2016

Signal Detection and Estimation for MIMO radar and Network Time Synchronization

Anand Srinivas Guruswamy
Lehigh University

SIGNAL DETECTION AND ESTIMATION FOR MIMO RADAR AND NETWORK TIME SYNCHRONIZATION

by

Anand Srinivas Guruswamy

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Lehigh University

January 2016

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Date

_____
Prof. Rick S. Blum
(Dissertation Director)

_____
Accepted Date

Committee Members:

_____
Prof. Rick S. Blum
(Committee Chair)

_____
Prof. Shalinee Kishore

_____
Prof. Liang Cheng

_____
Prof. Ping-Shi Wu

Acknowledgments

My PhD has been a humbling, enlightening experience that has served to open my mind in more ways than I can count. There are many people whom I would like to thank for helping me reach Lehigh, and for making my stay here so fruitful.

First and foremost, I would like to thank my advisor Prof. Rick Blum. Throughout my PhD, he has placed a great deal of trust in me, granting me considerable freedom in choosing the direction of my research. At the same time, he has provided me with gentle guidance whenever necessary, suggesting new ideas to explore and offering deeper insights into challenging topics. Most importantly, he has always provided me with the sharpest critique of my work, highlighting weakness in my research early to help me improve on my ideas. I have always felt that if Prof. Blum is satisfied with my ideas, then they have merit and can hold up against any scrutiny.

I would like to thank Prof. Shalinee Kishore, Prof. Liang Cheng and Prof. Ping-Shi Wu for agreeing to serve on my committee and providing feedback on my work. I would also like to thank Mark Bordogna and Prof. Kishore, for the wide-ranging insights and feedback they provided during the industry collaboration that formed the basis of much of my PhD research. My deepest thanks also goes to Prof. V.U. Reddy, my advisor at IIIT-Hyderabad, for teaching me many of the elements of being a good researcher. He had a huge role to play in my coming to Lehigh, along with the right mindset to succeed here.

I extend my thanks to my friends and colleagues at the SPCRL lab for their support, company and insightful discussions: Jiangfan Zhang, Basel Alnajjab, Raghu Srinivas Vishnubhotla, Joe Baker, David Saska, Qian He, Chuanming Wei and Yang Yang.

I would like to thank my close friends at Lehigh who made my life here memorable and

enjoyable: Abhishek Mishra, Anantha Krishna Karthik, Sonam Srivastava, Preeti Ashwin, Apratim Bhattacharya and Srinivas Mettu. I would also like to thank my many friends from the Indian community at Lehigh university who helped me countless times, and made my life here more interesting and enjoyable.

I would like to extend my deepest gratitude to my family, including my father Guruswamy Krishna, my mother Vijayalaxmi Guruswamy, my brother Aditya Krishna, and my sister-in-law Amrita Banerjee. Their boundless love and care are largely responsible making me the person I am today. They have always had the highest hopes for me, and have held an unwavering faith in my abilities.

Finally, I would like to thank my wife and my best friend, Akshaya Shankar, whom I first met when I came to Lehigh. Her love and support have helped me immeasurably in succeeding here. I consider the two greatest gifts that Lehigh has bestowed upon me to be her and my PhD.

Contents

List of Tables

List of Figures

Abstract

The theory of signal detection and estimation concerns the recovery of useful information from signals corrupted by random perturbations. This dissertation discusses the application of signal detection and estimation principles to two problems of significant practical interest: MIMO (multiple-input multiple output) radar, and time synchronization over packet switched networks. Under the first topic, we study the extension of several conventional radar analysis techniques to recently developed MIMO radars. Under the second topic, we develop new estimation techniques to improve the performance of widely used packet-based time synchronization algorithms.

The ambiguity function is a popular mathematical tool for designing and optimizing the performance of radar detectors. Motivated by Neyman-Pearson testing principles, an alternative definition of the ambiguity function is proposed under the first topic. This definition directly associates with each pair of true and assumed target parameters the probability that the radar will declare a target present. We demonstrate that the new definition is better suited for the analysis of MIMO radars that perform non-coherent processing, while being equivalent to the original ambiguity function when applied to conventional radars. Based on the nature of antenna placements, transmit waveforms and the observed clutter and noise, several types of MIMO radar detectors have been individually studied in literature. A second investigation into MIMO radar presents a general method to model and analyze the detection performance of such systems. We develop closed-form expressions for a Neyman-Pearson optimum detector that is valid for a wide class of radars. Further, general closed-form expressions for the detector SNR, another tool used to quantify radar performance, are derived. Theoretical and numerical results demonstrating the value of the proposed techniques to

optimize and predict the performance of arbitrary radar configurations are presented.

There has been renewed recent interest in the application of packet-based time synchronization algorithms such as the IEEE 1588 Precision Time Protocol (PTP), to meet challenges posed by next-generation mobile telecommunication networks. In packet based time synchronization protocols, clock phase offsets are determined via two-way message exchanges between a master and a slave. Since the end-to-end delays in packet networks are inherently stochastic in nature, the recovery of phase offsets from message exchanges must be treated as a statistical estimation problem. While many simple intuitively motivated estimators for this problem exist in the literature, in the second part of this dissertation we use estimation theoretic principles to develop new estimators that offer significant performance benefits. To this end, we first describe new lower bounds on the error variance of phase offset estimation schemes. These bounds are obtained by re-deriving two Bayesian estimation bounds, namely the Ziv-Zakai and Weiss-Weinstien bounds, for use under a non-Bayesian formulation. Next, we describe new minimax estimators for the problem of phase offset estimation, that are optimum in terms of minimizing the maximum mean squared error over all possible values of the unknown parameters. Minimax estimators that utilize information from past timestamps to improve accuracy are also introduced. These minimax estimators provide fundamental limits on the performance of phase offset estimation schemes. Finally, a restricted class of estimators referred to as L-estimators are considered, that are linear functions of order statistics. The problem of designing optimum L-estimators is studied under several hitherto unconsidered criteria of optimality. We address the case where the queuing delay distributions are fully known, as well as the case where network model uncertainty exists. Optimum L-estimators that utilize information from past observation windows to improve performance are also described. Simulation results indicate that significant performance gains over conventional

estimators can be obtained via the proposed optimum processing techniques.

Chapter 1

Introduction

1.1 MIMO Radar

A radar is a system that transmits electromagnetic signals, and analyzes the reflections returned from objects in its surroundings (referred to as *targets*) to deduce information such as the position, velocity and size of these objects. Until recently, most deployed radar systems could be classified as either *phased arrays* or *multistatic radars*. Phased arrays utilize several closely spaced transmitters and receivers, jointly designed to focus on a narrow region of space at any given time. This is achieved by transmitting scaled versions of a single waveform from different transmitters, and coherently processing the received signals. Multistatic radars consist of multiple independently operating phased array radars, whose decisions are fused, often suboptimally, to improve performance over that of any individual subsystem.

The term MIMO radar refers to a new, holistic approach to radar design, wherein multiple transmitters and receivers are jointly designed and operated. MIMO radars have received significant research interest in recent years, owing to advances in integrated circuit technol-

ogy and signal processing techniques that have improved the practical viability of the MIMO radar concept. By allowing greater design flexibility in the choice of transmit waveforms, the placement of transmit and receive antennas, as well the design of receiver processing algorithms, MIMO radars can exhibit significantly improved performance characteristics relative to conventional radars. MIMO radar configurations considered in literature can be broadly classified into two categories. In MIMO radars with closely spaced antennas [1, 2, 3], also known as colocated MIMO radar, arbitrarily correlated waveforms can be used to increase the degrees of freedom available for beamforming relative to phased arrays. In MIMO radars with widely separated antennas, significantly improved localization accuracy is possible under coherent processing [4, 5, 6], while under non-coherent processing [6, 7, 8, 9, 10] spatial diversity gains can be obtained when different transmit/receive paths illuminate different aspects of a target's radar cross section (RCS). In the first part of this dissertation, we seek to extend several conventional radar analysis and waveform design techniques to MIMO radars.

1.2 Packet-based Time Synchronization

In the modern technological era, there are a myriad of electronic devices that are critically dependent on the availability of a common time reference across many different physical locations. To this end, these devices typically perform timekeeping locally using clock hardware that exploits the periodicity of certain physical phenomena, such as the mechanical resonance of vibrating crystals (in low-cost quartz crystal oscillators), or electromagnetic transitions within cesium or rubidium atoms (in expensive atomic clocks). However, all such timekeeping techniques are subject to random errors that accumulate over large time scales, and the cost, size and complexity of timekeeping hardware are all typically proportional

to clock stability. As a result, there are often scenarios where it is impractical to locally maintain clock hardware required to achieve a desired level of stability, due to space or budget constraints. Network time synchronization algorithms address this problem by regularly correcting the output of low-cost, low-stability clocks (termed *slaves*) using measurements obtained from high-cost, high-stability clocks (termed *masters*) via an interconnecting network. These algorithms are widely used in areas such as industrial measurement and control, wireless sensor networks, telecommunications, smart grids, and internet-enabled applications such as voice and video telephony and financial communications.

Many network time synchronization protocols have been developed in the literature, addressing different types of networks. For instance, the Network Time Protocol [11] and the IEEE 1588 Precision Time Protocol (PTP) [12] are widely used in IP networks. For wireless sensor networks, protocols such as Reference Broadcast Synchronization [13], Post facto [14], Tiny-Sync and Mini-sync [15] have been developed. In the context of ad-hoc communication networks, Romer's time synchronization protocol [16] has been discussed. Though these protocols differ from each other in many aspects, a fundamental mechanism common to most synchronization protocols is the *two-way message exchange*; which refers to the exchange of messages between a pair of nodes to achieve clock synchronization.

During a two-way message exchange, a slave node exchanges a series of packets with a master node over an interconnecting network, and collects timestamps corresponding to the departure and arrival times of these packets. The slave then attempts to utilize these timestamps to correct it own clock, however this is hindered by the random queuing delays experienced by packets as they traverse the interconnecting network. The second part of this dissertation develops new approaches to combat the degrading effects of these random queuing delays on synchronization accuracy.

While the techniques we develop can be applied to any protocol that uses two-way message exchanges, we mainly discuss the applications of our results in the context of PTP applied to telecommunication networks. Packet-based time synchronization techniques based on the IEEE 1588 Precision Time Protocol (PTP) are being increasingly considered as a means of providing microsecond-level synchronization between cell towers in 4G LTE (Long Term Evolution) mobile networks [17, 18, 19, 20, 21]. Such a high degree of synchronization accuracy is a necessity in 4G networks since it helps ensure seamless handovers between cell towers, helps reduce inter-cell interference, and also enables the use of MIMO techniques to improve capacity [22]. As compared to GPS (global positioning system) based synchronization, packet-based synchronization is often more cost-effective since it utilizes the existing mobile backhaul network infrastructure that is used to interconnect cell towers. However, since backhaul networks are typically leased from commercial internet service providers (ISPs), mobile network operators must share their use with other commercial and residential users. Background traffic generated by these users often results in large random network delays that hinder packet-based synchronization. Overcoming this problem is key to the adoption of packet-based synchronization schemes in mobile backhaul networks, especially given that the synchronization accuracy requirements are only expected to grow more stringent in the future. In the second part of this dissertation, we attempt to address this challenge by developing new techniques to combat variable network delays, and demonstrate their performance in mobile backhaul networks.

1.3 Outline of the Dissertation

As mentioned in the previous sections, this dissertation focuses on two topics: MIMO radar, and network time synchronization. The dissertation consists of five chapters. Chapters 2 and 3 are related to the first topic of MIMO radar, while Chapters 4, 5 and 6 are related to the second topic of network time synchronization. A brief outline of each chapter is presented below.

In Chapter 2, we propose a new definition of the ambiguity function, an important tool used to analyze and optimize the performance of radar detectors. Motivated by Neyman-Pearson testing principles, we propose an alternative definition of the ambiguity function that directly associates with each pair of true and assumed target parameters the probability that the radar will declare a target present. We show that the original ambiguity function definition of Woodward and Davies for single antenna systems (and its extensions to multichannel systems that use coherent processing) are essentially equivalent to the proposed definition. Further, for radars that perform non-coherent processing, we show the extensions to Woodward's ambiguity function proposed in the literature are not equivalent to our proposed definition, and therefore may not accurately reflect detection performance. Simulations results demonstrate the differences between these different ambiguity function definitions for non-coherent radars. This research was published in in [23].

In Chapter 3, we study the general case where the transmit and receiver antennas have arbitrary separations, while also assuming that the transmit waveforms are arbitrarily correlated with one another. In this general scenario, we derive closed form expressions for the optimal Neyman-Pearson detector assuming white as well as colored clutter-plus-noise. We further obtain general expressions for the detector signal-to-noise ratio (SNR), a popular

measure of detection performance. It is shown that for radars with widely separated antennas, orthogonal waveforms maximize detector SNR at high received SNRs. A new scalar measure to characterize the nature of the waveform correlation matrix and the channel covariance matrix, termed the *normalized eigenvariance*, is also introduced. Simulation results are presented demonstrating the nature of the optimal transmit waveforms for various antenna separations. This research was published in [24].

In Chapter 4, we describe new lower bounds on error variance of phase offset estimation schemes used in PTP based synchronization. To this end, we re-derive two Bayesian estimation bounds, namely the Ziv-Zakai and Weiss-Weinstien bounds, for use under a non-Bayesian formulation. This enables us to apply these bounds to the problem of phase offset estimation. Simulation results compare the performance of existing estimation schemes against these lower bounds under a variety of different network scenarios. This research was published in [25].

In Chapter 5, we describe new minimax estimators for the problem of phase offset estimation in PTP based synchronization. These estimators are optimum in terms of minimizing the maximum mean squared error over all possible values of the unknown parameters. Minimax estimators that utilize information from past timestamps to improve accuracy are also introduced. These minimax estimators also provide fundamental limits on the performance of phase offset estimation schemes. Simulation results indicate that significant performance gains over conventional estimators can be obtained via such optimum processing techniques. This research was published in [26].

In Chapter 6, we consider a restricted class of estimators referred to as L-estimators, which are linear functions of order statistics. The problem of designing optimum L-estimators is studied under several hitherto unconsidered criteria of optimality. Our results address the

case where the queuing delay distributions are fully known, as well as the case where network model uncertainty exists. Optimum L-estimators that utilize information from past observation windows to improve performance are also described. The derived L-estimators have a much lower computational complexity than minimax estimators, and also require lesser statistical knowledge of the queuing delays. Simulation results indicate that L-estimators exhibit a mean squared estimation error very close to minimax estimators under many network scenarios. This research was published in [27].

Chapter 2

Ambiguity Functions for Non-Coherent MIMO Radars

2.1 Problem Motivation

The ambiguity function (AF) is often used to study the effect of system design choices such sensor geometry, waveform shape and receiver processing techniques on radar detection performance. It was first proposed by Woodward and Davies [28] in the context of single antenna radars, as a consequence of the nature of the optimal Neyman-Pearson detector. The optimal decision rule for such systems involved comparing a test statistic of the form

$$T = |r|^2 = |\alpha \chi_{\mathrm{W}}(\Delta\tau, \Delta\nu) + w|^2 \tag{2.1}$$

to a threshold, where r denotes the receiver's matched filter output, and α, w respectively represent (zero-mean) Gaussian random variables corresponding to the target's scattering

coefficient and receiver noise. $\chi_{\mathrm{W}}(\Delta\tau, \Delta\nu)$ represents Woodward's AF, defined as

$$\chi_{\mathrm{W}}(\Delta\tau, \Delta\nu) = \int_0^{T_s} s(t)s^*(t+\Delta\tau)\exp\left\{-j2\pi\Delta\nu t\right\}\mathrm{d}t$$

where $\Delta\tau$ and $\Delta\nu$ respectively denote the difference between the true and assumed values of the target's delay and Doppler frequency, and $s(t)$ is the transmitted waveform. Radar designers optimize $\chi_{\mathrm{W}}(\Delta\tau, \Delta\nu)$ to achieve a 'thumbtack' shape; the rationale for such optimization criteria is that maximizing $|\chi_{\mathrm{W}}(0,0)|$ maximizes the response of the detector to matched targets, while minimizing $|\chi_{\mathrm{W}}(\Delta\tau, \Delta\nu)|$ when $(\Delta\tau, \Delta\nu) \neq (0,0)$ minimizes the response of the detector to mismatched targets.

In most modern radar systems, multiple channels (established via multiplexing in the time, frequency, or spatial domains) exist for the propagation of signals between the transmitter(s) and the receiver(s). Consider a hypothetical two-channel radar that transmits distinct waveforms in each channel, and obtains the matched filter outputs

$$r_i = \alpha_i \chi_{\mathrm{W}}^{(i)}(\Delta\tau, \Delta\nu) + w_i \quad \text{for } i = 1, 2$$

where the noise terms w_1 and w_2 are i.i.d. zero-mean Gaussian random variables. In such a system, if the channel gains α_1 and α_2 are also zero mean Gaussian random variables, then their mutual correlation determines the nature of the optimal joint detector. For instance, when α_1 and α_2 are fully correlated (hence admitting the representation $\alpha_i = c_i\alpha$ for $i = 1, 2$, where c_i is a deterministic scalar), the optimum test statistic has the form

$$T = \left|\sum_{i=1}^{2} b_i r_i\right|^2 = \left|\alpha \sum_{i=1}^{2} b_i c_i \chi_{\mathrm{W}}^{(i)}(\Delta\tau, \Delta\nu) + \sum_{i=1}^{2} b_i w_i\right|^2 \tag{2.2}$$

where $\{b_i\}_{i=1}^{K}$ are deterministic weights. Such processing of the matched filter outputs is said to be *coherent*, and occurs in any radar where all channel gains can be jointly characterized

using a single Gaussian random variable, such as phased arrays or MIMO radars with closely spaced antennas [29]. Comparing (2.1) and (2.2), the equivalent of Woodward's AF for this system can be defined as

$$\hat{\chi}_W(\tau,\nu) = \sum_{i=1}^{K} b_i c_i \chi_{\mathrm{W}}^{(i)}(\Delta\tau,\Delta\nu) \ , \tag{2.3}$$

i.e., the overall AF is a weighted sum of the individual Woodward's AFs for each channel. When α_1 and α_2 are uncorrelated, then the test statistic has the form

$$T = \sum_{i=1}^{2} b_i |r_i|^2 = \sum_{i=1}^{2} b_i \left| \alpha_i \chi_{\mathrm{W}}^{(i)}(\Delta\tau,\Delta\nu) + w_i \right|^2 \tag{2.4}$$

Such processing is said to be *non-coherent*, and results whenever more than one Gaussian random variable is required to fully describe all channel gains. This scenario occurs in any radar that exploits diversity techniques to improve performance, including wideband radars, multi-pulse systems where the target RCS varies rapidly, and MIMO radar with widely separated antennas. Observe that in (2.4), the channel gains cannot be factored out from the individual Woodward's AFs as in (2.2), therefore it is no longer possible to define an AF analogous to (2.3) for the overall system.

Many recent research papers, especially in the context of multi-antenna radar, have attempted to describe an AF for systems that use non-coherent processing. For instance, in [30], the ambiguity function is defined as weighted sum of individual bistatic ambiguity functions. On the other hand, in [31], the AF is defined as the expected value of the test statistic. While both these definitions bear intuitive appeal, it is unclear whether they truly reflect the final performance goals of the system, which is to maximize the probability of detection of matched targets while minimizing the probability of detection of mismatched targets.

2.2 Proposed Ambiguity Function Definition

Consider a general radar model, in which the received signals are random processes whose statistics embed information about target parameters such as position and velocity. Assume, for simplicity, that exactly one target is present in the field of view (FOV) of the radar (our discussion can also be extended to multi-target scenarios). Let $\boldsymbol{\theta}_{\mathrm{T}}$ denote as a vector comprising the *true* values of all target parameters, and let $\boldsymbol{\theta}_{\mathrm{A}}$ represent the target parameter values assumed at the receiver. During detection, we typically consider the following hypothesis testing problem in each detection bin:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{Target present at } \boldsymbol{\theta}_{\mathrm{A}} \\ \mathcal{H}_1: &\quad \text{Target absent} \end{aligned} \tag{2.5}$$

The optimal Neyman-Pearson decision rule for this problem will have the form

$$T \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \eta \,, \tag{2.6}$$

where T represents the test statistic, with the detection threshold η chosen so that the observed false alarm rate $\Pr\{T > \eta \mid \mathcal{H}_1\}$ equals the the required false alarm rate ϵ.

When framing the hypothesis testing problem in (2.5), we ignore the possibility that the target may be present in a bin other the bin under test. Otherwise, we would have to consider the following hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{Target is present at } \boldsymbol{\theta}_{\mathrm{A}} \text{ (i.e. } \boldsymbol{\theta}_{\mathrm{T}} = \boldsymbol{\theta}_{\mathrm{A}}) \\ \mathcal{H}_1: &\quad \left[\begin{array}{l} \text{Either target absent, or target not} \\ \text{present at } \boldsymbol{\theta}_{\mathrm{A}} \text{ (i.e. } \boldsymbol{\theta}_{\mathrm{T}} \neq \boldsymbol{\theta}_{\mathrm{A}}) \end{array} \right] \end{aligned} \tag{2.7}$$

Note that the null hypothesis in (2.7) is composite. It can be shown that no uniformly most

powerful (UMP) test exists for this problem, therefore the preferred approach is to only consider the simple hypothesis testing problem in (2.5). Such an approximate hypothesis testing model will be valid as long as we ensure that the presence of mismatched targets under the null hypothesis does not significantly impact the performance of the detector.

The test statistic T is computed by processing the received signals (whose statistics depend on $\boldsymbol{\theta}_{\mathrm{T}}$ when a target is present in the FOV of the radar) using the assumed target parameter vector $\boldsymbol{\theta}_{\mathrm{A}}$. Therefore we may write

$$T = \begin{cases} T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) & \left[\begin{array}{l}\text{when target is present}\\ \text{in the radar's FOV,}\end{array}\right] \\ T_{\mathrm{Ab}}(\boldsymbol{\theta}_{\mathrm{A}}) & \left[\begin{array}{l}\text{when target is absent}\\ \text{from the radar's FOV.}\end{array}\right] \end{cases}$$

The probability of detection of mismatched targets is

$$P_{\mathrm{F}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) = \Pr\left\{T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) > \eta\right\} \tag{2.8}$$

whenever $\boldsymbol{\theta}_{\mathrm{T}} \neq \boldsymbol{\theta}_{\mathrm{A}}$, while the probability of detection of matched targets is

$$P_{\mathrm{D}}(\boldsymbol{\theta}_{\mathrm{A}}) = \Pr\left\{T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) > \eta \mid \boldsymbol{\theta}_{\mathrm{T}} = \boldsymbol{\theta}_{\mathrm{A}}\right\} \tag{2.9}$$

In order to minimize the impact of mismatched targets on the performance of the detector, we must minimize $P_{\mathrm{F}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}})$. Further, we would like to maximize the probability of detection $P_{\mathrm{D}}(\boldsymbol{\theta}_{\mathrm{A}})$ of matched targets. To state these optimization goals simultaneously we propose the following AF definition

$$\chi(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) = \Pr\left\{T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) > \eta\right\} = \begin{cases} P_{\mathrm{D}}(\boldsymbol{\theta}_{\mathrm{A}}) & \text{if } \boldsymbol{\theta}_{\mathrm{T}} = \boldsymbol{\theta}_{\mathrm{A}} \\ P_{\mathrm{F}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) & \text{if } \boldsymbol{\theta}_{\mathrm{T}} \neq \boldsymbol{\theta}_{\mathrm{A}} \end{cases}$$

Clearly, we must aim to minimize $\chi(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}})$ whenever $\boldsymbol{\theta}_{\mathrm{T}} \neq \boldsymbol{\theta}_{\mathrm{A}}$, and maximize it when $\boldsymbol{\theta}_{\mathrm{T}} = \boldsymbol{\theta}_{\mathrm{A}}$; this is similar to how Woodward's AF is optimized.

2.3 Ambiguity Function for Gaussian Signals

In this section, we further study the proposed AF under a Gaussian signal detection scenario. For ease of explanation, we consider a multi-antenna narrowband radar with M transmitters and N receivers, and assume a zero-velocity target. The form of optimal test statistic for this system will depend on nature of correlations between transmit waveforms, as well as sensor placements. Hence, we first derive the optimal detector under a general model that allows arbitrarily correlated waveforms and arbitrary radar geometries. We later consider specific coherent and non-coherent processing scenarios.

Let $\sqrt{E_s}s_m(t)$ $(0 \leq t \leq T_s)$ be the complex baseband signal transmitted by the m^{th} transmitter, where E_s is the total transmitted energy and $\sum_{m=1}^{M} \int_0^{T_s} |s_m(t)|^2 \, dt = 1$. The signal arriving at the n^{th} receiver can be modelled as

$$r_n(t) = \sqrt{E_r} \sum_{m=1}^{M} \alpha_{n,m} u_{n,m}(t, \boldsymbol{\theta}_{\mathrm{T}}) + w_n(t) \tag{2.10}$$

where $\alpha_{n,m}$ represents the channel gain between the m^{th} transmitter and the n^{th} receiver, E_r is the received energy along the first transmit-receive path ($\alpha_{1,1} = 1$), and $w_n(t)$ represents additive Gaussian receiver noise, with time-correlation

$$\mathrm{E}\left[w_n(t) w_{n'}^*(t')\right] = \frac{\sigma_n^2}{T_s} \delta(n - n') \delta(t - t')$$

Further, $\boldsymbol{\theta}_{\mathrm{T}}$ contains the target's coordinates, and $u_{n,m}(t, \boldsymbol{\theta}_{\mathrm{T}})$ represents the signal obtained by applying path delay effects to $s_m(t)$, given that it arrives at the n^{th} receiver after being reflected by a target located at $\boldsymbol{\theta}_{\mathrm{T}}$.

Let $\mathbf{H} = [\alpha_{n,m}]_{N\times M}$ and let $\boldsymbol{\alpha} = \text{vec}\{\mathbf{H}^T\}$, where vec{.} represents the vectorization operation. We assume a Gaussian vector channel, i.e. $\boldsymbol{\alpha} \sim C\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$, where $\boldsymbol{\Sigma}_\alpha$ is a known positive semidefinite matrix.

To compute the test statistic for the detection bin corresponding to target coordinates $\boldsymbol{\theta}_\text{A}$, we must first compute the matched filter outputs

$$r_{n,m} = \int_0^{T_s} r_n(t) u^*_{n,m}(t, \boldsymbol{\theta}_\text{A}) dt = \sqrt{E_r} \sum_{m'=1}^{M} \alpha_{n,m'} \xi_{n,m,m'}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) + w_{n,m} \tag{2.11}$$

for $n = 1, \cdots, N$ and $m = 1, \cdots, M$, where

$$\xi_{n,m,m'}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) = \int_0^{T_s} u_{n,m'}(t, \boldsymbol{\theta}_\text{T}) u^*_{n,m}(t, \boldsymbol{\theta}_\text{A}) \text{d}t \tag{2.12}$$

and

$$w_{n,m} = \int_0^{T_s} w_n(t) u^*_{n,m}(t, \boldsymbol{\theta}_\text{A}) \text{d}t$$

Define $\mathbf{R} = [r_{n,m}]_{N\times M}$ and $\mathbf{W} = [w_{n,m}]_{N\times M}$; and let $\mathbf{r} = \text{vec}\{\mathbf{R}\}$ and $\mathbf{w} = \text{vec}\{\mathbf{W}\}$. Further, define as the waveform correlation matrix

$$\boldsymbol{\xi}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) = \begin{bmatrix} \boldsymbol{\xi}_1(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\xi}_N(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) \end{bmatrix} \tag{2.13}$$

where

$$\boldsymbol{\xi}_n(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) = \begin{bmatrix} \xi_{n,1,1}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) & \cdots & \xi_{n,1,M}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) \\ \vdots & \ddots & \vdots \\ \xi_{n,M,1}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) & \cdots & \xi_{n,M,M}(\boldsymbol{\theta}_\text{T}, \boldsymbol{\theta}_\text{A}) \end{bmatrix}.$$

Then our model reduces to

$$\mathbf{r} = \sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}})\boldsymbol{\alpha} + \mathbf{w}$$

with $\mathbf{w} \sim C\mathcal{N}(\mathbf{0}, \sigma_w^2\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}}))$. In each bin, we can now frame the following hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{r} = \sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}})\boldsymbol{\alpha} + \mathbf{w} \quad \text{(Target present at } \boldsymbol{\theta}_{\mathrm{A}}\text{)} \\ \mathcal{H}_1: &\quad \mathbf{r} = \mathbf{w} \qquad\qquad\qquad\qquad\quad \text{(Target absent)} \end{aligned}$$

Clearly, under both hypotheses, $\mathbf{r}$ is a zero mean Gaussian random vector, with covariance matrix

$$\mathrm{E}\left[\mathbf{r}\mathbf{r}^{\dagger}\right] = \begin{cases} \boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}})\left[E_r\boldsymbol{\Sigma}_{\alpha}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}}) + \sigma_w^2\mathbf{I}\right] = \boldsymbol{\Sigma}_1 & \mathcal{H}_0 \\ \sigma_w^2\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}}) = \boldsymbol{\Sigma}_0 & \mathcal{H}_1 \end{cases}$$

where $()^{\dagger}$ represents the Hermitian transpose operation. Assuming $\boldsymbol{\Sigma_0}$ and $\boldsymbol{\Sigma_1}$ are positive definite, the test statistic can be written as the log-likelihood ratio

$$T = \mathbf{r}^{\dagger}\mathbf{D}(\boldsymbol{\theta}_{\mathrm{A}})\mathbf{r}$$

where

$$\mathbf{D}(\boldsymbol{\theta}_{\mathrm{A}}) = \boldsymbol{\Sigma_0}^{-1} - \boldsymbol{\Sigma_1}^{-1} = (\rho\boldsymbol{\Sigma}_{\alpha}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{A}}) + \mathbf{I})^{-1}\boldsymbol{\Sigma}_{\alpha} \tag{2.14}$$

and $\rho = E_r/\sigma_w^2$ denotes the received signal to noise ratio. Using low-rank matrix decompositions, it can be shown that the final expression on the right hand side of (2.14) will be valid even if $\boldsymbol{\Sigma_0}$ and $\boldsymbol{\Sigma_1}$ are rank deficient. We now consider two specific scenarios under this general Gaussian signal model.

2.3.1 Coherent Processing

Consider a scenario where a radar's sensors are closely spaced and the signals are narrowband in nature. Here the received signal can be modelled as [32]

$$r_n(t) = \sqrt{E_r} \sum_{m=1}^{M} s_m(t - \tau(\boldsymbol{\theta}_{\mathrm{T}})) \underbrace{\alpha_0 e^{j2\pi f_c[\tau(\boldsymbol{\theta}_{\mathrm{T}}) - \tau_{m,n}(\boldsymbol{\theta}_{\mathrm{T}})]}}_{\alpha_{n,m}} + w_n(t)$$

where α_0 represents the nominal channel gain (a zero-mean Gaussian random variable with $\mathrm{E}\left[|\alpha_0|^2\right] = 1$), $\tau(\boldsymbol{\theta}_{\mathrm{T}})$ represents the nominal path delay, and $\tau_{m,n}(\boldsymbol{\theta}_{\mathrm{T}})$ is a corrective factor used to make $\tau(\boldsymbol{\theta}_{\mathrm{T}}) - \tau_{m,n}(\boldsymbol{\theta}_{\mathrm{T}})$ equal to the delay along the $(m,n)^{\mathrm{th}}$ transmit-receive path (typically $|\tau_{m,n}(\boldsymbol{\theta}_{\mathrm{T}})| \ll |\tau(\boldsymbol{\theta}_{\mathrm{T}})|$). Comparing against (2.10), (2.11) and (2.12), for this system we have

$$\begin{aligned} u_{n,m}(t, \boldsymbol{\theta}_{\mathrm{T}}) &= s_m(t - \tau(\boldsymbol{\theta}_{\mathrm{T}})) \ , \\ \xi_{n,m,m'}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) &= \int_0^{T_s} s_m(t - \tau(\boldsymbol{\theta}_{\mathrm{T}})) s_{m}^{*}(t - \tau(\boldsymbol{\theta}_{\mathrm{A}})) \mathrm{d}t \end{aligned}$$

Further, the narrowband assumption allows us to write

$$\begin{aligned} \xi_{n,m,m'}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}) &\approx \int_0^{T_s} s_m(t) s_{m}^{*}(t + \tau(\boldsymbol{\theta}_{\mathrm{T}}) - \tau(\boldsymbol{\theta}_{\mathrm{A}})) \mathrm{d}t \\ &= \int_0^{T_s} s_m(t) s_{m}^{*}(t + \Delta\tau) \mathrm{d}t \triangleq \xi_{n,m,m'}(\Delta\tau) \end{aligned}$$

where $\Delta\tau = \tau(\boldsymbol{\theta}_{\mathrm{T}}) - \tau(\boldsymbol{\theta}_{\mathrm{A}})$. Hence we can reexpress $\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}})$ in (2.13) as $\boldsymbol{\xi}(\Delta\tau)$, where

$$\boldsymbol{\xi}(\Delta\tau) = \begin{bmatrix} \boldsymbol{\xi}_1(\Delta\tau) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\xi}_N(\Delta\tau) \end{bmatrix}$$

and

$$\boldsymbol{\xi}_n(\Delta\tau) = \begin{bmatrix} \xi_{n,1,1}(\Delta\tau) & \cdots & \xi_{n,1,M}(\Delta\tau) \\ \vdots & \ddots & \vdots \\ \xi_{n,M,1}(\Delta\tau) & \cdots & \xi_{n,M,M}(\Delta\tau) \end{bmatrix}.$$

We note that the channel gains $\alpha_{n,m}$ are all multiples of the Gaussian random variable α_0, hence the channel covariance matrix $\boldsymbol{\Sigma}_\alpha$ will be unit rank, and as a result $\mathbf{D}(\boldsymbol{\theta}_\mathrm{T})$ (in (2.14)) will also be unit rank. Thus we can write $\mathbf{D}(\boldsymbol{\theta}_\mathrm{T}) = \mathbf{b}\mathbf{b}^\dagger$, where $\mathbf{b}$ is a vector, and thereby reduce the test statistic to the form $T = \left|\mathbf{b}^\dagger\mathbf{r}\right|^2$. Such processing is termed coherent because the elements of the vector $\mathbf{r}$ are weighted and summed prior to the magnitude square operation. The proposed AF for this system is given as $\tilde{\chi}(\boldsymbol{\theta}_\mathrm{T}, \boldsymbol{\theta}_\mathrm{A}) = \Pr\left\{T_\mathrm{Pr}(\boldsymbol{\theta}_\mathrm{T}, \boldsymbol{\theta}_\mathrm{A}) > \eta\right\}$, where

$$T_\mathrm{Pr}(\boldsymbol{\theta}_\mathrm{T}, \boldsymbol{\theta}_\mathrm{A}) = \left|\mathbf{b}^\dagger\left[\sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_\mathrm{T}, \boldsymbol{\theta}_\mathrm{A})\boldsymbol{\alpha} + \mathbf{w}\right]\right|^2 \tag{2.15}$$

To show that this AF is equivalent to Woodward's AF, we note that $\mathbf{b}^\dagger\left[\sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_\mathrm{T}, \boldsymbol{\theta}_\mathrm{A})\boldsymbol{\alpha} + \mathbf{w}\right]$ is a zero-mean circular symmetric complex Gaussian random variable. If x_1 and x_2 are two random variables independently distributed as $C\mathcal{N}(0,1)$, then

$$\mathrm{E}\left[|x_1|^2\right] > \mathrm{E}\left[|x_2|^2\right] \iff \Pr\left\{|x_1|^2 > \eta\right\} > \Pr\left\{|x_2|^2 > \eta\right\} \quad \forall\, \eta \in \mathbb{R}$$

This property implies that under coherent processing, optimizing the proposed AF is equivalent to optimizing the expected value of the test statistic. Let $\boldsymbol{\Sigma}_\alpha = \mathbf{d}\mathbf{d}^\dagger$, where $\mathbf{d}$ is a vector.

Thus we may equivalently define the AF as

$$\begin{aligned}\tilde{\chi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}}) &= \mathrm{E}\left[T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}})\right] = \mathrm{E}\left[\left|\mathbf{b}^{\dagger}\left[\sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}})\boldsymbol{\alpha}+\mathbf{w}\right]\right|^2\right] \\ &= E_r\mathbf{b}^{\dagger}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}})\boldsymbol{\Sigma}_{\alpha}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}})\mathbf{b} + \mathbf{b}^{\dagger}\mathrm{E}\left[\mathbf{w}^{\dagger}\mathbf{w}\right]\mathbf{b} \\ &= E_r\left|\mathbf{b}^{\dagger}\boldsymbol{\xi}(\Delta\tau)\mathbf{d}\right|^2 + \mathbf{b}^{\dagger}\boldsymbol{\xi}(\mathbf{0})\mathbf{b} \end{aligned} \tag{2.16}$$

The second term on the right hand side of (2.16) is constant, and hence may be ignored. This leads us to the following equivalent AF definition

$$\hat{\chi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}}) = \mathbf{b}^{\dagger}\boldsymbol{\xi}(\Delta\tau)\mathbf{d}\ , \tag{2.17}$$

which corresponds exactly to extensions of Woodward's AF to multi-antenna systems with closely spaced antennas [32].

2.3.2 Non-Coherent Processing

For a multi-antenna radar with widely separated sensors, it can be shown [33] that $\boldsymbol{\Sigma}_{\alpha}$ will be full rank, and hence that $\mathbf{D}(\boldsymbol{\theta}_{\mathrm{A}})$ will also be full rank. When a target is present in the FOV of the radar, the test statistic will thus assume the form

$$T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}}) = \left|\left|\left[\mathbf{D}(\boldsymbol{\theta}_{\mathrm{A}})\right]^{1/2}\left[\sqrt{E_r}\boldsymbol{\xi}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}})\boldsymbol{\alpha}+\mathbf{w}\right]\right|\right|^2\ , \tag{2.18}$$

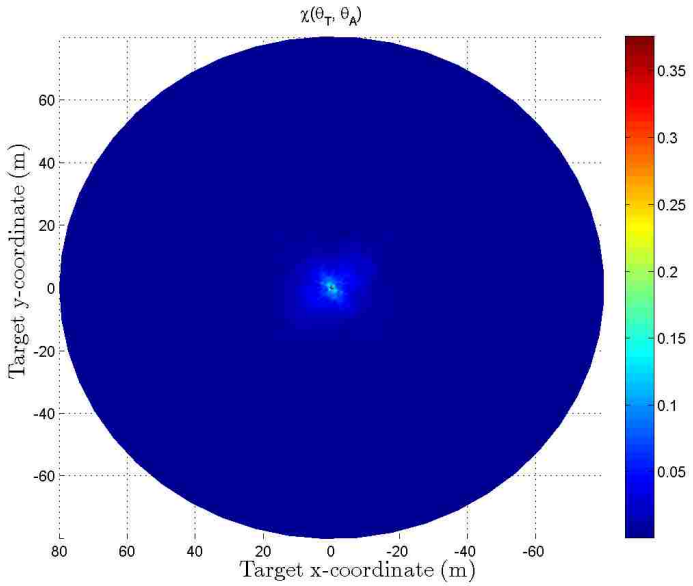
On the other hand, when the target is absent, we will have $T_{\mathrm{Ab}}(\boldsymbol{\theta}_{\mathrm{A}}) = \mathbf{w}^{\dagger}\mathbf{D}(\boldsymbol{\theta}_{\mathrm{A}})\mathbf{w}$. Let $\eta(\boldsymbol{\theta}_{\mathrm{A}})$ be the detection threshold chosen to achieve the required false alarm rate. Then our proposed AF is given as

$$\chi(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}}) = \Pr\left\{T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}},\boldsymbol{\theta}_{\mathrm{A}}) > \eta(\boldsymbol{\theta}_{\mathrm{A}})\right\} \tag{2.19}$$

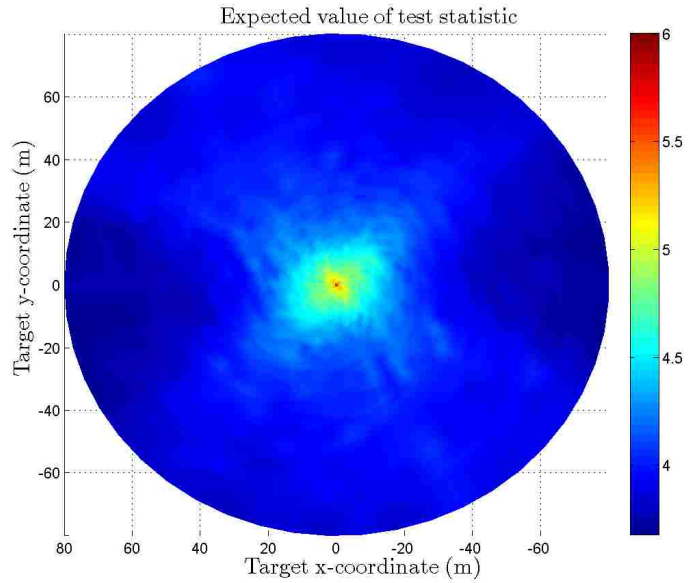
Since $\boldsymbol{\alpha}$ and $\mathbf{w}$ are zero mean Gaussian random vectors, $T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}})$ is distributed as a quadratic form in Gaussian random variables [34]. In general, if T_1 and T_2 are Gaussian quadratic forms, then $\mathrm{E}[T_1] > \mathrm{E}[T_2]$ does not necessarily imply that $\Pr\{T_1 > \eta\} > \Pr\{T_2 > \eta\}$. For example, if $T_1 = 1.9|x_1|^2 + 0.2|x_2|^2$ and $T_2 = |x_1|^2 + |x_2|^2$, where x_1, x_2 are independently distributed as $C\mathcal{N}(0,1)$, then $E[T_1] = 2.1 > 2 = E[T_2]$, while $\Pr\{T_1 > 1\} = 0.66 < 0.74 = \Pr\{T_2 > 1\}$. Therefore no reduction of the form of (2.16) is possible under non-coherent processing.

2.4 Simulation Results

For our simulations, we considered a radar with 6 transmitters and 6 receivers that were placed randomly on the circumference of a circle of radius 2000 m whose centre coincided with the origin of our coordinate system. A certain class of waveforms referred to as orthogonal phase coded transmit waveforms were used for $s_m(t)$. The target's true position $\boldsymbol{\theta}_{\mathrm{T}}$ was fixed at the origin. The channel covariance matrix $\boldsymbol{\Sigma}_{\alpha}$ was assumed to be an identity matrix. The required false alarm rate was set to $\eta = 10^{-4}$, and the SNR was set to 6 dB. The resulting probability of detection was $P_D = 0.38$. Figs. 2.1a and 2.1b show the ambiguity plots obtained using the proposed AF and the AF definition of [31] respectively. In these plots we observed multiple instances where for two assumed parameter values $\boldsymbol{\theta}_{\mathrm{A}}^{(1)}$ and $\boldsymbol{\theta}_{\mathrm{A}}^{(2)}$, we had $\mathrm{E}\left[T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}^{(1)})\right] < \mathrm{E}\left[T_{\mathrm{Pr}}(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}^{(2)})\right]$, while $\chi(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}^{(1)}) > \chi(\boldsymbol{\theta}_{\mathrm{T}}, \boldsymbol{\theta}_{\mathrm{A}}^{(2)})$. This illustrates that optimizing the proposed AF may lead to waveforms that are significantly different from those obtained using the the AF of [31].



(a) Proposed ambiguity function



(b) Ambiguity function definition of [31]

Figure 2.1: Plots of ambiguity functions

Chapter 3

Waveform Correlation Matrix Design for MIMO radar

3.0.1 Problem Motivation

The performance of any MIMO radar is mainly dependent on the following factors:

(a) The placement of transmit and receive antennas,

(b) The choice of the transmit waveforms,

(c) The nature of clutter-plus noise at the receivers, and

(d) The receiver processing algorithms.

In MIMO radar literature, typically only a few fixed choices for factors (a)-(c) are studied. In this chapter, we study the problem of target detection under arbitrary values for factors (a)-(c), while assuming optimum processing for factor (d).

When the transmit and receiver antennas may have arbitrary separations, the resulting channel gains along different transmit-receive paths will also have arbitrary correlations with one another. Our study assumes statistical knowledge of the channel gains, which is a more relaxed assumption than in studies where complete channel knowledge is assumed. In practice, this statistical knowledge (the covariance matrix of Gaussian channel gains and the clutter-plus-noise covariance matrix) has to be obtained via estimation techniques and used in conjugation with detection techniques such as the Generalized Likelihood Ratio test (GLRT). Our study essentially provides upper bounds on detection performance in this scenario.

For a MIMO radar with M transmitter and N receivers, a single $M \times M$ zero-lag waveform correlation matrix is sufficient to specify the detector when the transmit and receive antennas are closely spaced. In the general case, we show that for arbitrary antenna separations, the detector depends on the correlations between the received signals, i.e. the transmitted signals shifted by delays corresponding to a specific target location. Hence, N received waveform correlation matrices (each of size $M \times M$ with appropriate delays) are required to fully specify the detector for a specific target location.

With regard to the additive clutter-plus-noise occurring at the receiver, we assume a general scenario in this work, wherein the clutter might potentially be correlated along time and across different receive antennas, while the thermal receiver noise is temporally white and uncorrelated across receivers. The special case where the clutter-plus-noise as a whole is spatial and temporal white is also studied in detail, since it enables significantly simplifications in our analysis.

Thus, assuming that the channel covariance matrix and the waveform correlation matrix are arbitrary (but known) positive semidefinite matrices, we derive closed form expressions for the optimum Neyman-Pearson detector using low-rank matrix decomposition techniques.

While the detector expression in the case of spatially and temporally correlated clutter-plus-noise requires the computation of certain matrix decompositions, we show that this requirement can be eliminated in the case of spatially and temporally correlated white clutter-plus-noise. This is one of the main results of this work.

In this chapter, we further derive an expression for the detector SNR of a MIMO radar as a function of the channel covariance matrix, waveform correlation matrix and the input SNR. We show that this general expression reduces to known expressions for detector SNR for specific radar configurations as given in [10]. Further, we use this expression to study the nature of the waveform correlation matrices that maximize the detector SNR under various configurations at different input SNRs. Since it is difficult to find in closed form the waveform correlation matrix that maximizes the detector SNR for an arbitrary input SNR and channel covariance matrix, we use a stochastic optimization technique (the genetic algorithm) to find good sub-optimal solutions. The change in the nature of the optimized waveform correlation matrices at different input SNRs is also studied.

In order evaluate the performance of the detector, we use the Swerling-I extended target model discussed in [10] to generate channel covariance matrices as a function of target and antenna locations. In this model, the target is assumed to have a rectangular cross section composed of an infinite number of isotropic point scatterers. When the transmit/receive antennas have large separations relative to the target distance, an identity channel covariance matrix results under this model (which implies that all transmit-receive paths are uncorrelated), while small spacings result in an unit-rank channel covariance matrix (implying that all the transmit-receive paths are fully correlated). We also introduce a new scalar measure, termed the *normalized eigenvariance*, that can be used to express how close the channel covariance matrix corresponding to an arbitrary placement of antennas is to either of these two

extremes.

A short word on notation. Throughout this chapter, for a matrix $\mathbf{A}$, we use $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^\dagger$ and $\mathbf{A}^{1/2}$ to represent its transpose, conjugate, conjugate transpose and Hermitian square root respectively. Also, vec$\{\mathbf{A}\}$ is used to denote the vector obtained by stacking the columns of $\mathbf{A}$ below one another, while diag $\{\mathbf{a}\}$ denotes a diagonal matrix whose diagonal elements are taken from the vector $\mathbf{a}$. Further, $\circ$ and $\otimes$ are used refer to the Hadamard and Kronecker product operators respectively. Finally, $\delta(.)$ will be used to refer to the Dirac delta function.

3.0.2 System Model

We begin by considering a general MIMO radar model for any antenna placement. Let the radar contain M transmit and N receive antennas that are isotropic in nature[1]. Denote the locations of the of the m^{th} transmitter and n^{th} receiver on a 2-D plane[2] as $\mathbf{p}_{\text{t},m} = (x_{\text{t}m}, y_{\text{t}m})$ and $\mathbf{p}_{\text{r},n} = (x_{\text{r}n}, y_{\text{r}n})$ respectively. Further, denote the waveform transmitted by the m^{th} transmitter in complex baseband notation as $\sqrt{E_s}s_m(t)$ $(0 \leq t \leq T_s)$, where E_s is the total transmitted energy and $\sum_{m=1}^{M} \int_0^{T_s} |s_m(t)|^2 \, dt = 1$. Note that this implies a total transmit power constraint, but individual transmit powers need not be equal. Suppose that we are interested in testing for the presence of a target at a location $\mathbf{p}$ on the same 2-D plane. Let $\mathcal{H}_1$ and $\mathcal{H}_0$ denote hypotheses corresponding to the absence or presence of a target at this location, respectively. Define $\tau_{m,n} = (d_{\text{t}m} + d_{\text{r}n})/c$, where $d_{\text{t}m} = \|\mathbf{p}_{\text{t},m} - \mathbf{p}\|$, $d_{\text{r}n} = \|\mathbf{p}_{\text{r},n} - \mathbf{p}\|$ and c denotes the speed of light. We model the signal arriving at the n^{th} receiver (in complex

[1] Our analysis can be easily extended to accommodate arbitrary per-antenna beampatterns.

[2] For simplicity, we assume the target, transmitters and receivers all lie in the same 2-D plane. Extensions to 3-D models are straightforward.

baseband notation) as

$$r_n(t) = \begin{cases} w_n(t) & \mathcal{H}_1 \\ \sqrt{E_s}\sum_{m=1}^{M} \alpha_{m,n} s_m(t-\tau_{m,n}) + w_n(t) & \mathcal{H}_0 \end{cases}$$

where $\alpha_{m,n}$ models the combined attenuation due to path loss and reflection via the target for the signal transmitted by the m^{th} transmitter, and $w_n(t)$ is a zero-mean stationary Gaussian random process that models the sum of clutter and thermal noise at the n^{th} receiver. For simplicity, Doppler shift due to target motion is not introduced in this model; it is assumed to be separately estimated and accounted for.

A first step prior to detection is to reduce the continuous time received signals to a finite number of observation variables, via matched filtering. We thus obtain

$$\begin{aligned} r_{m,n} &= \int_{\tau_{m,n}}^{\tau_{m,n}+T_s} r_n(t) s_m^*(t-\tau_{m,n})\mathrm{d}t \;\; = \int_0^{T_s} r_n(t+\tau_{m,n}) s_m^*(t)\mathrm{d}t \\ &= \begin{cases} w_{m,n} & \mathcal{H}_1 \\ \sqrt{E_s}\sum_{\tilde{m}=1}^{M} \alpha_{\tilde{m},n}\xi_{m,n,\tilde{m}} + w_{m,n} & \mathcal{H}_0 \end{cases} \end{aligned}$$

where

$$\xi_{m,n,\tilde{m}} = \int_0^{T_s} s_{\tilde{m}}(t-\tau_{\tilde{m},n}+\tau_{m,n}) s_m^*(t)\mathrm{d}t \tag{3.1}$$

and

$$w_{m,n} = \int_0^{T_s} w_n(t+\tau_{m,n}) s_m^*(t)\mathrm{d}t \;. \tag{3.2}$$

In order to stack these matched filter outputs, we define the vectors

$$\mathbf{r} = [\ r_{1,1}\ r_{2,1}\ \cdots\ r_{M,1}\ r_{1,2}\ \cdots\ r_{M,N}\ ]^T$$

$$\boldsymbol{\alpha} = [\ \alpha_{1,1}\ \alpha_{2,1}\ \cdots\ \alpha_{M,1}\ \alpha_{1,2}\ \cdots\ \alpha_{M,N}\ ]^T$$

$$\mathbf{w} = [\ w_{1,1}\ w_{2,1}\ \cdots\ w_{M,1}\ w_{1,2}\ \cdots\ w_{M,N}\ ]^T$$

and the matrices

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\xi}_N \end{bmatrix}, \quad \boldsymbol{\xi}_n = \begin{bmatrix} \xi_{1,n,1} & \cdots & \xi_{1,n,M} \\ \vdots & \ddots & \vdots \\ \xi_{M,n,1} & \cdots & \xi_{M,n,M} \end{bmatrix}.$$

Here $\xi_{m,n,\tilde{m}}$ can be interpreted as the cross-correlation between $s_{\tilde{m}}(t - \tau_{\tilde{m},n})$ (the delayed version of the $\tilde{m}^{\text{th}}$ transmitter's signal which arrives at the n^{th} receiver), and the replica signal $s_m^*(t)$, computed at an appropriate delay $\tau_{m,n}$. We hence refer to $\boldsymbol{\xi}$ as the waveform correlation matrix. Note that due to our initial assumption of normalized total waveform energy, we have $\text{Tr}\{\boldsymbol{\xi}_n\} = 1$ for all n. Further, in order to aid in our mathematical analysis, we assume that $\boldsymbol{\xi}$ is Hermitian symmetric, i.e. $\xi_{m,n,\tilde{m}} = \xi_{\tilde{m},n,m}$ for all $m, n, \tilde{m}$. One way to ensure this property is by imposing the following constraint on the waveforms

$$s_m(t) = 0 \quad \text{for} \quad t \in (-\Delta\tau_{\max}, 0) \cup (T_s, T_s + \Delta\tau_{\max}),$$

where $\Delta\tau_{\max} = \max\limits_{m,n,m'} |\tau_{m,n} - \tau_{m',n}|$. This constraint basically means that a gap of $\Delta\tau_{\max}$ should be present between consecutive transmissions of the transmit waveforms.

Using the above notation, the matched filter outputs can be jointly represented as

$$\mathbf{r} = \begin{cases} \mathbf{w} & \mathcal{H}_1 \\ \sqrt{E_s}\boldsymbol{\xi}\boldsymbol{\alpha} + \mathbf{w} & \mathcal{H}_0 \end{cases} \tag{3.3}$$

Observe that $\mathbf{w}$ and $\boldsymbol{\alpha}$ are both zero-mean Gaussian random vectors due to our initial assumptions. Thus $\mathbf{r}$ is also distributed as a zero-mean Gaussian random vector under both hypotheses, with covariance matrix

$$\mathrm{E}\left[\mathbf{r}\mathbf{r}^{\dagger}\right] = \begin{cases} \sigma_w^2 \boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_0 & \mathcal{H}_1 \\ E_r \boldsymbol{\xi} \boldsymbol{\Sigma}_\alpha \boldsymbol{\xi}^{\dagger} + \sigma_w^2 \boldsymbol{\Sigma}_w = \boldsymbol{\Sigma}_1 & \mathcal{H}_0 \end{cases} \tag{3.4}$$

where

$$\sigma_w^2 = \mathrm{E}\left[\mathbf{w}^{\dagger}\mathbf{w}\right], \tag{3.5}$$

$$\boldsymbol{\Sigma}_w = \mathrm{E}\left[\mathbf{w}\mathbf{w}^{\dagger}\right] / \mathrm{E}\left[\mathbf{w}^{\dagger}\mathbf{w}\right], \tag{3.6}$$

$$E_r = E_s \mathrm{E}\left[\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}\right], \tag{3.7}$$

$$\boldsymbol{\Sigma}_\alpha = \mathrm{E}\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^{\dagger}\right] / \mathrm{E}\left[\boldsymbol{\alpha}^{\dagger}\boldsymbol{\alpha}\right] \ . \tag{3.8}$$

Here σ_w^2 and E_r respectively represent the total post-matched filtering energy in the noise plus clutter and transmit waveform components of the received signals. Further, the matrices $\boldsymbol{\Sigma}_\alpha$ and $\boldsymbol{\Sigma}_w$ represent the channel covariance matrix and the clutter plus noise covariance matrix respectively (with the matrix traces normalized to unity, i.e. $\mathrm{Tr}\left\{\boldsymbol{\Sigma}_\alpha\right\} = \mathrm{Tr}\left\{\boldsymbol{\Sigma}_w\right\} = 1$). Thus, while E_r captures information about the nominal path loss, the diagonal elements of $\boldsymbol{\Sigma}_\alpha$ capture information about the specific path loss (relative to nominal path loss) along each transmit-receive path. The off-diagonal elements of $\boldsymbol{\Sigma}_\alpha$ capture information about the correlation between different path gains. A similar interpretation applies to σ_w^2 and $\boldsymbol{\Sigma}_w$. We use such normalization in order to enable a study of detector performance with respect to the *receive SNR*, which we define as $\rho = E_r/\sigma_w^2$.

The nature of the optimum detector will depend on $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_\alpha$, which in turn depend on variables such as

(a) The positions of transmitters and receivers.

(b) The size and shape of the target.

(c) The duration and shape of the transmitted waveforms.

(d) The auto- and cross-correlation properties of clutter plus noise observed at different receivers.

We now briefly discuss the nature of this dependence.

3.0.3 Nature of Channel Covariance Matrix $\mathbf{\Sigma}_\alpha$

The channel covariance matrix $\mathbf{\Sigma}_\alpha$ is a function of the positions of the transmit and receive antennas, as well as the nature of the target. To characterize this dependence for our simulations, we use the Swerling-I target model, and assume (as in [10]) that the target has a rectangular cross-section of area $\Delta x \times \Delta y$, that is composed of an infinite number of random, isotropic and independent scatterers. A schematic representation of such a MIMO radar system is shown in Fig. 3.1. Let $\mathfrak{g}(x, y)$ denote the complex gain of the scatterer located at coordinates $(x_0 + x, y_0 + y)$. We assume that $\mathfrak{g}(x, y)$ is zero mean and that

$$\mathrm{E}\left[\mathfrak{g}(x,y)\mathfrak{g}^*(x',y')\right] = \frac{1}{\Delta x \Delta y}\delta(x-x')\delta(y-y')$$

where $\mathrm{E}[.]$ denotes expectation operation. Let f_c and λ_c represent the frequency and wavelength of the carrier signal that is modulated by the transmit waveforms. Further, let $\zeta_{t,m}$ and $\zeta_{r,n}$ denote the nominal path loss between the m^{th} transmitter and the target, and between the target and the n^{th} receiver respectively. Denote $\tau_{m,n}(\beta, \gamma)$ as the propagation delay along the path between m^{th} transmitter and n^{th} receiver via the scatterer located at coordinates

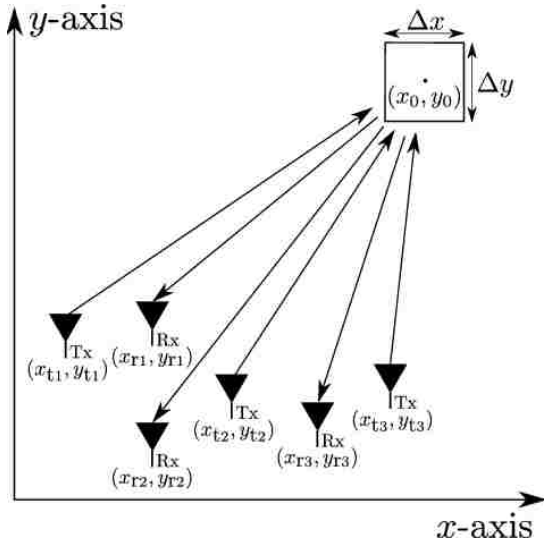


Figure 3.1: Schematic representation of the target, transmitters and receivers

$(x_0 + \beta, y_0 + \gamma)$. We assume that the transmitted signals is narrowband, in the sense that

$$\int_0^{T_s} s_m\big(t - \tau_{m,n}(\beta,\gamma)\big) s_{m'}\big(t - \tau_{m',n}\big)\, dt \;\approx \int_0^{T_s} s_m\big(t - \tau_{m,n}\big) s_{m'}\big(t - \tau_{m',n}\big)\, dt$$

where $\tau_{k,l} \triangleq \tau_{k,l}(0,0)$. Under these assumptions, we can show (following the derivation steps of [10]) that independent of the exact distribution of $\mathfrak{g}(x,y)$, the channel gains $\alpha_{m,n}$ will be zero mean complex Gaussian random variables, with mutual covariances given as

$$\begin{aligned} \mathrm{E}\left[\alpha_{m,n}\alpha^*_{m',n'}\right] = \;& \zeta_{t,m}\zeta_{r,n}\zeta_{t,m'}\zeta_{r,n'}\exp\left\{j2\pi f_c[\tau_{m,n} - \tau_{m',n'}]\right\} \\ & \times \mathrm{sinc}(\psi_x(m,n,m',n'))\mathrm{sinc}(\psi_y(m,n,m',n')) \end{aligned} \tag{3.9}$$

where

$$\psi_x(m,n,m',n') = \frac{\Delta x}{\lambda}\left[\frac{x_{\mathrm{t}m} - x_0}{d_{\mathrm{t}m}} - \frac{x_{\mathrm{t}m'} - x_0}{d_{\mathrm{t}m'}} + \frac{x_{\mathrm{r}n} - x_0}{d_{\mathrm{r}n}} - \frac{x_{\mathrm{r}n'} - x_0}{d_{\mathrm{r}n'}}\right]$$

and with $\psi_y(m,n,m',n')$ similarly defined using receive antenna coordinates. Now, define

$$\boldsymbol{\zeta}_t = \begin{bmatrix} \zeta_{t,1} & \cdots & \zeta_{t,M} \end{bmatrix}^T, \qquad \boldsymbol{\zeta}_r = \begin{bmatrix} \zeta_{r,1} & \cdots & \zeta_{r,N} \end{bmatrix}^T,$$

and let $\boldsymbol{\zeta} = \boldsymbol{\zeta}_r \otimes \boldsymbol{\zeta}_t$. Further define

$$\boldsymbol{\psi}_x(n_1,n_2) = \begin{bmatrix} \psi_x(1,n_1,1,n_2) & \cdots & \psi_x(1,n_1,M,n_2) \\ \vdots & \ddots & \vdots \\ \psi_x(M,n_1,1,n_2) & \cdots & \psi_x(M,n_1,M,n_2) \end{bmatrix},$$

$$\boldsymbol{\psi}_x = \begin{bmatrix} \psi_x(1,1) & \cdots & \psi_x(1,N) \\ \vdots & \ddots & \vdots \\ \psi_x(N,1) & \cdots & \psi_x(N,N) \end{bmatrix}$$

and similarly define $\boldsymbol{\psi}_y(n_1,n_2)$ and $\boldsymbol{\psi}_y$. Then we have

$$\mathrm{E}\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger\right] = ((\boldsymbol{\zeta}\circ\mathbf{b})(\boldsymbol{\zeta}\circ\mathbf{b})^\dagger)\circ\mathrm{sinc}(\boldsymbol{\psi}_x)\circ\mathrm{sinc}(\boldsymbol{\psi}_y) \tag{3.10}$$

where

$$\mathbf{b} = \exp\left\{ j2\pi f_c \begin{bmatrix} \tau_{1,1} & \tau_{2,1} & \cdots & \tau_{M,1} & \cdots & \tau_{M,N} \end{bmatrix}^T \right\},$$

with the $\mathrm{sinc}(\cdot)$ and $\exp\{\cdot\}$ functions applied elementwise on the matrix arguments. Further,

$$\mathrm{E}\left[\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}\right] = \mathrm{Tr}\left\{\mathrm{E}\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger\right]\right\} = \mathrm{Tr}\left\{(\boldsymbol{\zeta}\circ\mathbf{b})(\boldsymbol{\zeta}\circ\mathbf{b})^\dagger\right\} = \|\boldsymbol{\zeta}\circ\mathbf{b}\|^2 = \|\boldsymbol{\zeta}\|^2 \tag{3.11}$$

Using equations (3.10) and (3.11), the channel covariance matrix can be expressed as

$$\begin{aligned} \boldsymbol{\Sigma}_\alpha &= \mathrm{E}\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger\right] / \mathrm{E}\left[\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}\right] \\ &= \frac{1}{\|\boldsymbol{\zeta}\|^2}((\boldsymbol{\zeta}\circ\mathbf{b})(\boldsymbol{\zeta}\circ\mathbf{b})^\dagger)\circ\mathrm{sinc}(\boldsymbol{\psi}_x)\circ\mathrm{sinc}(\boldsymbol{\psi}_y) \end{aligned}$$

It has been shown (see [10] for more detailed derivations) that when the separations

between all the transmit antennas are small relative to the target distance, i.e. $d_{\mathrm{t}m} \approx d_{\mathrm{t}}$ and $\|\mathbf{p}_{\mathrm{t},m} - \mathbf{p}_{\mathrm{t},m'}\| \ll d_{\mathrm{t}}$ for all values of (m, m'), then

$$\frac{\Delta x}{\lambda}\left[\frac{x_{\mathrm{t}m} - x_0}{d_{\mathrm{t}m}} - \frac{x_{\mathrm{t}m'} - x_0}{d_{\mathrm{t}m'}}\right] \approx 0 \ ,$$
$$\frac{\Delta y}{\lambda}\left[\frac{y_{\mathrm{t}m} - y_0}{d_{\mathrm{t}m}} - \frac{y_{\mathrm{t}m'} - y_0}{d_{\mathrm{t}m'}}\right] \approx 0 \ .$$

Other the other hand, when the separations between all the transmit antennas are comparable to the target distance, then

$$\frac{\Delta x}{\lambda}\left[\frac{x_{\mathrm{t}m} - x_0}{d_{\mathrm{t}m}} - \frac{x_{\mathrm{t}m'} - x_0}{d_{\mathrm{t}m'}}\right] \gg 1 \ ,$$
$$\frac{\Delta y}{\lambda}\left[\frac{y_{\mathrm{t}m} - y_0}{d_{\mathrm{t}m}} - \frac{y_{\mathrm{t}m'} - y_0}{d_{\mathrm{t}m'}}\right] \gg 1 \ .$$

Similar comments can be made on the terms in $\psi_x(m, n, m', n')$ and $\psi_y(m, n, m', n')$ that depend on receive antenna coordinates.

3.0.4 Nature of Clutter-plus-Noise Covariance Matrix $\mathbf{\Sigma}_w$

Let $w_n(t) = \tilde{w}_n(t) + \hat{w}_n(t)$, where $\tilde{w}_n(t)$ represents receiver thermal noise while $\hat{w}_n(t)$ represents clutter. In typical practical scenarios, $\tilde{w}_n(t)$ is temporally and spatially (i.e. across different antennas) uncorrelated, while $\hat{w}_n(t)$ can be temporally and spatially correlated. Further, $\tilde{w}_n(t)$ and $\hat{w}_n(t)$ can be assumed independent of each other. Let $\mathrm{E}\left[\tilde{w}_{n_1}(t)\tilde{w}^*_{n_2}(t-\tau)\right] = \tilde{\sigma}^2_{n_1}\delta(\tau)\delta[n_1 - n_2]$ and define $c_{m_1,n_1,m_2,n_2} = \mathrm{E}\left[\hat{w}_{n_1}(t_1 + \tau_{m_1,n_1})\hat{w}^*_{n_2}(t_2 + \tau_{m_2,n_2})\right]$. Then one can write

$$\mathrm{E}\left[w_{n_1}(t_1 + \tau_{m_1,n_1})w^*_{n_2}(t_2 + \tau_{m_2,n_2})\right] = \tilde{\sigma}^2_{n_1}\delta(t_1 + \tau_{m_1,n_1} - t_2 - \tau_{m_2,n_2})\delta[n_1 - n_2] + c_{n_1,n_2}$$

From eq. (3.2), we have (under the assumption of stationary noise and clutter)

$$\mathrm{E}\left[w_{m_1,n_1}w^*_{m_2,n_2}\right] = \int_0^{T_s}\int_0^{T_s} s^*_{m_1}(t_1)s_{m_2}(t_2)\mathrm{E}\left[w_{n_1}(t_1+\tau_{m_1,n_1})w^*_{n_2}(t_2+\tau_{m_2,n_2})\right]\mathrm{d}t_1\mathrm{d}t_2$$
$$= \tilde{\sigma}^2_{n_1}\delta[n_1-n_2]\xi_{m_1,n_1,m_2} + c_{m_1,n_1,m_2,n_2}$$

and hence

$$\mathrm{E}\left[\mathbf{w}\mathbf{w}^\dagger\right] = \mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C}$$

where $\mathbf{G} = \mathrm{diag}\left\{[\ \tilde{\sigma}_1\ \cdots\ \tilde{\sigma}_N]^T\right\} \otimes \mathbf{I}_{MN}$, and

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_{1,1} & \cdots & \mathbf{c}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{c}_{N,1} & \cdots & \mathbf{c}_{N,N} \end{bmatrix}, \quad \text{where} \quad \mathbf{c}_{n_1,n_2} = \begin{bmatrix} c_{1,n_1,1,n_2} & \cdots & c_{1,n_1,M,n_2} \\ \vdots & \ddots & \vdots \\ c_{M,n_1,1,n_2} & \cdots & c_{M,n_1,M,n_2} \end{bmatrix}.$$

3.0.5 Neyman-Pearson Optimum Detection under General MIMO Radar Model

In this section, we first derive the Neyman-Pearson detector for the general MIMO radar hypothesis testing problem of (3.4) (restated here for convenience),

$$\mathrm{E}\left[\mathbf{r}\mathbf{r}^\dagger\right] = \begin{cases} \mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C} = \boldsymbol{\Sigma}_0 & \mathcal{H}_1 \\ E_r\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi} + \mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C} = \boldsymbol{\Sigma}_1 & \mathcal{H}_0 \end{cases} \tag{3.12}$$

To this end, we state and prove the following theorem.

Theorem 1. *Let rank$\{\boldsymbol{\Sigma}_0\} = P$, hence admitting the decomposition $\boldsymbol{\Sigma}_0 = \mathbf{V}\mathbf{V}^\dagger$, where $\mathbf{V}$ is a matrix of size $MN \times P$, and let $\tilde{\mathbf{r}} = (\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger\mathbf{r}$. Further, for any vector $\mathbf{x}$, define $\|\mathbf{x}\|_\mathbf{A} = \mathbf{x}^\dagger\mathbf{A}\mathbf{x}$ as the vector norm under matrix $\mathbf{A}$. Then the Neyman-Pearson optimum test statistic for the detection problem of (3.12) can be given as $T = \|\mathbf{r}\|_\mathbf{D}$, where $\mathbf{D} =$*

$\mathbf{V}[(\mathbf{V}^\dagger\mathbf{V})^{-2} - (\mathbf{V}^\dagger\mathbf{\Sigma}_1\mathbf{V})^{-1}]\mathbf{V}^\dagger$.

Proof. See Appendix 3.1.1. □

We now consider the special case where the clutter-plus-noise is spatially and temporally white i.e. $\mathbf{C} = 0$.

Theorem 2. *For the hypothesis testing problem of (3.12), the Neyman-Pearson optimum test statistic when* $\mathbf{C} = 0$ *is given as* $T = ||\mathbf{r}||_{\mathbf{D}}$*, where* $\mathbf{D} = (\mathbf{I}_{MN} + \rho\mathbf{P}^{-2}\mathbf{\Sigma}_\alpha\boldsymbol{\xi})^{-1}\mathbf{P}^{-2}\mathbf{\Sigma}_\alpha\mathbf{P}^{-2}$ *and* $\mathbf{P} = \sigma_w^{-1}\mathbf{G}$.

Proof. See Appendix 3.1.2. □

An interesting property of the detector of Theorem 2 is that it while the decompositions $\mathbf{\Sigma}_0 = \mathbf{V}\mathbf{V}^\dagger$ and $E_r\mathbf{G}^{-2}\mathbf{\Sigma}_\alpha\mathbf{G}^{-2} = \mathbf{U}\mathbf{U}^\dagger$ are used in the intermediate steps of the theorem, they are completely eliminated from the final expression for the test statistic. Also, the $\mathbf{G}^{-1}$ term in the expression for $\mathbf{D}$ is simple to compute since $\mathbf{G}$ is a diagonal matrix. Thus, Theorem 2 provides us for a simple general expression for the Neyman-Pearson optimum detector for a wide variety of radars, allowing us to study their performance under a general framework.

Under the conditions of either Theorem 1 or Theorem 2 described above, the test statistic T is a quadratic form in Gaussian random variables (since $\tilde{\mathbf{r}}$ and $\mathbf{r}$ are both Gaussian random vectors). Further, the form of the test is

$$T \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \delta \tag{3.13}$$

where δ represents the detection threshold. Recently, a closed-form expression for the CDF of such random variables has been presented [35]. We use the same method to compute the

CDF of T for our simulations. The inverse CDF of T, which is required to compute the threshold δ, can be evaluated by using any numerical search routine (such as binary search) over the CDF of T.

3.0.6 Detector SNR

Ideally, a radar's transmit waveforms should be optimized to maximize the probability of detection, given constraints on the transmit power, the received SNR, the channel and noise covariance matrices, and the maximum probability of false alarm. However, this is a difficult problem because of the non-linear relationships between the probability of detection and the design constraints. Hence, in order to facilitate analysis in this direction, simpler heuristic performance measures are often used. We now study one such measure, known as the detector SNR (or the *deflection coefficient*), which was used in [10] to compare the detection performance of different radar configurations. For a Neyman-Pearson statistic T, the detector SNR is computed as

$$\beta = \frac{\left|\mathrm{E}(T|\mathcal{H}_0) - \mathrm{E}(T|\mathcal{H}_1)\right|^2}{\frac{1}{2}[\mathrm{Var}(T|\mathcal{H}_0) + \mathrm{Var}(T|\mathcal{H}_1)]} \ . \tag{3.14}$$

The detector SNR essentially measures the separation between the probability distributions of the test statistic under the two hypotheses. Such a performance measure is useful since it allows the detection performance to be expressed through a single scalar value, and does not require a false alarm rate to be specified. In this section, we shall derive a simplified expression for this measure under our MIMO radar model, with two simplifying assumptions:

(a) The noise plus clutter is spatially and temporally uncorrelated ($\mathbf{C} = 0$).

(b) The noise plus clutter random processes at all receivers have equal power, i.e. $\tilde{\sigma}_i^2 = \tilde{\sigma}^2$ for all $i = 1, \cdots, N$.

Lemma 1. *For a MIMO radar with* $\mathbf{C} = 0$ *and* $\tilde{\sigma}_i^2 = \tilde{\sigma}^2$ *for all* $i = 1, \cdots, N$*, the detector's SNR is given as*

$$\beta = \frac{2\left(\mathrm{Tr}\left\{\mathbf{U} - 2\mathbf{I} + \mathbf{U}^{-1}\right\}\right)^2}{\mathrm{Tr}\left\{(\mathbf{U} - \mathbf{I})^2 + (\mathbf{U}^{-1} - \mathbf{I})^2\right\}} \tag{3.15}$$

where $\mathbf{U} = \rho N \boldsymbol{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{I}_{MN}$ *and* $\mathrm{Tr}\{.\}$ *denotes the trace operation.*

Proof. See Appendix 3.1.3. □

To demonstrate the generality as well as the correctness of our result, in Appendix 3.1.4 we show that the expression in (3.15) reduces exactly to specific expressions for detector SNR derived in [10] under three special detection scenarios.

The expression for detector SNR obtained from the above lemma can be used to find the waveform correlation matrix that maximizes detection performance for a given MIMO radar configuration. Specifically, we can find the matrix $\boldsymbol{\xi}$ that maximizes β given ρ and $\boldsymbol{\Sigma}_\alpha$. To this end, we first study the case of high received SNR under widely separated transmit and receive antennas.

Lemma 2. *Assume a MIMO radar with widely separated transmitters and receiver* ($\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}$)*. If*

$$\rho \gg \max(\sqrt{M^3 N}, M^2)$$

then the detector SNR

$$\beta = \frac{2\left(\mathrm{Tr}\left\{\mathbf{U} - 2\mathbf{I} + \mathbf{U}^{-1}\right\}\right)^2}{\mathrm{Tr}\left\{(\mathbf{U} - \mathbf{I})^2 + (\mathbf{U}^{-1} - \mathbf{I})^2\right\}}$$

(where $\mathbf{U} = \rho N \boldsymbol{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{I}_{MN}$*) will be maximum when* $\boldsymbol{\xi} = M^{-1}\mathbf{I}_{MN}$*.*

Proof. See Appendix 3.1.5. □

The solution $\boldsymbol{\xi} = M^{-1}\mathbf{I}_{MN}$ can be interpreted to mean that the delayed versions of the transmitted signals arriving at each receiver must be mutually orthogonal. Note also that the resultant maximum value of β will be $\beta_{Max} = 2MN$.

Further, if we assume $\tau_{m,n} \approx \tau_{\tilde{m},n}$ for all $m, \tilde{m}, n$, and hence that $\boldsymbol{\xi}_n \approx \hat{\boldsymbol{\xi}}$ for all n (see Appendix for more details), where $\hat{\boldsymbol{\xi}}$ represents the zero-lag correlation matrix of the transmit waveforms, then we have $\boldsymbol{\xi} = \mathbf{I}_N \otimes \hat{\boldsymbol{\xi}}$. Hence the solution $\boldsymbol{\xi} = M^{-1}\mathbf{I}_{MN}$ will imply $\hat{\boldsymbol{\xi}} = M^{-1}\mathbf{I}_M$, i.e. the transmit waveforms must be orthogonal (at zero lag).

3.0.7 Simulation Results

In this Section, we shall optimize the waveform correlation matrix $\boldsymbol{\xi}$ to maximize the detector SNR under specific practical scenarios. For simplicity, we make assumptions (a) and (b) listed in the Appendix. Then our problem of interest is

$$\begin{aligned} \max_{\hat{\boldsymbol{\xi}}} \quad & \beta = \frac{2\left(\mathrm{Tr}\left\{\mathbf{U} - 2\mathbf{I} + \mathbf{U}^{-1}\right\}\right)^2}{\mathrm{Tr}\left\{(\mathbf{U}-\mathbf{I})^2 + (\mathbf{U}^{-1}-\mathbf{I})^2\right\}} \\ \text{s.t.} \quad & \mathbf{U} = \tilde{\rho}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi} + \mathbf{I}_{MN}, \quad \boldsymbol{\xi} = \mathbf{I}_N \otimes \hat{\boldsymbol{\xi}}, \quad \hat{\boldsymbol{\xi}} \succeq 0 \end{aligned} \tag{3.16}$$

We consider the above optimization problem under the following four scenarios (refer to the appendix for details about how the form of $\boldsymbol{\Sigma}_\alpha$ is obtained):

(a) Large transmit and receive antenna separations ($\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}$)

(b) Small transmit and receive antenna separations ($\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_{MN}$)

(c) Large transmit antenna separations, small receive antenna separations ($\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_N \otimes \mathbf{I}_M$)

(d) Intermediate transmit and receive antenna separations ($\boldsymbol{\Sigma}_\alpha$ = a randomly generated positive semidefinite matrix)

Note here that $\mathbf{1}_L$ refer to a matrix of size $L \times L$ that contains all ones. While we obtained a closed form solution for case (a) under the high ρ regime in Section 3.0.5, is not easy to obtain general closed form solutions under scenarios (a) - (d) for arbitrary ρ values. Hence, in this section we present results obtained using a stochastic optimization technique (the genetic algorithm) to optimize $\hat{\boldsymbol{\xi}}$ under scenarios (a) - (d), under a range of values of ρ.

Since $\hat{\boldsymbol{\xi}}$ is a $M \times M$ matrix, it is difficult to visualize trends in the nature of the optimized values of $\hat{\boldsymbol{\xi}}$ as ρ is varied. Hence, we formulated a novel measure, which we term the *normalized eigenvariance*, that allows us to characterize the nature of $\hat{\boldsymbol{\xi}}$ with a single scalar value. The normalized eigenvariance $\sigma^2_{en}(\mathbf{A})$ for any positive semidefinite matrix $\mathbf{A}$ of size $L \times L$ is defined as

$$\begin{aligned}\sigma^2_{en}(\mathbf{A}) &= \left[\frac{L}{L-1}\right] \frac{\sum_{i=1}^{L}(\mu_i - \bar{\mu})^2}{\left(\sum_{i=1}^{L}\mu_i\right)^2} \\ &= \left[\frac{L}{L-1}\right] \left[\frac{\mathrm{Tr}\left\{\mathbf{A}^2\right\}}{\left(\mathrm{Tr}\left\{\mathbf{A}\right\}\right)^2} - \frac{1}{L}\right]\end{aligned}$$

where $\{\mu_i\}_{i=1}^{L}$ represent the L eigenvalues of $\mathbf{A}$ (this may include zero or repeated eigenvalues) and $\bar{\mu} = \sum_{i=1}^{L}\mu_i/L$ is the mean of its eigenvalues. $\sigma^2_{en}(\mathbf{A})$ basically represents the variance of the eigenvalues of $\mathbf{A}$, divided by the maximum variance possible for any positive semidefinite matrix that has the same trace as $\mathbf{A}$. Thus, $\sigma^2_{en}(\mathbf{A})$ measures the normalized spread of the eigenvalues of $\mathbf{A}$. It is easy to show that $\sigma^2_{en}(\mathbf{A})$ has a maximum value of 1 (that occurs when $\mathbf{A}$ is unit rank and hence has only one non-zero eigenvalue), and a minimum value of 0 (that occurs when $\mathbf{A}$ is a constant times an identity matrix and hence has all equal eigenvalues). When $\mathbf{A}$ is rank deficient, or full rank but with unequal eigenvalues, then $\sigma^2_{en}(\mathbf{A})$ lies between 0 and 1. When used in the context of the zero-lag transmit waveform correlation matrix $\hat{\boldsymbol{\xi}}$, this measure tells us how close the transmit waveforms corresponding to a given value of $\hat{\boldsymbol{\xi}}$ are to

orthogonal waveforms (where $\sigma^2_{en}(\hat{\boldsymbol{\xi}}) = 0$) or fully correlated waveforms (where $\sigma^2_{en}(\hat{\boldsymbol{\xi}}) = 1$). Similarly, $\sigma^2_{en}(\boldsymbol{\Sigma}_\alpha)$ tells how close the channel gains corresponding a given channel covariance matrix $\boldsymbol{\Sigma}_\alpha$ are to being either independent $\sigma^2_{en}(\boldsymbol{\Sigma}_\alpha) = 0$, or fully correlated $\sigma^2_{en}(\boldsymbol{\Sigma}_\alpha) = 1$.

In our simulations, we assumed $M = 4$ transmit and $N = 4$ receive antennas, and for each choice of $\boldsymbol{\Sigma}_\alpha$, we varied $\tilde{\rho} = \rho/M$ between -10 dB and 20 dB. The normalized eigenvariance, detector SNR and probability of detection P_D (at a false alarm rate of $P_{FA} = 10^{-6}$) corresponding to the optimal waveform correlation matrix were then plotted as a function of input SNR.

The results for case (a), i.e. large transmit and receive antenna separations, are shown in Fig. 3.2. It can be seen that the stochastic optimization routine results in fully correlated waveforms ($\sigma^2_{\text{en}}(\hat{\boldsymbol{\xi}}) = 1$) below $\tilde{\rho} = 2$ dB and orthogonal waveforms above $\tilde{\rho} = 12$ dB. Between these two thresholds, there is a transition region where partially correlated waveforms are chosen (as evidenced by the normalized eigenvariance curve). The P_D curves also exhibit a similar behaviour, thus showing the detector SNR is indeed a good measure of detection performance, while having the advantage of being easy to compute relative to P_D.

Next, the results for case (b) (closely spaced antennas) and case (c) (widely spaced transmitters and closely spaced receivers) are plotted in Figs. 3.3 and 3.4 respectively. For these two scenarios, the stochastic optimization routine results in fully correlated waveforms throughout the chosen input SNR range. For case (d), we randomly selected $\boldsymbol{\Sigma}_\alpha$ as a positive semidefinite matrix with $\sigma^2_{en}(\boldsymbol{\Sigma}_\alpha) = 0.5$, corresponding to a scenario in between cases (a) and (b). Here, the stochastic optimization routine results in (Fig. 3.5) partially correlated waveforms above $\tilde{\rho} = 4$ dB, and fully correlated waveforms are optimal below $\tilde{\rho} = 4$ dB.

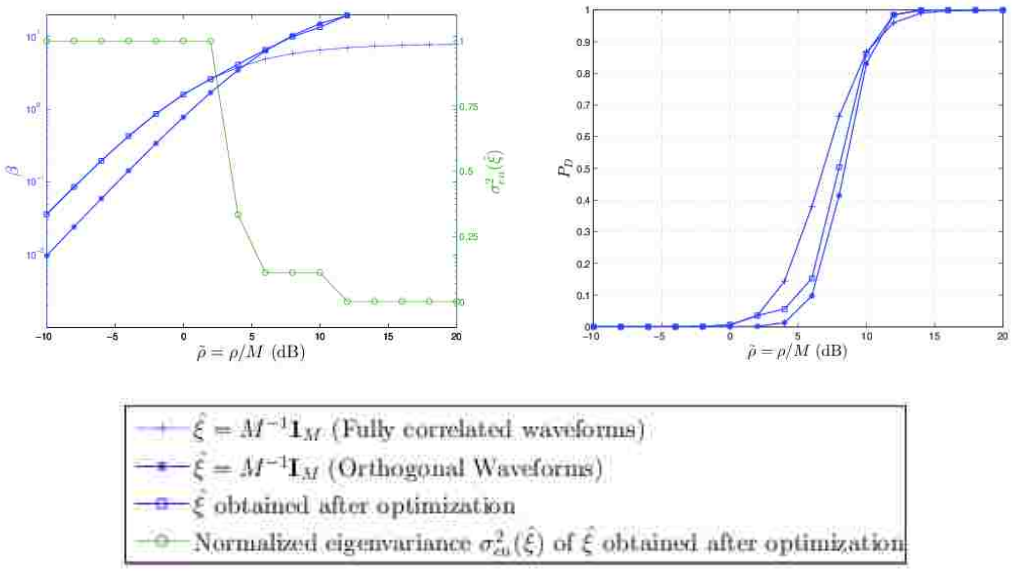


Figure 3.2: Normalized eigenvariance, detector SNR and P_D (at $P_{FA} = 10^{-6}$) using the $\hat{\boldsymbol{\xi}}$ matrix obtained after optimization at different SNRs for $\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}$

3.1 Appendix

3.1.1 Proof of Theorem 1

Proof. We begin by showing that while $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ may be rank deficient, they satisfy $\text{rank}(\boldsymbol{\Sigma}_0) = \text{rank}(\boldsymbol{\Sigma}_1)$. To this end, we use the kernel operator, defined for any $n \times n$ matrix $\mathbf{A}$ as

$$\ker(\mathbf{A}) = \{\mathbf{x} \in \mathbb{C}^n : \mathbf{A}\mathbf{x} = 0\}$$

We note that since $\mathbf{G} \succ 0$, and $\boldsymbol{\Sigma}_\alpha \succeq 0$, we have $\ker(\mathbf{G}\boldsymbol{\xi}\mathbf{G}) = \ker(\boldsymbol{\xi})$ and $\ker(\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi}) \subseteq \ker(\boldsymbol{\xi})$. Further, given matrices $\mathbf{A}, \mathbf{B} \succeq 0$, we can write

$$\ker(\mathbf{A} + \mathbf{B}) = \ker(\mathbf{A}) \cap \ker(\mathbf{B})$$

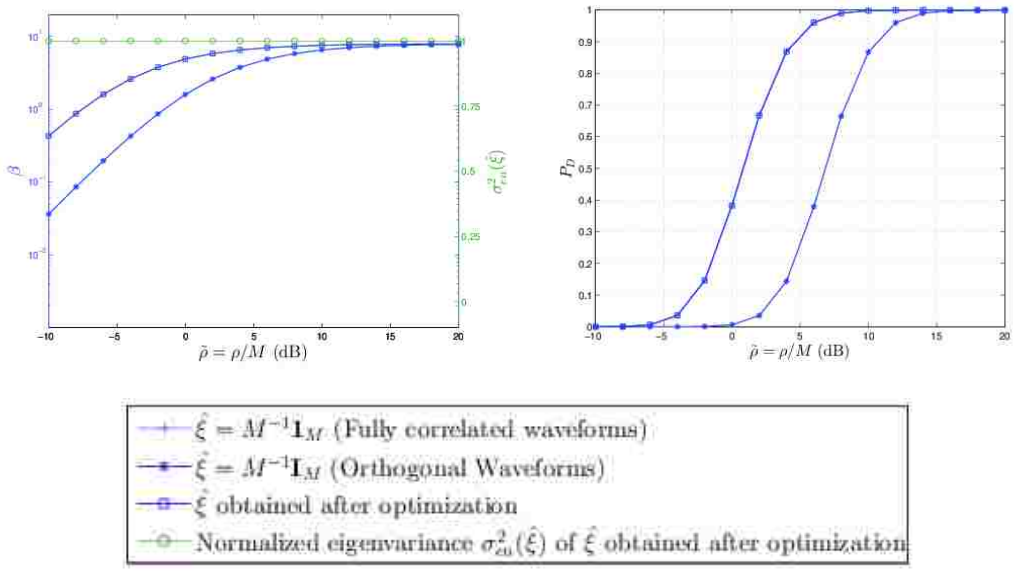


Figure 3.3: Normalized eigenvariance, detector SNR and P_D (at $P_{FA} = 10^{-6}$) using the $\hat{\boldsymbol{\xi}}$ matrix obtained after optimization at different SNRs for $\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_{MN}$

(since if $\mathbf{A} \succeq 0$, then $\mathbf{x}^\dagger \mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$, and $\mathbf{A}\mathbf{x} = \mathbf{0}$ is true if and only if $\mathbf{x}^\dagger \mathbf{A}\mathbf{x} = 0$). Hence we have

$$\begin{aligned}
\ker(\boldsymbol{\Sigma_1}) &= \ker(E_r \boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha \boldsymbol{\xi}) \cap \ker(\mathbf{G}\boldsymbol{\xi}\mathbf{G}) \cap \ker(\mathbf{C}) \\
&= \ker(\boldsymbol{\xi}) \cap \ker(\mathbf{C}) \\
&= \ker(\mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C}) = \ker(\boldsymbol{\Sigma_0})
\end{aligned}$$

and hence by the rank-nullity theorem, we have $\text{rank}(\boldsymbol{\Sigma}_0) = \text{rank}(\boldsymbol{\Sigma}_1)$. Then, since $\ker(\boldsymbol{\Sigma}_0) \subseteq \ker(\boldsymbol{\xi}) \subseteq \ker(\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi})$ and $\boldsymbol{\Sigma}_0 = \mathbf{V}\mathbf{V}^\dagger$, the decomposition $E_r\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi} = \mathbf{V}\mathbf{X}\mathbf{V}^\dagger$, where $\mathbf{X}$ is a

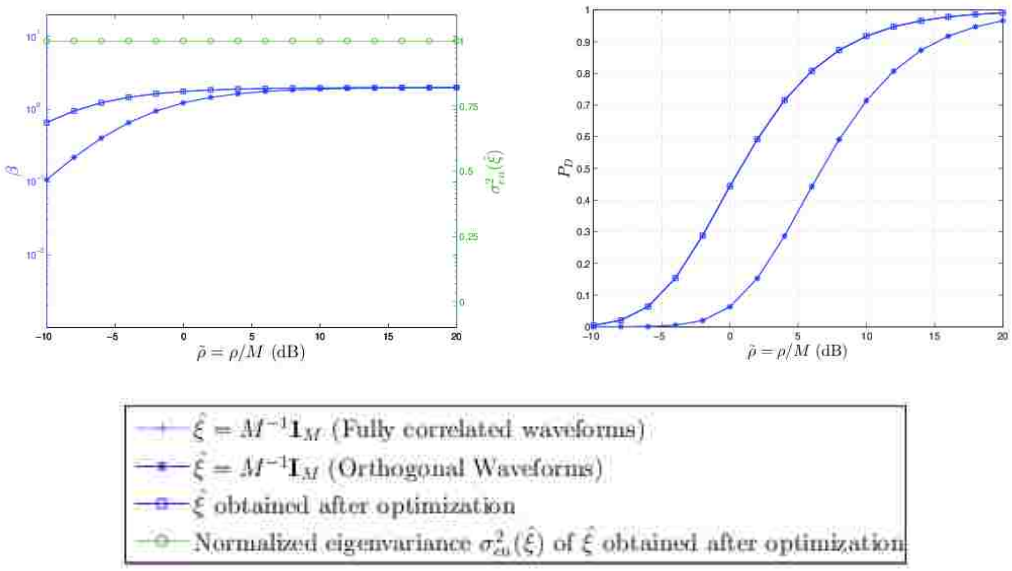


Figure 3.4: Normalized eigenvariance, detector SNR and P_D (at $P_{FA} = 10^{-6}$) using the $\hat{\boldsymbol{\xi}}$ matrix obtained after optimization at different SNRs for $\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_N \otimes \mathbf{I}_M$

positive semidefinite matrix, should exist. Solving for $\mathbf{X}$, we obtain

$$\begin{aligned}\mathbf{X} &= E_r(\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi}\mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-1} \\ &= (\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)\mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-1} = \mathbf{Z} - \mathbf{I}_P\end{aligned}$$

where $\mathbf{Z} = (\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-1}$. This in turn allows us to re-express the observation vector as $\mathbf{r} = \mathbf{V}\tilde{\mathbf{r}}$, where $\tilde{\mathbf{r}}$ is a $P \times 1$ zero-mean Gaussian random vector, with

$$\mathrm{E}\left[\tilde{\mathbf{r}}\tilde{\mathbf{r}}^\dagger\right] = \begin{cases} \mathbf{I}_P & \mathcal{H}_1 \\ \mathbf{I}_P + \mathbf{X} & \mathcal{H}_0 \end{cases}$$

Note that the value of $\tilde{\mathbf{r}}$ corresponding to any value of $\mathbf{r}$ can be obtained as $\tilde{\mathbf{r}} = (\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger\mathbf{r}$. Let $f(\tilde{\mathbf{r}}|\mathcal{H}_1)$ and $f(\tilde{\mathbf{r}}|\mathcal{H}_0)$ represent the probability density functions of $\tilde{\mathbf{r}}$ under the two hypotheses. Since $\mathbf{r}$ is a deterministic function of $\tilde{\mathbf{r}}$, the Neyman-Pearson optimum test

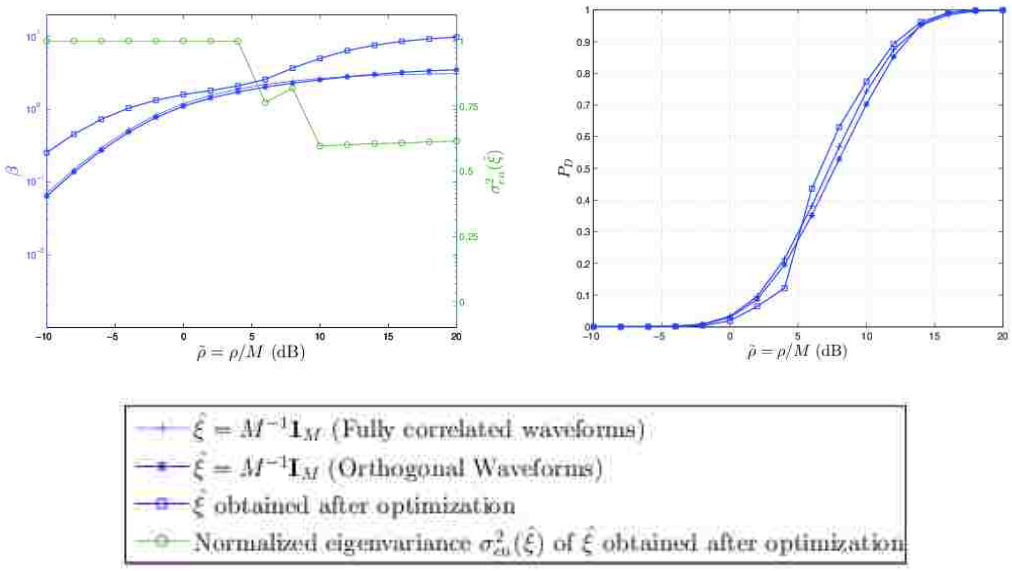


Figure 3.5: Normalized eigenvariance, detector SNR and P_D (at $P_{FA} = 10^{-6}$) using the $\hat{\boldsymbol{\xi}}$ matrix obtained after optimization at different SNRs for a randomly chosen channel covariance matrix with $\sigma^2_{\text{en}}(\boldsymbol{\Sigma}_\alpha) = 0.5$.

statistic is given as

$$\begin{aligned}
T &= \log \frac{f(\tilde{\mathbf{r}}|\mathcal{H}_0)}{f(\tilde{\mathbf{r}}|\mathcal{H}_1)} \\
&= \log \frac{\exp\left\{-\tilde{\mathbf{r}}^\dagger[\mathbf{I}_P + \mathbf{X}]^{-1}\tilde{\mathbf{r}}\right\}}{\exp\left\{-\tilde{\mathbf{r}}^\dagger \mathbf{I}_P^{-1}\tilde{\mathbf{r}}\right\}} + \log \frac{|\pi \mathbf{I}_P|}{|\pi(\mathbf{I}_P + \mathbf{X})|} \\
&= \tilde{\mathbf{r}}^\dagger[\mathbf{I}_P - (\mathbf{I}_P + \mathbf{X})^{-1}]\tilde{\mathbf{r}} \quad \text{(ignoring constant term)} \\
&= \tilde{\mathbf{r}}^\dagger[\mathbf{I}_P - \mathbf{Z}^{-1}]\tilde{\mathbf{r}} \\
&= \mathbf{r}^\dagger \mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-2}\mathbf{V}^\dagger\mathbf{r} - \mathbf{r}^\dagger\mathbf{V}[\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{V}]^{-1}\mathbf{V}^\dagger\mathbf{r} = \mathbf{r}^\dagger\mathbf{D}\mathbf{r}
\end{aligned}$$

where $\mathbf{D} = \mathbf{V}[(\mathbf{V}^\dagger\mathbf{V})^{-2} - (\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{V})^{-1}]\mathbf{V}^\dagger$. □

Note that the decomposition $\boldsymbol{\Sigma}_0 = \mathbf{V}\mathbf{V}^\dagger$ is unique upto a $P \times P$ unitary transformation matrix $\mathbf{W}$ (satisfying $\mathbf{W}^\dagger = \mathbf{W}^{-1}$), hence if $\tilde{\mathbf{V}} = \mathbf{V}\mathbf{W}$ then $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^\dagger = \mathbf{V}\mathbf{W}\mathbf{W}^\dagger\mathbf{V}^\dagger = \boldsymbol{\Sigma}_0$.

However, the final test statistic does not change under this transformation, since

$$\begin{aligned}
&\tilde{\mathbf{V}}[(\tilde{\mathbf{V}}^\dagger\tilde{\mathbf{V}})^{-2} - (\tilde{\mathbf{V}}^\dagger\boldsymbol{\Sigma}_1\tilde{\mathbf{V}})^{-1}]\tilde{\mathbf{V}}^\dagger \\
&= \mathbf{VW}[(\mathbf{W}^{-1}\mathbf{V}^\dagger\mathbf{VWW}^{-1}\mathbf{V}^\dagger\mathbf{VW})^{-2} \\
&\qquad - (\mathbf{W}^{-1}\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{VW})^{-1}]\mathbf{W}^{-1}\mathbf{V}^\dagger \\
&= \mathbf{VW}[\mathbf{W}^{-1}(\mathbf{V}^\dagger\mathbf{V})^{-2}\mathbf{W} - \mathbf{W}^{-1}(\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{V})^{-1}\mathbf{W}]\mathbf{W}^{-1}\mathbf{V}^\dagger \\
&= \mathbf{V}[(\mathbf{V}^\dagger\mathbf{V})^{-2} - (\mathbf{V}^\dagger\boldsymbol{\Sigma}_1\mathbf{V})^{-1}]\mathbf{V}^\dagger = \mathbf{D}
\end{aligned}$$

3.1.2 Proof of Theorem 2

Proof. Define the vector $\tilde{\mathbf{r}}$ and the matrices $\mathbf{X}$ and $\mathbf{V}$ as in Theorem 1. Since $\boldsymbol{\xi}$ is a block diagonal matrix, and $\mathbf{G}$ is a diagonal matrix whose entries are constant within each block of $\boldsymbol{\xi}$, we can write $\mathbf{G}\boldsymbol{\xi}\mathbf{G} = \boldsymbol{\xi}\mathbf{G}^2 = \mathbf{G}^2\boldsymbol{\xi}$. Therefore, when $\mathbf{C} = 0$, we have

$$\boldsymbol{\Sigma}_0 = \mathbf{G}\boldsymbol{\xi}\mathbf{G} = \mathbf{VV}^\dagger \quad \Rightarrow \quad \boldsymbol{\xi} = \mathbf{VV}^\dagger\mathbf{G}^{-2} = \mathbf{G}^{-2}\mathbf{VV}^\dagger$$

and hence

$$\begin{aligned}
\mathbf{X} &= E_r(\mathbf{V}^\dagger\mathbf{V})^{-1}\mathbf{V}^\dagger\mathbf{VV}^\dagger\mathbf{G}^{-2}\boldsymbol{\Sigma}_\alpha\mathbf{G}^{-2}\mathbf{VV}^\dagger\mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-1} \\
&= E_r\mathbf{V}^\dagger\mathbf{G}^{-2}\boldsymbol{\Sigma}_\alpha\mathbf{G}^{-2}\mathbf{V}
\end{aligned}$$

Further, let $\operatorname{rank}\{\boldsymbol{\Sigma}_\alpha\} = Q$, and let $E_r\mathbf{G}^{-2}\boldsymbol{\Sigma}_\alpha\mathbf{G}^{-2} = \mathbf{UU}^\dagger$, where $\mathbf{U}$ is a full rank matrix of size $MN \times Q$. Then, using the identity $\mathbf{I} - (\mathbf{I} + \mathbf{AB})^{-1}) = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}$, we can write

$$\begin{aligned}
\mathbf{I}_P - (\mathbf{I}_P + \mathbf{X})^{-1} &= \mathbf{I}_P - (\mathbf{I}_P + \mathbf{V}^\dagger\mathbf{UU}^\dagger\mathbf{V})^{-1} \\
&= \mathbf{V}^\dagger\mathbf{U}(\mathbf{I}_Q + \mathbf{U}^\dagger\mathbf{VV}^\dagger\mathbf{U})^{-1}\mathbf{U}^\dagger\mathbf{V}
\end{aligned}$$

Finally, applying the identity $\mathbf{B}(\mathbf{I}+\mathbf{AB})^{-1} = (\mathbf{I}+\mathbf{BA})^{-1}\mathbf{B}$, we obtain

$$\begin{aligned} T &= \tilde{\mathbf{r}}^\dagger \mathbf{V}^\dagger \mathbf{U}(\mathbf{I}_Q + \mathbf{U}^\dagger \mathbf{V}\mathbf{V}^\dagger \mathbf{U})^{-1}\mathbf{U}^\dagger \mathbf{V}\tilde{\mathbf{r}} \\ &= \mathbf{r}^\dagger \mathbf{U}(\mathbf{I}_Q + \mathbf{U}^\dagger \mathbf{V}\mathbf{V}^\dagger \mathbf{U})^{-1}\mathbf{U}^\dagger \mathbf{r} \\ &= \mathbf{r}^\dagger (\mathbf{I}_{MN} + \mathbf{U}\mathbf{U}^\dagger \mathbf{V}\mathbf{V}^\dagger)^{-1}\mathbf{U}\mathbf{U}^\dagger \mathbf{r} = \mathbf{r}^\dagger \mathbf{D}\mathbf{r} \end{aligned}$$

where $\mathbf{D} = (\mathbf{I}_{MN} + E_r \mathbf{G}^{-2}\boldsymbol{\Sigma}_\alpha \boldsymbol{\xi})^{-1}\mathbf{G}^{-2}\boldsymbol{\Sigma}_\alpha \mathbf{G}^{-2}$. Substituting $\mathbf{G} = \sigma_w \mathbf{P}$ and $\rho = E_r/\sigma_w^2$, and ignoring multiplicative constants, we obtain the result of the theorem. □

3.1.3 Proof of Lemma 1

Proof. Under the given assumptions, we have

$$\mathbf{G} = \tilde{\sigma}\mathbf{I}_{MN}, \qquad \boldsymbol{\Sigma}_0 = \mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C} = \tilde{\sigma}^2 \boldsymbol{\xi}\ ,$$

$$\boldsymbol{\Sigma}_1 = E_r \boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{G}\boldsymbol{\xi}\mathbf{G} + \mathbf{C} = E_r \boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha \boldsymbol{\xi} + \tilde{\sigma}^2 \boldsymbol{\xi}$$

and

$$\begin{aligned} \sigma_w^2 &= \mathrm{E}\left[\mathbf{w}^\dagger \mathbf{w}\right] = \mathrm{Tr}\left\{\boldsymbol{\Sigma}_0\right\} \\ &= \tilde{\sigma}^2 \mathrm{Tr}\left\{\boldsymbol{\xi}\right\} = \tilde{\sigma}^2 \sum_{i=1}^{N} \mathrm{Tr}\left\{\boldsymbol{\xi}_i\right\} = N\tilde{\sigma}^2, \end{aligned}$$

resulting in $\mathbf{P} = \sigma_w^{-1}\mathbf{G} = N^{-1/2}\mathbf{I}_{MN}$. Further, using Theorem 2, we obtain

$$\begin{aligned} \mathbf{D} &= (\mathbf{I}_{MN} + \rho \mathbf{P}^{-2}\boldsymbol{\Sigma}_\alpha \boldsymbol{\xi})^{-1}\mathbf{P}^{-2}\boldsymbol{\Sigma}_\alpha \mathbf{P}^{-2} \\ &= (\mathbf{I}_{MN} + \rho N \boldsymbol{\Sigma}_\alpha \boldsymbol{\xi})^{-1}\boldsymbol{\Sigma}_\alpha \end{aligned}$$

(ignoring multiplicative constants in $\mathbf{D}$). The mean of the test statistic is

$$\mathrm{E}\left[T\right] = \mathrm{E}\left[\mathbf{r}^\dagger \mathbf{D}\mathbf{r}\right] = \mathrm{E}\left[\mathrm{Tr}\left\{\mathbf{r}^\dagger \mathbf{D}\mathbf{r}\right\}\right] = \mathrm{Tr}\left\{\mathbf{D}\mathrm{E}\left[\mathbf{r}\mathbf{r}^\dagger\right]\right\} \tag{3.17}$$

and we hence have

$$\mathrm{E}\left[T|\mathcal{H}_0\right] - \mathrm{E}\left[T|\mathcal{H}_1\right] = \mathrm{Tr}\left\{\mathbf{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)\right\} = E_r \mathrm{Tr}\left\{\mathbf{D}\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi}\right\}$$

Towards obtaining the variance of T, we assume the decomposition $\mathbf{D} = \mathbf{A}^\dagger\mathbf{A}$, where $\mathbf{A}$ is a $R \times MN$ matrix ($R = \mathrm{rank}\left\{\mathbf{D}\right\}$). Setting $\mathbf{z} = \mathbf{A}\mathbf{r}$, we can write

$$\begin{aligned}
\mathrm{E}\left[T^2\right] &= \mathrm{E}\left[\left|\mathbf{z}^\dagger\mathbf{z}\right|^2\right] = \sum_{i=1}^{MN}\sum_{k=1}^{MN}\mathrm{E}\left[z_i z_i^* z_k z_k^*\right] \\
&= \sum_{i=1}^{MN}\sum_{k=1}^{MN}\left(\mathrm{E}\left[|z_i|^2\right]\mathrm{E}\left[|z_k|^2\right] + |\mathrm{E}\left[z_i z_k\right]|^2 + |\mathrm{E}\left[z_i z_k^*\right]|^2\right) \qquad (3.18) \\
&= \mathrm{E}^2[T] + \mathrm{Tr}\left\{\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right]\left(\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right]\right)^\dagger\right\}\mathrm{Tr}\left\{\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\left(\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\right)^\dagger\right\}
\end{aligned}$$

where (3.18) results from the fact that if x_1, x_2, x_3 and x_4 are Gaussian distributed complex random variables, then

$$\mathrm{E}\left[x_1x_2x_3x_4\right] = \mathrm{E}\left[x_1x_2\right]\mathrm{E}\left[x_3x_4\right] + \mathrm{E}\left[x_1x_3\right]\mathrm{E}\left[x_2x_4\right] + \mathrm{E}\left[x_1x_4\right]\mathrm{E}\left[x_2x_3\right]$$

It can be shown that $\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right] = 0$ under both $\mathcal{H}_1$ and $\mathcal{H}_0$, hence allowing us to express the variance of T as

$$\begin{aligned}
\mathrm{var}(T) = \mathrm{E}\left[T^2\right] - \mathrm{E}^2\left[T\right] &= \mathrm{Tr}\left\{\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\left(\mathrm{E}\left[\mathbf{z}\mathbf{z}^\dagger\right]\right)^\dagger\right\} \\
&= \mathrm{Tr}\left\{\mathbf{A}\mathrm{E}\left[\mathbf{r}\mathbf{r}^\dagger\right]\mathbf{A}^\dagger\mathbf{A}\mathrm{E}\left[\mathbf{r}\mathbf{r}^\dagger\right]\mathbf{A}^\dagger\right\} \\
&= \mathrm{Tr}\left\{(\mathbf{D}\mathrm{E}\left[\mathbf{r}\mathbf{r}^\dagger\right])^2\right\}
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\beta &= \frac{2\left(\mathrm{Tr}\left\{\mathbf{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)\right\}\right)^2}{\mathrm{Tr}\left\{(\mathbf{D}\boldsymbol{\Sigma}_1)^2\right\} + \mathrm{Tr}\left\{(\mathbf{D}\boldsymbol{\Sigma}_0)^2\right\}} \\
&= \frac{2\left(E_r\mathrm{Tr}\left\{\mathbf{D}\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi}\right\}\right)^2}{\mathrm{Tr}\left\{(E_r\mathbf{D}\boldsymbol{\xi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\xi} + \tilde{\sigma}^2\mathbf{D}\boldsymbol{\xi})^2\right\} + \mathrm{Tr}\left\{(\tilde{\sigma}^2\mathbf{D}\boldsymbol{\xi})^2\right\}}
\end{aligned}$$

Now, let $\mathbf{\Omega} = \mathbf{\Sigma}_\alpha \boldsymbol{\xi}$ and $\mathbf{U} = \rho N \mathbf{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{I}_{MN}$. Then we have $\mathbf{D} = \mathbf{U}^{-1}\mathbf{\Sigma}_\alpha$, and

$$\beta = \frac{2\left(E_r \mathrm{Tr}\left\{\mathbf{U}^{-1}\mathbf{\Omega}^2\right\}\right)^2}{\mathrm{Tr}\left\{(E_r \mathbf{U}^{-1}\mathbf{\Omega}^2 + \tilde{\sigma}^2 \mathbf{U}^{-1}\mathbf{\Omega})^2\right\} + \mathrm{Tr}\left\{(\tilde{\sigma}^2 \mathbf{U}^{-1}\mathbf{\Omega})^2\right\}} \tag{3.19}$$

Also, we can write

$$\begin{aligned} \mathbf{\Omega} &= \frac{1}{N\rho}(\mathbf{U} - \mathbf{I}) \\ \Rightarrow \mathbf{U}^{-1}\mathbf{\Omega} &= \frac{1}{N\rho}(\mathbf{I} - \mathbf{U}^{-1}) \qquad (3.20) \\ \Rightarrow \mathbf{U}^{-1}\mathbf{\Omega}^2 &= \frac{1}{N\rho}(\mathbf{\Omega} - \mathbf{U}^{-1}\mathbf{\Omega}) \\ &= \frac{1}{N^2\rho^2}(\mathbf{U} - 2\mathbf{I} + \mathbf{U}^{-1}) \qquad (3.21) \end{aligned}$$

Substituting (3.20) and (3.21) in (3.19) and refactoring, we obtain (3.15), thus proving Lemma 1. □

3.1.4 Simplification of Detector SNR expression

In this Appendix, we simplify the general expression for the detector SNR in (3.22),

$$\beta = \frac{2\left(\mathrm{Tr}\left\{\mathbf{U} - 2\mathbf{I} + \mathbf{U}^{-1}\right\}\right)^2}{\mathrm{Tr}\left\{(\mathbf{U} - \mathbf{I})^2 + (\mathbf{U}^{-1} - \mathbf{I})^2\right\}} \tag{3.22}$$

where ($\mathbf{U} = \rho N \mathbf{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{I}_{MN}$) under a few special cases commonly studied in MIMO radar literature. We show that the resulting expressions correspond exactly with the expressions derived in [10]. To this end, we first obtain simplified expressions for the channel covariance matrix $\mathbf{\Sigma}_\alpha$ under four types of antenna placements (refer to notation introduced in Section 3.0.3):

Case (i) *Large transmit and receive antenna separations*:

In this case, all the channel gains will be uncorrelated, $\mathbf{\Sigma}_\alpha = \|\boldsymbol{\zeta}\|^{-2}((\boldsymbol{\zeta} \circ \mathbf{b})(\boldsymbol{\zeta} \circ \mathbf{b})^\dagger) \circ \mathbf{I}_{MN}$. Here the rank of $\mathbf{\Sigma}_\alpha$ can be at most MN.

Case (ii) *Small transmit and receive antenna separations*:

Here we will have $\boldsymbol{\psi}_x = \boldsymbol{\psi}_y = \mathbf{0}_{MN}$, and hence $\mathbf{\Sigma}_\alpha = \|\boldsymbol{\zeta}\|^{-2}((\boldsymbol{\zeta} \circ \mathbf{b})(\boldsymbol{\zeta} \circ \mathbf{b})^\dagger)$. Thus, $\mathbf{\Sigma}_\alpha$ will always be unit rank in this case.

Case (iii) *Large transmit antenna separations, small receive antenna separations*:

In this case we have $\mathbf{\Sigma}_\alpha = \|\boldsymbol{\zeta}\|^{-2}((\boldsymbol{\zeta} \circ \mathbf{b})(\boldsymbol{\zeta} \circ \mathbf{b})^\dagger) \circ (\mathbf{1}_N \otimes \mathbf{I}_M)$. where $\mathbf{1}_N$ represents a $N \times N$ matrix all of whose elements are ones. Note that the rank of $\mathbf{\Sigma}_\alpha$ in this case is at most M.

Case (iv) *Small transmit antenna separations, large receive antenna separations*:

In this case we similarly have $\mathbf{\Sigma}_\alpha = \|\boldsymbol{\zeta}\|^{-2}((\boldsymbol{\zeta} \circ \mathbf{b})(\boldsymbol{\zeta} \circ \mathbf{b})^\dagger) \circ (\mathbf{I}_N \otimes \mathbf{1}_M)$, The rank of $\mathbf{\Sigma}_\alpha$ in this case is at most N.

For ease of demonstration, we make the following assumptions

(a) The vectors $\mathbf{b}$ and $\boldsymbol{\zeta}$ (that are of length MN) have all elements equal to 1, resulting in $\|\boldsymbol{\zeta}\|^{-2}((\boldsymbol{\zeta} \circ \mathbf{b})(\boldsymbol{\zeta} \circ \mathbf{b})^\dagger) = \frac{1}{MN}\mathbf{1}_{MN}$. This enables us to write $\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_{MN}$ in Case (i), $\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_N \otimes \mathbf{1}_M$ in Case (ii), $\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_N \otimes \mathbf{I}_M$ in Case (iii) and $\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}$ in Case (iv).

(b) We assume that the target is at a sufficient distance from the transmitters and receivers to allow us to write $\tau_{m,n} \approx \tau_{\tilde{m},n}$ for all $m, \tilde{m}, n$. Under this assumption we can write $\boldsymbol{\xi}_n \approx \hat{\boldsymbol{\xi}}$ for all n (and hence $\boldsymbol{\xi} = \mathbf{I}_N \otimes \hat{\boldsymbol{\xi}}$), where $\hat{\boldsymbol{\xi}}$ represents the zero-lag correlation

matrix of the transmit waveforms, i.e. $\hat{\boldsymbol{\xi}} = [\hat{\xi}_{m,\tilde{m}}]_{M\times M}$ with

$$\hat{\xi}_{m,\tilde{m}} = \int_0^{T_s} s_{\tilde{m}}(t) s_m^*(t) \mathrm{d}t$$

Further, to simplify the notation, we use $\mathbf{A}_{MN}$ and $\mathbf{A}_N$ for the matrix $\mathbf{A}$ of size $MN \times MN$ and $N \times N$, respectively, and not use any subscript when the size can be inferred from the context.

We now obtain simplifications for the detector SNR β under three common scenarios. Note that in [10], the received SNR is defined as $\tilde{\rho} = \rho/M$. In order to enable comparisons, we shall use definition of received SNR in the derivations below.

Large transmit and receive antenna separations, and orthogonal waveforms

In this case,

$$\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}, \quad \hat{\boldsymbol{\xi}} = \frac{1}{M}\mathbf{I}_M \quad \boldsymbol{\xi} = \mathbf{I}_N \otimes \hat{\boldsymbol{\xi}} = \frac{1}{M}\mathbf{I}_{MN} \tag{3.23}$$

In view of (3.23), we have

$$\begin{aligned} \mathbf{U} = \rho N \boldsymbol{\Sigma}_\alpha \boldsymbol{\xi} + \mathbf{I}_{MN} &= \left(\frac{\tilde{\rho}}{M} + 1\right)\mathbf{I}_{MN} \ , \\ \mathbf{U} - \mathbf{I}_{MN} = \frac{\tilde{\rho}}{M}\mathbf{I}_{MN}, \quad \mathbf{U}^{-1} - \mathbf{I}_{MN} &= -\frac{\tilde{\rho}}{M}\frac{\mathbf{I}_{MN}}{(1+\tilde{\rho}/M)} \end{aligned} \tag{3.24}$$

Substituting (3.24) in (3.22) and simplifying, we obtain for the numerator and denominator of (3.22), denoted by NUM and DEN, respectively, as follows.

$$NUM = \frac{2\tilde{\rho}^4 N^2}{M^2(1+\tilde{\rho}/M)^2} \tag{3.25}$$

$$DEN = \frac{2\tilde{\rho}^2 N(1+\tilde{\rho}^2/2M^2 + \tilde{\rho}/M)}{M(1+\tilde{\rho}/M)^2} \tag{3.26}$$

Substituting (3.25) and (3.26) in (3.22), we obtain

$$\beta = \frac{N\tilde{\rho}^2}{M(1+\tilde{\rho}^2/2M^2+\tilde{\rho}/M)} \tag{3.27}$$

which is same as the result given in [10].

Small transmit and receive antenna separations, and fully correlated waveforms

In this case

$$\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_{MN} \qquad \hat{\boldsymbol{\xi}} = \frac{1}{M}\mathbf{1}_M \tag{3.28}$$

Therefore,

$$\mathbf{\Sigma}_\alpha \boldsymbol{\xi} = \frac{1}{MN}\mathbf{1}_{MN}(\mathbf{I}_N \otimes \frac{1}{M}\mathbf{1}_M) = \frac{1}{MN}\mathbf{1}_{MN}$$

and $\mathbf{U} = \tilde{\rho}\mathbf{1}_{MN} + \mathbf{I}_{MN}$. Applying the matrix inversion lemma to $\mathbf{U}$, we can show that

$$\mathbf{U}^{-1} = \mathbf{I}_{MN} - \frac{1}{(MN+1/\tilde{\rho})}\mathbf{1}_{MN} \tag{3.29}$$

Substituting (3.29) in (3.22) and simplifying, we obtain

$$NUM = \frac{2\tilde{\rho}^2 M^4 N^4}{(MN+1/\tilde{\rho})^2} \tag{3.30}$$

$$DEN = \frac{(2M^2N^2 + \tilde{\rho}^2 M^4 N^4 + 2\tilde{\rho}M^3N^3)}{(MN+1/\tilde{\rho})^2} \tag{3.31}$$

and hence

$$\beta = \frac{NUM}{DEN} = \frac{\tilde{\rho}^2 M^2 N^2}{1+\tilde{\rho}^2 M^2 N^2/2 + \tilde{\rho}MN} \tag{3.32}$$

which is same as the result given in [10].

Large transmit antenna separations, small receive antenna separations, and orthogonal waveforms

In this case

$$\mathbf{\Sigma}_\alpha = \frac{1}{MN}\mathbf{1}_N \otimes \mathbf{I}_M, \quad \hat{\boldsymbol{\xi}} = \frac{1}{M}\mathbf{I}_M, \quad \boldsymbol{\xi} = \frac{1}{M}\mathbf{I}_{MN} \tag{3.33}$$

We then have

$$\mathbf{U} = \tilde{\rho}(\mathbf{1}_N \otimes \mathbf{I}_M)\frac{1}{M}\mathbf{I}_{MN} + \mathbf{I}_{MN} = \frac{\tilde{\rho}}{M}\mathbf{A} + \mathbf{I}_{MN} \tag{3.34}$$

where $\mathbf{A} = \mathbf{1}_N \otimes \mathbf{I}_M$. To find $\mathbf{U}^{-1}$, we use the following approach[3]. We let

$$\left(\frac{\tilde{\rho}}{M}\mathbf{A} + \mathbf{I}_{MN}\right)(\mathbf{I}_{MN} - \gamma\mathbf{A}) = \mathbf{I}_{MN} \tag{3.35}$$

and find γ which satisfies (3.35). Simplifying (3.35), we obtain

$$\frac{\tilde{\rho}}{M}\mathbf{A} - \frac{\tilde{\rho}}{M}N\gamma\mathbf{A} - \gamma\mathbf{A} = 0 \tag{3.36}$$

where we have used $\mathbf{A}^2 = N(\mathbf{1}_N \otimes \mathbf{I}_M) = N\mathbf{A}$. From (3.36), we get

$$\gamma = \frac{\tilde{\rho}}{(M + \tilde{\rho}N)} \tag{3.37}$$

We thus have

$$\mathbf{U} - \mathbf{I}_{MN} = \frac{\tilde{\rho}}{M}\mathbf{A}, \quad \mathbf{U}^{-1} - \mathbf{I}_{MN} = -\gamma\mathbf{A} \tag{3.38}$$

Combining (3.38) with (3.22) and using $\mathbf{A}^2 = N\mathbf{A}$, we obtain the following (after some manipulations) for the numerator and denominator of (3.22)

$$NUM = 2\frac{\tilde{\rho}^4 N^4}{(M + \tilde{\rho}N)^2} \tag{3.39}$$

$$DEN = \tilde{\rho}^2 N^2 \frac{(2M^2 + \tilde{\rho}^2 N^2 + 2\tilde{\rho}MN)}{M(M + \tilde{\rho}N)^2} \tag{3.40}$$

[3]This method has been suggested by KVS Hari of Indian Institute of Science, Bangalore.

From (3.39) and (3.40), we get

$$\beta = \frac{\tilde{\rho}^2 N^2}{M(1 + \tilde{\rho}^2 N^2/2M^2 + \tilde{\rho} N/M)} \tag{3.41}$$

which is same as the result given in [10].

3.1.5 Proof of Lemma 2

Proof. Let $\{\lambda_i\}_{i=1}^{MN}$ denote the eigenvalues of $\boldsymbol{\xi}$. Given $\boldsymbol{\Sigma}_\alpha = \frac{1}{MN}\mathbf{I}_{MN}$, we have $\mathbf{U} = \rho M^{-1}\boldsymbol{\xi} + \mathbf{I}_{MN}$, therefore the eigenvalues of $\mathbf{U}$ will be $\{\tilde{\rho}\lambda_i + 1)\}_{i=1}^{MN}$, where $\tilde{\rho} = \rho M^{-1}$. Since $\boldsymbol{\Sigma}_\alpha$ and $\boldsymbol{\xi}$ are Hermitian symmetric, $\mathbf{U}$ is also Hermitian symmetric, and we can hence write

$$\mathrm{Tr}\left\{\mathbf{U}^k\right\} = \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^k \quad \forall\ k \in \mathcal{Z}$$

Further, $\boldsymbol{\xi} \succeq 0$ implies that $\lambda_i \geq 0$ for all i. Hence, for any $k > 0$, we have

$$(\tilde{\rho}\lambda_i + 1)^{-k} < 1 \quad \Rightarrow \quad \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-k} < MN \tag{3.42}$$

Further,

$$\sum_{i=1}^{MN} \lambda_i = \mathrm{Tr}\{\boldsymbol{\xi}\} = \sum_{n=1}^{N} \mathrm{Tr}\{\boldsymbol{\xi}_n\} = N \tag{3.43}$$

Let $\beta = \mathrm{NUM/DEN}$, where

$$\begin{aligned}
\mathrm{NUM} &= 2\left[\sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1) - 2MN + \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-1}\right]^2 \\
&= 2\left[\tilde{\rho}N - MN + \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-1}\right]^2 \\
&= 2N^2\left[\tilde{\rho} - c_1\right]^2
\end{aligned}$$

and

$$\begin{aligned}
\text{DEN} &= \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-2} - 2\sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-1} + 2MN \\
&\quad - 2\sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1) + \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^2 \\
&= \sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-2} - 2\sum_{i=1}^{MN} (\tilde{\rho}\lambda_i + 1)^{-1} + MN \\
&\quad + \tilde{\rho}^2 \sum_{i=1}^{MN} \lambda_i^2 \\
&= c_2 + \tilde{\rho}^2 \sum_{i=1}^{MN} \lambda_i^2
\end{aligned}$$

where c_1 and c_2 are variables that are dependent on $\{\lambda_i\}_{i=1}^{MN}$, but are limited to the intervals $c_1 \in (0, M)$ and $c_2 \in (-MN, 2MN)$. Thus, when $\tilde{\rho} \gg \max(\sqrt{MN}, M)$ (which implies $\rho = M\tilde{\rho} \gg \max(\sqrt{M^3N}, M^2)$), then $\tilde{\rho} \gg c_1$ and $\tilde{\rho}^2 \gg c_2$, and we can write

$$\beta \approx \frac{2\tilde{\rho}^2 N^2}{\tilde{\rho}^2 \sum_{i=1}^{MN} \lambda_i^2}$$

Hence, maximizing β in this scenario is equivalent to minimizing $\sum_{i=1}^{MN} \lambda_i^2$ under the constraint of normalized power, i.e., $\sum_{i=1}^{MN} \lambda_i = N$. It can be shown using the method of Lagrangian multipliers that the minimum will occur only when $\lambda_i = N/MN = 1/M \; \forall i$. This implies $\boldsymbol{\xi} = M^{-1}\mathbf{I}_{MN}$, hence proving Lemma 2. □

Chapter 4

Estimation performance lower bounds for phase synchronization in IEEE 1588

4.1 Introduction

Packet-based time synchronization techniques based on the IEEE 1588 Precision Time Protocol (PTP) are being increasingly considered as a means of providing microsecond-level synchronization between cell towers in 4G LTE (Long Term Evolution) mobile networks [17, 18, 19, 20, 21]. Such a high degree of synchronization accuracy is a necessity in 4G networks since it helps ensure seamless handovers between cell towers, helps reduce inter-cell interference, and also enables the use of MIMO techniques to improve capacity [22]. As compared to GPS (global positioning system) based synchronization, packet-based synchronization is often more cost-effective since it utilizes the existing mobile backhaul network

infrastructure that is used to interconnect cell towers. However, since backhaul networks are typically leased from commercial internet service providers (ISPs), mobile network operators must share its use with other commercial and residential users. Background traffic generated by these users often results in random network delays that can hinder packet-based synchronization. Overcoming this problem is key to the adoption of packet-based synchronization schemes in mobile backhaul networks, especially given that the microsecond-level synchronization requirements are only expected to grow more stringent in the future.

The output of a typical computer clock can be modeled mathematically using a function $c(t)$ of the true current time t. A perfect clock will output $c(t) = t$, while in practice $c(t)$ is a random process with error $e(t) = |c(t) - t|$ that tends to grow over large time scales. Over short time scales, it is possible to model clock behavior using the linear approximation $c(t) = \phi t + \delta$, where $\phi - 1$ is the frequency offset, and δ is the phase offset. Typically, network time synchronization algorithms treat frequency and phase synchronization as two independent problems. In Chapters 4–6, we study to problem of improving the accuracy of phase synchronization schemes, while assuming that near-perfect frequency synchronization is already available. An obvious practical scenario where such an assumption can be made occurs when synchronous ethernet is used in conjunction with the precision time protocol (PTP). Here the PLL obtains frequency information from the physical layer signals of synchronous ethernet, while PTP messages are used for phase synchronization.

In PTP synchronization, messages traveling between the master and the slave encounter several intermediate switches and routers, accumulating random queuing delays at each such node. The resulting randomness in the overall network traversal times is referred to as packet delay variation (PDV). Further, the problem of estimating the phase offset of the slave clock while combating the randomness in the observations that occurs due to PDV is known as

phase offset estimation (POE). In this regard, the requirement arising from 4G backhaul networks is that neighboring cell towers need to be synchronized with absolute phase offset under 1.25 μs, in order to ensure efficient operation in the time division duplexing (TDD) mode.

In Chapters 4–6, we study the performance of POE schemes from a non-Bayesian perspective. Specifically, we treat the slave clock's phase offset as a unknown deterministic parameter that has to be estimated from the arrival/departure timestamps of synchronization packets. These timestamps are modeled as random variables whose probability distributions are influenced by the unknown phase offset as well as the minimum fixed delays in the network. Given the nature of the observations, POE falls under a class of estimation problems known as *location parameter problems*, wherein unknown parameters influence the observations by translating the probability density function (p.d.f.) of the observations, without affecting its shape. Location parameter problems occur in a wide range of practical applications, some examples include regression analysis [36] and the estimation of user position from pseudoranges in global positioning system (GPS) receivers [37].

The IEEE 1588 PTP standard [12] and related literature prescribe the use of simple POE schemes such as the sample mean, minimum and maximum filtering schemes. Several recent papers [17, 21, 22] have studied methods to improve the performance of these schemes. However, it is not well understood as to how close these POE schemes come to achieving the best possible performance. To address this issue, new lower bounds on the error variance of estimators for non-Bayesian location parameter problems are presented in this chapter. The bounds are obtained from two existing Bayesian performance bounds, namely the Weiss-Weinstien [38, 39] and the Ziv-Zakai [40] bounds. We also demonstrate how these lower bounds can be further simplified by exploiting the structure of the POE problem. Numerical

results are also presented that evaluate these lower bounds under a few representative network scenarios.

4.2 System Model

Consider a slave clock whose phase offset relative to its master clock is represented by δ. During *two-way message exchange* in PTP, the following series of packet exchanges are performed between the master and slave in order to determine δ:

1. The master initiates the message exchange by sending a *SYNC* packet to the slave at time t_1. The value of t_1 is later communicated to the slave via a *FOLLOW_UP* message.

2. The slave records the time of reception of the *SYNC* message as $t_2 = t + d_1 + \delta$, where d_1 is the end-to-end (ETE) network delay between the master and the slave.

3. The slave sends a *DELAY_REQ* message to the master, recording the time of transmission as t_3.

4. The master records the time of arrival of the *DELAY_REQ* packet as $t_4 = t_3 - \delta + d_2$, where d_2 is the ETE delay between the slave and the master. The value of t_4 is sent to the slave using a *DELAY_RESP* packet.

Thus, four timestamps (t_1, t_2, t_3 and t_4) are available to the slave at the end of each two-way packet exchange. In order to estimate δ, it is clearly sufficient to only retain the pair of timestamp differences

$$y_1 = t_2 - t_1 = d_1 + \delta \ , \quad y_2 = t_4 - t_3 = d_2 - \delta \tag{4.1}$$

In order to model the ETE delays d_1 and d_2, we note that packets traveling between the master and the slave hop across several intermediate nodes (switches or routers), that are typically part of a larger network that is shared among multiple users. Hence each intermediate node concurrently services background traffic generated by other network users, in addition to synchronization traffic. Assume for simplicity that a single common network path is taken by all packets traveling between the master and the slave. Then each ETE delay will be the sum of a fixed minimum delay component and a variable non-negative component. Here the minimum fixed delay component corresponds to constant propagation and processing delays, while the variable non-negative component corresponds to random queuing delays that occur due to contention for service with background traffic. Hence, we model the ETE delays as

$$d_1 = d_1^{\min} + w_1, \qquad d_2 = d_2^{\min} + w_2$$

where $d_1^{\min}$ and $d_2^{\min}$ represent the fixed minimum delay component, while w_1 and w_2 represent the variable part. Some key assumptions we make are as follows:

(a) The forward and reverse queuing delays w_1 and w_2 are assumed to be non-negative random variables with a finite maximum value. The finite maximum value assumption is reasonable since synchronization packet are typically assigned higher priority than packets of background traffic, hence the worst case queuing delay is bounded for a finite number of switches between the master and the slave (provided that packets of synchronization traffic are spaced sufficiently apart).

(b) The forward and reverse fixed delays are assumed to be equal, i.e. $d_1^{\min} = d_2^{\min} = d$. This is necessary since in general δ, $d_1^{\min}$ and $d_2^{\min}$ cannot be unambiguously (i.e. uniquely) determined from y_1 and y_2, as illustrated by the following example.

Example Consider two cases : $(\delta, d_1^{\min}, d_2^{\min}) = (1, 10, 20)$ and $(\delta, d_1^{\min}, d_2^{\min}) = (2, 9, 21)$. In both cases, we have $(d_1^{\min} + \delta, d_2^{\min} - \delta) = (11, 19)$. Thus, for a given distribution of variable delays w_1 and w_2, the observations $y_1 = d_1^{\min} + \delta + w_1$ and $y_2 = d_1^{\min} - \delta + w_2$ will be identically distributed in both cases!

To avoid this situation, it is necessary to assume that the relationship between $d_1^{\min}$ and $d_2^{\min}$ is known. For simplicity, we only consider the case of equal fixed delays in our system model, while noting that our results can be easily extended to also address the cases where the ratio or difference between $d_1^{\min}$ and $d_2^{\min}$ is known.

(c) Both d and δ are treated as deterministic unknown parameters that can assume any value on the real line $\mathbb{R}$. No prior distributions over either d or δ are assumed.

If δ and d remain constant over a sufficiently long duration of time, multiple two-way exchanges can be performed to obtain more data for POE. If P two-way exchanges are performed, then the slave obtains the $2P$ timestamp differences

$$y_1^{(i)} = d + \delta + w_1^{(i)}, \quad y_2^{(i)} = d - \delta + w_2^{(i)} \tag{4.2}$$

for $i = 1, \cdots, P$. For convenience, we now rewrite the observation model using vector notation. Define $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{T}} & \mathbf{y}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ and $\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} & \mathbf{w}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, where

$$\mathbf{y}_i = \begin{bmatrix} y_i^{(1)} & \cdots & y_i^{(P)} \end{bmatrix}^{\mathrm{T}}, \tag{4.3}$$

$$\mathbf{w}_i = \begin{bmatrix} w_i^{(1)} & \cdots & w_i^{(P)} \end{bmatrix}^{\mathrm{T}} \tag{4.4}$$

Further, let $\boldsymbol{\theta} = [\ \delta\ d\ ]^{\mathrm{T}}$. Then our observation model can be compactly written as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w} \tag{4.5}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{P\times 1} & \mathbf{1}_{P\times 1} \\ -\mathbf{1}_{P\times 1} & \mathbf{1}_{P\times 1} \end{bmatrix} \tag{4.6}$$

and $\mathbf{1}_{P\times 1}$ is a $P \times 1$ vector of ones. The conditional p.d.f. of the observation vector $\mathbf{y}$ is given as

$$f(\mathbf{y}|\boldsymbol{\theta}) = f_{\mathbf{W}}(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \tag{4.7}$$

where $f_{\mathbf{W}}(\mathbf{w})$ is the joint pdf of all the forward and reverse queuing delays. Given the above model, the problem of POE is to estimate δ from the observation vector $\mathbf{y}$.

In order to fully characterize the statistical nature of the queuing delays, in general it is necessary to specify the entire joint pdf $f_{\mathbf{W}}(\mathbf{w})$. An important special case occurs when all the queuing delays $w_1^{(1)}, \cdots, w_1^{(P)}, w_2^{(1)}, \cdots, w_2^{(P)}$ are mutually independent, and

$$w_1^{(i)} \overset{iid}{\sim} f_1\left(w_1^{(i)}\right), \quad w_2^{(i)} \overset{iid}{\sim} f_2\left(w_2^{(i)}\right) \tag{4.8}$$

for all $i = 1, \cdots, P$, allowing us to write

$$f_{\mathbf{W}}(\mathbf{w}) = \prod_{i=1}^{P} f_1\left(w_1^{(i)}\right) \prod_{i=1}^{P} f_2\left(w_2^{(i)}\right) \tag{4.9}$$

Such independent, identically distributed forward and reverse queuing delays occur when the stochastic behavior of the network is the same for every *SYNC* and *DELAY_REQ* packet within any observation window of P two-way exchanges. Let the background traffic characteristics include packet size distributions, arrival time distributions, load factors and flow patterns. Then (4.8) requires that the queue occupancy distributions of intermediate switches and the background traffic characteristics remain static over the observation window. It also requires that consecutive *SYNC* and *DELAY_REQ* packets be sufficiently separated in time

to ensure that dependence is not introduced between neighboring queuing delays.

4.3 Estimator Performance Bounds for Location Parameter Problems

In statistical literature, an estimation problem is said to be a *location parameter problem* if the value of the parameter of interest determines the *location* or *shift* of the distribution of the observations. In this section, we consider a generalization of location parameter problems to vector parameter scenarios, and derive lower bounds on the performance of estimators for such problems. The application of these bounds to the POE problem shall be described in the next section.

Definition 1. *Consider an estimation problem where an observation vector* $\mathbf{x} = [x_1\ x_2\ \cdots\ x_N]^{\mathrm{T}} \in \mathbb{R}^N$ *is influenced by a parameter vector* $\boldsymbol{\theta} \in \mathbb{R}^M$*, via the conditional p.d.f.* $f(\mathbf{x}|\boldsymbol{\theta})$*. If there exists an* $N \times M$ *matrix* $\mathbf{G}$ *and a function* $f_0(\cdot)$ *such that*

$$f(\mathbf{x}|\boldsymbol{\theta}) = f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\theta}) \tag{4.10}$$

then we shall refer to such an estimation problem as a location parameter problem.

Given a location parameter problem, assume we are interested in estimating a scalar of the form $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$, and let $g(\mathbf{x})$ denote any estimator for this problem. Assuming a quadratic loss function, the performance of $g(\mathbf{x})$ can be characterized using either the *conditional risk*

$$R(\boldsymbol{\theta}, g) = \int_{\mathbf{x}} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}]^2 f(\mathbf{x}|\boldsymbol{\theta}) \mathrm{d}\mathbf{x} \tag{4.11}$$

or the *maximum risk*

$$\mathcal{M}(g) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} R(\boldsymbol{\theta}, g) \tag{4.12}$$

or the *Bayes risk*

$$B(g, p) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} R(\boldsymbol{\theta}, g)p(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{4.13}$$

where $p(\boldsymbol{\theta})$ represents a prior distribution defined over the parameter space $\boldsymbol{\Theta}$.

In estimation problems where no prior distribution is known for $\boldsymbol{\theta}$, as is the case in our POE problem model, the conditional risk $R(\boldsymbol{\theta}, g)$ is the preferred measure used to characterize performance. Classical non-Bayesian estimation theory provides several techniques to lower bound $R(\boldsymbol{\theta}, g)$, such as the Cramer-Rao bound [41] and other related bounds [42][43]. Unfortunately, these bounds require the region of support of the conditional p.d.f. in (4.10) to be constant with respect to $\boldsymbol{\theta}$. This condition is clearly violated in our problem of interest.

Since typical non-Bayesian estimation bounds are inadmissible in our problem, we considered Bayesian bounds such as the Weiss-Weinstien bound (WWB) [38][39] and the Ziv-Zakai bound (ZZB) [40][44]. These bounds do not impose any regularity conditions on the observations, however they only provide lower bounds on the Bayes risk $B(g, p)$, which requires a prior distribution $p(\boldsymbol{\theta})$ to be defined. In this section, we describe novel techniques that repurpose these Bayesian estimation bounds to obtain lower bounds on $\mathcal{M}(g)$ for location parameter problems. Our techniques are encapsulated in the following two theorems whose proofs are provided in appendices 4.8.1 and 4.8.2.

Theorem 3. (Obtained from the Weiss-Weinstien Bound) *Given the location parameter problem of Definition 1, let* $\{s_i\}_{i=1}^K$ *and* $\{\mathbf{h}_i\}_{i=1}^K$ *be arbitrarily chosen scalars and* $M \times 1$

vectors respectively. Define

$$\begin{aligned}&\xi(\mathbf{h}_1, \mathbf{h}_2, s_1, s_2)\\&= \int_{\mathbf{x}} \left[\frac{f_0(\mathbf{x}+\mathbf{G}\mathbf{h}_1)}{f_0(\mathbf{x})}\right]^{s_1} \left[\frac{f_0(\mathbf{x}+\mathbf{G}\mathbf{h}_2)}{f_0(\mathbf{x})}\right]^{s_2} f_0(\mathbf{x})d\mathbf{x}\end{aligned} \tag{4.14}$$

Further, let $\mathbf{u}$ *be a* $K \times 1$ *vector whose* i^{th} *element is given as*

$$u_i = (\mathbf{c}^{\mathrm{T}}\mathbf{h}_i) \cdot \xi(\mathbf{h}_i, \mathbf{0}_{M\times 1}, 1-s_i, 0) \tag{4.15}$$

where $\mathbf{0}_{M\times 1}$ *is a* $M \times 1$ *vector of zeros, and let* $\mathbf{V}$ *be a* $K \times K$ *matrix whose* (i, j) *element is*

$$\begin{aligned}V_{ij} &= \xi(-\mathbf{h}_i, -\mathbf{h}_j, s_i, s_j) - \xi(-\mathbf{h}_i, \mathbf{h}_j, s_i, 1-s_j)\\&\quad - \xi(\mathbf{h}_i, -\mathbf{h}_j, 1-s_i, s_j) + \xi(\mathbf{h}_i, \mathbf{h}_j, 1-s_i, 1-s_j)\end{aligned} \tag{4.16}$$

If $\mathbf{V}$ *is positive definite, then any estimator* $g(\mathbf{x})$ *of* $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ *will satisfy*

$$\mathcal{M}(g) \geq \mathbf{u}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{u} \tag{4.17}$$

We note here that since the choice of K, $\{s_i\}_{i=1}^K$ and $\{\mathbf{h}_i\}_{i=1}^K$ in Theorem 3 are arbitrary, with each choice resulting in possibly different values for the bound, Theorem 3 actually provides a family of lower bounds rather than a single lower bound. The tightest lower bound from this family is obtained by maximizing $\mathcal{M}(g)$ over the choice of K, $\{s_i\}_{i=1}^K$ and $\{\mathbf{h}_i\}_{i=1}^K$. This is a difficult optimization problem to solve analytically. In this chapter, we evaluate the bounds using the approach of [38], where it is suggested that the tightest bound can be characterized closely by setting $s_i = 1/2$ for all i, and choosing a small number K of test points $\{\mathbf{h}_i\}_{i=1}^K$ from the set of values of $\mathbf{h}$ for which $f_0(\mathbf{x}+\mathbf{G}\mathbf{h})$ is non-zero. The improvement in tightness obtained by increasing K falls sharply with respect to K.

Theorem 4. (Obtained from the Ziv-Zakai Bound) *Given the location parameter problem*

of Definition 1, assume that $f_0(\mathbf{x})$ has bounded support (i.e there exists a $\mu > 0$ such that $x_i \geq \mu$ for all $i = 1, \cdots, N$ necessarily implies that $f_0(\mathbf{x}) = 0$). Define

$$L(\mathbf{x}, \mathbf{h}) = \log \left[\frac{f_0(\mathbf{x} - \mathbf{G}\mathbf{h})}{f_0(\mathbf{x})} \right] \tag{4.18}$$

and the valley filling function

$$\mathcal{V}\{f(h)\} = \max_{\xi > 0} f(h + \xi). \tag{4.19}$$

Then the inequality

$$\mathcal{M}(g) \geq \int_0^\infty \mathcal{V} \left\{ \max_{\mathbf{h}:\mathbf{c}^{\mathrm{T}}\mathbf{h}=h} Pr\{L(\mathbf{x}, \mathbf{h}) > 0\} \right\} h \ dh \tag{4.20}$$

holds for any estimator $g(\mathbf{x})$ of the scalar $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$.

While Theorems 3 and 4 only provide lower bounds on the maximum risk $\mathcal{M}(g)$, it is possible to extend the applicability of these bounds to the conditional risk $R(\boldsymbol{\theta}, g)$ if we consider a restricted class of estimators that are *shift invariant*. Given the location parameter problem of Definition 1, we say that an estimator $g(\mathbf{x})$ is *shift invariant* if for the same matrix $\mathbf{G}$ used in (4.10),

$$g(\mathbf{x} + \mathbf{G}\mathbf{h}) = g(\mathbf{x}) + \mathbf{c}^{\mathrm{T}}\mathbf{h} \quad \forall \ \ \mathbf{h} \in \mathbb{R}^{M \times 1} \tag{4.21}$$

This condition implies that a shift in the observation vector causes a corresponding shift in the estimate, with these shifts sharing a linear relationship with one another. A useful property of shift invariant estimators is that they have constant conditional risk in location parameter problems. To demonstrate this property, we note that if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are any two values of the parameter vector with $\mathbf{h} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$, then for any shift invariant estimator we

have

$$R(\boldsymbol{\theta}_1, g) = \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_1]^2 f(\mathbf{x}|\boldsymbol{\theta}_1)\mathrm{d}\mathbf{x} \tag{4.22}$$

$$= \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}(\boldsymbol{\theta}_2 + \mathbf{h})]^2 f_0(\mathbf{x} - \mathbf{G}(\boldsymbol{\theta}_2 + \mathbf{h}))\mathrm{d}\mathbf{x} \tag{4.23}$$

$$= \int_{\mathbb{R}^N} [g(\mathbf{x} - \mathbf{G}\mathbf{h}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_2]^2 f((\mathbf{x} - \mathbf{G}\mathbf{h})|\boldsymbol{\theta}_2)\mathrm{d}\mathbf{x} \tag{4.24}$$

$$= \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_2]^2 f(\mathbf{x}|\boldsymbol{\theta}_2)\mathrm{d}\mathbf{x} = R(\boldsymbol{\theta}_2, g) \tag{4.25}$$

Hence, for shift invariant estimators, we have $\mathcal{M}(g) = R(\boldsymbol{\theta}, g)$, since the conditional risk $R(\boldsymbol{\theta}, g)$ is constant with respect to $\boldsymbol{\theta}$. Thus, the lower bounds on $\mathcal{M}(g)$ in Theorems 3 and 4 are also lower bounds on $R(\boldsymbol{\theta}, g)$ for such estimators.

4.4 Application of Estimator Bounds to the Phase Offset Estimation Problem

Recall from Section 4.2 that our problem is to estimate $\delta = \mathbf{c}_0^{\mathrm{T}}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = [\delta\ d]^{\mathrm{T}}$ and $\mathbf{c}_0 = [1\ 0]^{\mathrm{T}}$, from the observation vector $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w}$. Since $f(\mathbf{y}|\boldsymbol{\theta})$ satisfies $f(\mathbf{y}|\boldsymbol{\theta}) = f_{\mathbf{W}}(\mathbf{y} - \mathbf{A}\boldsymbol{\theta})$, the POE problem can be classified as a location parameter problem. Hence, by setting $\mathbf{G} = \mathbf{A}$ and $f_0(\cdot) = f_{\mathbf{W}}(\cdot)$ in the result of Theorems 3 and 4, bounds can be obtained for the general POE problem where only the joint pdf $f_{\mathbf{W}}(\mathbf{w})$ is known. We now further simplify the results of Theorems 3 and 4 under the special case of i.i.d. forward and reverse queuing delays (as defined by eqns. (4.8) and (4.9) of Section 4.2).

4.4.1 Simplification of Theorem 1

Theorem 3 states that every estimator $\hat{\delta}(\mathbf{y})$ of δ will satisfy (4.17), provided $\mathbf{V} \succ 0$. Towards obtaining simplified expressions for the elements of $\mathbf{u}$ and $\mathbf{V}$, let $\{s_i\}_{i=1}^K$ and $\{\mathbf{h}_i\}_{i=1}^K$ be arbitrary scalars and $M \times 1$ vectors respectively. Further, let $h_i = \mathbf{c}_0^{\mathrm{T}}\mathbf{h}_i$, $\mathbf{a}_1 = [1\ 1]^{\mathrm{T}}$ and $\mathbf{a}_2 = [-1\ 1]^{\mathrm{T}}$. Assuming i.i.d. forward and reverse queing delays, it is easy to show (using the change of variables $\mathbf{w} = \mathbf{y} - \mathbf{A}\boldsymbol{\theta}$) that $\xi(\mathbf{h}_1, \mathbf{h}_2, s_1, s_2)$ simplifies to the form

$$\begin{aligned}
&\xi(\mathbf{h}_1, \mathbf{h}_2, s_1, s_2) \\
&= \int_{\mathbf{w}} \left[\frac{f_{\mathbf{W}}(\mathbf{w} + \mathbf{A}\mathbf{h}_1)}{f_{\mathbf{W}}(\mathbf{w})}\right]^{s_1} \left[\frac{f_{\mathbf{W}}(\mathbf{w} + \mathbf{A}\mathbf{h}_2)}{f_{\mathbf{W}}(\mathbf{w})}\right]^{s_2} f_{\mathbf{W}}(\mathbf{w}) d\mathbf{w} \qquad (4.26) \\
&= \prod_{k=1}^{2} \left[\mathrm{E}_k \left\{ \left[\frac{f_k\left(w + \mathbf{a}_k^{\mathrm{T}}\mathbf{h}_1\right)}{f_k\left(w\right)}\right]^{s_1} \left[\frac{f_k\left(w + \mathbf{a}_k^{\mathrm{T}}\mathbf{h}_2\right)}{f_k\left(w\right)}\right]^{s_2} \right\}\right]^P \qquad (4.27)
\end{aligned}$$

where $\mathrm{E}_k\{\cdot\}$ (for $k = 1, 2$) represents an expectation taken with respect to the density $f_k(w)$ (as defined in (4.8)). The final expressions for the elements of $\mathbf{u}$ and $\mathbf{V}$ (corresponding to eqns. (4.15) and (4.16)) are

$$u_i = h_i \cdot \xi(\mathbf{h}_i, \mathbf{0}_{2\times 1}, 1 - s_i, 0) \qquad (4.28)$$

where $\mathbf{0}_{2\times 1} = [0\ 0]^{\mathrm{T}}$, and

$$\begin{aligned}
V_{ij} = &\xi(-\mathbf{h}_i, -\mathbf{h}_j, s_i, s_j) - \xi(-\mathbf{h}_i, \mathbf{h}_j, s_i, 1 - s_j) \\
&- \xi(\mathbf{h}_i, -\mathbf{h}_j, 1 - s_i, s_j) + \xi(\mathbf{h}_i, \mathbf{h}_j, 1 - s_i, 1 - s_j)\ . \qquad (4.29)
\end{aligned}$$

As stated in Section 4.3, the values K, $\{s_i\}_{i=1}^K$ and $\{\mathbf{h}_i\}_{i=1}^K$ that lead to the tightest lower bound on error variance still need to be selected. For the POE problem, we observed that $s_i = 1/2$ for all i led to the tightest bounds. Further, depending on the nature of $f_1(w)$ and $f_2(w)$, between $K = 50$ and $K = 500$ test points sampled uniformly from the set

$\{[h_1\ h_2]^{\mathrm{T}} : (h_1+h_2) \in [0, M_1]$ and $(h_1-h_2) \in [0, M_2]\}$, were sufficient to generate maximally tight bounds (recall that $[0, M_1]$ and $[0, M_2]$ represent the support sets of $f_1(w)$ and $f_2(w)$ respectively).

4.4.2 Simplification of Theorem 2

Applying Theorem 2 to the POE problem, we obtain

$$\mathcal{M}(\hat{\delta}) \geq \int_0^{\infty} \mathcal{V}\left\{\max_{h_2}\ \Pr\left\{L(\mathbf{w}, [h_1\ h_2]^{\mathrm{T}}) > 0\right\}\right\} h_1\ \mathrm{d}h_1 \tag{4.30}$$

$$= \int_0^{\infty} \max_{\xi>0,\ h_2}\ \Pr\left\{L(\mathbf{w}, [(h_1+\xi)\ h_2]^{\mathrm{T}}) > 0\right\} h_1\ \mathrm{d}h_1 \tag{4.31}$$

where $\mathbf{w}$ is a random vector with p.d.f. $f_{\mathbf{W}}(\mathbf{w})$, as defined in (4.9). Further, assuming i.i.d. forward and reverse queuing delays, the expression for $L(\mathbf{w}, \mathbf{h})$ can be simplified as follows

$$L(\mathbf{w}, \mathbf{h}) = \sum_{i=1}^{P} \log\left[\frac{f_1\left(w_1^{(i)} - h_1 - h_2\right)}{f_1(w_1^{(i)})}\right] + \sum_{i=1}^{P} \log\left[\frac{f_2\left(w_2^{(i)} - h_1 + h_2\right)}{f_2(w_2^{(i)})}\right] \tag{4.32}$$

4.5 Simulation Results

We now compare the performance of existing POE schemes against the WWB and ZZB derived in Section 4.4. We shall characterize performance by plotting the estimation error variance against the observation window size P. We note that typically the time evolution of the phase error is used to characterize the performance of POE schemes [17][22]. Here the time required for the phase error to converge to a small neighborhood of zero is a metric of key interest. In the context of such performance evaluations, the WWB and ZZB tell us the minimum window size required to achieve a desired level of convergence with respect to

the variance of the phase error. Since the maximum window size in such problems is limited by the duration of time over which the phase offset and fixed delays can be assumed to be approximately constant, our bounds can help network designers evaluate if the convergence of the error variance to a desired level is feasible at all for a given network scenario. Further, given information about the background traffic conditions, our bounds can help network designers determine the maximum number of switches that can be allowed between the master and slave nodes for a desired level of POE performance.

The existing POE schemes we consider are the sample minimum, maximum, mean and median filtering schemes. Given the observation vector $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{T}} & \mathbf{y}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, these schemes use an estimator of the form

$$\hat{\delta} = \frac{1}{2} [\ g(\mathbf{y}_1) - g(\mathbf{y}_2)\ ] \tag{4.33}$$

where $g(.)$ may be either the sample minimum, maximum, mean or median functions. It is easy to show that all of these conventional estimators are shift invariant, and hence have a constant conditional risk that is independent of $\boldsymbol{\theta}$. Thus, in order to characterize the performance of these schemes, it is sufficient to determine the estimator error variance at $\boldsymbol{\theta} = [\delta\ d]^{\mathrm{T}} = [0\ 0]^{\mathrm{T}}$.

In order to evaluate the performance of these schemes, we assume a Gigabit ethernet network consisting of a cascade of N switches between a master and a slave node. Each switch is assumed to be a store-and-forward switch that implements strict priority queuing. We consider three types of background traffic flows in this network:

(a) *Cross Traffic Flows*: In cross traffic flows [45], fresh background traffic is injected at each switch along the master-slave path, and this traffic exits the master-slave path at the subsequent switch (see 4-switch example in Fig. 4.1a). The arrival times and sizes

of background traffic packets injected at each switch were assumed to be statistically independent of traffic at other switches.

(b) *Inline Traffic Flows*: In inline traffic flows [22], background traffic is injected only at the first switch along the master-slave path, and this traffic travels along the same path as synchronization traffic through the entire cascade of N switches (see 4-switch example in Fig. 4.1b).

(c) *Mixed Traffic Flows*: Here a mixture of cross and inline traffic flows are present in the network.

As discussed in Section I, mobile network operators typically lease the backhaul network from commercial ISPs, and the use of this backhaul network is shared among several users. In the context of such backhaul networks, traffic generated by other users of the network can be typically modeled as cross traffic flows, while inline traffic flows can be used to model non-synchonization traffic between the master and the slave nodes.

With regard to the packet size distributions of background traffic, we use Traffic Models 1 (TM1) and 2 (TM2)from the ITU-T recommendation G.8261 [45] for cross traffic flows, as described in Table 4.1. We also consider a third traffic model, where packet sizes are uniformly distributed between 64 and 1500 bytes [22] for inline traffic scenarios. Further, we consider both symmetric and asymmetric assumptions for the background traffic occurring along the forward path versus the reverse path.

We consider the following specific scenarios:

(i) Cross Traffic Flows, Symmetric Traffic, TM1.

(ii) Cross Traffic Flows, Symmetric Traffic, TM2.

Traf. Model	Packet Sizes (Bytes)	% of Load
TM1	$\{64, 576, 1518\}$	$\{80\%, 5\%, 15\%\}$
TM2	$\{64, 576, 1518\}$	$\{30\%, 10\%, 60\%\}$

Table 4.1: Models for composition of background traffic packets

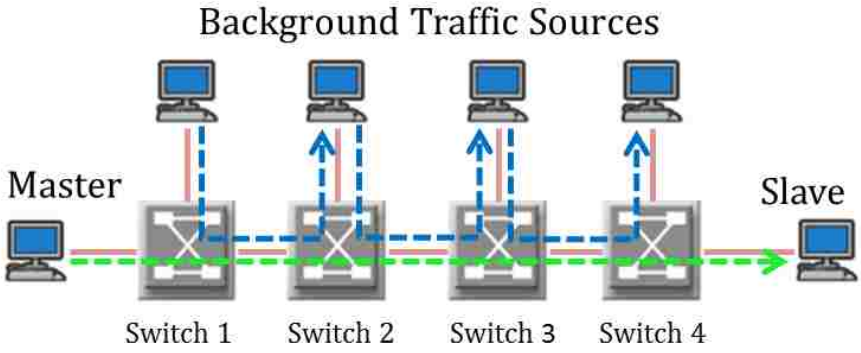
(iii) Inline Traffic Flows, Symmetric Traffic, TM2.

(iv) Mixed Traffic Flows, Symmetric Traffic, TM1 for cross traffic, uniform packet size distribution for inline traffic.

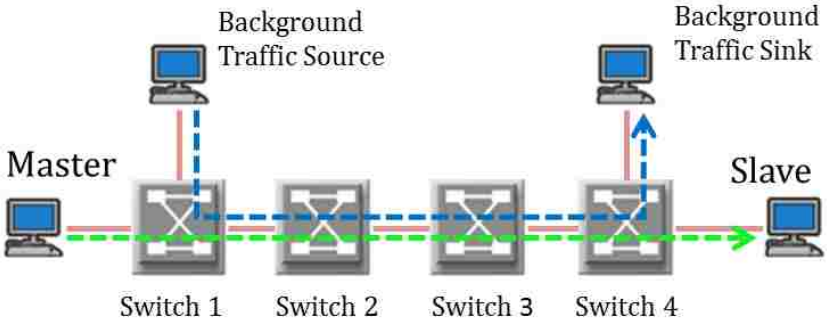
(v) Cross Traffic Flows, Asymmetric Traffic, TM1 for forward path and TM2 for reverse path.

For the load factor, i.e. the percentage of the link capacity consumed by background traffic, we consider values between 20 - 80% of the link capacity. We assume that the interarrival times between packets in all background traffic flows follow exponential distributions, and set the rate parameter of each exponential distribution to obtain the desired load factor.

Empirical pdfs of the queuing delays (Figs. 4.2a - 4.2f) were obtained using the OPNET network simulator [46] for cross traffic flows, and a custom MATLAB-based network simulator for inline and mixed traffic flows . Our simulations assumed that all switches were store-and-forward switches that implemented strict priority queuing. Without loss of generality, we assume that fixed minimum delay components of the ETE delays equal zero, hence the support of $f_{\mathrm{w}}(w)$ always begins at zero in the plots. These empirical densities are used to obtain lower bounds using the simplified expressions derived in Section 4.4. The WWB-based performance bound of Theorem 1 was evaluated numerically using Riemann sums, while the ZZB-based performance bound of Theorem 2 was evaluated using Monte-Carlo simulation techniques. The resulting bounds on the standard deviation of estimation error



(a) Cross Traffic Flows



(b) Inline Traffic Flows

Figure 4.1: Example of a four switch network with cross and inline traffic flows. Red lines indicate Gigabit ethernet links, dotted blue lines indicate the direction of background traffic flows, while the dotted green line represents the direction of synchronization traffic flow.

are compared with the error standard derivation of various conventional estimation schemes at different sample sizes in Figs. 4.4 - 4.10, with legend provided in Fig. 4.3).
In order to compare the results against the LTE synchronization requirement of 1.25 μs of synchronization accuracy, the estimation error standard deviation required so that the absolute estimation error lies under 1.25 μs with a 5-sigma level of certainty is also plotted over the curves. Here the 5-sigma level of certainty implies that on average, about 6 out of every 10 million estimates will have absolute estimation error that exceeds 1.25 μs.

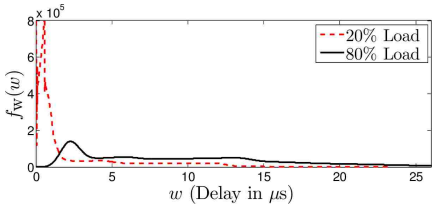
Based on the simulations results, we made the following observations:

1. *Estimator Standard Deviation vs Theoretical Limits*: Typically, for a fixed value of P, the standard deviation of estimator error is 10-80% higher for the best conventional estimation scheme over the tightest estimator bound, either the ZZB or the WWB.

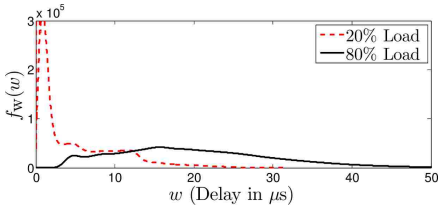
2. *Number of samples required vs Theoretical Limits*: The tightest estimator bound achieves the LTE synchronization requirement using 10-60% fewer samples than the best performing conventional estimator.

3. *Effect of increase in load factor*: The number of samples required both by the bounds as well as different conventional estimation schemes to achieve the LTE synchronization requirement tends to grow significantly (by roughly 4 to 80 times) between the 20% and 80% load cases. This can be attributed to the fact that at lower loads, a fair fraction of samples with zero queuing delays occur, while such samples disappear at higher loads.

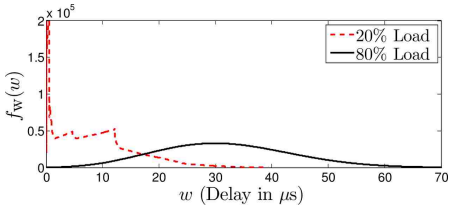
Our numerical results lead us to the following conclusions. Each conventional estimation scheme is particularly suited for a certain kind of queuing delay distribution, achieving near-optimum performance under a suitable scenario. Suppose we measure the closeness of an estimator to the lower bounds as the percentage increase in the number of samples required



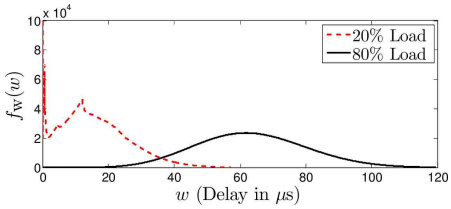
(a) 10 Switches, TM1



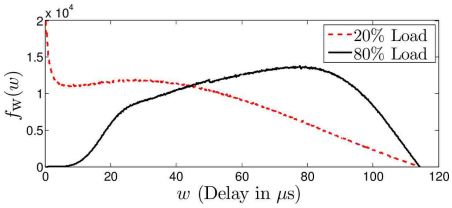
(b) 10 Switches, TM2



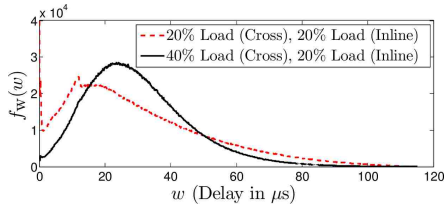
(c) 20 Switches, TM1



(d) 20 Switches, TM2



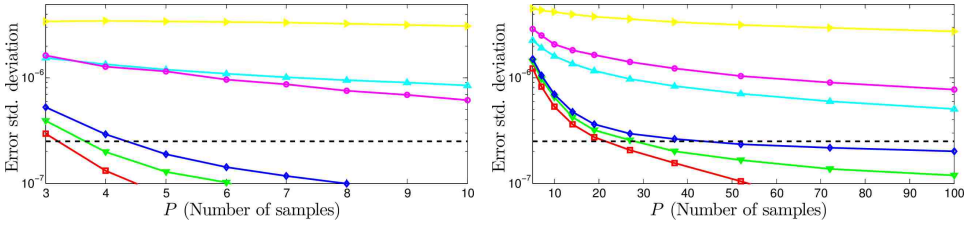
(e) 10 Switches, Inline Traffic



(f) 10 Switches, Mixed Traffic

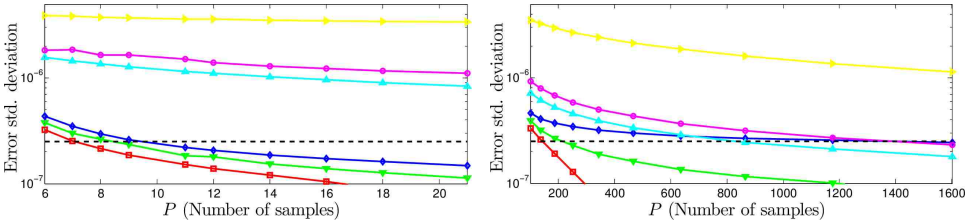
Figure 4.2: Plots of $f_0(x)$ under varying network conditions.

Figure 4.3: Legend common to Figs. 4.4 - 4.10.



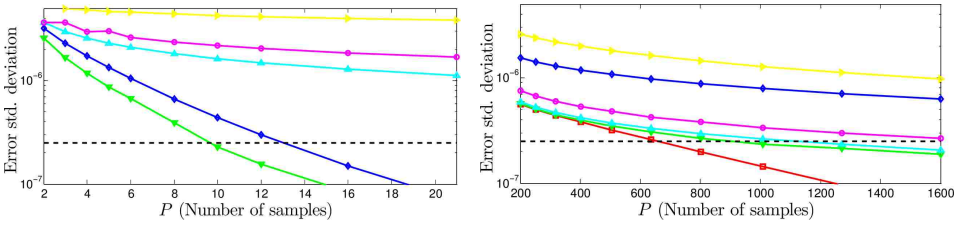
(a) 20% Load (b) 80% Load

Figure 4.4: Plots of the standard deviation of estimator error with 10 switches and cross traffic flows distributed according to TM1, under varying load factors.



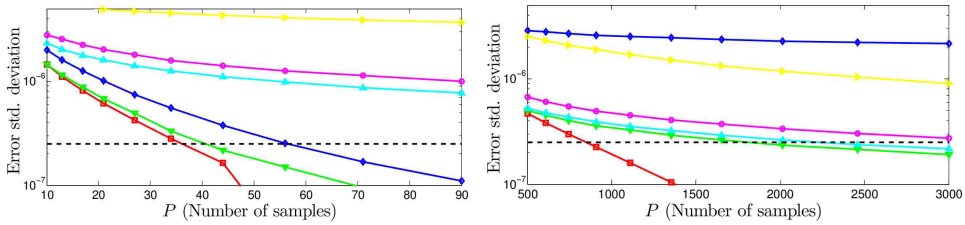
(a) 20% Load (b) 80% Load

Figure 4.5: Plots of the standard deviation of estimator error with 10 switches and cross traffic flows distributed according to TM2, under varying load factors.



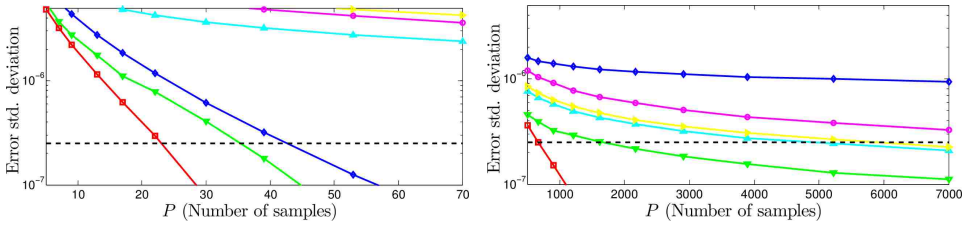
(a) 20% Load (b) 80% Load

Figure 4.6: Plots of the standard deviation of estimator error with 20 switches and cross traffic flows distributed according to TM1, under varying load factors.



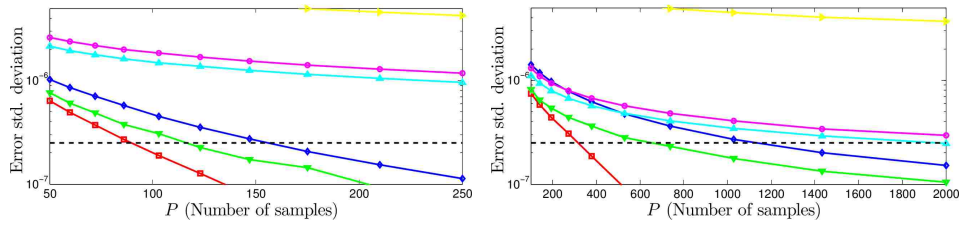
(a) 20% Load (b) 80% Load

Figure 4.7: Plots of the standard deviation of estimator error with 20 switches and cross traffic flows distributed according to TM2, under varying load factors.



(a) 20% Load (b) 80% Load

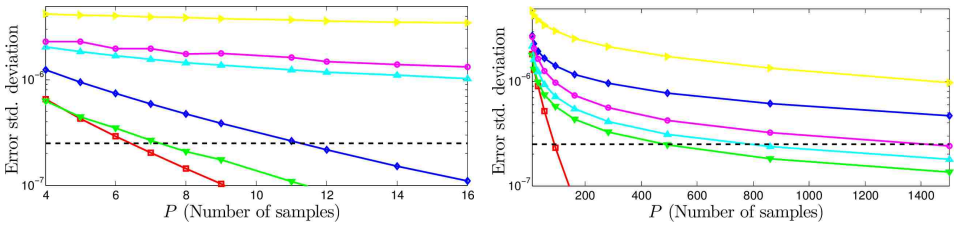
Figure 4.8: Plots of the standard deviation of estimator error with 10 switches and inline traffic flows, under varying load factors.



(a) 20% Load (Inline), 20% Load (Cross, TM2)

(b) 20% Load (Inline), 40% Load (Cross, TM2)

Figure 4.9: Plots of the standard deviation of estimator error with 10 switches and mixed traffic flows, under varying load factors.



(a) 20% Load (TM1), 20% Load (TM2) (b) 80% Load (TM1), 80% Load (TM2)

Figure 4.10: Plots of the standard deviation of estimator error with 10 switches and asymmetric cross traffic flows (TM1 for forward path and TM2 for reverse path), under varying load factors.

by the estimator to achieve the LTE synchronization threshold, as compared to the tightest lower bound. Then we observe that sample minimum filtering comes very close to achieving the lower bound in Figs. 4.4a and 4.5a. Here the combination of a low load factor and a small number of switches results in a significant concentration of probability mass near zero delay in the queuing delay pdf. On the other hand, sample mean filtering comes close to achieving the lower bound in Figs. 4.6b, 4.7b and 4.10b. These are specific high load scenarios that are well suited to sample-mean filtering. We similarly believe that there also exist network scenarios, other than those considered in this chapter, where the sample maximum and median schemes will come close to achieving the lower bounds. However, under many other low and high load scenarios, such as in Figs. 4.4b, 4.5b, 4.6a, 4.8a, 4.8b, 4.9a, 4.9b and 4.10a, no conventional estimation scheme comes close to the lower bounds. In these scenarios, our bounds indicate that significant performance gains could be achieved by considering other more suitable estimators.

4.6 Summary

In this chapter, we describe new lower bounds on the maximum risk of estimators for the POE problem. These new bounds, first described for a general vector parameter estimation problem, are subsequently simplified for the POE problem. Simulations compare the performance of several POE schemes against these bounds under different network conditions. Results indicate that while conventional estimators come close to achieving the bounds in some scenarios, there are also many scenarios where it may be possible to achieve significant performance gains via the use of better POE schemes. Our future work will aim to address whether or not these bounds are achievable.

4.7 Acknowledgment

The authors would like to thank OPNET for providing the network simulation software free of charge under the University License agreement.

4.8 Appendix

4.8.1 Proof of Theorem 3

We begin the proof with the straightforward inequality

$$\mathcal{M}(g) = \max_{\boldsymbol{\theta}} \ \mathrm{E}\left\{[g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}]^2 \mid \boldsymbol{\theta}\right\} \tag{4.34}$$

$$\geq \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{E}\left\{[g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}]^2 | \ \boldsymbol{\theta}\right\} p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = B(g,p) \tag{4.35}$$

i.e. the maximum risk exceeds the Bayes risk for any choice of prior distribution $p(\boldsymbol{\theta})$. Define

$$\tilde{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{f(\mathbf{x}|\boldsymbol{\theta}_1)p(\boldsymbol{\theta}_1)}{f(\mathbf{x}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2)} \tag{4.36}$$

The original WWB [38] states that

$$B(g, p) \geq \tilde{\mathbf{u}}^{\mathrm{T}} \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{u}} \tag{4.37}$$

provided that $\tilde{\mathbf{V}}$ is positive definite. In (4.37), $\tilde{\mathbf{u}}$ is a $K \times 1$ vector whose i^{th} element is

$$\tilde{u}_i = (\mathbf{c}^{\mathrm{T}} \mathbf{h}_i) \cdot \mathrm{E}\left[\tilde{L}^{1-s_i}(\boldsymbol{\theta} - \mathbf{h}_i, \boldsymbol{\theta})\right] \tag{4.38}$$

while $\tilde{\mathbf{V}}$ is a $K \times K$ matrix whose (i, j) element is

$$\begin{aligned} \tilde{V}_{ij} = \mathrm{E}\Big\{ & [\tilde{L}^{s_i}(\boldsymbol{\theta} + \mathbf{h}_i, \boldsymbol{\theta}) - \tilde{L}^{1-s_i}(\boldsymbol{\theta} - \mathbf{h}_i, \boldsymbol{\theta})] \\ & \cdot [\tilde{L}^{s_j}(\boldsymbol{\theta} + \mathbf{h}_j, \boldsymbol{\theta}) - \tilde{L}^{1-s_j}(\boldsymbol{\theta} - \mathbf{h}_j, \boldsymbol{\theta})] \Big\} , \end{aligned} \tag{4.39}$$

with expectations computed over both $\mathbf{x}$ and $\boldsymbol{\theta}$. Note that the definitions of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{V}}$ are obtained from [38], and are distinct from the definitions of $\mathbf{u}$ and $\mathbf{V}$ in the statement of Theorem 3. In [38], the vectors $\{\mathbf{h}_i\}_{i=1}^{K}$ in the definitions of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{V}}$ are referred to as *test points* . From (4.35) and (4.37), we see that the inequality $\mathcal{M}(g) \geq \tilde{\mathbf{u}}^{\mathrm{T}} \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{u}}$ holds for any prior distribution $p(\boldsymbol{\theta})$. In order to maximize the tightness of this lower bound, we should pick the $p(\boldsymbol{\theta})$ that maximizes $\tilde{\mathbf{u}}^{\mathrm{T}} \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{u}}$. However, the optimum choice of $p(\boldsymbol{\theta})$ is difficult to obtain analytically. Instead, here we consider a sequence of prior distributions $p(\boldsymbol{\theta})$ that result in monotonically increasing values for $\mathcal{M}(g)$, and determine the limiting value for $\mathcal{M}(g)$. To this end, consider a prior distribution $p(\boldsymbol{\theta})$ that is uniformly distributed over its support set

$\boldsymbol{\Theta}$, i.e.

$$p(\boldsymbol{\theta}) = \begin{cases} 1/S & \text{if } \boldsymbol{\theta} \in \boldsymbol{\Theta} \\ 0 & \text{otherwise} \end{cases} \tag{4.40}$$

where $S = \int_{\boldsymbol{\Theta}} d\boldsymbol{\theta}$. Define

$$\begin{aligned} &\xi(\mathbf{h}_1, \mathbf{h}_2, s_1, s_2) \\ &= \int_{\mathbb{R}^N} \left[\frac{f_0(\mathbf{x}+\mathbf{G}\mathbf{h}_1)}{f_0(\mathbf{x})}\right]^{s_1} \left[\frac{f_0(\mathbf{x}+\mathbf{G}\mathbf{h}_2)}{f_0(\mathbf{x})}\right]^{s_2} f_0(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{4.41}$$

Then the i^{th} element of $\tilde{\mathbf{u}}$ can be simplified as

$$\begin{aligned} \tilde{u}_i &= (\mathbf{c}^{\mathrm{T}}\mathbf{h}_i) \cdot \int_{\boldsymbol{\Theta}} \int_{\mathbb{R}^N} \left[\frac{f(\mathbf{x}|\boldsymbol{\theta}-\mathbf{h}_i)p(\boldsymbol{\theta}-\mathbf{h}_i)}{f(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}\right]^{1-s_i} \\ &\qquad \cdot f(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \end{aligned} \tag{4.42}$$

$$\begin{aligned} &= (\mathbf{c}^{\mathrm{T}}\mathbf{h}_i) \cdot \int_{\{\boldsymbol{\theta}:\boldsymbol{\theta},(\boldsymbol{\theta}-\mathbf{h}_i)\in\boldsymbol{\Theta}\}} \int_{\mathbb{R}^N} \\ &\quad \left[\frac{f_0(\mathbf{x}-\mathbf{G}\boldsymbol{\theta}+\mathbf{G}\mathbf{h}_i)}{f_0(\mathbf{x}-\mathbf{G}\boldsymbol{\theta})}\right]^{1-s_i} f_0(\mathbf{x}-\mathbf{G}\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \end{aligned} \tag{4.43}$$

$$\text{(Setting } \tilde{\mathbf{x}} = \mathbf{x} - \mathbf{G}\boldsymbol{\theta}\text{)}$$

$$\begin{aligned} &= (\mathbf{c}^{\mathrm{T}}\mathbf{h}_i) \cdot \int_{\{\boldsymbol{\theta}:\boldsymbol{\theta},(\boldsymbol{\theta}-\mathbf{h}_i)\in\boldsymbol{\Theta}\}} \int_{\mathbb{R}^N} \left[\frac{f_0(\tilde{\mathbf{x}}+\mathbf{G}\mathbf{h}_i)}{f_0(\tilde{\mathbf{x}})}\right]^{1-s_i} \\ &\qquad \cdot \ f_0(\tilde{\mathbf{x}}) p(\boldsymbol{\theta}) d\tilde{\mathbf{x}} d\boldsymbol{\theta} \end{aligned} \tag{4.44}$$

$$= \mathcal{D}_i \cdot (\mathbf{c}^{\mathrm{T}}\mathbf{h}_i) \cdot \xi(\mathbf{h}_i, \mathbf{0}_{M\times 1}, 1-s_i, 0) = \mathcal{D}_i u_i \tag{4.45}$$

where

$$\mathcal{D}_i = \int_{\{\boldsymbol{\theta}:\boldsymbol{\theta},(\boldsymbol{\theta}-\mathbf{h}_i)\in\boldsymbol{\Theta}\}} p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \frac{1}{S} \int_{\{\boldsymbol{\theta}:\boldsymbol{\theta},(\boldsymbol{\theta}-\mathbf{h}_i)\in\boldsymbol{\Theta}\}} d\boldsymbol{\theta}$$

and u_i is as defined in the statement of the theorem. Now consider a support set of the form

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M) \ : \ \theta_i \in [-B, B] \ \ \forall i\}$$

For any bounded set of test points $\{\mathbf{h}_i\}_{i=1}^K$, we have

$$\lim_{B\to\infty} \mathcal{D}_i = \lim_{B\to\infty} \frac{1}{S} \int_{\{\boldsymbol{\theta}:\boldsymbol{\theta},(\boldsymbol{\theta}-\mathbf{h}_i)\in\boldsymbol{\Theta}\}} d\boldsymbol{\theta} = 1$$

Hence, under this limiting prior distribution, we obtain $\tilde{u}_i = u_i$. Note that this argument does not require the existence of an uniform prior distribution over an infinite support set, it merely implies that the difference between $\tilde{u}_i$ and u_i can be made arbitrarily small by choosing larger and larger values for $(u_j - l_j)$.

Using a similar argument, it can also be shown that $\tilde{V}_{ij} = V_{ij}$ under the same limiting distribution, where V_{ij} is as defined in the statement of the theorem. Thus, we obtain $\mathcal{M}(g) \geq \mathbf{u}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{u}$, proving eqn. (4.17) in the statement of the theorem. This concludes the proof.

4.8.2 Proof of Theorem 4

As in the proof of Theorem 3, we begin with the straightforward inequality $\mathcal{M}(g) \geq B(g,p)$. Further, it is easy to show using the original (unmodified) ZZB ([44], eq. (32)) that

$$B(g,p) \geq \frac{1}{2}\int_0^\infty \mathcal{V}\Bigg\{ \max_{\mathbf{h}:\mathbf{c}^{\mathrm{T}}\mathbf{h}=h} \int_{\boldsymbol{\varphi}\in\boldsymbol{\Theta}} [p(\boldsymbol{\varphi}) + p(\boldsymbol{\varphi}+\mathbf{h})] \cdot P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi}+\mathbf{h}) \mathrm{d}\boldsymbol{\varphi} \Bigg\} h \ \mathrm{d}h \tag{4.46}$$

where

$$\mathcal{V}\{f(h)\} = \max_{\xi>0} f(h+\xi) \tag{4.47}$$

is the *valley filling function*, as defined in [44]. Here $P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi}+\mathbf{h})$ corresponds to the probability of error of the optimum decision rule for the Bayesian hypothesis testing problem

$$\begin{aligned} \mathcal{H}_1: \quad & \boldsymbol{\theta} = \boldsymbol{\varphi} && \mathbf{x} \sim f(\mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\varphi}) \\ \mathcal{H}_0: \quad & \boldsymbol{\theta} = \boldsymbol{\varphi} + \mathbf{h} && \mathbf{x} \sim f(\mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\varphi} + \mathbf{h}) \end{aligned} \tag{4.48}$$

where $\Pr(\mathcal{H}_1) = p(\boldsymbol{\varphi})$ and $\Pr(\mathcal{H}_0) = p(\boldsymbol{\varphi}+\mathbf{h})$.

Thus, we see that the inequality

$$\mathcal{M}(g) \geq \frac{1}{2}\int_0^\infty \mathcal{V}\Bigg\{ \max_{\mathbf{h}:\mathbf{c}^{\mathrm{T}}\mathbf{h}=h} \int_{\boldsymbol{\varphi}\in\boldsymbol{\Theta}} [p(\boldsymbol{\varphi}) + p(\boldsymbol{\varphi}+\mathbf{h})] \cdot P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi}+\mathbf{h}) \mathrm{d}\boldsymbol{\varphi} \Bigg\} h \ \mathrm{d}h \tag{4.49}$$

holds for any prior distribution $p(\boldsymbol{\theta})$. In order to maximize the tightness of this lower bound, we should pick the $p(\boldsymbol{\theta})$ that maximizes the right side of the above equation. However, the optimum choice of $p(\boldsymbol{\theta})$ is difficult to obtain analytically. Hence, as in the proof of Theorem 3, we consider a sequence of prior distributions $p(\boldsymbol{\theta})$ that result in monotonically increasing values for $\mathcal{M}(g)$, and determine the limiting value for $\mathcal{M}(g)$. To this end, consider again a prior distribution that is uniformly distributed over its support set $\boldsymbol{\Theta}$, i.e.

$$p(\boldsymbol{\theta}) = \begin{cases} 1/S & \text{if } \boldsymbol{\theta} \in \boldsymbol{\Theta} \\ 0 & \text{otherwise} \end{cases} \tag{4.50}$$

where $S = \int_{\boldsymbol{\Theta}} d\boldsymbol{\theta}$. Further, consider two cases:

(i) If either $\boldsymbol{\varphi} \notin \boldsymbol{\Theta}$ or $\boldsymbol{\varphi}+\mathbf{h} \notin \boldsymbol{\Theta}$:

Here a decision rule that always decides in favor of either $\mathcal{H}_0$ (if $\boldsymbol{\varphi} \notin \boldsymbol{\Theta}$ and $\boldsymbol{\varphi} + \mathbf{h} \in \boldsymbol{\Theta}$) or $\mathcal{H}_1$ (if $\boldsymbol{\varphi}+\mathbf{h} \notin \boldsymbol{\Theta}$) will have a zero error probability, hence we will have $P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi}+\mathbf{h}) = 0$.

(ii) If both $\boldsymbol{\varphi} \in \boldsymbol{\Theta}$ and $\boldsymbol{\varphi} + \mathbf{h} \in \boldsymbol{\Theta}$:

Here the optimum decision rule is given by the likelihood ratio test, whose probability of error is given as

$$P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h}) = \Pr\left\{\hat{L}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h}) > l \mid \mathcal{H}_1\right\} \tag{4.51}$$

where

$$\hat{L}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h}) = \log\left[\frac{f(\mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\varphi} + \mathbf{h})}{f(\mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\varphi})}\right] \tag{4.52}$$

$$= \log\left[\frac{f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\varphi} - \mathbf{G}\mathbf{h})}{f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\varphi})}\right] \tag{4.53}$$

and $l = \log\left[p(\boldsymbol{\varphi})/p(\boldsymbol{\varphi} + \mathbf{h})\right] = \log(1) = 0$. Define

$$I(p) = \begin{cases} 1 & \text{if } p > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.54}$$

Then $P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h})$ can be simplified as

$$P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h})$$

$$= \Pr\left\{\hat{L}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h}) > 0 \mid \mathcal{H}_1\right\} \tag{4.55}$$

$$= \int_{\mathbb{R}^N} I\left(\log\left[\frac{f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\varphi} - \mathbf{G}\mathbf{h})}{f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\varphi})}\right]\right) \tag{4.56}$$

$$\cdot f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\varphi})\mathrm{d}\mathbf{x} \tag{4.57}$$

$$= \int_{\mathbb{R}^N} I\left(\log\left[\frac{f_0(\mathbf{x} - \mathbf{G}\mathbf{h})}{f_0(\mathbf{x})}\right]\right) f_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{4.58}$$

$$= \Pr\{L(\mathbf{x}, \mathbf{h}) > 0\} = P_{\min}(\mathbf{h}) \tag{4.59}$$

where $L(\mathbf{x}, \mathbf{h})$ and $P_{\min}(\mathbf{h})$ are as defined earlier in this proof.

Substituting the values computed for $P_{\min}(\boldsymbol{\varphi}, \boldsymbol{\varphi} + \mathbf{h})$ under these two cases into (4.49) and

simplifying, we obtain

$$\mathcal{M}(g) \geq \frac{1}{2}\int_0^\infty \mathcal{V}\left\{ \max_{\mathbf{h}:\mathbf{c}^{\mathrm{T}}\mathbf{h}=h} P_{\min}(\mathbf{h}) \cdot \mathcal{D}(\mathbf{h}) \right\} h \ \mathrm{d}h \tag{4.60}$$

where

$$\mathcal{D}(\mathbf{h}) = \int_{\{\boldsymbol{\varphi}:\boldsymbol{\varphi},(\boldsymbol{\varphi}+\mathbf{h})\in\boldsymbol{\Theta}\}} [p(\boldsymbol{\varphi}) + p(\boldsymbol{\varphi}+\mathbf{h})] \mathrm{d}\boldsymbol{\varphi} \tag{4.61}$$

Now consider a support set $\boldsymbol{\Theta}$ for $p(\boldsymbol{\theta})$, of the form

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M) \ : \ \theta_i \in [-B, B] \ \ \forall i\} \tag{4.62}$$

We note since we assume the p.d.f. $f_0(\cdot)$ has finite support, $P_{\min}(\mathbf{h})$ becomes zero when the elements of $\mathbf{h}$ grow very large. Thus, the maximum values of the elements of $\mathbf{h}$ that must be considered in the right hand side of (4.60) are all finite (for a more rigorous argument, one can show that there exists a positive scalar h_0 such that $P_{\min}(\mathbf{h}) = 0$ for any $\mathbf{h}$ satisfying $\mathbf{c}^{\mathrm{T}}\mathbf{h} = h$ when $h > h_0$). For any bounded value of $\mathbf{h}$, it is easy to see that under the limiting prior distribution where $B \to \infty$, we obtain

$$\lim_{B\to\infty} \mathcal{D}(\mathbf{h}) = \lim_{B\to\infty} \frac{2}{S} \int_{\{\boldsymbol{\varphi}:\boldsymbol{\varphi},(\boldsymbol{\varphi}+\mathbf{h})\in\boldsymbol{\Theta}\}} \mathrm{d}\boldsymbol{\varphi} = 2 \tag{4.63}$$

Hence, in this limiting case, we obtain $\mathcal{D}(\mathbf{h}) = 2$, and (4.60) reduces to

$$\mathcal{M}(g) \geq \int_0^\infty \mathcal{V}\left\{ \max_{\mathbf{h}:\mathbf{c}^{\mathrm{T}}\mathbf{h}=h} P_{\min}(\mathbf{h}) \right\} h \ \mathrm{d}h \tag{4.64}$$

This concludes the proof.

Chapter 5

Minimax Estimators for Phase offset estimation in IEEE 1588

5.1 Introduction

In the IEEE 1588 precision time protocol (PTP), a master and a slave node exchange a series of packets to achieve phase synchronization. Packets traveling between the master and the slave encounter several intermediate network nodes such as switches or routers, accumulating random queuing delays at each node. The problem of finding the slave's phase offset from the timestamps of the exchanged packets, while combating the random queuing delays, is referred to as *phase offset estimation* (POE). The PTP standard and related literature prescribe the use of simple estimators such as the sample mean, minimum and maximum filters for POE. Several recent papers [17]–[50] have studied methods to improve the performance of these filters, especially in the presence of large queuing delays due to high network loads. However, it is not well understood as to how close these POE schemes come to achieving the best

possible estimation performance, measured in terms of the mean squared estimation error.

In this chapter, we derive optimum estimators for the problem of POE, which, to our knowledge, have not been described previously in literature. To this end, in Section 5.2 we begin by modeling POE as a non-Bayesian estimation problem. Specifically, we treat the phase offset as an unknown deterministic parameter to be estimated from timestamps that are also affected by the fixed delays along the forward and reverse network paths. We then consider three observation models, with varying degrees of information available about the fixed delays. Under the *known fixed delays model* (K-model), we assume complete knowledge of both the fixed delays, while under the *standard model* (S-model), we assume that only the *delay asymmetry* is known. Further, under the *multiblock model* (M-model), we assume known delay asymmetry, as well as the availability of additional past observations which contain the same fixed delays but different phase offsets. Under all three observation models, we show that POE falls under a general class of estimation problems known as *vector location parameter problems*. In statistical estimation theory, the *Pitman estimator* [51][52] is well known to be *minimax* optimum for location parameter problems, in the sense that it minimizes the maximum mean squared error (*maximum MSE*) over all values of the unknown parameters. While the original Pitman estimator was derived only for scalar location parameter problems, in Section 5.3 we rederive it in the more general context of vector location parameter problems. Other properties of the Pitman estimator, related to the estimation of linear combinations of parameters, are also rederived in this new context.

Our motivation in considering multiple observation models is to provide insight into the dependence between estimation performance and the amount of prior information available about the fixed delays. Specifically, we show that the minimax MSE (MSE of the minimax optimum estimator) under the M-model is guaranteed to be less than the minimax MSE

under S-model, and greater than the minimax MSE under the K-model, independent of the amount of past information available under the M-model. Hence, while the K-model is not as practical as the S-model and M-model, it helps us establish a useful limit on the performance gains that can be achieved under the M-model relative to the S-model.

In Section 5.4, we simplify the general minimax estimator for the problem of POE under each observation model. In Section 5.5, using the properties of the minimax estimator derived in Section 5.3, we show that under typical network assumptions, the MSE of the minimax estimator grows at least linearly with the number of intermediate nodes between the master and the slave. Our simulations in section 5.6 compare the performance of the new minimax estimates against conventional estimators under several network conditions. Results indicate that there are several network scenarios where conventional estimation schemes fall significantly short of achieving the maximum possible synchronization accuracy. Further, in asymmetric network traffic scenarios, we show that significant performance gains become available if we exploit information about fixed delays from past observations.

The results in this chapter extend on our the result presented in Chapter 4, where lower bounds on the maximum MSE of POE schemes under the second observation model were derived. In this chapter, we address more observational models, provide the tightest lower bounds on the maximum MSE of POE schemes under each model, and also specify the estimators that achieve these lower bounds.

5.2 System Model

Consider a scenario where the slave clock has a phase offset δ and zero frequency offset with respect to its master. To help the slave determine δ, the IEEE 1588 PTP protocol allows a

two-way message exchange between the master and slave. The steps involved in a two-way message exchange are as follows:

1. The message exchange is initiated by the master, when it sends a *SYNC* packet to the slave at a time t_1. A *FOLLOW_UP* message is used to communicate the value of t_1 to the slave.

2. The time of reception of the *SYNC* packet is recorded as $t_2 = t_1 + d_1 + \delta$, where d_1 denotes the end-to-end (ETE) network delay between the master and the slave.

3. The slave responds to the master with a *DELAY_REQ* packet, and records its time of transmission as t_3.

4. The time of arrival of the *DELAY_RESP* packet is recorded by the master as $t_4 = t_3 - \delta + d_2$, where d_2 denotes the ETE network delay between the slave and the master. The value of t_4 is sent to the slave via a *DELAY_REQ* packet.

In order to estimate δ, it is clearly sufficient for the slave to only retain the pair of timestamp differences

$$y_1 = t_2 - t_1 = d_1 + \delta \tag{5.1}$$

$$y_2 = t_4 - t_3 = d_2 - \delta \tag{5.2}$$

Here d_1 and d_2 denote the end-to-end (ETE) network delays in the master-slave and slave-master directions, respectively. Assume for simplicity that a common network path is taken by all packets traveling between the master and the slave and vice-versa. Then each ETE delay receives contributions from three factors:

(a) Constant propagation delays along network links between the master and the slave (or

vice-versa).

(b) Constant processing delays at intermediate nodes (such as switches or routers) along each network path.

(c) Random queuing delays at intermediate nodes along each network path.

Hence each ETE delay can be modeled as

$$d_1 = d_1^{\min} + w_1, \qquad d_2 = d_2^{\min} + w_2 \tag{5.3}$$

Here $d_1^{\min}$ and $d_2^{\min}$ denote fixed delays corresponding to the sum of the constant propagation and processing delays, while w_1 and w_2 model the random queuing delays.

Assuming the values of δ, $d_1^{\min}$ and $d_2^{\min}$ remain constant over the duration of P two-way message exchanges, we can collect multiple observation pairs (y_1, y_2) to help estimate δ. We denote these observations as

$$y_{i,1}^* = d_1^{\min} + \delta + w_{i,1} \ , \quad y_{i,2}^* = d_2^{\min} - \delta + w_{i,2} \tag{5.4}$$

for $i = 1, \cdots, P$. The accuracy with which we can estimate δ from the observations in (5.4) depends on the amount of knowledge we have about $d_1^{\min}$ and $d_2^{\min}$. We now consider three observation models, differentiated based on the amount of prior information available about $d_1^{\min}$ and $d_2^{\min}$:

1. *Known fixed delay model (K-model)*: Here we assume that $d_1^{\min}$ and $d_2^{\min}$ are fully known at the slave. Hence, setting $y_{i,k} = y_{i,k}^* - d_k^{\min}$, we obtain the compensated observations

$$y_{i,1} = \delta + w_{i,1} \ , \quad y_{i,2} = -\delta + w_{i,2} \tag{5.5}$$

for $i = 1, \cdots, P$. These observations can be collected to obtain the vector observation

model

$$\mathbf{y} = \delta\mathbf{e} + \mathbf{w} \tag{5.6}$$

where

$$\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}}\ \mathbf{y}_2^{\mathrm{T}}\right]^{\mathrm{T}},\ \mathbf{y}_k = [y_{1,k}\ \cdots\ y_{P,k}]^{\mathrm{T}} \tag{5.7}$$

$$\mathbf{w} = \left[\mathbf{w}_1^{\mathrm{T}}\ \mathbf{w}_2^{\mathrm{T}}\right]^{\mathrm{T}},\ \mathbf{w}_k = [w_{1,k}\ \cdots\ w_{P,k}]^{\mathrm{T}} \tag{5.8}$$

$$\mathbf{e} = [\mathbf{1}_P \quad (-\mathbf{1}_P)]^{\mathrm{T}} \tag{5.9}$$

and $\mathbf{1}_N$ is a $N \times 1$ vector with all elements equal to 1.

2. *Standard model (S-model)*: Here we assume that only the difference between $d_1^{\min}$ and $d_2^{\min}$, referred to as the *delay asymmetry*, is known to the slave. By compensating the observations as

$$y_{i,1} = y_{i,1}^*,\quad y_{i,2} = y_{i,2}^* - d_2^{\min} + d_1^{\min} \tag{5.10}$$

we obtain

$$y_{i,1} = d + \delta + w_{i,1}\ ,\quad y_{i,2} = d - \delta + w_{i,2} \tag{5.11}$$

for $i = 1, \cdots, P$, where $d = d_1^{\min}$. These observations can be denoted vectorially as

$$\mathbf{y} = d\mathbf{1}_{2P} + \delta\mathbf{e} + \mathbf{w} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w} \tag{5.12}$$

where $\mathbf{y}$ and $\mathbf{w}$ are as defined in (5.7) and (5.8), and

$$\boldsymbol{\theta} = [\theta_1 \ \theta_2]^{\mathrm{T}} = [d+\delta \ \ d-\delta]^{\mathrm{T}}, \tag{5.13}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_P & \mathbf{0}_P \\ \mathbf{0}_P & \mathbf{1}_P \end{bmatrix}, \tag{5.14}$$

with $\mathbf{1}_Q$, $\mathbf{0}_Q$ representing $Q \times 1$ vectors of ones and zeros, respectively.

Note that this model also covers the case of symmetric path delays, where $d_1^{\min} = d_2^{\min}$, and hence the delay asymmetry is zero. We further note that other cases where the relationship between the fixed delays is known, such as the case where the ratio $d_1^{\min}/d_2^{\min}$ is known, can also be handled using a model similar to (5.12). For brevity, only the case of known delay asymmetry is considered here.

3. *Multiblock model (M-model)*: Here we assume, as in the standard model, that the delay asymmetry is known to the slave. Suppose we refer to a set of P observation pairs as a *block*. In this model, we further assume that in addition to the current block, we have observation pairs from B previous blocks available to us. The phase offset δ is modeled as being constant for all observation pairs within each block, but varying between different blocks. The fixed delay d is modeled as constant across all $B+1$ blocks. This model is representative of scenarios where changes in the fixed delay occur over longer time scales than changes in phase offset. We denote observation pairs in past blocks using the notation

$$y_{i,j,1} = d + \delta_j + w_{i,j,1} \ , \quad y_{i,j,2} = d - \delta_j + w_{i,j,2} \tag{5.15}$$

and observation pairs in the current block as

$$y_{i,1} = d + \delta + w_{i,1} \ , \quad y_{i,2} = d - \delta + w_{i,2} \tag{5.16}$$

for $i = 1, \cdots, P$ and $j = 1, \cdots, B$. We thus obtain the vector observation model

$$\mathbf{y} = \mathbf{G}\boldsymbol{\theta} + \mathbf{w} \tag{5.17}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{T}} & \mathbf{y}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{5.18}$$

$$\mathbf{y}_k = [y_{1,k} \ \cdots \ y_{P,k} \ y_{1,1,k} \ y_{1,2,k} \ \cdots \ y_{B,P,k}]^{\mathrm{T}} \tag{5.19}$$

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} & \mathbf{w}_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{5.20}$$

$$\mathbf{w}_k = [w_{1,k} \ \cdots \ w_{P,k} \ w_{1,1,k} \ w_{1,2,k} \ \cdots \ w_{B,P,k}]^{\mathrm{T}} \tag{5.21}$$

$$\boldsymbol{\theta} = [d \ \delta \ \ \delta_1 \ \ \cdots \ \ \delta_B]^{\mathrm{T}}, \tag{5.22}$$

$$\mathbf{G} = [\mathbf{1}_{2BP} \quad \mathbf{Z} \otimes \mathbf{1}_P] \, , \quad \mathbf{Z} = [\mathbf{I}_B \quad (-\mathbf{I}_B)]^{\mathrm{T}} \tag{5.23}$$

and $\mathbf{I}_B$, $\otimes$ denote the identity matrix of size B and the Kronecker product operator, respectively.

It is easy to see that the K-model can be difficult to use in practice, since it requires that the fixed delays $d_1^{\min}$ and $d_2^{\min}$ both be known to the slave. One way to determine $d_1^{\min}$ and $d_2^{\min}$ would be via an initial calibration step, where a perfect time source is temporarily attached to the slave, and network transit times are measured in the absence of background traffic. In typical situations where such an expensive calibration step is not feasible, but it is known that the master-slave and slave-master path have identical fixed delays, the more practical S-model can be used. Further, whenever it is known that the fixed delays

remain constant over longer time intervals than the phase offsets, the M-model can be used instead of the S-model, since it contains additional observations that could be used to improve estimation performance. We still consider the K-model in this chapter since we later shows that it provides useful bounds on the estimation performance achievable under the M-model.

Given either of the observation models, the problem of POE is to estimate δ from the observation vector $\mathbf{y}$. Here we further make the following assumptions:

(i) All the queuing delays are strictly positive random variables that are mutually independent.

(ii) All forward queuing delays share a common pdf $f_1(w)$. Similarly the reverse queuing delays share a common pdf $f_2(w)$.

(iii) The maximum possible value for a forward or reverse queuing delay is finite.

(iv) All the unknown fixed delays and phase offsets are deterministic parameters, i.e. no probability distributions for these parameters are known a priori.

Note that in practice, it is often reasonable to assume that background traffic patterns remain constant over several minutes. Hence, the assumption that all queing delays share a common pdf is fairly realistic.

5.3 Minimax Estimators for Location Parameter Problems

We now consider a general class of estimation problems, where the effect of the unknown parameters is to shift the location of the pdf of the observations without modifying the underlying shape of the pdf. The POE problems under all three observation models considered in Section 5.2 belong to this general class of problems. The general results derived here shall

be applied to the POE models in Section 5.4. The proof of all the lemmas and theorems stated in this section are provided in the appendix.

We first define the general class of problems we are interested in studying.

Definition 2 (Vector Location Parameter Problem)**.** *Suppose we want to estimate a linear combination* $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ *of the unknown parameters contained in* $\boldsymbol{\theta} \in \mathbb{R}^M$ *(where* $\mathbf{c} \in \mathbb{R}^M$ *is a constant vector), based on observations* $\mathbf{x} \in \mathbb{R}^N$*. If the observations have a pdf of the form*

$$f(\mathbf{x}|\boldsymbol{\theta}) = f_0(\mathbf{x} - \mathbf{G}\boldsymbol{\theta}) \tag{5.24}$$

for some $N \times M$ *matrix* $\mathbf{G}$ *and function* $f_0(\cdot)$*, then we shall refer to such an estimation problem as a vector location parameter problem.*

All the definitions and theorems in the remainder of this section apply specifically to this vector location parameter problem. The results we derive further require that the function $f_0(\mathbf{x})$ be non-zero over a bounded, positive range of values of its arguments, as defined below.

Definition 3 (Finite Support)**.** *We say that* $f_0(\mathbf{x})$ *in (5.24) has finite support if there exists a finite* $L > 0$ *such that* $f_0(\mathbf{x}) = 0$ *whenever all the elements of the vector* $\mathbf{x}$ *lie outside the interval* $[0, L]$*.*

It is typical in statistical literature to characterize the performance of an estimator via the mean squared error (MSE) metric. There are three ways to define the MSE metric:

1. The *conditional MSE*

$$\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) = \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}]^2 f(\mathbf{x}|\boldsymbol{\theta}) \mathrm{d}\mathbf{x} \tag{5.25}$$

2. The *maximum MSE*

$$\mathcal{M}(g(\mathbf{x})) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^M} \mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) \tag{5.26}$$

3. The *average MSE*

$$\mathcal{B}(g(\mathbf{x}), p(\boldsymbol{\theta})) = \int_{\mathbb{R}^M} \mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{5.27}$$

where $p(\boldsymbol{\theta})$ is a prior distribution defined over $\boldsymbol{\theta} \in \mathbb{R}^M$.

In this section, we consider the problem of finding estimators that are optimum in terms of minimizing the maximum MSE, and refer to such estimators as *minimax* estimators. The definitions of the conditional and average MSEs shall be used in the proofs of the optimality of the minimax estimator.

We now consider a class of estimators known as shift invariant estimators, defined as follows.

Definition 4 (Shift Invariant Estimator)**.** *We say that an estimator $g(\mathbf{x})$ of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ is shift invariant if for the same matrix $\mathbf{G}$ used in (5.24),*

$$g(\mathbf{x} + \mathbf{G}\mathbf{h}) = g(\mathbf{x}) + \mathbf{c}^{\mathrm{T}}\mathbf{h} \tag{5.28}$$

for all $\mathbf{h} \in \mathbb{R}^M$.

While the conditional, maximum and average MSEs can be different for a estimator, for a shift invariant estimator they are always equal, as stated in the following lemma.

Lemma 3. *Any shift invariant estimator $g(\mathbf{x})$ of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ has a conditional MSE that is constant with respect to $\boldsymbol{\theta}$, and satisfies*

$$\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) = \mathcal{M}(g(\mathbf{x})) = \mathcal{B}(g(\mathbf{x}), p(\boldsymbol{\theta})) \tag{5.29}$$

for any choice of prior distribution $p(\boldsymbol{\theta})$.

We now give the expression for the minimax estimator and prove its optimality using Definition 4 and Lemma 3. We note that the following result is an extension of the Pitman estimator [51] to vector location parameter problems.

Theorem 5 (Minimax estimator)**.** *If* $f_0(\mathbf{x})$ *has finite support, then the estimator*

$$g^*(\mathbf{x}) = \frac{\int_{\mathbb{R}^M} [\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\theta}}] f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ d\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ d\hat{\boldsymbol{\theta}}} \tag{5.30}$$

satisfies the following properties:

(i) $g^*(\mathbf{x})$ *is shift invariant.*

(ii) $g^*(\mathbf{x})$ *is a minimax estimate of* $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$.

(iii) Among all estimators of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ *that are shift invariant,* $g^*(\mathbf{x})$ *achieves the minimum conditional MSE* $\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta})$.

(iv) $g^*(\mathbf{x})$ *is unbiased, i.e.* $E\left\{ \left[g^*(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta} \right] \mid \boldsymbol{\theta} \right\} = 0$.

An interesting property of the minimax estimator is that for a given set of observations, the minimax estimate of a linear combination of parameters is identical to the same linear combination of the minimax estimates of each of the parameters. Formally, this can be stated as follows.

Lemma 4. *Let* $\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_M]^T$, *and let* $g_i^*(\mathbf{x})$ *represent the minimax estimate of* θ_i. *If* $\mathbf{c} = [c_1 \ \cdots \ c_M]^{\mathrm{T}}$, *then the minimax estimate* $g^*(\mathbf{x})$ *of* $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ *satisfies*

$$g^*(\mathbf{x}) = \sum_{i=1}^{M} c_i g_i^*(\mathbf{x})$$

This property will allow us to simplify the form of the minimax estimator under the S-model in Section 5.4.

Another interesting property of the minimax estimator emerges when we consider multiple minimax estimates, each based on a different observation vector. Here we can show that the sum of the MSEs of the individual minimax estimates will always be less than the MSE of the minimax estimate based on sum of all the observation vectors. We note that this result is an extension of a similar result presented in [52] for scalar location parameter problems.

Theorem 6. *Let $\mathbf{x}_1, \cdots, \mathbf{x}_K$ be N-dimensional random vectors with pdfs of the form*

$$f(\mathbf{x}_k|\boldsymbol{\theta}_k) = f_k(\mathbf{x}_k - \mathbf{G}_k\boldsymbol{\theta}_k) \tag{5.31}$$

where $f_k(\cdot)$ has finite support for $k = 1, \cdots, K$. Assume that $\mathbf{x}_1, \cdots, \mathbf{x}_K$ are all mutually independent conditioned on the unknown parameters, i.e. the joint pdf $f(\mathbf{x}_{k_1}, \mathbf{x}_{k_2}|\boldsymbol{\theta}_{k_1}, \boldsymbol{\theta}_{k_2})$ satisfies

$$f(\mathbf{x}_{k_1}, \mathbf{x}_{k_2}|\boldsymbol{\theta}_{k_1}, \boldsymbol{\theta}_{k_2}) = f(\mathbf{x}_{k_1}|\boldsymbol{\theta}_{k_1})f(\mathbf{x}_{k_2}|\boldsymbol{\theta}_{k_2}) \tag{5.32}$$

for all values of k_1 and k_2. Let $h_k^*(\mathbf{x}_k)$ denote the minimax estimate of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_k$. Further, let $\mathbf{x} = \sum_{k=1}^K \mathbf{x}_k$, $\boldsymbol{\theta} = \sum_{k=1}^K \boldsymbol{\theta}_k$, and let $g^*(\mathbf{x})$ denote the minimax estimate of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ from $\mathbf{x}$. Then $g^*(\mathbf{x})$ satisfies

$$\mathcal{M}(g^*(\mathbf{x})) \geq \sum_{k=1}^K \mathcal{M}(h_k^*(\mathbf{x}_k)) \tag{5.33}$$

This property will be useful in proving certain properties of the minimax estimator for POE in Section 5.5.

5.4 Simplification of Minimax Estimator for the POE problem

We now use the results in Section 5.3 to obtain minimax optimum estimators under the three POE observation models discussed in Section 5.2, and simplify the resulting expressions.

1) *Known fixed delay model*: As stated in (5.6), the pdf of the observation vector $\mathbf{y}$ has the form

$$f(\mathbf{y}|\delta) = f_{\mathbf{W}}(\mathbf{y} - \delta\mathbf{e}) \tag{5.34}$$

where

$$f_{\mathbf{W}}(\mathbf{w}) = \prod_{i=1}^{P} f_1(w_{i,1}) f_2(w_{i,2}) \tag{5.35}$$

Hence, according to Definition 2, this is a vector location parameter problem. Thus, using Theorem 5, we obtain the minimax estimator of δ as

$$\hat{\delta}(\mathbf{y}) = \frac{\int_{\mathbb{R}} \delta f_{\mathbf{W}}(\mathbf{y} - \delta\mathbf{e}) \mathrm{d}\delta}{\int_{\mathbb{R}} f_{\mathbf{W}}(\mathbf{y} - \delta\mathbf{e}) \mathrm{d}\delta} \tag{5.36}$$

2) *Standard Model*: As stated in (5.12), here the pdf of the observation vector $\mathbf{y}$ has the form

$$f(\mathbf{y}|\boldsymbol{\theta}) = f_{\mathbf{W}}(\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \tag{5.37}$$

$$= f_{\mathbf{W},1}(\mathbf{y}_1 - \theta_1 \mathbf{1}_P) f_{\mathbf{W},2}(\mathbf{y}_2 - \theta_2 \mathbf{1}_P) \tag{5.38}$$

where

$$f_{\mathbf{W}}(\mathbf{w}) = \prod_{i=1}^{P} f_1(w_{i,1}) f_2(w_{i,2}) \ , \tag{5.39}$$

$$f_{\mathbf{W},k}(\mathbf{w}_k) = \prod_{i=1}^{P} f_k(w_{i,k}) \quad \text{for} \quad k = 1, 2 \tag{5.40}$$

Hence, according to Definition 2, this is a vector location parameter problem. Our goal

is to estimate $\delta = \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ (where $\mathbf{c} = [0.5 \ -0.5]^{\mathrm{T}}$) from the observation vector $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w}$. Hence, using Theorem 5, we obtain the minimax estimate

$$\hat{\delta}(\mathbf{y}) = \frac{\int_{\mathbb{R}^2} [\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}] f(\mathbf{y}|\boldsymbol{\theta}) \ \mathrm{d}\boldsymbol{\theta}}{\int_{\mathbb{R}^2} f(\mathbf{y}|\boldsymbol{\theta}) \ \mathrm{d}\boldsymbol{\theta}} \tag{5.41}$$

Using Lemma 4, we can further simplify the estimator as

$$\hat{\delta}(\mathbf{y}) = \frac{1}{2}\left[\frac{\int_{\mathbb{R}} \theta_1 f_{\mathbf{W},1}(\mathbf{y}_1 - \theta_1 \mathbf{1}_P) \ \mathrm{d}\theta_1}{\int_{\mathbb{R}} f_{\mathbf{W},1}(\mathbf{y}_1 - \theta_1 \mathbf{1}_P) \ \mathrm{d}\theta_1} - \frac{\int_{\mathbb{R}} \theta_2 f_{\mathbf{W},2}(\mathbf{y}_2 - \theta_2 \mathbf{1}_P) \ \mathrm{d}\theta_2}{\int_{\mathbb{R}} f_{\mathbf{W},2}(\mathbf{y}_2 - \theta_2 \mathbf{1}_P) \ \mathrm{d}\theta_2} \right] \tag{5.42}$$

3) *Multiblock Model*: As stated in (5.17), here the pdf of the observation vector $\mathbf{y}$ has the form

$$f(\mathbf{y}|\boldsymbol{\theta}) = f_{\mathbf{W}}(\mathbf{y} - \mathbf{G}\boldsymbol{\theta}) \tag{5.43}$$

where

$$f_{\mathbf{W}}(\mathbf{w}) = \prod_{i=1}^{B} \prod_{j=1}^{P} \prod_{k=1}^{2} f_k(w_{i,j,k}) \tag{5.44}$$

Hence, according to Definition 2, this is also a vector location parameter problem. Our goal is to estimate $\delta = \hat{\mathbf{c}}^{\mathrm{T}}\boldsymbol{\theta}$ from $\mathbf{y}$, where $\hat{\mathbf{c}} = [0 \ 1 \ \underbrace{0 \ \cdots \ 0}_{B-1 \text{ zeros}}]^{\mathrm{T}}$. Using Theorem 5, we obtain the minimax estimate

$$\hat{\delta}(\mathbf{y}) = \frac{\int_{\mathbb{R}} \delta \Gamma(\delta, \mathbf{y}) \mathrm{d}\delta}{\int_{\mathbb{R}} \Gamma(\delta, \mathbf{y}) \mathrm{d}\delta} \tag{5.45}$$

where

$$\Gamma(\delta, \mathbf{y}) = \int_{\mathbb{R}} \left[\prod_{i=1}^{P} \prod_{k=1}^{2} f_k(y_{i,k} - d + (-1)^k \delta) \right] \cdot \Omega(d, \mathbf{y}) \, \mathrm{d}(d) \tag{5.46}$$

$$\Omega_j(d, \mathbf{y}) = \prod_{j=1}^{B} \left[\int_{\mathbb{R}} \prod_{i=1}^{P} \prod_{k=1}^{2} f_k(y_{i,j,k} - d + (-1)^k \delta_j) \mathrm{d}\delta_j \right] \tag{5.47}$$

In scenarios where analytical expressions for the queuing delay pdfs $f_1(w)$ and $f_2(w)$ are known, it might be possible to further simplify the integrals in (5.36), (5.42) and (5.45)-(5.47). In the more general case of arbitrary pdfs $f_1(w)$ and $f_2(w)$, these integrals can be computed by approximating them with Riemann summations. In such cases, the computational complexity associated with the minimax estimators will depend on the number of bins used in the Riemann summations. Typically, this computational complexity is significantly higher than that of conventional estimators such as the sample minimum, mean, median or maximum estimators.

Due to the nature of the POE observation models, some comments regarding the *minimax MSE* (the MSE of the minimax optimum estimator) can be made directly, without requiring numerical evaluations. Firstly, the minimax MSE under the K-model is guaranteed to be lower than that under the S-model or M-model, since the nuisance parameter d is absent from the K-model. Further, the minimax MSE under the M-model is guaranteed to be lower than that under the S-model, since the M-model has additional information from past blocks available to it. This past information can be used to reduce the uncertainty associated with the nuisance parameter d, and hence improve the estimate of δ.

5.5 Minimax MSE under IID single-node queuing delays

The performance of the minimax estimators described in Section 5.4 depends on the nature of the network queuing delays, which in turn depends on the number of nodes present between the master and the slave. Theorem 6 can be used to obtain a simple relationship between the minimax MSE and the number of intermediate nodes, under certain network conditions. We state this relationship in the form of the following corollary to Theorem 6, with the proof provided in the appendix.

Corollary 1. *Consider a network consisting of a master and a slave separated by N nodes. Let $\rho(N)$ represent the minimax MSE associated with POE under the S-model in this scenario, for a fixed number of two-way message exchanges. Let the single-node queuing delay refer to the queuing delay experienced by packets at any single node*[1]*. Assume that the single-node queuing delays across all nodes in the forward direction are independent and identically distributed (i.i.d.). Assume that the same is true in the reverse direction as well. Then $\rho(N)$ satisfies*

$$\rho(KL) \geq K\rho(L) \tag{5.48}$$

where K and L are any two positive integers.

For $L = 1$, the relation in (5.48) reduces to $\rho(K) \geq K\rho(1)$, which essentially implies that *in networks with i.i.d. single-node queuing delays at all intermediate network nodes, the minimax MSE grows at least linearly with the number of nodes.* This interpretation can be especially useful for network designers, since it provides a computationally simple upper limit on the number of nodes that can be allowed between the master and the slave for a given

[1]Measurements of the single node queuing delay in the forward or reverse direction would corresponding to the proper entries of the vector $\mathbf{w}$ in (5.12) for the case where only N=1 node is involved.

synchronization accuracy requirement. A typical example where independent, identically distributed single-node queuing delay distributions can be assumed is a network in which only *cross traffic flows* (defined in Section 5.6) are present. Note that a relationship similar to (5.48) can also be derived under the K-model and the M-model. For brevity, only the S-model is considered in this section.

5.6 Simulation Results

We now compare the performance of conventional POE schemes against the newly derived minimax estimators. To this end, we consider a few network scenarios motivated by the ITU-T recommendation G.8261 [45]. The metric we use to quantify estimator performance is the maximum MSE. For brevity, we refer to the maximum MSE as simply the MSE throughout this section.

We consider four commonly used conventional POE schemes, namely the sample minimum, maximum, mean and median filtering schemes. Given the observation vector $\mathbf{y}$ of either the K-model or the S-model, these schemes use an estimator of the form

$$\hat{\delta} = \frac{\xi(\mathbf{y}_1) - \xi(\mathbf{y}_2)}{2} \tag{5.49}$$

where $\xi(\mathbf{x})$ denotes either the minimum, maximum, mean or median of the elements of the vector $\mathbf{x}$. Under the M-model, these estimators behave exactly as under the S-model, discarding information from past blocks since they have no means of utilizing it. It is easy to show that these estimators are shift invariant under all three observation models. They

also have an identical value for the MSE across all three models, given as

$$\mathcal{M}(\hat{\delta}) = \mathrm{E}\left\{ \hat{\delta}^2 \;\middle|\; \boldsymbol{\theta} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \right\} = \sigma^2 + \mu^2 \tag{5.50}$$

where

$$\sigma^2 = \mathrm{var}\left\{ \hat{\delta}^2 \;\middle|\; \boldsymbol{\theta} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \right\} \tag{5.51}$$

$$= \frac{1}{4}\big[\mathrm{var}\xi(\mathbf{w}_1) + \mathrm{var}\xi(\mathbf{w}_2)\big] \tag{5.52}$$

represents the estimator variance, while

$$\mu = \frac{1}{2}\Big[\mathrm{E}\left[\xi(\mathbf{w}_1)\right] - \mathrm{E}\left[\xi(\mathbf{w}_2)\right]\Big] \tag{5.53}$$

represents the estimator bias. Note that

$$E[g(\mathbf{w}_i)] = \int \xi(\mathbf{w}_i) f_{\mathbf{W}_i}(\mathbf{w}_i) \mathrm{d}\mathbf{w}_i \ , \tag{5.54}$$

$$\mathrm{var}\xi(\mathbf{w}_i) = \int_{\mathbf{w}_i} \left\{\xi(\mathbf{w}_i) - E[\xi(\mathbf{w}_i)]\right\}^2 f_{\mathbf{W}_i}(\mathbf{w}_i) \mathrm{d}\mathbf{w}_i \ , \tag{5.55}$$

$$f_{\mathbf{W}_i}(\mathbf{w}_i) = \prod_{j=1}^{P} f_i(w_{i,j}) \tag{5.56}$$

It is easy to see from (5.53) that when the forward and reverse queuing delay distributions $f_1(w)$ and $f_2(w)$ are not identical, μ can be non-zero, and hence have a significant contribution in the MSE expression in (5.50). This can be avoided by subtracting out the bias, to obtain the unbiased estimate

$$\tilde{\delta} = \hat{\delta} - \mu \tag{5.57}$$

Hence, in our results, we measure the performance of conventional estimators as their MSE

Traf. Model	Packet Sizes (Bytes)	% of Load
TM1	$\{64, 576, 1518\}$	$\{80\%, 5\%, 15\%\}$
TM2	$\{64, 576, 1518\}$	$\{30\%, 10\%, 60\%\}$

Table 5.1: Models for composition of background traffic packets

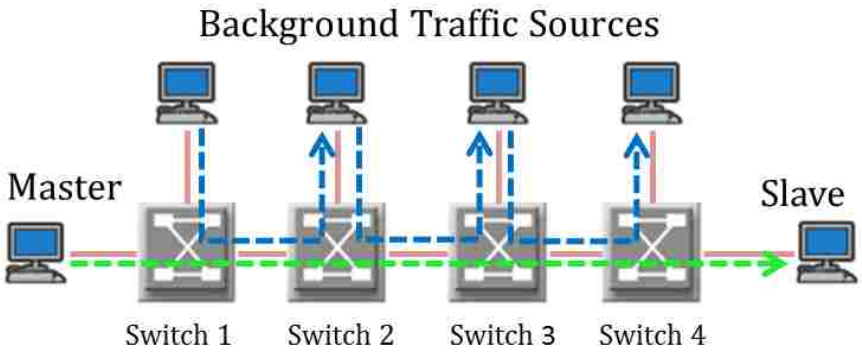
after their bias has been compensated.

In order to obtain the queuing delay distributions, we consider a Gigabit ethernet network consisting of a cascade of 20 switches between the master and slave nodes. Each switch is assumed to be a store-and-forward switch, which implements strict priority queuing. We consider two types of background traffic flows in this network:

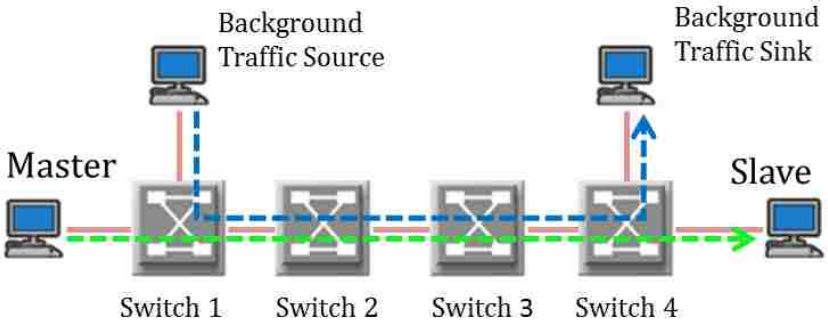
1. *Cross traffic flows*: In such traffic flows [45][53], fresh background traffic packets are injected at each node along the master-slave path, and these packets exit the master-slave path at the subsequent node (see 4-switch example in Fig. 5.1a). The arrival times and sizes of the packets injected at each switch are assumed to be statistically independent of that of packets injected at other switches.

2. *Mixed traffic flows*: Here a mixture of cross traffic flows and *inline traffic flows* are present in the network. Inline traffic flows [22] are characterized by packets that are injected only at the first switch along the master slave path, and that travel along the same path as synchronization traffic through the entire cascade of switches (see 4-switch example in Fig. 5.1b).

With regard to the distribution of packet sizes in background traffic, we consider Traffic Models 1 (TM1) and 2 (TM2) from the ITU-T recommendation G.8261 [45] for cross traffic flows, as specified in Table 5.1. For inline traffic flows, we consider a third traffic model where packet sizes are uniformly distributed between 64 and 1500 bytes [7]. We assume



(a) Cross traffic flows



(b) Inline traffic flows

Figure 5.1: Examples of four switch networks with cross and inline traffic flows. Red lines indicate network links, blue lines indicate the direction of background traffic flows, and green line represents the direction of synchronization traffic flows.

that the interarrival times between packets in all background traffic flows follow exponential distributions. We refer to the percentage of the link capacity consumed by background traffic as the *load*. In order to achieve a particular load, we accordingly set the rate parameter of each exponential distribution. The queuing delay distributions under a number of network scenarios are plotted in Fig. 5.3. These distributions were obtained empirically using low-level queue simulations. Without loss of generality, we assume that fixed delay components of the ETE delays equal zero, hence the support of the queuing delay distributions always begins at zero in the plots.

The MSE of various estimators under different observation models and network conditions are plotted versus the number of observation pairs/samples P in Figs. 5.4 - 5.7. In order to compute the minimax estimates, the integrals in (5.36), (5.42) and (5.45) were replaced with Riemann sums. The spacing between adjacent Riemann summation bins was set to 0.001 μs, to ensure that the additional error introduced due to the Riemann sum approximation is small relative to the MSE being computed. Further, to facilitate comparisons against the LTE synchronization requirement of 1.25 μs of synchronization accuracy, the estimation error standard deviation required so that the absolute estimation error lies under 1.25 μs with a 5-sigma level of certainty is also plotted over the curves. Here the 5-sigma level of certainty implies that on average, only about 6 out of 10^6 estimates will have absolute estimation error that exceeds 1.25 μs. Some key observations we can make from the results are:

1. *Performance under symmetric cross traffic (Fig. 5.4)*: Here, the gap between the K-model and S-model minimax estimators is negligible under all four loads (20%, 40%, 60%, 80%) considered. Hence, under these network scenarios, there is little performance to be gained from the additional knowledge about fixed delays that the K-model provides over the S-model. Further, while the sample minimum estimator performs near-optimally at 20%

load, at higher loads none of the conventional estimation schemes come close to achieving minimax optimal performance. In fact, at 80% load, the minimax estimators achieve the LTE synchronization requirement using only about 200 samples, while about 800 samples are required by the best conventional estimator.

2. *Performance under symmetric mixed traffic (Fig. 5.5)*: Here, there is a fair gap between the K-model and S-model minimax estimators under the lower load scenario of Fig. 5.5a, which disappears under the higher load scenario of Fig. 5.5b. Further, the S-model minimax estimator requires about 50% fewer samples than the best conventional estimator, to achieve the LTE synchronization requirement under the low load scenario. Interestingly, the sample mean filter performs near-optimally under the high load scenario. This indicates that the performance gap between the best conventional estimator and the minimax estimator may need to be studied on a per-case basis, and predicting general trends might be difficult.

3. *Performance under asymmetric traffic (Figs. 5.6 and 5.7)*: Here there is a significant gap between the K-model and S-model minimax estimators, with the K-model minimax estimator requiring about 90% and 22% fewer samples than the S-model estimator in Fig. 5.6 and Fig. 5.7, respectively, in order to meet the LTE synchronization requirement threshold. This is expected in cases where the queuing delay distribution in one network direction has significantly lower spread than in the other direction. In such cases, the MSE of conventional estimators, given by (5.52), is dominated by either the first or second term in (5.52) if one these variances is much larger than the other. On the other hand, the K-model estimator can utilize knowledge of the fixed delays to base its estimate on only the observations corresponding to the direction with lower variance, thereby eliminating

large contributions to its MSE caused by the queuing delay distribution that has higher variance.

Further, since the M-model estimator can use information from B past blocks to estimate the fixed delay, we expect it to achieve the performance of the K-model estimator in the limiting case where $B \to \infty$. In our simulations, we observe that M-model minimax estimator closely approaches the K-model minimax estimator in performance for fairly small values of B (between 5 and 20).

5.7 Summary

We derived minimax optimum estimators for a general class of location parameter problems, and applied them to the problem of phase offset estimation under multiple observation models. In cases where the pdf of the queuing delays are known, minimax estimators can be used to obtain the best possible estimation performance. The MSE curves of the minimax estimators can also serve as a design tool for practical synchronization deployments, by providing fundamental limits on POE performance for a given set of network conditions. Our simulation results indicate that conventional estimators can perform close to optimum in certain low-load scenarios with symmetric queuing delay distributions. However, optimum estimators appear to provide significant performance benefits in scenarios where the queuing delay distributions are asymmetric, a case that occurs frequently in practice. The results in this chapter could help guide the development of new POE schemes that address synchronization challenges arising in current and future generations of mobile networks.

5.8 Appendix

Proof of Lemma 3. For any shift invariant estimator $g(\mathbf{x})$, we can show that if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are any two values of the parameter vector with $\mathbf{h} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$, then

$$\begin{aligned}
&\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}_1) \\
&= \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}(\boldsymbol{\theta}_2 + \mathbf{h})]^2 f_0(\mathbf{x} - \mathbf{G}(\boldsymbol{\theta}_2 + \mathbf{h}))\mathrm{d}\mathbf{x} && (5.58)\\
&= \int_{\mathbb{R}^N} [g(\mathbf{x} - \mathbf{G}\mathbf{h}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_2]^2 f((\mathbf{x} - \mathbf{G}\mathbf{h})|\boldsymbol{\theta}_2)\mathrm{d}\mathbf{x} && (5.59)\\
&= \int_{\mathbb{R}^N} [g(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}_2]^2 f(\mathbf{x}|\boldsymbol{\theta}_2)\mathrm{d}\mathbf{x} && (5.60)\\
&\qquad \text{(using a change of variables)} \\
&= \mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}_2) && (5.61)
\end{aligned}$$

Hence $g(\mathbf{x})$ has constant conditional MSE w.r.t. $\boldsymbol{\theta}$. Further, using the definitions of the maximum and average MSEs, we obtain $\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) = \mathcal{M}(g(\mathbf{x})) = \mathcal{B}(g(\mathbf{x}), p(\boldsymbol{\theta}))$. □

Proof of Theorem 5. (i) It is simple to show that $g^*(\mathbf{x})$ is shift invariant, since

$$\begin{aligned}
g^*(\mathbf{x} + \mathbf{G}\mathbf{h}) &= \frac{\int_{\mathbb{R}^M} [\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\theta}}] f_0(\mathbf{x} + \mathbf{G}\mathbf{h} - \mathbf{G}\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f_0(\mathbf{x} + \mathbf{G}\mathbf{h} - \mathbf{G}\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}} && (5.62)\\
&= \frac{\int_{\mathbb{R}^M} [\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\theta}}] f(\mathbf{x}|\hat{\boldsymbol{\theta}} - \mathbf{h})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}} - \mathbf{h})\ \mathrm{d}\hat{\boldsymbol{\theta}}} && (5.63)\\
&= \frac{\int_{\mathbb{R}^M} [\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\theta}}] f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}} + \mathbf{c}^{\mathrm{T}}\mathbf{h} && (5.64)\\
&= g^*(\mathbf{x}) + \mathbf{c}^{\mathrm{T}}\mathbf{h} && (5.65)
\end{aligned}$$

(ii) For any choice of prior distribution $p(\boldsymbol{\theta})$, any estimator $g(\mathbf{x})$ of $\mathbf{c}^T\boldsymbol{\theta}$ satisfies

$$\mathcal{M}(g(\mathbf{x})) \geq \sup_{p(\boldsymbol{\theta})} \mathcal{B}(g(\mathbf{x}), p(\boldsymbol{\theta})) \tag{5.66}$$

$$\geq \sup_{p(\boldsymbol{\theta})} \inf_{\tilde{g}(\mathbf{x})} \mathcal{B}(\tilde{g}(\mathbf{x}), p(\boldsymbol{\theta})) = \mathcal{B}_0 \tag{5.67}$$

Further, it can be proved (by contradiction) that $\mathcal{M}(g(\mathbf{x})) = \mathcal{B}_0$ holds if and only if $g(\mathbf{x})$ is minimax. Now consider the estimator $g^*(\mathbf{x})$ of (5.30). From (5.67), we already have $\mathcal{M}(g^*(\mathbf{x})) \geq \mathcal{B}_0$. We shall now show that $\mathcal{B}_0 \geq \mathcal{M}(g^*(\mathbf{x}))$ also holds, hence proving that $\mathcal{B}_0 = \mathcal{M}(g^*(\mathbf{x}))$, and thus that $g^*(\mathbf{x})$ is minimax.

Consider a sequence of prior distributions $p_i(\boldsymbol{\theta})$, each uniformly distributed over a support set $\boldsymbol{\Theta}_i$ for $i = 1, 2, \cdots$, where

$$\boldsymbol{\Theta}_i = \{\boldsymbol{\theta} : (-i) \cdot \mathbf{1}_M \leq \boldsymbol{\theta} \leq i \cdot \mathbf{1}_M\} \tag{5.68}$$

Here the inequality $(-i) \cdot \mathbf{1}_M \leq \boldsymbol{\theta} \leq i \cdot \mathbf{1}_M$ implies that all the elements of the vector $\boldsymbol{\theta}$ lie in the interval $[-i, i]$. Given a prior distribution $p_i(\boldsymbol{\theta})$, the estimator that minimizes $\mathcal{B}(g(\mathbf{x}), p(\boldsymbol{\theta}))$ is the minimum mean square error (MMSE) estimator,

$$g_i(\mathbf{x}) = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_i} [\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}] f_i(\boldsymbol{\theta}|\mathbf{x}) \, \mathrm{d}\boldsymbol{\theta} \tag{5.69}$$

where $f_i(\boldsymbol{\theta}|\mathbf{x})$ represents the posterior pdf

$$f_i(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta}) p_i(\boldsymbol{\theta})}{\int_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_i} f(\mathbf{x}|\tilde{\boldsymbol{\theta}}) p_i(\tilde{\boldsymbol{\theta}}) \, \mathrm{d}\tilde{\boldsymbol{\theta}}} = \frac{f(\mathbf{x}|\boldsymbol{\theta})}{\int_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_i} f(\mathbf{x}|\tilde{\boldsymbol{\theta}}) \, \mathrm{d}\tilde{\boldsymbol{\theta}}} \tag{5.70}$$

Hence we can write

$$\mathcal{B}_0 = \sup_{p(\boldsymbol{\theta})} \inf_{\tilde{g}(\mathbf{x})} \mathcal{B}(\tilde{g}(\mathbf{x}), p(\boldsymbol{\theta}))$$

$$\geq \inf_{\tilde{g}} \mathcal{B}(\tilde{g}(\mathbf{x}), p_i(\boldsymbol{\theta})) = \mathcal{B}(g_i(\mathbf{x}), p_i(\boldsymbol{\theta})) \tag{5.71}$$

Further, since $f_0(\mathbf{x})$ has finite support, we have

$$\lim_{i\to\infty} g_i(\mathbf{x}) = \lim_{i\to\infty} \frac{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} [\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}] f(\mathbf{x}|\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_i} f(\mathbf{x}|\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}}$$
$$= \frac{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(\mathbf{x})} [\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}] f_0(\mathbf{x}-\mathbf{G}\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(\mathbf{x})} f_0(\mathbf{x}-\mathbf{G}\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}} = g^*(\mathbf{x})$$

where

$$\boldsymbol{\Theta}(\mathbf{x}) = \{\boldsymbol{\theta} : (\mathbf{x}-\mathbf{G}\boldsymbol{\theta}) > 0 \ \text{ and } \ (\mathbf{x}-\mathbf{G}\boldsymbol{\theta}) < L\cdot\mathbf{1}_N\} \tag{5.72}$$

and hence

$$\lim_{i\to\infty} \mathcal{B}(g_i(\mathbf{x}), p_i(\boldsymbol{\theta}))$$
$$= \lim_{i\to\infty} \mathcal{B}(g^*(\mathbf{x}), p_i(\boldsymbol{\theta})) \tag{5.73}$$
$$= \lim_{i\to\infty} \mathcal{M}(g^*(\mathbf{x})) \quad (\text{Since } g^*(\mathbf{x}) \text{ is shift invariant}) \tag{5.74}$$
$$= \mathcal{M}(g^*(\mathbf{x})) \tag{5.75}$$

From (5.71) and (5.75), we obtain $\mathcal{B}_0 \geq \mathcal{M}(g^*(\mathbf{x}))$, hence completing the proof.

(iii) Since $g^*(\mathbf{x})$ is shift invariant, from Lemma 3 we have $\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta}) = \mathcal{M}(g(\mathbf{x}))$. Further, since all shift invariant estimators have constant conditional MSE, and $g^*(\mathbf{x})$ minimizes $\mathcal{M}(g(\mathbf{x}))$, it also minimizes $\mathcal{R}(g(\mathbf{x}), \boldsymbol{\theta})$ for every value of $\boldsymbol{\theta}$.

(iv) We shall prove that $g^*(\mathbf{x})$ is unbiased by contradiction. Assume $g^*(\mathbf{x})$ is biased. Since $g^*(\mathbf{x})$ is shift invariant, its bias should be constant with respect to $\boldsymbol{\theta}$. Let

$$\beta = \mathrm{E}\left\{\left[g^*(\mathbf{x}) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}\right] \mid \boldsymbol{\theta}\right\} = \mathrm{E}\left\{g^*(\mathbf{x}) \mid \boldsymbol{\theta} = \mathbf{0}_M\right\} \tag{5.76}$$

denote this constant bias. Now consider a new estimator of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$, given as

$$\hat{g}(\mathbf{x}) = g^*(\mathbf{x}) - \beta \tag{5.77}$$

It is easy to show that $\hat{g}(\mathbf{x})$ is also shift invariant. Further,

$$\mathcal{M}\left(\hat{g}(\mathbf{x})\right) = \mathrm{E}\left\{\left[\hat{g}(\mathbf{x}) - \mathbf{c}^T\boldsymbol{\theta}\right]^2 \mid \boldsymbol{\theta}\right\} \tag{5.78}$$

$$= \mathrm{E}\left\{\left[g^*(\mathbf{x}) - \mathbf{c}^T\boldsymbol{\theta}\right]^2 \mid \boldsymbol{\theta}\right\} \tag{5.79}$$

$$- 2\beta\mathrm{E}\left\{\left[g^*(\mathbf{x}) - \mathbf{c}^T\boldsymbol{\theta}\right] \mid \boldsymbol{\theta}\right\} + \beta^2 \tag{5.80}$$

$$= \mathcal{M}\left(g^*(\mathbf{x})\right) - \beta^2 \quad < \quad \mathcal{M}\left(g^*(\mathbf{x})\right) \tag{5.81}$$

However, this is impossible since $g^*(\mathbf{x})$ has already been shown to minimize the maximum MSE. Thus, the assumption that $g^*(\mathbf{x})$ is biased is incorrect.

□

Proof of Lemma 4. Using theorem 5, we obtain

$$g_i^*(\mathbf{x}) = \frac{\int_{\mathbb{R}^M} \hat{\theta}_i f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}} \tag{5.82}$$

and

$$g^*(\mathbf{x}) = \frac{\int_{\mathbb{R}^M} [\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\theta}}] f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}} \tag{5.83}$$

$$= \frac{\sum_{i=1}^M c_i \int_{\mathbb{R}^M} \theta_i f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}}{\int_{\mathbb{R}^M} f(\mathbf{x}|\hat{\boldsymbol{\theta}})\ \mathrm{d}\hat{\boldsymbol{\theta}}} = \sum_{i=1}^M c_i g_i^*(\mathbf{x}) \tag{5.84}$$

hence proving the theorem. □

Proof of Theorem 6. Consider the problem of estimating $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$ given all the observations $\mathbf{x}_1, \cdots, \mathbf{x}_K$. It is easy to show that this problem is a vector location parameter problem as defined in Section 5.3. Hence, a minimax optimum estimator for this problem can be obtained via Theorem 5. Denote this minimax estimator as $h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)$. Further, note that $h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)$ and $g^*(\mathbf{x})$ are both estimators of $\mathbf{c}^{\mathrm{T}}\boldsymbol{\theta}$, but $h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)$ has more information available to it, since $\sum_{k=1}^{K} \mathbf{x}_k = \mathbf{x}$. Hence, we must have

$$\mathcal{M}(g^*(\mathbf{x})) \geq \mathcal{M}(h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)) \tag{5.85}$$

Further, using Lemma 4, it can be shown that

$$h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K) = \sum_{k=1}^{K} h_k^*(\mathbf{x}_k) \tag{5.86}$$

Since minimax estimators are shift invariant, we can write

$$\mathcal{M}(h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)) \tag{5.87}$$

$$= \sup_{\boldsymbol{\theta} \in \mathbb{R}^N} \mathrm{E}\left\{ \left[h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K) - \mathbf{c}^T\boldsymbol{\theta}\right]^2 \,\middle|\, \boldsymbol{\theta} \right\} \tag{5.88}$$

$$= \mathrm{E}\left\{ [h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)]^2 \,\middle|\, \boldsymbol{\theta} = \mathbf{0}_M \right\} \tag{5.89}$$

This can be further simplified as

$$\mathcal{M}(h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)) \tag{5.90}$$

$$= \mathrm{E}\left\{ \left[\sum_{k=1}^{K} h_k^*(\mathbf{x}_k) \right]^2 \;\middle|\; \boldsymbol{\theta}_1 = \mathbf{0}_M, \cdots, \boldsymbol{\theta}_K = \mathbf{0}_M \right\} \tag{5.91}$$

$$= \sum_{k=1}^{K} \mathrm{E}\left\{ [h_k^*(\mathbf{x}_k)]^2 \middle| \; \boldsymbol{\theta}_k = \mathbf{0}_M \right\} + \sum_{k_1=1}^{K} \sum_{\substack{k_2=1 \\ k_2 \neq k_1}}^{K} \Bigg[\mathrm{E}\left\{ h_{k_1}^*(\mathbf{x}_{k_1}) h_{k_2}^*(\mathbf{x}_{k_2}) \;\middle|\; \boldsymbol{\theta}_{k_1} = \mathbf{0}_M, \boldsymbol{\theta}_{k_2} = \mathbf{0}_M \right\} \Bigg] \tag{5.92}$$

We note that $h_1^*(\mathbf{x}_1), \cdots, h_K^*(\mathbf{x}_K)$ are all mutually independent conditioned on the unknown parameters, due to our initial assumption that $\mathbf{x}_1, \cdots, \mathbf{x}_K$ are mutually independent as per (5.32). Hence, we obtain

$$\mathcal{M}(h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)) \tag{5.93}$$

$$= \sum_{k=1}^{K} \mathrm{E}\left\{ [h_k^*(\mathbf{x}_k)]^2 \middle| \; \boldsymbol{\theta}_k = \mathbf{0}_M \right\} + \sum_{k_1=1}^{K} \sum_{\substack{k_2=1 \\ k_2 \neq k_1}}^{K} \mathrm{E}\left\{ h_{k_1}^*(\mathbf{x}_{k_1}) \middle| \; \boldsymbol{\theta}_{k_1} = \mathbf{0}_M \right\} \cdot \mathrm{E}\left\{ h_{k_2}^*(\mathbf{x}_{k_2}) \middle| \; \boldsymbol{\theta}_{k_2} = \mathbf{0}_M \right\} \tag{5.94}$$

Since $h_k^*(\mathbf{x}_k)$ is a minimax estimator, it is unbiased and shift invariant according to Theorem 5, and hence

$$\mathrm{E}\left\{ h_k^*(\mathbf{x}_k) \middle| \; \boldsymbol{\theta} = \mathbf{0}_M \right\} = 0 \;, \tag{5.95}$$

$$\mathrm{E}\left\{ [h_{k_1}^*(\mathbf{x}_{k_1})]^2 \middle| \; \boldsymbol{\theta}_{k_1} = \mathbf{0}_M \right\} = \mathcal{M}(h_k^*(\mathbf{x}_k)) \tag{5.96}$$

From (5.94), (5.95) and (5.96) we obtain

$$\mathcal{M}(h^*(\mathbf{x}_1, \cdots, \mathbf{x}_K)) = \sum_{k=1}^{K} \mathcal{M}(h_k^*(\mathbf{x}_k)) \tag{5.97}$$

Finally, from (5.85) and (5.97), we obtain

$$\mathcal{M}(g^*(\mathbf{x})) \geq \sum_{k=1}^{K} \mathcal{M}(h_k^*(\mathbf{x}_k)) \tag{5.98}$$

hence concluding the proof.

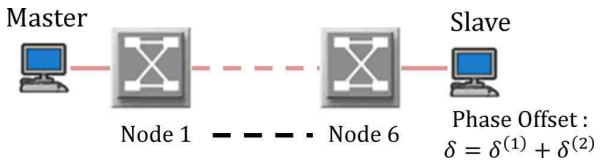
□

Proof of Corollary 1. We shall prove this corollary by applying Theorem 6 to POE under the S-model. To this end, assume $N = KL$, where K and L are both integers. For the N-node network, assuming P pairs of timestamp differences are collected per the S-model, the observation vector can be written similar to (5.12), as

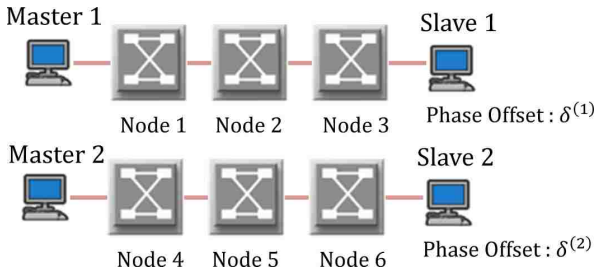
$$\mathbf{y} = d\mathbf{1}_{2P} + \delta\mathbf{e} + \mathbf{w} \tag{5.99}$$

where d and δ represent the unknown fixed delay and phase offset, while $\mathbf{w}$ represents the $2P \times 1$ vector of queuing delays.

Now suppose that the cascade of $N = KL$ nodes is split into K smaller cascades, each consisting of L nodes. Each cascade of L nodes is placed between a new master-slave pair, resulting in K new networks (see example in Fig. 5.2). Let the phase offset of the slave in the k^{th} network be $\delta^{(k)}$, and let the fixed delay in the k^{th} network be $d^{(k)}$. Assume that the



(a) Original network



(b) Networks obtained after splitting

Figure 5.2: Example of a network containing $N = 6$ intermediate nodes, that has been split into $K = 2$ networks, each containing $L = 3$ intermediate nodes.

phase offsets and fixed delays satisfy the relation

$$\sum_{k=1}^{K} \delta^{(k)} = \delta \ , \qquad \sum_{k=1}^{K} d^{(k)} = d \tag{5.100}$$

Assuming that P observation pairs are collected per the S-model, the observation vector for each L-node network can be written, similar to (5.12), as

$$\mathbf{y}^{(k)} = d^{(k)}\mathbf{1}_{2P} + \delta^{(k)}\mathbf{e} + \mathbf{w}^{(k)} \tag{5.101}$$

for $k = 1, \cdots, K$. Here $\mathbf{w}^{(k)}$ represents the $2P \times 1$ vector of queuing delays in the k^{th} network. Since the single-node queuing delays across all nodes are identically distributed, the minimax MSE associated with estimating $\delta^{(k)}$ from $\mathbf{y}^{(k)}$ will be identical, and equal $\rho(L)$ in all the L-node networks. Note that due to the shift invariance of the minimax estimator and the result in Lemma 3, the minimax MSE will remain unchanged regardless of the assumption in (5.100), since the minimax MSE does not depend on the value of $\delta^{(k)}$ or $d^{(k)}$.

In order to apply Theorem 6, we note that the queuing delay vector under the KL node network can be written as sum of the queuing delay vectors under each L-node network, i.e. $\mathbf{w} = \sum_{k=1}^{K} \mathbf{w}^{(k)}$. Further, due to the assumption that the single-node queuing delays are mutually independent, we have

$$f(\mathbf{y}^{(k_1)}, \mathbf{y}^{(k_2)} | \delta^{(k_1)}, d^{(k_1)}, \delta^{(k_2)}, d^{(k_2)}) = f(\mathbf{y}^{(k_1)} | \delta^{(k_1)}, d^{(k_1)}) f(\mathbf{y}^{(k_2)} | \delta^{(k_2)}, d^{(k_2)}).$$

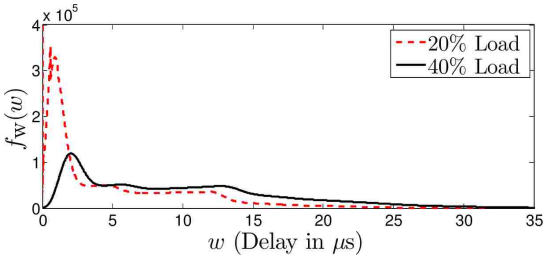
Due to the assumption in (5.100), we also have

$$\begin{aligned} \mathbf{y} &= d\mathbf{1}_{2P} + \delta\mathbf{e} + \mathbf{w} && (5.102) \\ &= \left[\sum_{k=1}^{K} d^{(k)}\right] \mathbf{1}_{2P} + \left[\sum_{k=1}^{K} \delta^{(k)}\right] \mathbf{e} + \left[\sum_{k=1}^{K} \mathbf{w}^{(k)}\right] && (5.103) \\ &= \sum_{k=1}^{K} \left[d^{(k)}\mathbf{1}_{2P} + \delta^{(k)}\mathbf{e} + \mathbf{w}^{(k)}\right] \quad = \quad \sum_{k=1}^{K} \mathbf{y}^{(k)} && (5.104) \end{aligned}$$

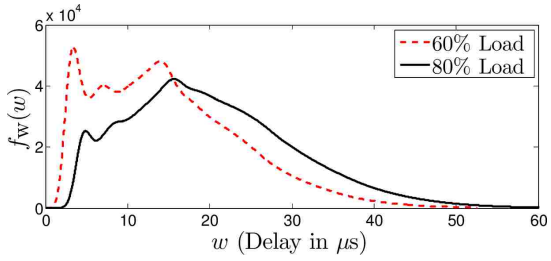
Noting the similarity in the relationships between $\mathbf{y}$, $\mathbf{y}^{(k)}$ and the vectors $\mathbf{x}$, $\mathbf{x}_k$ in Theorem 6, we can apply Theorem 6 to obtain the relation

$$\rho(KL) \geq K\rho(L) \tag{5.105}$$

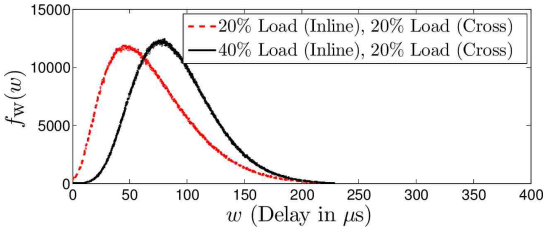
which concludes the proof. □



(a) TM1, $20 - 40\%$ Load

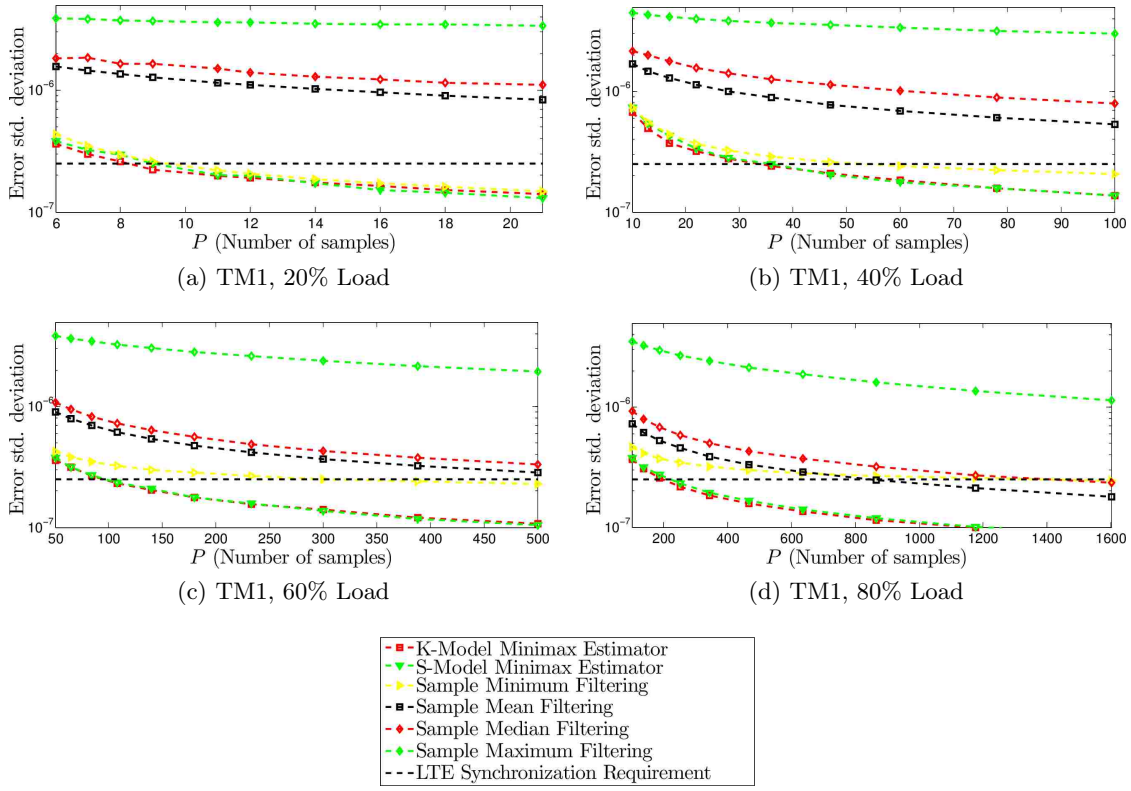


(b) TM1, $60 - 80\%$ Load



(c) Mixed Traffic, with TM2 for cross traffic and uniform packet size distribution for inline traffic.

Figure 5.3: Plots of queuing delay distributions under different network conditions



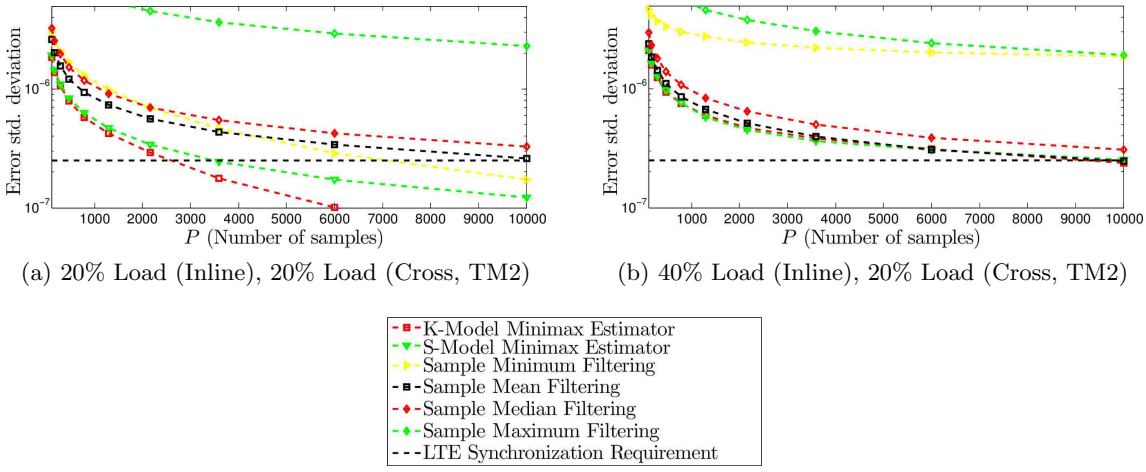
(a) TM1, 20% Load

(b) TM1, 40% Load

(c) TM1, 60% Load

(d) TM1, 80% Load

Figure 5.4: Performance comparison of different estimators under symmetric cross traffic.



(a) 20% Load (Inline), 20% Load (Cross, TM2)

(b) 40% Load (Inline), 20% Load (Cross, TM2)

Figure 5.5: Performance comparison of different estimators under symmetric mixed traffic.

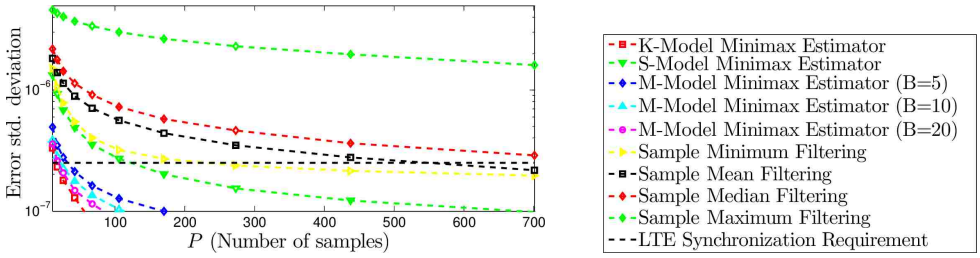


Figure 5.6: Performance comparison of different estimators under asymmetric cross traffic. Forward path: 80% Load (TM1), Reverse path: 20% Load (TM1).

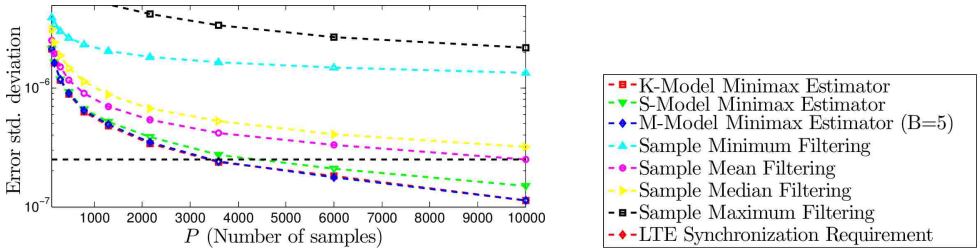


Figure 5.7: Performance comparison of different estimators under asymmetric mixed traffic. Traffic models used are TM2 for cross traffic and uniform packet size distribution for inline traffic. The forward path has 40% inline load and 20% cross load, while the reverse path has 20% inline load and 20% cross load.

Chapter 6

Optimum Design of L-Estimators for Phase offset estimation in IEEE 1588

6.1 Introduction

Many recent papers have studied techniques to improve the resilience of network time synchronization protocols against random queuing delays [22][54][55][50][56][47][17][21]. In Chapter 5 (also see [26]), we described new POE schemes that are optimum in terms of minimizing the worst-case mean squared error (MSE), and are hence termed minimax estimators. While these minimax estimators achieve optimum performance, they require complete knowledge of the probability density function (pdf) of the queuing delay to be designed, and have a high computational complexity. In this chapter, we describe new estimators with many practical advantages relative to these minimax estimators; they have a much lower computational

complexity, require lesser statistical knowledge of the queuing delays, and exhibit a mean squared estimation error very close to minimax estimators.

As discussed in Chapter 5, from a statistical perspective the problem of POE can be classified as a location parameter problem, wherein the effect of the unknown parameters is to shift the pdf of the observations, without changing the shape of the pdf. Further, the minimax estimators of [26] can be classified as M-estimators [57], defined as a class of estimators that can be obtained as the zeros of an estimating function [58]. Another class of estimators that are popular in the context of location parameter problems are L-estimators [59][60], which are estimators that are obtained as linear combinations of the order statistics of a set of observations. Here the order statistics of a set of observations refer simply to the same observations rearranged in nondecreasing order. Given that L-estimators are computed by sorting followed by a linear combination operation, they have a much lower complexity than the minimax estimators of [26], and a complexity not significantly greater than the sample minimum, mean, median or maximum filtering schemes. In fact, the latter four filtering schemes can all be described as L-estimator with fixed linear combination weights. In this chapter, we propose a number of novel L-estimator structures for POE, and solve the problem of optimizing the weights under different optimality criteria.

We note that L-estimators with optimized weights have been previously studied in the context of POE. In particular, [61] and [62] showed that in the case of exponentially distributed queuing delays, both the maximum likelihood estimator (MLE) and the minimum variance unbiased estimator (MVUE) of phase offset are obtained as L-estimators. Further, in [62], the best linear unbiased estimator using order statistics (BLUE-OS) was also derived. In this chapter, we extend on the BLUE-OS estimator of [62] by constructing optimum L-estimators under many different POE observation models and optimality criteria that have

not been considered previously in literature.

In order to study the problem of POE, we model the end-to-end (ETE) delays along the forward and reverse network paths between the master and slave as the sum of a fixed minimum delay and a random queuing delay. We then consider two models for observations available to the slave. Under the *known fixed delays model* (K-model), we assume that the fixed delays in both the forward and reverse directions are known to the slave. Under the *standard model* (S-model), we assume that the fixed delays are unknown, but identical in the forward and reverse directions. Our key new contributions in this chapter are as follows:

1. *L-estimators under known queuing delay distributions*: Given perfect statistical knowledge of the queuing delays, we derive optimum L-estimators under the K-model as well as the S-model. Results are derived for the general case where the order statistics of the forward and reverse queuing delays are correlated, and further simplified for the special case where they are uncorrelated.

2. *L-estimators under network model uncertainty*: We consider a scenario where perfect statistical knowledge of the queuing delays is not available. To model this problem, we assume that we are given a finite set of distributions from which the queuing delays may arise. We then derive optimum L-estimators that minimize the worst-case estimation error within the given set of distributions, for both the K-model and S-model.

3. *L-estimators that exploit information from past observation windows*: We first define an observation window as a set of observations across which the slave's phase offset can be assumed to be constant. While our previous results assume that all the observations belong to a single observation window, here we also describe and optimize L-estimator structures that can utilize past observation windows to improve estimation performance.

In order to study the efficiency of the proposed L-estimators, we compare their performance against the minimax optimum estimators of [53] under network scenarios motivated by the ITU-T recommendation G.8261 [45]. Results indicate that their MSE closely approximates that of minimax estimators under the tested network conditions.

6.2 System Model

Consider a synchronization problem where the slave has a phase offset δ relative to its master. To help determine δ, a two-way message exchange (Fig. 6.1) is performed between the master and slave, which involves the following steps:

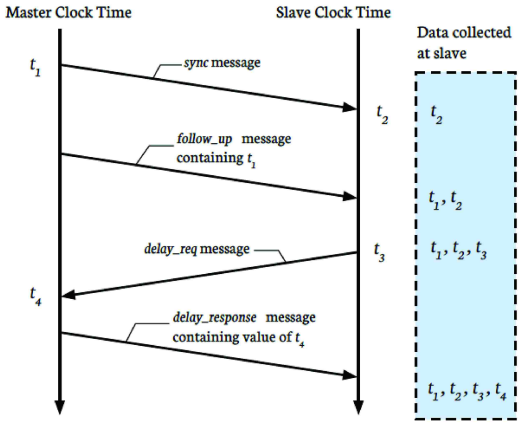


Figure 6.1: Two-way Synchronization between a master and slave

1. The master initiates the exchange by sending a *SYNC* message to the slave at time t_1. A *FOLLOW_UP* message later communicates the value of t_1 to the slave.

2. The slave records the time of reception of the *SYNC* message as $t_2 = t_1 + d_1 + \delta$, where d_1 denotes the ETE network delay between the master and the slave.

3. The slave responds with a *DELAY_REQ* message, and records its time of transmission as t_3.

4. The master records the time of arrival of the *DELAY_REQ* message as $t_4 = t_3 - \delta + d_2$, where d_2 denotes the ETE network delay between the slave and the master. The value of t_4 is sent to the slave via a *DELAY_RESP* message.

While four timestamps (t_1, t_2, t_3, t_4) are available after each two-way exchange, in order to estimate δ it is clearly sufficient for the slave to only retain the pair of timestamp differences

$$y_1 = t_2 - t_1 = d_1 + \delta \tag{6.1}$$

$$y_2 = t_4 - t_3 = d_2 - \delta \tag{6.2}$$

In a typical network, the ETE delays d_1 and d_2 receive contributions from three factors:

(a) Constant propagation delays along network links between the master and the slave (or vice-versa).

(b) Constant processing delays at intermediate nodes (such as switches or routers) along each network path.

(c) Random queuing delays at intermediate nodes along each network path.

Hence, each ETE delay can be modeled as

$$d_1 = d_1^{\min} + w_1, \qquad d_2 = d_2^{\min} + w_2 \tag{6.3}$$

Here $d_1^{\min}$ and $d_2^{\min}$ denote fixed delays corresponding to the sum of the constant propagation and processing delays, while w_1 and w_2 model the random queuing delays.

Assuming the values of δ, $d_1^{\min}$ and $d_2^{\min}$ remain constant over the duration of P two-way message exchanges, we can collect multiple observation pairs (y_1, y_2) to help estimate δ. We

denote these observations as

$$y^*_{i,1} = d_1^{\min} + \delta + w_{i,1} \ , \quad y^*_{i,2} = d_2^{\min} - \delta + w_{i,2} \tag{6.4}$$

for $i = 1, \cdots, P$. We refer to the problem of estimating δ from multiple observations of the timestamp differences (y_1, y_2) as the problem of *phase offset estimation* (POE). Further, we shall refer to a set of P observations of (y_1, y_2) that share the same values of $d_1^{\min}$, $d_2^{\min}$ and δ as an *observation window*. To study the problem of POE, we consider two models for the observations:

1. *Known fixed delay model (K-model)*: Under this model we assume that $d_1^{\min}$ and $d_2^{\min}$ are fully known at the slave. Hence, setting $y_{i,j} = y^*_{i,j} - d_j^{\min}$, we can obtain the simpler observation model

   $$y_{i,1} = \delta + w_{i,1} \ , \quad y_{i,2} = -\delta + w_{i,2} \tag{6.5}$$

   for $i = 1, \cdots, P$. These observations can be collected and described by the vector observation model

   $$\begin{aligned} \mathbf{y}_1 &= \delta \mathbf{1}_P + \mathbf{w}_1 \\ \mathbf{y}_2 &= -\delta \mathbf{1}_P + \mathbf{w}_2 \end{aligned} \tag{6.6}$$

   where

   $$\mathbf{y}_j = [y_{1,j} \ \cdots \ y_{P,j}]^{\mathrm{T}} \tag{6.7}$$

   $$\mathbf{w}_j = [w_{1,j} \ \cdots \ w_{P,j}]^{\mathrm{T}} \tag{6.8}$$

   for $j = 1, 2$ and $\mathbf{1}_P$ is a $P \times 1$ vector with all elements equal to 1.

2. *Standard model (S-model)*: Here we assume that only the difference between $d_1^{\min}$ and

$d_2^{\min}$, referred to as the *delay asymmetry*, is known to the slave. One example where this occurs is when the forward and reverse network paths are identical, and hence it is known that $d_1^{\min} - d_2^{\min} = 0$. By compensating the observations as

$$y_{i,1} = y_{i,1}^*, \quad y_{i,2} = y_{i,2}^* - d_2^{\min} + d_1^{\min} \tag{6.9}$$

we obtain

$$y_{i,1} = d + \delta + w_{i,1} \ , \quad y_{i,2} = d - \delta + w_{i,2} \tag{6.10}$$

for $i = 1, \cdots, P$, where $d_1^{\min} = d$. These observations can be denoted vectorially as

$$\begin{aligned} \mathbf{y}_1 &= d\mathbf{1}_P + \delta\mathbf{1}_P + \mathbf{w}_1 \\ \mathbf{y}_2 &= d\mathbf{1}_P - \delta\mathbf{1}_P + \mathbf{w}_2 \end{aligned} \tag{6.11}$$

where (6.7) and (6.8) were employed.

Let $\hat{\delta}$ denote any estimator of δ based on the observation vector $\mathbf{y}$. Following the approach adopted in Chapters 4 and 5, here we assume that both δ and d are deterministic under both observation models, i.e. no prior probability distributions are defined over either δ over d. Given such an assumption, a typical statistical approach is to define estimator optimality in the robust sense [57][59], where the performance metric is defined as

$$\mathcal{M}(\hat{\delta}) = \max_{\delta,d} \ \mathrm{MSE}(\hat{\delta} \mid \delta, d) \tag{6.12}$$

where $\mathrm{MSE}(\hat{\delta} \mid \delta, d)$ denotes the mean squared error (MSE),

$$\mathrm{MSE}(\hat{\delta} \mid \delta, d) = \mathrm{E}\left[(\hat{\delta} - \delta)^2 \mid \delta, d\right], \tag{6.13}$$

$$= \mathrm{Bias}(\hat{\delta} \mid \delta, d)^2 + \mathrm{var}(\hat{\delta} \mid \delta, d)^2 \tag{6.14}$$

where

$$\text{Bias}(\hat{\delta} \mid \delta, d) = \text{E}\left[\hat{\delta} - \delta \mid \delta, d\right], \tag{6.15}$$

$$\text{var}(\hat{\delta} \mid \delta, d) = \text{E}\left[\left(\hat{\delta} - \text{E}[\hat{\delta}]\right)^2 \mid \delta, d\right] \tag{6.16}$$

respectively represent the estimator bias and variance. The expectations in (6.13), (6.15) and (6.16) are taken with respect to the conditional pdf $f(\mathbf{y}|\delta)$ (under the K-model) or $f(\mathbf{y}|\delta, d)$ (under the S-model). In this chapter, we shall focus on the optimization of L-estimators to minimize $\mathcal{M}(\hat{\delta})$. It is easy to show for L-estimators that unless $\text{Bias}(\hat{\delta} \mid \delta, d)$ is a constant independent of both δ and d, we will have $\mathcal{M}(\hat{\delta}) = \infty$. Hence, in our analysis, we shall only consider estimators whose bias is constant with respect to both δ and d. Such estimators will also have a MSE and variance that is constant with respect to both δ and d, hence the conditioning of the quantities in (6.13) - (6.16) on δ and d can be dropped.

6.3 Optimum L-Estimators when statistics of queuing delays are known

Given a vector $\mathbf{x} = [x_1 \cdots x_N]^{\text{T}}$, the i^{th} order statistic of this vector is defined as the i^{th} largest element of the vector. We shall refer to the vector containing all the order statistics of $\mathbf{x}$, ordered from smallest to largest, as the *order statistic vector* of $\mathbf{x}$, and denote it as $\lfloor \mathbf{x} \rfloor$. Now consider a POE scheme which is a linear combination of order statistics,

$$\hat{\delta} = \mathbf{c}_1^{\text{T}} \lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\text{T}} \lfloor \mathbf{y}_2 \rfloor + \eta \tag{6.17}$$

where $\mathbf{c}_1$, $\mathbf{c}_2$ are weight vectors and η is a scalar constant. It is easy to see that $\hat{\delta}$ corresponds to conventional estimators for the following values of $\mathbf{c}_1$ and $\mathbf{c}_2$ (with $\eta = 0$):

$$\textit{Sample minimum estimator: } \mathbf{c}_1 = \mathbf{c}_2 = [0.5\ \mathbf{0}_{P-1}^{\mathrm{T}}]^{\mathrm{T}} \tag{6.18}$$

$$\textit{Sample mean estimator: } \mathbf{c}_1 = \mathbf{c}_2 = P^{-1}\mathbf{1}_P \tag{6.19}$$

Sample median estimator:

$$\mathbf{c}_1 = \mathbf{c}_2 = \begin{cases} [\mathbf{0}_{\frac{P}{2}-1}^{\mathrm{T}}\ 1\ \mathbf{0}_{\frac{P}{2}-1}^{\mathrm{T}}]^{\mathrm{T}} & \text{for } P \text{ odd} \\ [\mathbf{0}_{\frac{P}{2}-1}^{\mathrm{T}}\ 1\ 1\ \mathbf{0}_{\frac{P}{2}-1}^{\mathrm{T}}]^{\mathrm{T}} & \text{for } P \text{ even} \end{cases} \tag{6.20}$$

$$\textit{Sample maximum estimator: } \mathbf{c}_1 = \mathbf{c}_2 = [\mathbf{0}_{P-1}^{\mathrm{T}}\ 0.5]^{\mathrm{T}} \tag{6.21}$$

We now consider the problem of designing $\mathbf{c}_1$, $\mathbf{c}_2$ and η to minimize the mean squared error under the constraint of constant bias. Define

$$\boldsymbol{\mu}_j = \mathrm{E}\left[\lfloor \mathbf{w}_j \rfloor\right], \quad \mathbf{S}_j = \mathrm{cov}\left\{\lfloor \mathbf{w}_j \rfloor\right\} \tag{6.22}$$

$$\mathbf{S}_{12} = \mathrm{cov}\left\{\lfloor \mathbf{w}_1 \rfloor, \lfloor \mathbf{w}_2 \rfloor\right\} \tag{6.23}$$

for $j = 1, 2$, and let

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \\ \mathbf{c}_2 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & -\mathbf{S}_{12} \\ \\ -\mathbf{S}_{12}^{\mathrm{T}} & \mathbf{S}_2 \end{bmatrix}. \tag{6.24}$$

We shall show that the optimum values of $\mathbf{c}_1$, $\mathbf{c}_2$ and η are functions of these mean vectors and covariance matrices. To this end, we first state the following general identity. The proof of the theorem and all its corollaries are provided in the Appendix.

Theorem 7 (**Quadratic programming problem**). *The value of the vector* $\mathbf{z}$ *that solves*

the constrained quadratic optimization problem

$$\begin{aligned} \min_{\mathbf{z}} \quad & \mathbf{z}^{\mathrm{T}}\mathbf{H}\mathbf{z} \\ \textit{subject to} \quad & \mathbf{G}\mathbf{z} = \mathbf{s} \ , \end{aligned} \tag{6.25}$$

is given as

$$\mathbf{z}^{*} = \mathbf{H}^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{H}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.26}$$

provided that $\mathbf{H}$ *is positive definite, and the system of linear equalities* $\mathbf{G}\mathbf{z} = \mathbf{s}$ *has a non-empty set of solutions.*

Theorem 7 can be used to optimize the estimator of (6.17) under both the K-model and S-model, as stated in the following corollaries.

Corollary 2 (**Optimum L-estimator under K-model**). *Under the K-model, the values of* $\mathbf{c}_1$, $\mathbf{c}_2$ *and* η *that minimize the MSE of an estimator of* δ *of the form*

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}}\lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}}\lfloor \mathbf{y}_2 \rfloor + \eta \tag{6.27}$$

given the constraint of constant bias are

$$\begin{bmatrix} \mathbf{c}_1^{*} \\ \mathbf{c}_2^{*} \end{bmatrix} = \frac{\mathbf{S}^{-1}\mathbf{1}_{2P}}{\mathbf{1}_{2P}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{1}_{2P}}, \tag{6.28}$$

$$\eta^{*} = \boldsymbol{\mu}_2^{\mathrm{T}}\mathbf{c}_2^{*} - \boldsymbol{\mu}_1^{\mathrm{T}}\mathbf{c}_1^{*} \ . \tag{6.29}$$

The resultant optimum estimator $\hat{\delta}^{*}$ *has an MSE*

$$\mathrm{MSE}(\hat{\delta}^{*}) = (\mathbf{1}_{2P}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{1}_{2P})^{-1} \tag{6.30}$$

Further, if $\lfloor \mathbf{w}_1 \rfloor$ *and* $\lfloor \mathbf{w}_2 \rfloor$ *are uncorrelated, then the optimum weights can be simplified as*

$$\mathbf{c}_j^* = \frac{\mathbf{S}_j^{-1}\mathbf{1}_P}{\mathbf{1}_P^{\mathrm{T}}\mathbf{S}_1^{-1}\mathbf{1}_P + \mathbf{1}_P^{\mathrm{T}}\mathbf{S}_2^{-1}\mathbf{1}_P} \tag{6.31}$$

for $j = 1, 2$ *and*

$$\mathrm{MSE}(\hat{\delta}^*) = (\mathbf{1}_P^{\mathrm{T}}\mathbf{S}_1^{-1}\mathbf{1}_P + \mathbf{1}_P^{\mathrm{T}}\mathbf{S}_2^{-1}\mathbf{1}_P)^{-1}. \tag{6.32}$$

is the associated optimum MSE.

Corollary 3 (**Optimum L-estimator under S-model**)**.** *Under the S-model, the values of* $\mathbf{c}_1$, $\mathbf{c}_2$ *and* η *that minimize the MSE of an estimator of* δ *of the form*

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}}\lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}}\lfloor \mathbf{y}_2 \rfloor + \eta \tag{6.33}$$

given the constraint of constant bias are

$$\begin{bmatrix} \mathbf{c}_1^* \\ \mathbf{c}_2^* \end{bmatrix} = \mathbf{S}^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{S}^{-1}\mathbf{A}^{\mathrm{T}})^{-1}\boldsymbol{\gamma} \ , \tag{6.34}$$

$$\eta^* = (\mathbf{c}_1^*)^{\mathrm{T}}\boldsymbol{\mu}_2 - (\mathbf{c}_2^*)^{\mathrm{T}}\boldsymbol{\mu}_1 \ , \tag{6.35}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \\ \mathbf{1}_P^{\mathrm{T}} & -\mathbf{1}_P^{\mathrm{T}} \end{bmatrix}, \qquad \boldsymbol{\gamma} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \ . \tag{6.36}$$

The resultant optimum estimator $\hat{\delta}^*$ *has an MSE*

$$\mathrm{MSE}(\hat{\delta}^*) = \boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{A}\mathbf{S}^{-1}\mathbf{A}^{\mathrm{T}})^{-1}\boldsymbol{\gamma} \tag{6.37}$$

Further, if $\lfloor \mathbf{w}_1 \rfloor$ *and* $\lfloor \mathbf{w}_2 \rfloor$ *are uncorrelated, then*

$$\mathbf{c}_j^* = \frac{1}{2} \frac{\mathbf{S}_j^{-1} \mathbf{1}_P}{\mathbf{1}_P^{\mathrm{T}} \mathbf{S}_j^{-1} \mathbf{1}_P} \tag{6.38}$$

for $j = 1, 2$ *and*

$$\mathrm{MSE}(\hat{\delta}^*) = (\mathbf{1}_P^{\mathrm{T}} \mathbf{S}_1^{-1} \mathbf{1}_P)^{-1} + (\mathbf{1}_P^{\mathrm{T}} \mathbf{S}_2^{-1} \mathbf{1}_P)^{-1}. \tag{6.39}$$

is the associated the optimum MSE.

For a fixed pair of forward and reverse queuing delay distributions, it is easy to see that the optimum MSE under the S-model in (6.37) necessarily exceeds that under the K-model in (6.30), since

$$\begin{aligned} &\boldsymbol{\gamma}^{\mathrm{T}} (\mathbf{A} \mathbf{S}^{-1} \mathbf{A}^{\mathrm{T}})^{-1} \boldsymbol{\gamma} \\ &= [(\mathbf{1}_{2P}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{1}_{2P}) - (\mathbf{e}_{2P}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{1}_{2P})^2 (\mathbf{e}_{2P}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{e}_{2P})^{-1}]^{-1} &(6.40) \\ &\geq (\mathbf{1}_{2P}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{1}_{2P})^{-1} &(6.41) \end{aligned}$$

where $\mathbf{e}_{2P} = [\mathbf{1}_P^{\mathrm{T}} \ \ (-\mathbf{1}_P^{\mathrm{T}})]^{\mathrm{T}}$. An intuitive explanation for this result is that compared to the K-model, the presence of the additional nuisance parameter d in the S-model increases the uncertainty associated with estimating δ.

6.4 Minimax Optimum L-Estimators under network model uncertainty

In Section 6.3, we derived optimum L-estimators given perfect knowledge of the mean vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and covariance matrices $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_{12}$. Now we consider a the case where these vectors and matrices are not known perfectly due to network model uncertainty. To this

end, we assume that there are K possible network scenarios, with $(\boldsymbol{\mu}_{1,k}, \boldsymbol{\mu}_{2,k}, \mathbf{S}_{1,k}, \mathbf{S}_{2,k}, \mathbf{S}_{12,k})$ denoting the mean vectors and covariance matrices under the k^{th} network scenario. Denote the bias, variance, and MSE of any estimator $\hat{\delta}$ of δ under the k^{th} network scenario as

$$\text{Bias}_k(\hat{\delta}) = \text{E}_k\left[\hat{\delta} - \delta\right], \tag{6.42}$$

$$\text{var}_k(\hat{\delta}) = \text{E}_k\left[\left(\hat{\delta} - \text{E}[\hat{\delta}]\right)^2\right], \tag{6.43}$$

$$\text{MSE}_k(\hat{\delta}) = \text{E}_k\left[(\hat{\delta} - \delta)^2\right]. \tag{6.44}$$

Here $\text{E}_k[\cdot]$ represents an expectation computed by utilizing the mean vectors and covariance matrices corresponding to the k^{th} network scenario. While there are K different MSE values associated with any estimator $\hat{\delta}$, we require a single scalar performance metric to characterize and subsequently optimize estimation performance. To address this issue, we consider a new metric we term the *weighted maximum MSE*, defined as

$$\text{WMaxMSE}(\hat{\delta}) = \max_{k \in \{1,\cdots,K\}} \beta_k \text{MSE}_k(\hat{\delta}) \tag{6.45}$$

where β_k are constant positive weights for $k = 1, \cdots, K$. We shall consider the design of L-estimators to minimize $\text{WMaxMSE}(\hat{\delta})$ in this section. Note that $\beta_1, \cdots, \beta_K$ in (6.45) represent design parameters, the values of which depend upon the design philosophy adopted by the system designer. For example, when $\beta_k = 1$ for all k, $\text{WMaxMSE}(\hat{\delta})$ corresponds to the *maximum MSE* or *worst-case MSE*

$$\text{MaxMSE}(\hat{\delta}) = \max_{k \in \{1,\cdots,K\}} \text{MSE}_k(\hat{\delta}) \tag{6.46}$$

The optimum estimator under such a choice of weights focuses on the worst-case performance, which has intuitive appeal. However, when the set of K network scenarios considered are highly disparate in terms of achievable performance, it may be more prudent to choose the

weights as $\beta_k = (\text{MMSE}_k)^{-1}$, where MMSE_k denotes the minimum achievable MSE under the k^{th} network scenario. This corresponds to a metric we term the *maximum efficiency* or *worst-case efficiency*, defined as

$$\text{MaxEFF}(\hat{\delta}) = \max_{k\in\{1,\cdots,K\}} \frac{\text{MSE}_k(\hat{\delta})}{\text{MMSE}_k} \tag{6.47}$$

We now consider the problem of optimizing L-estimators to minimize $\text{WMaxMSE}(\hat{\delta})$ given arbitrary positive values for the weights $\beta_1, \cdots, \beta_K$. To this end, we note that (6.45) can be equivalently written as

$$\text{WMaxMSE}(\hat{\delta}) = \max_{\boldsymbol{\lambda}\in\Lambda} \sum_{k=1}^{K} \lambda_k \beta_k \text{MSE}_k(\hat{\delta}) \tag{6.48}$$

where $\boldsymbol{\lambda} = [\lambda_1 \ \cdots \ \lambda_K]^{\text{T}}$, and

$$\Lambda = \left\{ \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_K \end{bmatrix} \in \mathbb{R}^K \ \middle| \ \left(\sum_{k=1}^{K} \lambda_k = 1 \right) \wedge (\lambda_k \geq 0 \ \forall \ k) \right\} \tag{6.49}$$

The problem of designing L-estimators to minimize (6.48) can be addressed using a well known result from minimax optimization theory, restated here convenience[1].

Theorem 8 (**Property of minimax optimization problems**). *Let $\mathcal{C}$ and $\mathcal{D}$ be convex sets in $\mathcal{R}^m$ and $\mathcal{R}^n$ respectively. Let $f(\mathbf{u}, \mathbf{v})$ be a continuous function that is concave with respect to $\mathbf{u} \in \mathcal{C}$, and convex with respect to $\mathbf{v} \in \mathcal{D}$. If either $\mathcal{C}$ or $\mathcal{D}$ is bounded, then we have*

$$\inf_{\mathbf{v}\in\mathcal{D}} \sup_{\mathbf{u}\in\mathcal{C}} f(\mathbf{u}, \mathbf{v}) = \sup_{\mathbf{u}\in\mathcal{C}} \inf_{\mathbf{v}\in\mathcal{D}} f(\mathbf{u}, \mathbf{v}) \tag{6.50}$$

[1] See corollaries 37.3.1 and 37.3.2 of [63] for a proof of this result.

and the function

$$f^*(\mathbf{u}) = \inf_{\mathbf{v}\in\mathcal{D}} \; f(\mathbf{u},\mathbf{v}) \tag{6.51}$$

is concave with respect to $\mathbf{u} \in \mathcal{C}$.

Theorem 8 can be used to derive L-estimators under both the K-model and S-model, as stated in the following corollaries.

Corollary 4. ***(WMaxMSE-optimum L-estimator under K-model)*** *Under the K-model, the values of* $\mathbf{c}_1$, $\mathbf{c}_2$ *and* η *that minimize* $\mathrm{WMaxMSE}(\hat{\delta})$ *for an estimator* $\hat{\delta}$ *of* δ *of the form*

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}} \lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}} \lfloor \mathbf{y}_2 \rfloor + \eta \tag{6.52}$$

given the constraint of constant bias are

$$\begin{bmatrix} \mathbf{c}_1^* \\ \mathbf{c}_2^* \end{bmatrix} = \frac{1}{\mathbf{1}_{2P}^{\mathrm{T}} (\mathbf{M}^*)^{-1} \mathbf{1}_{2P}} (\mathbf{M}^*)^{-1} \mathbf{1}_{2P} \; , \tag{6.53}$$

$$\eta^* = -\sum_{k=1}^{K} \lambda_k^* \beta_k \left((\mathbf{c}_1^*)^{\mathrm{T}} \boldsymbol{\mu}_{1,k} - (\mathbf{c}_2^*)^{\mathrm{T}} \boldsymbol{\mu}_{2,k} \right) , \tag{6.54}$$

where

$$\mathbf{M}^* = \sum_{k=1}^{K} \lambda_k^* \beta_k (\mathbf{S}_k + \hat{\boldsymbol{\mu}}_k \hat{\boldsymbol{\mu}}_k^{\mathrm{T}}) \; , \tag{6.55}$$

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{S}_{1,k} & -\mathbf{S}_{12,k} \\ -\mathbf{S}_{12,k}^{\mathrm{T}} & \mathbf{S}_{2,k} \end{bmatrix} , \tag{6.56}$$

$$\hat{\boldsymbol{\mu}}_k = \begin{bmatrix} \boldsymbol{\mu}_{1,k} - \frac{\sum_{k'=1}^{K} \lambda_{k'}^* \beta_{k'} \boldsymbol{\mu}_{1,k'}}{\sum_{k'=1}^{K} \lambda_{k'}^* \beta_{k'}} \\ -\boldsymbol{\mu}_{2,k} + \frac{\sum_{k'=1}^{K} \lambda_{k'}^* \beta_{k'} \boldsymbol{\mu}_{2,k'}}{\sum_{k'=1}^{K} \lambda_{k'}^* \beta_{k'}} \end{bmatrix} . \tag{6.57}$$

Here $\boldsymbol{\lambda}^* = (\lambda_1^*, \cdots, \lambda_K^*)$ *is the solution to the convex minimization problem*

$$\min_{\boldsymbol{\lambda} \in \Lambda} \ \mathbf{1}_{2P}^{\mathrm{T}} \mathbf{M}^{-1} \mathbf{1}_{2P} \tag{6.58}$$

where

$$\mathbf{M} = \sum_{k=1}^{K} \lambda_k \beta_k (\mathbf{S}_k + \hat{\boldsymbol{\mu}}_k \hat{\boldsymbol{\mu}}_k^{\mathrm{T}}) \tag{6.59}$$

and the set Λ *is as defined in (6.49). The resultant optimum estimator* $\hat{\delta}^*$ *has*

$$\mathrm{WMaxMSE}(\hat{\delta}^*) = [\mathbf{1}_{2P}^{\mathrm{T}} (\mathbf{M}^*)^{-1} \mathbf{1}_{2P}]^{-1} \ , \tag{6.60}$$

$$\mathrm{MSE}_k(\hat{\delta}^*) = \frac{\mathbf{1}_{2P}^{\mathrm{T}} (\mathbf{M}^*)^{-1} (\mathbf{S}_k + \hat{\boldsymbol{\mu}}_k \hat{\boldsymbol{\mu}}_k^{\mathrm{T}}) (\mathbf{M}^*)^{-1} \mathbf{1}_{2P}}{[\mathbf{1}_{2P}^{\mathrm{T}} (\mathbf{M}^*)^{-1} \mathbf{1}_{2P}]^2} \tag{6.61}$$

for $k = 1, \cdots, K$.

Corollary 5. *(**WMaxMSE-optimum L-estimator under S-model**) Under the S-model, the values of* $\mathbf{c}_1$, $\mathbf{c}_2$ *and* η *that minimize* $\mathrm{WMaxMSE}(\hat{\delta})$ *for an estimator* $\hat{\delta}$ *of* δ *of the form*

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}} \downarrow \mathbf{y}_1 \downarrow - \mathbf{c}_2^{\mathrm{T}} \downarrow \mathbf{y}_2 \downarrow + \eta \tag{6.62}$$

given the constraint of constant bias are

$$\begin{bmatrix} \mathbf{c}_1^* \\ \mathbf{c}_2^* \end{bmatrix} = (\mathbf{M}^*)^{-1} \mathbf{G}^{\mathrm{T}} (\mathbf{G} (\mathbf{M}^*)^{-1} \mathbf{G}^{\mathrm{T}})^{-1} \mathbf{s} \ , \tag{6.63}$$

$$\eta^* = - \sum_{k=1}^{K} \lambda_k^* \beta_k \left((\mathbf{c}_1^*)^{\mathrm{T}} \boldsymbol{\mu}_{1,k} - (\mathbf{c}_2^*)^{\mathrm{T}} \boldsymbol{\mu}_{2,k} \right) \ , \tag{6.64}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \\ \mathbf{1}_P^{\mathrm{T}} & -\mathbf{1}_P^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \ , \tag{6.65}$$

and $\mathbf{M}^*$ *is as defined in (6.55). Here* $(\lambda_1^*, \cdots, \lambda_K^*)$ *is the solution to the concave maximization*

problem

$$\max_{\boldsymbol{\lambda}\in\Lambda} \ \mathbf{s}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.66}$$

where $\mathbf{M}$ *is as defined in (6.59), and the set* Λ *is as defined in (6.49). The resultant optimum estimator* $\hat{\delta}^*$ *has*

$$\mathrm{WMaxMSE}_k(\hat{\delta}^*) = \mathbf{s}^{\mathrm{T}}(\mathbf{G}(\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.67}$$

$$\begin{aligned}\mathrm{MSE}_k(\hat{\delta}^*) = \mathbf{s}^{\mathrm{T}}&(\mathbf{G}(\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{G}(\mathbf{M}^*)^{-1} \\ &\cdot (\mathbf{S}_k + \hat{\boldsymbol{\mu}}_k\hat{\boldsymbol{\mu}}_k^{\mathrm{T}})(\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}(\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s}\end{aligned} \tag{6.68}$$

for $k = 1, \cdots, K$.

Note that the optimization problems that need to be solved to determine $(\lambda_1^*, \cdots, \lambda_K^*)$ in (6.58) and (6.66) do not permit closed form solutions, but are respectively convex and concave. Hence, gradient descent techniques can be used to rapidly find globally optimum solutions to these problems. In the results section of this chapter, we utilized the `fmincon()` routine in MATLAB to solve these optimization problems. Also note that while we only consider a finite number of network scenarios in Corollaries 4 and 5, they can be applied to a continuous family of network scenarios by sampling to obtain a finite number of network scenarios.

6.5 L-estimators that use past observation windows to improve performance

Recall that the K-model and S-model assume that only a single observation window is available, where an observation window is defined as a set of P consecutive observation pairs over

which the fixed delays and phase offsets are constant. In certain scenarios, past observation windows that contain phase offsets distinct from that of the current observation window can be used to improve estimation performance. To demonstrate this claim, we define two new observation models that assume the availability of information from past observation windows, and derive optimum L-estimators under these models.

1. *Extended K-model*: Here we consider an extension to the K-model where in addition to the current observation window, we also have B past observation windows available. We assume that past observation windows contain different phase offsets, but have the same queuing delay distribution as the current observation window. Denote observations from the current window as

$$\mathbf{y}_1 = \delta \mathbf{1}_P + \mathbf{w}_1 \tag{6.69}$$

$$\mathbf{y}_2 = -\delta \mathbf{1}_P + \mathbf{w}_2 \tag{6.70}$$

and observations from the past window as

$$\mathbf{y}_1^{(i)} = \delta_i \mathbf{1}_P + \mathbf{w}_1^{(i)} \tag{6.71}$$

$$\mathbf{y}_2^{(i)} = -\delta_i \mathbf{1}_P + \mathbf{w}_2^{(i)} \tag{6.72}$$

for $i = 1, \cdots, B$. Here δ_i represents the phase offset in the i^{th} past observation window.

2. *Extended S-model*: Here we consider a similar extension to the S-model where B additional past observation windows are available. We assume that past observation windows contain different phase offsets, but have the same fixed delay and queuing delay distribution as the current observation window. We denote observations from the

current window as

$$\mathbf{y}_1 = d\mathbf{1}_P + \delta\mathbf{1}_P + \mathbf{w}_1 \tag{6.73}$$

$$\mathbf{y}_2 = d\mathbf{1}_P - \delta\mathbf{1}_P + \mathbf{w}_2 \tag{6.74}$$

and observations from the past window as

$$\mathbf{y}_1^{(i)} = d\mathbf{1}_P + \delta_i\mathbf{1}_P + \mathbf{w}_1^{(i)} \tag{6.75}$$

$$\mathbf{y}_2^{(i)} = d\mathbf{1}_P - \delta_i\mathbf{1}_P + \mathbf{w}_2^{(i)} \tag{6.76}$$

for $i = 1, \cdots, B$. Here δ_i represents the phase offset in the i^{th} past observation window.

Now consider the problem of designing L-estimators to minimize the MSE under a single network scenario, similar to the problems we studied in Section 6.3. Under such an optimality criterion, it is easy to see that the optimum L-estimator under the extended K-model would simply discard information from past observation windows, since they contain no information relevant to the estimation of the phase offset δ of the current observation window. However, under the extended S-model, past observation windows contain information about the nuisance parameter d, the knowledge of which could help form a better estimate of δ. Based on this reasoning, we consider the design of L-estimators to minimize the MSE under extended S-model, and obtain the result stated in the following corollary (proof provided in the appendix).

Corollary 6 (**Optimum L-estimator for extended S-model**). *Given the extended S-model, assume for simplicity that the queuing delays in different windows are mutually inde-*

pendent and identically distributed. Then the MSE of an estimator of δ of the form

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}} \lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}} \lfloor \mathbf{y}_2 \rfloor + \sum_{i=1}^{B} \left[\tilde{\mathbf{c}}_{1,i}^{\mathrm{T}} \lfloor \mathbf{y}_1^{(i)} \rfloor - \tilde{\mathbf{c}}_{2,i}^{\mathrm{T}} \lfloor \mathbf{y}_2^{(i)} \rfloor \right] + \eta \tag{6.77}$$

is minimized under the constraint of constant bias when $(\mathbf{c}_1, \mathbf{c}_2, \eta) = (\mathbf{c}_1^*, \mathbf{c}_2^*, \eta^*)$*, and* $(\tilde{\mathbf{c}}_{1,i}, \tilde{\mathbf{c}}_{2,i}) = (\tilde{\mathbf{c}}_1^*, \tilde{\mathbf{c}}_2^*)$ *for all* $i = 1, \cdots, B$*, where*

$$\left[(\mathbf{c}_1^*)^{\mathrm{T}} \ (\mathbf{c}_2^*)^{\mathrm{T}} \ (\tilde{\mathbf{c}}_1^*)^{\mathrm{T}} \ (\tilde{\mathbf{c}}_2^*)^{\mathrm{T}} \right]^{\mathrm{T}} = \tilde{\mathbf{S}}^{-1} \mathbf{G}^{\mathrm{T}} (\mathbf{G} \tilde{\mathbf{S}}^{-1} \mathbf{G}^{\mathrm{T}})^{-1} \mathbf{s} \ , \tag{6.78}$$

$$\eta^* = (\mathbf{c}_2^* + B\tilde{\mathbf{c}}_2^*)^{\mathrm{T}} \boldsymbol{\mu}_2 - (\mathbf{c}_1^* + B\tilde{\mathbf{c}}_1^*)^{\mathrm{T}} \boldsymbol{\mu}_1 \ , \tag{6.79}$$

and

$$\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & B\mathbf{S} \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & -\mathbf{S}_{12} \\ -\mathbf{S}_{12}^{\mathrm{T}} & \mathbf{S}_2 \end{bmatrix} \tag{6.80}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{1}_P & \mathbf{1}_P & \mathbf{0}_P & \mathbf{0}_P \\ \mathbf{0}_P & \mathbf{0}_P & \mathbf{1}_P & \mathbf{1}_P \\ \mathbf{1}_P & -\mathbf{1}_P & B\mathbf{1}_P & -B\mathbf{1}_P \end{bmatrix} \ , \qquad \mathbf{s} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \ . \tag{6.81}$$

The resultant optimum estimator

$$\begin{aligned} \hat{\delta}^* = {} & (\mathbf{c}_1^*)^{\mathrm{T}} \lfloor \mathbf{y}_1 \rfloor - (\mathbf{c}_2^*)^{\mathrm{T}} \lfloor \mathbf{y}_2 \rfloor \\ & + (\tilde{\mathbf{c}}_1^*)^{\mathrm{T}} \left[\sum_{i=1}^{B} \lfloor \mathbf{y}_1^{(i)} \rfloor \right] - \left[(\tilde{\mathbf{c}}_2^*)^{\mathrm{T}} \sum_{i=1}^{B} \lfloor \mathbf{y}_2^{(i)} \rfloor \right] + \eta^* \end{aligned} \tag{6.82}$$

has an MSE

$$\mathrm{MSE}(\hat{\delta}^*) = \mathbf{s}^{\mathrm{T}} (\mathbf{G} \tilde{\mathbf{S}}^{-1} \mathbf{G}^{\mathrm{T}})^{-1} \mathbf{s} \tag{6.83}$$

Next we consider the problem of designing L-estimators to minimize the weighted max-

imum MSE when K possible network scenarios can occur, as we studied in Section 6.4. Assume that the queuing delays in each observation window are independent of that of other observation windows, and that the queuing delays in all windows arise from a common distribution (though we do not have prior knowledge about which of the K possible distribution have occurred). Here, under the extended K-model, past windows contain information that could be used to better deduce which of the K network scenarios have occurred, and hence help improve the estimation of δ. Under the extended S-model, past windows contain information about the nuisance parameter d, as well as information of which network scenario has occurred. Based on this reasoning, we consider the design of L-estimator to minimize the weighted maximum MSE under the extended K-model and the extended S-model, and obtain the result stated in the following corollary.

Corollary 7. *(**WMaxMSE-optimum L-estimator under extended K- and S- models**) Under the either the extended K-model or extended S-model, assume for simplicity that the queuing delays in different windows are mutually independent and identically distributed. Then an estimator of δ of the form*

$$\hat{\delta} = \mathbf{c}_1^{\mathrm{T}} \lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}} \lfloor \mathbf{y}_2 \rfloor + \sum_{i=1}^{B} \left[\tilde{\mathbf{c}}_{1,i}^{\mathrm{T}} \lfloor \mathbf{y}_1^{(i)} \rfloor - \tilde{\mathbf{c}}_{2,i}^{\mathrm{T}} \lfloor \mathbf{y}_2^{(i)} \rfloor \right] + \eta \tag{6.84}$$

minimizes $\mathrm{WMaxMSE}(\hat{\delta})$ *under the constraint of constant bias when* $(\mathbf{c}_1, \mathbf{c}_2, \eta) = (\mathbf{c}_1^*, \mathbf{c}_2^*, \eta^*)$,

and $(\tilde{\mathbf{c}}_{1,i}, \tilde{\mathbf{c}}_{2,i}) = (\tilde{\mathbf{c}}_1^*, \tilde{\mathbf{c}}_2^*)$ *for all* $i = 1, \cdots, B$, *where*

$$\begin{aligned}
&\left[(\mathbf{c}_1^*)^{\mathrm{T}}\ (\mathbf{c}_2^*)^{\mathrm{T}}\ (\tilde{\mathbf{c}}_1^*)^{\mathrm{T}}\ (\tilde{\mathbf{c}}_2^*)^{\mathrm{T}}\right]^{\mathrm{T}} \\
&\qquad = (\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}(\mathbf{M}^*)^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s}, \qquad (6.85)\\
&\eta^* = -\left(\sum_{k=1}^{K}\lambda_k^*\beta_k\right)^{-1}\left\{\sum_{k=1}^{K}\lambda_k^*\beta_k\left[\boldsymbol{\mu}_{1,k}^{\mathrm{T}}\Big(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\Big)\right.\right. \\
&\qquad\qquad \left.\left. - \boldsymbol{\mu}_{2,k}^{\mathrm{T}}\Big(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\Big)\right]\right\}, \qquad (6.86)\\
&\mathbf{M}^* = \sum_{k=1}^{K}\lambda_k^*\beta_k\left\{\begin{bmatrix}\mathbf{S}_k & \mathbf{0}\\ \mathbf{0} & B\mathbf{S}_k\end{bmatrix} + \begin{bmatrix}\hat{\boldsymbol{\mu}}_k\\ B\hat{\boldsymbol{\mu}}_k\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\mu}}_k\\ B\hat{\boldsymbol{\mu}}_k\end{bmatrix}^{\mathrm{T}}\right\}, \qquad (6.87)\\
&\mathbf{S}_k = \begin{bmatrix}\mathbf{S}_{1,k} & -2\mathbf{S}_{12,k}\\ -2\mathbf{S}_{12,k}^{\mathrm{T}} & \mathbf{S}_{2,k}\end{bmatrix},\ \hat{\boldsymbol{\mu}}_k = \begin{bmatrix}\hat{\boldsymbol{\mu}}_{1,k}\\ \hat{\boldsymbol{\mu}}_{2,k}\end{bmatrix}, \qquad (6.88)\\
&\hat{\boldsymbol{\mu}}_{1,k} = \boldsymbol{\mu}_{1,k} - \frac{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}\boldsymbol{\mu}_{1,k'}}{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}}, \qquad (6.89)\\
&\hat{\boldsymbol{\mu}}_{2,k} = \boldsymbol{\mu}_{2,k} - \frac{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}\boldsymbol{\mu}_{2,k'}}{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}}, \qquad (6.90)
\end{aligned}$$

and $\boldsymbol{\lambda}^* = (\lambda_1^*, \cdots, \lambda_K^*)$ *is the solution to the convex maximization problem*

$$\max_{\boldsymbol{\lambda}\in\Lambda}\ \mathbf{s}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \qquad (6.91)$$

Here we have

$$\mathbf{G} = \begin{cases} \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} \\ \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \end{bmatrix} & \textit{For ext. K-model} \\ \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} \\ \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \\ \mathbf{1}_P^{\mathrm{T}} & -\mathbf{1}_P^{\mathrm{T}} & B\mathbf{1}_P^{\mathrm{T}} & -B\mathbf{1}_P^{\mathrm{T}} \end{bmatrix} & \textit{For ext. S-model} \end{cases} \tag{6.92}$$

$$\mathbf{s} = \begin{cases} \begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathrm{T}} & \textit{For ext. K-model} \\ \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^{\mathrm{T}} & \textit{For ext. S-model} \end{cases} \tag{6.93}$$

6.6 Simulation Results

We now compare the performance of the newly proposed L-estimators versus various existing POE techniques. In order to generate queuing delay distributions, we considered a few network scenarios motivated by the ITU-T recommendation G.8261 [45]. Specifically, we consider a Gigabit ethernet network consisting a cascade of 10 switches between the master and slave nodes. Each switch is assumed to be a store-and-forward switch, which implements strict priority queuing. We assumed cross traffic flows of background traffic to be present in this network. In such traffic flows [53][45], fresh background traffic packets are injected at each node along the master-slave path, and these packets exit the master-slave path at the subsequent node (see 4-switch example in Fig. 6.2). The arrival times and sizes of the packets injected at each switch were assumed to be statistically independent of that of packets injected at other switches. With regard to the distribution of packet sizes in background traffic, we consider Traffic Models 1 (TM1) and 2 (TM2) from the ITU-T recommendation G.8261 [45]

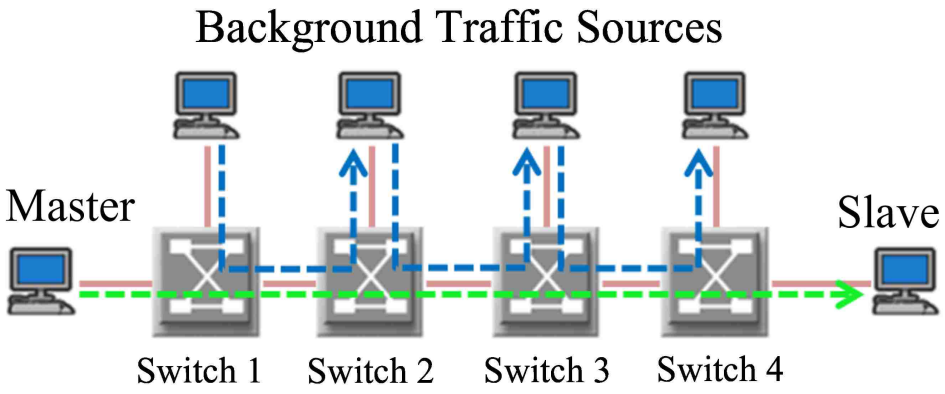


Figure 6.2: Example of a four switch network with cross and traffic flows. Red lines indicate network links, blue lines indicate the direction of background traffic flows, and green line represents the direction of synchronization traffic flows.

Traf. Model	**Packet Sizes (Bytes)**	**% of Load**
TM1	$\{64, 576, 1518\}$	$\{80\%, 5\%, 15\%\}$
TM2	$\{64, 576, 1518\}$	$\{30\%, 10\%, 60\%\}$

Table 6.1: Models for composition of background traffic packets

for cross traffic flows, as specified in Table 6.1.

We assume that the interarrival times between packets in all background traffic flows follow exponential distributions. We refer to the percentage of the link capacity consumed by background traffic as the *load*. In order to achieve a particular load, we accordingly set the rate parameter of each exponential distribution. For simplicity, we assumed in our simulations that the queuing delays on the forward path are statistically independent of the queuing delays on the reverse path.

In order to evaluate the performance of various L-estimators, the mean vectors and covariance matrices of the order statistics of queuing delays were first obtained using low-level queue simulations. Then the MSE of the L-estimators presented in Corollaries 2 - 7 were computed using these mean vectors and covariance matrices. In order to compute the

MSE of the minimax optimum estimator of [26] under the K-model and S-model, the pdfs of the queuing delays were obtained empirically from low-level queue simulations, and Riemann sums were use to evaluate estimator performance. The standard deviation of the estimation error (square root of the mean squared estimation error) for various estimators under TM1 and TM2 for various loads are plotted in Figs. 6.3 - 6.6. Note that Figs. 6.3 and 6.4 consider scenarios where the distribution of queuing delays in the forward and reverse paths are symmetrical while, Figs. 6.5 and 6.6 consider scenarios where they are asymmetrical.

In order to facilitate comparisons against a typical synchronization requirement of 1.5 μs of synchronization accuracy that arises in LTE networks [64], the estimation error standard deviation required so that the absolute estimation error lies under 1.5 μs with a 5-sigma level of certainty is also plotted over the curves. Here the 5-sigma level of certainty implies that on average, only about 6 out of 10^6 estimates will have absolute estimation error that exceeds 1.25 μs. Some key observations we can make from the figures are as follows:

1. *Performance gap between L-estimators and the minimax estimators of [26]*: We observe that the optimum L-estimators of Corollaries 1 and 2 have a MSE performance very close to the minimax optimum estimators under all the network scenarios considered. This indicates that the loss in POE performance when we are restricted to only using L-estimators is quite negligible.

2. *Performance relative to conventional estimators*: While the sample minimum estimator performs near optimally at low network loads, at high loads conventional estimators (sample minimum, mean, median and maximum) have a MSE that is significantly larger that of the optimized L-estimators.

3. *Performance differences between asymmetric and symmetric background traffic conditions*:

The performance gains obtained by using K-model model estimator of Corollary 1 versus the S-model estimator of Corollary 2 are significantly larger under asymmetric network conditions. This performance gap can be bridged by using the extended S-model estimators of Corollary 6.

4. *Performance difference between using* MaxMSE *and* MaxEFF *as an optimization metric*: We observe that when MaxMSE is used as an optimization metric for the estimators of Corollaries 4, 5 or 7 , the MSE curves tend to be flat across the range of loads considered. In contrast, when MaxEFF is used as the optimization metric, a little estimation performance is ceded at high loads in order to gain significantly improved estimation performance at low loads.

5. *Performance improvements under extended K- and S- models*: We observe that by utilizing a sufficient number past observation windows, the MSE penalty associated with network model uncertainty can be effectively eliminated under both traffic models considered. For example, we observe that the MaxEFF-Optimum L-estimator under the extended S-model (described in Corollary 7), achieves a MSE performance closely mirroring that of the MSE-Optimum L-estimator under the S-model (described in Corollary 3).

From the perspective of complexity, it is easy to see that when Riemann sums are used to evaluate the minimax estimator of [26], $\mathcal{O}(PN_B)$ multiplications and $\mathcal{O}(N_B)$ additions are required per estimate, where N_B denotes the number of Riemann sum bins utilized. On the other hand, the L-estimators of Corollaries 1-6 have a much lower run-time computational complexity. Specifically, it is well known that all the order statistics of P observations can be obtained using $\mathcal{O}(P\log(P))$ comparisons via the Quicksort algorithm [65]. Computing the weighted summation of P order statistics requires $\mathcal{O}(P)$ multiplications and $\mathcal{O}(P)$ additions.

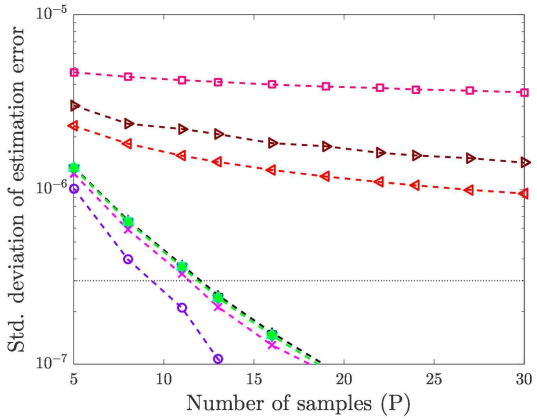
Further, it can be easily shown the estimators of Corollaries 6 and 7 admit a sliding filter implementation which requires $\mathcal{O}(P\log(P))$ comparisons, $\mathcal{O}(P)$ multiplications and $\mathcal{O}(P)$ additions, independent of the number B of past observation blocks considered. This makes L-estimators very attractive from a practical perspective.

6.7 Summary

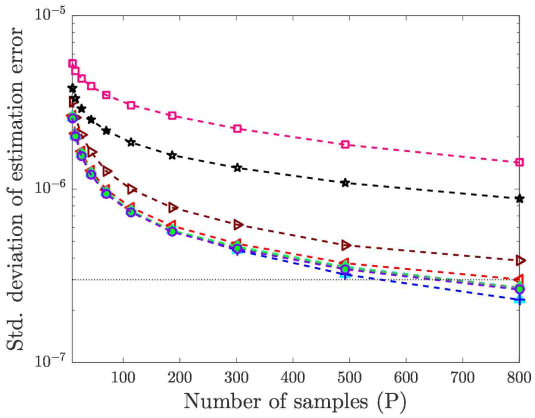
In this chapter, we solve the problem of designing optimum L-estimators of phase offset under various novel criteria of optimality. Two observation models were considered. L-estimators that minimize the MSE for a known network scenario, L-estimators that minimize the worst-cast MSE under network model uncertainly and L-estimator that utilizes past information to improve estimation performance were derived. The proposed estimators have a low computational complexity and offer many performance benefits relative to both conventional estimators and the minimax estimators of [26]. Of all the results described in this chapter, the authors believe that Corollary 7 offers the greatest practical utility, describing estimators that appear to achieve near-optimum performance even when the exact network model is not known.

6.8 Appendix

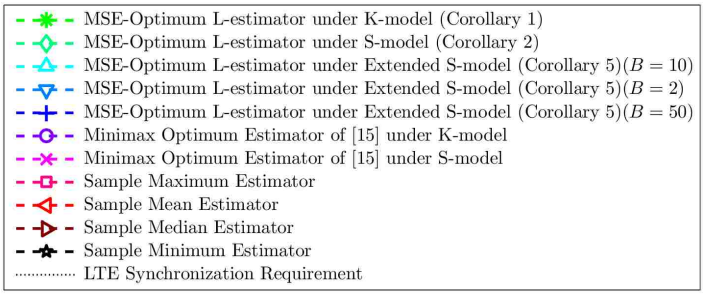
Proof of Theorem 7. This theorem can be obtained as a direct consequence of the Gauss-Markov theorem [66]. For the convenience of the readers, we also provide a short proof herein. It is easy to show that the function $\mathbf{z}^{\mathrm{T}}\mathbf{H}\mathbf{z}$ is convex in $\mathbf{z}$ since $\mathbf{H}$ is positive definite, and that the set of values of $\mathbf{z}$ that satisfy $\mathbf{G}\mathbf{z} = \mathbf{s}$ is a convex set. Hence, (6.25) has a unique local minimum that is also the global minimum. This minimum can be easily obtained via



(a) Forward path: 20% Load (TM2), Reverse path: 20% Load (TM2)

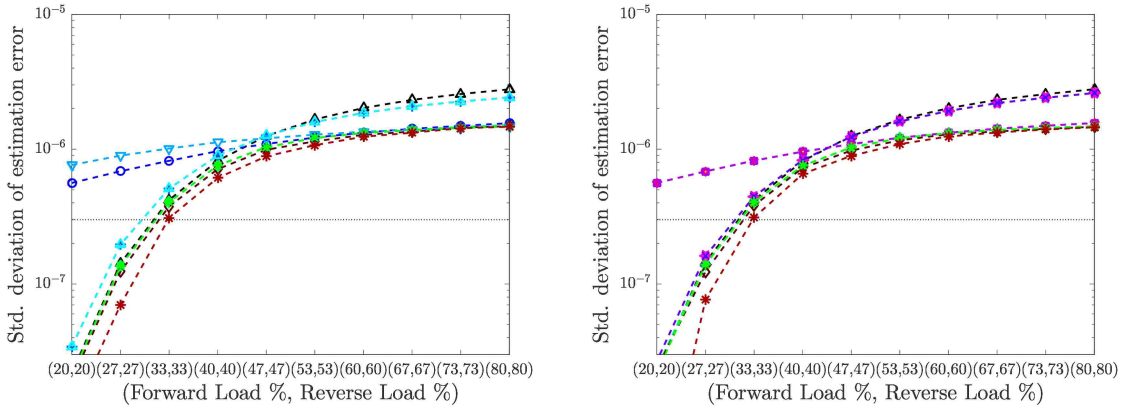


(b) Forward path: 80% Load (TM2), Reverse path: 80% Load (TM2)



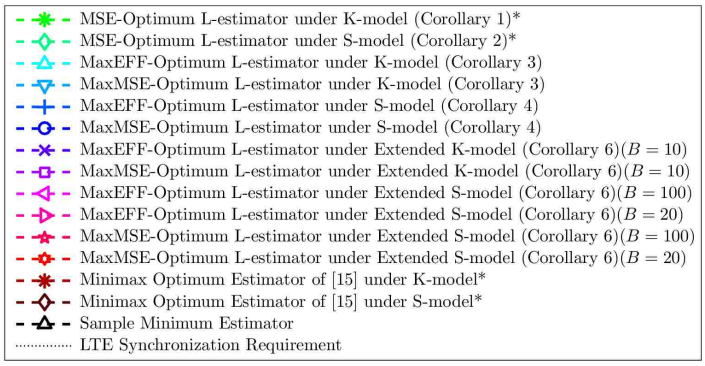
(c) Legend

Figure 6.3: Performance of various estimators plotted versus the number of two way message exchanges P, under symmetric traffic conditions.



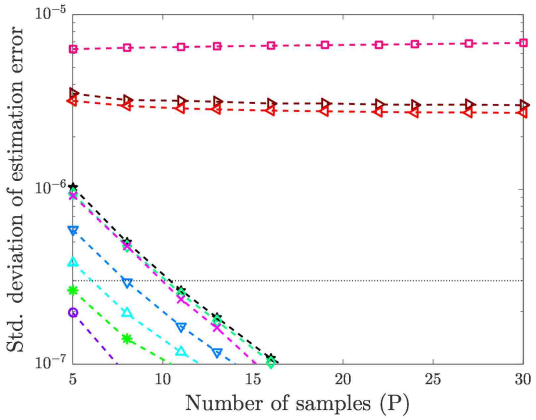
(a) Estimators of Corollaries 1, 2 and 5

(b) Estimators of Corollaries 3, 4 and 6

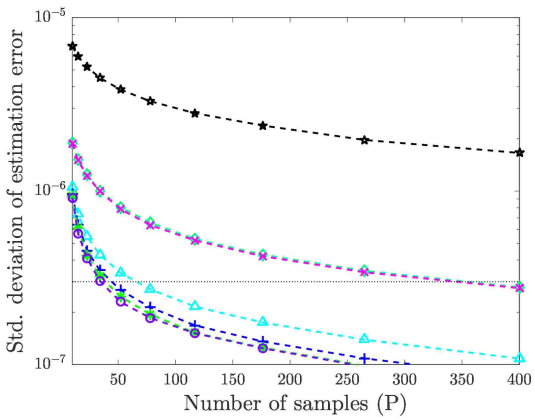


(c) Legend

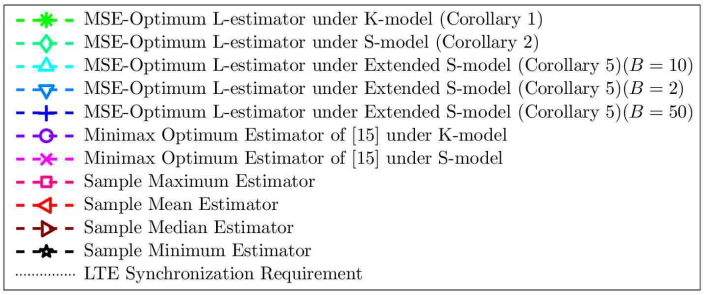
Figure 6.4: Performance of various estimators given $P = 30$ two way message exchanges. Background traffic on the forward and reverse links are assumed to be symmetrically distributed per TM2. $K = 10$ network scenarios are considered, obtained by stepping the forward and reverse path loads uniformly between 20% and 80%.



(a) Forward path: 20% Load (TM1), Reverse path: 80% Load (TM2)

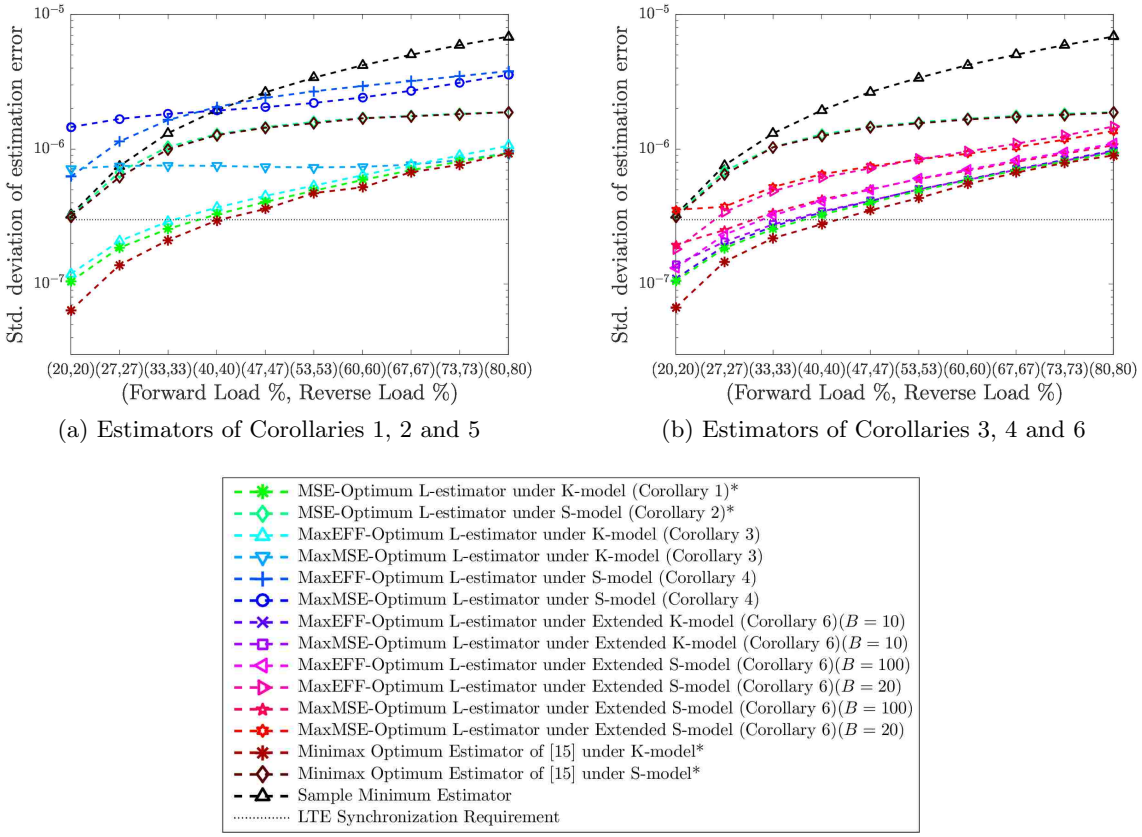


(b) Forward path: 80% Load (TM1), Reverse path: 80% Load (TM2)



(c) Legend

Figure 6.5: Performance of various estimators plotted versus the number of two way message exchanges P, under asymmetric traffic conditions.



(a) Estimators of Corollaries 1, 2 and 5

(b) Estimators of Corollaries 3, 4 and 6

(c) Legend

Figure 6.6: Performance of various estimators given $P = 10$ two way message exchanges, under an asymmetric traffic scenario. Background traffic on the forward link is assumed to be distributed per TM1 and traffic on the reverse link is assumed to be distributed per TM2. $K = 10$ network scenarios are considered, obtained by stepping the forward and reverse path loads uniformly between 20% and 80%. Estimators annotated with a '*' in the legend are assumed to have exact knowledge of which queuing delay distribution has occurred, while the other estimators only have knowledge of the set of 10 possible queuing delay distributions.

the method of Lagrangian multipliers. To this end, we construct the Lagrangian

$$\Omega = \mathbf{z}^{\mathrm{T}}\mathbf{H}\mathbf{z} - \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{G}\mathbf{z} - \mathbf{s}) \tag{6.94}$$

and differentiate it with respect to $\mathbf{z}$ and $\boldsymbol{\lambda}$ to obtain the equations

$$2\mathbf{H}\mathbf{z} - \mathbf{G}^{\mathrm{T}}\boldsymbol{\lambda} = 0, \tag{6.95}$$

$$\mathbf{G}\mathbf{z} - \mathbf{s} = 0 \tag{6.96}$$

which can be solved to obtain

$$\boldsymbol{\lambda}^{*} = 2(\mathbf{G}\mathbf{H}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s}, \tag{6.97}$$

$$\mathbf{z}^{*} = \mathbf{H}^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{H}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.98}$$

hence proving the theorem. □

Proof of Corollary 2. Under the K-model, we have

$$\lfloor \mathbf{y}_1 \rfloor = \delta \mathbf{1}_P + \lfloor \mathbf{w}_1 \rfloor \,, \quad \lfloor \mathbf{y}_2 \rfloor = -\delta \mathbf{1}_P + \lfloor \mathbf{w}_2 \rfloor \tag{6.99}$$

Hence, for the estimator of (6.27), we have

$$\mathrm{Bias}(\hat{\delta}) = \mathrm{E}\left[(\mathbf{c}_1^{\mathrm{T}}\lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\mathrm{T}}\lfloor \mathbf{y}_2 \rfloor + \eta) \ - \delta\right] \tag{6.100}$$

$$= \delta[(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P - 1] + \mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_2 + \eta \,. \tag{6.101}$$

From (6.101), we see that in order for $\hat{\delta}$ to have constant bias, we require

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1 \tag{6.102}$$

Further, assuming $\hat{\delta}$ satisfies (6.102), its variance has the form

$$\mathrm{var}(\hat{\delta}) = \mathrm{E}\left[(\mathbf{c}_1^{\mathrm{T}}(\lfloor \mathbf{w}_1 \rfloor - \boldsymbol{\mu}_1) - \mathbf{c}_2^{\mathrm{T}}(\lfloor \mathbf{w}_2 \rfloor - \boldsymbol{\mu}_2))^2\right] \tag{6.103}$$

$$= \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 \tag{6.104}$$

and the MSE is

$$\mathrm{MSE}(\hat{\delta}) = \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 + (\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_2 + \eta)^2 \tag{6.105}$$

Under the constraint of constant bias, both the variance and MSE are hence constant with respect to δ. Further, for any choice of $\mathbf{c}_1$ and $\mathbf{c}_2$, $\mathrm{MSE}(\hat{\delta})$ can be minimized by making $\hat{\delta}$ unbiased by setting $\eta = -(\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_2)$. The residual problem of choosing $\mathbf{c}_1$ and $\mathbf{c}_2$ to minimize the MSE under the constraint of constant bias can be stated as

$$\begin{aligned} \min_{\mathbf{c}_1,\mathbf{c}_2} \quad & \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 \\ \text{s.t.} \quad & (\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1 \end{aligned} \tag{6.106}$$

or equivalently as

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{S}\mathbf{c} \\ \text{s.t.} \quad & \mathbf{1}_{2P}^{\mathrm{T}}\mathbf{c} = 1 \end{aligned} \tag{6.107}$$

where $\mathbf{S}$ and $\mathbf{c}$ are defined in (6.24). Applying Theorem 7 to (6.107), we obtain the solution specified in (6.28). Further, when $\lfloor \mathbf{w}_1 \rfloor$ and $\lfloor \mathbf{w}_2 \rfloor$ are uncorrelated, we have $\mathbf{S}_{12} = \mathbf{0}$ and

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{S}_1^{-1} & 0 \\ 0 & \mathbf{S}_2^{-1} \end{bmatrix} \tag{6.108}$$

By substituting (6.108) in (6.28) and (6.30), the results in (6.31) and (6.32) are obtained. □

Proof of Corollary 3. Under the S-model, we have

$$\lfloor \mathbf{y}_1 \rfloor = d\mathbf{1}_P + \delta\mathbf{1}_P + \lfloor \mathbf{w}_1 \rfloor \ , \tag{6.109}$$

$$\lfloor \mathbf{y}_2 \rfloor = d\mathbf{1}_P - \delta\mathbf{1}_P + \lfloor \mathbf{w}_2 \rfloor \tag{6.110}$$

Hence, for the estimator of (6.33), we have

$$\text{Bias}(\hat{\delta}) = \text{E}\left[(\mathbf{c}_1^{\text{T}}\lfloor \mathbf{y}_1 \rfloor - \mathbf{c}_2^{\text{T}}\lfloor \mathbf{y}_2 \rfloor + \eta) \ - \delta\right] \tag{6.111}$$

$$= d(\mathbf{c}_1 - \mathbf{c}_2)^{\text{T}}\mathbf{1}_P + \delta[(\mathbf{c}_1 + \mathbf{c}_2)^{\text{T}}\mathbf{1}_P - 1] + \mathbf{c}_1^{\text{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\text{T}}\boldsymbol{\mu}_2 + \eta \ . \tag{6.112}$$

From (6.112), we see that in order for $\hat{\delta}$ to have constant bias with respect to both d and δ, we require

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\text{T}}\mathbf{1}_P = 1, \qquad (\mathbf{c}_1 - \mathbf{c}_2)^{\text{T}}\mathbf{1}_P = 0 \tag{6.113}$$

Further, assuming $\hat{\delta}$ satisfies (6.113), its variance has the form

$$\text{var}(\hat{\delta}) = \text{E}\left[(\mathbf{c}_1^{\text{T}}(\lfloor \mathbf{w}_1 \rfloor - \boldsymbol{\mu}_1) + \mathbf{c}_2^{\text{T}}(\lfloor \mathbf{w}_2 \rfloor - \boldsymbol{\mu}_2))^2\right] \tag{6.114}$$

$$= \mathbf{c}_1^{\text{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\text{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\text{T}}\mathbf{S}_{12}\mathbf{c}_2 \tag{6.115}$$

and its MSE is

$$\text{MSE}(\hat{\delta}) = \mathbf{c}_1^{\text{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\text{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\text{T}}\mathbf{S}_{12}\mathbf{c}_2 + (\mathbf{c}_1^{\text{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\text{T}}\boldsymbol{\mu}_2 + \eta)^2 \tag{6.116}$$

Under the constraint of constant bias, both the variance and MSE are hence constant with respect to both δ and d. Further, for any choice of $\mathbf{c}_1$ and $\mathbf{c}_2$, $\text{MSE}(\hat{\delta})$ can be minimized by making $\hat{\delta}$ unbiased by setting $\eta = -(\mathbf{c}_1^{\text{T}}\boldsymbol{\mu}_1 - \mathbf{c}_2^{\text{T}}\boldsymbol{\mu}_2)$. The residual problem of choosing $\mathbf{c}_1$ and

$\mathbf{c}_2$ to minimize the MSE under the constraint of constant bias can be stated as

$$\begin{aligned} \min_{\mathbf{c}_1,\mathbf{c}_2} \quad & \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 \\ \text{s.t.} \quad & \begin{cases} (\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1, \\ (\mathbf{c}_1 - \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 0 \end{cases} \end{aligned} \tag{6.117}$$

or equivalently as

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{S}\mathbf{c} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{c} = \boldsymbol{\gamma} \end{aligned} \tag{6.118}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \\ \mathbf{1}_P^{\mathrm{T}} & -\mathbf{1}_P^{\mathrm{T}} \end{bmatrix}, \qquad \boldsymbol{\gamma} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6.119}$$

Applying Theorem 7 to (6.118), we obtain the solution specified in (6.34). Further, when $\downharpoonleft\mathbf{w}_1\downharpoonright$ and $\downharpoonleft\mathbf{w}_2\downharpoonright$ are uncorrelated, we have $\mathbf{S}_{12} = \mathbf{0}$ and

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{S}_1^{-1} & 0 \\ 0 & \mathbf{S}_2^{-1} \end{bmatrix} \tag{6.120}$$

By substituting (6.120) in (6.34) and (6.37), the results in (6.38) and (6.39) are obtained. □

Proof of Corollary 4. Under the K-model, we have

$$\mathrm{Bias}_k(\hat{\delta}) = \delta[(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P - 1] + \mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_{2,k} + \eta \tag{6.121}$$

It is easy to see that the constant bias condition can be ensured under all K network scenarios by setting

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1 \tag{6.122}$$

Under this constraint, we have

$$\mathrm{MSE}_k(\hat{\delta}) = \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{1,k}\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_{2,k}\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 + (\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_{2,k} + \eta)^2 \tag{6.123}$$

Thus, the problem of minimizing $\mathrm{WMaxMSE}(\hat{\delta})$ under the constraint of constant bias can be stated as

$$\min_{\left[\begin{smallmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{smallmatrix}\right]\in\mathcal{C},\ \eta\in\mathbb{R}} \ \max_{\boldsymbol{\lambda}\in\Lambda} \ \sum_{k=1}^{K} \lambda_k\beta_k\mathrm{MSE}_k(\hat{\delta}) \tag{6.124}$$

where

$$\mathcal{C} = \left\{ \begin{bmatrix} \mathbf{c}_1 \\ \\ \mathbf{c}_2 \end{bmatrix} \in \mathbb{R}^{2P} \ \middle| \ (\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1 \right\} \tag{6.125}$$

It is easy to see that the sets $\mathcal{C}\times\mathbb{R}$ and Λ are both convex, and that Λ is bounded. Further, it is easy to show that the cost function in (6.124) satisfies the concave-convex property specified in Theorem 8. Hence, applying Theorem 8 to (6.124), we can restate it as

$$\max_{\boldsymbol{\lambda}\in\Lambda} \min_{\left[\begin{smallmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{smallmatrix}\right]\in\mathcal{C}} \ \min_{\eta\in\mathbb{R}} \ \sum_{k=1}^{K} \lambda_k\beta_k\mathrm{MSE}_k(\hat{\delta}) \tag{6.126}$$

For any fixed values of $\mathbf{c}_1$, $\mathbf{c}_2$ and $\boldsymbol{\lambda}$, the value of η that solves the innermost minimization problem in (6.126) can be obtained via differentiation as

$$\eta = \frac{-\sum_{k=1}^{K}\lambda_k\beta_k\left(\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_{2,k}\right)}{\sum_{k=1}^{K}\lambda_k\beta_k} \tag{6.127}$$

Substituting this optimum value of η in (6.126), we are left with the optimization problem

$$\max_{\boldsymbol{\lambda}\in\Lambda} \min_{\mathbf{c}\in\mathcal{C}} \ \mathbf{c}^{\mathrm{T}}\mathbf{M}\mathbf{c} \tag{6.128}$$

where $\mathbf{c} = [\mathbf{c}_1^{\mathrm{T}}\ \mathbf{c}_2^{\mathrm{T}}]^{\mathrm{T}}$, and

$$\mathbf{M} = \sum_{k=1}^{K} \lambda_k \beta_k (\mathbf{S}_k + \hat{\boldsymbol{\mu}}_k \hat{\boldsymbol{\mu}}_k^{\mathrm{T}}), \tag{6.129}$$

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{S}_{1,k} & -\mathbf{S}_{12,k} \\ -\mathbf{S}_{12,k}^{\mathrm{T}} & \mathbf{S}_{2,k} \end{bmatrix}, \tag{6.130}$$

$$\hat{\boldsymbol{\mu}}_k = \begin{bmatrix} \boldsymbol{\mu}_{1,k} - \frac{\sum_{k'=1}^{K} \lambda_{k'} \beta_{k'} \boldsymbol{\mu}_{1,k'}}{\sum_{k'=1}^{K} \lambda_{k'} \beta_{k'}} \\ -\boldsymbol{\mu}_{2,k} + \frac{\sum_{k'=1}^{K} \lambda_{k'} \beta_{k'} \boldsymbol{\mu}_{2,k'}}{\sum_{k'=1}^{K} \lambda_{k'} \beta_{k'}} \end{bmatrix}, \tag{6.131}$$

Using Theorem 7, the solution to the inner minimization in (6.128) is obtained as

$$\mathbf{c}^* = \frac{1}{\mathbf{1}_{2P}^{\mathrm{T}} \mathbf{M}^{-1} \mathbf{1}_{2P}} \mathbf{M}^{-1} \mathbf{1}_{2P} \tag{6.132}$$

and (6.128) reduces to the concave maximization problem

$$\max_{\boldsymbol{\lambda} \in \Lambda} (\mathbf{1}_{2P}^{\mathrm{T}} \mathbf{M}^{-1} \mathbf{1}_{2P})^{-1} \tag{6.133}$$

which is equivalent to (6.58). Further, the results in (6.61) and (6.60) can be obtained by substituting the optimum weights specified by (6.132) in (6.123). □

Proof of Corollary 5. Under the S-model, we have

$$\mathrm{Bias}_k(\hat{\delta}) = d[(\mathbf{c}_1 - \mathbf{c}_2)^{\mathrm{T}} \mathbf{1}_P] + \delta[(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}} \mathbf{1}_P - 1] \tag{6.134}$$

$$+ \mathbf{c}_1^{\mathrm{T}} \boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}} \boldsymbol{\mu}_{2,k} + \eta \tag{6.135}$$

It is easy to see that the constant bias condition can be ensured under all K network scenarios by setting

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}} \mathbf{1}_P = 1\ , \quad (\mathbf{c}_1 - \mathbf{c}_2)^{\mathrm{T}} \mathbf{1}_P = 0 \tag{6.136}$$

Under this constraint, we have

$$\mathrm{MSE}_k(\hat{\delta}) = \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{1,k}\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_{2,k}\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 + (\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_{2,k} + \eta)^2 \tag{6.137}$$

Thus, the problem of minimizing $\mathrm{WMaxMSE}(\hat{\delta})$ under the constraint of constant bias can be stated as

$$\min_{\left[\begin{smallmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{smallmatrix}\right]\in\mathcal{C},\ \eta\in\mathbb{R}} \ \max_{\boldsymbol{\lambda}\in\Lambda} \ \sum_{k=1}^{K} \lambda_k\beta_k \mathrm{MSE}_k(\hat{\delta}) \tag{6.138}$$

where

$$\mathcal{C} = \left\{ \begin{bmatrix}\mathbf{c}_1^{\mathrm{T}} & \mathbf{c}_2^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{2P} \ \middle| \ ((\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1) \wedge ((\mathbf{c}_1 - \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 0) \right\} \tag{6.139}$$

It is easy to see that the sets $\mathcal{C} \times \mathbb{R}$ and Λ are both convex, and that Λ is bounded. Hence, applying Theorem 8 to (6.138), we can restate it as

$$\max_{\boldsymbol{\lambda}\in\Lambda} \min_{\left[\begin{smallmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{smallmatrix}\right]\in\mathcal{C}} \ \min_{\eta\in\mathbb{R}} \mathrm{MSE}_k(\hat{\delta}) \tag{6.140}$$

For any fixed values of $\mathbf{c}_1$, $\mathbf{c}_2$ and $\boldsymbol{\lambda}$, the value of η that solves the inner minimization problem can be obtained via differentiation as

$$\eta = \frac{-\sum_{k=1}^{K} \lambda_k\beta_k \left(\mathbf{c}_1^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \mathbf{c}_2^{\mathrm{T}}\boldsymbol{\mu}_{2,k}\right)}{\sum_{k=1}^{K} \lambda_k\beta_k} \tag{6.141}$$

Substituting this optimum value of η in (6.140), we are left with the optimization problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\Lambda} \ \min_{\mathbf{c}} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{M}\mathbf{c} \\ \text{s.t.} \quad & \mathbf{G}\mathbf{c} = \mathbf{s} \end{aligned} \tag{6.142}$$

where $\mathbf{c} = [\mathbf{c}_1^{\mathrm{T}}\ \mathbf{c}_2^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{M}$ is as defined in (6.129), and

$$\mathbf{G} = \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \\ \mathbf{1}_P^{\mathrm{T}} & -\mathbf{1}_P^{\mathrm{T}} \end{bmatrix},\ \mathbf{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6.143}$$

Using Theorem 7, the solution to the inner minimization in (6.142) is obtained as

$$\mathbf{c}^* = \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.144}$$

and (6.142) reduces to the concave maximization problem

$$\max_{\boldsymbol{\lambda}\in\Lambda}\ \mathbf{s}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.145}$$

Further, the results in (6.68) and (6.67) can be obtained by substituting the optimum weights specified by (6.144) in (6.137). □

Proof of Corollary 6. Under the extended S-model, for the estimator of (6.77), we have

$$\begin{aligned} \mathrm{Bias}(\hat{\delta}) = d \left[\mathbf{c}_1 - \mathbf{c}_2 + \sum_{i=1}^{B}(\tilde{\mathbf{c}}_{1,i} - \tilde{\mathbf{c}}_{2,i})\right]^{\mathrm{T}} \mathbf{1}_P + \delta\left[(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P - 1\right] \\ + \sum_{i=1}^{B}\delta_i(\tilde{\mathbf{c}}_{1,i} + \tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P + \left(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_1 - \left(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_2 + \eta\ . \end{aligned} \tag{6.146}$$

From (6.146), we see that in order for $\hat{\delta}$ to have constant bias, we require

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1\ , \tag{6.147}$$

$$(\tilde{\mathbf{c}}_{1,i} + \tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P = 0 \quad \forall i = 1, \cdots, B \tag{6.148}$$

$$\left[\mathbf{c}_1 - \mathbf{c}_2 + \sum_{i=1}^{B}(\tilde{\mathbf{c}}_{1,i} - \tilde{\mathbf{c}}_{2,i})\right]^{\mathrm{T}} \mathbf{1}_P = 0\ . \tag{6.149}$$

Assuming $\hat{\delta}$ satisfies (6.147)-(6.149), its variance has the form

$$\mathrm{var}(\hat{\delta}) = \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_2 + \sum_{i=1}^{B}\left[\mathbf{c}_{1,i}^{\mathrm{T}}\mathbf{S}_1\mathbf{c}_{1,i} + \mathbf{c}_{2,i}^{\mathrm{T}}\mathbf{S}_2\mathbf{c}_{2,i} - 2\mathbf{c}_{1,i}^{\mathrm{T}}\mathbf{S}_{12}\mathbf{c}_{2,i}\right] \quad (6.150)$$

Thus, under the constraint of constant bias, the variance and consequently the MSE of $\hat{\delta}$ are constant with respect to δ and d. Further, for any choice of linear combination weights, $\mathrm{MSE}(\hat{\delta})$ can be minimized by making $\hat{\delta}$ unbiased, by setting

$$\eta = \left(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_2 - \left(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_1 \quad (6.151)$$

Now consider the residual problem of choosing the weight vectors $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_{1,i}$, $\mathbf{c}_{2,i}$ (for $i = 1, \cdots, B$) to minimize (6.150) under the constraints specified (6.147)-(6.149). It is easy to show using the method of Lagrangian multipliers that the values of $\mathbf{c}_{1,i}$ and $\mathbf{c}_{2,i}$ that solve this minimization problem are constant with respect to i. Hence, setting $\tilde{\mathbf{c}}_{1,i} = \tilde{\mathbf{c}}_1$ and $\tilde{\mathbf{c}}_{2,i} = \tilde{\mathbf{c}}_2$ for all i, we obtain the simpler problem

$$\begin{aligned} \min_{\mathbf{c}} \quad & \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{S}}\mathbf{c} \\ \text{s.t.} \quad & \mathbf{G}\mathbf{c} = \mathbf{s} \end{aligned} \quad (6.152)$$

where $\mathbf{c} = [\mathbf{c}_1^{\mathrm{T}}\ \mathbf{c}_2^{\mathrm{T}}\ \tilde{\mathbf{c}}_1^{\mathrm{T}}\ \tilde{\mathbf{c}}_2^{\mathrm{T}}]^{\mathrm{T}}$,

$$\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & B\mathbf{S} \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & -\mathbf{S}_{12} \\ -\mathbf{S}_{12}^{\mathrm{T}} & \mathbf{S}_2 \end{bmatrix} \quad (6.153)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{1}_P & \mathbf{1}_P & \mathbf{0}_P & \mathbf{0}_P \\ \mathbf{0}_P & \mathbf{0}_P & \mathbf{1}_P & \mathbf{1}_P \\ \mathbf{1}_P & -\mathbf{1}_P & B\mathbf{1}_P & -B\mathbf{1}_P \end{bmatrix} \qquad \mathbf{s} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \quad (6.154)$$

From Theorem 7, the solution to this problem is given as

$$\mathbf{c}^* = \tilde{\mathbf{S}}^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\tilde{\mathbf{S}}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.155}$$

which proves the corollary. □

Proof of Corollary 7. We first prove the result under the extended K-model. Under this model we have

$$\begin{aligned}\mathrm{Bias}_k(\hat{\delta}) = \delta\left[(\mathbf{c}_1+\mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P - 1\right] + \sum_{i=1}^{B}\delta_i(\tilde{\mathbf{c}}_{1,i}+\tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P \\ + \left(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \left(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_{2,k} + \eta\ .\end{aligned} \tag{6.156}$$

The constant bias condition property can be ensured under all K network scenarios by setting

$$(\mathbf{c}_1+\mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1\ , \tag{6.157}$$

$$(\tilde{\mathbf{c}}_{1,i}+\tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P = 0 \quad \forall i = 1,\cdots,B \tag{6.158}$$

Under this constraint, we have

$$\mathrm{MSE}_k(\hat{\delta}) = \mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{1,k}\mathbf{c}_1 + \mathbf{c}_2^{\mathrm{T}}\mathbf{S}_{2,k}\mathbf{c}_2 - 2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12,k}\mathbf{c}_2 \tag{6.159}$$

$$\begin{aligned}&+ \sum_{i=1}^{B}\left[\mathbf{c}_{1,i}^{\mathrm{T}}\mathbf{S}_{1,k}\mathbf{c}_{1,i} + \mathbf{c}_{2,i}^{\mathrm{T}}\mathbf{S}_{2,k}\mathbf{c}_{2,i} - 2\tilde{\mathbf{c}}_{1,i}^{\mathrm{T}}\mathbf{S}_{12,k}\tilde{\mathbf{c}}_{2,i}\right] \\ &+ \left[\boldsymbol{\mu}_{1,k}^{\mathrm{T}}\left(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\right) - \boldsymbol{\mu}_{2,k}^{\mathrm{T}}\left(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\right) + \eta\right]^2\end{aligned} \tag{6.160}$$

Let $\hat{\mathbf{c}} = [\mathbf{c}_1^{\mathrm{T}}\ \mathbf{c}_2^{\mathrm{T}}\ \tilde{\mathbf{c}}_{1,1}^{\mathrm{T}}\ \cdots\ \tilde{\mathbf{c}}_{1,B}^{\mathrm{T}}\ \tilde{\mathbf{c}}_{2,1}^{\mathrm{T}}\ \cdots\ \tilde{\mathbf{c}}_{2,B}^{\mathrm{T}}]^{\mathrm{T}}$. Then the problem of minimizing $\mathrm{WMaxMSE}(\hat{\delta})$ under the constraint of constant bias can be stated as

$$\min_{\hat{\mathbf{c}}\in\mathcal{C},\ \eta\in\mathbb{R}}\ \max_{\boldsymbol{\lambda}\in\Lambda}\ \sum_{k=1}^{K}\lambda_k\beta_k\mathrm{MSE}_k(\hat{\delta}) \tag{6.161}$$

where $\mathcal{C}$ is the set of values of $\hat{\mathbf{c}}$ for which (6.157) - (6.158) are satisfied. It is easy to see that Theorem 8 can be applied to (6.161), to obtain the equivalent problem

$$\max_{\boldsymbol{\lambda}\in\Lambda}\min_{\mathbf{c}\in\mathcal{C}}\min_{\eta\in\mathbb{R}}\sum_{k=1}^{K}\lambda_k\beta_k\mathrm{MSE}_k(\hat{\delta}) \tag{6.162}$$

For any fixed values of $\hat{\mathbf{c}}$ and $\boldsymbol{\lambda}$, the value of η that solves the inner minimization problem can be obtained via differentiation as

$$\begin{aligned}\eta = -\left(\sum_{k=1}^{K}\lambda_k\beta_k\right)^{-1}\Bigg\{\sum_{k=1}^{K}\lambda_k\beta_k\Bigg[\boldsymbol{\mu}_{1,k}^{\mathrm{T}}\Big(\mathbf{c}_1+\sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\Big)\\ -\boldsymbol{\mu}_{2,k}^{\mathrm{T}}\Big(\mathbf{c}_2+\sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\Big)\Bigg]\Bigg\}\end{aligned} \tag{6.163}$$

Substituting this optimum value of η in (6.162), we are left with the optimization problem

$$\begin{aligned}\max_{\boldsymbol{\lambda}\in\Lambda}\min_{\hat{\mathbf{c}}\in\mathcal{C}}\sum_{k=1}^{K}\lambda_k\beta_k\Bigg\{&\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{1,k}\mathbf{c}_1+\mathbf{c}_2^{\mathrm{T}}\mathbf{S}_{2,k}\mathbf{c}_2-2\mathbf{c}_1^{\mathrm{T}}\mathbf{S}_{12,k}\mathbf{c}_2\\ &+\sum_{i=1}^{B}\Big[\tilde{\mathbf{c}}_{1,i}^{\mathrm{T}}\mathbf{S}_{1,k}\tilde{\mathbf{c}}_{1,i}+\tilde{\mathbf{c}}_{2,i}^{\mathrm{T}}\mathbf{S}_{2,k}\tilde{\mathbf{c}}_{2,i}-2\tilde{\mathbf{c}}_{1,i}^{\mathrm{T}}\mathbf{S}_{12,k}\tilde{\mathbf{c}}_{2,i}\Big]\\ &+\Big[\hat{\boldsymbol{\mu}}_{1,k}^{\mathrm{T}}\big(\mathbf{c}_1+\sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\big)-\hat{\boldsymbol{\mu}}_{2,k}^{\mathrm{T}}\big(\mathbf{c}_2+\sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\big)\Big]^2\Bigg\}\end{aligned} \tag{6.164}$$

$$\text{s.t.}\begin{cases}(\mathbf{c}_1+\mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P=1\ ,\\ (\tilde{\mathbf{c}}_{1,i}+\tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P=0\quad\forall i=1,\cdots,B\\ \Big[\mathbf{c}_1-\mathbf{c}_2+\sum_{i=1}^{B}(\tilde{\mathbf{c}}_{1,i}-\tilde{\mathbf{c}}_{2,i})\Big]^{\mathrm{T}}\mathbf{1}_P=0\ .\end{cases} \tag{6.165}$$

where

$$\hat{\boldsymbol{\mu}}_{1,k}=\boldsymbol{\mu}_{1,k}-\frac{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}\boldsymbol{\mu}_{1,k'}}{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}}, \tag{6.166}$$

$$\hat{\boldsymbol{\mu}}_{2,k}=\boldsymbol{\mu}_{2,k}-\frac{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}\boldsymbol{\mu}_{2,k'}}{\sum_{k'=1}^{K}\lambda_{k'}\beta_{k'}}, \tag{6.167}$$

It is easy to show using the method of Lagrangian multipliers that for any value of $\boldsymbol{\lambda}$, the values of $\mathbf{c}_{1,i}$ and $\mathbf{c}_{2,i}$ that solve the inner minimization in (6.164) are constant with respect to i. Hence, setting $\tilde{\mathbf{c}}_{1,i} = \tilde{\mathbf{c}}_1$ and $\tilde{\mathbf{c}}_{2,i} = \tilde{\mathbf{c}}_2$ for all i, we obtain the simpler problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\Lambda} \min_{\mathbf{c}} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{M}\mathbf{c} \\ \text{s.t.} \quad & \mathbf{G}\mathbf{c} = \mathbf{s} \end{aligned} \tag{6.168}$$

where $\mathbf{c} = [\mathbf{c}_1^{\mathrm{T}}\ \mathbf{c}_2^{\mathrm{T}}\ \tilde{\mathbf{c}}_1^{\mathrm{T}}\ \tilde{\mathbf{c}}_2^{\mathrm{T}}]^{\mathrm{T}}$,

$$\mathbf{M} = \sum_{k=1}^{K} \lambda_k \beta_k \left\{ \begin{bmatrix} \mathbf{S}_k & \mathbf{0} \\ \mathbf{0} & B\mathbf{S}_k \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{\mu}}_k \\ B\hat{\boldsymbol{\mu}}_k \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}}_k \\ B\hat{\boldsymbol{\mu}}_k \end{bmatrix}^{\mathrm{T}} \right\}, \tag{6.169}$$

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{S}_{1,k} & -2\mathbf{S}_{12,k} \\ -2\mathbf{S}_{12,k} & \mathbf{S}_{2,k} \end{bmatrix},\ \hat{\boldsymbol{\mu}}_k = \begin{bmatrix} \hat{\boldsymbol{\mu}}_{1,k} \\ \hat{\boldsymbol{\mu}}_{2,k} \end{bmatrix}, \tag{6.170}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} \\ \mathbf{0}_P^{\mathrm{T}} & \mathbf{0}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} & \mathbf{1}_P^{\mathrm{T}} \end{bmatrix},\ \mathbf{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{6.171}$$

Using Theorem 7, the solution to the inner minimization problem in (6.168) is obtained as

$$\mathbf{c} = \mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.172}$$

and (6.168) reduces to the concave maximization problem

$$\max_{\boldsymbol{\lambda}\in\Lambda}\ \mathbf{s}^{\mathrm{T}}(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{s} \tag{6.173}$$

This proves the result for the extended K-model.

Under the extended S-model, we can use a similar proof, with different values for $\mathbf{G}$ and $\mathbf{s}$, since the constraints resulting from the constant bias condition are different relative to the

extended K-model. In particular, under the extended S-model, we have

$$\begin{aligned}\mathrm{Bias}_k(\hat{\delta}) = d\left[\mathbf{c}_1 - \mathbf{c}_2 + \sum_{i=1}^{B}(\tilde{\mathbf{c}}_{1,i} - \tilde{\mathbf{c}}_{2,i})\right]^{\mathrm{T}} \mathbf{1}_P + \delta\left[(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P - 1\right] \\ + \sum_{i=1}^{B}\delta_i(\tilde{\mathbf{c}}_{1,i} + \tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P + \left(\mathbf{c}_1 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{1,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_{1,k} - \left(\mathbf{c}_2 + \sum_{i=1}^{B}\tilde{\mathbf{c}}_{2,i}\right)^{\mathrm{T}}\boldsymbol{\mu}_{2,k} + \eta\ .\end{aligned} \tag{6.174}$$

The constant bias condition property can be ensured under all K network scenarios by setting

$$(\mathbf{c}_1 + \mathbf{c}_2)^{\mathrm{T}}\mathbf{1}_P = 1\ , \tag{6.175}$$

$$(\tilde{\mathbf{c}}_{1,i} + \tilde{\mathbf{c}}_{2,i})^{\mathrm{T}}\mathbf{1}_P = 0 \quad \forall i = 1, \cdots, B \tag{6.176}$$

$$\left[\mathbf{c}_1 - \mathbf{c}_2 + \sum_{i=1}^{B}(\tilde{\mathbf{c}}_{1,i} - \tilde{\mathbf{c}}_{2,i})\right]^{\mathrm{T}}\mathbf{1}_P = 0\ . \tag{6.177}$$

The concludes the proof. □

Chapter 7

Conclusions

In this dissertation, we developed novel techniques to improve detection and estimation in MIMO radars and network time synchronization schemes. Here we present some concluding remarks about work presented in Chapters 2 through 6. In Chapter 2, we developed a new definition of the ambiguity function for radars that perform non-coherent processing. We showed that the conventional definition of the ambiguity function suffers drawbacks when applied to non-coherent radars, and that our new definition was a better ambiguity measure. The new definition is especially relevant to many MIMO radars that are currently being studied. In Chapter 3, we developed detection and waveform design techniques under a general MIMO radar model that encompasses many existing MIMO radar configurations. The tools developed in this chapter can aid radar designers by removing restrictions on design parameters such as antenna separations and waveform correlations that are imposed by fixed MIMO radar configurations studied in literature.

In Chapter 4, we presented new lower bounds on the mean squared error of phase offset estimation schemes in PTP. These lower bounds help provide new insights into cases where

conventional estimation schemes perform well, and where significant scope for performance improvements may exist. In Chapter 5, we derived novel minimax optimum estimators under various observation models for the problem of phase offset estimation schemes in PTP. The estimators, while exhibiting a high implementation complexity, are guaranteed to provide the best possible performance under any network scenario. Hence, these estimators are better suited for obtaining lower bounds on achievable estimation performance, than as practical estimation procedures. To address the issue of designing practical estimators that exhibit near-optimum performance, in Chapter 6, we solve the problem of designing optimum L-estimators of phase offset. These estimators, obtained as linear combinations of order statistics, have a low implementation complexity and exhibit near-optimum performance under many network scenarios. We further demonstrate that L-estimators can be designed to be robust across a wide range of network scenarios. We believe that the L-estimators we developed show sufficient merit for their use in practical PTP implementations.

Bibliography

[1] J. Li and P. Stoica, "MIMO radar with colocated antennas," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 106–114, 2007.

[2] D. Fuhrmann and G. San Antonio, "Transmit beamforming for MIMO radar systems using signal cross-correlation," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 1, pp. 171–186, 2008.

[3] J. Li and P. Stoica, "MIMO radar–diversity means superiority," in *Proceedings of the 14th Adaptive Sensor Array Processing Workshop (ASAP06)*, 2009, pp. 1–6.

[4] N. Lehmann, A. Haimovich, R. Blum, and L. Cimini, "High resolution capabilities of MIMO radar," in *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on.* IEEE, 2006, pp. 25–30.

[5] M. Haleem and A. Haimovich, "On the distribution of ambiguity levels in MIMO radar," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on.* IEEE, 2008, pp. 198–202.

[6] A. Haimovich, R. Blum, and L. Cimini, "MIMO radar with widely separated antennas," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 116–129, 2008.

[7] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO radar: An idea whose time has come," in *Radar Conference, 2004. Proceedings of the IEEE*. IEEE, 2004, pp. 71–78.

[8] H. Godrich, A. Haimovich, and R. Blum, "Concepts and applications of a MIMO radar system with widely separated antennas," *MIMO Radar Signal Processing*, pp. 365–410, 2008.

[9] T. Aittomaki and V. Koivunen, "Performance of MIMO radar with angular diversity under Swerling scattering models," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 1, pp. 101–114, 2010.

[10] E. Fishler, A. Haimovich, R. Blum, L. Cimini Jr, D. Chizhik, and R. Valenzuela, "Spatial diversity in radars-models and detection performance," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 823–838, 2006.

[11] D. L. Mills, *Network Time Protocol (Version 3) Specification, Implementation and Analysis*, Internet Request for Comments, 1992.

[12] "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," *IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002)*, pp. c1–269, 2008.

[13] J. Elson, L. Girod, and D. Estrin, "Fine-grained network time synchronization using reference broadcasts," *ACM SIGOPS Operating Systems Review*, vol. 36, no. SI, pp. 147–163, 2002.

[14] J. Elson and D. Estrin, "Time synchronization for wireless sensor networks," in *Proceedings of the 15th International Parallel &Amp; Distributed Processing Symposium*,

ser. IPDPS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 186–. [Online]. Available: http://dl.acm.org/citation.cfm?id=645609.662464

[15] M. L. Sichitiu and C. Veerarittiphan, "Simple, accurate time synchronization for wireless sensor networks," in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, vol. 2. IEEE, 2003, pp. 1266–1273.

[16] K. Römer, "Time synchronization in ad hoc networks," in *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*. ACM, 2001, pp. 173–182.

[17] I. Hadzic and D. R. Morgan, "Adaptive packet selection for clock recovery," in *2010 International IEEE Symposium on Precision Clock Synchronization for Measurement Control and Communication (ISPCS)*. IEEE, 2010, pp. 42–47.

[18] I. Hadzic, D. R. Morgan, and Z. Sayeed, "A synchronization algorithm for packet mans," *Communications, IEEE Transactions on*, vol. 59, no. 4, pp. 1142–1153, 2011.

[19] K. Wolter, P. Reinecke, and A. Mittermaier, "Model-based evaluation and improvement of PTP syntonisation accuracy in packet-switched Backhaul networks for mobile applications," in *Computer Performance Engineering*. Springer, 2011, pp. 219–234.

[20] B. Mochizuki and I. Hadzic, "Improving IEEE 1588v2 clock performance through controlled packet departures," *Communications Letters, IEEE*, vol. 14, no. 5, pp. 459–461, 2010.

[21] T. Murakami, Y. Horiuchi, and K. Nishimura, "A packet filtering mechanism with a packet delay distribution estimation function for IEEE 1588 time synchronization in a

congested network," in *Precision Clock Synchronization for Measurement Control and Communication (ISPCS), 2011 International IEEE Symposium on.* IEEE, 2011, pp. 114–119.

[22] M. Anyaegbu, C.-X. Wang, and W. Berrie, "A sample-mode packet delay variation filter for IEEE 1588 synchronization," in *ITS Telecommunications (ITST), 2012 12th International Conference on.* IEEE, 2012, pp. 1–6.

[23] A. Guruswamy and R. Blum, "On a definition of the ambiguity function for non-coherent radars," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, June 2012, pp. 141–144.

[24] A. Guruswamy, V. Reddy, and R. Blum, "Mimo radar under general waveform correlations/antenna spacing: Optimum detectors, detection performance and waveform design," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, March 2012, pp. 1–6.

[25] A. Guruswamy, R. Blum, S. Kishore, and M. Bordogna, "Performance lower bounds for phase offset estimation in ieee 1588 synchronization," *Communications, IEEE Transactions on*, vol. 63, no. 1, pp. 243–253, Jan 2015.

[26] A. Guruswamy, R. Blum, S. Kishore, and M. Bordogna, "Minimax optimum estimators for phase synchronization in ieee 1588," *Communications, IEEE Transactions on*, vol. 63, no. 9, pp. 3350–3362, Sept 2015.

[27] A. Guruswamy, R. Blum, S. Kishore, and M. Bordogna, "On the optimum design of l-estimators for phase offset estimation in ieee 1588," *Communications, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[28] P. M. Woodward and I. L. Davies, *Probability and Information Theory With Applications to Radar*. London, U.K.: Pergamon, 1953.

[29] J. Li and P. Stoica, "MIMO Radar with Colocated Antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106 –114, Sept. 2007.

[30] I. Bradaric, G. Capraro, and M. Wicks, "Waveform Diversity for Different Multistatic Radar Configurations," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov. 2007, pp. 2038 –2042.

[31] T. Derham, S. Doughty, C. Baker, and K. Woodbridge, "Ambiguity Functions for Spatially Coherent and Incoherent Multistatic Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 230 –245, Jan. 2010.

[32] G. San Antonio, D. Fuhrmann, and F. Robey, "MIMO Radar Ambiguity Functions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 167 –177, June 2007.

[33] A. Haimovich, R. Blum, and L. Cimini, "MIMO Radar with Widely Separated Antennas," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 116 –129, 2008.

[34] T. Al-Naffouri and B. Hassibi, "On the distribution of indefinite quadratic forms in Gaussian random variables," in *IEEE International Symposium on Information Theory*, July 2009, pp. 1744 –1748.

[35] T. Y. Al-Naffouri and B. Hassibi, "On the distribution of indefinite quadratic forms in Gaussian randomvariables," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on.* IEEE, 2009, pp. 1744–1748.

[36] N. R. Draper and H. Smith, "Applied regression analysis 2nd ed." 1981.

[37] G. Blewitt, "Basics of the gps technique: observation equations," *Geodetic applications of GPS*, pp. 10–54, 1997.

[38] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *Information Theory, IEEE Transactions on*, vol. 34, no. 2, pp. 338–342, 1988.

[39] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai, "A fresh look at the bayesian bounds of the weiss-weinstein family," *Signal Processing, IEEE Transactions on*, vol. 56, no. 11, pp. 5334–5352, 2008.

[40] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, 1969.

[41] C. Radhakrishna Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[42] E. Barankin, "Locally best unbiased estimates," *The Annals of Mathematical Statistics*, pp. 477–501, 1949.

[43] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhyā: The Indian Journal of Statistics (1933-1960)*, vol. 8, no. 1, pp. 1–14, 1946.

[44] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended ziv-zakai lower bound for vector parameter estimation," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 624–637, 1997.

[45] "Timing and Synchronization Aspects in Packet Networks," *Telecommunication Standardization Sector, International Telecommunication Union (ITU), ITU-T Recommendation G.8261*, Apr 2008.

[46] "OPNET Modeler," www.opnet.com, [Online; accessed Dec 2013].

[47] J. Peng, L. Zhang, and D. McLernon, "On the Clock Offset Estimation in an Improved IEEE 1588 Synchronization Scheme," in *Proceedings of the 2013 19th European Wireless Conference (EW)*, April 2013, pp. 1–6.

[48] G. Giorgi and C. Narduzzi, "Performance analysis of Kalman-filter-based clock synchronization in IEEE 1588 networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2902–2909, 2011.

[49] A. Bletsas, "Evaluation of Kalman filtering for network time keeping," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 52, no. 9, pp. 1452–1460, 2005.

[50] C. Iantosca, C. Heitz, and H. Weibel, "Synchronizing IEEE 1588 clocks under the presence of significant stochastic network delays," in *Conference on IEEE 1588*, CH Winterthur, 2005.

[51] E. J. G. Pitman, "The Estimation of the Location and Scale Parameters of a Continuous Population of any Given Form," *Biometrika*, vol. 30, no. 3/4, pp. pp. 391–421, 1939. [Online]. Available: http://www.jstor.org/stable/2332656

[52] A. Kagan, T. Yu, A. Barron, and M. Madiman, "Contribution to the theory of pitman estimators," *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.(POMI)*, vol. 408, pp. 245–267, 2012.

[53] A. Guruswamy, R. Blum, S. Kishore, and M. Bordogna, "Performance Lower Bounds for Phase Offset Estimation in IEEE 1588 Synchronization," *IEEE Transactions on Communications*, vol. PP, no. 99, pp. 1–1, 2014.

[54] P. Yang, H. Tang, X. Chen, and M. Tian, "New algorithm for IEEE 1588 time synchronization under the presence of significant delay variation," in *Consumer Electronics, Communications and Networks (CECNet), 2013 3rd International Conference on*. IEEE, 2013, pp. 419–422.

[55] M. Anyaegbu, C. Wang, and W. Berrie, "Dealing with Packet Delay Variation in IEEE 1588 Synchronization Using a Sample-Mode Filter," *IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 4, pp. 20–27, 2013.

[56] R. Subrahmanyan, "Timing recovery for IEEE 1588 applications in telecommunications," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, no. 6, pp. 1858–1868, 2009.

[57] P. J. Huber, *Robust statistics*. Springer, 2011.

[58] V. P. Godambe and B. K. Kale, "Estimating functions: an overview," *Estimating functions*, vol. 7, pp. 3–20, 1991.

[59] P. J. Huber, *Robust Statistical Procedures*. SIAM, 1996.

[60] P. P. Gandhi and S. A. Kassam, "Design and performance of combination filters for signal restoration," *Signal Processing, IEEE Transactions on*, vol. 39, no. 7, pp. 1524–1540, 1991.

[61] D. R. Jeske, "On maximum-likelihood estimation of clock offset," *Communications, IEEE Transactions on*, vol. 53, no. 1, pp. 53–54, 2005.

[62] Q. M. Chaudhari, E. Serpedin, and Y.-C. Wu, "Improved estimation of clock offset in sensor networks," in *Communications, 2009. ICC'09. IEEE International Conference on.* IEEE, 2009, pp. 1–4.

[63] R. T. Rockafellar, *Convex analysis.* Princeton university press, 1970, no. 28.

[64] D. Bladsjo, M. Hogan, and S. Ruffini, "Synchronization aspects in lte small cells," *Communications Magazine, IEEE*, vol. 51, no. 9, pp. 70–77, 2013.

[65] C. A. Hoare, "Quicksort," *The Computer Journal*, vol. 5, no. 1, pp. 10–16, 1962.

[66] R. A. Johnson, D. W. Wichern *et al.*, *Applied multivariate statistical analysis.* Prentice hall Englewood Cliffs, NJ, 1992, vol. 4.

Vita

Anand Srinivas Guruswamy was born in Rourkela, Odisha, India on 8th April 1986, to Guruswamy Krishna and Vijayalaxmi Guruswamy. He received the B.Tech. (Honors) degree in Electronics and Communications Engineering from the International Institute of Information Technology (IIIT), Hyderabad, India, in 2008. Between 2008 and 2010, he worked with Qualcomm India as an Associate Engineer, and subsequently as a Project Associate at the Communications Research Center at IIIT-Hyderabad. Since 2010, he has been working towards the Ph.D. degree in Electrical Engineering at Lehigh University in Bethlehem, PA, USA. During the summer of 2012, he interned at Qualcomm Inc at San Diego, CA, USA.