - s_n)^2}{2\sigma^2_{\text{HMS}k}} - |s_n|^\beta_{\text{HMS}k}) ds_n},$$  \tag{7.24a}

$$\hat{\alpha}_{\text{HMS}k+1} = \left( \frac{\beta}{N} \|\hat{s}_{\text{MS}k+1}\|_2^2 \right)^{\frac{1}{\beta}},$$  \tag{7.24b}

$$\hat{\sigma}_{\text{HMS}k+1}^2 = \frac{1}{N} \|x - \hat{s}_{\text{MS}k+1}\|_2^2.$$  \tag{7.24c}
where $\hat{s}_{nMSk}$ is the $n^{th}$ element of $\hat{s}_{MSk}$. In (7.24a), we calculate $\hat{s}_{nMSk+1}$ for $n = 1, \cdots, N$, so we have $\hat{s}_{MSk} = [\hat{s}_{1MSk}, \cdots, \hat{s}_{NMSk}]^T$. Similar to the iterative MLE method, we initialize this algorithm by $\hat{\alpha}_{HMS0} = \hat{\alpha}_M$ and $\hat{\sigma}_{HMS0}^2 = \hat{\sigma}_M^2$ and consider $\|\hat{s}_{MSk+1} - \hat{s}_{MSk}\|_2^2 < \epsilon$ for some small $\epsilon > 0$ as the halt condition of (7.24). Thus, the SFET using (7.24) named SFET$_{MS}$ rejects $H_0$ if $T_{MS}(x) \Delta \frac{\hat{\sigma}_{HMS}^2}{\hat{\sigma}_{HMS}^2} > \eta_{MS}$, where $\eta_{MS}$ is set to satisfy the false alarm requirement, $T_{MS}(x)$ denotes the SFET$_{MS}$ statistic, $\hat{\sigma}_{HMS}^2$ and $\hat{\alpha}_{HMS}$ denote the converged value of $\hat{\sigma}_{HMSk}^2$ and $\hat{\alpha}_{HMSk}$ in (7.24).

The iterative method in (7.24) requires a high computational cost for estimating $s$ in (7.24a). In next method, we exploit the sparsity property to reduce the computational cost.

**SFET via MAPE**

For estimating $\alpha$ and $\sigma^2$, we can use the MAPE of $s$ which maximizes (7.23) as

$$\hat{s}_{MA} = \arg\min_{s \in \mathbb{R}^N} \frac{1}{2\sigma^2} \|x - s\|_2^2 + \frac{1}{\alpha^\beta} \|s\|_\beta^\beta. \quad (7.25)$$

Since (7.25) also depends on $\alpha$ and $\sigma^2$, we propose iterative methods to jointly estimate $s$, $\alpha$ and $\sigma^2$. Several techniques exist for solving (7.25). However, they are not applicable here as they require the values of $\alpha$ and $\sigma^2$.

We use $\hat{\alpha}_M$ and $\hat{\sigma}_M^2$ for the initialization and iteratively minimize the objective function $\hat{\text{Obj}}(s) \triangleq \frac{1}{2\hat{\sigma}_M^2} \|x - s\|_2^2 + \frac{1}{\hat{\alpha}_M^\beta} \|s\|_\beta^\beta$ defined in (7.25) in order to improve the estimate of $s$ and thereby we obtain enhanced estimate of $\rho(\theta) = \frac{\alpha^2}{\sigma^2}$. The steepest decent method for $\hat{\text{Obj}}(s)$ often diverges for $\beta < 1$, because the gradient becomes unbounded as any component of $s$ tends to zero. Hence we propose two numerical methods based on fixed-point method and Newton’s method. The first derivative
conditions implies for the optimal solution of (7.25), we either have \( \hat{s}_n = 0 \) or 
\[
\frac{\beta}{\hat{\sigma}_n^2} |\hat{s}_n|^{\beta-1} \text{sgn}(\hat{s}_n) = (\frac{1}{\hat{\sigma}_M} |x_n|) \text{sgn}(x_n).
\]
Thus, the sign of the solution is given by 
\[\text{sgn}(\hat{s}_n) = \text{sgn}(x_n).\]
We propose the following fixed-point iterative method to solve this equation,
\[
|\hat{s}_{nFk+1}^n| = \begin{cases} 
0, & \text{if } |x_n| < x^*, \\
(|x_n| - \frac{\beta \hat{\alpha}_M^2}{\hat{\beta}_M} |\hat{s}_{nFk}|^{\beta-1})^+, & \text{otherwise},
\end{cases}
\]  
(7.26)
where \( \hat{s}_{nFk} \) is the \( n^{th} \) element of \( \hat{s}_{Fk} \) which is the estimation of \( s \) using fixed-point method in the \( k^{th} \) iteration. Initializing with \( \hat{s}_{nF0} = x_n \) where 
\[ s^* = \frac{\Delta}{\beta \hat{\alpha}_M^2} |s|^{\frac{1}{1-\beta}} \]
and \( x^* = s^* + \frac{\beta \hat{\alpha}_M^2}{\hat{\beta}_M} s^{|s|^{\beta-1}}, \) for \( \beta \leq 1 \) and \( x^* = 0 \) for \( \beta > 1. \) This method guarantees exponentially convergence to the optimal solution for \( \beta \leq 1. \) Because \( \zeta(s) = (|x_n| - \frac{\beta \hat{\alpha}_M^2}{\hat{\beta}_M} |s|^{\beta-1})^+ \) is a contraction mapping function and (7.26) is obtained by 
\[ |\hat{s}_{nFk+1}| = \zeta(|\hat{s}_{nFk}|). \]
By virtue of the fixed-point theorem the sequence \( |\hat{s}_{nFk}| \) exponentially converges to the unique fixed-point, i.e., the optimal solution \( \hat{s}_{nF} \) obtained from \( \zeta(|\hat{s}_{nF}|) = |\hat{s}_{nF}|. \) It is easy to show that for \( \beta < 1 \) and \( |x_n| < x^*, \) we have \( \hat{s}_{nk} = 0. \)
Note that for \( \beta = 1, \) this method converges only in one iteration to 
\[ |\hat{s}_n| = |x_n| - \frac{\beta \hat{\alpha}_M^2}{\hat{\beta}_M}. \]
Using MLE of unknown parameters and the fixed-point solution \( \hat{s}, \) the estimation of SF is 
\[ T_F(x) \Delta \frac{(\frac{\beta}{\hat{\sigma}_M^2})^\frac{1}{\beta} \hat{\alpha}_M^2}{\|x - \hat{s}\|^2}. \]

Thus, the test using the proposed fixed-point method named SFET\(_F\) rejects \( H_0 \) if 
\[ T_F(x) > \eta_F, \] where \( \eta_F \) is set by the probability of false alarm \( P_{fa}. \)

As alternative, we minimize \( \hat{\text{Obj}}(\hat{s}) \) using Newton's method after some simple
manipulations as follows

\[
\hat{s}_{n_{MA_{k+1}}} = \hat{s}_{n_{MA_k}} - \mu \frac{\beta \hat{s}_{n_{MA_k}}}{\sigma_M^2} - \frac{(x_n - \hat{s}_{n_{MA_k}})}{\sigma_M^2} \frac{2 - \beta}{\beta (\beta - 1)} \frac{\hat{s}_{n_{MA_k}}^2}{\sigma_M^2} + \frac{1}{\alpha_M} \| \hat{s}_{n_{MA_k}} \|^\beta - \frac{1}{2 \sigma_M^2} |x_n - \hat{s}_{n_{MA_k}}|^2 \]

(7.27)

where \( \mu \in (0, 1] \) is the step size, \( \hat{s}_{MA_k} \triangleq [\hat{s}_{1MA_k}, \ldots, \hat{s}_{NMA_k}]^T \) and \( \hat{s}_{MA_0} = x \) is the initial point where \( Q_{MA_k} \triangleq \frac{1}{2 \sigma_M^2} \| x - \hat{s}_{MA_k} \|^2 + \frac{1}{\alpha_M} \| \hat{s}_{MA_k} \|^\beta \) is the objective function in the \( k \)th iteration. Note that, in contrast to the gradient decent method all involved terms in (7.27) remain bounded and converging due to algebraic simplifications. The halt condition of (7.27) is given by \( |Q_{MA_{k+1}} - Q_{MA_k}| < \epsilon \) for a given positive \( \epsilon \). Assume that \( \hat{s}_{MAC} \triangleq [\hat{s}_{1MAC}, \ldots, \hat{s}_{NMAC}]^T \) is the converged vector of signal using (7.27). Since the objective function is not differentiable at \( s_n = 0 \) for all \( n \in \{1, \ldots, N\} \), then to find the minimum point of \( \hat{Obj}(\hat{s}) \) named \( \hat{s}_{MA} \triangleq [\hat{s}_{1MA}, \ldots, \hat{s}_{NMA}]^T \), we replace \( s_n = 0 \) and \( \hat{s}_{nMAC} \) in \( \hat{Obj}(\hat{s}) \) to find the minimum value of \( \hat{Obj}(\hat{s}) \), i.e.,

\[
\hat{s}_{MA} = \begin{cases} 
0, & \text{if } \frac{1}{2 \sigma_M^2} |x_n - \hat{s}_{nMAC}|^2 + \frac{1}{\alpha_M} |\hat{s}_{nMAC}|^\beta \geq \frac{1}{2 \sigma_M^2} |x_n|^2 \\
\hat{s}_{nMAC}, & \text{otherwise}.
\end{cases}
\]

The hybrid estimator of SF which is \( \rho(\theta) = \frac{\sigma^2}{\sigma^2} \), using MLE of unknown parameters and the MAPE of \( s \) using Newton’s method is \( T_{MA}(x) \triangleq \frac{1}{\alpha_M} \| x - \hat{s}_{MA} \|^\beta \), where \( \hat{s}_{MA} = [\hat{s}_{1MA}, \ldots, \hat{s}_{NMA}]^T \). Thus, the SFET using the proposed Newton’s method named \( SFET_{MA} \) rejects \( H_0 \) if \( T_{MA}(x) > \eta_{MA} \), where \( \eta_{MA} \) is set by \( P_{fa} \).

### 7.2.4 CFAR Property

In this section, first, we show that the proposed detectors are CFAR. Then we find the optimal performance bound of CFAR tests. In order to show the CFAR property
of detectors, we find an MIG for (7.12) then we show that SFETs are invariant with respect to the MIG. An MIG denoted \( Q_m \) is a subgroup of \( Q \), such that for any \( \theta \) and \( \theta' \) in \( \Theta_0 \), there exists only a unique \( \bar{Q}_m(\cdot) \in \bar{Q}_m \), such that \( \theta = \bar{Q}_m(\theta') \), where \( \bar{Q}_m \) is the induced group of transformation associated with \( Q_m \) [126]. Moreover, the optimal CFAR performance bound is given by the LR of a maximal invariant of the MIG as the optimal benchmark for SFETs.

**Minimally invariance property**

Consider a group of transformations defined as \( Q_m = \{ q_m(\cdot)|q_m(x) = \xi x, \xi > 0 \} \). Since \( Q_m \subset Q \), then (7.12) is invariant with respect to \( Q_m \), where \( Q \) is defined in Section 7.2.1. The following proposition implies the invariance property of the proposed SFETs with respect to \( Q_m \).

**Proposition 7** The proposed SFETs are invariant with respect to \( Q_m \).

**Proof 25** See Section D.2

In the following the CFAR property is proven. Any invariant test with respect to \( Q_m \) is a function of \( x \) only through the maximal invariant of \( Q_m \) named \( m(x) \).

In this problem, \( m(x) = \left[ \frac{x_1}{|x_N|}, \cdots, \frac{x_N}{|x_N|} \right]^T \), because from the definition of maximal invariant 1) for any \( \xi > 0, \ m(\xi x) = m(x) \), and 2) if \( m(x) = \left[ \frac{x_1}{|x_N|}, \cdots, \frac{x_N}{|x_N|} \right]^T = \ m(y) = \left[ \frac{y_1}{|y_N|}, \cdots, \frac{y_N}{|y_N|} \right]^T \), then \( x = \xi y \), where \( \xi = \frac{|x_N|}{|y_N|} \). Since any invariant function is a function of \( m(x) \), then the pdf of the statistic of any invariant test with respect to \( Q_m \) depends on the pdf of \( m(x) \). In the following, we show that the pdf of \( m(x) \) under \( H_0 \) does not depend on \( \sigma^2 \). The conditional pdf of \( m(x) \) given \( x_N \) under \( H_0 \) is \( f_m(m|x_N; H_0) = \frac{|x_N|^{N-1}}{(2\pi\sigma^2)^{N-2}} \exp\left(\frac{-x_N^2}{2\sigma^2} \sum_{n=1}^{N-1} m_n^2 \right) \times \delta(m_N - \frac{x_N}{|x_N|}) \). Thus \( f_m(m; H_0) \) is
given by
\[
f_m(m; \mathcal{H}_0) = \int_{-\infty}^{\infty} f_m(m|x_N; \mathcal{H}_0) \frac{e^{-x_N^2/2\sigma^2}}{(2\pi\sigma^2)^{N/2}} dx_N = \frac{\Gamma(N/2)}{2^N \pi^{N/2} \|m\|_2^N}, \tag{7.28}
\]
where \(\Gamma(\cdot)\) is the gamma function. Since \(f_m(m; \mathcal{H}_0)\) does not depend on \(\sigma^2\) and the proposed SFETs are invariant with respect to \(Q_m\) then \(P_{fa}\) for the proposed SFETs does not depend on \(\sigma^2\), i.e., the proposed tests are CFAR.

Optimal CFAR bound

In Section 7.2.1, it is shown that the problem is invariant with respect to \(Q\). Then the induced group with respect to \(Q\) under \(\mathcal{H}_0\) is
\[
\mathcal{Q}_0 = \{q(\cdot)|q([0, \sigma^2]^T) = [0, \xi^2 \sigma^2]^T, [0, \sigma^2]^T \in \Theta_0\}.
\]
Consider \(Q_m = \{q(\cdot)|q(x) = \xi x, \xi > 0\}\), and the corresponding induced group of transformations under \(\mathcal{H}_0\) as \(\mathcal{Q}_m = \{\overline{q}_m([0, \sigma^2]^T) = [0, \xi^2 \sigma^2]^T, [0, \sigma^2]^T \in \Theta_0\}\). For any \(\theta = [0, \sigma^2]\) and \(\theta' = [0, \sigma'^2]\) in \(\Theta_0\), defining \(\xi = \frac{\sigma^2}{\sigma'^2} > 0\), we have \(\theta = \overline{q}_m(\theta')\). Based on the defined \(\xi\), since \(\xi\) is the only positive coefficient satisfying \(\theta = \overline{q}_m(\theta')\), then \(Q_m\) is an MIG for (7.12). In [126], it is shown that the optimal performance bound of all CFAR tests is given by the LR of the maximal invariant with respect to MIG.

As mentioned in Subsection 7.2.4, the maximal invariant with respect to \(Q_m\) is \(m(x) = [\overline{x_1}/|x_N|, \cdots, \overline{x_N}/|x_N|]^T\). In Appendix D.3, it is shown that the pdf of \(m(x)\) under \(\mathcal{H}_1\) is given by (D.3c). The integral in (D.3c) does not have a closed form hence, in the simulations, a Monte Carlo integration method is applied to calculate (D.3c) using generating the random vector \(s\) with generalized Gaussian distribution, with true
parameters $\alpha$ and $\beta$. The optimal CFAR performance bound named MP-CFAR test rejects $H_0$ if $f_{m|H_i}(m; H_0) > \eta_0$ where $f_{m}(m; H_0)$ is given by (7.28) and $\eta_0$ is set to satisfy the false alarm requirement. Note that this test is the UMP-CFAR test when the value of $\alpha/\sigma$ is assumed to be known while in this case this ratio is unknown. Hence this test is not a practical test and we use this test as an optimal bound to evaluate the performance of other suboptimal tests in the simulations. In the simulations, we replace the true value of $\alpha/\sigma$ in the MP-CFAR statistic test to find the optimal bound.

7.2.5 Extension to Matrix Observation

In some applications, the observation is arranged in a matrix structure. Here, we extend our results to matrix of observation and generalize the proposed SFETs in Section 7.2.1 for sparse matrix. Let $X = S + W$ be sparse signal $S$ in additive Gaussian noise $W$, where $X = [x_1 \cdots x_I]$, $S = [s_1 \cdots s_I]$ and $W = [w_1 \cdots w_I]$, the length of $x_i$ is $N$ and $i \in \{1, \cdots, I\}$. Assume that each column is modeled by a nuisance sparse signal as $x_i = s_i + w_i$, where the pdf of $s_i$ is $f_s(s_i, \alpha_i)$. Since $\lim_{\alpha_i \to 0} f_s(s_i, \alpha_i) = \delta(s_i)$, then $\alpha_i = 0$ for all $i$ represents the case $s_i = 0$. So using (7.11), the pdf of observation matrix $X$ is given by

$$f_X(X; \alpha, \sigma^2) = \prod_{i=1}^{I} \frac{\beta^N \int_{R^N} \exp\left(-\frac{1}{2\sigma_i^2} \|x_i - s\|^2 - \frac{1}{\alpha_i} \|s\|^\beta \right) ds}{(2\sqrt{2\pi \alpha_i \sigma_i} \Gamma(1/\beta))^N},$$

where $\alpha \triangleq [\alpha_1, \alpha_2, \cdots, \alpha_I]^T$ and $\sigma^2 \triangleq [\sigma_1^2, \sigma_2^2, \cdots, \sigma_I^2]^T$. Hence, the signal activity hypothesis testing problem of (7.12) is extended by $\Theta_0 = \{(\alpha, \sigma^2) | \alpha = 0, \sigma_i^2 > 0, i \in \{1, \cdots, I\}\}$ and $\Theta_1 = \{(\alpha, \sigma^2) | \alpha \in R^+ \setminus 0, \sigma_i^2 > 0, i \in \{1, \cdots, I\}\}$, where
$\mathbb{R}^{+I}$ is the set of all real vectors with length of $I$ containing non-negative elements. Similar to the first part of Section 7.2.1, this problem is invariant with respect to $Q = \{q(\cdot)|q(x_i) = c_iP_ix_i, c_i \in \mathbb{C}, P_i \in \mathbb{P}, i \in \{1, \cdots, I\}\}$, then the induced group of transformations over the unknown parameters is given by $Q = \{\tilde{q}(\cdot)|\tilde{q}([\alpha_i, \sigma_i^2]^T) = [\vert c_i \vert \alpha_i, \vert c_i \vert ^2 \sigma_i^2]^T, i \in \{1, \cdots, I\}\}$. Thus the induced maximal invariant is given by $\rho(\alpha, \sigma^2) = \left[\frac{\alpha_i^2}{\sigma_i^2}, \frac{\sigma_i^2}{\sigma_i^2}, \cdots, \frac{\alpha_I^2}{\sigma_I^2}\right]^T$. Based on Theorem 2 of [75], any increasing function of $\rho(\alpha, \sigma^2)$ provides an SF. In this problem, we consider the summation of elements of $\rho(\alpha, \sigma^2)$ as an SF, i.e., $SF(\alpha, \sigma^2) = \sum_{i=1}^I \frac{\alpha_i^2}{\sigma_i^2}$. Based on the result of [90], this SF is the asymptotically optimal SF. Applying the estimation methods in Section 7.2.1, we can generalize $SF_{ET}^M, SF_{ET}^{ML}, SF_{ET}^{MS}, SF_{ET}^{MA}$ and $SF_{ET}^{F}$ for the proposed SF($\alpha, \sigma^2$).

The proposed detectors for the matrix model are CFAR. This is because, the pdf of $\frac{\tilde{\alpha}_i^2}{\sigma_i^2}$ does not depend on $\sigma_i^2$, where $\tilde{\alpha}_i$ and $\tilde{\sigma}_i^2$ are any estimates (previously proposed estimation of this section) of $\alpha$ and $\sigma^2$ respectively. Similar to Subsection 7.2.4, the MIG for this problem is $Q_m = \{q(\cdot)|q(X) = [\xi_1x_1, \cdots, \xi_Ix_I], \xi_1 > 0, \cdots, \xi_I > 0\}$. Hence, a maximal invariant for $Q_m$ is $M(X) = [m(x_1), \cdots, m(x_I)]$, where $m(\cdot)$ is the maximal invariant proposed in Subsection 7.2.4. Since $x_i$’s are independent, the pdf of $M(X)$ under $H_0$ and $H_1$, are given by $\prod_{i=1}^I f_m(m_i; H_0)$ and $\prod_{i=1}^I f_m(m_i; \alpha_i, \sigma_i^2H_1)$, where $f_m(m; H_0)$ and $f_m(m; \alpha, \sigma^2, H_1)$ are given by (7.28) and (D.3c) respectively, where $m_i$ is the $i$th column of $M(X)$. Thus the MP-CFAR test for matrix model hypothesis testing problem is given by $\frac{\prod_{i=1}^I f_m(m_i; \alpha_i, \sigma_i^2H_1)}{\prod_{i=1}^I f_m(m_i; H_0)} \lessgtr_{H_0} \eta_0$, where $\eta_0$ is set to satisfy $P_{fa}$.
7.2.6 Simulation Result

The performance of the proposed SFETs and the optimal CFAR bound are evaluated in this section. The Signal-to-Noise-Ratio for (7.12) is
\[
\text{SNR} = \frac{E(s^2)}{E(w_n^2)} = \frac{\Gamma(3/\beta)\alpha^2}{\Gamma(1/\beta)\sigma^2}.
\]
In the first part of simulations, we evaluate the performances for a constant \(\beta\). Figure 7.3a depicts the probability of detection versus the probability of false alarm for \(N = 100, \beta = 1\) and SNR = 0.02 (or -16.98dB). Figure 7.3b shows the probability of miss detection versus the probability of false alarm for \(N = 100, \beta = 1\) and SNR = 0.5 (or -3.01dB). Figure 7.4 shows the detection performance of SFETs versus SNR for \(P_{fa} = 0.01, N = 100, \) and \(\beta = 1\). We observe that SFET\(_{MA}\) has the best performance, SFET\(_{MS}\) is in the second place, and SFET\(_M\) has the worst performance amongst all SFETs. Except SFET\(_M\) and SFET\(_{ML}\), they perform comparably. For \(P_{fa} = 0.01, \beta = 1\) and \(N = 100\), the performance gap between SFET\(_{MA}\) and the optimal MP-CFAR bound is less than 0.3dB in SNR. The performance of SFET\(_M\) is worst, because other detectors make use of enhanced estimation of sparse signal.

In the second part of simulations, we study the convergence speed of the proposed estimators. Figure 7.5 shows the sample mean of estimation of \(\alpha\) using \(\hat{\alpha}_{ML}\) and \(\hat{\alpha}_{HMS}\) calculated over 500 independent runs versus the iteration index. We observe that \(\hat{\alpha}_{ML}\) and \(\hat{\alpha}_{HMS}\) estimators converge in less than 10 iterations for \(\beta = 1\) or \(\beta = 0.1\) and different values of \(\alpha\). Figure 7.5 shows that the speed of convergence increases for smaller value of \(\beta\). Furthermore, the convergence speed of \(\hat{s}_{MAk}\) and \(\hat{s}_{Fk}\) are illustrated in Figure 7.6. The same objective function \(\hat{\text{Obj}}(s)\) which is minimized to find \(\hat{s}_{MAk}\) and \(\hat{s}_{Fk}\) is plotted in Figure 7.6. From Figure 7.6, we observe that after convergence, the lowest objective function is achieved by \(\hat{s}_{MAk}\) for \(\beta = 0.9\) while for \(\beta = 0.7, \hat{s}_{Fk}\) minimized the objective function. In both cases, the \(\beta\)-norm of \(\hat{s}_{MAk}\) is less than \(\hat{s}_{Fk}\).
and the convergence speed of $\hat{s}_{F_k}$ is higher than $\hat{s}_{MA_k}$. Figures 7.5 and 7.6 reveal faster convergence of SFET$_{ML}$ and SFET$_{MS}$ compared to SFET$_{MA}$. From [75], this result justifies the performance gap between these detectors. Moreover, Figures 7.6a and 7.6b show that these algorithms also converge faster as $\beta$ decreases, i.e., as the sparsity of signal increases.

Figure 7.7a depicts the probability of detection versus $\beta$ for $\alpha = 0.1$ and $N = 100$. We observe that the MP-CFAR bound and the performance of SFET$_{MA}$ reduces as $\beta \rightarrow 2$. This is obviously because at $\beta = 2$, the generalized Gaussian becomes a Gaussian distribution. Hence around $\beta = 2$, the noise and signal have very similar family of distributions and thereby the signal does not exhibit distinguishable statistics from the noise and cannot be detected. Figure 7.7b shows the asymptotically optimality of the proposed detectors, because their performance tend to the MP-CFAR bound as the number of observations increases.

Radar Target Detection

In this subsection, we apply the results for detecting radar target detection using sparse matrix detection. Figure 7.8 shows the performance of SFET$_{MA}$ and the MP-CFAR bound for a sparse matrix detection for a target radar detection with 20 cells. We consider three scenarios with 2, 5, and 10 targets. Tacking a 40 point Fast Fourier Transformation over each cell provides a matrix of observations. We only evaluate the extension of the SFET$_{MA}$, since it outperforms other SFETs. This simulation shows the probability of miss detection versus the probability of false alarm.
Voice Activity Detection and Speech Enhancement

Here, we apply SFET\textsubscript{MA} as a VAD in the presence of white stationary additive noise. Since the speech signal is not sparse in time domain [106,109], we first apply Discrete Cosine Transform (DCT) on a zero padded window of signal with a length of 5 milliseconds to convert to a sparse signal, where successive windows have %50 overlap. Let $x_i$ denote the output of the DCT of $i^{th}$ window. The number of zeros padded in each window is equivalent to 2.5 milliseconds.

The proposed method follows the structure in [105] and provides a hard decision output using a desired $P_{fa}$ and a soft detector output using the LR and the Markov model in [105]. The probability of false alarm is set to $10^{-4}$ and the noise variance is estimated using recent intervals detected as silence intervals. The initial hard decision outputs to build the initial noise library is given by

$$y_i = \left( \frac{\hat{\sigma}^2 ||\tilde{s}_{i,\text{MA}}||^2}{\bar{\sigma}^2} \right)^{\frac{1}{2}} \text{H}_1 \geq \eta_{\text{Pre}}, \quad (7.29)$$

where $\eta_{\text{Pre}}$ is set to satisfy the false alarm requirement using the Monte Carlo method under $\text{H}_0$ and from [106,109], we consider $\beta = 1$ and 0.7. The noise library includes 20 most recent DCT vectors detected as speech free signal. Note that (7.29) is SFET\textsubscript{MA} while the estimate of the noise variance $\hat{\sigma}^2$ is obtained from the noise library. This hard detector does not exploit the collected information from the adjacent windows. To further enhance this detector following [105], we propose a soft VAD which estimates the probability of speech be active signal within $i^{th}$ window named $P_i$. Assuming $q_i$ be a priori probability of speech be active, we have

$$P_i = \frac{L_i(y_i)q_i}{L_i(y_i)q_i + (1-q_i)}, \quad q_{i+1} = \Pi_{01}(1 - P_i) + \Pi_{11}P_i, \quad (7.30)$$
where $L_i(y_i)$ is the LR of $y_i \overset{A}{=} \left( \frac{\hat{\omega}}{\hat{\sigma}} \| \hat{s}_{MA} \|_2^2 \right)^{\hat{k}_1}$. In our simulation, we set the transition probabilities at $\Pi_{01} = 0.2$ and $\Pi_{11} = 0.8$. Experimentally, we observed that the data in (7.29) has a pdf similar to gamma pdf. Thus, we propose to estimate the LR $L_i(y_i)$ from the decision statistic in (7.29) using gamma pdf as

$$L_i(y_i) = \frac{\Gamma(\hat{k}_0) \hat{\varrho}_{0}^{\hat{k}_1} \exp(y_i/\hat{\varrho}_1)}{\Gamma(\hat{k}_1) \hat{\varrho}_{0}^{\hat{k}_1} \exp(y_i/\hat{\varrho}_0)}, \tag{7.31}$$

where $\hat{\varrho}_0$, $\hat{\varrho}_1$, $\hat{k}_0$ and $\hat{k}_1$ are MLE estimates and are (can be updated once in a while) from two libraries of 50 milliseconds each, one containing 20 DCT vectors of silence intervals and the other containing 20 active intervals. Let $y_0$ and $y_1$ be the concatenated vectors of the silence and active libraries, respectively. The LR in (7.31) is calculated using the MLEs $\hat{\varrho}_p = \left\| y_p \right\|^2$ and $\hat{k}_p = \frac{1}{l_p} y_p l_p$, where $l_p$ is the length of $y_p$.

Figure 7.9 shows the outputs of (7.29) and $P_i$ in (7.30) as well as the cleaned signal $\hat{s}_i$ with input at two SNRs -5.5dB and 3.46dB. In this part, we calculate the input SNR by $\frac{\| s \|^2}{\text{length of } s}$, where $s$ is the clean speech signal over the total duration of simulation, $\sigma^2$ is the variance of noise and length of $s$ is the number of samples of the clean speech signal which are detected to have active voice. We obtained the length of $s$ by employing the proposed method on the noise free signal. To evaluate the ability of the proposed method in signal enhancement we define $\text{SE}_s \overset{A}{=} \frac{\| s - \hat{s}_{MA} \|^2}{\text{length of } s}$ as the squared error of the estimated signal per sample. Our results show $\text{SE}_s = -38.9149$dB using $\beta = 1$ at SNR = 3.4dB. Moreover, using $\beta = 0.7$ results in $\text{SE}_s = -40.7656$dB which exhibits some improvement. Note that the proposed algorithm converges with a single iteration for $\beta = 1$ where as for $\beta = 0.7$ requires a small number of iterations. Figures 7.9b, d, f and h show similar trends for SNR = -5.5dB exhibiting $\text{SE}_s = -34.7550$dB and $\text{SE}_s = -36.3821$dB for $\beta = 1$ and $\beta = 0.7$, respectively.
7.2. SPARSE SIGNAL DETECTION AND VAD

Figure 7.3: a) Performance comparison of the SFETs and MP-CFAR versus $P_{fa}$, for $\alpha = 0.1, \sigma^2 = 1, N = 100$. b) Probability of miss detection of the SFETs and MP-CFAR versus $P_{fa}$, for $\beta = 1, \alpha = 0.5, \sigma^2 = 1, N = 100$.

Figure 7.4: Detection performance of the SFETs and MP-CFAR versus SNR, for $P_{fa} = 0.01, N = 100$ and $\beta = 1$. 
7.2. SPARSE SIGNAL DETECTION AND VAD

Figure 7.5: Sample mean of the estimation of $\alpha$ using (7.21) and (7.24) versus the number of iterations for $\alpha = 0.5$ and $\alpha = 1$.

Figure 7.6: The objective function $\hat{\text{Obj}}(s)$ and the $l_\beta$ norm of sparse signal estimator using (7.26) and (7.27) versus the number of iterations, a. for $\alpha = 0.2, \beta = 0.7$ and $N = 100$ and, b. for $\alpha = 1, \beta = 0.9$ and $N = 100$. 
Figure 7.7: a. Performance evaluation of the MP-CFAR test and SFET\textsubscript{MA} versus $\beta$ for $\alpha = 0.1$. b. Probability of detection versus the number of observations for $P_{fa} = 0.01$, $\alpha = 1$ and $\beta = 1.5$.

Figure 7.8: Probability of miss detection versus the probability of false alarm for SFET\textsubscript{MA} for detecting a sparse vector generated with a 40 point FFT of a received target signal when the number of targets are 2, 5 and 10, number of cells is 20.
Figure 7.9: Hard and Soft detection process for $\beta = 1$, sampling frequency is 48000, (a) input clean signal, (b) noisy signal at SNR = $-5.5$dB, (c) noisy signal at SNR = $3.46$dB, (d) decision statistic $y_t$ out put of (7.29) for SNR = $-5.5$dB (e) $y_t$ out put of (7.29) for SNR = $3.46$dB, (f,g) soft decision (7.30), (h,i) enhanced speech signal $\hat{s}_{IMA}$. 
7.3 Signal Detection in Uniform Linear Antennas

The narrowband signal detection problem using a linear array of antennas is addressed in this section. We assume that all parameters such as the signal amplitude, phase, frequency, Direction-of-Arrival (DoA) and noise variance are unknown. The criterion to find a detector is the Neyman-Pearson criterion. The optimal detector in this criterion maximizes the probability of detection $P_d$ subject to the Probability of false alarm $P_{fa}$ is less than a desired value. In [10], it is shown that the optimal detector using the Neyman-Pearson is given by comparing the LR with a threshold while the threshold satisfies $P_{fa}$. Unfortunately, in narrowband signal detection problem with unknown parameters the optimal detector depends on the unknown parameters [36], hence we cannot use the LR detector in this problem. Our approach to find a suboptimal detector is the SFET [75] and the GLRT.

In this section, we derive the AOSF using the FIM of unknown parameters and the invariance property of the hypothesis testing problem. It is shown that, any SFET using AOSF named AOSFET, provides an asymptotically optimal SFET. The AOSFET maximizes the probability of detection when the number of observations tends to infinity [90]. To derive the AOSFET, we estimate the unknown parameters and then replace in to AOSF. In this section we apply MLE and also propose another estimator called Outlier Protector MLE (OP-MLE). The MLE is a parametric estimation method applied for DoA estimation. The MLE based methods have a high computational complexity for estimating the DoA. To decrease the complexity of MLE, the the eigenvalue analysis of covariance matrix is proposed. Although the eigenvalue based estimators provide a less complexity, the MLE generally outperforms the eigenvalue based estimators [127]. In this section, the OP-MLE estimates the DoA
and the frequency of received signal while the proposed estimator removes the outliers and also provides a better performance comparing with MLE while the search region is smaller. As a benchmark, we compare the performance of the proposed AOSFETs with the GLRT.

In this section, a group of transformations called $Q$ is achieved such that the hypothesis testing problem is invariant with respect to $Q$. Then, using the invariance property of the hypothesis testing, a maximal invariant is applied to remove the unknown parameter under null hypothesis and preserve the maximum information of observations [86]. It is shown than, after applying the maximal invariant, all unknown parameters shrink into a function of unknown parameters called induced maximal invariant which is known under null hypothesis. Then, an AOSF is proposed using the Fisher information function of the induced maximal invariant. The invariance property plays a significant role in signal processing. Using this property, we can find robust statistics which are invariant with respect to the variation of unknown parameters.

### 7.3.1 Signal Model and Invariant Groups

Consider the received signal in the $n$th snapshot in a linear array with $M$ receiver antennas as $\mathbf{x}(n) = [x_0(n), \ldots, x_{M-1}(n)]^T$, where $x_m(n)$ is the sample of $m$th antenna at $n$th snapshot and $^T$ is the transpose operator. In the presence of a narrowband signal $\mathbf{x}(n)$ is given by $\mathbf{x}(n) = \mathbf{w}(n) + \mathbf{s}(n)$, where $\mathbf{s}(n) = A e^{j\omega n T_s} [1, e^{-j\omega \tau}, \ldots, e^{-j\omega (M-1)\tau}]^T$, $A$ is a complex amplitude, $\omega \in (-\pi/T_s, \pi/T_s)$ is the carrier frequency of the target signal, and $T_s$ is the period of sampling. The time delay at the $m$th receiver is given by $(m-1)\tau$. We define the time delay between two first elements by $\tau = l \cos(\phi)/c$, where
c is the speed of wave propagation, $\phi \in (0, \pi)$ is the DoA of target signal, $l \in [0, \lambda/2]$ is the distance between first and second antenna and $\lambda$ is the wavelength of receiving signal. In this section, we assume that the distance of adjacent antennas are identical. Moreover, the additive noise in the signal model $w(n)$ is a white complex Gaussian noise with zero mean and covariance matrix $\sigma^2 I_M$, i.e., $w(n) \sim \mathcal{N}(0, \sigma^2 I_M)$, and $I_M$ is the $M \times M$ identical matrix. Concatenating the snapshots in a matrix structure, we have $X = [x(0), \cdots, x(N - 1)]$. Hence the probability density function (pdf) of $X$ is given by

$$f_X(X; \theta) = \frac{\exp \left( -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left| x_i(n) - A e^{j\omega T_s} e^{-j\vartheta m} \right|^2 \right)}{(\pi \sigma^2)^{MN}},$$

(7.32)

where $\vartheta \overset{\Delta}{=} \omega \tau = 2\pi \cos(\phi) \frac{1}{\lambda} \in [-\pi, \pi)$ and the vector of unknown parameters is defined by $\theta = [A, \omega, \vartheta, \sigma^2]^T$. The target detection hypothesis testing problems using the observation matrix $X$ is given by $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, where $\Theta_1 = \{ \theta = [A, \omega, \vartheta, \sigma^2]^T | A \in \mathbb{C} - \{0\}, \omega \in (-\pi/T_s, \pi/T_s), \vartheta \in \left( -\frac{\alpha \pi}{2}, \frac{\alpha \pi}{2} \right), \sigma^2 > 0 \}$ and $\Theta_0 = \{ \theta = [A, \omega, \vartheta, \sigma^2]^T | A = 0, \omega \in (-\pi/T_s, \pi/T_s), \vartheta \in (-\pi, \pi), \sigma^2 > 0 \}$ and $\mathbb{C}$ is the set of complex numbers. The latest inequality is achieved by this fact that $l$ is less than half of wavelength to prevent the ambiguity in the phase measuring of the achieved signal.

In the following, the general group of transformations are proposed for this problem. The Hypothesis testing problem is invariant with respect to

$$Q = \{ q(\cdot) | q(X) = c(A \odot X), c \in \mathbb{C}, [A]_{n,m} = e^{j(\omega \tau T_s - \vartheta m)} \}. \quad (7.33)$$
7.3. SIGNAL DETECTION IN UNIFORM LINEAR ANTENNAS

To show the invariance property of problem with respect to the proposed group $Q$, we prove that, for any pdf of $X$ in $\Theta_i$, the pdf of $Y = q(X)$ belongs to $\Theta_i$ for $i = 0, 1$. Assume that the pdf of $X$ giving in (7.32) has the parameters $\theta = [A, \omega, \vartheta, \sigma^2]^T$, where if $\theta \in \Theta_0$ then $A = 0$ else $A \neq 0$. Hence, the pdf of $Y = q(X)$ is $f_X(Y, \theta)$, where $\theta = [cA, \mod_{2\pi}(T_s(\omega + \omega_d))^{T_s}, \mod_{2\alpha \pi}(\vartheta + \vartheta_d), |c|^2 \sigma^2]^T$, where $\mod_b(\cdot)$ returns the remainder after division by $b$. Clearly $\theta \in \Theta_0$ if $A = 0$ and $\theta \in \Theta_1$ if $A \neq 0$. The induced group of transformation denoted $Q$ is the set of all transformations from $\theta$ to $\bar{\theta}$. Hence we have

$$Q = \left\{ \bar{\theta}(\cdot)|\bar{\theta}([A, \omega, \vartheta, \sigma^2]^T) = [cA, \mod_{2\pi}(T_s(\omega + \omega_d))^{T_s}, \mod_{2\alpha \pi}(\vartheta + \vartheta_d), |c|^2 \sigma^2]^T; \right\}$$

$$A, c \in \mathbb{C}; \omega, \omega_d \in (-\pi^{T_s}, \pi^{T_s}); \vartheta, \vartheta_d \in (-\frac{\alpha \pi}{2}, \frac{\alpha \pi}{2}); \sigma^2 > 0 \right\}.$$

The induced maximal invariant of this problem denoted by $\rho = \rho(\theta)$ is the maximal invariant of $\theta$ with respect to $Q$. In general a maximal invariant of $\theta$ is a function of $\theta$ such that satisfies two conditions, first, for any $\bar{\theta} \in Q$, $\rho(\bar{\theta}(\theta)) = \rho(\theta)$ and second condition is that, if for any two $\theta, \theta'$, $\rho(\theta) = \rho(\theta')$, then there exists a $\bar{\theta} \in Q$, such that $\theta' = \bar{\theta}(\theta)$. For $Q$ in (7.34), the induced maximal invariant is given by

$$\rho(\theta) = \frac{|A|^2}{\sigma^2}.$$

the proposed induced maximal invariant meets two conditions mentioned above.
7.3. SIGNAL DETECTION IN UNIFORM LINEAR ANTENNAS

7.3.2 Asymptotically Optimal SF

The algorithm to find the Asymptotically Optimal SF has three steps which are summarized as follows [90]:

1) Finding the FIM of \( \theta \); called \( J(\theta) \),

2) Finding the Fisher Information function of \( \rho \); called \( J(\rho) \) using \( J(\rho) = \Delta^T J(\theta) \Delta \), where \( \Delta \) is the Jacobian matrix of \( \theta \) with respect to \( \rho(\theta) \),

3) Solving \( \frac{\partial \varrho(\rho)}{\partial \rho} = \sqrt{J(\rho)} \) to find \( \varrho(\rho) \),

4) The Asymptotically Optimal SF is given by \( |\varrho(\rho) - \varrho(\rho_0)|^2 \), where \( \rho_0 = \rho(\theta_0) \) and \( \theta_0 \in \Theta_0 \).

First step is finding \( J(\theta) \). From (7.32), the FIM of \( \theta \) is given by

\[
J(\theta) = \begin{bmatrix}
\frac{1}{\sigma^2} & -\frac{(N-1)T_A A_I}{\sigma^2} & \frac{(M-1)A_I}{\sigma^2} & 0 \\
-(N-1)T_A A_I & \frac{(N-1)(2N-1)T_A^2 |A|^2}{3\sigma^2} & (M-1)T_s l A_R & 0 \\
\frac{(M-1)A_I}{\sigma^2} & (M-1)T_s l A_R & \frac{(M-1)(2M-1)|A|^2}{3\sigma^2} & 0 \\
0 & 0 & 0 & \frac{1}{\sigma^2}
\end{bmatrix},
\]

where \( A_R \) and \( A_I \) are real and imaginary parts of \( A \). In the second step, we calculate the fisher information function of \( \rho \) as follows. The Jacobian matrix \( \theta \) with respect to \( \rho \) named \( \Delta \) is given by \( \Delta = \left[ \frac{\sigma^2}{2\Delta} , 0 , 0 , 0 \right]^T \). Hence the Fisher information function of \( \rho \) is achieved by \( J(\rho) = \Delta^T J(\theta) \Delta = \frac{MN}{4\rho} \).

Based on the third step, \( \varrho \) is given by solving \( \frac{\partial \varrho(\rho)}{\partial \rho} = \sqrt{J(\rho)} = \sqrt{\frac{MN}{4\rho}} \) which is \( \varrho(\rho) = \sqrt{MN\rho} + k \), where \( k \) is the integral constant. Then the Asymptotically optimal SF using the forth step is AOSF = \( |\varrho(\rho) - \varrho(\rho_0)| = |\sqrt{MN\rho} - \sqrt{MN\rho_0}|^2 = MN|\rho|, \)
where $\rho_0 = \rho(\theta_0) = 0$ since $\theta \in \Theta_0$.

In this section, using the proposed AOSF in the previous section, an AOSFET is achieved. The AOSFET is a test given by replacing an estimation of $\rho$ in to AOSF which is called $\hat{\text{AOSF}}$ and comparing $\hat{\text{AOSF}}$ with a threshold as follows

$$\hat{\text{AOSF}} = M N |\hat{\rho}| \nabla_{H_1} \geq H_0 \eta,$$

(7.37)

where $\eta$ is set to satisfy the required probability of false alarm.

In this section, we apply two estimators for $\rho$ which are MLE and an Outlier Protector MLE (OP-MLE). In the following, it is shown that the MLE is given by an exhaustive search to find the maximum of the absolute of the two dimension Discrete Fourier Transformation (DFT) of $X$, since $X$ is a noisy signal, the exhaustive search may pick an outlier having the maximum value of the absolute of the two dimension DFT of $X$. This kind of error motivates proposing another method called OP-MLE in this section. The OP-MLE is a smart MLE improving the performance of MLE based on exploiting an outlier remover method.

To find the MLE, taking the derivative of $f_X(X; \theta)$ with respect to $\theta$ and setting equal to zero, the MLE of unknown parameters are given as follows

$$[\hat{\omega}, \hat{\vartheta}]^T = \arg \max_{\omega, \vartheta} |\mathcal{X}(\omega T_s, -\vartheta)|^2$$

(7.38)

$$\hat{A} = \frac{1}{MN} \mathcal{X}(\hat{\omega} T_s, -\hat{\vartheta})$$

(7.39)

$$\hat{\sigma}^2 = \frac{1}{MN} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} |x_i(n) - \hat{A} e^{j\hat{\omega} n T_s e^{-j\hat{\vartheta} m}}|^2$$

(7.40)

where $\mathcal{X}(\omega, \vartheta) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x_m(n) e^{-j\omega n} e^{-j\vartheta m}$ is a two dimension DFT of $X$. Hence, the AOSFET using MLE called AOSFET\text{MLE} is given by $MN |\hat{A}| \nabla_{H_1} \geq \hat{\sigma}^2 \nabla_{H_0} \eta_{\text{MLE}}$, where $\eta_{\text{MLE}}$
is set to satisfy the false alarm requirement.

In the following, we propose the OP-MLE to improve the performance of MLE using an outlier remover method. The main idea in this section is that, using a pre-estimation of $\omega$ and $\vartheta$, we find a subset of all possible pair of $(\omega, \vartheta)$ to restrict the search region in (7.38). The new search region is considered as $\mathcal{A} = [\tilde{\omega}_{\text{OP}} - \delta_{\omega}, \tilde{\omega}_{\text{OP}} + \delta_{\omega}] \times [\tilde{\vartheta}_{\text{OP}} - \delta_{\vartheta}, \tilde{\vartheta}_{\text{OP}} + \delta_{\vartheta}]$, where $\tilde{\omega}_{\text{OP}}$ and $\tilde{\vartheta}_{\text{OP}}$ are the pre-estimation of $\omega$ and $\vartheta$ respectively. $\delta_{\omega}$ and $\delta_{\vartheta}$ are two guard bands to deal with the inaccuracy of $\tilde{\omega}_{\text{OP}}$ and $\tilde{\vartheta}_{\text{OP}}$. The pre-estimation $\tilde{\omega}_{\text{OP}}$ is given by taking an average over the estimation of $\omega$ using each antenna and $\tilde{\vartheta}_{\text{OP}}$ is achieved by taking an average over the estimation of $\vartheta$ using two by two adjacent antennas. The proposed OP-MLE of $\rho$ and its corresponding detector using the proposed AOSF are illustrated as follows:

1) Estimating $\hat{\vartheta}_{\text{OP}}_m$ using $x_m(n)$ and $x_{m+1}(n)$, i.e.,

$$e^{j\hat{\vartheta}_{\text{OP}}_m} = \hat{R}_{m,m+1} \triangleq \frac{1}{N} \sum_{n=0}^{N-1} x_m(n)x^*_{m+1}(n),$$

where * shows the conjugate operator and $\hat{R}_{m,m+1}$ shows the estimation of cross-correlation of received signal in the $m$th and $m+1$th antennas. Note that $\hat{\vartheta}_{\text{OP}}_m$ is given by a “one-to-one inverse tangent” which is an inverse tangent as $\hat{\vartheta}_{\text{OP}}_m = \tan^{-1}(\frac{\text{Re}(\hat{R}_{m,m+1})}{\text{Im}(\hat{R}_{m,m+1})})$ while we consider the sign of $\text{Re}(\hat{R}_{m,m+1})$ and $\text{Im}(\hat{R}_{m,m+1})$, i.e., $\hat{\vartheta}_{\text{OP}}_m$ is in the first quadrant if $\text{Re}(\hat{R}_{m,m+1})$ and $\text{Im}(\hat{R}_{m,m+1})$ are positive, in the second quadrant if $\text{Re}(\hat{R}_{m,m+1})$ is negative and $\text{Im}(\hat{R}_{m,m+1})$ is positive, in the third quadrant if $\text{Re}(\hat{R}_{m,m+1})$ is negative and $\text{Im}(\hat{R}_{m,m+1})$ is negative and finally in the forth quadrant if $\text{Re}(\hat{R}_{m,m+1})$ is positive and $\text{Im}(\hat{R}_{m,m+1})$ is negative. Then proposing $\tilde{\vartheta}_{\text{OP}} = \frac{1}{M-1} \sum_{m=0}^{M-2} \hat{\vartheta}_m$ as a pre-estimation of $\vartheta$. 
2) Estimating $\hat{\omega}_{\text{OP}}$ using $x_m(n)$, i.e., $\hat{\omega}_{\text{OP}} = \arg \max_{\omega \in (-\pi Ts, \pi Ts)} \sum_{n=0}^{N-1} x_m(n)e^{-j\omega nTs}$ and proposing $\bar{\omega}_{\text{OP}} = \frac{1}{M-1} \sum_{m=0}^{M-1} \hat{\omega}_m$.

3) Defining a subset of all possible $\omega$ and $\vartheta$ called $A$ to eliminate the outliers which are not close to the true value of $\omega$ and $\vartheta$ as $A = [\bar{\omega}_{\text{OP}} - \delta \omega, \bar{\omega}_{\text{OP}} + \delta \omega] \times [\bar{\vartheta}_{\text{OP}} - \delta \vartheta, \bar{\vartheta}_{\text{OP}} + \delta \vartheta]$. Then estimating the unknown parameters as follows,

$$[\hat{\omega}_{\text{OP}}, \hat{\vartheta}_{\text{OP}}]^T = \arg \max_{(\omega, \vartheta) \in A} |X(\omega Ts, -\vartheta)|^2 \quad (7.41)$$

$$A_{\text{OP}} = \frac{1}{MN} X(\hat{\omega}_{\text{OP}} Ts, -\hat{\vartheta}_{\text{OP}}) \quad (7.42)$$

$$\hat{\sigma}^2_{\text{OP}} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |x_i(n) - A_{\text{OP}} e^{j\hat{\omega}_{\text{OP}} nTs} e^{-j\hat{\vartheta}_{\text{OP}} m}|^2 \quad (7.43)$$

The OP-MLE of $\rho$ is $\frac{|A_{\text{OP}}|^2}{\sigma^2_{\text{OP}}}$.

The AOSFET using OP-MLE called AOSFET$_{\text{OP}}$ is given by $MN|\frac{A_{\text{OP}}}{\sigma^2_{\text{OP}}} H_1|_1 \geq \eta_{\text{OP}}$, where $\eta_{\text{OP}}$ is set to satisfy the false alarm requirement. The proposed OP-MLE exploits smart removing outliers and decreases the exhaustive search region which improves the complexity of OP-MLE comparing with MLE. Moreover, since we search to find $\omega$ and $\vartheta$ in a restricted region provided by a pre-estimation of the pair $(\omega, \vartheta)$, most likely the OP-MLE provides a more accurate estimation, while the MLE may pick an outlier which is not close to the true values of $\omega$ and $\vartheta$. The simulation results show that AOSFET$_{\text{OP}}$ outperforms AOSFET$_{\text{MLE}}$ and the GLRT. The GLRT is given by replacing the MLE of unknown parameters under each hypothesis in to the LR. We can show that the MLE of unknown parameters under $H_1$ named are given by (7.38), (7.39) and (7.40), while the MLE of $\sigma^2$ under $H_0$ is $\hat{\sigma}^2_0 = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |x_m(n)|^2}{MN}$. Hence, The GLRT rejects $H_0$ if $\frac{f_X(X;\hat{\theta})}{f_X(X;\bar{\theta})} > \eta_{\text{GLRT}}$, where $\hat{\theta} = [\hat{A}, \hat{\omega}, \hat{\vartheta}, \hat{\sigma}^2]^T$, $\bar{\theta}_0 = [0, \omega, \vartheta, \hat{\sigma}^2_0]^T$ and $\eta_{\text{GLRT}}$ is set to satisfy the probability of false alarm.
7.3. SIGNAL DETECTION IN UNIFORM LINEAR ANTENNAS

7.3.3 Simulations

Consider a linear array of antennas with $M = 10$ and $N = 100$. In Figure 7.10, the probability of detection versus the probability of false alarm for AOSFET\(_{\text{MLE}}\), AOSFET\(_{\text{OP}}\) and the GLRT are considered, when $A = 0.1$ and $\sigma^2 = 1$ (or equivalently $\rho = 0.01$) and $\phi$ and $\omega T_s$ are considered as $\frac{\pi}{4}$ and $\frac{\pi}{3}$ respectively. Figure 7.10 shows the GLRT and AOSFET\(_{\text{MLE}}\) are equivalent and the AOSFET\(_{\text{OP}}\) outperforms GLRT and AOSFET\(_{\text{MLE}}\). In the second simulation the probability of miss detection versus the number of snapshots $N$ is considered. Figure 7.10 shows the asymptotically optimality of AOSFET\(_{\text{MLE}}\) and AOSFET\(_{\text{OP}}\) and GLRT. In this simulation we consider $\rho = 0.04$ and $P_{fa} = 0.01$ and the rest of parameters are considered same as the previous simulation. This simulation shows the miss detection of all detectors tends to zero and the AOSFET\(_{\text{OP}}\) provides a quicker convergence comparing with AOSFET\(_{\text{MLE}}\) and GLRT. Moreover this simulation show the GLRT and AOSFET\(_{\text{MLE}}\) are equivalent. In Figure 7.12, the Mean Square Error (MSE) of estimation $\rho$ using MLE and OP-MLE versus the number of snapshots is studied. It is seen that, the OP-MLE provides a less MSE comparing with MLE, which is the main reason of the better performance of AOSFET\(_{\text{OP}}\).
7.3. SIGNAL DETECTION IN UNIFORM LINEAR ANTENNAS

Figure 7.10: Probability of detection versus probability of false alarm for AOSFET_{MLE}, AOSFET_{OP} and the GLRT. The simulation parameters are $\rho = 0.01$, $\phi = \frac{\pi}{4}$ and $\omega T_s = \frac{\pi}{3}$.

Figure 7.11: Probability of missed-detection versus the number of snapshots for AOSFET_{MLE}, AOSFET_{OP} and the GLRT. The simulation parameters are $\rho = 0.04$, $\phi = \frac{\pi}{4}$, $\omega T_s = \frac{\pi}{3}$ and $P_{fa} = 0.01$. 
Figure 7.12: MSE versus the number of snapshots for AOSFET_{MLE}, AOSFET_{OP} and the GLRT. The simulation parameters are $\rho = 0.04$, $\phi = \frac{\pi}{4}$ and $\omega T_s = \frac{\pi}{3}$.
Chapter 8

Conclusion and Future Work

In the first part, we studied the relationship between the UMP test and the MVUE. In particular, we showed that the UMP test in a 1D, one-sided problem is performed by comparing the MVUE of the unknown parameter with a threshold. Then, we extended this result for a general class of problems with one or multiple unknown parameters, i.e., we proved that the UMPU test (if it exist) is performed by comparing the MVUE of a corresponding SF with a threshold. The SF is a function which maps the unknown parameter space to two disjoint real subsets. We proved that the SF exists for a wide class of problems which includes most of engineering applications. For the cases where the MVUE does not exist, we proposed a new suboptimal detection approach namely SFET which is easy to derive. This approach is to find an accurate estimate of the SF and use it as the decision statistics. Our results illustrate that improving the accuracy of this estimator results in improvement in the detection performance. Simulation results confirm that the detection behavior of the proposed SFETs can be predicted using the estimation behavior of the employed estimators for the SF.

In the second part, we introduced the concepts of the minimally IFD and the
MIG. We showed that the unknown parameters of the pdf of any invariant function does not depend on the unknown parameters if and only if the family of distribution is minimally invariant. We showed that the maximal invariant of any MIG eliminates the unknown parameters and preserves the maximum information of the observation. We proved that any invariant test with respect to the MIG is CFAR, conversely, for any CFAR test there exists an invariant statistic with respect to an MIG. Furthermore, if the support of pdf of observations does not depend on the unknown parameters then any CFAR test is also invariant with respect to an MIG. This result allowed us to directly prove that the GLRT for any minimally IFD is CFAR. We showed that for any CFAR test, there exists a function of observations and an MIG such that this function of observations is invariant with respect to the achieved MIG. This result allows to easily derive an enhanced CFAR test using a given one.

We introduced the UMP-CFAR test as the optimal bound for all CFAR tests. We proved that for the minimally invariant hypothesis testing problem, the UMP-CFAR test is given by the LR of the maximal invariant of the MIG. In many problems, this statistic depends on the unknown parameters (of $H_1$), therefore we say the the UMP-CFAR does not exist. In such cases the achieved test is defined by MP-CFAR test. However, this statistic always provides an upper-bound performance benchmark allowing to evaluate all suboptimal CFAR tests. We also proposed three suboptimal CFAR tests (ALM-CFAR, SFET-CFAR and CFAR-GLRT) where the SFET-CFAR is proven to be asymptotically optimal.

To illustrate the application of our results, we considered a general class of signal detection in Gaussian noise as an example and derived the MP-CFAR as an upper bound of CFAR tests. Moreover six CFAR SFETs and a GLRT are derived as
suboptimal tests.

In the third part, we considered the composite hypothesis testing problem after reduction using MIG. It is shown that, since the MLE of the induced maximal invariant of MIG is invariant with respect to the MIG, any SFET using MLE is CFAR. Hence we proposed GLRT-CFAR, Wald-CFAR, and Rao-CFAR as the suboptimal CFAR tests. Based on the invariance property of the MLE of induced maximal invariant, we considered the class of SFETs using MLE after reduction. Furthermore, the relationship between the GLRT-CFAR, Wald-CFAR, and Rao-CFAR and SFETs is studied. We have shown that under some mild conditions Wald-CFAR is an SFET, hence for iid observations the Wald-CFAR is asymptotically optimal. Moreover, the Rao-CFAR and the GLRT-CFAR are asymptotically SFET under some conditions. Based on the achieved results, the GLRT-CFAR, the Wald-CFAR and the Rao-CFAR are asymptotically optimal under the conditions. To improve the performance of SFETs after reduction, we proposed an SF named AOSF asymptotically maximizing the PoD in terms of separating functions. The AOSFET is applied using the MLE of the proposed AOSF. From the invariance property of MLE, the CFAR property of AOSFET is guaranteed. It is shown that maximizing the asymptotical PoD and a lower bound of PoD provide the same AOSF. The asymptotic PoD depends on the SF, the derivative of SF and the FIM of the family of distributions after reduction using MIG. We show that if the FIM does not depend on the unknown parameters the AOSFET and the Wald-CFAR are equivalent. In the case that FIM depends on the unknown parameters, the AOSF is given by the distance of a transformation on the unknown space under each hypothesis, where the gradient of the transformation gives a decomposition of FIM. The resulting AOSFET, generalizes the Wald-CFAR
and improves the performance. We show that all achieved results remind true without reducing by MIG for the composite hypothesis testing problem without main unknown parameters. Furthermore, using the proposed lower bound, a new lower bound of the PoD is derived for the LR test in simple hypothesis testing problems, when the number of observations tends to infinity.

In the forth part, an extension of SFET to the $M$-ary problems is proposed. For an $M$-ary problem, the SF is a vector of SFs such that the $m^{th}$ element of vector is an SF for the $\mathcal{H}_m$ versus the rest of hypotheses. It is shown that the MLE of SFs and the UMET (if exists) are equivalent. In many composite problems, the UMET does not exist, while the proposed SFET called CM-SFET is asymptotically optimal. Although the CM-SFET is asymptotically optimal, if the Lebesgue measure of at least one of unknown parameters sets is zero then the performance of CM-SFET decreases for finite observation. Using relaxing the hypothesis testing problem, the NDP is resolved. Two methods are proposed for relaxing the $M$-ary problems. One is based on the mixture modeling of pdfs of all hypotheses and another is based on the EEF modeling for the pdfs. The simulation result shows that the CM-SFET using EEF outperforms for low SNRs while the CM-SFET provides a lower probability of error when the SNR increases.

In this thesis, we applied the proposed method on practical problems such as detecting the target using MIMO radar, sparse signal detection in Gaussian noise and detecting a narrowband signal using a linear array. The simulation results show that the proposed tests significantly outperforms the GLRT, Wald and Rao tests.

Up to now, a new perspective is opened in detection theory, using applying MIG and the relationship between estimation and detection. In this thesis we study the
parametric estimation and detection and all methods are based on the parametric processing. In this case the list of possible future work are listed as follows

- Considering the Bayesian or semi-parametric problems and applying the concept of SFET.

- Finding the optimal SF for the case of finite observations.

- Finding a transformation to remove the unknown parameters under $\mathcal{H}_0$ and preserving the maximum information when the hypothesis testing problem is not invariant.

- Applying the proposed methods for practical problems to find new aspect of SFET, MIG SFET-CFAR and GLRT-CFAR.

- Applying the concept of SF for the problems that the pdf is unknown.
BIBLIOGRAPHY

Bibliography


detection with structured and unstructured covariance matrices,” *IEEE Trans.

identification using pilot tone signature,” in *IEEE International Workshop on
Signal processing*, May 2011, pp. 95–98.


[84] V. Friedman, “A zero crossing algorithm for the estimation of the frequency of
a single sinusoid in white noise,” *IEEE Trans. on Signal Processing*, vol. 42,

[85] A. M Haimovich, R. S. Blum, and L. J. Cimin, “MIMO radar with widely
separated antennas,” *IEEE Signal Processing Letters*, vol. 25, pp. 116–129,

[86] A. Ghobadzadeh, S. Gazor, M.R. Taban, A.A. Tadaion, and S.M. Moshtaghioun,
“Invariance and optimality of cfar detectors in binary composite
3535, July 2014.

criteria for determining the number of signals in high resolution array processing,”
*IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 11,


Appendix A

Appendix to Chapter 3

A.1 Proof of Theorem 2

Proof 26 This proof has two steps. We first show that for any \( \theta_1 \in \Theta_1 \) and any \( \theta_0 \in \Theta_0 \) the log-LR is a monotonic increasing function in \( T_g(x) \) which shows that the UMPU test can be given by \( \psi(x) = u(T_g(x) - \eta) \). In the second step, we show that \( E_{\theta_1}(\psi(x)) = \cdots = E_{\theta_{2N}}(\psi(x)) \) and that the threshold \( \eta \) shall be set by \( E_{\theta_i}(\psi(x)) = \alpha \) for \( i = 1, 2, \cdots, 2N \) to adjust the false alarm probability to the predetermined value \( \alpha \).

Step 1: The Cramér-Rao theorem for \( \lambda \overset{\Delta}{=} g(\theta) \) and \( \hat{\lambda} = T_g(x) \) yields

\[
\frac{\partial \ln(f(x; \theta))}{\partial \lambda} = I_g(\theta) (T_g(x) - g(\theta)), \quad (A.1)
\]

where \( I_g(\theta) = |\frac{dg(\theta)}{d\theta}|^2 I(\theta) \) [25]. Similar to the proof of Theorem 1, the log-LR for any \( \theta_1 \in \Theta_1 \) and any \( \theta_0 \in \Theta_0 \) is \( \ln(\text{LR}(x)) = (I_g(\theta_1) - I_g(\theta_0))T_g(x) - (H_g(\theta_1) - H_g(\theta_0)) \), where \( I_g(\cdot) \) and \( H_g(\cdot) \) are the indefinite integrals of \( I_g(\theta) \) and \( g(\theta)I_g(\theta) \) with respect to \( \lambda \), respectively, i.e., \( \frac{dI_g(\theta)}{d\lambda} = I_g(\theta) \) and \( \frac{dH_g(\theta)}{d\lambda} = g(\theta)I_g(\theta) \). Since \( I_g(\theta) > 0 \), \( I_g(\cdot) \)
is an increasing function with respect to \( \lambda = g(\theta) \). On the other hand from \( \theta_1 \in \Theta_1 \) and \( \theta_0 \in \Theta_0 \), we have \( g(\theta_1) > g(\theta_0) \); hence, \( \mathcal{I}_g(\theta_1) - \mathcal{I}_g(\theta_0) > 0 \). This means that the log-LR \( \ln(LR(x)) \) is an increasing function in \( T_g(x) \). Thus the UMPU test may exist and is given by \( \psi(x) = u(T_g(x) - \eta) \).

**Step 2:** To show that \( \eta \) does not depend on the unknown parameter, we consider a new hypothesis testing problem: \( \tilde{H}_0 : \theta = \vartheta_0 \) versus \( \tilde{H}_1 : \theta \in \{ \theta \in \Theta_0 \cup \Theta_1 | g(\theta) > g(\vartheta_0) \} \) for any given \( \vartheta_0 \in \Theta_0 \cup \Theta_1 \). The log-LR for this problem is the same as in Step 1 and thus is increasing in \( T_g(x) \). Therefore, the size-\( \alpha \) UMPU test for this problem is \( u(T_g(x) - \eta) \), where \( \eta \) is set to \( E_{\vartheta_0}(u(T_g(x) - \eta)) = \alpha \). Since this is an unbiased test, its probability of detection \( E_\theta(u(T_g(x) - \eta)) \) is greater than \( E_{\vartheta_0}(u(T_g(x) - \eta)) \) for all \( \theta \in \{ \theta \in \Theta_0 \cup \Theta_1 | g(\theta) > g(\vartheta_0) \} \) [10], i.e.,

\[
E_{\vartheta_0}(u(T_g(x) - \eta)) \leq E_\theta(u(T_g(x) - \eta)). \tag{A.2}
\]

Now for the main problem \( H_0 : \theta \in \Theta_0 \) against \( H_1 : \theta \in \Theta_1 \), we define the function \( \Pi : \Theta_0 \cup \Theta_1 \to [0, 1] \) by \( \Pi(\vartheta_0) = E_{\vartheta_0}(u(T_g(x) - \eta)) = \int_{|x| T_g(x) \geq \eta} f(x; \vartheta_0) d\mu_N \). This is a continuous function, since \( f(x; \vartheta_0) \) is continuous with respect to \( \vartheta_0 \). Thus, we have

\[
\sup\{\Pi(\vartheta_0) | \forall \vartheta_0; g(\vartheta_0) < 0\} = \sup\{\Pi(\vartheta_0) | \forall \vartheta_0; g(\vartheta_0) \leq 0\} \tag{128, Theorem 8.8}. \]

From \( g(\theta_i) = 0 \) for \( i = 1, 2, \ldots, 2N \), we obtain \( \Pi(\theta_i) \in \{ \Pi(\theta_0) | \forall \vartheta_0; g(\vartheta_0) \leq 0\} \) (A.2) provided that for all \( \vartheta_0 \in \Theta_0 \) we have \( E_{\vartheta_0}(u(T_g(x) - \eta)) \leq E_{\theta_i}(u(T_g(x) - \eta)) \), i.e.,

\[
\sup\{\Pi(\vartheta_0) | \forall \vartheta_0; g(\vartheta_0) < 0\} = \max_{i \in \{1, \ldots, 2N\}} E_{\theta_i}(u(T_g(x) - \eta)). \tag{A.3}
\]
In addition from (A.2), we obtain

\[
\sup_{\theta_0 \in \Theta_0} E^{\theta_0}(u(T_g(x) - \eta)) \leq \inf_{\theta \in \Phi} E^{\theta}(u(T_g(x) - \eta)), \tag{A.4}
\]

where \(\Phi = \{\theta \in \Theta_0 \cup \Theta_1 | g(\theta) > 0\}\). Similarly using [128, Theorem 8.8], it is easy to show

\[
\inf_{\theta \in \Phi} E^{\theta}(u(T_g(x) - \eta)) = \inf_{\theta \in \Phi} E^{\theta}(u(T_g(x) - \eta)), \tag{A.5}
\]

where \(\Phi = \Phi \cup \{\theta_1, \ldots, \theta_{2N}\}\). From (A.2) we have \(E^{\theta_i}(u(T_g(x) - \eta)) \leq E^{\theta}(u(T_g(x) - \eta))\), for all \(i \in \{1, 2, \ldots, 2N\}\) and all \(\theta\) such that \(g(\theta) > 0\). Hence \(\inf_{\theta \in \Phi} E^{\theta}(u(T_g(x) - \eta)) = \min_{i \in \{1, 2, \ldots, 2N\}} E^{\theta_i}(u(T_g(x) - \eta))\). Therefore, (A.4) and (A.5) provide

\[
\sup_{\theta_0 \in \Theta_0} E^{\theta_0}(u(T_g(x) - \eta)) \leq \min_{i \in \{1, 2, \ldots, 2N\}} E^{\theta_i}(u(T_g(x) - \eta)). \tag{A.6}
\]

By comparing (A.3) and (A.6), we obtain \(\sup_{\theta_0 \in \Theta_0} E^{\theta_0}(u(T_g(x) - \eta)) = E^{\theta_1}(u(T_g(x) - \eta)) = E^{\theta_2}(u(T_g(x) - \eta)) = \cdots = E^{\theta_{2N}}(u(T_g(x) - \eta))\). Thus to achieve the UMPU test, the threshold \(\eta\) must satisfy \(\sup_{\theta_0 \in \Theta_0} E^{\theta_0}(u(T_g(x) - \eta)) = \alpha\) which is independent of the unknowns parameter \(\theta \in \Theta_0\).

### A.2 Proof of Theorem 3

**Proof 27** Similarly the proof has two major steps. First, we show that the log-LR is increasing in \(T_g(x)\), which shows that the UMP may exist and is given by \(\psi(x) = u(T_g(x) - \eta)\) and in the second step it is shown that, the threshold \(\eta\) is independent to unknown parameters and it is set by \(E^{\theta_g}(\psi(x)) = \alpha\), where \(\theta_g = \{\theta | g(\theta) = 0\}\).
Step 1: We first show that the log-LR is a monotone function. For a differentiable pdf, the derivative of \( \ln(f(x; \theta)) \) is \( \frac{df(x; \theta)}{d\theta} \). Using Cramér-Rao theorem \([25]\), we have 

\[
\lambda \sum_{i=1}^{M} \frac{\partial \ln(f(x; \theta))}{\partial \theta_i} d\theta_i = \int \frac{\partial \ln(f(x; \theta))}{\partial \theta_i} \sum_{i=1}^{M} \frac{\partial f(x; \theta)}{\partial \theta_i} d\theta_i. \quad (A.7)
\]

Let \( \lambda \triangleq g(\theta) \), substituting in (A.7), the log-likelihood function is 

\[
\ln(f(x; \theta)) + c(x) = \sum_{i=1}^{M} \int \frac{\partial \ln(f(x; \theta))}{\partial \theta_i} d\theta_i.
\]

Step 2: Similar to proof of Theorem 2, to show that \( \eta \) does not depend on the unknown parameter, we consider a new hypothesis testing problem: \( \widetilde{H}_0 : \theta = \theta_0 \) against \( \widetilde{H}_1 : \theta \in \{ \theta \in \Theta \cup \Theta_1 | g(\theta) > g(\theta_0) \} \) for any given \( \theta_0 \in \Theta \cup \Theta_1 \). Step 1 implies that the log-LR is increasing in \( T_g(x) \). Therefore, the size-\( \alpha \) UMPU test for this problem is 

\[
u(T_g(x) - \eta) \]

where \( \eta \) satisfies \( E_{\theta_0}(u(T_g(x) - \eta)) = \alpha \). Similar to Step 2 in the proof of Theorem 2, we have 

\[
E_{\theta_0}(u(T_g(x) - \eta)) \leq E_{\theta}(u(T_g(x) - \eta)). \quad (A.8)
\]
for all \( \theta \in \{ \theta \in \Theta_0 \cup \Theta_1 | g(\theta) > g(\vartheta_0) \} \).

Now for the problem \( H_0 : \theta \in \Theta_0 \) against \( H_1 : \theta \in \Theta_1 \), it is clear that \([128, Theorem 8.8]\) \( \sup \{ E_{\vartheta_0}(u(T_g(x) - \eta)) \forall \theta_0; g(\theta_0) < 0 \} = \sup \{ E_{\vartheta_0}(u(T_g(x) - \eta)) \forall \theta_0; g(\theta_0) \leq 0 \} \). For all \( \theta_g \in g^{-1}\{0\} \) we have \( g(\theta_g) = 0 \) hence,

\[
E_{\theta_g}(u(T_g(x) - \eta)) \in \{ E_{\vartheta_0}(u(T_g(x) - \eta)) \forall \theta_0; g(\theta_0) \leq 0 \}.
\]

According to (A.8), for all \( \vartheta_0 \in \Theta_0 \) we have \( E_{\vartheta_0}(u(T_g(x) - \eta)) \leq E_{\theta_g}(u(T_g(x) - \eta)) \).

Hence

\[
\sup \{ E_{\vartheta_0}(u(T_g(x) - \eta)) \forall \theta_0; g(\theta_0) < 0 \} = \sup_{\theta_g \in g^{-1}(\{0\})} E_{\theta_g}(u(T_g(x) - \eta)). \quad (A.9)
\]

On the other hand, from (A.8), we obtain

\[
\sup_{\theta_0 \in \Theta_0} E_{\theta_0}(u(T_g(x) - \eta)) \leq \inf_{\theta \in \{ \theta \in \Theta_0 \cup \Theta_1 | g(\theta) > 0 \}} E_{\theta}(u(T_g(x) - \eta)).
\]

Similarly, \([128, Theorem 8.8]\) provides

\[
\inf_{\theta \in \Theta_1} E_{\theta}(u(T_g(x) - \eta)) = \inf_{\theta \in g^{-1}(\{0, +\infty\})} E_{\theta}(u(T_g(x) - \eta)). \quad (A.10)
\]

Also, from (A.8) we have \( E_{\theta_g}(u(T_g(x) - \eta)) \leq E_{\theta}(u(T_g(x) - \eta)) \) for all \( \theta_g \in g^{-1}(\{0\}) \) and all \( \theta \) such that \( g(\theta) > 0 \); hence

\[
\inf_{\theta \in \{ \theta \in \Theta_0 \cup \Theta_1 | g(\theta) \geq 0 \}} E_{\theta}(u(T_g(x) - \eta)) = \inf_{\theta_g \in g^{-1}(\{0\})} E_{\theta_g}(u(T_g(x) - \eta)).
\]
Therefore, we have
\[ \sup_{\vartheta_0 \in \Theta_0} E_{\vartheta_0}(u(T_g(x) - \eta)) \leq \inf_{\theta_g \in g^{-1}(\{0\})} E_{\theta_g}(u(T_g(x) - \eta)). \tag{A.11} \]

By comparing (A.9) and (A.11) we must have
\[ \sup_{\vartheta_0 \in \Theta_0} E_{\vartheta_0}(u(T_g(x) - \eta)) = E_{\theta_g}(u(T_g(x) - \eta)) \]
for all \( \theta_g \in g^{-1}(\{0\}) \). Thus to achieve the UMPU test, the threshold \( \eta \) must satisfy \( E_{\theta_g}(u(T_g(x) - \eta)) = \alpha \), that is independent of unknowns parameters \( \theta \in \Theta_0 \).

**A.3 Proof of Lemma 1**

**Proof** Without loss of generality we can assume \( m_M(\Theta_0) > 0 \), since otherwise a constant function fulfils the required conditions. If \( \Theta_1 - \partial \Theta_1 = \emptyset \), define the function \( g_0 : \Theta \to \mathbb{R} \) as a constant function \( g_0(\theta) = c \) for all \( \theta \in \Theta \). Otherwise, let \( \theta_0 \in \Theta_1 - \partial \Theta_1 \). Then, the sets \( \{\theta_0\} \) and \( \partial \Theta_1 \) are two disjoint closed subsets of the perfectly normal space \( \mathbb{R}^M \). Hence, we can apply Lemma 3 at the end of this appendix to these two sets. Thus, there exists a continuous function \( g_0 \) such that \( g_0 : \Theta \to [c, \beta] \) with \( g_0(\theta_0) = \beta \) and \( \partial \Theta_1 = g_0^{-1}(c) \), for some \( c < \beta \). Therefore, for any \( \theta \in \Theta_1 - \partial \Theta_1 \), we have \( g_0(\theta) > c \).

We define a sequence of sets as \( A_n \triangleq \{\theta \in \mathbb{R}^M | \inf_{\theta_1 \in \Theta_1} d(\theta, \theta_1) > \frac{1}{n}\} \) for \( n = 1, 2, \cdots \) where \( d \) denotes an Euclidian metric. Therefore, \( \Theta_0 \subseteq \cup_{n=1}^{\infty} A_n \) almost everywhere, this is because \( \Theta_0 \cap \overline{\Theta_1} \) has a zero Lebesgue measure. Let \( \partial A_n \) denote the boundary set of \( A_n \). Then, \( \partial A_n \subseteq \{\theta | \inf_{\theta_1 \in \Theta_1} d(\theta, \theta_1) = \frac{1}{n}\} \). Furthermore, \( \partial A_n \) and \( \partial A_m \) are disjoint closed sets for all non-equal arbitrary \( m \) and \( n \). By Lemma 3, there
exist continuous functions \( h_n : \mathbb{R}^M \to \left[ c - \frac{1}{n-1}, c - \frac{1}{n} \right] \) for \( n = 2, 3, \cdots \) such that \( h_n^{-1}\left(\left\{ c - \frac{1}{n-1} \right\} \right) = \partial A_{n-1}, h_n^{-1}\left(\left\{ c - \frac{1}{n} \right\} \right) = \partial A_n \) for \( n \geq 2 \), and \( h_1 : \mathbb{R}^M \to \mathbb{R} \), where \( h_1(\theta) = c - 1 \) for all \( \theta \). We define functions \( r_n(\cdot) \) by

\[
 r_n(\theta) = \begin{cases} 
 h_1(\theta), & \theta \in A_1, \\
 h_2(\theta), & \theta \in A_2 - A_1, \\
 \vdots & \vdots \\
 h_n(\theta), & \theta \in A_n - A_{n-1}, \\
 c - \frac{1}{n}, & \theta \in A_n^c.
\end{cases} 
\] (A.12)

According to the Pasting lemma [79, Theorem 3.7] \( r_n \) is a continuous function. Furthermore, \( r_n \) is increasing in \( n \), i.e., \( r_{n+1}(\theta) \geq r_n(\theta) \), and \( r_n \) is bounded by the constant \( c \). The Weierstrass M-test implies that \( r_n \) uniformly converges to a continuous function \( g_1 \overset{\Delta}{=} \lim_{n \to \infty} r_n \). We define

\[
 g(\theta) = \begin{cases} 
 g_1(\theta), & \theta \in \bigcup_{n=1}^{\infty} A_n, \\
 g_0(\theta), & \theta \in \overline{\Theta_1}.
\end{cases} 
\] (A.13)

according to the Pasting lemma, \( g(\theta) \) is a continuous function. Hence, we have \( g(\theta) < c \) for almost all \( \theta \in \Theta_0 \), and \( g(\theta) \geq c \) for \( \theta \in \overline{\Theta_1} \) and \( \partial \Theta_1 = g^{-1}(\{c\}) \).

Therefore, the proof is complete by the restriction of \( g \) to the set \( \Theta = \Theta_0 \cup \Theta_1 \).

**Lemma 3** Let \( \mathcal{X} \) be a normal space, \( A \) and \( B \) be two disjoint closed subsets in \( \mathcal{X} \) and \( [\alpha, \beta] \) be a closed interval in the real line. Then there exists a continuous map \( f : \mathcal{X} \to [\alpha, \beta] \) such that \( f(x) = \alpha, \forall x \in A \) and \( f(x) = \beta, \forall x \in B \). Furthermore, if \( \mathcal{X} \) is a perfectly normal space, the function \( f \) can be chosen such that \( A = f^{-1}(\alpha) \)
and $B = f^{-1}(\beta)$.

**Proof 29** The space $\mathcal{X}$ is said to be normal if for any disjoint closed subsets $A$, $B$ of $\mathcal{X}$, there exist two disjoint open subsets of $\mathcal{X}$ containing $A$ and $B$, respectively. The first part of Lemma 3 is the Urysohn’s lemma, e.g., see [79, Urysohn’s lemma, Section 4] and the second part comes from a natural extension of normal spaces to the concept of perfectly normal spaces.

### A.4 Proof of Theorem 4

**Proof 30** Consider two reliable detectors $\psi_1$ and $\psi_2$ with the false alarm probability of $P_{fa1} = P_{fa2} = \alpha$. For $i = 1, 2$, $j = 0, 1$ and $\Gamma_{i,j} = \{x | \psi_i(x) = j\}$, Lemmas 1 and 2 reveal that there exist $\tilde{h}_i(x)$ and $\tilde{c}_i(\alpha)$ where $\psi_i(x) = u(\tilde{h}_i(x) - \tilde{c}_i(\alpha))$. Since the tests are reliable, $E_{\theta}(\psi_i) = 1 - F_{\tilde{h}_i}(\tilde{c}_i(\alpha); \theta)$ is monotonically increasing in $\theta$, in which $F_{\tilde{h}_i}(\cdot; \theta)$ is the Cumulative Distribution Function (CDF) of $\tilde{h}_i(x)$. Thus $\sup_{\theta \in \Theta} E_{\theta}(\psi_i) = E_{\theta_b}(\psi_i) = \alpha$. Therefore $\tilde{c}_i(\alpha)$ must satisfy $F_{\tilde{h}_i}(\tilde{c}_i(\alpha); \theta_b) = 1 - \alpha$. This reveals that $\tilde{c}_i(\alpha)$ is a monotonically decreasing function of $\alpha$, hence $-\tilde{c}_i^{-1}(\cdot)$ is a monotonically increasing function. By applying $-\tilde{c}_i^{-1}(\cdot)$ on both sides of the tests we have $\psi_i(x) = u(\tilde{h}_{c,i}(x) - x)$ where $\tilde{h}_{c,i}(x) \overset{\Delta}{=} -\tilde{c}_i^{-1}(\tilde{h}_i(x))$ and $x \overset{\Delta}{=} -\alpha$ (i.e., $\psi_i$ rejects $\mathcal{H}_0$ if $\tilde{h}_{c,i}(x) > x$).

The assumption $P_{d1} > P_{d2}$ means $F_{\tilde{h}_{c,1}}(x; \theta) < F_{\tilde{h}_{c,2}}(x; \theta)$ for all $\theta > \theta_b$ and $x \in [-1, 0]$. In addition, these CDFs are increasing in $x$ and tend to 1 as $x \to 0$. Definition 11 implies that the inflection points of these CDFs are countable and bounded. Hence, we can define $x^* \overset{\Delta}{=} \max\{-1, x_{k,i}\}$, where $x_{k,i}$’s are the inflection points of $F_{\tilde{h}_{c,i}}(\cdot; \theta)$. Since both functions $F_{\tilde{h}_{c,1}}(x; \theta)$ and $F_{\tilde{h}_{c,2}}(x; \theta)$ are monotonically increasing in $x$ and tend to 1 and have no inflection point over the interval $[x^*, 0]$,
the function \( F_{\tilde{h}_{c,i}}(x;\theta) - F_{\tilde{h}_{c,i}}(x;\theta) \geq 0 \) is a monotonically decreasing function in \( x \in [x^*,0] \) and tends to zero as \( x \to 0 \).

On the other hand, we have \( 1 - P_{d_i} = F_{\tilde{h}_{c,i}}(x;\theta) \). Definition 11 implies that \( F_{\tilde{h}_{c,i}}(x;\theta) \) is monotonically decreasing in \( \theta \), so the inverse of the CDFs with respect to \( \theta \) exist and we can denote it by \( \theta = F_{\tilde{h}_{c,i}}^{-1}(x;1-P_{d_i}) = F_{\tilde{h}_{c,i}}^{-1}(x;1-P_{d_i}) \) for \( x \in [-1,0] \) and \( \theta > \theta_b \). Note that in \( \theta = F_{\tilde{h}_{c,i}}^{-1}(x;1-P_{d_i}) \), \( \theta \) is an increasing of \( x \) for any given \( P_{d_i} \). Now, we define \( \theta = \zeta(x) \triangleq F_{\tilde{h}_{c,i}}^{-1}(x/|x^*|;1-P_{d_i}) \) for \( x \in [x^*,0] \) and \( \zeta(x) \triangleq x + \theta_b - x^* \) for \([-1,x^*)\). Since \( \zeta(x) \) is an increasing function, it is clear that \( \psi_i(x) = u(\zeta(\tilde{h}_{c,i}(x)) - \zeta(x)) \).

We now show that \( h_1(x) \triangleq \zeta(\tilde{h}_{c,1}(x)) \) is a more \( \epsilon \)-accurate estimator of \( \theta \) than \( h_2(x) \triangleq \zeta(\tilde{h}_{c,2}(x)) \). Since \( F_{\tilde{h}_{c,2}}(x;\theta) - F_{\tilde{h}_{c,1}}(x;\theta) \) is decreasing in \( x \) for all \( x > x^* \), we have \( F_{\tilde{h}_{c,2}}(\zeta^{-1}(\theta+\epsilon);\theta) - F_{\tilde{h}_{c,2}}(\zeta^{-1}(\theta-\epsilon);\theta) < F_{\tilde{h}_{c,1}}(\zeta^{-1}(\theta+\epsilon);\theta) - F_{\tilde{h}_{c,1}}(\zeta^{-1}(\theta-\epsilon);\theta) \), where \( \theta > \theta_b \) and \( \zeta^{-1}(\theta-\epsilon) > x^* \) or equally \( \epsilon \leq \epsilon_{\text{max}} \triangleq \theta - \theta_b \). As \( \zeta(\cdot) \) is an increasing function, for all \( \theta > \theta_b \) the above leads to \( \Pr_{\theta}(|h_2(x) - \theta| < \epsilon) = \Pr_{\theta}(\theta - \epsilon < \zeta(\tilde{h}_{c,2}(x)) < \theta + \epsilon) < \Pr_{\theta}(\theta - \epsilon < \zeta(\tilde{h}_{c,1}(x)) < \theta + \epsilon) = \Pr_{\theta}(|h_1(x) - \theta| < \epsilon) \).

### A.5 Proof of Theorem 6

**Proof 31** An invariant test depends on observation vector \( x \) only via a maximal invariant of \( x \) with respect to \( Q \), denoted by \( m(x) \) [10]. The pdf of \( m(x) \) depends on \( \theta \) only via the maximal invariant of \( \theta \in \Theta \) with respect to \( Q \) that denoted by \( \rho(\theta) \) [10]. If all \( \rho_l(\theta) \)'s are constant for all \( l = 1, \cdots, j - 1, j + 1, \cdots, L \), then Lemma 4 at the end of this appendix proves that \( E_{\theta}(\psi) \) is increasing (or decreasing) in \( \rho_j(\theta) \) for all invariant tests \( \psi(x) \). We can easily redefine the maximal invariant such that \( E_{\theta}(\psi) \) is increasing in \( \rho_j(\theta) \) for all \( j \), e.g., if it is a decreasing in any element \( \rho_j(\theta) \), we
could substitute that element by \(1/\rho_j(\theta)\).

The UMPI test is an invariant and unbiased test \([10,80]\). Thus its probability of detection is greater than its false alarm probability, i.e., \(E_{\theta \in \Theta_1}(\psi_{\text{UMPI}}) > E_{\theta \in \Theta_0}(\psi_{\text{UMPI}})\). On the other hand, Lemma 4 guarantees that \(E_{\theta}(\psi_{\text{UMPI}})\) is increasing in \(\rho(\theta)\). As \(\rho_{\text{tot}}(\theta)\) is increasing in \(\rho_l(\theta)\) for all \(l\) and \(E_{\theta}(\psi)\) is increasing in \(\rho_l(\theta)\), then \(E_{\theta}(\psi)\) is increasing in \(\rho_{\text{tot}}(\theta)\). Then, as \(\rho_{\text{tot}}(\theta)\) is increasing in \(\rho_{\text{tot}}(\theta)\) for all \(\theta\) and \(\theta_1 \in \Theta_1\) we have \(\rho_{\text{tot}}(\theta_1) > \rho_{\text{tot}}(\theta_0)\). Therefore \(g(\theta)\) defined in (3.5) is an SF, i.e., \(g(\theta) > 0\) for all \(\theta \in \Theta_1\) and \(g(\theta) \leq 0\) for all \(\theta \in \Theta_0\).

**Lemma 4** Assuming that the problem in (2.1) is invariant under group \(Q\) and \(\rho(\theta)\) is a maximal invariant of \(\theta \in \Theta\) with respect to induced group transformation \(Q\). Then for any \(\psi(x) = u(h(x) - \eta)\) invariant test where \(\eta\) is set to the false alarm probability requirements and \(h(x)\) is the test statistic, the function \(E_{\theta}(\psi(x))\) is a one to one with respect to \(\rho(\theta)\).

**Proof 32** We must show that from \(E_{\theta_1}(\psi(x)) = E_{\theta_2}(\psi(x))\) we get \(\rho(\theta_1) = \rho(\theta_2)\), where \(\theta_1, \theta_2 \in \Theta\). Consider the random variable \(h(x)\) with the pdf of \(f(h; \vartheta)\), where \(\vartheta\) is the parameter of the distribution. Note that \(\vartheta\) is a function of \(\theta \in \Theta\). Defining the indicator function of a set \(A\) as \(i_A(h) = 1\) if \(h \in A\) and \(i_A(h) = 0\) otherwise, we have \(E_{\theta}(\psi(x)) = \int_R i_{(\eta, \infty)}(h)f(h; \vartheta)d\mu_1\), where \(\theta \in \Theta_1\) and \(\mu_1\) is Lebesgue measure in \(\mathbb{R}\). Let \(\vartheta_1\) and \(\vartheta_2\) correspond to \(\theta_1\) and \(\theta_2\) respectively. We first assume that \(E_{\theta_2}(\psi(x)) = E_{\theta_2}(\psi(x))\) for all \(P_\text{fa} \in [0,1]\) or similarly for all \(\eta \in \mathbb{R}\) and prove that \(f(h; \vartheta_1) = f(h; \vartheta_2)\) for all \(h \in \mathbb{R}\). From

\[
\int_R i_{(\eta, \infty)}(h)f(h; \vartheta_1)d\mu_1 = \int_R i_{(\eta, \infty)}(h)f(h; \vartheta_2)d\mu_1, \quad (A.14)
\]
for all $\eta_{1k}^{(n)} < \eta_{2k}^{(n)} \in \mathbb{R}$, we obtain

\[
\int_{\mathbb{R}} i_{(\eta_{1k}^{(n)}, \eta_{2k}^{(n)})}(h)[f(h; \vartheta_1) - f(h; \vartheta_2)] dm_1 = 0. \tag{A.15}
\]

On the other hand, for any arbitrary Borel measurable function $w(h) > 0$, a sequence of simple functions defined by $s_n(h) = \sum_k a_k^{(n)} i_{(\eta_{1k}^{(n)}, \eta_{2k}^{(n)})}(h)$ exists such that $s_n(h) \leq s_{n+1}(h) \leq w(\cdot)$ and $\lim_{n \to \infty} s_n(\cdot) = w(\cdot)$ \cite[Ch.1, Th.17]{74}. Using (A.15), we get

\[
\int_{\mathbb{R}} s_n(h)[f(h; \vartheta_1) - f(h; \vartheta_2)] dm_1 = 0. \quad \text{Thus as } n \to \infty \text{ \cite[Ch. 1, Th. 26]{74} we obtain}
\]

\[
\int_{\mathbb{R}} w(h)(f(h; \vartheta_1) - f(h; \vartheta_2)) dm_1 = 0. \quad \tag{A.16}
\]

In this case, Theorem 39 in \cite[Ch. 1]{74} implies that $f(h; \vartheta_1) = f(h; \vartheta_2)$ for all $h \in \mathbb{R}$ (almost everywhere).

On the other hand, if $\psi(x)$ is an invariant test, we have $h(q(x)) = h(x)$ for all $q \in Q$. Hence for a Borel set $\Upsilon \in \mathbb{R}$, we can write

\[
\int_{\Upsilon} f(h; \vartheta_1) dm_1 = \Pr_{\theta_1}(h(x) \in \Upsilon) = \Pr_{\theta_1}(h(q(x)) \in \Upsilon)
\]

\[
= \Pr_{\theta_1}(q(x) \in h^{-1}(\Upsilon)) = \Pr_{q(\theta_1)}(x \in h^{-1}(\Upsilon)), \tag{A.17}
\]

where $h^{-1}(A) = \{x|h(x) \in A\}$. Similarly, we have

\[
\int_{\Upsilon} f(h; \vartheta_2) dm_1 = \Pr_{\theta_2}(x \in h^{-1}(\Upsilon)). \tag{A.18}
\]

Assuming $f(h; \vartheta_1) = f(h; \vartheta_2)$, (A.17) and (A.18) are equivalent, i.e., $\Pr_{q(\theta_1)}(x \in h^{-1}(\Upsilon)) = \Pr_{\theta_2}(x \in h^{-1}(\Upsilon))$. Since $\Upsilon$ is a Borel set, using similar arguments in
(A.14) and (A.16), we conclude that \( q(\theta_1) = \theta_2 \). As \( \rho(\cdot) \) is maximal invariant with respect to the induced group, by applying \( \rho(\cdot) \) on both sides of this equation, we have \( \rho(\theta_1) = \rho(\theta_2) \).
Appendix B

Appendix to Chapter 4

B.1 Proof of Proposition 2

Proof 33 Let \( m_{m_1}(x) \) be the maximal invariant of \( Q_{m_1} \), defining an invertible transformation \( \zeta' \) from \( Q_{m_1} \) to \( Q_{m_2} \) by \( m_{m_2} = \zeta'(m_{m_1}) \overset{\Delta}{=} h_2^{-1}(h_1(m_{m_1})) \), where \( h_1 : Q_{m_1} \to Q \), \( h_1(m_{m_1}) = \overline{m_{m_1}} \) and \( h_2 : Q_{m_2} \to Q \), \( h_2(m_{m_2}) = \overline{m_{m_2}} \) are the monomorphism transformations and \( m_{m_1} \in Q_{m_1} \) and \( m_{m_2} \in Q_{m_2} \). We claim that \( m_{m_2}(x) = m_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(x))) \) is the maximal invariant with respect to \( Q_{m_2} \), where \( I_{Q_{m_2}} \) is the identity element of \( Q_{m_2} \). The conditions of maximal invariant for all \( m_{m_2} \in Q_{m_2} \) are

1) \( m_{m_2}(m_{m_2}(x))m_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(m_{m_2}(x)))) = m_{m_1}(m_{m_1}(I_{Q_{m_1}}(x))) = m_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(x))) = m_{m_2}(x) \). Note that \( q_{m_1}(I_{Q_{m_1}}(x)) = q_{m_2}(I_{Q_{m_2}}(x)) \).

2) if \( m_{m_2}(x_1) = m_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(x_1))) = m_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(x_2))) = m_{m_2}(x_2) \) then there exists a \( q_{m_1} \in Q_{m_1} \) such that \( \zeta'^{-1}(I_{Q_{m_2}}(x_1)) = q_{m_1}(\zeta'^{-1}(I_{Q_{m_2}}(x_2))) \), thus \( x_1 = q_{m_2}(x_2) \), where \( q_{m_2} \) is given by \( q_{m_2} = \zeta'(q_{m_1}) \), and the identity element of \( Q_{m_1} \) is given by \( I_{Q_{m_1}} = \zeta'^{-1}(I_{Q_{m_2}}(x)) \).

Defining \( \zeta(x) \overset{\Delta}{=} \zeta'^{-1}(I_{Q_{m_2}}(x)) \), Proposition 2 is proved.
Proof 34 Remark 1 ensures the existence of a continuous function $h(x)$ and a threshold $\eta \in \mathbb{R}$ such that $\psi(x) = 1$ for $h(x) \geq \eta$ and $\psi(x) = 0$ for $h(x) < \eta$. Hence, the probability of false alarm is $P_{fa} = 1 - F_h(\eta)$, where $F_h(\cdot)$ is the cdf of random variable $h(x)$. Since the test is CFAR, the cdf of $h(x)$ under $\mathcal{H}_0$ should not depend on the unknown null parameters $\theta \in \Theta_0$. For an arbitrary $\theta \in \Theta_0$ and a $\overline{q}_m \in \overline{Q_m}$, we define $\theta' = \overline{q}_m(\theta) \in \Theta_0$ which gives null pdf $f(\cdot; \theta')$. Under $\mathcal{H}_0$, the function $h$ maps $\cup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta))$ to $h(\cup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)))$. Since the distribution of $h(x)$ does not depend on the unknown null parameters, we can write

$$
\Pr_{\theta}\{h(x) \in \gamma\} = \Pr_{\theta'}\{h(x) \in \gamma\}, \ i.e.,
$$

$$
\int_{h^{-1}(\gamma)} f_x(x; \theta) dm_N(x) = \int_{h^{-1}(\gamma)} f_x(\overline{q}_m(\theta)) dm_N(x), \quad (B.1)
$$

and $\gamma \subset h(\cup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)))$ is a Borel set. Theorem 26.7 in [74] provides

$$
\int f(x; \overline{q}_m(\theta)) dm_N(x) = \int f(q_m(x); \overline{q}_m(\theta)) J_{q_m} dm_N(x), \quad (B.2)
$$

where, $J_{q_m}$ is the absolute of Jacobian of $q_m(x)$ with respect to $x$. Comparing $(B.1)$ and $(B.2)$, the relation $f(q_m(x); \overline{q}_m(\theta)) J_{q_m} = f_x(x; \theta)$ [64] leads to

$$
\int_{h^{-1}(\gamma)} f_x(x; \theta) dm_N(x) = \int_{q_m^{-1}(h^{-1}(\gamma))} f_x(x; \theta) dm_N(x).
$$

As $\gamma$ is an arbitrary measurable set, the above implies $h^{-1}(\gamma) = q_m^{-1}(h^{-1}(\gamma))$, i.e., $h(q_m(h^{-1}(\gamma))) = \gamma$, almost everywhere. Since $h$ is an onto function, we have [128, Ch.
B.3. PROOF OF THEOREM 12

1] \[ h(q_m(h^{-1}(\gamma))) = h(h^{-1}(\gamma)). \]

By denoting \( A \triangleq h^{-1}(\gamma) \subset C^N \), we obtain \( h(q_s(A)) = h(A) \). Now, we aim to prove that \( \psi(\cdot) \) is invariant under \( Q_m \) almost everywhere, i.e., \( \psi(q_m(x)) = \psi(x) \) for all \( x \in \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \) and all \( q_m \in Q_m \). Since the rejection region of \( \psi(\cdot) \) is given by \( \Gamma_1 = h^{-1}([\eta, \infty)) \), we use \( \gamma = [\eta, \infty) \cap h(\bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta))) \) in \( h(q_s(A)) = h(A) \), and obtain \( A = h^{-1}(\gamma) = h^{-1}(h(\bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)))) \cap \Gamma_1 \). It is obvious that \( h^{-1}(h(\bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)))) \supseteq \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \). Thus for all \( x \in \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \cap h^{-1}([\eta, \infty)) \), we have \( h(q_m(x)) \in h(A) \subset [\eta, \infty) \) which means \( h(q_m(x)) > \eta \), i.e., we get \( \psi(q_m(x)) = 1 = \psi(x) \). Similarly, for all \( x \in \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \cap h^{-1}((-\infty, \eta)) \), we prove that the acceptance region of \( \psi(x) \), \( \Gamma_0 = h^{-1}((-\infty, \eta)) \), is identical to that of \( \psi(q_m(x)) \). Using \( \gamma = (-\infty, \eta) \cap h(\bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta))) \) in \( h(q_s(A)) = h(A) \) and for all \( x \in \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \cap \Gamma_0 \), we see that \( h(q_m(x)) \in (-\infty, \eta) \), i.e., \( h(q_m(x)) < \eta \) and \( \psi(q_m(x)) = 0 = \psi(x) \). In summary, the theorem is proven as we have \( \psi(q_s(x)) = \psi(x) \) for all \( x \in \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \) since the invariance property of \( \psi(\cdot) \) is extended for all \( x \) by assuming that the support \( \bigcup_{\theta \in \Theta_0} \text{supp}(f(\cdot; \theta)) \) does not depend on the unknowns.

B.3 Proof of Theorem 12

Proof 35 According to Theorem 9, if the support of pdf does not depend on unknown parameters, any CFAR test for invariant hypothesis problems with respect to an MIG is an invariant, i.e., \( \Omega \subseteq \Omega_T \) where \( \Omega \) hereafter denotes the set of all possible CFAR
tests and $\Omega_I$ denotes all invariant tests with respect to $Q_m$. Moreover, as $Q_m$ is MIG, Theorem 10 guaranties $\Omega_I \subseteq \Omega$. Then, $\Omega_I = \Omega$ and so the optimal CFAR test of $\Omega$ is given by comparing the LR of maximal invariant $m_m(x)$ with a threshold $\eta$, which is the optimal test amongst all invariant tests. Now, suppose that the support of pdf depends on the unknown parameters. We denote the probability of false alarm by $\alpha \in [0, 1]$ and define $\pi_\alpha \triangleq \sup_{\psi \in \Omega} P_{d,\psi|\alpha}$ as the highest detection probability for a test size of $\alpha$, where $P_{d,\psi|\alpha}$ is the detection probability of $\psi$. The value $\pi_\alpha$ exists as $P_{d,\psi|\alpha} \in [0, 1]$. Let divide the interval $\alpha \in [0, 1]$ by choosing $n$ arbitrary points as $0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_n = 1$.

A CFAR test denoted by $\psi_{k,n}(\cdot)$ exists with a size of $\alpha_{k,n}$ and power of $\pi_{\alpha_{k,n}}$. Theorem 9 implies that $\psi_{k,n}(\cdot)$ is invariant, for all $x \in \text{supp}(f(\cdot;\theta))$ and $\theta \in \Theta_0$. Hence the test is invariant over $\text{supp}(f(\cdot;\theta))$ and can be expressed as a function of maximal invariant $m_m(x)$. Furthermore, $x \notin \text{supp}(f(\cdot;\theta))$ means that $H_0$ is not correct and the optimal CFAR test must declare $H_1$. Thus, $\psi_{k,n}(\cdot)$ depends on $x$ through $m_m(x)$ for $x \in \text{supp}(f(\cdot;\theta))$. Hence, $\psi_{k,n}(\cdot)$ is an invariant test with respect to $Q_m$.

Let the arbitrary size $\alpha$ fall within the interval $[\alpha_{k-1,n}, \alpha_{k,n}]$ and denote $q = \frac{\alpha - \alpha_{k-1,n}}{\alpha_{k,n} - \alpha_{k-1,n}}$. We define the output of $\psi_n(x)$ to be randomly be equal to that of $\psi_{k-1,n}(x)$ with probability of $P_q$ and to $\psi_{k,n}(x)$ with probability of $1 - P_q$. Obviously, the test $\psi_n(x)$ has a power of $\pi_{\alpha,n} = P_q \pi_{\alpha_{k-1,n}} + (1 - P_q) \pi_{\alpha_{k,n}}$ and a size of $\alpha$. The curve $(\alpha, P_q \pi_{\alpha_{k-1,n}} + (1 - P_q) \pi_{\alpha_{k,n}})$ is a continuous piecewise linear function interconnecting the set of points $\{(\alpha_{k,n}, \pi_{\alpha_{k,n}})\}$. We choose $n = 2^b$ and $\alpha_{k,n} = \frac{k}{2^b}$ for $k = 0, 1, \cdots, 2^b$. As $b$ increases, this curve converges to the function $\pi_\alpha$. This simply means that there exists an invariant test converging to the optimal CFAR bound. Obviously,
\( \pi_{\alpha,2^b} \) is increasing in \( b \), and converges to its upper-bound \( \pi_{\alpha} \). Since \( \psi_{2^b}(\cdot) \) is invariant it is also CFAR, thus its power \( \pi_{\alpha,n} \) is less than \( \pi_{\alpha} \) for a given size of \( \alpha \). We have \( \pi_{\alpha} = \sup_{\psi \in \Omega} P_{\psi,\alpha} = \sup_{\psi} P_{\psi,\alpha} \). On the other hand, the invariance of \( \psi_{2^b} \) and Theorem 10 implies that \( \Omega_\pi \subseteq \Omega \), then \( \pi_{\alpha} = \sup_{\psi \in \Omega_\pi} P_{\psi,\alpha} \leq \sup_{\psi \in \Omega_\pi} P_{\psi,\alpha} \leq \sup_{\psi \in \Omega} P_{\psi,\alpha} = \pi_{\alpha} \). It means that \( \pi_{\alpha} = \sup_{\psi \in \Omega_\pi} P_{\psi,\alpha} \). It is known that for all test \( \psi_i \in \Omega_\pi \), each \( \psi_i \) depends on \( x \) just through \( m_m(x) \), hence based on the NP lemma [1] the optimal test is obtained by comparing the LR of \( m_m(x) \) with a threshold.

**B.4 Derivation of (4.7)**

Denoting \( x_{N-1} = r \exp(j\phi) \) and \( s = [s_0, \ldots, s_{N-1}]^T \), the conditional pdf of \( m_m = [m_0, \ldots, m_{N-1}]^T \) given \( r \) and \( \phi \) under \( \mathcal{H}_1 \) is Gaussian, i.e.,

\[
f(m_m | r, \phi; a, \sigma^2, \mathcal{H}_1) = \frac{\exp \left( -\frac{r^2}{\pi \sigma^2} \sum_{n=0}^{N-2} |m_n - a_n|^2 \right)}{\pi^N \sigma^{2N}} \times \delta(\text{Re}(m_{N-1}) - \cos(\phi)) \delta(\text{Im}(m_{N-1}) - \sin(\phi)). \tag{B.3}
\]

The Dirac delta function part of (B.3) is because \( m_{N-1} = \exp(j\phi) \). The pdf of \( m_m \) under \( \mathcal{H}_1 \) is given by \( \int_0^\infty \int_0^{2\pi} f(m_m | r, \phi; a, \sigma^2, \mathcal{H}_1) f(r, \phi; a, \sigma^2, \mathcal{H}_1) d\phi dr \), where \( f(r, \phi; \mathcal{H}_1) = \frac{r}{\pi \sigma^2} \exp(-\frac{|r\exp(j\phi) - a_{N-1}|^2}{\sigma^2}) \) is the joint pdf of \( r \) and \( \phi \). Integrating over \( \phi \), assuming \( \|s\|^2 = 1 \), defining \( \beta_1 = \|m_m\| \) and \( \beta_2 = \text{Re}(m_m^H as) \), we obtain

\[
f(m_m; a, \sigma^2, \mathcal{H}_1) = \int_0^\infty \frac{r^{2N-1}}{\pi^N \sigma^{2N}} e^{\frac{-1}{\sigma^2}(r^2 \beta_1^2 - 2r \beta_2 + |a|^2)} dr. \tag{B.4}
\]

Using \( r^2 \beta_1^2 - 2r \beta_2 + |a|^2 = (\beta_1 r - \frac{\beta_2}{\beta_1})^2 - (\frac{\beta_2}{\beta_1})^2 + |a|^2 \) and \( u = \frac{\beta_1 r - \beta_2}{\beta_1 \sigma} \), we have

\[
f(m_m; a, \sigma^2, \mathcal{H}_1) = \frac{e^{\frac{-1}{\sigma^2} |a|^2}}{\pi^N \beta_1^{2N}} \int_{-\infty}^{\infty} \frac{e^{\frac{-1}{\sigma^2} (u + \frac{\beta_2}{\sigma})^2}}{\pi^N \beta_1^{2N}} e^{-u^2} du. \]

Expanding \( (u + \frac{\beta_2}{\sigma})^{2N-1} \), we
rewrite $f(m_m; a, \sigma^2, \mathcal{H}_1)$ as (4.7) where $\rho \triangleq \frac{\sigma}{\sigma}$ and $\tau(n, x) \triangleq \int_{-\infty}^{\infty} u^{n-1} e^{-u^2} du$. Using $z = u^2$ this integral is given by $\tau(n, x) = \frac{1}{2} \Gamma(n/2, x^2)$ for $x \leq 0$ and $\tau(n, x) = \frac{1}{2} \Gamma(n/2, 0) + \frac{(-1)^{n-1}}{2} (\Gamma(n/2, 0) - \Gamma(n/2, x^2))$ for $x > 0$, where $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function.
Appendix C

Appendix to Chapter 5

C.1 Proof of Proposition 3

Proof 36 Assumptions yield \( f_x(x; \hat{\theta}(x)) = \max_{\theta \in \Theta} f_x(x; \theta) \). Thus for all \( q_m \in Q_m \), we have \( f_x(q_m(x); \hat{\theta}(q_m(x)))J = \max_{\theta \in \Theta} f_x(q_m(x); \theta)J \), where \( J \) is the jacobian of \( q_m(x) \) with respect to \( x \). Since \( f_x(x; \theta) = f_x(q_m(x); \bar{q}_m(\theta))J \), then \( f_x(x; \bar{q}_m^{-1}(\hat{\theta}(q_m(x)))) = \max_{\theta \in \Theta} f_x(x; \bar{q}_m(\theta)). \) The right hand side is \( \max_{\theta \in \Theta} f_x(x; \bar{q}_m(\theta)) = \max_{\theta \in \Theta} f_x(x; \theta) \) since \( \bar{q}_m^{-1}(\hat{\theta}(q_m(x))) \) is the MLE of \( \theta \), i.e., \( \bar{q}_m^{-1}(\hat{\theta}(q_m(x))) = \hat{\theta}(x) \). The invertible property of \( q_m \) provides that \( \hat{\theta}(q_m(x)) = \bar{q}_m^{-1}(\hat{\theta}(x)) \). We have \( \rho(\hat{\theta}(q_m(x))) = \rho(\bar{q}_m^{-1}(\hat{\theta}(x))) = \rho(\hat{\theta}(x)) \) since \( \rho \) is the maximal invariant of \( \theta \) under \( \bar{q}_m \).

C.2 Proof of Proposition 4

Proof 37 From \( \Pr(\hat{\rho} = \rho_0) = 0 \), we have \( \hat{\rho} = \hat{\rho}_1 \) almost everywhere where \( \hat{\rho} = \arg\max_{\rho \in \Omega \cup \{\rho_0\}} f_x(x; \rho) \) and \( \hat{\rho}_1 = \arg\max_{\rho \in \Omega} f_x(x; \rho) \). Hence, \( (\hat{\rho}_1 - \rho_0)^T \mathbf{I}(\hat{\rho}_1)(\hat{\rho}_1 - \rho_0) = (\hat{\rho} - \rho_0)^T \mathbf{I}(\hat{\rho})(\hat{\rho} - \rho_0) \) almost everywhere. Therefore defining \( g_W(\rho) \equiv (\rho - \rho_0)^T \mathbf{I}(\hat{\rho})(\hat{\rho} - \rho_0) \) almost everywhere.
\( \rho_0 \)\(^T I(\rho)(\rho - \rho_0) \), the Wald test in (5.1b) is equivalent to \( g_W(\hat{\rho}) \geq_{H_0} ^{H_1} \eta_W \) almost everywhere. We show that \( g_W(\rho) \) is an SF for (2.1), i.e., \( g_W(\rho) \) is continuous in \( \theta \) and \( g_W(\rho) = 0 \) for \( \theta \in \Theta_0 \) and \( g_W(\rho) > 0 \) for \( \theta \in \Theta_1 \).

The function \( g_W(\rho) \) is continuous. The existence of FIM implies that the second order derivatives \( \frac{\partial^2 f_{m}(m, \rho)}{\partial \rho^2} \) exist where \( \rho_l \) is the \( l \)th element of \( \rho \). Thus the gradient \( \frac{\partial f_{m}(m, \rho)}{\partial \rho} \) is continuous in \( \rho \). This means that \( I(\rho) = E_{\rho}(\frac{\partial f_{m}(m, \rho)}{\partial \rho} \frac{\partial f_{m}(m, \rho)}{\partial \rho})^T \) is continuous which implies that \( g_W(\rho) \) is continuous as product of continuous functions. Moreover, we have \( \lim_{\rho \to \rho_0} g_W(\rho) = 0 \). And \( g_W(\rho) > 0 \) since \( I(\rho) \) is a positive semidefinite.

C.3 Proof of Proposition 6

**Proof 38** The derivative of the expression of \( k(\gamma) \) leads to \( \frac{\partial k(\gamma)}{\partial \gamma} = \int_R \alpha f_x(x; \gamma)dx = E_\gamma(\alpha), \frac{\partial^2 k(\gamma)}{\partial \gamma^2} = E_\gamma(\alpha^2) - E_\gamma(\alpha)^2 \), where \( \alpha \triangleq \ln(f_x(x|H_1)/f_x(x|H_0)) \). Thus \( \frac{\partial^2 k(\gamma)}{\partial \gamma^2} \) is the variance of the log-likelihood function under \( f_x(x; \gamma) \) and is non-negative. Therefore \( \frac{\partial k(\gamma)}{\partial \gamma} \) is a non-decreasing function of \( \gamma \). The MLE of \( \gamma \) is a solution of \( \frac{\partial}{\partial \gamma} \ln(f(x; \gamma))|_{\gamma=\hat{\gamma}} = 0 \) which using (5.4) leads to \( \frac{\partial k(\gamma)}{\partial \gamma} |_{\gamma=\hat{\gamma}} = \ln\left(\frac{f(x|H_1)}{f(x|H_0)}\right) \), i.e., the MLE of \( \frac{\partial k(\gamma)}{\partial \gamma} \) is the log-likelihood function. Moreover, the function \( g_{NP}(\gamma) = \gamma - \frac{1}{2} \) is a SF, because \( g_{NP}(\gamma) = \gamma - \frac{1}{2} > 0 \) for \( \gamma > \frac{1}{2} \) and \( g_{NP}(\gamma) = \gamma - \frac{1}{2} \leq 0 \) for \( \gamma \leq \frac{1}{2} \). Thus, \( g_{NP}(\gamma) = \gamma - \frac{1}{2} \) is also an increasing function of \( \frac{\partial k(\gamma)}{\partial \gamma} \). Therefore this SFET is equivalent to compare the MLE of \( \frac{\partial k(\gamma)}{\partial \gamma} \) with a threshold which is the NP test.
C.4 Derivation of (5.24)

The pdfs of \( m \) given \( x_N \) under each hypothesis are

\[
f_m(m|x_n; H_1) = \frac{\delta(m_N - \frac{x_N}{|x_N|})|x_N|^{N-1}e^{-\frac{-(x_N-m)^2}{2\sigma^2}}}{(2\pi(\sigma^2 + \theta^2))^{N/2}}, \tag{C.1}
\]

where \( m_n \) is the \( m^{th} \) member of \( m \). The pdfs of \( m \) under \( H_0 \) are obtained by substituting \( \theta = 0 \) in the pdfs under \( H_1 \). Defining \( \sigma^2 \triangleq \mu^2 + \sigma^2 \), we have

\[
f_m(m; \rho, H_1) = \int_{-\infty}^{\infty} f_m(m|x_n; H_1) e^{-\frac{-(x_N-x)^2}{2\sigma^2}} \frac{1}{(2\pi\sigma^2)^{1/2}} dx_N, \tag{C.2}
\]

Replacing (C.1) into above relations and because of \( \delta(m_N - \frac{x_n}{|x_N|}) \), we have \( f_m(m; \rho, H_1) \) and \( f_m(m; H_0) \) are zero for \( m_N \neq \pm 1 \), and for \( m_N = \pm 1 \), \( \delta(m_N - \frac{x_n}{|x_N|}) \) implies that \( x_n = m_N|x_N| \), hence, in this case the pdfs is given by

\[
f_m(m; \rho, H_1) = \frac{2\exp\left(\frac{\rho^2}{2} \frac{(\sum_{n=1}^{N-1} m_n)^2}{\sum_{n=1}^{N-1} m_n^2} - 1\right)}{(2\pi\sigma^2)^{N/2}} \times \int_0^{\infty} x_N^{N-1} \exp(-A(x_N - B)^2) dx_N, \tag{C.3}
\]
where \( \rho' = \frac{\mu}{\sigma'} = \frac{|\rho|}{\sqrt{1+\rho^2}} \), \( A = \frac{\sum_{n=1}^{N} m_n^2}{2\sigma'^2} \) and \( B = \frac{\mu \sum_{n=1}^{N} m_n}{\sum_{n=1}^{N} m_n} \). Defining \( z = \sqrt{A}(x_N - B) \), we have

\[
 f_m(m; \rho, \mathcal{H}_1) = \frac{2 \exp \left( \frac{\rho'^2}{2} \frac{(\sum_{n=1}^{N} m_n^2)}{\sum_{n=1}^{N} m_n^2} - 1 \right)}{(2\pi\sigma'^2)^{N/2}} \times \int_{-\infty}^{\infty} \frac{B}{A^{N-1}} \exp(\frac{z^2}{2}) \, dz.
\]

(C.4)

where \( \tau(n, x) \triangleq \int_{-x}^{x} z^{N-1-i} \exp(-z^2) \, dz \). By substituting \( \theta = 0 \), we have

\[
 f_m(m; \mathcal{H}_0) = \frac{2\Gamma(N/2)}{(\pi \|m\|^2)^{N/2}}.
\]

(C.5)

The ratio of (C.4) and \( f_m(m; \mathcal{H}_0) \) is expressed in (5.24) is achieved.
Appendix D

Appendix to Section 7.2

D.1 Derivation of $\hat{\alpha}_{ML_{k+1}}$ and $\hat{\sigma}_{ML_{k+1}}^2$ when $\hat{\alpha}_{ML_k}$ is zero

In this appendix, we derive $\hat{\alpha}_{ML_{k+1}}$ and $\hat{\sigma}_{ML_{k+1}}^2$ when $\hat{\alpha}_{ML_k}$ tends to zero. We show that $\lim_{\hat{\alpha}_{ML_k} \to 0} \hat{\alpha}_{ML_{k+1}} = 0$ and $\lim_{\hat{\alpha}_{ML_k} \to 0} \hat{\sigma}_{ML_{k+1}}^2 = \frac{\|x\|_2^2}{N}$. We first prove that

$$\lim_{\hat{\alpha}_{ML_k} \to 0} \hat{\alpha}_{ML_{k+1}} = 0 \quad \text{and} \quad \lim_{\hat{\alpha}_{ML_k} \to 0} \hat{\sigma}_{ML_{k+1}}^2 = \frac{\|x\|_2^2}{N}. \quad (D.1)$$

Proof 39 Consider

$$\beta \int_{-\infty}^{\infty} \exp\left(\frac{-(x_n-s_n)^2}{2\hat{\sigma}^2} - \frac{|s_n|^\beta}{\hat{\alpha}^{\beta}}\right) ds_n = \frac{-x_n^2}{2\hat{\sigma}^2} \ast \frac{|s_n|^\beta}{\hat{\alpha}^{\beta}} \sqrt{2\pi \hat{\sigma} \hat{\alpha} \Gamma\left(\frac{1}{\beta}\right)} = \frac{-x_n^2}{2\hat{\sigma}^2} \ast \frac{|s_n|^\beta}{\hat{\alpha}^{\beta}} \sqrt{2\pi \hat{\sigma} \hat{\alpha} \Gamma\left(\frac{1}{\beta}\right)},$$

where $\ast$ is the convolution operator. Theorem 14.8 of [129] and $\lim_{\hat{\alpha} \to 0} \frac{\beta \exp\left(-|s_n|^\beta\right)}{2\hat{\alpha} \Gamma(1/\beta)} = \delta(x_n)$, imply that

$$\lim_{\hat{\alpha} \to 0} \frac{\beta \int_{-\infty}^{\infty} \exp\left(\frac{-(x_n-s_n)^2}{2\hat{\sigma}^2} - \frac{|s_n|^\beta}{\hat{\alpha}^{\beta}}\right) ds_n}{\sqrt{2\pi \hat{\sigma} \hat{\alpha} \Gamma(1/\beta)}} = \frac{\exp\left(-x_n^2/2\hat{\sigma}^2\right)}{\sqrt{2\pi \hat{\sigma}}}.$$
Now, we use (D.1) in the denominator of (7.20a) as follows:

$$\lim_{\hat{\alpha}_{ML} \to 0} \hat{\alpha}_{ML_k+1} = \lim_{\hat{\alpha}_{ML} \to 0} \left( \frac{\beta^2}{2NT(1/\beta)} \right)^{\frac{1}{\beta}} \times$$

$$\times \left( \sum_{n=1}^{N} \int_{-\infty}^{\infty} \frac{|s_n|^\beta \exp \left( \frac{-|s_n|^2}{2\hat{\sigma}^2_{ML_k}} \right)}{\frac{|s_n|^\beta \hat{\alpha}_{ML_k}}{\hat{\sigma}^2_{ML_k}}} ds_n \right)^{\frac{1}{\beta}}.$$

Since $$\lim_{\hat{\alpha}_{ML} \to 0} \frac{\beta \exp \left( -|s_n|^\beta \right)}{2\hat{\sigma}^2_{ML_k} \Gamma(1/\beta)} = \delta(s_n)$$, Theorem 14.8 of [129] implies that

$$\lim_{\hat{\alpha}_{ML} \to 0} \hat{\alpha}_{ML_k+1} = \left( \frac{\beta^2}{2NT(1/\beta-1)} \sum_{n=1}^{N} \int_{-\infty}^{\infty} |s_n|^\beta \delta(s_n) ds_n \right)^{\frac{1}{\beta}} = 0.$$

Now, we derive $$\lim_{\hat{\alpha}_{ML} \to 0} \hat{\sigma}^2_{ML+1}$$. Applying (D.1), on (7.20b), we have

$$\lim_{\hat{\alpha}_{ML} \to 0} \hat{\sigma}^2_{ML_k+1} =$$

$$\lim_{\hat{\alpha}_{ML} \to 0} \sum_{n=1}^{N} \frac{\beta \int_{-\infty}^{\infty} (x_n-s_n)^2 \exp \left( -\frac{(x_n-s_n)^2}{2\hat{\sigma}^2_{ML_k}} \right) \frac{|s_n|^\beta}{\hat{\sigma}^2_{ML_k}} ds_n}{2\hat{\sigma}^2_{ML_k} \Gamma(\beta-1) \hat{\alpha}_{ML_k}} = \frac{\|x\|^2}{2N}.$$
prove the invariance of $SFET_{ML}$, $SFET_{MS}$ and $SFET_{MA}$ using mathematical induction. Hence

- first we assume that for a given $k$, $\hat{\alpha}_{ML_k}(\xi x) = \xi \hat{\alpha}_{ML_k}(x)$, $\hat{\sigma}_{ML_k}^2(\xi x) = \xi^2 \hat{\sigma}_{ML_k}^2(x)$, $\hat{s}_{MS_k}(\xi x) = \xi \hat{s}_{MS_k}(x)$, $\hat{\alpha}_{MS_k}(\xi x) = \xi \hat{\alpha}_{MS_k}(x)$, $\hat{\sigma}_{MS_k}^2(\xi x) = \xi^2 \hat{\sigma}_{MS_k}^2(x)$, where

$$
\hat{\sigma}_{ML_k}^2(x) = \hat{\sigma}_{ML_k}^2,
$$

$\hat{\sigma}_{MS_k}^2(x) = \hat{\sigma}_{MS_k}^2$, are defined in (7.21a) and (7.24).

- From (7.21a) and (7.24) for $(k+1)^{\text{th}}$ iteration, we have $\hat{\alpha}_{ML_{k+1}}(\xi x) = \xi \hat{\alpha}_{ML_{k+1}}(x)$, $\hat{\sigma}_{ML_{k+1}}^2(\xi x) = \xi^2 \hat{\sigma}_{ML_{k+1}}^2(x)$, $\hat{s}_{MS_{k+1}}(\xi x) = \xi \hat{s}_{MS_{k+1}}(x)$, $\hat{\alpha}_{HMS_{k+1}}(\xi x) = \xi \hat{\alpha}_{HMS_{k+1}}(x)$, $\hat{\sigma}_{HMS_{k+1}}^2(\xi x) = \xi^2 \hat{\sigma}_{HMS_{k+1}}^2(x)$.

- For the initial points for $SFET_{ML}$ and $SFET_{MS}$ are $\hat{\alpha}_{M}$ and $\hat{\sigma}_{M}^2$. moreover since $\hat{\alpha}_M(\xi x) = \xi \hat{\alpha}_M(x)$ and $\hat{\sigma}_M^2(\xi x) = \xi^2 \hat{\sigma}_M^2(x)$, then for all arbitrary $k$, we have

$$
\frac{(\hat{\alpha}_{ML_k}(\xi x))^2}{\hat{\sigma}_{ML_k}^2(x)} = \frac{(\hat{\alpha}_{MS_k}(\xi x))^2}{\hat{\sigma}_{MS_k}^2(x)} = \frac{(\hat{\alpha}_{MS_k}(\xi x))^2}{\hat{\sigma}_{MS_k}^2(x)}.
$$

These results show that $SFET_{ML}$ and $SFET_{MS}$ are invariant with respect to $G_m$. The $SFET_{MA}$ and $SFET_F$ are obtained by solving

$$
\hat{s}_{MA}(x) \triangleq \arg\min_{s \in \mathbb{R}^N} \frac{|x-s|^2}{2\hat{s}_M^2(x)} + \frac{\|s\|^2}{\alpha_M(x)^2}, \text{ which yields }
$$

$$
\hat{s}_{MA}(\xi x) = \arg\min_{s \in \mathbb{R}^N} \frac{|x-s/\xi|^2}{2\hat{s}_M(\xi x)} + \frac{\|s/\xi\|^2}{\alpha_M(\xi x)^2} = \xi \hat{s}_M(x).
$$

This implies that for $SFET_{MA}$ and $SFET_F$ (i.e., the estimators of $\hat{\alpha}_M^2/\hat{\sigma}_M^2$) are invariant with respect to the scale parameter, i.e., $\frac{\hat{\alpha}_{HMA}(\xi x)}{\hat{\sigma}_{HMA}^2(\xi x)} = \frac{\hat{\alpha}_{HMA}(x)}{\hat{\sigma}_{HMA}^2(x)}$ and $\frac{\hat{\alpha}_F(\xi x)}{\hat{\sigma}_F^2(\xi x)} = \frac{\hat{\alpha}_F(x)}{\hat{\sigma}_F^2(x)}$. 


\[ f_m(m|s, \mathcal{H}_1) = \int_{-\infty}^{\infty} |x_N|^{N-1} \exp\left(-\frac{|x_N|^2}{2\sigma^2}\right) \frac{N^{-1}}{(2\pi\sigma^2)^{(N-1)/2}} \delta(m_N - \frac{x_N}{|x_N|}) \]

\[ \exp\left(-\frac{|x_N - x_N|^2}{2\sigma^2}\right) \frac{\sqrt{2\pi\sigma^2}}{\sqrt{\beta}} \right) \frac{\beta^{N}}{(2\alpha\Gamma(1/\beta))} \exp\left(-\frac{|s|^2}{\alpha\beta}\right) ds. \quad (D.3c) \]

**D.3 Proof of (D.3c)**

The pdf of \( x \) given \( s \) is Gaussian with mean \( s \) and variance \( \sigma^2 \). Thus the pdf of \( m_n = \frac{x_n}{|x_N|} \) given \( s \) and \( x_N \) is Gaussian with mean \( \frac{s_n}{|x_N|} \) and variance \( |x_N|^2\sigma^2 \), for \( n = 1, \cdots, N - 1 \), and \( m_N \) is a deterministic value \( \frac{x_N}{|x_N|} \) given \( s \) and \( x_N \). Thus the pdf of \( m(x) \) given \( s \) is obtained by integrating over the pdf of \( x_N \) as in (D.3a), where \( \delta(\cdot) \) is delta function. The delta function in integral implies \( x_N = m_N|x_N| \) which simplifies to left hand side of (D.3b). Using the expression for the absolute moment of a Gaussian random variable, the later equation is further simplified to the right hand side of (D.3b), where \( _1F_1(\cdot; \cdot; \cdot) \) is a confluent hypergeometric function. Using total probability rule, the pdf of \( m \) under \( \mathcal{H}_1 \) achieving from \( \int_{\mathbb{R}^N} f_m(m|s, \mathcal{H}_1)f(s; \alpha)ds \) is calculated in (D.3c).