2016

Multimodal Moving Target Detection and Data Processing Via Distributed Sensors

Fuat Cogun
University of Rhode Island, fuat.cogun@gmail.com

MULTIMODAL MOVING TARGET DETECTION AND DATA PROCESSING VIA DISTRIBUTED SENSORS

BY

FUAT COGUN

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2016

DOCTOR OF PHILOSOPHY DISSERTATION

OF

FUAT COGUN

APPROVED:

Dissertation Committee:

Major Professor Steven M. Kay

Peter F. Swaszek

Jean-Yves Herve

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2016

ABSTRACT

The detection problem of a target based on radio-frequency (RF) and infrared (IR) data sources is addressed in this dissertation. The target is assumed to radiate an RF signal to multiple widely distributed sensors in space and is concurrently imaged using multiple frames of an IR sensor whose field-of-view is fixed with respect to the background. The observations contain additive noise in both sources due to the physical nature of the problem and sensor imperfections.

First, we define accurate signal models to be used in hypothesis testing for the RF and IR data sources. Second, the generalized likelihood ratio test (GLRT) statistics for both RF and IR detection problem are derived. Once the GLRT statistics are found, the optimal integration of test statistics is done to obtain the integrated GLRT detector. Although the integration is done optimally, the integrated GLRT requires; the joint maximization of a highly nonlinear statistic over target motion parameters, which is computationally expensive, and the submission of all RF data observed at local sensors to the central processor (CP), which requires high communication bandwidth and a CP having large memory. The random basis functions (RBF) approach is proposed as a computationally efficient method to reduce the computational complexity and data compression techniques are proposed for the distributed detection of the RF signal.

The RBF approach is applied to IR image sequences and it is shown that the reduction in computations is substantial as the dimension of the unknown parameter space is high. However, it causes some performance loss when compared to GLRT detector, thus, the approach requires higher signal-to-noise ratio to operate. Another possible use of the RBF approach is to apply it to reduce down the search space for the RF detector instead of applying it until convergence. Then, the GLRT for the RF data is run over the reduced down search space to obtain

the multimodal estimates of the target motion parameters. This alternative representation of the multimodal detector is also implemented and the localization performance is tested. It is seen that the multimodal detector is more robust to partial occlusion when compared to detection using RF sensors only.

For the RF signal detection problem, the GLRT requires the submission of all data obtained at local sensors to the CP where the maximization takes place. Although this classical centralized detector has asymptotically optimal detection performance, submission of all observed data to the CP is practically infeasible as the RF detection problem is typically based on large data records. Therefore, distributed detection methods are considered for this problem and novel approaches based on Taylor expansions are proposed. Observed data is compressed into local test statistics at each sensor and transmitted to the CP for the formation of overall test statistic. The GLRT detector performance is used as the upper bound to assess the performance of the proposed compression techniques in this work.

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my advisor Dr. Steven Kay for his guidance and continuous support of my Ph.D. studies. I am grateful for the time and effort he spent on me. I could not have imagined having a better advisor and mentor for my Ph.D study. Without his immense knowledge and motivation, this dissertation would not have been possible.

Besides my advisor, I would like to thank the rest of my thesis committee members: Dr. Peter F. Swaszek, Dr. Jean-Yves Herve, Dr. Ramdas Kumaresan and Dr. Tom Bella for their help and efforts in participating in my comprehensive exam and dissertation defense.

I would also like to thank my parents and my sister for their continuous moral and financial support during my study. They have always been very supportive in my life and I am so blessed to have such a wonderful family.

Last but not the least, I would like to thank my beloved wife Tugce who has always been there standing by me through the good times and bad. Her endless encouragement, patience and love made everything possible. I would also like to thank my son Can, who is a source of joy, for always making me smile and cheering me up.

PREFACE

This dissertation is constructed in the manuscript format and consists of five manuscripts.

In Chapter 1, the IR and RF signal models considered in this research are presented and the GLRT statistics for both RF and IR data are found. The optimal integration of the GLRT detectors is done to obtain the multimodal GLRT detector, assuming that the observation noise is independent between each dataset. Additionally, the RBF approach is introduced as a potential solution to decrease the computational complexity associated with the integrated GLRT detector.

Chapter 2 focuses on target detection based on IR data. The detailed derivation of the GLRT detector is included and detector performance analysis is done. An occluded IR scene is studied and the performance loss caused by occlusion is given. RBF approach is fully implemented and the detection performance of the RBF-based detector is presented. To show the computational gain obtained by the use of the RBF approach, computational count of the RBF approach is compared to the straightforward grid-search. Finally, the performance comparison of the GLRT and RBF-based detectors is given.

An alternative approach to the multimodal target detection problem is considered in Chapter 3. As the integrated GLRT detector requires the computation of the maximum likelihood estimates of target motion parameters, the problem may also be viewed as a localization problem. In this chapter, first, the RBF approach is applied for a predetermined number of iterations (not until convergence) to reduce down the search space for the RF detector. Second, the GLRT for the RF data is run over the reduced down search space to obtain multimodal estimates. The robustness of this suboptimal multimodal detector to RF occlusion is presented by simulations.

In Chapter 4, the distributed detection problem of a signal embedded in white Gaussian noise is considered. The GLRT detector is derived for both linear and nonlinear (NL) signal model. It is shown that the GLRT requires all observed data to be submitted the central processor which should be avoided in practice. For that matter, we propose data compression techniques based on Taylor approximation for efficient distributed detection. The GLRT detector performance is used as the upper bound to assess the performance of the proposed detectors.

Chapter 5 addresses the generalized distributed detection problem for the case when a NL signal is a function of an unknown parameter vector varying from sensor to sensor. Note that the RF LPI signal detection problem considered in this study is covered under this generalized NL signal model. We propose a data compression technique based on the second-order Taylor approximation of the local log-likelihood functions and provide detailed derivations. The performance of the proposed detector is compared to the GLRT detector for the linear frequency modulation (LFM) sweep signal detection problem, since LFM sweep is commonly used in radar signal processing.

Finally, some possible future research direction ideas are given in Chapter 6.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Integrated Sensor Detection/Localization for Multi-Source Data

by

Steven Kay and Fuat Cogun

Dept. of Electrical, Computer and Biomedical Engineering

University of Rhode Island, Kingston, RI, USA

published in IEEE Radar Conference, 2014.

Abstract

In this paper, the detection/localization of a target based on radio-frequency (RF) and infra-red (IR) data sources problem is addressed. The target is assumed to radiate an RF signal to multiple widely distributed sensors in space and is imaged using multiple frames of an IR sensor. The goal is to integrate RF and IR data to reliably detect and localize the target. The generalized likelihood ratio test (GLRT) approach is employed to find the detector. In order to reduce the computation required by a straightforward GLRT, the random basis function (RBF) approach is used.

1.1 Introduction

The main concern of this work is to detect and localize a constant velocity ground moving vehicle emitting an unknown signal. The target is assumed to radiate an RF signal to multiple widely distributed sensors in space. In addition, the target is assumed to be imaged using multiple frames of an IR sensor at the same time.

Some of the difficulties encountered in detecting and localizing the target are the following. The received RF low probability of intercept (LPI) signal is unknown and it may be subject to multipath [1]. The received IR image signal may be temporarily occluded by trees, buildings, etc., or the vehicle may stop and appear as part of background. It is clear that these difficulties cause lower detection performance if a single RF or IR detector is considered for the problem. Therefore, fusion of RF and IR data is necessary to overcome these difficulties and leads to better performing detectors.

In this paper, we find individual GLRTs for both RF and IR signal models, then integrate them to find a GLRT detector based on the fusion of RF and IR data.

1.2 Signal Model

Signal modeling is crucial in designing a detector well-suited to the problem. Selected models should be as accurate as possible to minimize the modeling errors. In this section, we introduce the signal models used for the RF and IR signals and the reasoning behind them.

1.2.1 RF Signal Model

The hypothesis test for the RF signal is

$$\begin{aligned} \mathcal{H}_0 &: \tilde{x}_i[n] = \tilde{w}_i[n] \\ \mathcal{H}_1 &: \tilde{x}_i[n] = \tilde{A}_i \tilde{s}[n - n_i] e^{\frac{j2\pi k_i (n-n_i)}{N}} + \tilde{w}_i[n] \end{aligned} \tag{1}$$

where $\tilde{w}_i[n] \sim \mathcal{CN}(0, \sigma^2)$, $n = 0, 1, \ldots, N-1$ and $i = 0, 1, \ldots, M-1$. A tilde ($\sim$) is used throughout to indicate that we are dealing with complex data. Here, $\tilde{s}[n]$ is the n^{th} time sample of the low probability of intercept (LPI) transmit signal, $\tilde{x}_i[n]$ is the n^{th} time sample observed, $\tilde{A}_i$ is the complex gain/attenuation, n_i is the time delay, k_i/N is the Doppler shift and $\tilde{w}_i[n]$ is the n^{th} time sample of the noise at i^{th} receiver. It is important to note that $\tilde{A}_i$, $\tilde{s}[n]$, n_i and k_i are all unknown. Here $\mathcal{H}_0$ signifies no target, while $\mathcal{H}_1$ signifies a target present. This problem is one of composite hypothesis testing.

Now the data model can be written in more compact form as

$$\tilde{\mathbf{x}}_i = \tilde{A}_i \mathbf{P}^{n_i} \tilde{\mathbf{W}}^{k_i} \tilde{\mathbf{s}} + \tilde{\mathbf{w}}_i \tag{2}$$

which is the observed data at the i^{th} sensor over a time interval of length N samples. That is $\tilde{\mathbf{x}}_i = [\tilde{x}_i[0]\ \tilde{x}_i[1]\ \ldots\ \tilde{x}_i[N-1]]^T$. Similarly, $\tilde{\mathbf{w}}_i = [\tilde{w}_i[0]\ \tilde{w}_i[1]\ \ldots\ \tilde{w}_i[N-1]]^T$ and $\tilde{\mathbf{s}} = [\tilde{s}[0]\ \tilde{s}[1]\ \ldots\ \tilde{s}[N-1]]^T$. $\mathbf{P}$ is an $N \times N$ permutation matrix providing time shifts and $\tilde{\mathbf{W}} = \text{diag}(\omega^0, \omega^1, \ldots, \omega^{N-1})$ providing Doppler shifts where $\omega =$

$\exp\left(\frac{j2\pi}{N}\right)$. Then the composite hypothesis testing problem becomes:

$$\mathcal{H}_0 : \tilde{\mathbf{s}} = \mathbf{0}$$
$$\mathcal{H}_1 : \tilde{\mathbf{s}} \neq \mathbf{0} \tag{3}$$

1.2.2 IR Signal Model

The hypothesis test for the IR signal is

$$\mathcal{H}_0 : x[m,n,k] = B[m,n] + w[m,n,k]$$
$$\mathcal{H}_1 : x[m,n,k] = As[m,n,k] + B[m,n] + w[m,n,k] \tag{4}$$

for $m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$ (image pixels) and $k = 0, 1, \ldots, K-1$ (frame number). In the IR signal model, all data is real, not complex as in RF signal model.

The background $B[m,n]$ is constant with frame and unknown. Additive noise $w[m,n,k]$ is assumed to be white Gaussian noise with variance σ^2 and independent from pixel-to-pixel and frame-to-frame. We can explicitly write $s[m,n,k]$ as:

$$s[m,n,k] = g[m - m_0 - v_x k, n - n_0 - v_y k] \tag{5}$$

where (m_0, n_0) is the starting position of the target and (v_x, v_y) is its velocity, both of which are unknown. The velocity of the target is assumed to be constant throughout the data collection time interval. The target is assumed to be point-like so the signal $s[m,n,k]$ is represented as one pixel. Note that the assumed IR signal hypothesis is also a composite hypothesis testing problem.

1.3 Generalized Likelihood Ratio Test (GLRT)

Asymptotically, the GLRT is the uniformly most powerful test among all tests that are invariant. The GLRT replaces the unknown parameters in the Neyman-Pearson optimum test statistic with their maximum likelihood estimates (MLEs).

The GLRT decides $\mathcal{H}_1$ if [2]

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_0, \mathcal{H}_0)} > \gamma \tag{6}$$

where p is the probability distribution function (PDF) , $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}_1$ assuming $\mathcal{H}_1$ is true and $\hat{\boldsymbol{\theta}}_0$ is the MLE of $\boldsymbol{\theta}_0$ assuming $\mathcal{H}_0$ is true. In our problem, $\boldsymbol{\theta}_1$, the unknown parameter vector under $\mathcal{H}_1$, consists of the emitted signal, propagation/sensor gains, target position and velocity.

Although there is no optimality associated with the GLRT, in practice it appears to work quite well. We make use of GLRTs for both the RF and IR signal in this work.

1.3.1 GLRT for the RF signal

The GLRT for the data observed at M sensors for a time interval of length N is [1]

$$\max_{(\mathbf{r}_0, \mathbf{v})} \lambda_{\max}(\tilde{\mathbf{C}}(\mathbf{r}_0, \mathbf{v})) \tag{7}$$

where $\mathbf{r}_0$ is initial target position, $\mathbf{v}$ is velocity, and $\lambda_{\max}$ is the maximum eigenvalue of the $M \times M$ complex ambiguity matrix (CAM) $\tilde{\mathbf{C}}$ defined by

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_0 & \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_1 & \cdots & \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_{M-1} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_0 & \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_1 & \cdots & \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_{M-1} \end{bmatrix} \tag{8}$$

where $\tilde{\mathbf{y}}_i = (\mathbf{P}^{n_i} \tilde{\mathbf{W}}^{k_i})^H \tilde{\mathbf{x}}_i$. Note that $\tilde{\mathbf{y}}_i$ is the aligned signal in delay and Doppler at sensor i.

In this case, once the maximum of $\lambda_{\max}(\tilde{\mathbf{C}})$ is found, then estimates (MLEs) of $\mathbf{r}_0$ (initial position) and $\mathbf{v}$ (velocity) are also available. The delays and Doppler shifts are a function of the initial target position and velocity, i.e., $\mathbf{n} = \mathbf{f}(x_0, y_0)$ and $\mathbf{k} = \mathbf{g}(x_0, y_0, v_x, v_y)$. The values $(\hat{x}_0, \hat{y}_0, \hat{v}_x, \hat{v}_y)$ that maximize the maximum eigenvalue $\lambda_{\max}(\tilde{\mathbf{C}}(x_0, y_0, v_x, v_y))$ are the MLEs for (x_0, y_0, v_x, v_y), respectively. It

is difficult to compute these MLEs analytically (maximization of $\lambda_{\max}(\tilde{\mathbf{C}})$ must be done). Any existing non-linear optimization technique or brute-force search technique can be used.

1.3.2 GLRT for the IR signal

The GLRT with the unknown parameters (m_0, n_0) and (v_x, v_y) is

$$\max_{(m_0,n_0),(v_x,v_y)} 2\ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} \tag{9}$$

where $\mathbf{X}$ is Gaussian, $\hat{\mathbf{B}}_0$ is the MLE of the background under $\mathcal{H}_0$, $\hat{A}$ is the MLE of the signal amplitude under $\mathcal{H}_1$, and $\hat{\mathbf{B}}_1$ is the MLE of the background under $\mathcal{H}_1$.

In order to utilize the multivariate Gaussian PDF for the image frame data, it is necessary to roll out the columns to form a column vector. A simple example is to roll out the columns of

$$\mathbf{X} = \begin{bmatrix} x[0,0] & x[0,1] \\ x[1,0] & x[1,1] \\ x[2,0] & x[2,1] \end{bmatrix} \quad 3 \times 2 \text{ image}$$

to yield

$$\mathbf{x} = \begin{bmatrix} x[0,0] & x[1,0] & x[2,0] & x[0,1] & x[1,1] & x[2,1] \end{bmatrix}^T.$$

The problem here is that neighboring pixels become widely dispersed. It becomes difficult to realize the detector and also adds to coding complexity. Note that the problem becomes even worse when the data is 3-dimensional as in $x[m, n, k]$. Hence, it is better to retain the $\mathbf{X}$ structure and more convenient to use the matrix normal PDF. Hence, we use instead

$$\mathbf{X}_k = \begin{bmatrix} x[0,0,k] & x[0,1,k] \\ x[1,0,k] & x[1,1,k] \\ x[2,0,k] & x[2,1,k] \end{bmatrix} \quad 3 \times 2 \text{ image for } k\text{th frame.}$$

The PDF of $\{\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{K-1}\}$ becomes

$$p(\mathbf{X}; A, \mathbf{B}, \mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{MNK/2}} \exp\left[-\frac{1}{2\sigma^2} J(A, \mathbf{B})\right]$$

where

$$J(A, \mathbf{B}) = \sum_{k=0}^{K-1} \mathrm{Tr}[(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})^T]$$

Also a more compact expression for J is

$$J = \sum_{k=0}^{K-1} \langle \mathbf{X}_k - A\mathbf{S}_k - \mathbf{B}, \mathbf{X}_k - A\mathbf{S}_k - \mathbf{B} \rangle \tag{10}$$

where $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}\mathbf{Y}^T)$. The latter is a 2-dimensional inner product on an inner product space of $M \times N$ images. The inner product formulation can easily be extended to 3-dimensional "data cubes" to accommodate frames as well.

The required MLE under $\mathcal{H}_0$ is

$$\hat{\mathbf{B}}_0 = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{X}_k \tag{11}$$

and of $A, \mathbf{B}$ under $\mathcal{H}_1$

$$\hat{\mathbf{B}}_1 = \frac{1}{K} \sum_{k=0}^{K-1} (\mathbf{X}_k - \hat{A}\mathbf{S}_k) = \bar{\mathbf{X}} - \hat{A}\bar{\mathbf{S}} \tag{12}$$

$$\hat{A} = \frac{\sum_{k=0}^{K-1} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=0}^{K-1} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \tag{13}$$

Finally the GLRT is

$$2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} = \frac{\sum_{k=0}^{K-1} ||\mathbf{S}_k - \bar{\mathbf{S}}||^2}{\sigma^2} \left[\frac{\sum_{k=0}^{K-1} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=0}^{K-1} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \right]^2 \tag{14}$$

For unknown starting position and velocity we must maximize this over (m_0, n_0) and (v_x, v_y) ($\mathbf{S}_k$'s depend on this).

The power of the GLRT is that the MLEs not only provide information that summarizes the presence of a target, i.e., used for detection, but also provide information that can be used for estimation of target trajectory and can be used for classification.

Here are the steps for finding the MLEs and GLRT, providing a verbal explanation of the procedure:

1. Estimate background (constant with frame number) by using the sample mean of the image frames

2. Subtract out the background estimate from all image frames

3. Subtract out the sample mean of signal images (since we can't distinguish between the background and a zero velocity target)

4. Correlate the residual image against expected signal (will change with the frame since target is assumed moving)

5. Add some normalization to allow the PDF of the GLRT to be determined and hence the thresholding required.

Note that we have assumed an inner product space of images (2-dimensional) with independent noise from frame to frame. But it can be extended to a 3-dimensional inner product space to allow representation of an entire video stream, where $\mathbf{X}$ and $\mathbf{Y}$ becomes $M \times N \times K$ data cubes, and the inner product becomes

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} x[m,n,k] y[m,n,k]. \tag{15}$$

This approach employs the coordinate-free approach to linear models. [3]

1.4 Integration of RF and IR GLRTs

Assuming the observation noise is independent between each data set we can derive the integrated GLRT from the RF and IR GLRTs. Let $\tilde{\mathbf{x}}_1$ be the complex RF data vector and $\mathbf{X}_2$ be the real IR data cube, and σ_1^2, σ_2^2 be the noise powers, respectively. Then the overall GLRT can be shown to be

$$L_G(\tilde{\mathbf{x}}_1, \mathbf{X}_2) = \max_{\mathbf{r}_0, \mathbf{v}} \left[\frac{\lambda_{\max}(\tilde{\mathbf{C}}(\tilde{\mathbf{x}}_1))}{\sigma_1^2} + \frac{\langle \mathbf{X}_2 - \bar{\mathbf{X}}_2, \mathbf{S} - \bar{\mathbf{S}} \rangle^2}{2\sigma_2^2 \langle \mathbf{S} - \bar{\mathbf{S}}, \mathbf{S} - \bar{\mathbf{S}} \rangle} \right] \tag{16}$$

where $\tilde{\mathbf{C}}$ is the complex ambiguity function matrix and $\mathbf{r}_0, \mathbf{v}$ are the initial target position and velocity vector. Note that the inner product $\langle \mathbf{X}, \mathbf{Y} \rangle$ is defined in (15).

The likelihood ratios of the joint RF and IR problem decouples since the noise is the only stochastic part and is independent between the RF and IR sensors. Realize that the overall statistic is maximized jointly over target motion parameters. This makes the computations very difficult. Therefore, an approach that will decrease the computational complexity is required.

1.5 RBF approach

The random basis function (RBF) approach converts difficult non-linear least squares problems to a succession of linear least squares problems. It is a bisection approach and there are no convergence issues. The performance is suboptimal since the approach requires higher SNR, but it can reduce the computation by orders of magnitude.

Consider the problem of fitting a signal with unknown amplitude and non-linear parameters to data. The non-linear least squares method tries to minimize the error function

$$J(A, \boldsymbol{\theta}) = \sum_{n=0}^{N-1} (x[n] - As[n; \boldsymbol{\theta}])^2 \tag{17}$$

over amplitude A and non-linear parameters $\boldsymbol{\theta}$. If we consider frequency estimation as an example, the signal is $s[n] = A\cos(2\pi\theta n)$, where $\theta = f_0$ is the non-linear signal parameter. Estimating f_0 usually requires a brute force search such as an FFT. To have a precise frequency estimate, a very large FFT size is required.

The *random basis function approach* [4] proceeds as follows. Consider $s[n; \boldsymbol{\theta}]$ as a basis function with a random parameter $\boldsymbol{\theta}$. The *expected least squares* criterion is defined as

$$J(A) = E_{\boldsymbol{\theta}} \left[\sum_{n=0}^{N-1} (x[n] - As[n; \boldsymbol{\theta}])^2 \right] \tag{18}$$

Note that we need to minimize over A only since this criterion gives the average amplitude over all possible values of $\boldsymbol{\theta}$. The main idea of the RBF approach is to split $\boldsymbol{\theta}$ into two disjoint regions, then apply the expected least squares criterion and

find the minimizing A's corresponding to the intervals. Then we choose the band with the larger A^2 and repeat subdividing the tentative band until convergence. If we apply the approach to the frequency estimation example, we need to:

1. Split the possible frequency band into two disjoint intervals

2. Assign a uniform PDF to the random frequency for each interval

3. Minimize over A_1 and A_2 the expected least squares

$$J(A_1, A_2) = E_{\theta_1,\theta_2}\left[\sum_{n=0}^{N-1}\left(x[n] - A_1\cos(2\pi\theta_1 n) - A_2\cos(2\pi\theta_2 n)\right)^2\right]$$

4. Choose the band with the larger A^2 and repeat subdividing the tentative band until convergence.

5. Choose the estimate of θ as midpoint of the last subdivided interval.

The RBF approach can be applied to image sequences. The data is given as

$$x[m, n, k] = As[m - m_0 - v_x k, n - n_0 - v_y k] + w[m, n, k] \tag{19}$$

Here the unknown non-linear parameter is $\boldsymbol{\theta} = [m_0\, n_0\, v_x\, v_y]^T$. As an example, consider the simpler problem of 1-D motion. In this case the data becomes

$$x[m, k] = As[m - m_0 - v_x k] + w[m, k] \tag{20}$$

where we let $\boldsymbol{\theta} = [m_0\, v_x]^T$. As there are more than one non-linear parameter, we cycle between the parameters at each iteration. That is, the RBF approach is applied to one parameter at a time. After choosing the interval for that parameter, we continue to operate on the other parameter until all the parameters are visited on $\boldsymbol{\theta}$. This ends one iteration and we iterate until all non-linear parameters converge. An illustration of cycling is given in Fig. 1 for the 1-D motion example. In this particular case, the unknown parameter v_x (target velocity) is assumed to be in

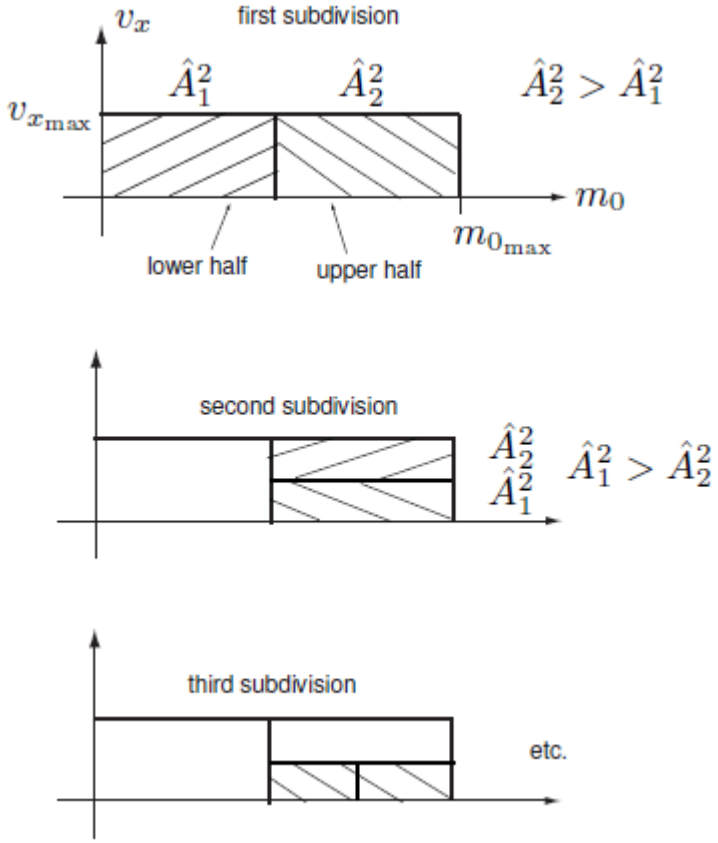


Figure 1: RBF approach - Cycling between m_0 and v_x non-linear parameters

the interval $(-1, 1)$, i.e., $v_x \in (-1, 1) \subset \mathcal{R}$, and the unknown parameter m_0 (target initial position) is assumed to be in the interval $[-60, 60]$, i.e., $m_0 \in [-60, 60] \subset \mathcal{Z}$. Note that the target velocity can be any real value and the target initial position can be any integer in the corresponding intervals. For clarity, the first 5 iterations of the RBF approach are shown in Fig. 2. The estimates converge to the true target velocity and initial target position ($v_x = -0.3, m_0 = 35$) in a few iterations. The convergence of the RBF approach is 15 times faster when compared with the straightforward non-linear least squares method. Hence, the approach decreases the computational complexity significantly. It is especially important to note that when the unknown parameter space dimension increases, the reduction in computations becomes much more substantial. In Fig. 3, the performance analysis of

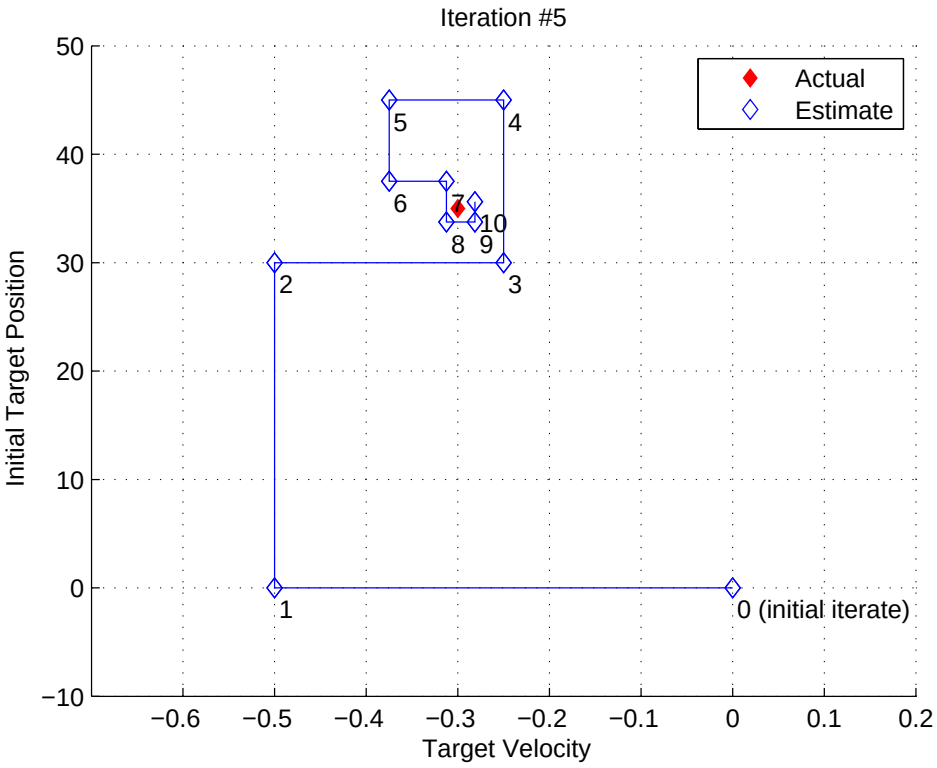


Figure 2: RBF approach - First 5 iterations (10 cycles)

the approach for the v_x parameter is given. Note that the MSE vs. SNR curves are obtained assuming unknown v_x, known m_0 parameter, since straightforward non-linear least squares method is computationally inefficient and a reduction in the unknown parameter space dimension was necessary to run enough realizations to obtain smooth curves.

1.6 Conclusion

In this paper, RF and IR signal models are used to determine the individual GLRTs. Then the optimal integration of RF and IR data for target detection/localization is obtained by using an integrated GLRT. It is seen that a straightforward GLRT requires a very large computation to maximize over the initial position and velocity. Therefore, we make use of the random basis function (RBF) approach. Since the RBF approach is not optimal, a higher SNR is needed

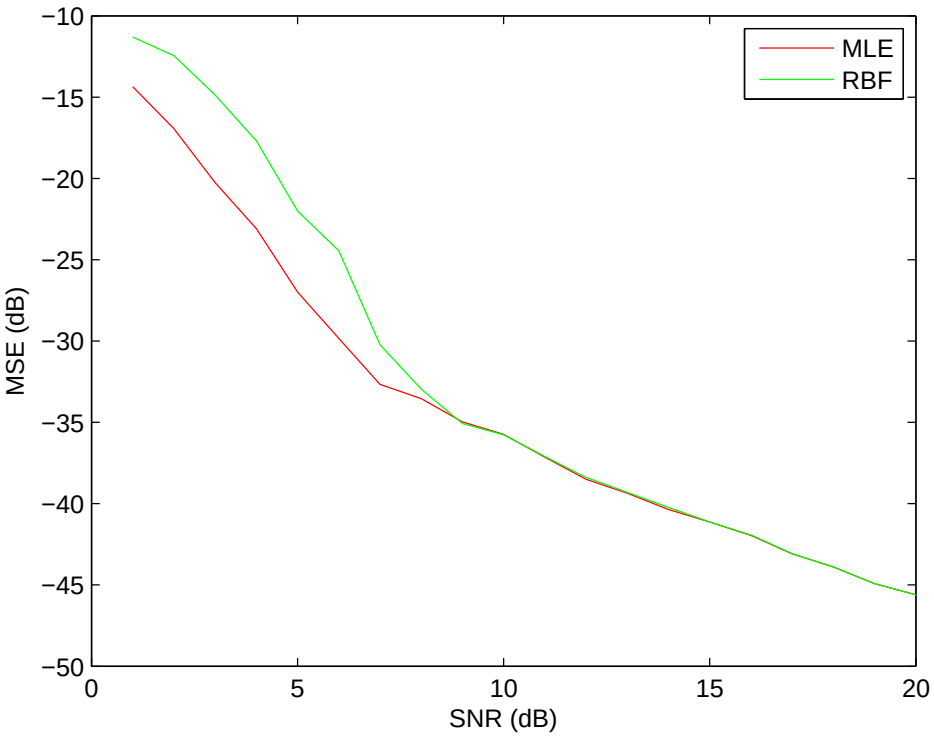


Figure 3: Performance analysis on v_x parameter in 1-D motion example

when compared to the true GLRT. However, the RBF approach can reduce the computation by orders of magnitude.

List of References

[1] N. Vankayalapati and S. Kay, "Asymptotically optimal detection of low probability of intercept signals using distributed sensors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 1, pp. 737–748, 2012.

[2] S. M. Kay, *Fundamentals of statistical signal processing, volume II: Detection theory*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

[3] M. J. Wichura, *The coordinate-free approach to linear models*. Cambridge University Press, 2006, vol. 19.

[4] S. Kay, "A computationally efficient nonlinear least squares method using random basis functions," *IEEE Signal Processing Letters*, vol. 7, no. 20, pp. 721–724, 2013.

CHAPTER 2

Detection of Moving Dim Target in IR Video

by

Fuat Cogun and Steven Kay

Dept. of Electrical, Computer and Biomedical Engineering

University of Rhode Island, Kingston, RI, USA

submitted to IEEE Transactions on Aerospace and Electronic Systems.

Abstract

In this paper, the detection/localization problem of a moving single dim target in infra-red (IR) video is considered. First, the generalized likelihood ratio test (GLRT) is employed to obtain the straightforward GLRT detector and its performance is found by using simulated data based on modelling of real IR data. Second, the random basis functions (RBF) approach is exploited to reduce the computational complexity associated with the straightforward GLRT detector for higher signal-to-noise ratio (SNR) scenarios. Monte-Carlo simulations are run for a point target to obtain the performance of the RBF-based detector.

2.1 Introduction

Detection of dim moving targets relative to the background clutter still remains an active research area. The basic formulation of the detection problem is as follows. A moving dim target is immersed in background clutter in a scene and the scene is imaged via a remote sensor whose field-of-view is fixed with respect to the background. The imaged scene contains additive noise due to the physical nature of the problem and sensor imperfections. The standard approach for target detection processes the imaged scene to separate the target from the clutter and noise. There are many approaches proposed in the literature addressing this detection problem.

The simplest approaches to the problem are based on detect-before-track (DBT) approach [1], [2], [3]. The target is detected at each temporal frame, then the detections are used to estimate target trajectory. These methods require the target intensity in the imaged scene be adequate for detection. Although it is easy to implement these detectors, their performance at low signal-to-noise (SNR) scenarios is poor [4] and much of the temporal information is lost since a threshold is applied at each frame. Thus, they cannot be used in many real-world target

detection problems.

In the track-before-detect (TBD) approach, the observed data are processed over frames before any target detection is declared. In addition to better performance in low SNR scenarios, the TBD approach provides the estimated target trajectory if a target is detected at the scene. A three-dimensional (3-D) matched filtering approach [5] [6] based on the maximization of the spatio-temporal SNR has been proposed by Reed et al. The approach works well if the target signature matches the 3-D filter coefficients. A frequency domain approach is given in [7] where a bank of directional filters performs moving target signature matched filtering for all possible directions in the Fourier domain. Mohanty [8] proposed an adaptive algorithm based on maximum-likelihood (ML) ratio which requires all possible trajectories be introduced to the detector prior to detection. Note that all of these methods have high computational complexity since the number of possible straight trajectories would be huge in practice and exhaustive matched filtering is impractical.

A computationally efficient well-known suboptimal approach makes use of the dynamic programming [9] and the detector performance is given in [10]. However, it is hard to implement this detector and the performance is reduced for dim targets under severe clutter [11]. Reed et al. proposed a moving target indicator (MTI) algorithm [12] to obtain computational efficiency for the same problem in [5] [6] by using 2-D transforms only, resulting in a degraded performance in the presence of velocity mismatch.

The main concern of this work is the detection of a single dim target, moving at a constant velocity in IR image sequences. The background clutter plus observation noise obscures the target; thus, it is not possible to use a DBT approach. It is assumed that the intensity, initial position and velocity of the target are all

unknown. For this problem, we propose a detector based on the generalized-likelihood-ratio-test (GLRT) which is known to be asymptotically uniformly most powerful (UMP) test among all tests that are invariant [13]. In order to reduce the computational complexity of the straightforward GLRT detector, the random basis functions (RBF) approach [14] is also used. Note that the RBF approach is an efficient non-linear least squares method [15] and is easy to implement.

The outline of this paper is as follows. In the next section, the signal model and modeling assumptions are given. Section 2.3 discusses the GLRT approach taken to derive the straightforward detector and explains the detection procedure in detail. The performance of the derived straightforward GLRT detector is given in Section 2.4. The RBF approach which is applied to reduce the computational complexity associated with the straightforward GLRT detector is introduced in Section 2.5. Section 2.6 presents the computational counts for both straightforward grid-search and RBF approaches to show the computational gain obtained by the use of the RBF approach. The performance of the RBF-based detector is given in Section 2.7. Finally, the paper concludes with some discussions and conclusions.

2.2 Modeling assumptions

In this section, we introduce the signal model used and the reasoning behind it. The hypothesis test for the IR signal is

$$\begin{aligned} \mathcal{H}_0 &: x[m,n,k] = B[m,n] + w[m,n,k] \\ \mathcal{H}_1 &: x[m,n,k] = As[m,n,k] + B[m,n] + w[m,n,k] \end{aligned} \tag{21}$$

where $m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$ (image pixels) and $k = 1, 2, \ldots, K$ (frame number). The target intensity A and background $B[m,n]$ are constant with frame and unknown. Additive noise $w[m,n,k]$ is assumed to be white Gaussian noise with variance σ^2 and independent from pixel-to-pixel and frame-to-frame. In

this study, the signal-to-noise ratio (SNR) is defined as:

$$\text{SNR} = 10 \log_{10} \frac{A^2}{\sigma^2} \tag{22}$$

Note that it is possible to explicitly write $s[m, n, k]$ as:

$$s[m, n, k] = g[m - m_0 - v_x k, n - n_0 - v_y k] \tag{23}$$

where (m_0, n_0) is the starting position of the target and (v_x, v_y) is its velocity, both of which are unknown. The velocity of the target is assumed to be constant throughout the data collection time interval. The function g is the target function which characterizes the size of the target. If the target is point-like, the signal $s[m, n, k]$ is represented as one pixel; thus, g is the Kronecker delta function. Note that the assumed IR signal hypothesis problem in this paper is a composite hypothesis testing problem.

2.3 Detection approach

Asymptotically, the GLRT is the UMP test among all tests that are invariant. The GLRT replaces the unknown parameters in the Neyman-Pearson optimum test statistic with their maximum likelihood estimates (MLEs). The GLRT detector decides $\mathcal{H}_1$ if [16]

$$L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\boldsymbol{\theta}}_0, \mathcal{H}_0)} > \gamma \tag{24}$$

where p is the probability distribution function (PDF) , $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}_1$ assuming $\mathcal{H}_1$ is true and $\hat{\boldsymbol{\theta}}_0$ is the MLE of $\boldsymbol{\theta}_0$ assuming $\mathcal{H}_0$ is true. In our problem, $\boldsymbol{\theta}_1$, the unknown parameter vector under $\mathcal{H}_1$, consists of target intensity, target position and velocity.

Although there is no optimality associated with the GLRT, in practice it appears to work quite well. In this paper, the GLRT with the unknown parameters

(m_0, n_0) and (v_x, v_y) is

$$\max_{(m_0,n_0),(v_x,v_y)} 2\ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} \tag{25}$$

where $\mathbf{X}$ is Gaussian, $\hat{\mathbf{B}}_0$ is the MLE of the background under $\mathcal{H}_0$, $\hat{A}$ is the MLE of the target intensity under $\mathcal{H}_1$, and $\hat{\mathbf{B}}_1$ is the MLE of the background under $\mathcal{H}_1$. Note that the dimension of $\mathbf{X}$ is $M \times N \times K$, the dimension of $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{B}}_1$ is $M \times N$.

In order to utilize the multivariate Gaussian PDF for the image frame data, it is customary [8] to roll out the columns to form a column vector. A simple example is to roll out the columns of

$$\mathbf{X} = \begin{bmatrix} x[0,0] & x[0,1] \\ x[1,0] & x[1,1] \\ x[2,0] & x[2,1] \end{bmatrix} \quad 3 \times 2 \text{ image}$$

to yield

$$\mathbf{x} = \begin{bmatrix} x[0,0] & x[1,0] & x[2,0] & x[0,1] & x[1,1] & x[2,1] \end{bmatrix}^T.$$

The problem here is that neighboring pixels become widely dispersed. It becomes difficult to realize the detector and also adds to coding complexity. Additionally, the intuition behind the final detector can become lost. Note that the problem becomes even worse when the data is 3-dimensional as in $x[m, n, k]$. Hence, it is better to retain the matrix $\mathbf{X}$ structure and more convenient to use the matrix normal PDF. Hence, we use instead

$$\mathbf{X}_k = \begin{bmatrix} x[0,0,k] & x[0,1,k] \\ x[1,0,k] & x[1,1,k] \\ x[2,0,k] & x[2,1,k] \end{bmatrix} \quad 3 \times 2 \text{ image for } k\text{th frame.}$$

The PDF of $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K\}$ under $\mathcal{H}_1$ becomes

$$p(\mathbf{X}; A, \mathbf{B}, \mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{MNK/2}} \exp\left[-\frac{1}{2\sigma^2} J(A, \mathbf{B})\right]$$

where

$$J(A, \mathbf{B}) = \sum_{k=1}^{K} \mathrm{Tr}[(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})^T]$$

Also a more compact expression for J is

$$J = \sum_{k=1}^{K} \langle \mathbf{X}_k - A\mathbf{S}_k - \mathbf{B}, \mathbf{X}_k - A\mathbf{S}_k - \mathbf{B} \rangle \tag{26}$$

where $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}\mathbf{Y}^T)$. The latter is a 2-dimensional inner product on an inner product space of $M \times N$ images. The inner product formulation can easily be extended to 3-dimensional "data cubes" to accommodate frames as well.

The required MLE under $\mathcal{H}_0$ is (see Appendix)

$$\hat{\mathbf{B}}_0 = \frac{1}{K} \sum_{k=1}^{K} \mathbf{X}_k \tag{27}$$

and MLE of $\mathbf{B}$ under $\mathcal{H}_1$ is

$$\hat{\mathbf{B}}_1 = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{X}_k - \hat{A}\mathbf{S}_k) = \bar{\mathbf{X}} - \hat{A}\bar{\mathbf{S}} \tag{28}$$

where $\bar{\mathbf{X}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{X}_k$ and $\bar{\mathbf{S}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{S}_k$. The MLE of A under $\mathcal{H}_1$ is

$$\hat{A} = \frac{\sum_{k=1}^{K} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=1}^{K} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \tag{29}$$

Finally, the GLRT is found as (see Appendix 2A)

$$2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} = \frac{\sum_{k=1}^{K} ||\mathbf{S}_k - \bar{\mathbf{S}}||^2}{\sigma^2} \left[\frac{\sum_{k=1}^{K} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=1}^{K} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \right]^2 \tag{30}$$

For unknown starting position and velocity we must maximize this over (m_0, n_0) and (v_x, v_y) ($\mathbf{S}_k$'s depend on this).

The power of the GLRT is that the MLEs not only provide information that summarizes the presence of a target, i.e., used for detection, but also provide information that can be used for estimation of target trajectory and can be used for classification.

Here are the steps for finding the MLEs and GLRT, providing a verbal explanation of the procedure:

1. Estimate background (constant with frame number) by using the sample mean of the image frames

2. Subtract out the background estimate from all image frames

3. Subtract out the sample mean of signal images (since we can't distinguish between the background and a zero velocity target)

4. Correlate the residual image against expected signal (will change with the frame since target is assumed moving)

5. Add some normalization to allow the PDF of the GLRT to be determined and hence the thresholding required to maintain constant false alarm.

Note that we have assumed an inner product space of images (2-dimensional) with independent noise from frame to frame. But it can be extended to a 3-dimensional inner product space to allow representation of an entire video stream, where $\mathbf{X}$ and $\mathbf{Y}$ becomes $M \times N \times K$ data cubes, and the inner product becomes

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{k=1}^{K} x[m,n,k] y[m,n,k]. \tag{31}$$

This approach employs the coordinate-free approach to linear models. [17]

2.4 Performance of the GLRT Detector

The performance of the GLRT detector is obtained by applying the detector to simulated data based on modelling of real IR data. Note that access to real IR data was not possible due to its classified nature; therefore, data based on modelling of real IR data provided by Michigan Tech Research Institute [18] is used. The pixels of the scene including the target is shown in Fig. 4. There is no added noise to the scene in Fig. 4 to maximize the visibility of the target at scene. In Fig. 5 the detailed target region for a fixed frame number can be seen. Note

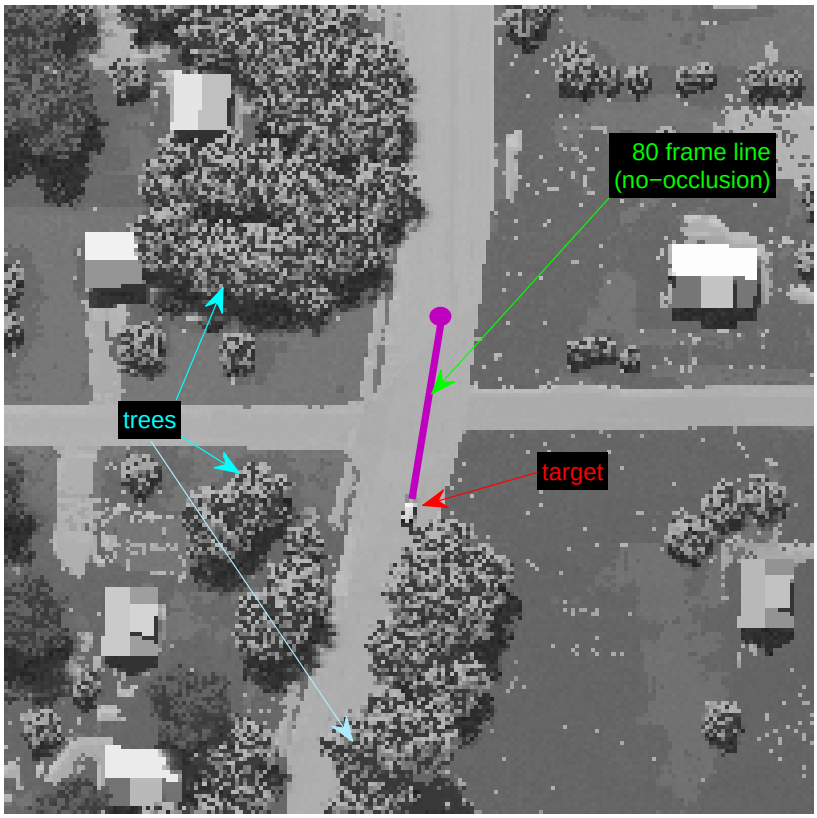


Figure 4: Pixels of the scene - no occlusion occurs during target movement

that the target is scintillating and consists of multiple (6 to 8) pixels of varying intensity which are the usual cases in real-world IR target detection problem. To accommodate the fluctuation in the simulations, signal amplitude is chosen by averaging the amplitudes of target pixels throughout its movement, then WGN is added to satisfy the desired SNR defined by (22).

As explained earlier in Section 2.1, as the SNR becomes lower, DBT methods or standard image processing methods cannot detect the presence of a target; thus, cannot be used. In Fig. 6, the scene is shown for an SNR equal to 10dB. As the target is dim, even at SNR = 10dB, it is not possible to pick the target. It is not

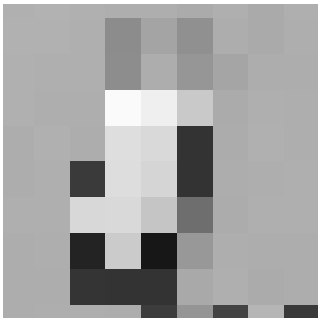


Figure 5: The target

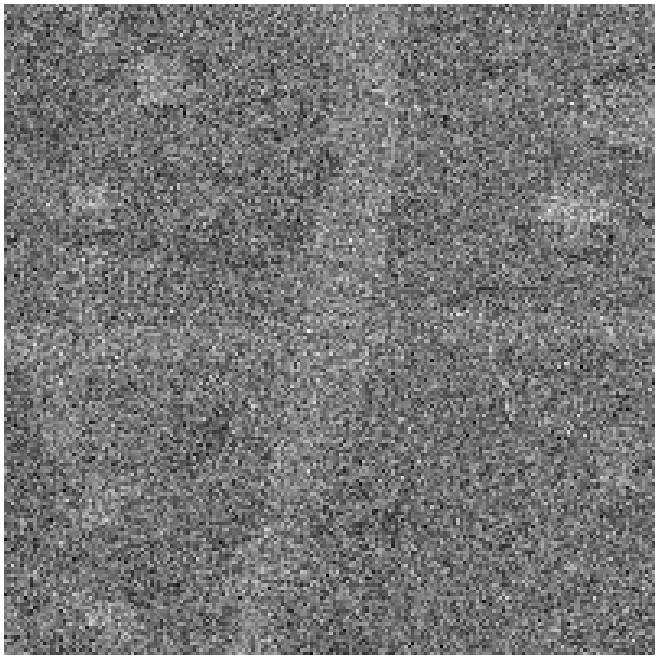


Figure 6: The scene at 10dB - Not possible to pick target

uncommon to deal with SNRs as low as 3dB in real-world IR scenarios; therefore, the need to use successive frames (TBD approach) is apparent.

As our performance metric for the GLRT detector, receiver operating characteristic (ROC) curves are used. For a given SNR value, 1000 realizations of the target-at-scene case ($\mathcal{H}_1$) were used to estimate the P_D and 1000 realizations of

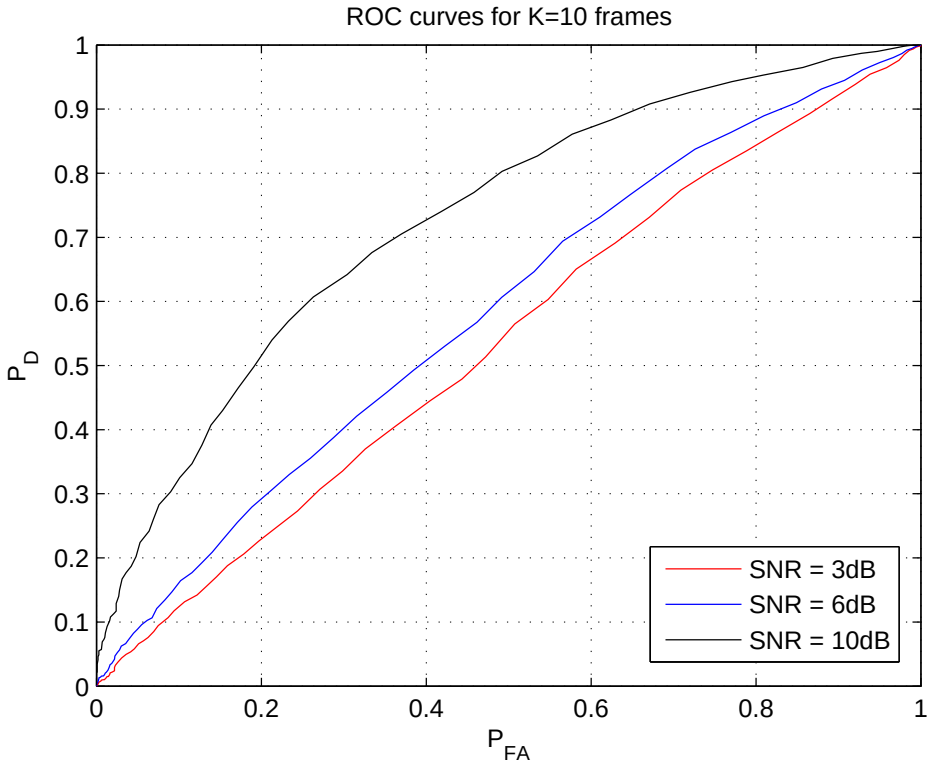


Figure 7: ROC curves for 10 frames

the no-target-at-scene case ($\mathcal{H}_0$) were used to estimate the P_{FA}; therefore, a total of 2000 realizations for each (P_D, P_{FA}) point on the ROC curve were generated. Note that 100 different thresholds are applied, thus, 100 points on the ROC curves are obtained to plot smooth curves.

The ROC curves in this section are found at SNRs 3dB, 6dB and 10dB for frame lengths $K = 10$, 40 and 80 and are plotted in Fig. 7, Fig. 8, Fig. 9, respectively. It is seen that as K increases, the GLRT detector performs better as expected. For an SNR level as low as 6dB, it is possible to obtain $P_D = 0.9$ for a given $P_{FA} = 10^{-2}$ for $K = 80$ frames which corresponds to about 4 seconds of processing. It is deemed reasonable amount of time for constant velocity target. In order to accomplish the same performance at lower SNR values, there is need to have more frames available for detection.

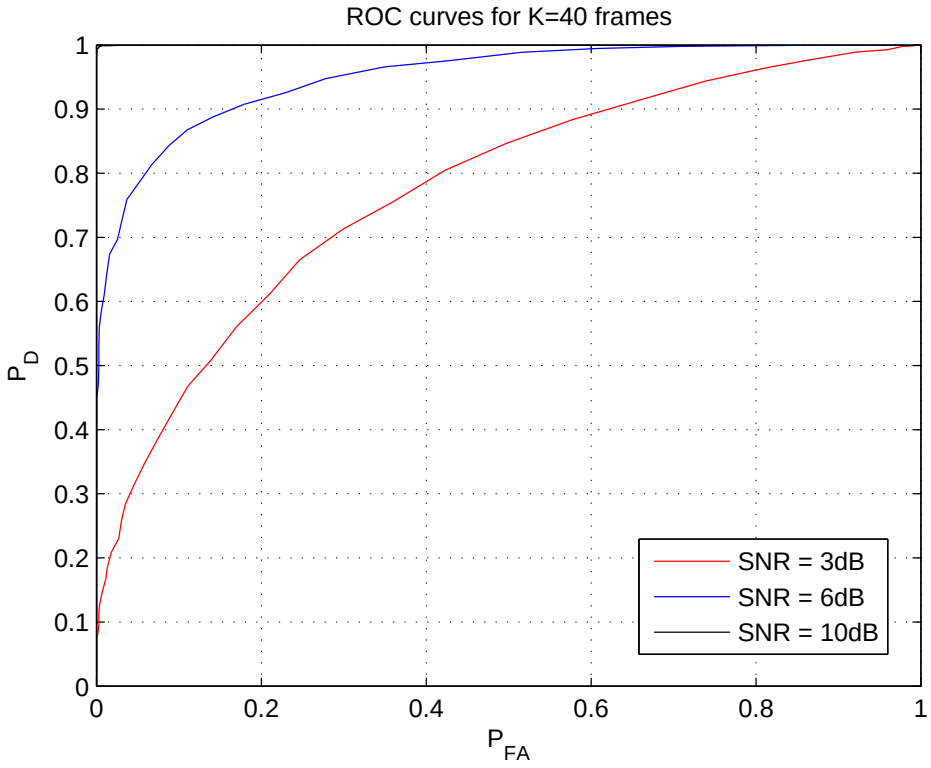


Figure 8: ROC curves for 40 frames

Previous results are all obtained by using a part of the scene such that there is no occlusion occurring throughout the target movement. Therefore, the best possible detector performances were presented. To find the detector performance for an occluded scene, the GLRT detector is applied to a part of the scene including occlusion (trees obscure the target, see Fig. 10) and the simulations are rerun. The performance loss caused by occlusion for $K = 40$ frames can be analyzed in Fig. 11. Clearly, to obtain the same detector performance as in the no-occlusion case at a fixed SNR, there is need to include more frames.

Note that the GLRT detector in this paper requires a grid-search over all possible combinations of position and velocity. Thus, the computational complexity of it is high and it becomes formidable as the search space increases. Fortunately, it is possible to decrease the computational complexity of the straightforward GLRT

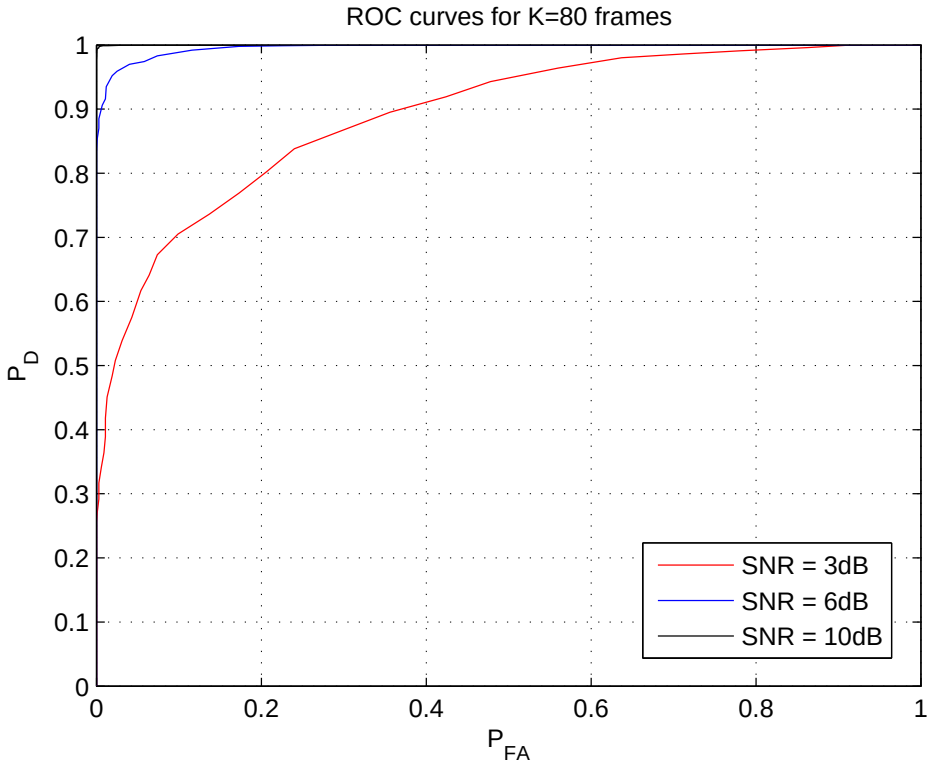


Figure 9: ROC curves for 80 frames

by orders of magnitude by using random basis functions (RBF) approach. Although the performance of RBF approach is suboptimal, at higher SNR values, it performs as well as grid-search [19]. This is illustrated further in the next section.

2.5 RBF Approach to Reduce Computation

The random basis function (RBF) approach converts difficult non-linear least squares problems to a succession of linear least squares problems. It is a bisection approach and thus, there are no convergence issues. The performance is suboptimal since the approach requires higher SNR, but it can reduce the computation by orders of magnitude.

Consider the problem of fitting a signal with unknown amplitude and non-linear parameters to data. The non-linear least squares method tries to minimize

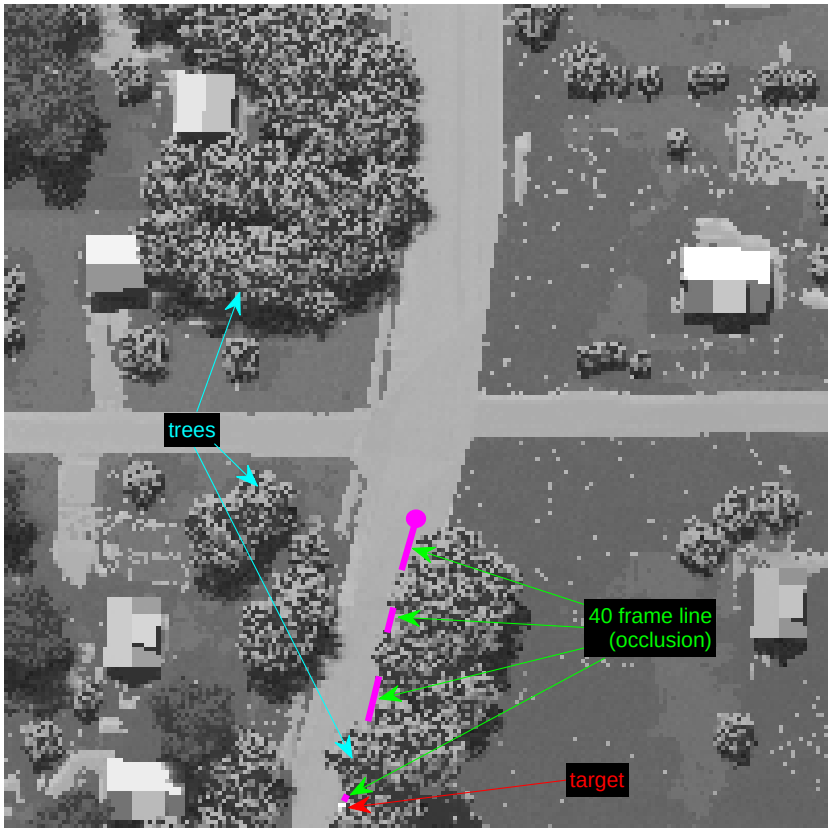


Figure 10: Pixels of the scene - occlusion occurs during target movement

the error function

$$J(A, \boldsymbol{\theta}) = \sum_{n=0}^{N-1} (x[n] - As[n; \boldsymbol{\theta}])^2 \tag{32}$$

over amplitude A and non-linear signal parameters $\boldsymbol{\theta}$. If we consider frequency estimation as an example, the signal is $s[n] = A\cos(2\pi\theta n)$, where $\theta = f_0$ is the non-linear signal parameter. Estimating f_0 usually requires a brute force search such as an FFT. To have a precise frequency estimate, a very large FFT size is required.

The RBF approach proceeds as follows. Consider $s[n; \boldsymbol{\theta}]$ as a basis function

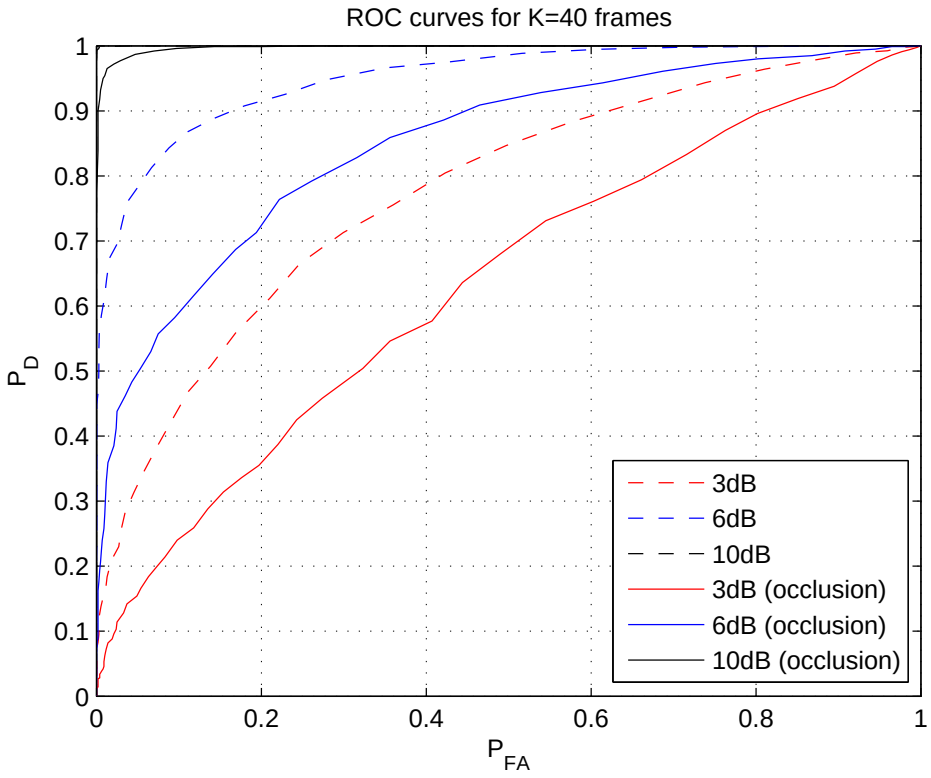


Figure 11: Performance loss caused by occlusion for $K = 40$ frames

with a random parameter $\boldsymbol{\theta}$. The *expected least squares* criterion is defined as

$$J(A) = E_{\boldsymbol{\theta}}\left[\sum_{n=0}^{N-1}(x[n] - As[n;\boldsymbol{\theta}])^2\right] \tag{33}$$

The main idea of the RBF approach is to split $\boldsymbol{\theta}$ into two disjoint regions, then apply the expected least squares criterion and find the minimizing A's corresponding to the intervals. Then we choose the band with the larger A^2 and repeat subdividing the tentative band until it provides a small enough interval. If we apply the approach to the frequency estimation example, we need to:

1. Split the possible frequency band into two disjoint intervals

2. Assign a uniform PDF to the random frequency for each interval

3. Minimize over A_1 and A_2 the expected least squares

$$J(A_1, A_2) = E_{\theta_1,\theta_2}\left[\sum_{n=0}^{N-1}\left(x[n] - A_1\cos(2\pi\theta_1 n) - A_2\cos(2\pi\theta_2 n)\right)^2\right]$$

4. Choose the band with the larger A^2 and repeat subdividing the tentative band until chosen band is small enough to yield an accurate estimate.

5. Choose the estimate of θ as midpoint of the last subdivided interval.

The RBF approach can be easily applied to image sequences. The data is given as

$$x[m, n, k] = As[m - m_0 - v_x k, n - n_0 - v_y k] + w[m, n, k] \tag{34}$$

Here the unknown non-linear parameter is $\boldsymbol{\theta} = [m_0\, n_0\, v_x\, v_y]^T$. As there are more than one non-linear parameter, we cycle between the parameters at each iteration. That is, the RBF approach is applied to one parameter at a time. After choosing the interval for that parameter, we continue to operate on the other parameter until all the parameters are visited on $\boldsymbol{\theta}$. This ends one iteration and we iterate until all non-linear parameters converge. An illustration of cycling is given in Fig. 12 for the 1-D motion example defined as

$$x[m, k] = As[m - m_0 - v_x k] + w[m, k] \tag{35}$$

where we let $\boldsymbol{\theta} = [m_0\, v_x]^T$. As an example, consider the case where the unknown parameter v_x (target velocity) is assumed to be in the interval $(-1, 1)$, i.e., $v_x \in (-1, 1) \subset \mathcal{R}$, and the unknown parameter m_0 (target initial position) is assumed to be in the interval $[-60, 60]$, i.e., $m_0 \in [-60, 60] \subset \mathcal{Z}$. For clarity, the first 5 iterations of the RBF approach are shown in Fig. 13. The estimates converge to the true target velocity and initial target position ($v_x = -0.3, m_0 = 35$) in a few iterations.

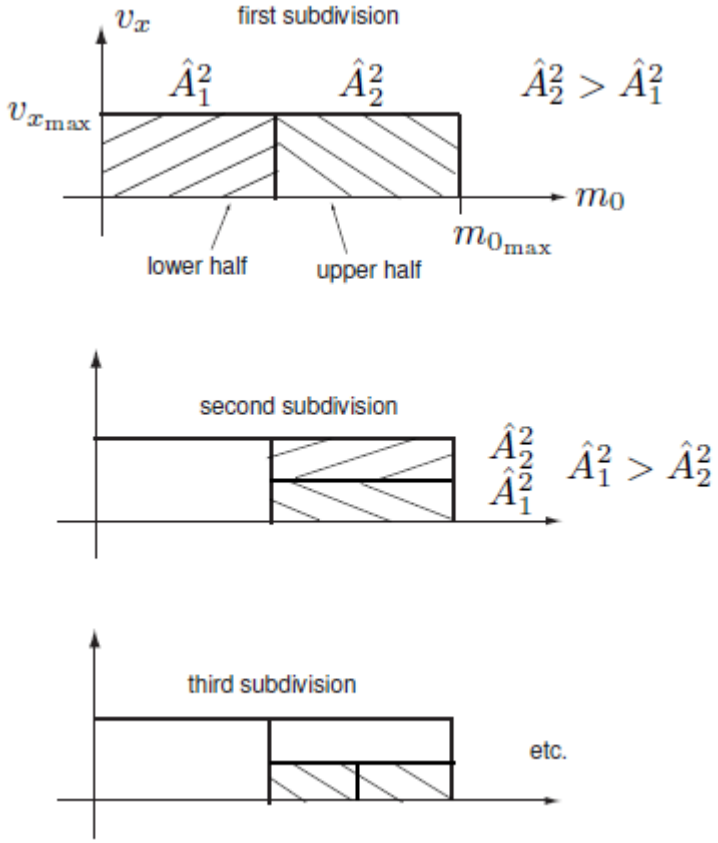


Figure 12: RBF approach - Cycling between m_0 and v_x non-linear parameters

2.6 Computational Counts

In this section, the computational counts for both the RBF approach and the straightforward grid-search are given to show the computational gain obtained by the use of the RBF approach. It is seen that the total number of multiplications required for both approaches are roughly the same and the multiplicative complexity is found as $O(\mathcal{MNV}^2K)$, where $\mathcal{M}$ and $\mathcal{N}$ denote the number of grid points for vertical and horizontal components of the initial position, respectively, $\mathcal{V}$ denotes the total number of grid points for velocity in each direction and K denotes the number of frames. On the other hand, it is found that the total number of addition operations involved in the grid-search increases with $O(\mathcal{MNV}^2K^2)$ whereas in RBF approach it increases with $O(\mathcal{MNV}^2K)$. Therefore, we get a speed-up of

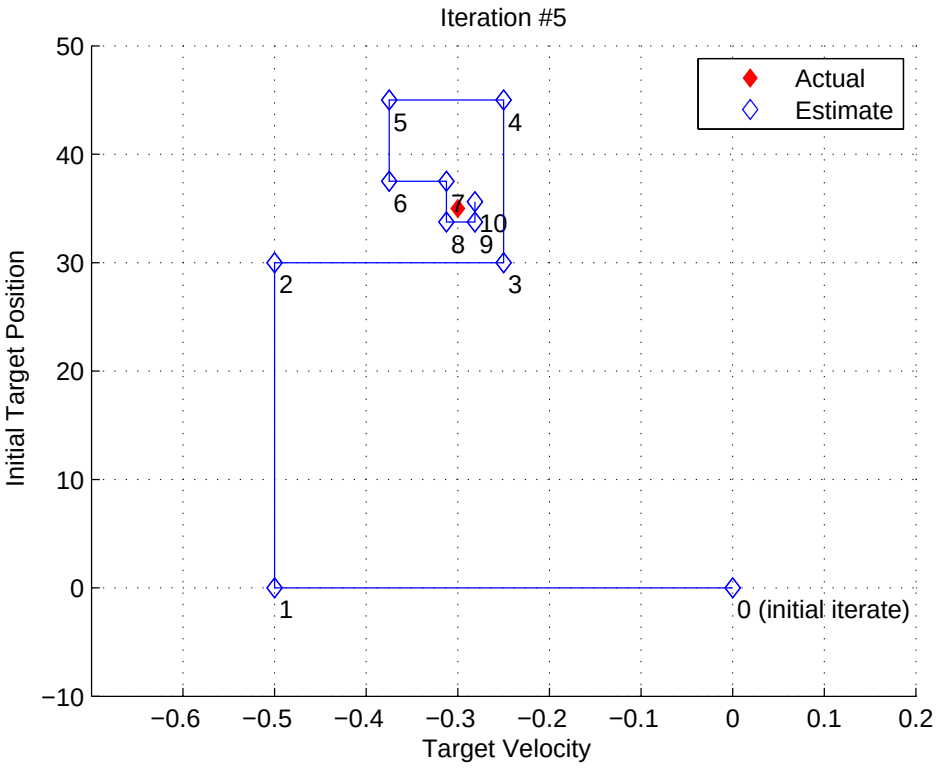


Figure 13: RBF approach - First 5 iterations (10 cycles)

about K for the RBF over the grid-search approach. The significance of this computational gain becomes much more evident when K increases. As an example, consider the case where $\mathcal{M} = \mathcal{N} = 25$ and $\mathcal{V} = 21$. The RBF approach finds the estimates of initial position and velocity 9 times faster than the straightforward grid-search for $K = 10$ as seen in Table 1. As K increases, the CPU execution time of the grid-search increases quadratically and it is not practical to use a grid-search for $K > 20$. In contrast, the CPU execution time of the RBF approach increases linearly.

2.7 Performance of the RBF-based Detector

The performance of the RBF-based detector is obtained via Monte-Carlo simulations in this paper. We used ROC curves as our performance metric as we did for the GLRT detector. In order to plot smooth ROC curves, 2000 realizations

Table 1: CPU execution times for Grid-search and RBF approach

K	**Grid-search (sec.)**	**RBF approach (sec.)**
5	8.37	1.81
10	32.6	3.62
20	271	7.98
40	1784	16.8
80	8961	34.5
160	*	69.7

**Unable to implement.*

for each (P_D, P_{FA}) point on the ROC curve were generated in particular. For a given threshold, 1000 realizations of the target-at-scene case were simulated to estimate the P_D and 1000 realizations of the no-target-at-scene case were simulated to estimate the P_{FA}, so that we could obtain a (P_D, P_{FA}) point on the ROC curve. The realizations were done for a total of 120 different thresholds, therefore, 120 points on the ROC curve were used to plot the curves seen in Fig. 14. To keep the run-time of the simulation at a reasonable level, a point target is used and the grid points were selected as $\mathcal{M} = 8$, $\mathcal{N} = 8$, $\mathcal{V} = 21$ and the total number of frames was selected as $K = 13$. Note that a privilege was given to $\mathcal{V}$ by assigning a higher number of grid points, since the MSE in the estimated target trajectory is much more sensitive to an error made in the velocity (v_x, v_y) estimates. In Fig. 14, it is seen that the performance of the detector increases when the SNR increases as expected. In order to compare the performance of the RBF-based detector and the GLRT detector, Fig. 15 is plotted. Although the detection performance of the GLRT detector is better for higher P_{FA} values, it is typical to focus on lower P_{FA} values in the target detection problem that is considered. It is seen that for this region, the detector performance of the RBF-based detector is comparable to the GLRT detector with only an SNR difference of 2dB.

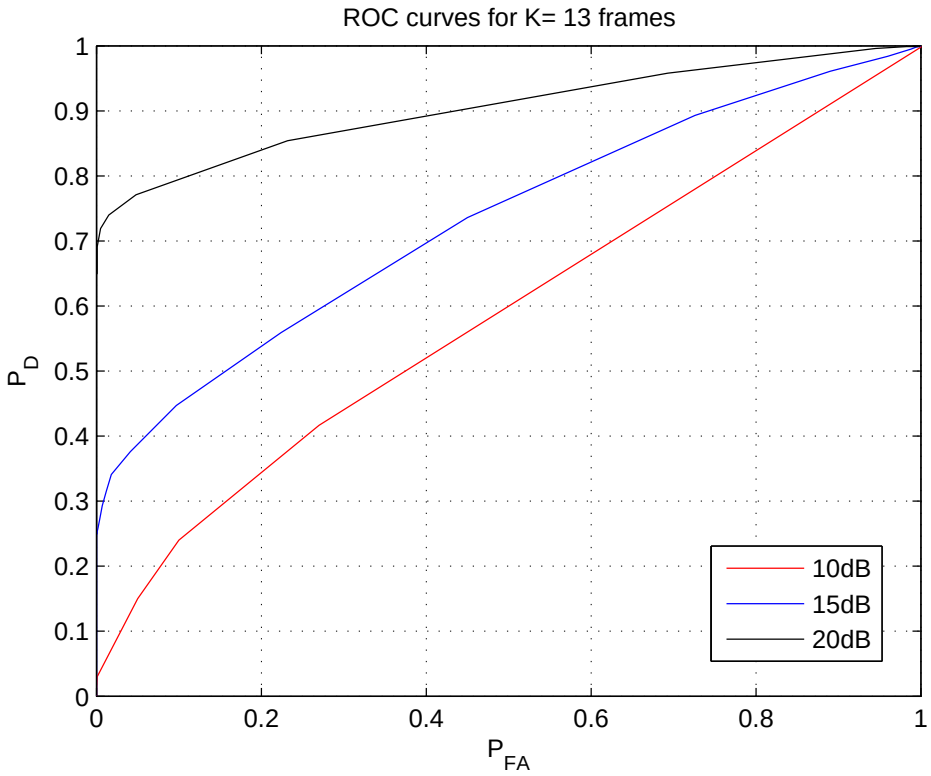


Figure 14: The performance of the RBF-based detector for $K = 13$ frames

An illustration of the detector for a given SNR (10dB) is shown in Figs. 16, 17 and 18. In Fig. 16, the simulated target signal is shown at different frame numbers. Additive WGN is added to every pixel at each frame satisfying SNR = 10dB. As seen in Fig. 17, the target is not visible to the naked eye. Running the detector on the observed data, the target is successfully detected at each frame as shown by a red circle in Fig. 18.

2.8 Discussion and conclusions

In this paper, the derivation of the GLRT detector is done assuming that the target velocity is constant throughout the target's movement. An important consideration might be the case when the velocity of the target changes over a sequence of frames. It is clear that a direct application of the proposed detectors would not perform well in this scenario since the assumed signal model is incorrect.

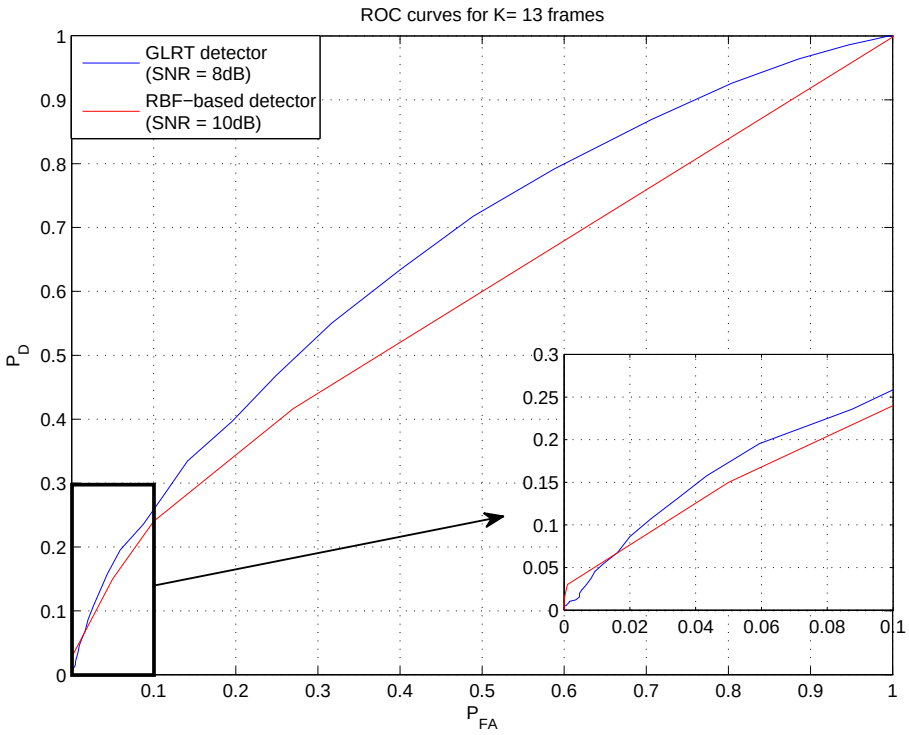
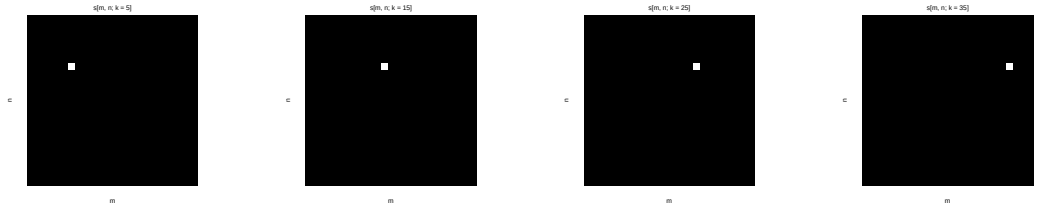


Figure 15: The performance comparison of the GLRT and RBF-based detectors for $K = 13$ frames

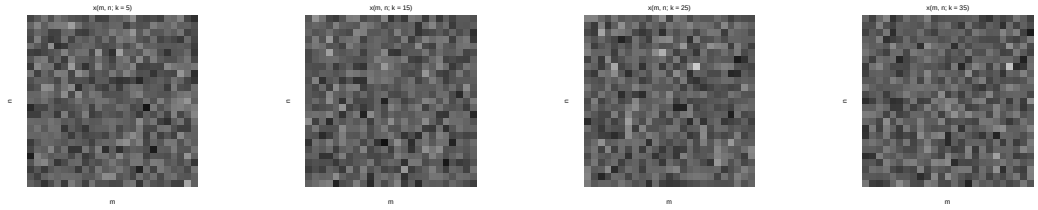


(a) k = 5 (b) k = 10 (c) k = 15 (d) k = 20

Figure 16: Target Signal $s[m, n, k]$

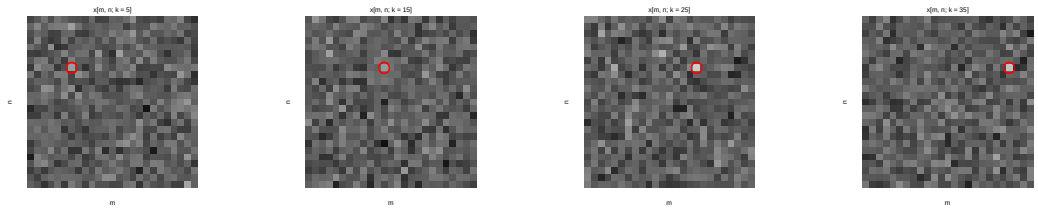
A workaround approach is to apply the detector over few frames for which target velocity is approximately constant first, then, use the results from the first step to reduce the search area for initial position and velocity for the next set of frames. The selection of the number of frames is an issue by itself and is subject to further research.

The proposed detectors are designed for the detection of a single dim target.



(a) k = 5 (b) k = 10 (c) k = 15 (d) k = 20

Figure 17: Observed data $x[m, n, k]$



(a) k = 5 (b) k = 10 (c) k = 15 (d) k = 20

Figure 18: Target Detected

But they can easily be extended to detect multiple targets moving at constant velocity. After the detection of the first target, it is blackened out from the original sequence of frames and then the detector is applied to find the second target. An illustration of a two target detection problem by using the RBF-based detector and the trajectories of the targets are shown in Fig. 19. Note that the same procedure can be carried on until all targets are detected in the multiple target detection problem. By this approach, all of the targets are detected one by one. If it is desired to detect all of the targets at the same time, the signal model should be updated to include multiple targets.

It should be noticed that the straightforward GLRT detector requires a grid-search over all possible combinations of position and velocity. Thus, the computational complexity of it is very high and it becomes formidable as the size of the search region increases. In this work, no prior knowledge of the observed scene was assumed. But it may be possible to reduce down the search region if there

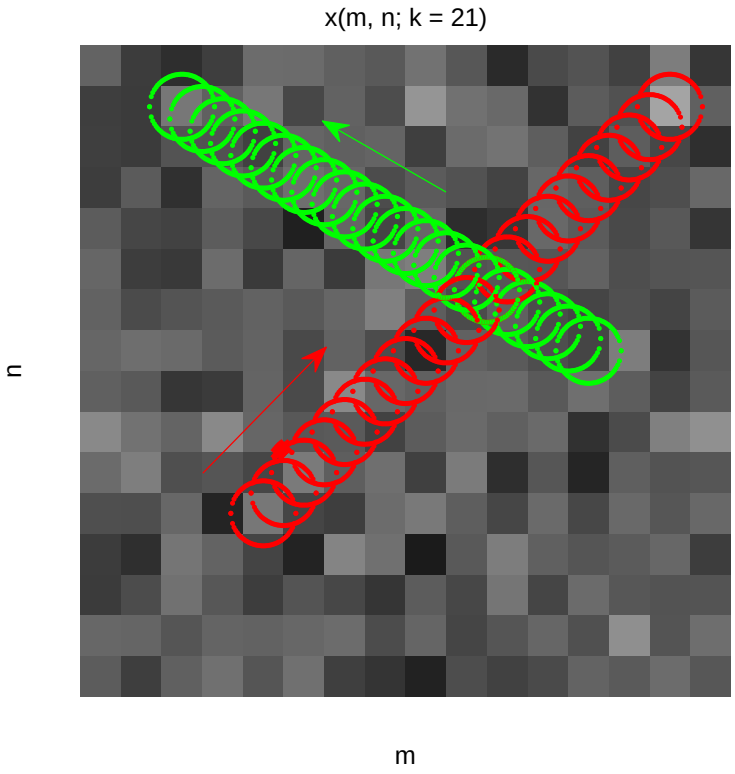


Figure 19: Two target detection/tracking problem at SNR=10dB

is some known information about the observed scene (roads, geography, etc.). If not, as discussed earlier, the RBF-based detector can be applied to reduce the computational complexity. However, in this case, the overall performance of the detector would be poorer if the SNR is low. Therefore, it is advised to use the RBF-based detector when the SNR is moderate to high.

One final note is on the performance of the GLRT detector. In this paper, ROC curves were obtained via Monte-Carlo simulations. However, it may also be possible to find the performance of the detector analytically. Note that this requires the use of order statistics because of the maximization operation involved in the test statistic. Once P_{FA} is found as a function of the threshold, we can realize a constant false alarm rate (CFAR) detector and obtain P_D vs. SNR curves. The analytical performance of the GLRT detector is considered as future work.

Appendix 2A. Derivation of the GLRT Detector

In this section, we will derive the GLRT detector used in this paper. Remember that general form of GLRT detector was given as

$$\max_{(m_0,n_0),(v_x,v_y)} 2\ln\frac{p(\mathbf{X};\hat{A},\hat{\mathbf{B}}_1,\mathcal{H}_1)}{p(\mathbf{X};\hat{\mathbf{B}}_0,\mathcal{H}_0)} \tag{36}$$

Now, lets find the PDFs as functions of $A,\mathbf{B}$

$$p(\mathbf{X};A,\mathbf{B},\mathcal{H}_1)=\prod_{k=1}^{K}p(\mathbf{X}_k;A,\mathbf{B},\mathcal{H}_1) \tag{37}$$

where

$$p(\mathbf{X}_k;A,\mathbf{B},\mathcal{H}_1)=\frac{1}{(2\pi\sigma^2)^{MN/2}}\exp\left[-\frac{1}{2\sigma^2}\mathrm{Tr}[(\mathbf{X}_k-A\mathbf{S}_k-\mathbf{B})(\mathbf{X}_k-A\mathbf{S}_k-\mathbf{B})^T]\right]$$

Thus,

$$\begin{aligned}p(\mathbf{X};A,\mathbf{B},\mathcal{H}_1)=\quad & \frac{1}{(2\pi\sigma^2)^{MNK/2}}\\ & \cdot\exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\mathrm{Tr}[(\mathbf{X}_k-A\mathbf{S}_k-\mathbf{B})(\mathbf{X}_k-A\mathbf{S}_k-\mathbf{B})^T]\right]\end{aligned}$$

Similarly,

$$p(\mathbf{X};\mathbf{B},\mathcal{H}_0)=\prod_{k=1}^{K}p(\mathbf{X}_k;\mathbf{B},\mathcal{H}_0) \tag{38}$$

where

$$p(\mathbf{X}_k;A,\mathbf{B},\mathcal{H}_0)=\frac{1}{(2\pi\sigma^2)^{MN/2}}\exp\left[-\frac{1}{2\sigma^2}\mathrm{Tr}[(\mathbf{X}_k-\mathbf{B})(\mathbf{X}_k-\mathbf{B})^T]\right]$$

Thus,

$$p(\mathbf{X};\mathbf{B},\mathcal{H}_0)=\frac{1}{(2\pi\sigma^2)^{MNK/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\mathrm{Tr}[(\mathbf{X}_k-\mathbf{B})(\mathbf{X}_k-\mathbf{B})^T]\right]$$

To determine the MLEs of the unknown parameters $A,\mathbf{B}$, we need to find the values maximizing $p(\mathbf{X};\mathbf{B},\mathcal{H}_0)$ and $p(\mathbf{X};A,\mathbf{B},\mathcal{H}_1)$ separately.

Determining $\hat{\mathbf{B}}_0$ and $p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)$

To find $\hat{\mathbf{B}}_0$, need to find $\mathbf{B}$ maximizing $p(\mathbf{X}; \mathbf{B}, \mathcal{H}_0)$ or equivalently minimizing

$$J_0(\mathbf{B}) = \sum_{k=0}^{K-1} \mathrm{Tr}[(\mathbf{X}_k - \mathbf{B})(\mathbf{X}_k - \mathbf{B})^T] \tag{39}$$

Now let

$$\hat{\mathbf{B}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{X}_k = \bar{\mathbf{X}} \tag{40}$$

Then we can write

$$\begin{aligned} J_0(\mathbf{B}) &= \sum_{k=1}^{K} \mathrm{Tr}\left[\left((\mathbf{X}_k - \bar{\mathbf{X}}) + (\bar{\mathbf{X}} - \mathbf{B})\right)\left((\mathbf{X}_k - \bar{\mathbf{X}}) + (\bar{\mathbf{X}} - \mathbf{B})\right)^T\right] \\ &= \sum_{k=1}^{K} [\mathrm{Tr}\left((\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T\right) + \underbrace{\mathrm{Tr}\left((\mathbf{X}_k - \bar{\mathbf{X}})(\bar{\mathbf{X}} - \mathbf{B})^T\right)}_{1} \\ &+ \underbrace{\mathrm{Tr}\left((\bar{\mathbf{X}} - \mathbf{B})(\mathbf{X}_k - \bar{\mathbf{X}})^T\right)}_{2} + \mathrm{Tr}\left((\bar{\mathbf{X}} - \mathbf{B})(\bar{\mathbf{X}} - \mathbf{B})^T\right)] \end{aligned} \tag{41}$$

Terms 1 and 2 are equal since $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$. Expanding term 1,

$$\sum_{k=1}^{K} \mathrm{Tr}\left((\mathbf{X}_k - \bar{\mathbf{X}})(\bar{\mathbf{X}} - \mathbf{B})^T\right) = \mathrm{Tr}\left[\underbrace{\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})}_{=0}(\bar{\mathbf{X}} - \mathbf{B})^T\right] = 0$$

The cross-terms 1 and 2 are zero and we are just left with the quadratic terms. Thus, it is proven that $\hat{\mathbf{B}}_0 = \bar{\mathbf{X}}$ minimizes $J_0(\mathbf{B})$. Putting $\hat{\mathbf{B}}_0$ into $J_0(\mathbf{B})$, we get

$$J_0(\hat{\mathbf{B}}_0) = \sum_{k=0}^{K-1} \mathrm{Tr}[(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T] \tag{42}$$

Thus,

$$p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{MNK/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^{K} \mathrm{Tr}[(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T]\right] \tag{43}$$

Determining $\hat{A}, \hat{\mathbf{B}}_1$ and $p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)$

To find $\hat{A}$ and $\hat{\mathbf{B}}_1$, need to find A and $\mathbf{B}$ maximizing $p(\mathbf{X}; A, \mathbf{B}, \mathcal{H}_1)$ or equivalently minimizing

$$J_1(A, \mathbf{B}) = \sum_{k=1}^{K} \mathrm{Tr}[(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})(\mathbf{X}_k - A\mathbf{S}_k - \mathbf{B})^T] \tag{44}$$

Now let

$$\mathbf{Y}_k = \mathbf{X}_k - A\mathbf{S}_k \tag{45}$$

$J_1(A, \mathbf{B})$ becomes

$$J_1(A, \mathbf{B}) = \sum_{k=1}^{K} \mathrm{Tr}[(\mathbf{Y}_k - \mathbf{B})(\mathbf{Y}_k - \mathbf{B})^T] \tag{46}$$

To minimize over $\mathbf{B}$ we use previous result in subsection A:

$$\hat{\mathbf{B}}_1 = \frac{1}{K} \sum_{k=1}^{K} \mathbf{Y}_k = \bar{\mathbf{Y}} \tag{47}$$

Note that $\mathbf{Y}_k$ depends on A. Now we need to minimize

$$J_1(A, \hat{\mathbf{B}}_1) = \sum_{k=1}^{K} \mathrm{Tr}[(\mathbf{Y}_k - \bar{\mathbf{Y}})(\mathbf{Y}_k - \bar{\mathbf{Y}})^T]$$

But,

$$\begin{aligned} \mathbf{Y}_k - \bar{\mathbf{Y}} &= \mathbf{X}_k - A\mathbf{S}_k - \left[\frac{1}{K} \sum_{k=1}^{K} (\mathbf{X}_k - A\mathbf{S}_k) \right] \\ &= \mathbf{X}_k - \bar{\mathbf{X}} - A(\mathbf{S}_k - \bar{\mathbf{S}}) \end{aligned}$$

So we can write

$$J_1(A, \hat{\mathbf{B}}_1) = \sum_{k=1}^{K} \mathrm{Tr}\left[(\mathbf{X}_k - \bar{\mathbf{X}} - A(\mathbf{S}_k - \bar{\mathbf{S}}))(\mathbf{X}_k - \bar{\mathbf{X}} - A(\mathbf{S}_k - \bar{\mathbf{S}}))^T \right]$$

Now let $\mathbf{U}_k = \mathbf{X}_k - \bar{\mathbf{X}}$ and $\mathbf{V}_k = \mathbf{S}_k - \bar{\mathbf{S}}$, we are left with

$$J_1(A, \hat{\mathbf{B}}_1) = \sum_{k=1}^{K} \mathrm{Tr}\left[(\mathbf{U}_k - A\mathbf{V}_k)(\mathbf{U}_k - A\mathbf{V}_k)^T \right]$$

Taking it derivative with respect to A

$$\begin{aligned}
\frac{\partial J_1}{\partial A} &= \sum_{k=1}^{K} \mathrm{Tr}\left[(\mathbf{U}_k - A\mathbf{V}_k)(-\mathbf{V}_k)^T + (-\mathbf{V}_k)(\mathbf{U}_k - A\mathbf{V}_k)^T\right] \\
&= \sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{V}_k(\mathbf{U}_k - A\mathbf{V}_k)^T\right] + \mathrm{Tr}\left[\mathbf{V}_k(\mathbf{U}_k - A\mathbf{V}_k)^T\right] \\
&= \sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{V}_k(\mathbf{U}_k - A\mathbf{V}_k)^T\right] \\
&= \sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{V}_k\mathbf{U}_k^T - A\mathbf{V}_k\mathbf{V}_k^T\right] \qquad (48)
\end{aligned}$$

and setting it equal to zero:

$$\begin{aligned}
\hat{A} &= \frac{\mathrm{Tr}\left[\sum_{k=1}^{K} \mathbf{V}_k\mathbf{U}_k^T\right]}{\mathrm{Tr}\left[\sum_{k=1}^{K} \mathbf{V}_k\mathbf{V}_k^T\right]} \\
&= \frac{\mathrm{Tr}\left[\sum_{k=1}^{K} (\mathbf{S}_k - \bar{\mathbf{S}})(\mathbf{X}_k - \bar{\mathbf{X}})^T\right]}{\mathrm{Tr}\left[\sum_{k=1}^{K} (\mathbf{S}_k - \bar{\mathbf{S}})(\mathbf{S}_k - \bar{\mathbf{S}})^T\right]} \qquad (49)
\end{aligned}$$

Putting $\hat{A}$ and $\hat{\mathbf{B}}_1$ into $J_1(A, \mathbf{B})$

$$\begin{aligned}
J_1(\hat{A}, \hat{\mathbf{B}}_1) &= \sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{U}_k(\mathbf{U}_k - \hat{A}\mathbf{V}_k)^T\right] - \hat{A}\underbrace{\sum_{k=1}^{K} \mathrm{Tr}\left[\mathbf{V}_k(\mathbf{U}_k - \hat{A}\mathbf{V}_k)^T\right]}_{=0} \\
&= \mathrm{Tr}\left[\sum_{k=1}^{K} \mathbf{U}_k\mathbf{U}_k^T\right] - \hat{A}\ \mathrm{Tr}\left[\sum_{k=1}^{K} \mathbf{U}_k\mathbf{V}_k^T\right] \\
&= \mathrm{Tr}\left[\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T\right] - \hat{A}\ \mathrm{Tr}\left[\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{S}_k - \bar{\mathbf{S}})^T\right] \\
&= J_0(\hat{\mathbf{B}}_0) - \hat{A}\ \mathrm{Tr}\left[\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{S}_k - \bar{\mathbf{S}})^T\right] \qquad (50)
\end{aligned}$$

Thus,

$$p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{MNK/2}} \exp\left[-\frac{1}{2\sigma^2} J_1(\hat{A}, \hat{\mathbf{B}}_1)\right] \qquad (51)$$

Finding the Test Statistic

It is easier to deal with the log-likelihood since both $p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)$ and $p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)$ are Gaussian. Using the previous results,

$$\begin{aligned}
2 \ln L_G(\mathbf{X}) &= 2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} \\
&= -\frac{2}{2\sigma^2} \left[J_1(\hat{A}, \hat{\mathbf{B}}) - J_0(\hat{\mathbf{B}}_0) \right] \\
&= -\frac{2}{2\sigma^2} \left[-\hat{A} \ \mathrm{Tr} \left(\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{S}_k - \bar{\mathbf{S}})^T \right) \right] \\
&= \frac{1}{\sigma^2} \left[\hat{A} \ \mathrm{Tr} \left(\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{S}_k - \bar{\mathbf{S}})^T \right) \right] \\
&= \frac{1}{\sigma^2} \left[\frac{\mathrm{Tr}^2 \left(\sum_{k=1}^{K} (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{S}_k - \bar{\mathbf{S}})^T \right)}{\mathrm{Tr} \left(\sum_{k=1}^{K} (\mathbf{S}_k - \bar{\mathbf{S}})(\mathbf{S}_k - \bar{\mathbf{S}})^T \right)} \right] \\
&= \frac{\sum_{k=0}^{K-1} ||\mathbf{S}_k - \bar{\mathbf{S}}||^2}{\sigma^2} \left[\frac{\sum_{k=0}^{K-1} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=0}^{K-1} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \right]^2 \qquad (52)
\end{aligned}$$

where $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}\mathbf{Y}^T)$. Since the starting position and velocity are unknown, we must maximize this over (m_0, n_0) and (v_x, v_y).

$$\max_{(m_0,n_0),(v_x,v_y)} 2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} \qquad (53)$$

List of References

[1] S. D. Blostein and T. S. Huang, "Detecting small, moving objects in image sequences using sequential hypothesis testing," *Signal Processing, IEEE Transactions on*, vol. 39, no. 7, pp. 1611–1629, 1991.

[2] D. S. Chan, D. A. Langan, and D. A. Staver, "Spatial-processing techniques for the detection of small targets in ir clutter," in *OE/LASE'90, 14-19 Jan., Los Angeles, CA*. International Society for Optics and Photonics, 1990, pp. 53–62.

[3] C. D. Wang, "Adaptive spatial/temporal/spectral filters for background clutter suppression and target detection," *Optical Engineering*, vol. 21, no. 6, pp. 216 033–216 033, 1982.

[4] D. S. Chan, "Unified framework for ir target detection and tracking," in *Aerospace Sensing*. International Society for Optics and Photonics, 1992, pp. 66–76.

[5] I. Reed, R. Gagliardi, and H. Shao, "Application of three-dimensional filtering to moving target detection," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 6, pp. 898–905, 1983.

[6] I. Reed, R. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3-d matched filtering," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 327–336, 1988.

[7] B. Porat and B. Friedlander, "A frequency domain algorithm for multiframe detection and estimation of dim targets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 4, pp. 398–401, 1990.

[8] N. Mohanty, "Computer tracking of moving point targets in space," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 5, pp. 606–611, 1981.

[9] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 1, pp. 144–156, 1985.

[10] Y. Barniv and O. Kella, "Dynamic programming solution for detecting dim moving targets part ii: analysis," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 6, pp. 776–788, 1987.

[11] S. M. Tonissen and R. J. Evans, "Peformance of dynamic programming techniques for track-before-detect," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 32, no. 4, pp. 1440–1451, 1996.

[12] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "A recursive moving-target-indication algorithm for optical image sequences," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 26, no. 3, pp. 434–440, 1990.

[13] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*, 3rd ed. Springer Science & Business Media, 2005.

[14] S. Kay, "Signal fitting with uncertain basis functions," *IEEE Signal Processing Letters*, vol. 6, no. 18, pp. 383–386, 2011.

[15] S. Kay, "A computationally efficient nonlinear least squares method using random basis functions," *IEEE Signal Processing Letters*, vol. 7, no. 20, pp. 721–724, 2013.

[16] S. M. Kay, *Fundamentals of statistical signal processing, volume II: Detection theory*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

[17] M. J. Wichura, *The coordinate-free approach to linear models*. Cambridge University Press, 2006, vol. 19.

[18] Michigan Tech Research Institute (MTRI), 3600 Green Court, Suite 100, Ann Arbor, MI 48105, 2015.

[19] S. Kay and F. Cogun, "Integrated sensor detection/localization for multi-source data," in *Radar Conference, 2014 IEEE*. IEEE, 2014, pp. 0708–0711.

CHAPTER 3

Multimodal Target Detection via Integrated GLRT

by

Steven Kay and Fuat Cogun

Dept. of Electrical, Computer and Biomedical Engineering

University of Rhode Island, Kingston, RI, USA

published in IEEE Military Communications Conference, 2015.

Abstract

The focus of this paper is the multimodal detection and localization of a ground moving target based on radio frequency (RF) and infrared (IR) data. The target radiates a low probability of intercept (LPI) RF signal received by multiple passive RF sensors at scene and is imaged by using a stationary IR camera concurrently. To obtain the multimodal detector proposed in this paper, first, the generalized likelihood ratio test (GLRT) is employed to derive the RF and IR detectors individually. Then, the RF and IR detectors are integrated optimally to get the integrated GLRT. In order to avoid the computational complexity associated with the integrated GLRT, a suboptimal multimodal detector is implemented by applying the random basis functions (RBF) approach to the IR image sequence to reduce down the search space for the RF detector. It is shown that the suboptimal multimodal detector has better localization performance compared to the RF detector when one of the RF sensors is partially occluded.

3.1 Introduction

There are different types of difficulties encountered in the detection and localization of a target for radio frequency (RF) and infrared (IR) detectors. For an RF detector, the received RF low probability of intercept (LPI) signal is not known a priori and the received signal may be subject to multipath [1]. For an IR detector, temporary occlusion of the received IR image signal by trees, buildings, etc. causes problems and lowers detection performance. Another issue with the IR detector is the possibility of the vehicle stopping and appearing as part of the background. In this case, the IR detector can neither detect the presence of a target nor localize it. Note that when a single RF or IR detector is considered, even one of these difficulties will result in lower detection and localization performance. Therefore, the fusion of RF and IR data is essential to solve these difficulties and

a multimodal detector leads to better performing detectors.

In this paper, the detection and localization of a ground moving target emitting a low probability of intercept (LPI) RF signal is the objective. The target is imaged by using a stationary IR camera and the emitted LPI RF signal is received by multiple passive RF sensors at scene. Note that we derived the multimodal detector for this problem in [2] by finding the individual GLRTs for both RF and IR signal models first, then integrating them to find a GLRT detector based on the fusion of RF and IR data. For the sake of clarity, the signal models, GLRTs and optimal integration of the GLRTs are summarized in this paper. The interested reader is advised to refer to that paper for details. This paper is concentrated on the suboptimal implementation of the multimodal detector.

The outline of this paper is as follows. In the next section, the signal models and modeling assumptions are reviewed. Section 3.3 summarizes the derived GLRTs for the RF and IR signal models and Section 3.4 gives the integrated GLRT detector for the multimodal detection problem. The IR camera model used for simulation purposes and the suboptimal integration of RF and IR data are given in Section 3.5. Section 3.6 presents the obtained simulation results. Finally, the paper ends with conclusions.

3.2 Signal Models

In this section, signal models assumed for the RF and IR signals are summarized.

3.2.1 RF Signal Model

The hypothesis test for the RF signal is

$$\begin{aligned}\mathcal{H}_0 &: \tilde{x}_i[n] = \tilde{w}_i[n] \\ \mathcal{H}_1 &: \tilde{x}_i[n] = \tilde{A}_i \tilde{s}[n-n_i] e^{\frac{j2\pi k_i (n-n_i)}{N}} + \tilde{w}_i[n]\end{aligned}$$

where $\tilde{w}_i[n] \sim \mathcal{CN}(0, \sigma^2)$, $n = 0, 1, \ldots, N-1$ and $i = 0, 1, \ldots, M-1$. Note that we use a tilde ($\sim$) to indicate that we are handling complex data. Here, $\tilde{s}[n]$ is the n^{th} time sample of the LPI transmit signal, $\tilde{x}_i[n]$ is the n^{th} time sample of the observed signal, $\tilde{w}_i[n]$ is the n^{th} time sample of the noise, $\tilde{A}_i$ is the complex gain/attenuation, n_i is the time delay and k_i/N is the Doppler shift at i^{th} receiver. Note that $\tilde{A}_i$, $\tilde{s}[n]$, n_i and k_i are all unknown. Here $\mathcal{H}_0$ symbolizes no target and $\mathcal{H}_1$ symbolizes a target present case. The data model can also be written as

$$\tilde{\mathbf{x}}_i = \tilde{A}_i \mathbf{P}^{n_i} \tilde{\mathbf{W}}^{k_i} \tilde{\mathbf{s}} + \tilde{\mathbf{w}}_i \tag{54}$$

which is the observed data at the i^{th} sensor over a time interval of length N samples. That is $\tilde{\mathbf{x}}_i = [\tilde{x}_i[0]\ \tilde{x}_i[1]\ \ldots\ \tilde{x}_i[N-1]]^T$. Similarly, $\tilde{\mathbf{w}}_i = [\tilde{w}_i[0]\ \tilde{w}_i[1]\ \ldots\ \tilde{w}_i[N-1]]^T$ and $\tilde{\mathbf{s}} = [\tilde{s}[0]\ \tilde{s}[1]\ \ldots\ \tilde{s}[N-1]]^T$. $\mathbf{P}$ is an $N \times N$ permutation matrix producing time shifts and $\tilde{\mathbf{W}} = \text{diag}(\omega^0, \omega^1, \ldots, \omega^{N-1})$ producing Doppler shifts where $\omega = \exp\left(\frac{j2\pi}{N}\right)$. Then the composite hypothesis testing problem becomes:

$$\mathcal{H}_0 : \tilde{\mathbf{s}} = \mathbf{0}$$

$$\mathcal{H}_1 : \tilde{\mathbf{s}} \neq \mathbf{0}$$

3.2.2 IR Signal Model

The hypothesis test for the IR signal is

$$\mathcal{H}_0 : x[m, n, k] = B[m, n] + w[m, n, k]$$

$$\mathcal{H}_1 : x[m, n, k] = As[m, n, k] + B[m, n] + w[m, n, k]$$

for $m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$ (image pixels) and $k = 0, 1, \ldots, K-1$ (frame number). Note that in the IR signal model, we are handling real data as opposed to the complex RF signal model. The background $B[m, n]$ is assumed to be constant with frame and unknown. $w[m, n, k]$ is additive white Gaussian noise (AWGN) with variance σ^2 and independent from pixel-to-pixel and frame-to-frame.

The target signal $s[m, n, k]$ can be expressed as:

$$s[m, n, k] = g[m - m_0 - v_x k, n - n_0 - v_y k] \tag{55}$$

where (m_0, n_0) is the starting position and (v_x, v_y) is the velocity of the target, both of which are unknown. The velocity of the target is assumed to be constant throughout the data collection time interval. The function g is the target function which characterizes the size of the target. In addition, the target is assumed to be point-like so the signal $s[m, n, k]$ is represented as one pixel.

3.3 Generalized Likelihood Ratio Test (GLRT)

Even though there is no optimality affiliated with the GLRT, we make use of GLRTs for both the RF and IR signals since in practice it works quite well. Further, the power of the GLRT is that once the maximum likelihood estimates (MLEs) are found to provide the information used in target detection, the information necessary for target localization is also available automatically.

3.3.1 GLRT for the RF signal

The GLRT with the unknown parameters $\mathbf{r}_0 = [x_0 \ y_0]^T$ and $\mathbf{v}$ for the data observed at M sensors for a time interval of length N is [1]

$$\max_{(\mathbf{r}_0, \mathbf{v})} \lambda_{\max}(\tilde{\mathbf{C}}(\mathbf{r}_0, \mathbf{v})) \tag{56}$$

where $\mathbf{r}_0$ is the initial target position, $\mathbf{v}$ is the velocity of the target, and $\lambda_{\max}$ is the maximum eigenvalue of the $M \times M$ complex ambiguity matrix (CAM) $\tilde{\mathbf{C}}$ defined by

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_0 & \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_1 & \cdots & \tilde{\mathbf{y}}_0^H \tilde{\mathbf{y}}_{M-1} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_0 & \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_1 & \cdots & \tilde{\mathbf{y}}_{M-1}^H \tilde{\mathbf{y}}_{M-1} \end{bmatrix} \tag{57}$$

where $\tilde{\mathbf{y}}_i = (\mathbf{P}^{n_i} \tilde{\mathbf{W}}^{k_i})^H \tilde{\mathbf{x}}_i$. Note that $\tilde{\mathbf{y}}_i$ is the aligned signal in delay and Doppler at sensor i.

Note that once the maximum of $\lambda_{\max}(\tilde{\mathbf{C}})$ is found, then MLEs of $\mathbf{r}_0$ and $\mathbf{v}$ are available which are necessary for localization. The delays and Doppler shifts are a function of the initial target position and velocity, i.e., $\mathbf{n} = \mathbf{f}(\mathbf{r}_0)$ and $\mathbf{k} = \mathbf{g}(\mathbf{r}_0, \mathbf{v})$. The values $(\hat{\mathbf{r}}_0, \hat{\mathbf{v}})$ that maximize the maximum eigenvalue $\lambda_{\max}(\tilde{\mathbf{C}}(\mathbf{r}_0, \mathbf{v}))$ are the MLEs for $(\mathbf{r}_0, \mathbf{v})$, respectively.

3.3.2 GLRT for the IR signal

The GLRT with the unknown parameters (m_0, n_0) and (v_x, v_y) is

$$\max_{(m_0,n_0),(v_x,v_y)} 2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} \tag{58}$$

where $\mathbf{X}$ is Gaussian, $\hat{\mathbf{B}}_0$ is the MLE of the background under $\mathcal{H}_0$, $\hat{A}$ is the MLE of the signal amplitude under $\mathcal{H}_1$, and $\hat{\mathbf{B}}_1$ is the MLE of the background under $\mathcal{H}_1$. The required MLE under $\mathcal{H}_0$ is found as

$$\hat{\mathbf{B}}_0 = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{X}_k \tag{59}$$

and of $A, \mathbf{B}$ under $\mathcal{H}_1$

$$\hat{\mathbf{B}}_1 = \frac{1}{K} \sum_{k=0}^{K-1} (\mathbf{X}_k - \hat{A}\mathbf{S}_k) = \bar{\mathbf{X}} - \hat{A}\bar{\mathbf{S}} \tag{60}$$

$$\hat{A} = \frac{\sum_{k=0}^{K-1} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=0}^{K-1} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \tag{61}$$

where $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}\mathbf{Y}^T)$ and $\mathbf{X}_k, \mathbf{S}_k$ are the k^{th} observed IR image and target frame, respectively. Replacing the MLEs back, the sufficient statistic is obtained as

$$2 \ln \frac{p(\mathbf{X}; \hat{A}, \hat{\mathbf{B}}_1, \mathcal{H}_1)}{p(\mathbf{X}; \hat{\mathbf{B}}_0, \mathcal{H}_0)} = \frac{\sum_{k=0}^{K-1} ||\mathbf{S}_k - \bar{\mathbf{S}}||^2}{\sigma^2} \left[\frac{\sum_{k=0}^{K-1} \langle \mathbf{X}_k - \bar{\mathbf{X}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle}{\sum_{k=0}^{K-1} \langle \mathbf{S}_k - \bar{\mathbf{S}}, \mathbf{S}_k - \bar{\mathbf{S}} \rangle} \right]^2 \tag{62}$$

The GLRT requires the maximization of (62) over unknown starting position (m_0, n_0) and velocity (v_x, v_y) ($\mathbf{S}_k$'s depend on this). Note that the used 2D inner product space of images can easily be extended to 3D inner product space of

data cubes. If we let $\mathbf{X}$ and $\mathbf{Y}$ be $M \times N \times K$ data cubes, the 3D inner product space is defined as

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} x[m,n,k] y[m,n,k]. \tag{63}$$

This approach employs the coordinate-free approach to linear models. [3]

3.4 Integration of RF and IR GLRTs

The integrated GLRT is derived from the RF and IR GLRTs assuming that the observation noise is independent between RF and IR data sets. Letting $\tilde{\mathbf{x}}_1$ be the complex RF data vector and $\mathbf{X}_2$ be the real IR data cube, and σ_1^2, σ_2^2 be the noise powers, respectively, the overall GLRT is found as

$$L_G(\tilde{\mathbf{x}}_1, \mathbf{X}_2) = \max_{\mathbf{r}_0, \mathbf{v}} \left[\frac{\lambda_{\max}(\tilde{\mathbf{C}}(\tilde{\mathbf{x}}_1))}{\sigma_1^2} + \frac{\langle \mathbf{X}_2 - \bar{\mathbf{X}}_2, \mathbf{S} - \bar{\mathbf{S}} \rangle^2}{2\sigma_2^2 \langle \mathbf{S} - \bar{\mathbf{S}}, \mathbf{S} - \bar{\mathbf{S}} \rangle} \right] \tag{64}$$

where $\tilde{\mathbf{C}}$ is the complex ambiguity function matrix and $\mathbf{r}_0, \mathbf{v}$ are the initial target position and velocity vector of the target. Notice that the 3D inner product space defined in (63) is used here.

3.5 Simulation Setup

For simulation purposes, assume that we have a ground moving emitter (target) at initial position E = $(1800, 2200)$ and 3 stationary RF receivers at locations RX1 = $(0, 0)$, RX2 = $(0, 4000)$ and RX3 = $(4000, 0)$ as seen in Fig. 20. The target is moving with constant velocity $v = (20, -10)$.

To simultaneously generate the IR data of the observed scene, an IR camera is assumed to be located at $(2000, 2000)$ which is the center of the scene. The height of the IR camera from the ground is denoted as x_3. Note that the IR camera is required to do a mapping from 3D coordinates to 2D coordinates in order to generate an IR video sequence. In this work, the "pinhole camera model" is used as our camera model which will be described in the next section.

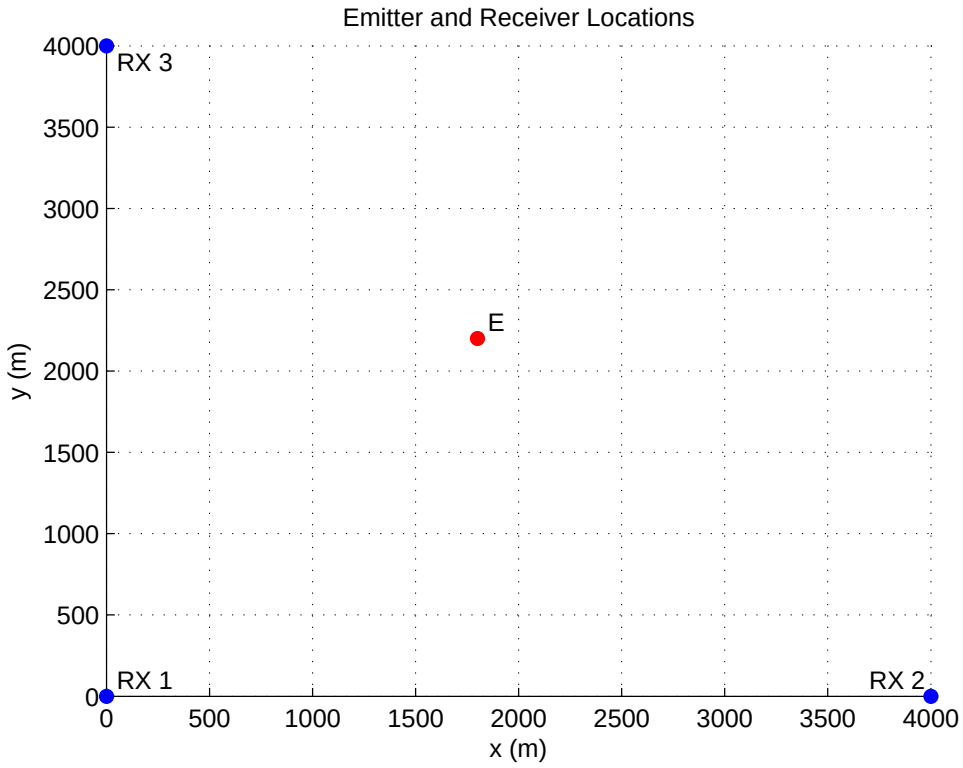


Figure 20: RF emitter and sensors locations

3.5.1 Pinhole Camera Model

The pinhole camera model describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane of an ideal pinhole camera, where the camera aperture is described as a point and no lenses are used to focus light [4]. The camera model is used as a first order approximation of the mapping from the 3D observed scene to the 2D IR image plane.

Fig. 21 shows the geometry of the pinhole camera model used in this paper. The camera aperture is located at origin O. 3D space is represented by X1, X2, X3 axes where X3 axis is pointing in the viewing direction of the camera, called the "principal axis", and (X1,X2) plane is called the "principal plane". The image plane is (Y1,Y2) plane which is located at distance f from the origin O in the negative direction of the X3 axis.

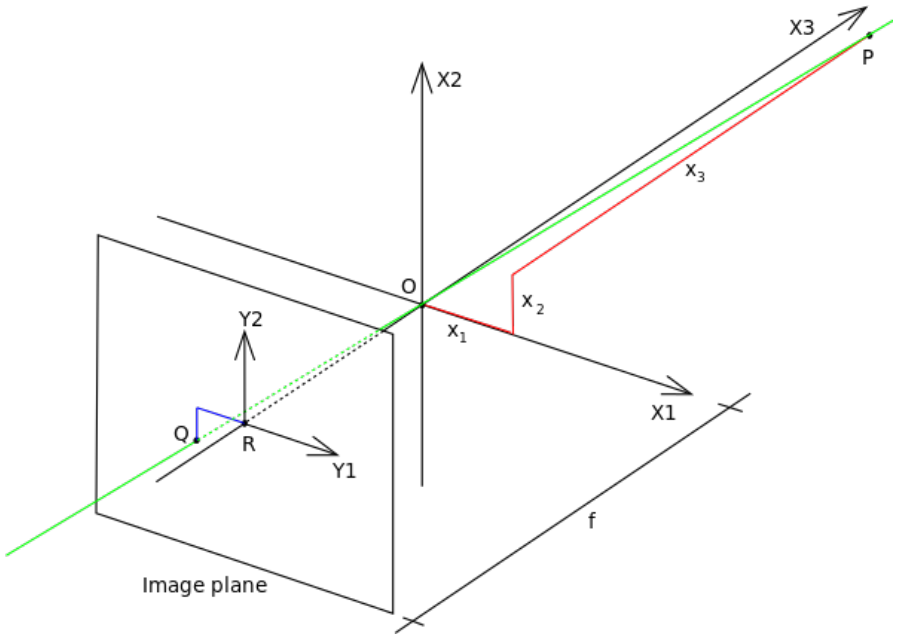


Figure 21: The geometry of the pinhole camera [4]

Now, assume that we have a target at point P somewhere on the scene at coordinate (x_1, x_2, x_3) relative to the axes X1, X2, X3. The projection of point P onto the image plane, denoted Q, has coordinates (y_1, y_2) relative to the axes Y1, Y2 and the mapping is given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = -\frac{f}{x_3} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

For simulation purposes, the mapping parameters are selected as $x_3 = 1000$ and $f = 10$. By applying the mapping to the observed scene, the IR video sequence is generated as illustrated in Fig. 22. The mapped IR frame size is chosen as 64 x 64 in this particular case. Having the multi-modal (RF/LPI signal and IR video sequence) data of the observed scene, it is possible to apply the RF and IR detectors simultaneously and see the performance improvement gained by the integration of both detectors.

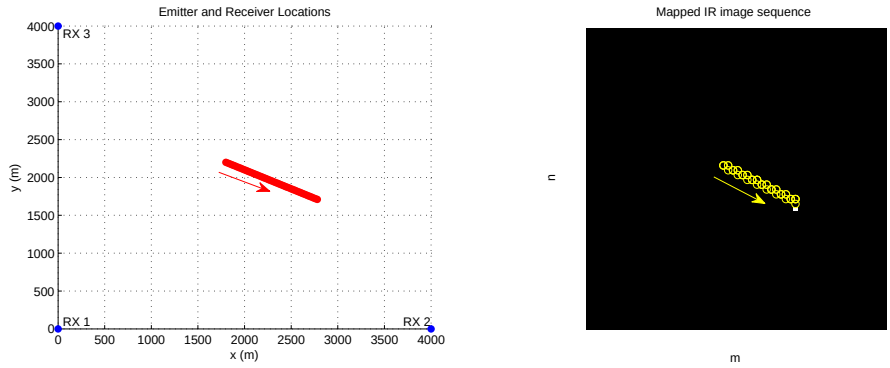


Figure 22: Mapping from 3D observed scene onto 2D IR image plane

3.5.2 Suboptimal Integration of RF and IR detectors

Although the optimal integration of RF and IR GLRTs (integrated GLRT) is done as given in (64), the straightforward implementation of the integrated GLRT detector is computationally very expensive since the overall statistic is maximized jointly over target motion parameters. Hence, a suboptimal approach is taken at this point to obtain preliminary results. First, the random basis functions (RBF) approach [5] is applied to the IR image sequences [2] for a predetermined number of iterations. By doing so, we reduce down the search space of the RF detector for the unknown target motion parameters. Finally, GLRT for the RF data is run over the reduced down search space to obtain the multi-modal estimates of the target motion parameters.

3.6 Simulation Results

Monte Carlo simulations are carried out to obtain error vectors where an error vector is defined as $\boldsymbol{\epsilon} = \mathbf{r}_0 - \hat{\mathbf{r}}_0$ for one realization. Note that the scatter of position estimates is used as the performance metric to assess detector performance in this study. We run R realizations of the same target detection problem under noise and each $\hat{\mathbf{r}}_0$ is plotted on a scatter diagram.

To better illustrate the performance improvement (or robustness) gained by

the joint use of RF and IR detectors, we take the following approach. First, the error vectors of the individual RF detector are computed for *i)* partial occlusion in one of the RF sensors and *ii)* no occlusion scenarios. Then, the error vectors of the integrated detector are computed for both scenarios. Finally, the error vectors of the individual (RF) and integrated (RF+IR) detectors are compared.

For 50 realizations, the performance of the individual RF detector is shown in Fig. 23. The black stars and blue circles on the scatter plot represent the initial position estimates of the target for no occlusion and 90 percent occlusion at RF receiver #2, respectively. It is seen that an occlusion at one of the RF receivers (sensors) results in large error vectors. Thus, we have lower RF detector performance. By incorporating the IR detector, the performance of the integrated detector is illustrated as shown in Fig. 24. Two iterations are run on IR detector (via RBF approach) to reduce the search region for RF detector down to 1000x1000. Note that more than two iterations could also be run to further reduce down the search region if it is known that the SNR of the IR image sequence is high. It is observed in Fig. 24 that although the no-occlusion detector performance is not improved significantly, the partial-occlusion detector performance is much better when the integrated detector is used. Therefore, the integrated detector is more robust and has better detector performance when compared to the individual RF detector.

3.7 Conclusion

In this paper, the multimodal RF-IR detection problem is addressed. The RF and IR signal models are used to determine the individual GLRTs. Then the optimal integration of RF and IR data for target detection/localization is obtained by using an integrated GLRT. For simulation purposes, the IR video sequence is generated from the observed 3D scene by using the pinhole camera model. As

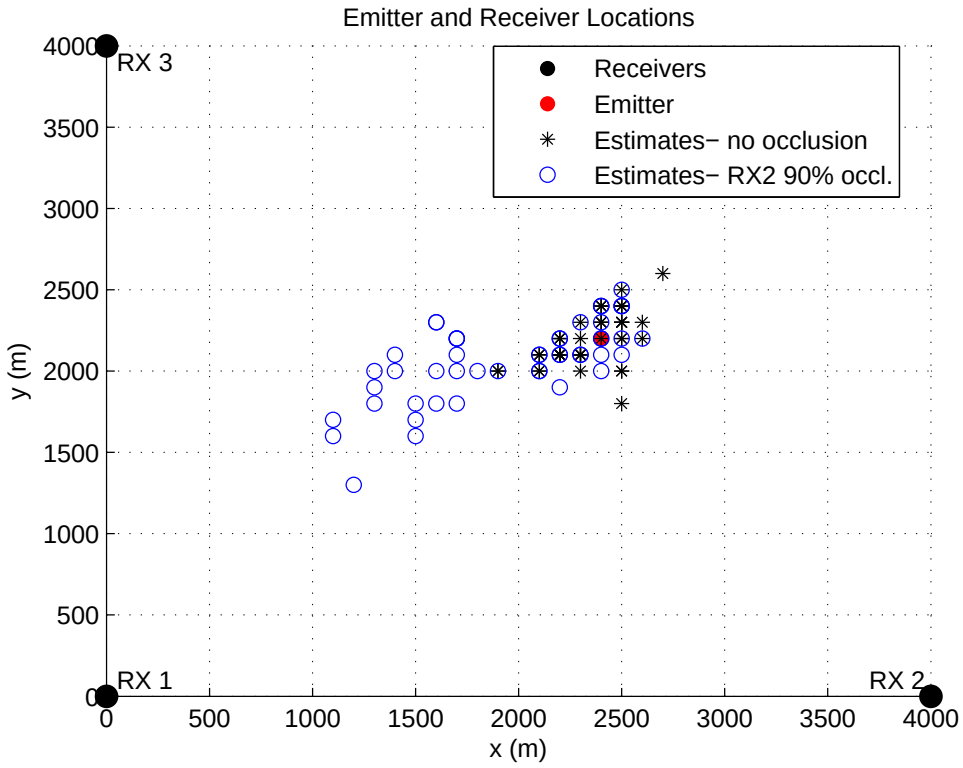


Figure 23: Initial position estimates of the individual RF detector

the straightforward implementation of the integrated GLRT detector is computationally very expensive, the multimodal detector is obtained in a suboptimal way. The random basis function (RBF) approach is applied on the IR image sequences to reduce down the search space of the RF detector, then, the GLRT for the RF data is run over the reduced down search space to obtain the multimodal estimates of the target motion parameters. The error vectors are used as the performance criterion and it is observed that the multimodal detector is more robust and has better detection performance than the individual RF detector, especially for the case which one of the RF sensors is partially occluded. The implementation of the actual integrated GLRT is considered as future work.

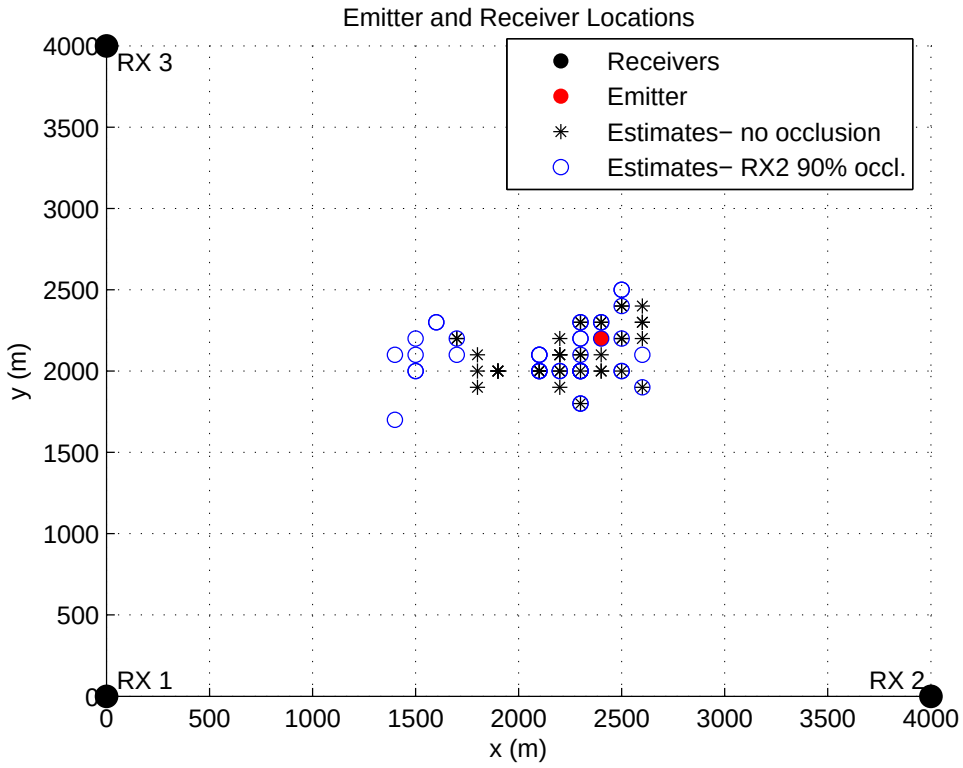


Figure 24: Initial position estimates of the joint RF + IR detector

List of References

[1] N. Vankayalapati and S. Kay, "Asymptotically optimal detection of low probability of intercept signals using distributed sensors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 1, pp. 737–748, 2012.

[2] S. Kay and F. Cogun, "Integrated sensor detection/localization for multi-source data," in *Radar Conference, 2014 IEEE*. IEEE, 2014, pp. 0708–0711.

[3] M. J. Wichura, *The coordinate-free approach to linear models*. Cambridge University Press, 2006, vol. 19.

[4] "Pinhole camera model." Oct. 2014. [Online]. Available: https://en.wikipedia.org/wiki/Pinhole_camera_model

[5] S. Kay, "A computationally efficient nonlinear least squares method using random basis functions," *IEEE Signal Processing Letters*, vol. 7, no. 20, pp. 721–724, 2013.

CHAPTER 4

Alternative Approaches to Data Compression for Distributed Detection

by

Fuat Cogun and Steven Kay

Dept. of Electrical, Computer and Biomedical Engineering

University of Rhode Island, Kingston, RI, USA

published in IEEE Radar Conference, 2016.

Abstract

In this paper, the distributed detection problem of linear and nonlinear signals embedded in white Gaussian noise (WGN) is considered. First, the asymptotically optimal generalized likelihood ratio test (GLRT) detector is derived for both signal models. It is found that the GLRT detector requires the submission of all observed data to the central processor which is practically infeasible. Thus, several test statistics based on compressing the observed data at each sensor are proposed. Monte Carlo simulations are carried out to plot the receiver operating characteristic (ROC) curves in order to compare the performance of the proposed detectors for a nonlinear signal example.

4.1 Introduction

Multi-sensor distributed detection problem has received great interest since it was first introduced by Tenney [1]. A detailed introduction and overview of the decentralized detection problem can be found in [2].

In classical centralized data fusion, all the sensor observations are transmitted to a central processor to be combined to obtain a global decision. Although centralized data fusion achieves the optimum performance, it requires communication links having high bandwidth and a central processor having large memory. Most of the detection problems are based on large data records, therefore, the data fusion method cannot be implemented practically. Instead, distributed detection methods are considered to address the issue [3].

In the distributed detection problem, the main goal is to derive the detector having minimum performance loss compared to the optimal centralized data fusion detector. There has been recent research on wireless sensor networks (WSN) addressing this problem which concentrates on two different approaches: hard decision [4] [5] and soft decison fusion [6] [7]. In the hard decision fusion approach,

each sensor submits its hard decision (0 or 1) based on local observations with the advantage of a low communication bandwidth requirement. However, the detection performance of these distributed detectors is poor. In the soft decision fusion approach, each distributed sensor determines a soft decision (a value between 0 and 1) based on observations rather than a hard decision (0 or 1) and submits it to central processor, usually after quantization [8]. This approach achieves better performance than the hard decision fusion approach, but still exhibits degraded performance compared to the optimal centralized detector.

The contribution of this paper is the compression of data into local statistics at each sensor and submitting them to construct the test statistic at the central processor for the global decision making. This is different from what has been done up until now. Instead of submitting local decisions based on measurements at each node, we are submitting compressed data, ie. local statistics, introducing a new approach to the distributed detection problem.

4.2 Signal Models

Linear and nonlinear signal models observed by distributed sensors are used in this work. These signal models are introduced next.

4.2.1 Linear Signal Model

The following linear model is used:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_i = \mathbf{w}_i \\ \mathcal{H}_1 &: \mathbf{x}_i = \mathbf{h}_i\theta + \mathbf{w}_i \end{aligned} \tag{65}$$

for $i = 1, \ldots, M$ where $\mathbf{x}_i, \mathbf{h}_i, \mathbf{w}_i$ are $N \times 1$ vectors, θ is the unknown scalar parameter. Note that $\mathbf{w}_i$'s are independent and identically distributed (iid) with the distribution $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 is assumed to be known. The hypothesis

testing problem can also be seen as a parameter testing problem:

$$\mathcal{H}_0 : \theta = 0$$
$$\mathcal{H}_1 : \theta \neq 0 \tag{66}$$

4.2.2 Nonlinear Signal Model

The following model is used for the nonlinear signal:

$$\begin{aligned} \mathcal{H}_0 &: x_i[n] = w_i[n] \\ \mathcal{H}_1 &: x_i[n] = s[n;\theta] + w_i[n] \end{aligned}$$

for $n = 0, \ldots, N-1$ and $i = 1, \ldots, M$, or equivalently,

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_i = \mathbf{w}_i \\ \mathcal{H}_1 &: \mathbf{x}_i = \mathbf{s}(\theta) + \mathbf{w}_i \end{aligned} \tag{67}$$

for $i = 1, \ldots, M$ where $\mathbf{x}_i, \mathbf{s}, \mathbf{w}_i$ are $N \times 1$ vectors, θ is the unknown scalar parameter. $\mathbf{w}_i$'s are independent and identically distributed (iid) with the distribution $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 is known. Assuming that $\theta = 0 \Leftrightarrow \mathbf{s}(\theta) = \mathbf{0}$, the hypothesis testing problem can be represented as a parameter testing problem:

$$\mathcal{H}_0 : \theta = 0$$
$$\mathcal{H}_1 : \theta \neq 0 \tag{68}$$

4.3 Test Statistics

In this section, first, the asymptotically optimal GLRT [9] is used to obtain detectors for both linear and nonlinear signals. Then, data compression is considered to obtain the test statistics which are used in the efficient distributed detection.

4.3.1 Linear Model

GLRT for linear signal model

The GLRT statistic for the linear signal model is found as:

$$T_G(\mathbf{x}) = \max_{\theta} \frac{1}{\sigma^2} \sum_{i=1}^{M} \left\{ 2\mathbf{h}_i^T \mathbf{x}_i \theta - \mathbf{h}_i^T \mathbf{h}_i \theta^2 \right\} \tag{69}$$

Computing $T_G(\mathbf{x})$ requires the submission of all data ($\mathbf{x}_i$'s for $i = 1, \ldots, M$) to the central processor which we want to avoid. It is desired to compress the data at each sensor before sending it, which will be considered next.

Exact Sufficient Statistics

By finding the MLE (maximum likelihood estimate) of θ at sensor i based on $\mathbf{x}_i$ only, denoted as $\hat{\theta}_i = \frac{\mathbf{h}_i^T \mathbf{x}_i}{\mathbf{h}_i^T \mathbf{h}_i}$ and the observed Fisher information of $\hat{\theta}_i$ as $\hat{I}(\hat{\theta}_i) = \mathbf{h}_i^T \mathbf{h}_i / \sigma^2$ (in this case $\hat{I}(\hat{\theta}_i) = I(\hat{\theta}_i)$) we can write the test statistic as

$$T_G(\mathbf{x}) = \max_{\theta} \sum_{i=1}^{M} \left\{ 2\hat{I}(\hat{\theta}_i)\hat{\theta}_i \theta - \hat{I}(\hat{\theta}_i)\theta^2 \right\} \tag{70}$$

Here $\hat{\theta}_i$ and $\hat{I}(\hat{\theta}_i)$ are the *exact* sufficient statistics to be submitted for sensor i, i.e., it is possible to compress data at each sensor without losing any information, thus; lossless compression is possible. By analytically maximizing over θ at the central processor, the global MLE, $\hat{\theta}$, is found as

$$\hat{\theta} = \frac{\sum_{i=1}^{M} \hat{I}(\hat{\theta}_i)\hat{\theta}_i}{\sum_{i=1}^{M} \hat{I}(\hat{\theta}_i)} \tag{71}$$

By replacing (71) into (70), the test statistic at the central processor becomes:

$$T_G(\mathbf{x}) = \frac{\left(\sum_{i=1}^{M} \hat{I}(\hat{\theta}_i)\hat{\theta}_i \right)^2}{\sum_{i=1}^{M} \hat{I}(\hat{\theta}_i)} \tag{72}$$

4.3.2 Nonlinear Model

GLRT for NL signal model

The GLRT statistic for the nonlinear signal model is found as:

$$T_G(\mathbf{x}) = \max_{\theta} \frac{1}{\sigma^2} \sum_{i=1}^{M} \left\{ 2\mathbf{x}_i^T \mathbf{s}(\theta) - \mathbf{s}(\theta)^T \mathbf{s}(\theta) \right\} \tag{73}$$

Computing $T_G(\mathbf{x})$ requires the submission of all data ($\mathbf{x}_i$'s for $i = 1, \ldots, M$) to the central processor which we want to avoid. It is desired to compress the data at each sensor before sending it. Lossless compression is not possible for the nonlinear model, so we will consider lossy compression methods which result in some detector performance loss next.

2^{nd} Order Expansion of Log-likelihoods

In this section, we consider the second order Taylor expansion of $\ln p(\mathbf{x}_i; \theta)$ at local MLEs $\hat{\theta}_i$ for each sensor $i = 1, \ldots, M$, where $\hat{\theta}_i$ is found by

$$\hat{\theta}_i = \arg_\theta \min \left[(\mathbf{x}_i - \mathbf{s}(\theta))^T (\mathbf{x}_i - \mathbf{s}(\theta)) \right]$$

which satisfies $\frac{\partial \ln p(\mathbf{x}_i;\theta)}{\partial \theta}\Big|_{\hat{\theta}_i} = 0$. Thus, if we apply the 2^{nd} order Taylor expansion around $\hat{\theta}_i$, we obtain the following approximation:

$$\ln p(\mathbf{x}_i; \theta) \approx \ln p(\mathbf{x}_i; \hat{\theta}_i) + \frac{1}{2} \left. \frac{\partial^2 \ln p(\mathbf{x}_i; \theta)}{\partial \theta^2} \right|_{\hat{\theta}_i} (\theta - \hat{\theta}_i)^2$$

By definition, the observed Fisher information of $\hat{\theta}_i$ is $\hat{I}(\hat{\theta}_i) = -\frac{\partial^2 \ln p(\mathbf{x}_i;\theta)}{\partial \theta^2}\Big|_{\hat{\theta}_i}$. Therefore,

$$\ln p(\mathbf{x}_i; \theta) \approx \ln p(\mathbf{x}_i; \hat{\theta}_i) - \frac{1}{2} \hat{I}(\hat{\theta}_i)(\theta - \hat{\theta}_i)^2$$

By using the second order approximation at sensors $i = 1, \ldots, M$, the approximate GLRT can be written as:

$$\begin{aligned} T_A(\mathbf{x}) &= \max_{\theta} \sum_{i=1}^{M} \left\{ \ln \left(\frac{p(\mathbf{x}_i; \hat{\theta}_i)}{p(\mathbf{x}_i; 0)} \right) - \frac{1}{2} \hat{I}(\hat{\theta}_i)(\theta - \hat{\theta}_i)^2 \right\} \\ &= \max_{\theta} \sum_{i=1}^{M} \left\{ \frac{1}{\sigma^2} \left[2\mathbf{x}_i^T \mathbf{s}(\hat{\theta}_i) - \mathbf{s}(\hat{\theta}_i)^T \mathbf{s}(\hat{\theta}_i) \right] - \frac{1}{2} \hat{I}(\hat{\theta}_i)(\theta - \hat{\theta}_i)^2 \right\} \end{aligned} \tag{74}$$

As in the linear model, each sensor needs to submit $\hat{\theta}_i$ and $\hat{I}(\hat{\theta}_i)$ in addition to $\mathbf{x}_i^T\mathbf{s}(\hat{\theta}_i)$ which are *approximately* sufficient statistics (not exact). At the central processor, it is possible to do the maximization over θ analytically to find $\hat{\theta}$, which produces (71) as in the linear case. Replacing (71) into (74) the test statistic at the central processor becomes:

$$T_A(\mathbf{x}) = \sum_{i=1}^{M} \frac{1}{\sigma^2}\left[2\mathbf{x}_i^T\mathbf{s}(\hat{\theta}_i) - \mathbf{s}(\hat{\theta}_i)^T\mathbf{s}(\hat{\theta}_i)\right] + \frac{1}{2}\frac{\left(\sum_{i=1}^{M}\hat{I}(\hat{\theta}_i)\hat{\theta}_i\right)^2}{\sum_{i=1}^{M}\hat{I}(\hat{\theta}_i)} - \frac{1}{2}\sum_{i=1}^{M}\hat{I}(\hat{\theta}_i)\hat{\theta}_i^2 \quad (75)$$

Variance Stabilizing Transformation

As the asymptotic distribution of $\hat{\theta}_i$ is $\hat{\theta}_i \sim \mathcal{N}(\theta, I^{-1}(\theta))$ the second derivative of the log-likelihood function depends on the unknown parameter θ. By applying the following variance-stabilizing transformation (VST):

$$\beta = g(\theta) = \int_{\theta} \frac{1}{\sqrt{I^{-1}(\theta')}} d\theta'$$

the asymptotic distribution of $\hat{\beta}_i$, which is the MLE of the transformed parameter β at sensor i, becomes $\hat{\beta}_i \sim \mathcal{N}(\beta, 1)$. Note that after the VST, the second derivative of the log-likelihood function does not depend on β. Therefore, the third and higher order derivatives of the log-likelihood function are all zero and we can truncate the expansion at the second order. Even though the distribution may be far from its asymptotic form, applying the VST makes the log-likelihood function become *more* quadratic in the transformed parameter space, thus, the second order Taylor approximation becomes better.

After applying the VST, the second order Taylor approximation of the log-likelihood function at the i^{th} sensor can be written as

$$\ln p(\mathbf{x}_i; \beta) \approx \ln p(\mathbf{x}_i; \hat{\beta}_i) - \frac{1}{2}(\beta - \hat{\beta}_i)^2$$

for $i = 1, \ldots, M$. Thus, the approximate test statistic obtained by the VST becomes:

$$T_{VST}(\mathbf{x}) = \max_{\beta} \sum_{i=1}^{M} \left\{ \frac{1}{\sigma^2} 2\mathbf{x}_i^T \mathbf{s}(g^{-1}(\hat{\beta}_i)) - \mathbf{s}(g^{-1}(\hat{\beta}_i))^T \mathbf{s}(g^{-1}(\hat{\beta}_i)) - \frac{1}{2}(\beta - \hat{\beta}_i)^2 \right\} \quad (76)$$

Each sensor only needs to submit $\hat{\beta}_i$ in this case.

K^{th} order Expansion of the Signal

An alternative approach is to apply the Taylor approximation directly to the nonlinear signal, instead of the log-likelihood function. The advantage of this approach is the selection of the point of expansion. In the previous approaches (log-likelihood expansions), the point of expansion is unknown and estimated at each sensor. On the contrary, in this approach, the point of expansion is naturally chosen as $\theta = 0$ with the small signal assumption as we are considering the following parameter testing problem in (68). By using a K^{th} order Taylor expansion of the nonlinear signal $s[n;\theta]$ about $\theta = 0$, the signal is approximated as

$$s[n;\theta] \approx s[n;0] + \left.\frac{\partial s[n;\theta]}{\partial \theta}\right|_{\theta=0} \theta + \cdots + \frac{1}{K!} \left.\frac{\partial^K s[n;\theta]}{\partial \theta^K}\right|_{\theta=0} \theta^K$$

for $n = 0, \ldots, N-1$. Thus, it is possible to write the log-likelihood function at the i^{th} sensor as

$$\ln p(\mathbf{x}_i;\theta) \approx c - \frac{\mathcal{E}(\theta)}{2\sigma^2} + \frac{1}{\sigma^2} \sum_{k=0}^{K} t_k(\mathbf{x}_i) \frac{\theta^k}{k!}$$

where c is a constant, $\mathcal{E}(\theta) = \sum_{n=0}^{N-1} s^2[n;\theta]$ and

$$t_k(\mathbf{x}_i) = \sum_{n=0}^{N-1} x_i[n] \left.\frac{\partial^k s[n;\theta]}{\partial \theta^k}\right|_{\theta=0}$$

$k = 0, \ldots, K$ are the approximate locally sufficient statistics (LSS) for the i^{th} sensor. The test statistic using approximate LSS is found as

$$T_{LSS}(\mathbf{x}) = \max_{\theta} \left[\frac{1}{\sigma^2} \sum_{i=1}^{M} \sum_{k=0}^{K} t_k(\mathbf{x}_i) \frac{\theta^k}{k!} - M \frac{\mathcal{E}(\theta)}{2\sigma^2} \right] \quad (77)$$

Thus, the i^{th} sensor needs to submit $\{t_0(\mathbf{x}_i), \ldots, t_K(\mathbf{x}_i)\}$ only and the maximization over θ is done at the central processor.

4.4 Simulation Results

The simulations are run for the nonlinear signal model, as lossless compression is obtained for the linear model resulting in the same detection performance of the GLRT. The following nonlinear signal example is used for simulations in this paper:

$$s[n;\theta] = \theta^{n+1}$$

where $0 < \theta < 1$ for $n = 0, \ldots, N-1$, which is a damped exponential. Note that this signal model satisfies the assumption $\theta = 0 \Leftrightarrow \mathbf{s}(\theta) = \mathbf{0}$. The observed Fisher information at the i^{th} sensor is found as

$$\hat{I}(\hat{\theta}_i) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(n+1)n\hat{\theta}_i^{n-1}\left(\hat{\theta}_i^{n+1} - x_i[n]\right) + (n+1)^2\hat{\theta}_i^{2n}$$

which is used in the approximate GLRT $T_A(\mathbf{x})$. The Fisher information of θ

$$I(\theta) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(n+1)^2\theta^{2n}$$

is used to find the VST, which is

$$\beta = g(\theta) = \frac{1}{\sigma}\int_0^{\theta}\sqrt{\sum_{n=0}^{N-1}(n+1)^2\theta^{2n}}\,d\theta'$$

and is found numerically and plotted in Fig. 25. Note that for the damped exponential signal,

$$\frac{\partial^k s[n;\theta]}{\partial\theta^k} = \frac{(n+1)!}{(n+1-k)!}\theta^{n+1-k}$$

$$\mathcal{E}(\theta) = \frac{\theta^2(1-\theta^{2N})}{1-\theta^2}$$

which are used to find $T_{LSS}(\mathbf{x})$. To obtain receiver operating characteristic (ROC) curves, Monte Carlo simulations are run. $R = 1000$ realizations are run for $M = 10, N = 30, \sigma^2 = 5$ and $\theta = 0.5$. The GLRT detector performance is compared to other lossy detectors in Fig. 26. It is seen that the second order expansion of the log-likelihood function at each sensor (T_A) results in the worst detector

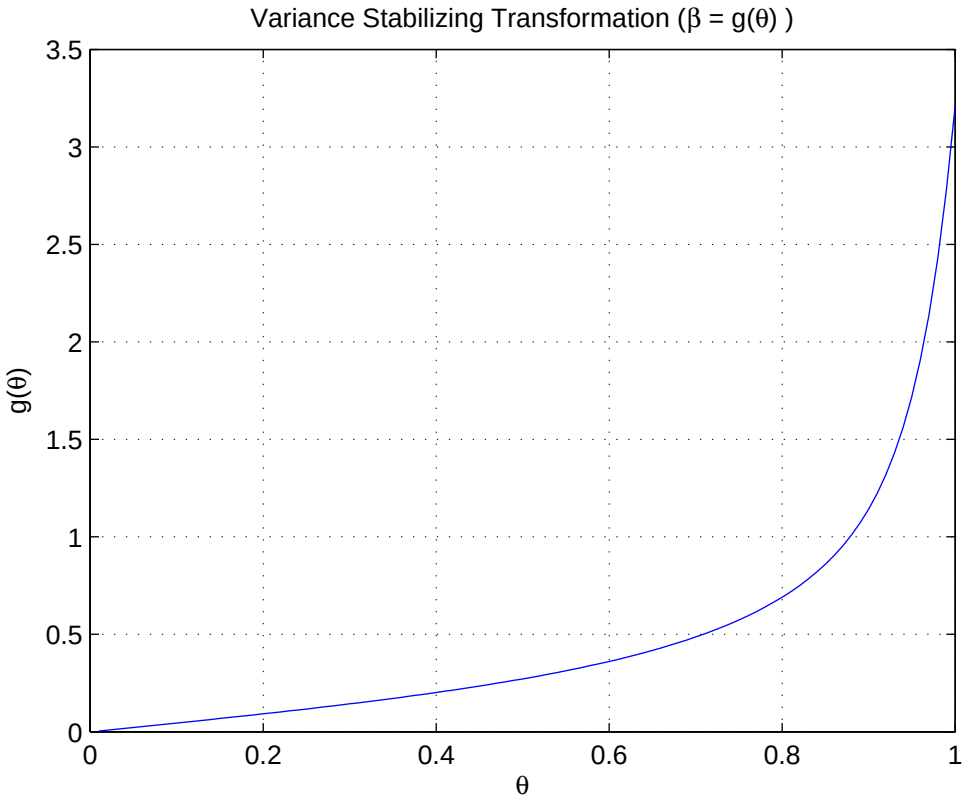


Figure 25: Variance Stabilizing Transformation

performance. Although the VST (T_{VST}) improves the detector performance a little bit, the performance difference is not significant. However, the 3^{rd} order expansion of the nonlinear signal ($T_{LSS}, K = 3$) around $\theta = 0$ is sufficient to obtain a similar performance compared to the asymptotically optimal GLRT detector (T_G). Therefore, the least performance loss caused by data compression is obtained by the expansion of the nonlinear signal in this case. The reason behind this result is due to the low energy to noise ratio at each sensor causing inaccurate MLEs of θ and poor log-likelihood approximations. Fig. 27 shows that when $\theta = 0.9$, for

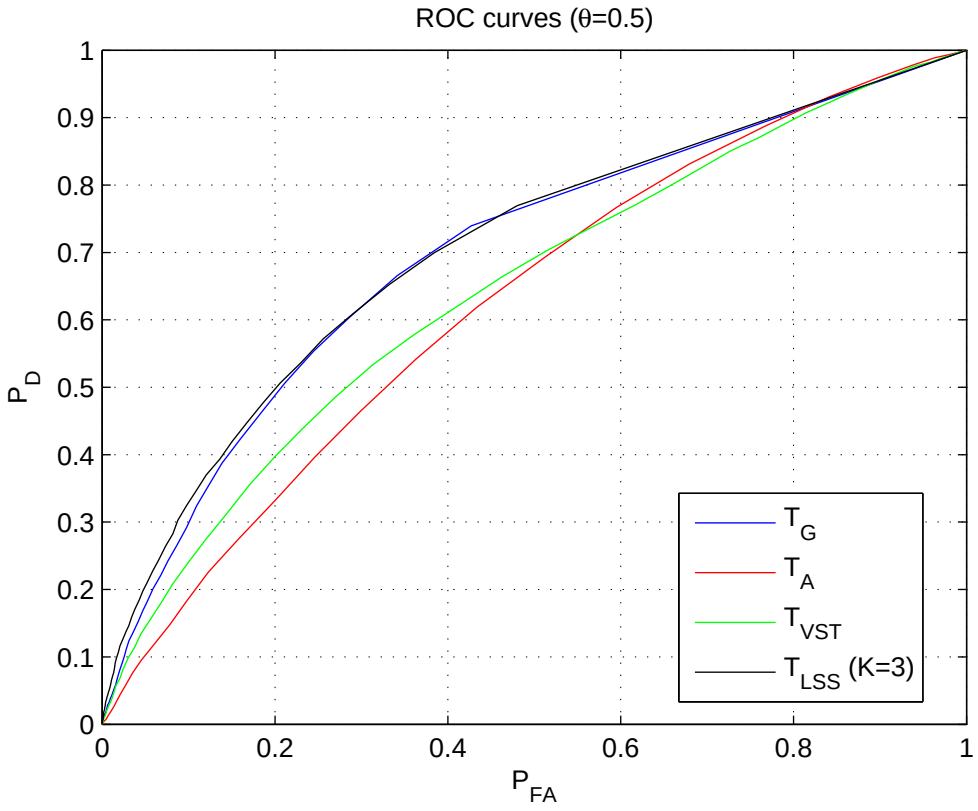


Figure 26: ROC comparison of the proposed detectors

the LSS detector to obtain similar performance as the GLRT detector, K needs to be increased to 9. In other words, more terms are needed to approximate the nonlinear signal accurately since the point of expansion ($\theta = 0$) is distant from the true parameter value ($\theta = 0.9$). Therefore, it is clear that the order of expansion

needed for good performance depends directly on the nonlinear signal. However, note that for the small signal assumption, θ is assumed to be close to 0, thus; K does not need to be very large in practice.

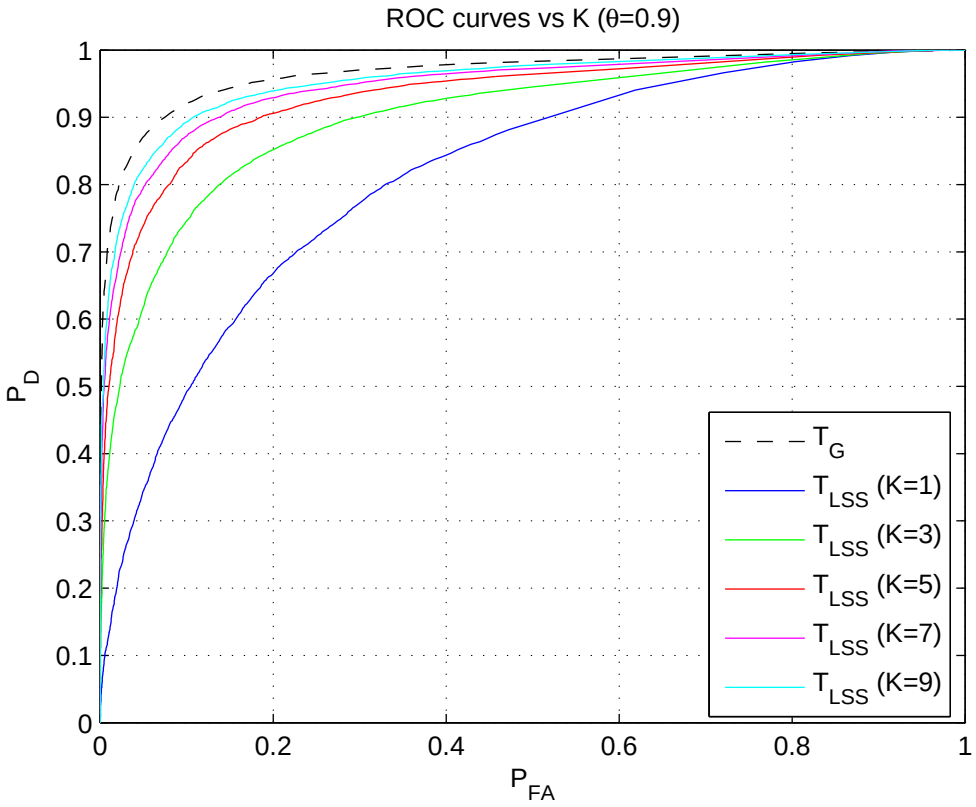


Figure 27: ROC comparison of the GLRT and LSS detectors for $\theta = 0.9$

4.5 Discussion and Conclusions

It is seen that for the linear model, lossless compression of data is possible and the sufficient statistics are the MLE of θ and the Fisher information of θ at each sensor. For the nonlinear model, lossy compression techniques are considered since lossless compression is not possible. The lossy data compression techniques proposed are the second order expansion of the log-likelihood function at each sensor, application of the VST prior to the second order expansion of the log-likelihood functions and the K^{th} order expansion of the nonlinear signal.

For the K^{th} order expansion of the nonlinear signal, we assumed a small signal and chose $\theta = 0$ as the point of expansion. This approach is advantageous over the log-likelihood expansions especially for the low energy to noise ratio case which is simulated in this paper, since there is no need to find the MLEs of θ at each sensor. However, if the energy to noise ratio is high, it may be possible to attain just as good performance as the GLRT by using log-likelihood expansions, thus; submitting less statistics compared to the signal expansion.

Another consideration might be to use a higher order expansion of the log-likelihood function at each sensor. In that case, although the log-likelihood approximation would be better, the computational load at each sensor and central processor would increase drastically. There is need for further analysis to decide if it would be worth going to higher orders or not, considering the ease of the K^{th} order signal expansion and its performance. The extension to an unknown parameter vector and nuisance parameters involved in the detection problem will be considered in future work.

List of References

[1] R. R. Tenney and N. R. Sandell Jr, "Detection with distributed sensors," *IEEE Transactions on Aerospace Electronic Systems*, vol. 17, pp. 501–510, 1981.

[2] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[3] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 1, pp. 24–30, 1987.

[4] G. Ferrari and R. Pagliari, "Decentralized binary detection with noisy communication links," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, no. 4, pp. 1554–1563, 2006.

[5] P. K. Varshney, *Distributed detection and data fusion*. Springer-Verlag, 1997.

[6] A. M. Aziz, "A new adaptive decentralized soft decision combining rule for distributed sensor systems with data fusion," *Information Sciences*, vol. 256, pp. 197–210, 2014.

[7] A. Aziz, "A soft-decision fusion approach for multiple-sensor distributed binary detection systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 3, pp. 2208–2216, 2011.

[8] V. Aalo and R. Viswanathan, "Multilevel quantisation and fusion scheme for the decentralised detection of an unknown signal," in *Radar, Sonar and Navigation, IEE Proceedings*, vol. 141, no. 1. IET, 1994, pp. 37–44.

[9] S. M. Kay, *Fundamentals of statistical signal processing, volume II: Detection theory*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

CHAPTER 5

A Data Compression Technique for Distributed Detection of Nonlinear Signals

by

Fuat Cogun and Steven Kay

Dept. of Electrical, Computer and Biomedical Engineering

University of Rhode Island, Kingston, RI, USA

submitted to IEEE Transactions on Aerospace and Electronic Systems.

Abstract

In this paper, the distributed detection of nonlinear (NL) signals embedded in white Gaussian noise (WGN) is considered. First, we derive the asymptotically optimal generalized likelihood ratio test (GLRT) detector which requires the transmission of all observed data to the central processor (CP). As the transmission of all data to the CP is impractical due to limited communication channel bandwidth and CP memory, we propose a data compression technique based on the second-order Taylor approximation of the local log-likelihood functions. The proposed technique is also applied to the well-known subclass of separable NL signals. Monte Carlo simulations are run to analyze the performance of the proposed detector and compare it to the GLRT detector for the linear frequency modulation (LFM) sweep signal detection problem where all of the parameters are unknown.

5.1 Introduction

The multi-sensor distributed detection problem was first introduced in [1] and it has received great interest since then. It is confronted in a many fields, such as radar signal detection, medical diagnosis, and organisational decision making [2]. A good source of information to the multi-sensor distributed detection problem is [3] which contains a detailed introduction and an extended overview.

In classical centralized detection, observations from all the sensors are transmitted to a central processor (CP) where all the data is used to obtain a decision. Even though a centralized detector achieves optimum performance, it requires communication links having high bandwidth and a CP having large memory. Most of the detection problems are based on large data records, thus, centralized detection is not implemented in practice. Instead, distributed detection methods are considered to deal with the issue [4].

In the distributed detection problem, the aim is to derive the detector having

minimum performance loss relative to the optimal centralized detector. There has been much research on wireless sensor networks (WSN) considering this problem which focuses on two alternative approaches: hard decision [5] [6] [7] [8] and soft decision fusion [9] [10]. In hard decision fusion, each sensor transmits its hard decision (0 or 1) based on local observations. Although this approach has the advantage of requiring a narrow communication bandwidth, the detection performance of these detectors is very poor. In the soft decision fusion approach, each distributed sensor makes a soft decision (a value between 0 and 1) based on local observations in contrast to a hard decision (0 or 1) and transmits it to the CP, usually after quantizing it [11] [12]. This approach achieves slightly better performance than the hard decision fusion approach, however, it still displays degraded performance compared to the optimal centralized detector.

The motivation behind this work is based on the distributed detection of a target emitting an unknown radio-frequency (RF) low probability of intercept (LPI) signal which we addressed in [13]. The derived optimal detector in [13] requires the submission of all local observations to the CP, which is hard to implement as the distributed sensors generate very large data records in practice. Moreover, hard or soft decision fusion is not applicable for military applications since these methods result in an unacceptable poor detection performance. The contribution of this paper is the compression of local observations into local statistics at each sensor and transmitting them to construct the test statistic at the central processor for decision making, instead of transmitting local decisions based on measurements at each node. This is a new approach to the distributed detection problem that has not been applied before to the best of our knowledge. We also concentrate our efforts around nonlinear (NL) signal detection using distributed sensors as the RF LPI signals are designated as NL signals. As a subclass of NL signals, we

handle the specific class of separable NL signals, which is introduced in [14], since the LFM sweep signal, which is used as a base signal in radar signal processing, belongs to the class of separable NL signals.

It should also be noted that this paper extends a portion of our work presented in [15]. In [15], the distributed detection of a common NL signal was considered for a scalar unknown parameter. We assumed that the unknown parameter is the same for all sensors which may not apply to many NL signal detection problems in practice. Furthermore, the detector using the second order expansion of log-likelihood functions (the approximate GLRT detector) was given without derivations. In this paper, we generalize the results to the case when the NL signal is a function of an unknown parameter vector varying from sensor to sensor. Therefore, the detectors derived in this paper apply to all distributed NL signal detection problems. We also provide detailed step-by-step derivations of the detectors.

The outline of this paper is as follows. The next section introduces the general NL signal model that is used in our derivations. Section 5.3 derives the GLRT and Approximate GLRT detectors for the general NL signal model. The class of separable NL signals is introduced and the detectors are found for this special class in Section 5.4. The simulation model, parameters and results are presented in Section 5.5. Section 5.6 ends the paper with conclusions and discussions.

5.2 The Nonlinear Signal Model

We have the following problem:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_i = \mathbf{w}_i \\ \mathcal{H}_1 &: \mathbf{x}_i = s(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) + \mathbf{w}_i \end{aligned} \tag{78}$$

for $i = 1, \ldots, M$. The data length at each sensor is $N \times 1$, i.e, $\mathbf{x}_i, \mathbf{s}, \mathbf{w}_i$ are $N \times 1$ vectors. The $\boldsymbol{\alpha}_i$ depends on sensor while $\boldsymbol{\beta}$ is common to all sensors. $\mathbf{w}_i$'s are

independent and identically distributed (iid) with the distribution $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 is known. If we let

$$\boldsymbol{\theta}_i = \begin{bmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\beta} \end{bmatrix} \tag{79}$$

we have

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_i = \mathbf{w}_i \\ \mathcal{H}_1 &: \mathbf{x}_i = s(\boldsymbol{\theta}_i) + \mathbf{w}_i \end{aligned} \tag{80}$$

for $i = 1, \ldots, M$. Note that $\boldsymbol{\theta}_i$ is the unknown parameter vector at the i^{th} sensor.

5.3 Test Statistics

In this section, first, the asymptotically optimal GLRT [16] is given for the general NL signal detection problem introduced in (80). Second, data compression is considered to obtain the "Approximate GLRT" statistics which are used in efficient distributed detection. The detailed derivations are given in Appendices 5A and 5B.

5.3.1 The GLRT statistic

The GLRT statistic is found as

$$T(\mathbf{x}) = \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\boldsymbol{\theta}_i)\mathbf{x}_i - s^T(\boldsymbol{\theta}_i)s(\boldsymbol{\theta}_i) \right] \right\} \tag{81}$$

where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_M \\ \boldsymbol{\beta} \end{bmatrix}$$

Note that to find the GLRT statistic $T(\mathbf{x})$, all observed data $\mathbf{x}_i$ for $i = 1, \ldots, M$ need to be transmitted to the CP where the maximization takes place.

5.3.2 The Approximate GLRT statistic

The Approximate GLRT statistic is found as follows:

$$T_A(\mathbf{x}) = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\hat{\boldsymbol{\theta}}_i)\mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i)s(\hat{\boldsymbol{\theta}}_i) \right] - \frac{1}{2} \sum_{i=1}^{M} \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right)^T \hat{\mathbf{I}}_i \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right) \quad (82)$$

where $\hat{\boldsymbol{\theta}}_i$ is the local maximum-likelihood estimate (MLE) of $\boldsymbol{\theta}_i$ found by [17]

$$\hat{\boldsymbol{\theta}}_i = \arg_{\boldsymbol{\theta}_i} \max \{ \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i) \} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i \\ \hat{\boldsymbol{\beta}}_i \end{bmatrix}$$

Note that $\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)$ is the local log-likelihood function under $\mathcal{H}_1$ and $\hat{\boldsymbol{\alpha}}_i$ and $\hat{\boldsymbol{\beta}}_i$ are the local MLEs of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$, respectively, based on $\mathbf{x}_i$ only. We define $\hat{\mathbf{I}}_i$ as

$$\hat{\mathbf{I}}_i = \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) - \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \hat{\mathbf{I}}^{-1}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \quad (83)$$

where the second term is the information loss due to having estimated $\hat{\boldsymbol{\alpha}}_i$. The matrices in (83) are partitions of the observed Fisher information matrix (FIM), which is defined by

$$\hat{\mathbf{I}}(\boldsymbol{\theta}_i) = \hat{\mathbf{I}}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) = \begin{bmatrix} \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) & \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \\ \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) & \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \end{bmatrix}$$

Finally, the overall MLE of $\boldsymbol{\beta}$ is found at the central processor as

$$\boldsymbol{\beta}^* = \left[\sum_{i=1}^{M} \hat{\mathbf{I}}_i \right]^{-1} \left[\sum_{i=1}^{M} \hat{\mathbf{I}}_i \hat{\boldsymbol{\beta}}_i \right] \quad (84)$$

To find the Approximate GLRT statistic, the i^{th} node needs to transmit $s^T(\hat{\boldsymbol{\theta}}_i)\mathbf{x}_i$, $s^T(\hat{\boldsymbol{\theta}}_i)s(\hat{\boldsymbol{\theta}}_i)$, $\hat{\boldsymbol{\beta}}_i$ and $\hat{\mathbf{I}}_i$ only, instead of $\mathbf{x}_i$. We call these terms the approximate local sufficient statistics, inspired by [18]. Letting $\dim(\boldsymbol{\beta}) = p$, only two scalars, a $p \times 1$ vector and a $p \times p$ matrix needs to be transmitted to the CP instead of an $N \times 1$ vector, noting that $N >> p$.

5.4 Class of Separable NL signals

Now, let's consider the case when the non-linear signal is separable. That is

$$s(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i \quad (85)$$

for $i = 1, \ldots, M$. In other words, the linear and nonlinear parts of the signal do separate. In this case, we have the following hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_i = \mathbf{w}_i \\ \mathcal{H}_1 &: \mathbf{x}_i = \mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i + \mathbf{w}_i \end{aligned} \tag{86}$$

for $i = 1, \ldots, M$. $\mathbf{x}_i, \mathbf{s}, \mathbf{w}_i$ are $N \times 1$ vectors. $\mathbf{w}_i$'s are independent and identically distributed (iid) with the distribution $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ where σ^2 is known.

5.4.1 The GLRT statistic

The GLRT statistic is derived in Appendix 5C as

$$T(\mathbf{x}) = \max_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{M} \frac{1}{2\sigma^2} \mathbf{y}_i^T(\boldsymbol{\beta}) \mathbf{y}_i(\boldsymbol{\beta}) \right\} \tag{87}$$

where we define

$$\mathbf{y}_i(\boldsymbol{\beta}) = \mathbf{P}(\boldsymbol{\beta})\mathbf{x}_i \tag{88}$$

Note that $\mathbf{P}(\boldsymbol{\beta})$ is the projection matrix found as

$$\mathbf{P}(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta}) \left[\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right]^{-1} \mathbf{H}^T(\boldsymbol{\beta}) \tag{89}$$

Note that to find the GLRT statistic $T(\mathbf{x})$, the CP still requires all observed data $\mathbf{x}_i$ for $i = 1, \ldots, M$ to be transmitted as $\mathbf{y}_i(\boldsymbol{\beta})$ depends on $\mathbf{x}_i$. However, the maximization is done over $\boldsymbol{\beta}$ instead of $\boldsymbol{\theta}$ for this special case. This reduction in the search space is possible since the MLE of $\boldsymbol{\alpha}_i$'s can be found analytically as a function of $\boldsymbol{\beta}$.

5.4.2 The Approximate GLRT statistic

The final form of the Approximate GLRT statistic for a separable NL signal is found in Appendix 5D as:

$$T_A(\mathbf{x}) = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \mathbf{y}_i^T(\hat{\boldsymbol{\beta}}_i) \mathbf{y}_i(\hat{\boldsymbol{\beta}}_i) - \frac{1}{2} \sum_{i=1}^{M} \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right)^T \hat{\mathbf{I}}_i \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right) \tag{90}$$

where

$$\hat{\boldsymbol{\beta}}_i = \arg_{\boldsymbol{\beta}} \max \left\{ \mathbf{y}_i^T(\boldsymbol{\beta}) \mathbf{y}_i(\boldsymbol{\beta}) \right\} \tag{91}$$

$$\hat{\boldsymbol{\alpha}}_i = \left[\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i) \mathbf{H}(\hat{\boldsymbol{\beta}}_i) \right]^{-1} \mathbf{H}^T(\hat{\boldsymbol{\beta}}_i) \mathbf{x}_i \tag{92}$$

Note that $\hat{\mathbf{I}}_i$ and $\boldsymbol{\beta}^*$ are defined in (83) and (84), respectively, and derived in Appendix 5D in detail. For this signal model, the i^{th} node needs to transmit $\mathbf{y}_i^T(\hat{\boldsymbol{\beta}}_i)\mathbf{y}_i(\hat{\boldsymbol{\beta}}_i)$, $\hat{\boldsymbol{\beta}}_i$ and $\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}_i)$ only, instead of $\mathbf{x}_i$. When compared to the Approximate GLRT detector for the general case given in (82), one less local statistic is required to be transmitted.

5.5 Simulations

In our simulations, as the NL signal of interest, we used the classical Linear Frequency Modulation (LFM) sweep signal which is commonly used in radar signal processing since it provides a very high range and velocity resolution [19].

5.5.1 Simulation Model

We define the hypothesis testing problem as follows

$$\begin{aligned} \mathcal{H}_0 &: x_i[n] = w_i[n] \\ \mathcal{H}_1 &: x_i[n] = A_i \cos(2\pi f_0 n + \pi m n^2 + \phi_i) + w_i[n] \end{aligned} \tag{93}$$

for $n = 0, 1, \ldots, N-1$ and $i = 1, \ldots, M$ where noise is WGN, $\mathbf{w}_i \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, and independent from sensor to sensor. Let

$$s[n; \tilde{\boldsymbol{\theta}}_i] = A_i \cos(2\pi f_0 n + \pi m n^2 + \phi_i)$$

where

$$\tilde{\boldsymbol{\theta}}_i = \begin{bmatrix} A_i \\ \phi_i \\ f_0 \\ m \end{bmatrix}$$

is the unknown parameter vector at each sensor. Note that the unknown amplitude and phase differ from sensor to sensor but the starting frequency and sweep rate is the same for all sensors. If we apply the following invertible transformation:

$$\alpha_i^{(1)} = A_i \cos(\phi_i)$$
$$\alpha_i^{(2)} = -A_i \sin(\phi_i)$$

we can represent the NL signal in the transformed parameter space as

$$s[n; \boldsymbol{\theta}_i] = \alpha_i^{(1)} \cos(2\pi f_0 n + \pi m n^2) + \alpha_i^{(2)} \sin(2\pi f_0 n + \pi m n^2) \tag{94}$$

where

$$\boldsymbol{\theta}_i = \begin{bmatrix} \alpha_i^{(1)} \\ \alpha_i^{(2)} \\ f_0 \\ m \end{bmatrix}$$

By letting $\boldsymbol{\alpha}_i = \begin{bmatrix} \alpha_i^{(1)} \\ \alpha_i^{(2)} \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} f_0 \\ m \end{bmatrix}$, we have

$$\boldsymbol{\theta}_i = \begin{bmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\beta} \end{bmatrix} \tag{95}$$

Now, the NL signal can be represented in the following vector form:

$$\mathbf{s}(\boldsymbol{\theta}_i) = \mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i \tag{96}$$

where

$$\mathbf{H}(\boldsymbol{\beta}) = \begin{bmatrix} 1 & 0 \\ \cos(2\pi f_0 + \pi m) & \sin(2\pi f_0 + \pi m) \\ \vdots & \vdots \\ \cos(2\pi f_0(N-1) + \pi m(N-1)^2) & \sin(2\pi f_0(N-1) + \pi m(N-1)^2) \end{bmatrix}$$

Note that this is the exact form given in (85). Thus, the LFM sweep signal model used in this work belongs to the class of separable NL signals as previously stated in the introduction. For the hypothesis testing problem introduced in (93), we can apply the GLRT and the Approximate GLRT detectors derived in (87) and (90), respectively.

5.5.2 Simulation Parameters

Monte Carlo simulations are run for 5000 realizations for the following parameters:

- $N = 200$
- $M = 10$
- $\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_{10} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.38 \\ 0.36 \\ \vdots \\ 0.22 \end{bmatrix}$
- $\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{10} \end{bmatrix} = \begin{bmatrix} 0 \\ -\pi/36 \\ -\pi/18 \\ \vdots \\ -\pi/4 \end{bmatrix}$
- $f_0 = 0.1$
- $m = (0.5 - f_0)/N$

It is implicitly assumed that Sensor #1 has the best signal-to-noise ratio (SNR) among others, observing the maximum amplitude and minimum phase. The SNR decreases gradually as the sensor ID increases numerically. The observed signal at Sensor #1, $s[n; \tilde{\boldsymbol{\theta}}_1]$, is illustrated in Fig. 28.

5.5.3 Simulation Results
Receiver Operating Characteristic Analysis

The receiver operating characteristic (ROC) curves shown in Fig. 29 are obtained for the GLRT detector when $\sigma^2 = 1, 2, 3$ and 5. It is seen that the

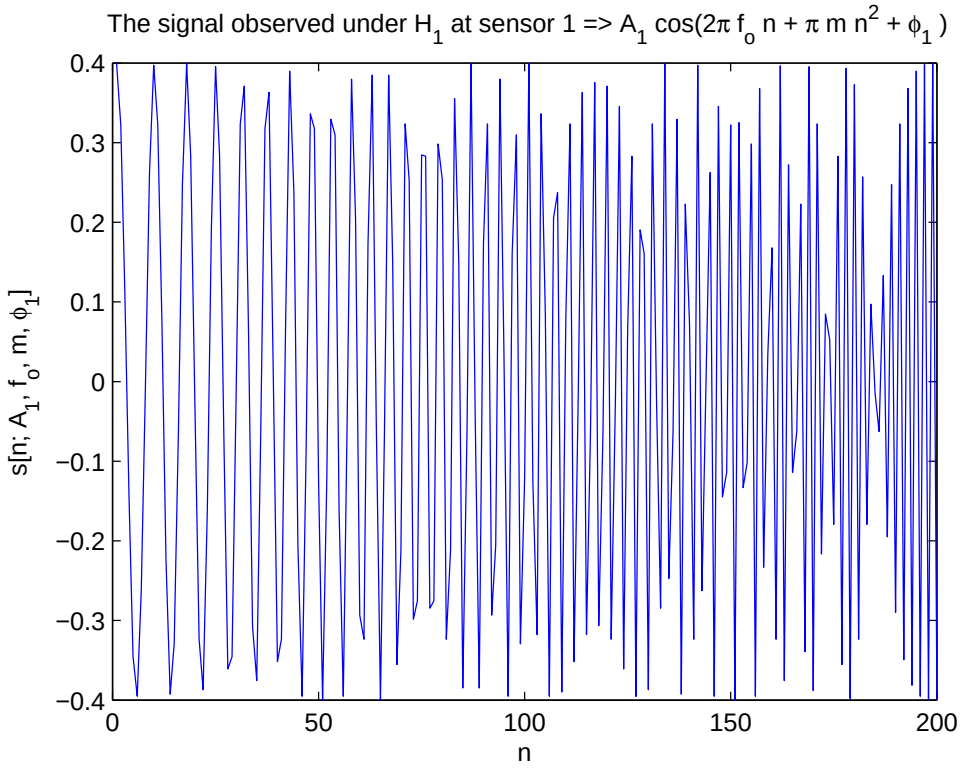


Figure 28: LFM sweep signal at sensor 1 (no noise)

performance of the GLRT detector is quite good even for high noise variances. In Fig. 30, the performance of the Approximate GLRT detector for $\sigma^2 = 0.5, 1, 1.5$ and 3 is shown. As expected, the detection performance of both detectors degrades as σ^2 increases. Note that the performance of the GLRT detector for $\sigma^2 = 2$ is nearly the same as the performance of the Approximate GLRT detector for $\sigma^2 = 1$. Therefore, only about a 3dB performance loss is encountered in using the Approximate GLRT.

A performance comparison of the GLRT and Approximate GLRT detectors is given in Fig. 31 for $\sigma^2 = 1$ and $\sigma^2 = 2$. Note that the ROC curves of the GLRT detector form the upper bound for the Approximate GLRT. It is seen that for $\sigma^2 = 1$, the GLRT detector has perfect detection performance. As σ^2 increases, the performance gap between the detectors increases. In other words, the Approximate

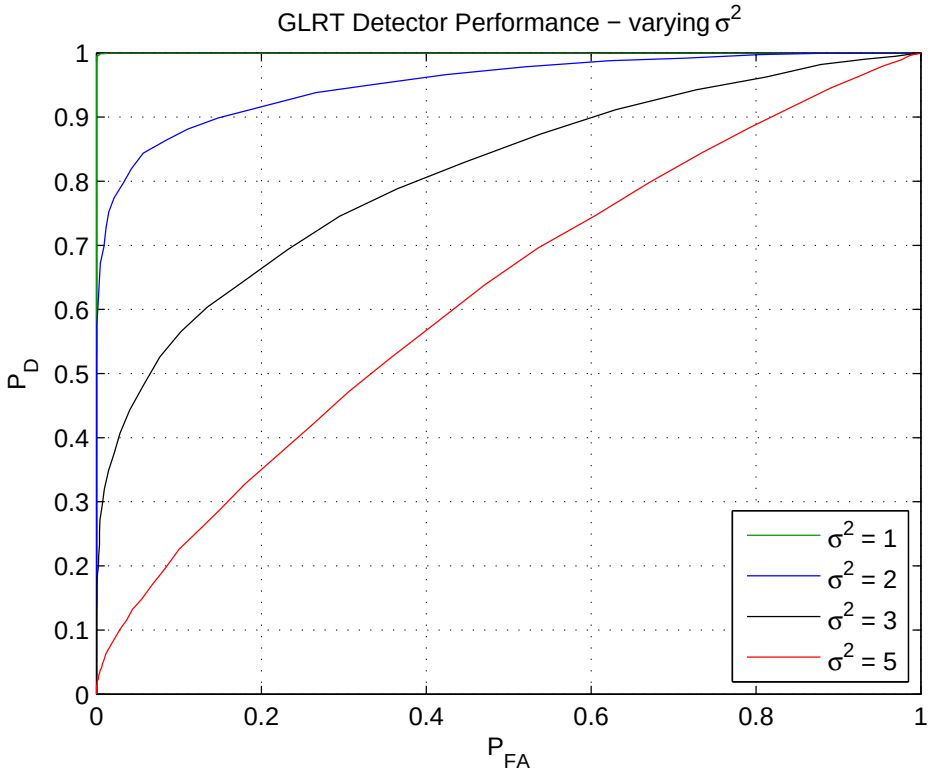


Figure 29: Performance of the GLRT detector

GLRT is more susceptible to noise when compared to the GLRT.

Scatter Plots of the Local MLE

Now, let's consider the local MLEs of $\boldsymbol{\beta}$ at different nodes for fixed σ^2. To see the effect of the way we setup the simulations, consider Sensors #1 and #10. Remember that since we let $A_1 = 0.4$ and $A_{10} = 0.22$, the SNR at Sensor #1 is much (about 5.2dB) higher than the SNR at Sensor #10. For fixed $\sigma^2 = 0.5$, we obtain the scatter plot of the local MLEs ($\hat{\boldsymbol{\beta}}_i$'s) shown in Fig. 32 for $i = 1$ and 10 by running 100 realizations. It is seen that the local MLE obtained at Node #1 is very accurate. 91 out of 100 times, the estimate is equal to the true parameter value and other estimates are close to the true value except 3 of them. However, the local MLE at Node #10 is poor and sparsely distributed due to the low SNR. Only 17 out of 100 times the true parameter value is correctly estimated.

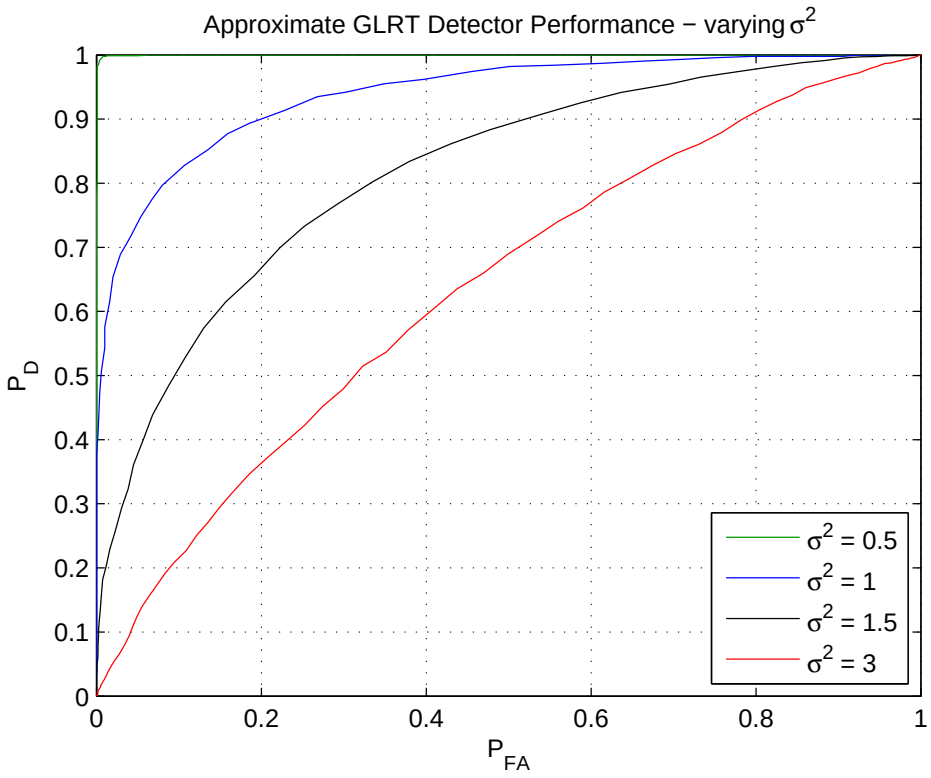


Figure 30: Performance of the Approximate GLRT Detector

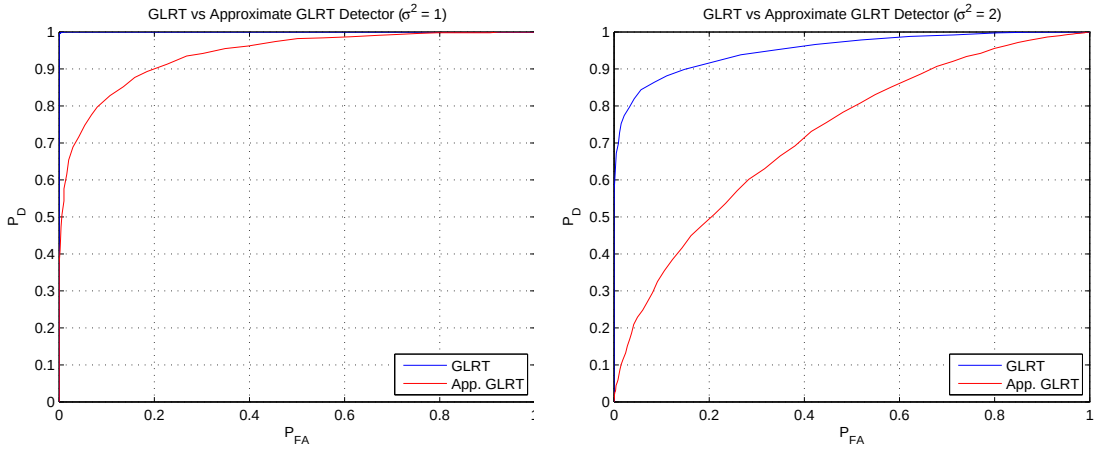


Figure 31: Comparison of detector performances for $\sigma^2 = 1$ and $\sigma^2 = 2$

To see the effect of the noise variance on the local MLE, consider Sensor #1 ($A_1 = 0.4$). We compare the scatter plots of $\hat{\boldsymbol{\beta}}_1$ for the cases $\sigma^2 = 0.5$ and $\sigma^2 = 3$. Again, 100 realizations are run to obtain the scatter plots in Fig. 33. Note that

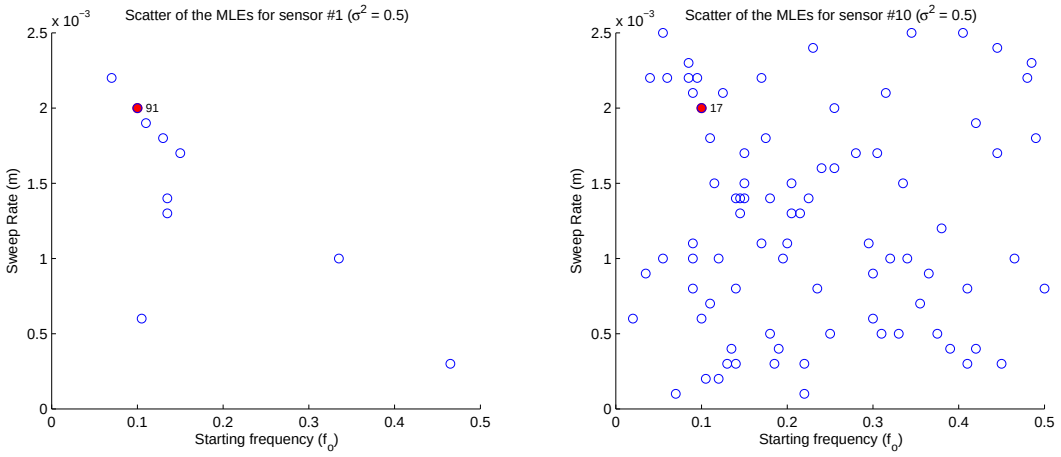


Figure 32: Scatter Plots of the Local MLEs - Sensor 1 compared to Sensor 10

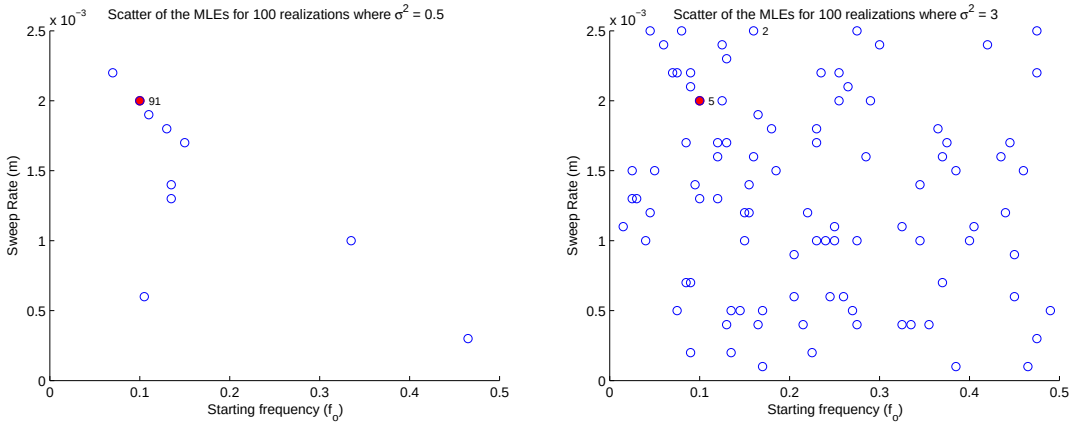


Figure 33: Scatter Plots of the Local MLEs at Sensor 1 for $\sigma^2 = 0.5$ and $\sigma^2 = 3$

the true parameter value is indicated by the red filled circle. It is observed that when the noise variance σ^2 is increased from 0.5 to 3, the local MLE at the sensor become sparsely distributed.

Scatter Plots of the Overall MLE

In this section, we consider plotting the overall MLE ($\boldsymbol{\beta}^*$) of the Approximate GLRT detector for 150 realizations when $\sigma^2 = 1.5$. The scatter of the overall MLE is plotted on top of the ambiguity function contours of the LFM sweep signal in Fig. 34. It is seen that most of the $\boldsymbol{\beta}^*$'s reside within the ambiguity function

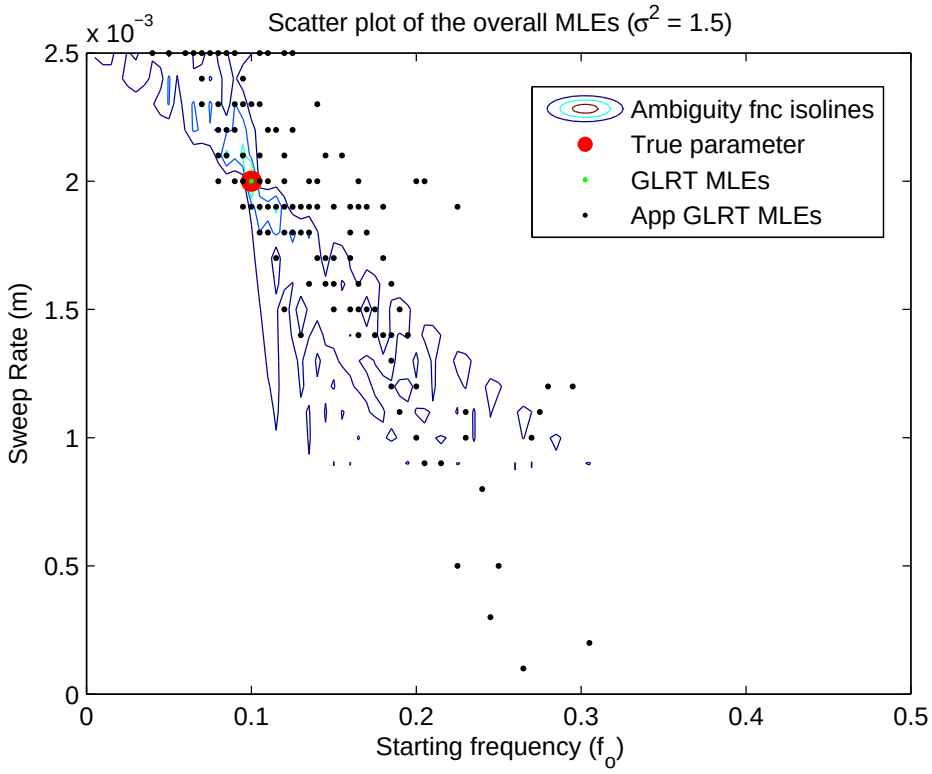


Figure 34: Scatter plot of the overall MLEs

contours and the distribution follows the ambiguity ridge. Also note that, for all realizations, the MLE of the GLRT detector (marked by a green dot) is equal to the true parameter.

Scatter Plots of the Local MLEs and the Overall MLE

Now, we consider a different perspective on viewing the scatter plots. The scatter plots shown in Fig. 35 are obtained by using 1 realization only. This time, the local MLEs of all nodes, i.e, $\{\hat{\boldsymbol{\beta}}_i : i = 1, \ldots, M\}$, are plotted with the resulting overall MLE ($\boldsymbol{\beta}^*$). The relation between the overall MLE and local MLEs is given by (84). It is seen in Fig. 35 that the approximate GLRT detector has good local estimates except for two outliers. Thus, a good overall estimate is obtained when $\sigma^2 = 1$. However, when $\sigma^2 = 3$, the overall estimate becomes very poor as the number of outliers increases. Note that the increase in the number of outliers when

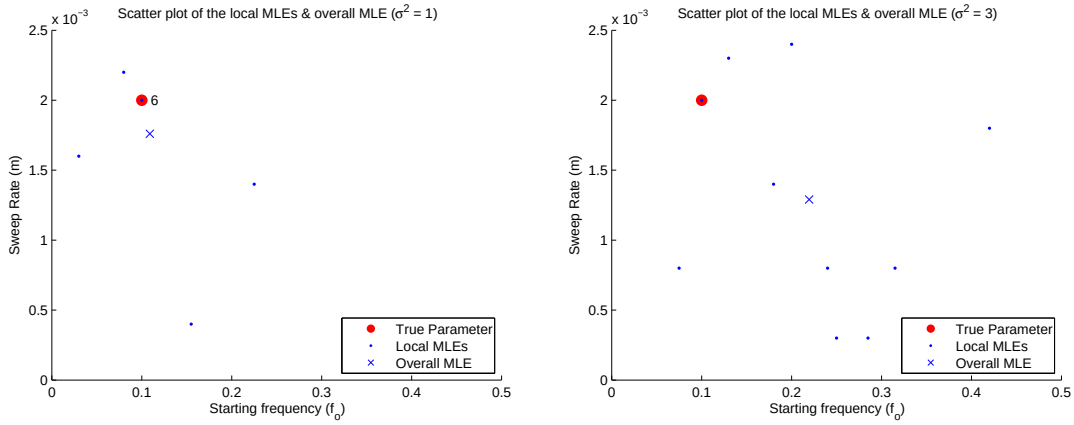


Figure 35: Scatter Plots of the local MLEs and overall MLE for $\sigma^2 = 1$ and $\sigma^2 = 3$

$\sigma^2 = 3$ becomes evident observing different realizations, even though it not shown in here.

ROC curves of the Approximate GLRT - Varying M

In order to obtain the same SNR at each sensor, the simulation parameters are modified as $A_i = 0.4,\ \phi_i = 0$ for $i = 1, \ldots, M$. The other parameters are kept the same. We plot the ROC curves of the Approximate GLRT detector versus M for $\sigma^2 = 1$ and $\sigma^2 = 2$ as shown in Fig. 36. It is seen that as the number of sensors, M, increases, the performance of the detector improves for both σ^2 values as anticipated.

5.6 Conclusions and Discussions

In this paper, the distributed detection of a NL signal which is a function of an unknown parameter vector varying from sensor to sensor is considered. The asymptotically optimal centralized GLRT detector is derived and its performance is used as an upper bound to the proposed Approximate GLRT detector. The idea behind the proposed Approximate GLRT detector can be summarized as:

- The compression of local observations into approximate local sufficient statistics at each node,

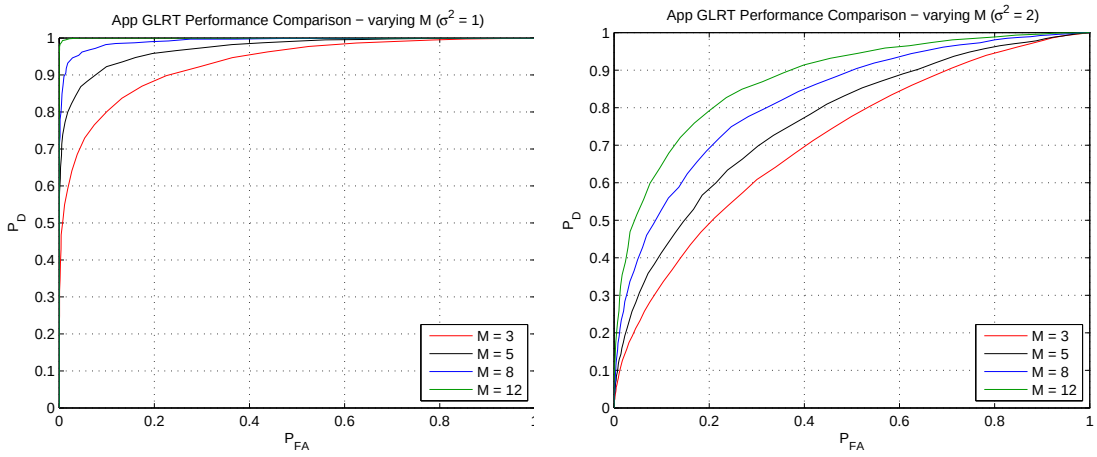


Figure 36: Appr. GLRT ROC comparison vs M for $\sigma^2 = 1$ and $\sigma^2 = 2$

- The transmission of the compressed local statistics to the CP,
- The formation of the Approximate GLRT statistic at the central processor using local statistics.

The approximate local sufficient statistics are found to be the local MLE, observed Fisher information matrix evaluated at the local MLE, and the inner product of the locally estimated NL signal with itself and with the observed data.

The class of separable NL signals is also investigated and both detectors are derived for this special case. The proposed Approximate GLRT detector does not need to transmit the inner product of the locally estimated NL signal with itself, thus, even less local statistics are needed when dealing with a separable NL signal. Simulations are run for the separable NL LFM sweep signal and ROC curves, scatter plots of the local MLEs and the overall MLE are presented for various σ^2 and M values. It is seen that the detection performance and the accuracy of the local MLEs, and thus, the overall MLE of the proposed detector, increases as σ^2 decreases. It is also observed that additional sensors result in an improved detection performance.

Another important consideration is the case when all local MLEs of the unknown parameter vector are the same. In that case, the second term in the Approximate GLRT test statistic disappears and it becomes equivalent to the GLRT statistic. Practically, this only occurs when the SNR at each sensor is very high and all nodes perfectly estimate the unknowns. As a result, the performance of the Approximate GLRT detector attains the performance of the GLRT detector, which is the upper bound. Conversely, it is seen in our simulations that high noise variance (low SNR) at the distributed sensors results in poor local MLEs, thus, degraded performance. Note that the degraded performance is caused by the second order Taylor approximation of the local log-likelihoods becoming invalid as the point of expansion moves further away from the true parameter.

As a final remark, it may also be possible to improve the performance of the Approximate GLRT detector by detecting and discarding the outlier local MLEs and the local statistics associated with them at the CP. But this is out of this paper's scope and is left as future work.

Appendix 5A - Derivation of the GLRT statistic

The GLRT statistic is defined as [16]

$$\begin{aligned} T(\mathbf{x}) &= \max_{\boldsymbol{\theta}} \left\{ \ln \frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\mathcal{H}_0)} \right\} \\ &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \ln p(\mathbf{x}_i;\boldsymbol{\theta}_i) - \ln p(\mathbf{x}_i;\mathcal{H}_0) \right\} \end{aligned} \tag{97}$$

where $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix}$, $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_M \\ \boldsymbol{\beta} \end{bmatrix}$. Under $\mathcal{H}_0$, we have

$$p(\mathbf{x}_i;\mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \mathbf{x}_i^T \mathbf{x}_i \right\}$$

taking the natural logarithm of both sides,

$$\ln p(\mathbf{x}_i;\mathcal{H}_0) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathbf{x}_i^T\mathbf{x}_i \tag{98}$$

Under $\mathcal{H}_1$, the local likelihood function at the i^{th} node is

$$p(\mathbf{x}_i;\boldsymbol{\theta}_i) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right]^T \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right] \right\}$$

and the local log-likelihood function can be written as

$$\ln p(\mathbf{x}_i;\boldsymbol{\theta}_i) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right]^T \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right] \tag{99}$$

Using (98) and (99), we can write the GLRT statistic as

$$\begin{aligned} T(\mathbf{x}) &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} -\frac{1}{2\sigma^2} \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right]^T \left[\mathbf{x}_i - s(\boldsymbol{\theta}_i)\right] + \frac{1}{2\sigma^2}\mathbf{x}_i^T\mathbf{x}_i \right\} \\ &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} -\frac{1}{2\sigma^2} \left[\mathbf{x}_i^T\mathbf{x}_i - 2s^T(\boldsymbol{\theta}_i)\mathbf{x}_i + s^T(\boldsymbol{\theta}_i)s(\boldsymbol{\theta}_i) - \mathbf{x}_i^T\mathbf{x}_i\right] \right\} \\ &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\boldsymbol{\theta}_i)\mathbf{x}_i - s^T(\boldsymbol{\theta}_i)s(\boldsymbol{\theta}_i)\right] \right\} \end{aligned} \tag{100}$$

Appendix 5B - Derivation of the Approximate GLRT statistic

We begin our derivation of the Approximate GLRT statistic by denoting the local maximum-likelihood estimate (MLE) of $\boldsymbol{\theta}_i$ as $\hat{\boldsymbol{\theta}}_i$, which is found by [17]

$$\hat{\boldsymbol{\theta}}_i = \arg_{\boldsymbol{\theta}_i} \max \{\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\} \tag{101}$$

The local log-likelihood function $\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)$ under $\mathcal{H}_1$ is given in (99). $\hat{\boldsymbol{\theta}}_i$ is defined as

$$\hat{\boldsymbol{\theta}}_i = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_i \\ \hat{\boldsymbol{\beta}}_i \end{bmatrix} \tag{102}$$

Note that $\hat{\boldsymbol{\alpha}}_i$ and $\hat{\boldsymbol{\beta}}_i$ are the local MLEs of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$, respectively, based on $\mathbf{x}_i$ only. Want to approximate $\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)$ around $\hat{\boldsymbol{\theta}}_i$ up to the 2nd order, which is

$$\begin{aligned} \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i) \;\approx\; & \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) + (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \nabla \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\Big|_{\hat{\boldsymbol{\theta}}_i} \\ & + \frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{H} \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\Big|_{\hat{\boldsymbol{\theta}}_i} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \end{aligned}$$

for $i = 1, \ldots, M$. But by the definition of MLE

$$\nabla \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\Big|_{\hat{\boldsymbol{\theta}}_i} = \mathbf{0}$$

Thus, the approximation simplifies to:

$$\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i) \approx \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) + \frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{H} \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\Big|_{\hat{\boldsymbol{\theta}}_i} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \tag{103}$$

Defining the observed Fisher information matrix (FIM) evaluated at $\hat{\boldsymbol{\theta}}_i$ as

$$\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}_i) = -\mathbf{H} \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i)\Big|_{\hat{\boldsymbol{\theta}}_i} \tag{104}$$

we can simplify (103) as

$$\ln p(\mathbf{x}_i; \boldsymbol{\theta}_i) \approx \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) - \frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \tag{105}$$

for $i = 1, \ldots, M$. The overall log-likelihood function $\ln p(\mathbf{x}; \boldsymbol{\theta})$ can be approximated by using (105):

$$\begin{aligned} \ln p(\mathbf{x}; \boldsymbol{\theta}) &= \sum_{i=1}^{M} \ln p(\mathbf{x}_i; \boldsymbol{\theta}_i) \\ &\approx \sum_{i=1}^{M} \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) - \frac{1}{2} \sum_{i=1}^{M} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \end{aligned} \tag{106}$$

If we let

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{M} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \tag{107}$$

(106) becomes

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) \approx \sum_{i=1}^{M} \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) - \frac{1}{2} J(\boldsymbol{\theta}) \tag{108}$$

Applying the approximation (108) to the GLRT statistic defined in (97), the approximate GLRT statistic can be written as

$$T_A(\mathbf{x}) = \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \ln \frac{p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i)}{p(\mathbf{x}_i; \mathcal{H}_0)} - \frac{1}{2} J(\boldsymbol{\theta}) \right\} \tag{109}$$

Lets simplify the first term in curly brackets by using (98) and (99):

$$\begin{aligned} \sum_{i=1}^{M} \ln \frac{p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i)}{p(\mathbf{x}_i; \mathcal{H}_0)} &= \sum_{i=1}^{M} \ln p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_i) - \ln p(\mathbf{x}_i; \mathcal{H}_0) \\ &= \sum_{i=1}^{M} -\frac{1}{2\sigma^2} \left[\mathbf{x}_i - s(\hat{\boldsymbol{\theta}}_i) \right]^T \left[\mathbf{x}_i - s(\hat{\boldsymbol{\theta}}_i) \right] + \frac{1}{2\sigma^2} \mathbf{x}_i^T \mathbf{x}_i \\ &= \sum_{i=1}^{M} -\frac{1}{2\sigma^2} \left[\mathbf{x}_i^T \mathbf{x}_i - 2 s^T(\hat{\boldsymbol{\theta}}_i) \mathbf{x}_i + s^T(\hat{\boldsymbol{\theta}}_i) s(\hat{\boldsymbol{\theta}}_i) - \mathbf{x}_i^T \mathbf{x}_i \right] \\ &= \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2 s^T(\hat{\boldsymbol{\theta}}_i) \mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i) s(\hat{\boldsymbol{\theta}}_i) \right] \end{aligned} \tag{110}$$

Replacing (110) into (109)

$$\begin{aligned} T_A(\mathbf{x}) &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2 s^T(\hat{\boldsymbol{\theta}}_i) \mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i) s(\hat{\boldsymbol{\theta}}_i) \right] - \frac{1}{2} J(\boldsymbol{\theta}) \right\} \\ &= \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2 s^T(\hat{\boldsymbol{\theta}}_i) \mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i) s(\hat{\boldsymbol{\theta}}_i) \right] - \frac{1}{2} J(\boldsymbol{\theta}^*) \end{aligned} \tag{111}$$

where $J(\boldsymbol{\theta}^*)$ is found by

$$J(\boldsymbol{\theta}^*) = \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \tag{112}$$

or alternatively,

$$\boldsymbol{\theta}^* = \arg_{\boldsymbol{\theta}} \min J(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\alpha}_1^* \\ \vdots \\ \boldsymbol{\alpha}_M^* \\ \boldsymbol{\beta}^* \end{bmatrix} \tag{113}$$

We would like to find $J(\boldsymbol{\theta}^*)$ analytically and substitute into (111). Replacing (95) and (102) into (107) and applying partitioning to $\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}_i)$:

$$\begin{aligned} J(\boldsymbol{\theta}) &= J(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \boldsymbol{\beta}) \\ &= \sum_{i=1}^{M} \begin{bmatrix} \boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i \end{bmatrix}^T \begin{bmatrix} \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) & \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \\ \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) & \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i \end{bmatrix} \\ &= \sum_{i=1}^{M} (\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i) \\ &\quad + 2(\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ &\quad + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \end{aligned} \tag{114}$$

Note that for $k \in \{1, 2, \ldots, M\}$

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k) = 2\hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k) \\ &\frac{\partial}{\partial \boldsymbol{\alpha}_k}(\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) = \hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) \\ &\frac{\partial}{\partial \boldsymbol{\alpha}_k}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) = \mathbf{0} \end{aligned}$$

Now we can take the partial derivative of (114) with respect to $\boldsymbol{\alpha}_k$, which is found as:

$$\frac{\partial J}{\partial \boldsymbol{\alpha}_k} = 2\hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k) + 2\hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) \tag{115}$$

To find the $\boldsymbol{\alpha}_k^*$'s minimizing J for $k = 1, \ldots, M$:

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\alpha}_k}\Big|_{\boldsymbol{\alpha}_k^*} &= \mathbf{0} \\ &= \hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\alpha}_k^* - \hat{\boldsymbol{\alpha}}_k) + \hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) \end{aligned} \tag{116}$$

From (116), it is found that

$$\boldsymbol{\alpha}_k^* - \hat{\boldsymbol{\alpha}}_k = -\hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k}^{-1}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_k\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_k) \tag{117}$$

for $k = 1, \ldots, M$. It follows from (114) that:

$$\begin{aligned} J(\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_M^*, \boldsymbol{\beta}) &= \sum_{i=1}^{M} (\boldsymbol{\alpha}_i^* - \hat{\boldsymbol{\alpha}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\alpha}_i^* - \hat{\boldsymbol{\alpha}}_i) \\ &\quad + 2(\boldsymbol{\alpha}_i^* - \hat{\boldsymbol{\alpha}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ &\quad + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \end{aligned} \tag{118}$$

Replacing (117) into (118)

$$\begin{aligned} & J(\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_M^*, \boldsymbol{\beta}) \\ = & \sum_{i=1}^{M} \left[-\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \right]^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \\ & \qquad \cdot \left[-\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \right] \\ & + 2 \left[-\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \right]^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ & + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ = & \sum_{i=1}^{M} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ & - 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ & + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ = & \sum_{i=1}^{M} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ & - (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \\ = & \sum_{i=1}^{M} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \left[\hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) - \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \right] \\ & \qquad \cdot (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \end{aligned}$$

For the tractability of further mathematical operations, let

$$\hat{\mathbf{I}}_i = \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) - \hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\alpha}_i}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}^{-1}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) \tag{119}$$

We get

$$J(\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_M^*, \boldsymbol{\beta}) = \sum_{i=1}^{M} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_i (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_i) \tag{120}$$

(120) is quadratic in $\boldsymbol{\beta}$. Thus, it is possible to find $\boldsymbol{\beta}^*$ by:

$$\begin{aligned} \frac{\partial J(\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_M^*, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}^*} &= \mathbf{0} \\ &= \sum_{i=1}^{M} 2\hat{\mathbf{I}}_i (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_i) \end{aligned} \tag{121}$$

Equivalently,

$$\left[\sum_{i=1}^{M} \hat{\mathbf{I}}_i\right] \boldsymbol{\beta}^* = \sum_{i=1}^{M} \hat{\mathbf{I}}_i \hat{\boldsymbol{\beta}}_i$$

Therefore,

$$\boldsymbol{\beta}^* = \left[\sum_{i=1}^{M} \hat{\mathbf{I}}_i\right]^{-1} \left[\sum_{i=1}^{M} \hat{\mathbf{I}}_i \hat{\boldsymbol{\beta}}_i\right] \tag{122}$$

Replacing $\boldsymbol{\beta}^*$ into (120):

$$\begin{aligned} J(\boldsymbol{\alpha}_1^*, \ldots, \boldsymbol{\alpha}_M^*, \boldsymbol{\beta}^*) &= \sum_{i=1}^{M} (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_i)^T \hat{\mathbf{I}}_i (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_i) \\ &= J(\boldsymbol{\theta}^*) \end{aligned} \tag{123}$$

Replacing (123) into (111), the final form of the Approximate GLRT statistic is found as follows:

$$T_A(\mathbf{x}) = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\hat{\boldsymbol{\theta}}_i)\mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i)s(\hat{\boldsymbol{\theta}}_i)\right] - \frac{1}{2} \sum_{i=1}^{M} \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^*\right)^T \hat{\mathbf{I}}_i \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^*\right) \tag{124}$$

where $\hat{\mathbf{I}}_i$ is defined in (119).

Appendix 5C - Derivation of the GLRT statistic for a Separable NL Signal

Replacing (85) into (100)

$$T(\mathbf{x}) = \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2\boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{x}_i - \boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i \right] \right\} \tag{125}$$

Note that the argument is a quadratic function of $\boldsymbol{\alpha}_i$'s. Therefore, we can find the maximizing $\boldsymbol{\alpha}_i$'s analytically. Let

$$J(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) = \frac{1}{2\sigma^2} \left[\boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i - 2\boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{x}_i \right] \tag{126}$$

for $i = 1, \ldots, M$. Thus,

$$\begin{aligned} T(\mathbf{x}) &= \max_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} -J(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \right\} = \min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{M} J(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \right\} \\ &= \min_{\boldsymbol{\beta}} \left\{ \min_{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M} \sum_{i=1}^{M} J(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \right\} \\ &= \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{M} \min_{\boldsymbol{\alpha}_i} J(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) \right\} \\ &= \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{M} J(\hat{\boldsymbol{\alpha}}_i, \boldsymbol{\beta}) \right\} \end{aligned} \tag{127}$$

since $\hat{\boldsymbol{\alpha}}_i = \arg_{\boldsymbol{\alpha}_i} \min J(\boldsymbol{\alpha}_i, \boldsymbol{\beta})$ and $\hat{\boldsymbol{\alpha}}_i$ is found by:

$$\begin{aligned} \frac{\partial J(\boldsymbol{\alpha}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}_i} \Big|_{\hat{\boldsymbol{\alpha}}_i} &= \mathbf{0} \\ &= \frac{1}{2\sigma^2} \left[2\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \hat{\boldsymbol{\alpha}}_i - 2\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{x}_i \right] \end{aligned} \tag{128}$$

Equivalently,

$$\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \hat{\boldsymbol{\alpha}}_i = \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{x}_i$$

Therefore,

$$\hat{\boldsymbol{\alpha}}_i = \left[\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right]^{-1} \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{x}_i \tag{129}$$

Now, we can evaluate

$$\begin{aligned} J(\hat{\boldsymbol{\alpha}}_i, \boldsymbol{\beta}) &= \frac{1}{2\sigma^2}\left[\hat{\boldsymbol{\alpha}}_i^T \mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\hat{\boldsymbol{\alpha}}_i - 2\hat{\boldsymbol{\alpha}}_i^T \mathbf{H}^T(\boldsymbol{\beta})\mathbf{x}_i\right] \\ &= \frac{1}{2\sigma^2}\left[\mathbf{x}_i^T \mathbf{H}(\boldsymbol{\beta})\left[\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right]^{-1}\mathbf{H}^T(\boldsymbol{\beta})\mathbf{x}_i\right] \\ &- \frac{1}{2\sigma^2}\left[2\mathbf{x}_i^T \mathbf{H}(\boldsymbol{\beta})\left[\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right]^{-1}\mathbf{H}^T(\boldsymbol{\beta})\mathbf{x}_i\right] \\ &= -\frac{1}{2\sigma^2}\mathbf{x}_i^T \mathbf{H}(\boldsymbol{\beta})\left[\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right]^{-1}\mathbf{H}^T(\boldsymbol{\beta})\mathbf{x}_i \end{aligned} \tag{130}$$

We define the projection matrix $\mathbf{P}(\boldsymbol{\beta})$ as

$$\mathbf{P}(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\left[\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right]^{-1}\mathbf{H}^T(\boldsymbol{\beta}) \tag{131}$$

Then, (130) can be written as

$$\begin{aligned} J(\hat{\boldsymbol{\alpha}}_i, \boldsymbol{\beta}) &= -\frac{1}{2\sigma^2}\mathbf{x}_i^T \mathbf{P}(\boldsymbol{\beta})\mathbf{x}_i \\ &= -\frac{1}{2\sigma^2}\mathbf{x}_i^T \mathbf{P}(\boldsymbol{\beta})\mathbf{P}(\boldsymbol{\beta})\mathbf{x}_i \\ &= -\frac{1}{2\sigma^2}\mathbf{x}_i^T \mathbf{P}^T(\boldsymbol{\beta})\mathbf{P}(\boldsymbol{\beta})\mathbf{x}_i \\ &= -\frac{1}{2\sigma^2}\mathbf{y}_i^T(\boldsymbol{\beta})\mathbf{y}_i(\boldsymbol{\beta}) \end{aligned} \tag{132}$$

where we define

$$\mathbf{y}_i(\boldsymbol{\beta}) = \mathbf{P}(\boldsymbol{\beta})\mathbf{x}_i \tag{133}$$

Note that we used the following properties of the projection matrix:

$$\mathbf{P}(\boldsymbol{\beta}) = \mathbf{P}^T(\boldsymbol{\beta})$$

$$\mathbf{P}(\boldsymbol{\beta})\mathbf{P}(\boldsymbol{\beta}) = \mathbf{P}(\boldsymbol{\beta})$$

Substituting (132) into (127)

$$\begin{aligned} T(\mathbf{x}) &= \min_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{M} -\frac{1}{2\sigma^2}\mathbf{y}_i^T(\boldsymbol{\beta})\mathbf{y}_i(\boldsymbol{\beta})\right\} \\ &= \max_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{M} \frac{1}{2\sigma^2}\mathbf{y}_i^T(\boldsymbol{\beta})\mathbf{y}_i(\boldsymbol{\beta})\right\} \end{aligned} \tag{134}$$

Appendix 5D - Derivation of the Approximate GLRT statistic for a Separable NL Signal

We can write

$$s(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) = \mathbf{H}(\hat{\boldsymbol{\beta}}_i)\hat{\boldsymbol{\alpha}}_i \tag{135}$$

for $i = 1, \ldots, M$. If we plug (135) back into the first term of (124), we get

$$\begin{aligned}
&\sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\hat{\boldsymbol{\theta}}_i)\mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i)s(\hat{\boldsymbol{\theta}}_i)\right] \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\mathbf{x}_i - s^T(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)s(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i)\right] \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2\hat{\boldsymbol{\alpha}}_i^T\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i - \hat{\boldsymbol{\alpha}}_i^T\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{H}(\hat{\boldsymbol{\beta}}_i)\hat{\boldsymbol{\alpha}}_i\right] \qquad (136)
\end{aligned}$$

where $\hat{\boldsymbol{\alpha}}_i$ is given in (129). Note that

$$\begin{aligned}
\mathbf{H}(\hat{\boldsymbol{\beta}}_i)\hat{\boldsymbol{\alpha}}_i &= \mathbf{H}(\hat{\boldsymbol{\beta}}_i)\left[\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{H}(\hat{\boldsymbol{\beta}}_i)\right]^{-1}\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i \\
&= \mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i \qquad (137) \\
&= \mathbf{y}_i(\hat{\boldsymbol{\beta}}_i)
\end{aligned}$$

Substituting (137) into (136)

$$\begin{aligned}
&\sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2s^T(\hat{\boldsymbol{\theta}}_i)\mathbf{x}_i - s^T(\hat{\boldsymbol{\theta}}_i)s(\hat{\boldsymbol{\theta}}_i)\right] \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2\mathbf{x}_i^T\mathbf{P}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i - \mathbf{x}_i^T\mathbf{P}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i\right] \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \left[2\mathbf{x}_i^T\mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i - \mathbf{x}_i^T\mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i\right] \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2}\mathbf{x}_i^T\mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2}\mathbf{x}_i^T\mathbf{P}^T(\hat{\boldsymbol{\beta}}_i)\mathbf{P}(\hat{\boldsymbol{\beta}}_i)\mathbf{x}_i \\
&\qquad = \sum_{i=1}^{M} \frac{1}{2\sigma^2}\mathbf{y}_i^T(\hat{\boldsymbol{\beta}}_i)\mathbf{y}_i(\hat{\boldsymbol{\beta}}_i) \qquad (138)
\end{aligned}$$

To evaluate $\mathbf{I}_i$ defined in (119), we need to find:

- $\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i}$
- $\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}}$
- $\hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}}$

which are defined as

$$\begin{aligned}
\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i} &= -\frac{\partial}{\partial\boldsymbol{\alpha}_i}\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \\
\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}} &= -\frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \\
\hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -\frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial\boldsymbol{\beta}^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta})
\end{aligned}$$

for the i^{th} sensor, where the local log-likelihood is

$$\begin{aligned}
\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta}) &= c - J(\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \\
&= c - \frac{1}{2\sigma^2}\left[\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i - 2\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i\right] \qquad (139)
\end{aligned}$$

Note that c is a constant not depending on $\boldsymbol{\alpha}_i$ or $\boldsymbol{\beta}$. First, let's derive

$$\begin{aligned}
\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i} &= -\frac{\partial}{\partial\boldsymbol{\alpha}_i}\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \\
&= \frac{1}{2\sigma^2}\frac{\partial}{\partial\boldsymbol{\alpha}_i}\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\left\{\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i - 2\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\boldsymbol{\alpha}_i\right\} \\
&= \frac{1}{\sigma^2}\frac{\partial}{\partial\boldsymbol{\alpha}_i}\left\{\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) - \mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\right\} \\
&= \frac{1}{\sigma^2}\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) \qquad (140)
\end{aligned}$$

Second,

$$
\begin{aligned}
\hat{\mathbf{I}}_{\boldsymbol{\alpha}_i\boldsymbol{\beta}} &= -\frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta}) \\
&= \frac{\partial}{\partial\boldsymbol{\beta}}\left\{-\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\ln p(\mathbf{x}_i;\boldsymbol{\alpha}_i,\boldsymbol{\beta})\right\} \\
&= \frac{\partial}{\partial\boldsymbol{\beta}}\left\{\frac{1}{\sigma^2}\left[\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})-\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\right]\right\} \\
&= \frac{1}{\sigma^2}\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})-\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\right\} \\
&= \frac{1}{\sigma^2}\left[\underbrace{\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\boldsymbol{\alpha}_i^T\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right\}}_{\text{Appendix 5E-1}}-\underbrace{\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\right\}}_{\text{Appendix 5E-2}}\right] \\
&= \frac{1}{\sigma^2}\begin{bmatrix}\boldsymbol{\alpha}_i^T & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}_i^T\end{bmatrix}\begin{bmatrix}\mathbf{F}^{(1)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})+\mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(1)}(\boldsymbol{\beta})\\ \vdots \\ \mathbf{F}^{(p)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})+\mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(p)}(\boldsymbol{\beta})\end{bmatrix} \\
&- \frac{1}{\sigma^2}\begin{bmatrix}\mathbf{x}_i^T & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{x}_i^T\end{bmatrix}\begin{bmatrix}\mathbf{G}^{(1)}(\boldsymbol{\beta})\\ \vdots \\ \mathbf{G}^{(p)}(\boldsymbol{\beta})\end{bmatrix} \qquad (141)
\end{aligned}
$$

where

$$\mathbf{F}^{(k)}(\boldsymbol{\beta}) = \frac{\partial}{\partial\beta_k}\mathbf{H}^T(\boldsymbol{\beta}) \qquad (142)$$

$$\mathbf{G}^{(k)}(\boldsymbol{\beta}) = \frac{\partial}{\partial\beta_k}\mathbf{H}(\boldsymbol{\beta}) \qquad (143)$$

for $k = 1, \ldots, p$. Finally,

$$\begin{aligned}
\hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \ln p(\mathbf{x}_i; \boldsymbol{\alpha}_i, \boldsymbol{\beta}) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ -\ln p(\mathbf{x}_i; \boldsymbol{\alpha}_i, \boldsymbol{\beta}) \right\} \\
&= \frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i - 2\mathbf{x}_i^T \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i \right\} \\
&= \frac{1}{2\sigma^2} \left[\underbrace{\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \boldsymbol{\alpha}_i^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i \right\}}_{\text{Appendix 5E-3}} - 2 \underbrace{\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \mathbf{x}_i^T \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}_i \right\}}_{\text{Appendix 5E-4}} \right] \\
&= \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\alpha}_i^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}_i^T \end{bmatrix} \begin{bmatrix} \mathbf{D}^{(1,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{D}^{(1,p)}(\boldsymbol{\beta}) \\ \vdots & \ddots & \vdots \\ \mathbf{D}^{(p,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{D}^{(p,p)}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}_i \end{bmatrix} \\
&- \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{x}_i^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}_i^T \end{bmatrix} \begin{bmatrix} \mathbf{E}^{(1,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{E}^{(1,p)}(\boldsymbol{\beta}) \\ \vdots & \ddots & \vdots \\ \mathbf{E}^{(p,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{E}^{(p,p)}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}_i \end{bmatrix} \qquad (144)
\end{aligned}$$

where

$$\mathbf{D}^{(l,k)}(\boldsymbol{\beta}) = \mathbf{F}^{(l)}(\boldsymbol{\beta}) \mathbf{G}^{(k)}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \qquad (145)$$

and

$$\mathbf{E}^{(l,k)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \beta_l} \mathbf{G}^{(k)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \beta_l} \frac{\partial}{\partial \beta_k} \mathbf{H}(\boldsymbol{\beta}) \qquad (146)$$

for $l = 1, \ldots, p$ and $k = 1, \ldots, p$. Using (140), (141) and (144), we can compute

$$\hat{\mathbf{I}}_i = \left(\hat{\mathbf{I}}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i \boldsymbol{\beta}}^T \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i}^{-1} \hat{\mathbf{I}}_{\boldsymbol{\alpha}_i \boldsymbol{\beta}} \right) \Big|_{\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i} \qquad (147)$$

where

$$\hat{\boldsymbol{\beta}}_i = \arg_{\boldsymbol{\beta}} \max \left\{ \mathbf{y}_i^T(\boldsymbol{\beta}) \mathbf{y}_i(\boldsymbol{\beta}) \right\} \qquad (148)$$

$$\hat{\boldsymbol{\alpha}}_i = \left[\mathbf{H}^T(\hat{\boldsymbol{\beta}}_i) \mathbf{H}(\hat{\boldsymbol{\beta}}_i) \right]^{-1} \mathbf{H}^T(\hat{\boldsymbol{\beta}}_i) \mathbf{x}_i \qquad (149)$$

Substituting (138) and (147) into (124), the final form of the Approximate GLRT statistic for a separable NL signal is found as:

$$T_A(\mathbf{x}) = \sum_{i=1}^{M} \frac{1}{2\sigma^2} \mathbf{y}_i^T(\hat{\boldsymbol{\beta}}_i) \mathbf{y}_i(\hat{\boldsymbol{\beta}}_i) - \frac{1}{2} \sum_{i=1}^{M} \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right)^T \hat{\mathbf{I}}_i \left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \right) \qquad (150)$$

Appendix 5E - Derivation of the Matrix Derivatives

For convenience, we drop sensor indices (i's) in the following derivations. We let

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix}_{q\times 1}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix}_{p\times 1}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}_{N\times 1}$$

Thus, $\mathbf{H}(\boldsymbol{\beta})$ is $N \times q$.

Appendix 5E-1. Derivation of $\frac{\partial}{\partial \boldsymbol{\beta}} \left\{ \boldsymbol{\alpha}^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right\}$

Let

$$\begin{aligned} \mathbf{f}(\boldsymbol{\beta}) &= \boldsymbol{\alpha}^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \\ &= \begin{bmatrix} \alpha_1 & \dots & \alpha_q \end{bmatrix}_{1\times q} \left[\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right]_{q\times q} \\ &= \begin{bmatrix} \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{i1} & \dots & \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{iq} \end{bmatrix} \end{aligned}$$

Now lets take the derivative with respect to β_k for $k = 1, \dots, p$

$$\begin{aligned} \frac{\partial}{\partial \beta_k} \mathbf{f}(\boldsymbol{\beta}) &= \frac{\partial}{\partial \beta_k} \begin{bmatrix} \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{i1} & \dots & \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{iq} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \beta_k} \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{i1} & \dots & \frac{\partial}{\partial \beta_k} \sum_{i=1}^q \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{iq} \end{bmatrix} \\ &= \begin{bmatrix} \underbrace{\sum_{i=1}^q \alpha_i \frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{i1}}_{:= d_1^{(k)}(\boldsymbol{\beta})} & \dots & \underbrace{\sum_{i=1}^q \alpha_i \frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{iq}}_{:= d_q^{(k)}(\boldsymbol{\beta})} \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
d_1^{(k)}(\boldsymbol{\beta}) &= \sum_{i=1}^{q} \alpha_i \frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right)_{i1} \\
&= \sum_{i=1}^{q} \alpha_i \frac{\partial}{\partial \beta_k} \left\{ \sum_{m=1}^{q} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{m1} \right\} \\
&= \sum_{i=1}^{q} \alpha_i \sum_{m=1}^{q} \left\{ \underbrace{\frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im}}_{:=\left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\right)_{im}} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{m1} + \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \underbrace{\frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{m1}}_{:=\left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{m1}} \right\} \\
&= \sum_{i=1}^{q} \alpha_i \sum_{m=1}^{q} \left\{ \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\right)_{im} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{m1} + \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{m1} \right\} \\
&= \sum_{i=1}^{q} \alpha_i \left\{ \sum_{m=1}^{q} \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\right)_{im} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{m1} + \sum_{m=1}^{q} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{m1} \right\} \\
&= \sum_{i=1}^{q} \alpha_i \left\{ \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right)_{i1} + \left(\mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{i1} \right\} \\
&= \sum_{i=1}^{q} \alpha_i \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{i1}
\end{aligned}
$$

Similarly,

$$
d_s^{(k)}(\boldsymbol{\beta}) = \sum_{i=1}^{q} \alpha_i \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{is}
$$

for $s = 2, \ldots, q$. Thus,

$$
\begin{aligned}
\frac{\partial}{\partial \beta_k}\mathbf{f}(\boldsymbol{\beta}) &= \left[\sum_{i=1}^{q} \alpha_i \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{i1} \ \ldots \right. \\
&\qquad \left. \sum_{i=1}^{q} \alpha_i \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{iq} \right] \\
&= \left[\begin{array}{ccc} \alpha_1 & \ldots & \alpha_q \end{array} \right] \left[\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right] \\
&= \boldsymbol{\alpha}^T \left(\mathbf{F}^{(k)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(k)}(\boldsymbol{\beta})\right) \qquad (151)
\end{aligned}
$$

for $k = 1, \ldots, p$.

Finally,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}) &= \begin{bmatrix} \frac{\partial}{\partial \beta_1}\mathbf{f}(\boldsymbol{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p}\mathbf{f}(\boldsymbol{\beta}) \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\alpha}^T\left(\mathbf{F}^{(1)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(1)}(\boldsymbol{\beta})\right) \\ \vdots \\ \boldsymbol{\alpha}^T\left(\mathbf{F}^{(p)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(p)}(\boldsymbol{\beta})\right) \end{bmatrix}_{p\times q} \\
&= \begin{bmatrix} \boldsymbol{\alpha}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^T \end{bmatrix}_{p\times pq} \begin{bmatrix} \mathbf{F}^{(1)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(1)}(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{F}^{(p)}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta})\mathbf{G}^{(p)}(\boldsymbol{\beta}) \end{bmatrix}_{pq\times q} \qquad (152)
\end{aligned}$$

Appendix 5E-2. Derivation of $\frac{\partial}{\partial \boldsymbol{\beta}}\left\{\mathbf{x}_i^T\mathbf{H}(\boldsymbol{\beta})\right\}$

Let

$$\begin{aligned}
\mathbf{f}(\boldsymbol{\beta}) &= \mathbf{x}^T\mathbf{H}(\boldsymbol{\beta}) \\
&= \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix}_{1\times N}\left[\mathbf{H}(\boldsymbol{\beta})\right]_{N\times q} \\
&= \begin{bmatrix} \sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{n1} & \dots & \sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{nq} \end{bmatrix}
\end{aligned}$$

Now, lets take the derivative of $\mathbf{f}(\boldsymbol{\beta})$ with respect to β_k for $k = 1, \dots, p$

$$\begin{aligned}
\frac{\partial}{\partial \beta_k}\mathbf{f}(\boldsymbol{\beta}) &= \frac{\partial}{\partial \beta_k}\begin{bmatrix} \sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{n1} & \dots & \sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{nq} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_k}\sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{n1} & \dots & \frac{\partial}{\partial \beta_k}\sum_{n=1}^N x_n\left(\mathbf{H}(\boldsymbol{\beta})\right)_{nq} \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{\sum_{n=1}^N x_n\frac{\partial}{\partial \beta_k}\left(\mathbf{H}(\boldsymbol{\beta})\right)_{n1}}_{c_1^{(k)}(\boldsymbol{\beta})} & \dots & \underbrace{\sum_{n=1}^N x_n\frac{\partial}{\partial \beta_k}\left(\mathbf{H}(\boldsymbol{\beta})\right)_{nq}}_{c_q^{(k)}(\boldsymbol{\beta})} \end{bmatrix}
\end{aligned}$$

where

$$c_s^{(k)}(\boldsymbol{\beta}) = \sum_{n=1}^N x_n \underbrace{\frac{\partial}{\partial \beta_k}\left(\mathbf{H}(\boldsymbol{\beta})\right)_{ns}}_{\left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{ns}}$$

for $s = 1, \ldots, q$.

Thus,

$$\begin{aligned}
\frac{\partial}{\partial \beta_k}\mathbf{f}(\boldsymbol{\beta}) &= \left[\begin{array}{ccc} \sum_{n=1}^{N} x_n \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{n1} & \ldots & \sum_{n=1}^{N} x_n \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{nq} \end{array} \right] \\
&= \left[\begin{array}{ccc} x_1 & \ldots & x_N \end{array} \right] \mathbf{G}^{(k)}(\boldsymbol{\beta}) \\
&= \mathbf{x}^T \mathbf{G}^{(k)}(\boldsymbol{\beta}) \qquad (153)
\end{aligned}$$

for $k = 1, \ldots, p$. Finally,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{f}(\boldsymbol{\beta}) &= \left[\begin{array}{c} \frac{\partial}{\partial \beta_1}\mathbf{f}(\boldsymbol{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p}\mathbf{f}(\boldsymbol{\beta}) \end{array} \right] \\
&= \left[\begin{array}{c} \mathbf{x}^T \mathbf{G}^{(1)}(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{x}^T \mathbf{G}^{(p)}(\boldsymbol{\beta}) \end{array} \right]_{p \times q} \\
&= \left[\begin{array}{ccc} \mathbf{x}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}^T \end{array} \right]_{p \times pN} \left[\begin{array}{c} \mathbf{G}^{(1)}(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{G}^{(p)}(\boldsymbol{\beta}) \end{array} \right]_{pN \times q} \qquad (154)
\end{aligned}$$

Appendix 5E-3. Derivation of $\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \boldsymbol{\alpha}^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha} \right\}$

Let

$$\begin{aligned}
f(\boldsymbol{\beta}) &= \boldsymbol{\alpha}^T \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha} \\
&= \left[\begin{array}{ccc} \alpha_1 & \ldots & \alpha_q \end{array} \right] \left[\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right] \left[\begin{array}{c} \alpha_1 \\ \ldots \\ \alpha_q \end{array} \right] \\
&= \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{H}(\boldsymbol{\beta}) \right)_{ij} \alpha_j
\end{aligned}$$

First, lets take the derivative of $f(\boldsymbol{\beta})$ with respect to β_k

$$
\begin{aligned}
\frac{\partial}{\partial \beta_k} f(\boldsymbol{\beta}) &= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})\right)_{ij} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \frac{\partial}{\partial \beta_k} \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \sum_{m=1}^{N} \left\{ \frac{\partial}{\partial \beta_k} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right. \\
&\qquad \left. + \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \sum_{m=1}^{N} \left\{ \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{jm} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mi} \right. \\
&\qquad \left. + \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{jm} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mi} \right. \\
&\qquad \left. + \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{jm} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mi} \right\} \alpha_j \\
&\quad + \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{j=1}^{q}\sum_{i=1}^{q} \alpha_j \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{jm} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mi} \right\} \alpha_i \\
&\quad + \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ 2 \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \underbrace{\frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{mj}}_{\left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{mj}} \right\} \alpha_j \\
&= \sum_{i=1}^{q}\sum_{j=1}^{q} \alpha_i \left\{ 2 \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta})\right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{mj} \right\} \alpha_j
\end{aligned}
$$

for $k = 1, \ldots, p$.

Second, take the derivative of $\frac{\partial}{\partial \beta_k} f(\boldsymbol{\beta})$ wrt $\frac{\partial}{\partial \beta_l}$ for $l = 1, \ldots, p$:

$$
\begin{aligned}
& \frac{\partial}{\partial \beta_l} \frac{\partial}{\partial \beta_k} f(\boldsymbol{\beta}) \\
= & \frac{\partial}{\partial \beta_l} \left\{ \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i \left[2 \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj} \right] \alpha_j \right\} \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i \left[2 \sum_{m=1}^{N} \frac{\partial}{\partial \beta_l} \left\{ \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj} \right\} \right] \alpha_j \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i (2 \sum_{m=1}^{N} \underbrace{\frac{\partial}{\partial \beta_l} \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im}}_{\left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \right)_{im}} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj} \\
& \qquad + \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im} \underbrace{\frac{\partial}{\partial \beta_l} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj}}_{:= \left(\mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)_{mj}}) \alpha_j \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i \left[2 \sum_{m=1}^{N} \left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj} \right. \\
& \qquad \left. + \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)_{mj} \right] \alpha_j \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i 2 \left[\sum_{m=1}^{N} \left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{mj} \right. \\
& \qquad \left. + \sum_{m=1}^{N} \left(\mathbf{H}^T(\boldsymbol{\beta}) \right)_{im} \left(\mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)_{mj} \right] \alpha_j \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i 2 \left[\left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \mathbf{G}^{(k)}(\boldsymbol{\beta}) \right)_{ij} + \left(\mathbf{H}^T(\boldsymbol{\beta}) \mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)_{ij} \right] \alpha_j \\
= & \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i 2 \left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \mathbf{G}^{(k)}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)_{ij} \alpha_j \\
= & \boldsymbol{\alpha}^T 2 \underbrace{\left(\mathbf{F}^{(l)}(\boldsymbol{\beta}) \mathbf{G}^{(k)}(\boldsymbol{\beta}) + \mathbf{H}^T(\boldsymbol{\beta}) \mathbf{E}^{(l,k)}(\boldsymbol{\beta}) \right)}_{:= \mathbf{D}^{(l,k)}(\boldsymbol{\beta})} \boldsymbol{\alpha} \\
= & \boldsymbol{\alpha}^T 2 \mathbf{D}^{(l,k)}(\boldsymbol{\beta}) \boldsymbol{\alpha} \qquad (155)
\end{aligned}
$$

for $k = 1, \ldots, p$ and $l = 1, \ldots, p$.

Finally,

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} f(\boldsymbol{\beta}) \\
= &\begin{bmatrix} \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_1^2} & \cdots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_p^2} \end{bmatrix} \\
= &\begin{bmatrix} \boldsymbol{\alpha}^T \left[2\mathbf{D}^{(1,1)}(\boldsymbol{\beta})\right] \boldsymbol{\alpha} & \cdots & \boldsymbol{\alpha}^T \left[2\mathbf{D}^{(1,p)}(\boldsymbol{\beta})\right] \boldsymbol{\alpha} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}^T \left[2\mathbf{D}^{(p,1)}(\boldsymbol{\beta})\right] \boldsymbol{\alpha} & \cdots & \boldsymbol{\alpha}^T \left[2\mathbf{D}^{(p,p)}(\boldsymbol{\beta})\right] \boldsymbol{\alpha} \end{bmatrix}_{p \times p} \\
= &\begin{bmatrix} \boldsymbol{\alpha}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^T \end{bmatrix} 2 \begin{bmatrix} \mathbf{D}^{(1,1)}(\boldsymbol{\beta}) & \cdots & \mathbf{D}^{(1,p)}(\boldsymbol{\beta}) \\ \vdots & \ddots & \vdots \\ \mathbf{D}^{(p,1)}(\boldsymbol{\beta}) & \cdots & \mathbf{D}^{(p,p)}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha} \end{bmatrix}
\end{aligned}
\tag{156}
$$

Appendix 5E-4. Derivation of $\frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{\mathbf{x}^T \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha}\right\}$

Let

$$
\begin{aligned}
f(\boldsymbol{\beta}) &= \mathbf{x}^T \mathbf{H}(\boldsymbol{\beta}) \boldsymbol{\alpha} \\
&= \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} [\mathbf{H}(\boldsymbol{\beta})] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{q} x_i \left(\mathbf{H}(\boldsymbol{\beta})\right)_{ij} \alpha_j
\end{aligned}
$$

First, lets find the derivative of $f(\boldsymbol{\beta})$ with respect to β_k for $k = 1, \ldots, p$

$$
\begin{aligned}
\frac{\partial}{\partial \beta_k} f(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \sum_{j=1}^{q} x_i \underbrace{\frac{\partial}{\partial \beta_k} \left(\mathbf{H}(\boldsymbol{\beta})\right)_{ij}}_{\left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{ij}} \alpha_j \\
&= \sum_{i=1}^{N} \sum_{j=1}^{q} x_i \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{ij} \alpha_j
\end{aligned}
$$

Now, take the derivative with respect to β_l, $l = 1, \ldots, p$

$$\begin{aligned}
\frac{\partial}{\partial\beta_l}\frac{\partial}{\partial\beta_k} &= \frac{\partial}{\partial\beta_l}\left\{\sum_{i=1}^{N}\sum_{j=1}^{q} x_i \left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{ij} \alpha_j\right\} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{q} x_i \underbrace{\frac{\partial}{\partial\beta_l}\left(\mathbf{G}^{(k)}(\boldsymbol{\beta})\right)_{ij}}_{\left(\mathbf{E}^{(l,k)}(\boldsymbol{\beta})\right)_{ij}} \alpha_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{q} x_i \left(\mathbf{E}^{(l,k)}(\boldsymbol{\beta})\right)_{ij} \alpha_j \\
&= \begin{bmatrix} x_1 & \ldots & x_N \end{bmatrix} \left[\mathbf{E}^{(l,k)}(\boldsymbol{\beta})\right] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix} \\
&= \mathbf{x}^T \mathbf{E}^{(l,k)}(\boldsymbol{\beta})\boldsymbol{\alpha} \qquad (157)
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial\boldsymbol{\beta}^T} f(\boldsymbol{\beta}) \qquad (158) \\
&= \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{\beta})}{\partial\beta_1^2} & \cdots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial\beta_1\partial\beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{\beta})}{\partial\beta_p\partial\beta_1} & \cdots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial\beta_p^2} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{x}^T\mathbf{E}^{(1,1)}(\boldsymbol{\beta})\boldsymbol{\alpha} & \ldots & \mathbf{x}^T\mathbf{E}^{(1,p)}(\boldsymbol{\beta})\boldsymbol{\alpha} \\ \vdots & \ddots & \vdots \\ \mathbf{x}^T\mathbf{E}^{(p,1)}(\boldsymbol{\beta})\boldsymbol{\alpha} & \ldots & \mathbf{x}^T\mathbf{E}^{(p,p)}(\boldsymbol{\beta}) \end{bmatrix}_{p\times p} \\
&= \begin{bmatrix} \mathbf{x}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}^T \end{bmatrix} \begin{bmatrix} \mathbf{E}^{(1,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{E}^{(1,p)}(\boldsymbol{\beta}) \\ \vdots & \ddots & \vdots \\ \mathbf{E}^{(p,1)}(\boldsymbol{\beta}) & \ldots & \mathbf{E}^{(p,p)}(\boldsymbol{\beta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha} \end{bmatrix} \qquad (159)
\end{aligned}$$

List of References

[1] R. R. Tenney and N. R. Sandell Jr, "Detection with distributed sensors," *IEEE Transactions on Aerospace Electronic Systems*, vol. 17, pp. 501–510, 1981.

[2] P. K. Varshney, "Multisensor data fusion," *Electronics & Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.

[3] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[4] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 1, pp. 24–30, 1987.

[5] G. Ferrari and R. Pagliari, "Decentralized binary detection with noisy communication links," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, no. 4, pp. 1554–1563, 2006.

[6] P. K. Varshney, *Distributed detection and data fusion*. Springer-Verlag, 1997.

[7] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 28, no. 3, pp. 639–653, 1992.

[8] S. A Thomopoulos, R. Viswanathan, and D. Bougoulias, "Optimal distributed decision fusion," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 25, no. 5, pp. 761–765, 1989.

[9] A. M. Aziz, "A new adaptive decentralized soft decision combining rule for distributed sensor systems with data fusion," *Information Sciences*, vol. 256, pp. 197–210, 2014.

[10] A. Aziz, "A soft-decision fusion approach for multiple-sensor distributed binary detection systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 3, pp. 2208–2216, 2011.

[11] V. Aalo and R. Viswanathan, "Multilevel quantisation and fusion scheme for the decentralised detection of an unknown signal," in *Radar, Sonar and Navigation, IEE Proceedings*, vol. 141, no. 1. IET, 1994, pp. 37–44.

[12] S. A. Aldosari and J. M. Moura, "Detection in decentralized sensor networks," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 2. IEEE, 2004, pp. ii–277.

[13] S. Kay and F. Cogun, "Integrated sensor detection/localization for multi-source data," in *Radar Conference, 2014 IEEE*. IEEE, 2014, pp. 0708–0711.

[14] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on numerical analysis*, vol. 10, no. 2, pp. 413–432, 1973.

[15] F. Cogun and S. M. Kay, "Alternative approaches to data compression for distributed detection."

[16] S. M. Kay, *Fundamentals of statistical signal processing, volume II: Detection theory*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

[17] S. M. Kay, *Fundamentals of statistical signal processing, volume I: Estimation theory*. Prentice Hall, 1993.

[18] P. McCullagh, "Local sufficiency," *Biometrika*, vol. 71, no. 2, pp. 233–244, 1984.

[19] M. A. Richards, *Fundamentals of radar signal processing*. Tata McGraw-Hill Education, 2005.

CHAPTER 6

Future Work

Some of the assumptions and methods used in this research can be extended in the future:

- Our research was concentrated on the multimodal detection/localization of a single target. The derivations can be extended to include multiple targets and the detectors can be modified accordingly. This may in particular be useful for the specific example of detecting multiple vehicles in an urban area.

- A direct line of sight from the emitter to all RF receivers is assumed in our RF signal model. In practice, this might not be the case and the receivers may be subject to multipath. An RF detector dealing with multipath interference may be useful in practice.

- The performance of the GLRT statistic for IR data is found using Monte Carlo simulations in this work. It may also be possible to find the performance of this detector analytically, which requires the use of order statistics due to the maximization operation involved in the test statistic. Finding analytical expressions will enable us to realize a constant false alarm rate (CFAR) detector.

- Note that no prior knowledge of the observed scene was assumed in this study. It is possible to reduce down the search region if there is some known information about the observed scene (roads, geography, etc.). This information can be used to reduce computational complexity, as grid-search over all possible combinations of position and velocity will be avoided.

- The derivation of the GLRT detector is done assuming that the target velocity is constant throughout the movement of the target. An important consideration might be the case when the velocity of the target changes over a sequence of frames. In addition, the moving target may change its direction during its movement. Abrupt model change detection may be considered to increase overall detection performance.

- In the distributed detection problem of a NL signal, we used the second-order Taylor approximation of the log-likelihood function at each sensor. Another consideration might be to use a higher order expansion. Although the log-likelihood approximation would be better at each node, the computational load at each node and central processor would increase. There is need for further analysis.

- The extension of Approximate GLRT detector to the vector parameter case including nuisance parameters is done. However, the extension of the LSS GLRT detector has not been done. In future, the extension of the LSS GLRT detector which uses NL signal expansion may be considered. It may also be possible to use some alternative signal representation besides polynomials, which may lead to some other local statistics that may give better detection performance.

BIBLIOGRAPHY

"Pinhole camera model." Oct. 2014. [Online]. Available: https://en.wikipedia.org/wiki/Pinhole_camera_model

A Thomopoulos, S., Viswanathan, R., and Bougoulias, D., "Optimal distributed decision fusion," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 25, no. 5, pp. 761–765, 1989.

Aalo, V. and Viswanathan, R., "Multilevel quantisation and fusion scheme for the decentralised detection of an unknown signal," in *Radar, Sonar and Navigation, IEE Proceedings*, vol. 141, no. 1. IET, 1994, pp. 37–44.

Aldosari, S. A. and Moura, J. M., "Detection in decentralized sensor networks," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 2. IEEE, 2004, pp. ii–277.

Aziz, A., "A soft-decision fusion approach for multiple-sensor distributed binary detection systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 3, pp. 2208–2216, 2011.

Aziz, A. M., "A new adaptive decentralized soft decision combining rule for distributed sensor systems with data fusion," *Information Sciences*, vol. 256, pp. 197–210, 2014.

Barniv, Y., "Dynamic programming solution for detecting dim moving targets," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 1, pp. 144–156, 1985.

Barniv, Y. and Kella, O., "Dynamic programming solution for detecting dim moving targets part ii: analysis," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 6, pp. 776–788, 1987.

Blostein, S. D. and Huang, T. S., "Detecting small, moving objects in image sequences using sequential hypothesis testing," *Signal Processing, IEEE Transactions on*, vol. 39, no. 7, pp. 1611–1629, 1991.

Chan, D. S., "Unified framework for ir target detection and tracking," in *Aerospace Sensing*. International Society for Optics and Photonics, 1992, pp. 66–76.

Chan, D. S., Langan, D. A., and Staver, D. A., "Spatial-processing techniques for the detection of small targets in ir clutter," in *OE/LASE'90, 14-19 Jan., Los Angeles, CA*. International Society for Optics and Photonics, 1990, pp. 53–62.

Cogun, F. and Kay, S. M., "Alternative approaches to data compression for distributed detection."

Ferrari, G. and Pagliari, R., "Decentralized binary detection with noisy communication links," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 42, no. 4, pp. 1554–1563, 2006.

Golub, G. H. and Pereyra, V., "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on numerical analysis*, vol. 10, no. 2, pp. 413–432, 1973.

Kay, S., "Signal fitting with uncertain basis functions," *IEEE Signal Processing Letters*, vol. 6, no. 18, pp. 383–386, 2011.

Kay, S., "A computationally efficient nonlinear least squares method using random basis functions," *IEEE Signal Processing Letters*, vol. 7, no. 20, pp. 721–724, 2013.

Kay, S. and Cogun, F., "Integrated sensor detection/localization for multi-source data," in *Radar Conference, 2014 IEEE*. IEEE, 2014, pp. 0708–0711.

Kay, S. M., *Fundamentals of statistical signal processing, volume I: Estimation theory*. Prentice Hall, 1993.

Kay, S. M., *Fundamentals of statistical signal processing, volume II: Detection theory*. Prentice Hall Upper Saddle River, NJ, USA:, 1998.

Lehmann, E. L., Romano, J. P., and Casella, G., *Testing statistical hypotheses*, 3rd ed. Springer Science & Business Media, 2005.

McCullagh, P., "Local sufficiency," *Biometrika*, vol. 71, no. 2, pp. 233–244, 1984.

Michigan Tech Research Institute (MTRI), 3600 Green Court, Suite 100, Ann Arbor, MI 48105, 2015.

Mohanty, N., "Computer tracking of moving point targets in space," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 5, pp. 606–611, 1981.

Papastavrou, J. D. and Athans, M., "Distributed detection by a large team of sensors in tandem," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 28, no. 3, pp. 639–653, 1992.

Porat, B. and Friedlander, B., "A frequency domain algorithm for multiframe detection and estimation of dim targets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 4, pp. 398–401, 1990.

Reed, I. S., Gagliardi, R. M., and Stotts, L. B., "A recursive moving-target-indication algorithm for optical image sequences," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 26, no. 3, pp. 434–440, 1990.

Reed, I., Gagliardi, R., and Stotts, L. B., "Optical moving target detection with 3-d matched filtering," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 327–336, 1988.

Reed, I., Gagliardi, R., and Shao, H., "Application of three-dimensional filtering to moving target detection," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 6, pp. 898–905, 1983.

Reibman, A. R. and Nolte, L. W., "Optimal detection and performance of distributed sensor systems," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 1, pp. 24–30, 1987.

Richards, M. A., *Fundamentals of radar signal processing*. Tata McGraw-Hill Education, 2005.

Tenney, R. R. and Sandell Jr, N. R., "Detection with distributed sensors," *IEEE Transactions on Aerospace Electronic Systems*, vol. 17, pp. 501–510, 1981.

Tonissen, S. M. and Evans, R. J., "Peformance of dynamic programming techniques for track-before-detect," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 32, no. 4, pp. 1440–1451, 1996.

Vankayalapati, N. and Kay, S., "Asymptotically optimal detection of low probability of intercept signals using distributed sensors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 1, pp. 737–748, 2012.

Varshney, P. K., *Distributed detection and data fusion*. Springer-Verlag, 1997.

Varshney, P. K., "Multisensor data fusion," *Electronics & Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.

Viswanathan, R. and Varshney, P. K., "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

Wang, C. D., "Adaptive spatial/temporal/spectral filters for background clutter suppression and target detection," *Optical Engineering*, vol. 21, no. 6, pp. 216 033–216 033, 1982.

Wichura, M. J., *The coordinate-free approach to linear models*. Cambridge University Press, 2006, vol. 19.