

Syracuse University

**SURFACE**

---

Electrical Engineering and Computer Science -  
Dissertations

College of Engineering and Computer Science

---

5-2013

## Collaborative Estimation in Distributed Sensor Networks

Swarnendu Kar  
*Syracuse University*

Follow this and additional works at: [https://surface.syr.edu/eecs\\_etd](https://surface.syr.edu/eecs_etd)



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Kar, Swarnendu, "Collaborative Estimation in Distributed Sensor Networks" (2013). *Electrical Engineering and Computer Science - Dissertations*. 333.

[https://surface.syr.edu/eecs\\_etd/333](https://surface.syr.edu/eecs_etd/333)

This Dissertation is brought to you for free and open access by the College of Engineering and Computer Science at SURFACE. It has been accepted for inclusion in Electrical Engineering and Computer Science - Dissertations by an authorized administrator of SURFACE. For more information, please contact [surface@syr.edu](mailto:surface@syr.edu).

# ABSTRACT

Networks of smart ultra-portable devices are already indispensable in our lives, augmenting our senses and connecting our lives through real time processing and communication of sensory (e.g., audio, video, location) inputs. Though usually hidden from the user's sight, the engineering of these devices involves fierce tradeoffs between energy availability (battery sizes impact portability) and signal processing / communication capability (which impacts the "smartness" of the devices). The goal of this dissertation is to provide a fundamental understanding and characterization of these tradeoffs in the context of a sensor network, where the goal is to estimate a common signal by coordinating a multitude of battery-powered sensor nodes. Most of the research so far has been based on two key assumptions – "distributed processing" and "temporal independence" – that lend analytical tractability to the problem but otherwise are often found lacking in practice. This dissertation introduces novel techniques to relax these assumptions – leading to vastly efficient energy usage in typical networks (up to 20% savings) and new insights on the quality of inference. For example, the phenomenon of sensor drift is ubiquitous in applications such as air-quality monitoring, oceanography and bridge monitoring, where calibration is often difficult and costly. This dissertation provides an analytical framework linking the state of calibration to the overall uncertainty of the inferred parameters.

In distributed estimation, sensor nodes locally process their observed data and send the resulting messages to a sink, which combines the received messages to produce a final estimate of the unknown parameter. In this dissertation, this problem is generalized and called "collaborative estimation", where some sensors can potentially have access to the observations from neighboring sensors and use that information to enhance the quality of their messages sent to the sink, while using the same (or lower) energy resources. This is motivated by the fact that inter-sensor communication may be possible if sensors are

geographically close. As demonstrated in this dissertation, collaborative estimation is particularly effective in "energy-skewed" and "information-skewed" networks, where some nodes may have larger batteries than others and similarly some nodes may be more informative (less noisy) compared to others. Since the node with the largest battery is not necessarily also the most informative, the proposed inter-sensor collaboration provides a natural framework to route the relevant information from low-energy-high-quality nodes to high-energy-low-quality nodes in a manner that enhances the overall power-distortion tradeoff.

This dissertation also analyzes how time-correlated measurement noise affects the uncertainties of inferred parameters. Imperfections such as baseline drift in sensors result in a time-correlated additive component in the measurement noise. Though some models of drift have been reported in the literature earlier, none of the studies have considered the effect of drifting sensors on an estimation application. In this dissertation, approximate measures of estimation accuracy (Cramér-Rao bounds) are derived as a function of physical properties of sensors – namely the drift strength, correlation (Markov) factor and the time-elapsd since last calibration. For stationary drift (Markov factor less than one), it is demonstrated that the first order effect of drift is asymptotically equivalent to scaling the measurement noise by an appropriate factor. When the drift is non-stationary (Markov factor equal to one), it is established that the constant part of a signal can only be estimated inconsistently (with non-zero asymptotic variance). The results help quantify the notions that measurements taken sooner after calibration result in more accurate inference.

# COLLABORATIVE ESTIMATION IN DISTRIBUTED SENSOR NETWORKS

By

Swarnendu Kar

B.Tech. Indian Institute of Technology, Kharagpur, 2004  
M.S. Syracuse University, 2009

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in Electrical Engineering

Syracuse University  
May 2013

Copyright © 2013 Swarnendu Kar

All rights reserved

# ACKNOWLEDGMENTS

I would like to take this opportunity to thank my advisor Prof Pramod Varshney for his continued intellectual, motivational and financial support through the roller-coaster ride called doctoral studies. Immense gratitude is due to my parents and family members for their sustained cheer-leading and support during all these years I spent away from them. Special thanks to my wife Chandrani, whom I came to know mid-way through this program and who has since supported me in every manner possible, including keeping up with my night schedules. I am also indebted to my Sensor Fusion group colleagues and other friends, including Arijit De, Kapil Borle, Priyadip Ray, Min Xu, Hao Chen, Onur Ozdemir, Ruixin Niu, Arun Subramanian, Ashok Sundaresan, Satish Iyengar, Sid Nadendla, Aditya Vempaty, Swastik Brahma, Raghed Ahmad El Bardan, Thakshila Wimalajeewa, Nianxia Cao, Bhavya Kailkhura, Hao He, Yujiao Zheng, Sijia Liu and Xiaojing Shen, with whom I had many enlightening discussions.

# Contents

List of Tables . . . . .	xi
List of Figures . . . . .	xii
<b>1 Introduction</b>	<b>1</b>
1.1 Distributed and collaborative estimation . . . . .	2
1.2 Problems addressed and thesis organization . . . . .	4
1.3 Bibliographic notes . . . . .	7
<b>2 Background and literature survey</b>	<b>9</b>
2.1 System description . . . . .	9
2.2 Past research . . . . .	10
2.3 Taxonomy . . . . .	14
2.3.1 Signal and noise . . . . .	14
2.3.2 Quantization . . . . .	16
2.3.3 Modulation for digital communication . . . . .	18
2.3.4 Channel models . . . . .	19
2.3.5 Fusion-rules and distortion metrics . . . . .	20
2.4 Major contributions . . . . .	23
<b>3 Quantizer design for distributed estimation</b>	<b>25</b>
3.1 Introduction . . . . .	25
3.1.1 Identical one-bit quantizers . . . . .	26
3.1.2 Error-free reception . . . . .	28

3.1.3	Performance metric . . . . .	28
3.1.4	Previous work . . . . .	30
3.2	Main results . . . . .	31
3.3	Design of antisymmetric minimax-CRB quantizer . . . . .	36
3.4	Illustrative examples . . . . .	39
3.5	Summary . . . . .	41
<b>4</b>	<b>Bit allocation for collaborative estimation using spatial whitening</b>	<b>43</b>
4.1	Introduction . . . . .	43
4.2	Problem statement . . . . .	45
4.3	Iterative algorithm for spatial whitening . . . . .	49
4.4	Bit-allocation for collaborative estimation . . . . .	50
4.5	Summary . . . . .	53
<b>5</b>	<b>Power allocation for analog forwarding based collaborative estimation</b>	<b>56</b>
5.1	Introduction . . . . .	56
5.2	Problem formulation . . . . .	60
5.2.1	Linear sensing model . . . . .	60
5.2.2	Linear spatial collaboration . . . . .	61
5.2.3	Coherent multiple access channel . . . . .	63
5.2.4	Linear minimum mean square estimation . . . . .	64
5.2.5	Problem statement . . . . .	66
5.2.6	Solution methodology . . . . .	67
5.2.7	Performance metric - Collaboration gain . . . . .	68
5.3	Main results - Ideal collaborative power allocation: . . . . .	70
5.3.1	Explicit formulation w.r.t. non-zero weights . . . . .	70
5.3.2	Cumulative power constraint . . . . .	73
5.3.3	Individual power constraints . . . . .	80
5.3.4	Collaboration gain for a homogeneous network . . . . .	89



5.4	Collaboration with finite costs . . . . .	101
5.4.1	Individual power constraint . . . . .	101
5.4.2	Cumulative power constraint . . . . .	103
5.4.3	Numerical simulations . . . . .	104
5.5	Summary . . . . .	105
<b>6</b>	<b>Analog forwarding based collaborative estimation of dynamic parameters</b>	<b>107</b>
6.1	Introduction . . . . .	107
6.2	Problem formulation . . . . .	109
6.2.1	Collaboration strategies . . . . .	111
6.2.2	Partially connected networks . . . . .	113
6.2.3	Ornstein-Uhlenbeck process . . . . .	114
6.3	Main results . . . . .	117
6.3.1	Single snapshot estimation . . . . .	117
6.3.2	Time varying process estimation . . . . .	120
6.4	Summary . . . . .	127
<b>7</b>	<b>Signal estimation using sensors with baseline drift</b>	<b>129</b>
7.1	Introduction . . . . .	129
7.1.1	Inference with drifting sensors . . . . .	130
7.1.2	Models of noise and drift . . . . .	131
7.1.3	Deterministic signal model . . . . .	133
7.1.4	Related work . . . . .	134
7.2	Problem formulation . . . . .	135
7.2.1	Sensor calibration and noise covariance . . . . .	137
7.2.2	Maximum-Likelihood estimation and Cramér-Rao bound . . . . .	139
7.3	Main result: single sensor case . . . . .	140
7.3.1	Inverse of disturbance covariance . . . . .	140

7.3.2	Computation of Fisher Information Matrix . . . . .	143
7.3.3	Cramér-Rao bounds for $\rho < 1$ . . . . .	144
7.3.4	Cramér-Rao bounds for $\rho = 1$ . . . . .	147
7.3.5	Numerical results - Maximum relative error . . . . .	150
7.4	Multiple sensors . . . . .	153
7.4.1	Numerical results . . . . .	154
7.5	Quantized observations . . . . .	157
7.5.1	Numerical results . . . . .	159
7.6	Summary . . . . .	160
<b>8</b>	<b>Conclusions and suggestions for future work</b>	<b>161</b>
<b>A</b>	<b>Appendix - Proofs of various results</b>	<b>166</b>
A.1	Proof of Lemma 3.2.5 . . . . .	166
A.2	Proof of Theorem 3.2.6 . . . . .	167
A.3	Derivation for Example 3.2.7 . . . . .	167
A.4	Proof of Proposition 3.2.8 . . . . .	168
A.5	Derivation of limits in Table 3.1 . . . . .	169
A.6	Proof of Proposition 5.3.2 . . . . .	170
A.7	An equality involving positive definite matrices . . . . .	172
A.8	Proof of Proposition 5.3.4 . . . . .	173
A.9	Proof of Proposition 5.3.5 . . . . .	174
A.10	Proof of Proposition 5.3.7 . . . . .	177
A.11	Proof of Lemma 5.3.8 . . . . .	179
A.12	Proof of Proposition 5.3.9 . . . . .	180
A.13	Proof of Theorem 6.3.1 . . . . .	182
A.14	Explicit expressions for Rayleigh distributed gains . . . . .	184
A.15	Proof of Theorem 6.3.2 . . . . .	186
A.16	Proof of Proposition 7.3.1 . . . . .	188

A.17 Polynomial Approximation of $\mathbf{X}'\mathbf{A}\mathbf{X}$ . . . . .	190
A.18 FIM and CRB for $\rho = 1$ . . . . .	194
<b>Bibliography</b>	<b>197</b>

# List of Tables

2.1	Summary of literature in distributed estimation with a fixed fusion center.	13
2.2	Summary of our contributions on distributed and collaborative estimation.	24
3.1	Lower bound on maximum-CRB using Sine Quantizers. . . . .	35
5.1	Main results: cumulative power constraint (Examples 1, 2 and 3) . . . . .	99
5.2	Main results: individual power constraint (Examples 4, 5 and 6) . . . . .	100
A.1	Values of several constants helpful to derive $\mathcal{A}_{k,l,i}$ and $\alpha_{k,l}^{(\tau)}$ . . . . .	194

# List of Figures

1.1	Distributed estimation (left) and estimation with spatial collaboration among sensors (right). . . . .	3
2.1	System diagram of decentralized estimation. . . . .	10
3.1	System diagram of identical binary quantization based distributed estimation.	26
3.2	Minimum Fisher information $(\phi(\gamma, f)^{-1})$ for Threshold, Sine, Dithering and AUPL quantizers for (a) Laplacian and (b) Gaussian noise. . . . .	39
3.3	Designed probability $\gamma_P(x) = P(Y = \mathcal{S}_1 x)$ for AUPL quantizer. . . . .	41
3.4	Overall probability $g(\theta) = \mathbb{E}_W(\gamma_P(\theta + W))$ for AUPL quantizer. . . . .	42
4.1	System diagram of spatial whitening based distributed estimation. . . . .	46
4.2	Random Geometric Graph used for example in Section 4.4. Edges are shown of pairwise distance less than 0.18. . . . .	54
4.3	Distortion reduction achieved by spatial whitening. . . . .	55
5.1	Sensor network performing collaborative estimation. . . . .	57
5.2	A typical power-distortion curve illustrating collaboration gain. . . . .	70
5.3	Matrices $\mathbf{G}$ and $\mathbf{\Omega}_P$ assuming columnwise ordering of elements in $\mathbf{W}$ . . .	72
5.4	Directed cycle graphs, $(K - 1)$ connected. . . . .	79
5.5	Collaboration gain in a homogeneous sensor network with skewed power constraints. . . . .	94

5.6	Random Geometric Graph with 20 nodes, used for example in Section 5.3.4. Edges are shown for pairwise distance less than 0.2. The radius of collaboration is depicted for sensors 1,2,3 and 17. . . . .	96
5.7	Improvement in distortion due to increasing collaboration among 20 nodes in a random geometric graph. . . . .	98
5.8	Estimation with finite collaboration cost - an example with 10 sensor nodes.	105
6.1	Wireless sensor network performing collaborative estimation. . . . .	109
6.2	Example of $Q$ -cliques. . . . .	113
6.3	Example of two practical collaboration topologies for a $N = 20$ -node network. Bidirectional links are shown without arrows. . . . .	115
6.4	Periodically sampled Ornstein-Uhlenbeck process. . . . .	116
6.5	Energy-constrained estimation with single snapshot spatial sampling. . . .	121
6.6	Power-constrained estimation of OU process . . . . .	124
6.7	Power-constrained estimation of OU process - Average variance . . . . .	126
7.1	Estimation of a polynomial signal using sensors with baseline drift. . . . .	130
7.2	Sample size $T_\epsilon$ required for 95% accurate performance prediction . . . . .	152
7.3	Average performance for multi-sensor systems for $\rho \in [0.85, 0.95]$ . . . . .	155
7.4	Average performance for multi-sensor systems for $\rho = 1$ . . . . .	156
7.5	Cramér-Rao bounds with quantized observations. . . . .	160
A.1	Example $Q = 2$ -clique topology with $N = 4$ nodes. Illustration of $\mathbf{\Omega}$ and $\mathbf{G}$ matrices assuming row-wise ordering of elements in $\mathbf{W}$ . . . . .	183
A.2	Illustrating $\mathbf{p}\mathbf{p}^T$ and the summations along diagonals in various quadrants	187

# Chapter 1

## Introduction

Over the last decade, advances in low-power micro-electro-mechanical system (MEMS) technology and wireless communications have led to the emergence of wireless sensor networks (WSNs) [2]. A wireless sensor network consists of a large number of sensors that can communicate with each other to achieve a specific task. Individually, each sensor in a sensor network is characterized by a low power constraint which translates to limited computation and communication capabilities. However, when suitably deployed in large scale, potentially powerful networks can be constructed to accomplish various high-level tasks, and this makes wireless sensor networks a promising technology for a wide range of applications. Examples of current and potential applications of wireless sensor networks include environment monitoring, military sensing, traffic surveillance, health care, and smart homes [2, 19, 13].

A common goal in many WSN applications is to reconstruct or infer regarding the underlying physical phenomenon based on sensor measurements. The estimation of unknown parameters by a set of distributed sensor nodes and a fusion center has become an important topic in signal processing research for wireless sensor networks [95, 52, 67]. In distributed parameter estimation, sensor nodes collect real-valued data, locally process the data, and send the resulting messages to the fusion center (FC), which combines all the received messages to produce a final estimate of the unknown parameter. Both the local processing of data and its subsequent transmission are done in accordance with the

limited energy resources inside each node and usually involves some loss of information. The purpose of this thesis is to understand the tradeoff between this loss of information and energy requirements for various hitherto-unexplored but practical scenarios of interest in the context of estimation.

In addition to local information processing, estimation using a WSN may also involve inter-sensor communication if sensors are geographically close. In fact, a major challenge in WSN research is the integrated design of local signal processing operations and strategies for inter-sensor communication and networking so as to achieve a desirable trade-off among resource efficiency, system performance and ease of implementation.

We illustrate the ideas of distributed estimation and collaborative estimation in Figure 1.1. In the traditional centralized estimation, the sensors receive noisy observations based on the signal of interest (dotted lines imply information loss) and communicate the full-precision observations to the fusion center (bold lines indicate no information loss). In case of distributed estimation, the sensors perform some resource-conscious local processing before transmission to the fusion center. In collaborative estimation, the sensors communicate the observations among their neighbors before deciding on the exact data to be transmitted to the fusion center. The local communications for the purpose of spatial collaboration are assumed to be lossless in this thesis, since this can mostly be ensured for sensors with sufficient geographical proximity. However, in a possibly more general setting, this assumption may need to be relaxed. The act of collaboration is important since it enables efficient allocation of energy resources throughout the network.

## 1.1 Distributed and collaborative estimation

Enabling distributed and collaborative estimation applications in the context of a sensor network may involve several challenges, some of which are discussed below.

- *Resource allocation:* Sensor networks may comprise of a wide variety of sensors with a varying degree of reliability or precision. In those situations, it is important to assign



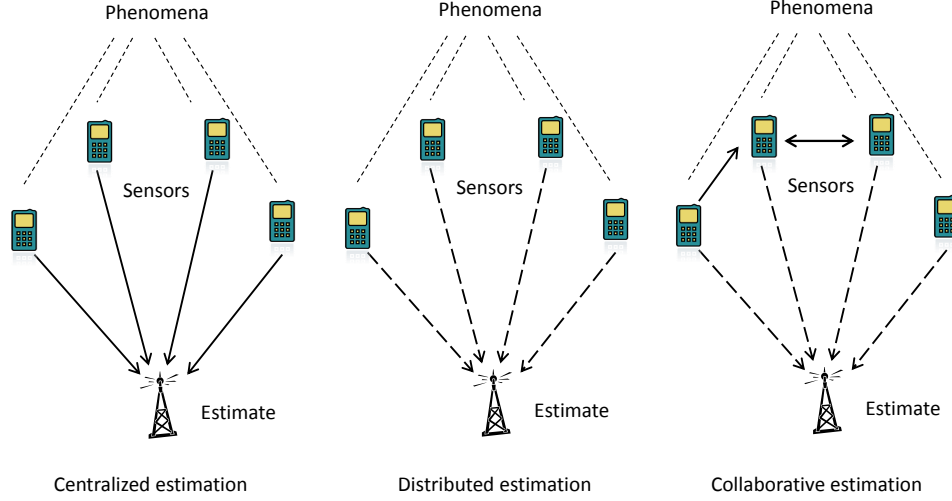


Figure 1.1: Distributed estimation (left) and estimation with spatial collaboration among sensors (right).

resources to each of these sensors in a judicious manner. By resource, we mean either the power required for the transmission of analog values or the number of bits for quantization (assuming a fixed cost for transmitting each bit). Intuitively, a sensor with better observation quality should be assigned a higher amount of resources. However, finding the *optimal* strategy is a difficult problem for most scenarios. The problem becomes more complicated when the sensors have correlated observations, in which case even the earlier intuition is invalid, i.e., better individual observation quality does not necessarily imply allocating more resources. For example, if two sensors produce absolutely identical observations, then it makes sense to transmit only one of them. The other sensor in this case is assigned zero resources, even though it is equally informative. Significant correlation among observation noise may be exhibited when the sensors are geographically close.

- *Quantizer design:* Constraints on sensor cost, bandwidth, and energy budget dictate that low-quality sensor observations may have to be aggressively quantized (e.g., down to a few bits per observation per sensor). Furthermore, local compression at a sensor node depends not only on the quality of sensor observation, but also on the quality of the wireless communication channels between sensor nodes and the fusion center.

Thus, estimators must be developed based on severely quantized versions of very noisy observations. Designing a quantizer involves finding the optimum mapping between the observation space and quantizer outputs and is usually a difficult problem.

- *Different signal and noise conditions:* Owing to the difficulty (in general) of quantizer design and resource allocation problems, distributed and collaborative signal processing algorithms are intricately tailored to the kind of signal and noise conditions present in the network. Motivated by realistic application scenarios, different assumptions may be made about the observation characteristics. For example, the signal of interest may be constant with time (e.g. temperature of a region within the span of a few minutes), or may be time-varying based on some predictable temporal dynamics (e.g. gas concentration in a region with a constant rate of pollutants), or may vary randomly with time based on some known statistics (e.g. temperature of a region within the span of a several days). Similarly, the noise conditions also play an important role. Based on the spatio-temporal correlation properties of noise and the state of calibration of the sensors, different compression/transmission/estimation techniques may be required.

## 1.2 Problems addressed and thesis organization

In this thesis we address several problems on centralized, distributed and collaborative estimation. Though the topics of centralized and distributed estimation are well researched in the literature, collaborative estimation is still a nascent area and a majority of the work in this thesis is directed to that topic. In Chapter 2, we formally introduce the distributed signal processing framework in wireless sensor networks and present an extensive literature review of centralized and distributed estimation problems.

Among the various possibilities for power saving local signal processing, one-bit (or binary) quantization is particularly appealing due to its ease of implementation. Traditional one-bit quantization involves fixing a threshold, below which the encoded value is “0” and

above which it is “1”. In Chapter 3, we consider *probabilistic quantizers*, where quantization is defined not just by a simple threshold but instead defined as an arbitrary function that maps the noisy observations to its probability of being encoded as “0” (say). We assume the parameter to be range limited and design a quantization rule to minimize the maximum distortion over this parameter range. We also identify a broad class of noise distributions for which a simple threshold quantizer is also the optimal probabilistic quantizer (in some sense). For a wide range of noise distributions, we demonstrate the superior performance of the probabilistic quantizer - particularly in the moderate to high-SNR regime.

Designing resource allocation strategies for power constrained sensor networks in the presence of correlated data often gives rise to intractable problem formulations. In such situations, applying well-known strategies derived from conditional-independence assumption may turn out to be fairly suboptimal. In Chapter 4, we address this issue for the situation where neighboring nodes are able to collaborate among themselves prior to transmission. In a scheme which we refer to as *spatial whitening*, each sensor updates their observations in a coordinated manner such that the updated observations are maximally whitened, i.e., their covariance matrix is as close to a diagonal matrix as possible. We demonstrate that existing bit allocation schemes, when applied on this updated observation space, yield superior estimation performance. We also comment on the computational limitations for obtaining the optimal whitening transformation, and propose an iterative optimization scheme to achieve the same for large networks.

In Chapter 5, we explore the still-nascent research topic of *collaborative estimation* in the context of single-snapshot estimation of a random parameter. By single-snapshot estimation, we mean that the FC makes an inference based on observations collected at one particular instant. In contrast to the distributed estimation problem, this collaborative framework allows the sensor nodes to update their observations by (linearly) combining observations from other adjacent nodes. The updated observations are communicated to the FC by transmitting through a wireless channel. The optimal collaborative strategy is obtained by minimizing the expected mean-square-error subject to power constraints at

the sensor nodes. Each sensor can utilize its available power for both collaboration with other nodes and transmission to the FC. Two kinds of constraints, namely the cumulative and individual power constraints are considered. The effects due to imperfect information about observation and channel gains are also investigated. The resulting performance improvement is illustrated analytically through the example of a homogeneous network with equi-correlated parameters. Assuming random geometric graph topology for collaboration, numerical results demonstrate a significant reduction in distortion even for a moderately connected network, particularly in the low local-SNR regime.

In Chapter 6, we build on our research on collaborative estimation by exploring two important extensions. Firstly, for the single-snapshot estimation, we gain further insights into partially connected collaboration networks (nearest-neighbor and random geometric graphs for example) through the analysis of a family of topologies with regular structure. Secondly, we explore the estimation problem by adding the dimension of time, where the goal is to estimate a time-varying *dynamic signal* in a power-constrained network. To model the time dynamics, we consider the stationary Gaussian process with exponential covariance (sometimes referred to as Ornstein-Uhlenbeck process) as our representative signal. For such a signal, we show that it is always beneficial to sample as frequently as possible, despite the fact that the samples get increasingly noisy due to the power-constrained nature of the problem. Simulation results are presented to corroborate our analytical results.

In Chapter 7, we seek to characterize the estimation performance of a sensor network where the individual sensors exhibit the phenomenon of *drift*, i.e., a gradual change of the bias. Though estimation in the presence of random errors has been extensively studied in the literature, the loss of estimation performance due to systematic errors like drift have rarely been looked into. In Chapter 7, we derive closed-form bounds on the estimation accuracy of drift-corrupted signals. We assume a polynomial time-series as the representative signal and an autoregressive process model for the drift. When the Markov parameter for drift  $\rho < 1$ , we show that the first-order effect of drift is asymptotically equivalent

to scaling the measurement noise by an appropriate factor. For  $\rho = 1$ , i.e., when the drift is non-stationary, we show that the constant part of a signal can only be estimated inconsistently (non-zero asymptotic variance). Practical usage of the results are demonstrated through the analysis of bandwidth limited multi-sensor network communicating only quantized observations.

Chapter 8 summarizes the research work and the results presented in this dissertation. Some concluding remarks and suggestions for future work are also provided.

## 1.3 Bibliographic notes

Most of the research work appearing in this dissertation are either already published or in several stages of publication at various venues, as listed below.

### Chapter 3:

- [42] S. Kar, H. Chen, and P. K. Varshney, “Optimal identical binary quantizer design for distributed estimation,” *Signal Processing, IEEE Transactions on*, 60(7):3896–3901, July 2012.

### Chapter 4:

- [46] S. Kar, P. K. Varshney, and H. Chen, “Spatial whitening framework for distributed estimation,” In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 Proc. 4th IEEE International Workshop on*, pages 293–296, Dec. 2011.

### Chapter 5:

- [45] S. Kar and P. K. Varshney, “On linear coherent estimation with spatial collaboration,” In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1448–1452, July 2012.

- [43] S. Kar and P. K. Varshney, “Linear coherent estimation with spatial collaboration,” available online <http://dx.doi.org/10.1109/TIT.2013.2248876>, to appear in *IEEE Transactions on Information Theory*.

## **Chapter 6:**

- [44] S. Kar and P. K. Varshney, “Controlled collaboration for linear coherent estimation in wireless sensor networks,” In *Proc. 50th Annual Allerton Conference on Communication, Control and Computing 2012*, Monticello, IL, Oct. 2012.

## **Chapter 7:**

- [47] S. Kar, P. K. Varshney, and M. Palaniswami, “Cramér-Rao bounds for polynomial signal estimation using sensors with AR(1) drift,” *Signal Processing, IEEE Transactions on*, 60(10):5494–5507, Oct. 2012.

# Chapter 2

## Background and literature survey

Estimation in resource constrained sensor networks usually involves some local processing/compression at the sensor nodes. If the act of local processing is done independently by all nodes, then the problem is called distributed estimation. On the other hand, if the local processing involves cooperation with other neighboring sensors, the problem is called collaborative estimation. While the research on distributed estimation is fairly mature, collaborative estimation is a relatively new area of research. In this thesis, we make contributions to both of these areas. Specifically, Chapters 3 and 7 deal with distributed estimation while Chapters 4, 5 and 6 addresses some collaborative estimation problems. In this chapter, we formalize the problem of distributed estimation and present a methodical overview of the literature on this topic. The literature on collaborative estimation will be presented later in the appropriate chapters.

### 2.1 System description

Consider a dense sensor network that includes  $N$  sensor nodes and a fusion center (FC) to observe and estimate an unknown parameter  $\theta$ . The general distributed estimation architecture consists of several blocks, as depicted in Figure 2.1. The signal of interest  $\theta$  is observed through noise-corrupted observations at each sensor node. For the  $n^{\text{th}}$  sensor node, we denote the noise as  $\epsilon_n$  and observation as  $x_n$ . Because of transmission-power

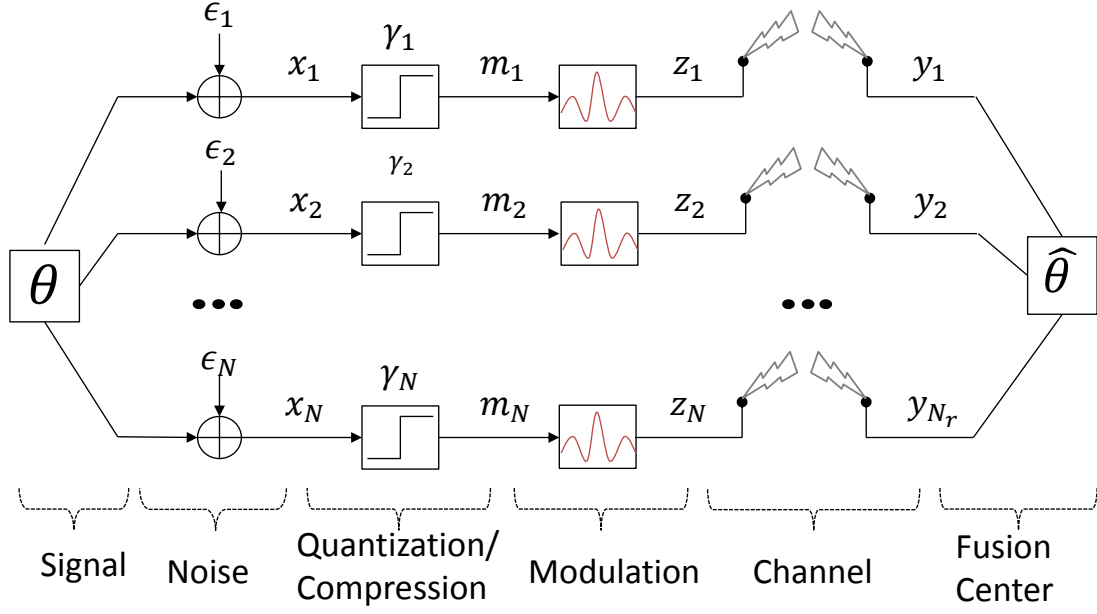


Figure 2.1: System diagram of decentralized estimation.

constraints at the sensor nodes, these observations must undergo power-efficient transformations before they are transmitted through wireless channels. The transformation can be digital (quantization followed by modulation) or analog (e.g., amplify and forward). We denote the transformed variable as  $z_n$ . In case of quantization, we denote the quantizer function as  $\gamma_n(\cdot)$  and the quantized message as  $m_n$ . The received signal at the FC is denoted by the  $N_r$ -dimensional signal  $\mathbf{y} = [y_1, y_2, \dots, y_{N_r}]$ . Note that for orthogonal channels,  $N_r = N$ , so that each sensor can be thought of having separate channels for themselves. The FC then applies a fusion-rule to combine the observations in  $\mathbf{y}$  and obtain an estimate  $\hat{\theta}$ .

## 2.2 Past research

The entire distributed estimation system presents a rich field for potential research and several aspects of this problem have been considered by researchers, especially over the past decade. It should be mentioned here that the literature on distributed estimation is too diverse and in this chapter, we focus only on estimation methods where all sensors fully



observe the parameter of interest and there is a fixed fusion-center. For example, we do not survey distributed estimation of a random field [4, 21], or distributed Karhunen-Loève transform [29]. Both are examples where each sensor observes the parameter of interest only partially. We also exclude consensus algorithms [18] and estimation fusion in a chain-type topology [40] from this survey, both being examples of systems without a fixed fusion center. Another active research area that is not surveyed is dimensionality-reduction (in case of a vector parameter) as a preprocessing step before quantization [22].

In Table 2.1, we attempt to summarize the research progress so far, listing several references and the frameworks addressed therein, as relevant to the research in this thesis. The literature is summarized with respect to the specific assumptions and schemes considered under each architectural block, (namely signal, noise, quantization, modulation, channel and fusion-rule) and the design problems that are addressed. The literature on distributed estimation is broadly split into two groups based on the communication framework used - *digital* and *analog*.

- In the *analog communication framework*, a general *non-linear transformation*  $z_n = h_n(x_n)$  is applied to the observations which are then directly transmitted. If the transformation  $h_n(\cdot)$  is linear, then it is also known as the *amplify and forward* scheme.
- The *digital communication framework* is implemented through quantization and modulation, where the unit of information is *bits*. The  $n^{\text{th}}$  sensor quantizes its observation  $x_n$  into  $b_n$  bits and uses  $P_n$  units of power to transmit the signal using a modulation scheme (e.g., BPSK). The transmitted signal is then attenuated and noise-corrupted as it passes through the channel and this may incur an erroneous reception. Hence there are two venues where loss of information can potentially take place - first during quantization and second during noisy reception. Sometimes, while analyzing a distributed estimation application, the distortion due to erroneous reception is ignored. This assumption, which is only valid in the high-power regime, sub-classifies the body of literature on digital-communication based distributed estimation, as depicted in Table 2.1.

Comparative studies of these two (analog and digital) schemes can be found in [26, 28, 29] and we will not describe this in detail here. The summary is that though the analog framework can provide better energy-distortion (system-wide) tradeoff for several simple examples, a digital framework is often preferred from an implementation perspective. This is because several practical aspects like error-correction techniques, resistance to channel impairments, multiple-access strategies and security are better handled in a digital framework [31].

The objective of distributed estimation problems is to minimize the overall distortion of the estimated parameter at the FC, subject to various considerations, e.g., limited resources, robustness to failure, etc. Due to stringent power and bandwidth limitations at each sensor node, several problems address the efficient allocation of such resources. We call them *bit-allocation* or *power-allocation* problems. The problem of how to optimally quantize/compress/transform an observed signal to meet a pre-specified bandwidth constraint is also an interesting problem. This is coupled with the problem of how to optimally/easily estimate the parameter in the FC, and we refer to these problems as *quantizer design* or *estimator design*. Some research questions are aimed at understanding the performance of a specific quantization or transmission scheme, we call them *performance-analysis* problems.

We next provide a taxonomy for the various schemes under each system block (signal, noise, quantization, modulation, channel and fusion-rule) that are mentioned in Table 2.1. For ease of reference, we assign each of the schemes/assumptions an identifying-name, which we subsequently use in Table 2.1.

Signal	Noise	Quantization/ Compression	Modulation	Channel	Fusion Rule	Objective	Ref.
Centralized Case							
unknown	density	full-precision	noiseless		MLE	optimal fusion	[49]
density	density				MMSE		
unknown	covariance				BLUE	optimal linear fusion	
variance	covariance				LMMSE		
Analog Communication Framework							
variance	ind-variance	amplify-and-forward		path-loss, MAC-AWGN	LMMSE	power-allocation	[96]
variance	ind-variance			path-loss, AWGN	LMMSE		[14]
joint-covariance				path-loss, AWGN	LMMSE	[20]	
density	iid-density			nonlinear-transformation	MAC-AWGN	MMSE	non-linearity design
Digital Communication Framework: Error-free Reception							
range	iid-Gaussian	binary-dithering, binary,multi-level	noiseless		MLE	quantizer and estimator design	[69]
range Gaussian	iid-Gaussian, iid-unknown	binary, multi-level			MLE, moment-based		[73],[74]
range	iid-range	binary-equispaced			moment-based	[58],[57]	
range	iid-M-moment	identical-binary- probabilistic			MLE	[9]	
range	zero-noise				performance analysis	[10]	
range	ind-variance-range	uniform-probabilistic	noiseless	path-loss,AWGN	max-quant-BLUE	bit-allocation	[53]
						power-allocation	[95]
	covariance-range		MQAM,variable-const MQAM,fixed-const MQAM,variable-const MQAM,fixed-const		unquantized-BLUE		[54]
							[51]
Digital Communication Framework: Erroneous Reception							
unknown	iid-density	identical-binary	BPSK	identical-AWGN	MLE	quantizer and	[3]
	Gaussian	identical-binary		Rayleigh,AWGN	moment-based	estimator design	[94]

Table 2.1: Summary of literature in distributed estimation with a fixed fusion center.

## 2.3 Taxonomy

### 2.3.1 Signal and noise

Depending on the particular application, several models of signal are assumed in the literature. The estimation application can either be single-snapshot or dynamic. In the single-snapshot problem, energy-constrained spatial sampling is performed at one particular instant and the inference is made using those samples. For the dynamic scenario, the goal is to estimate a time-varying parameter process for all time instants. In contrast to the single snapshot framework, this involves obtaining multiple samples in time and computing the filtered estimates for any desired time instants, including time instants where observation samples are not available.

The signal of interest may either be constant or vary slightly over the spatial field of interest. Noise captures both the randomness associated with the sensing circuitry (which is uncorrelated across sensors) and the error due to unexplained variables describing the phenomena of interest (which is likely to be correlated across sensors). In some applications, the phenomena may be best described by a joint model of the signal and noise. In addition, the signal and noise can have temporal variations.

#### Signal models (single-snapshot/static)

- *unknown*: Unknown deterministic scalar signal.  $\theta \in \mathbb{R}$ .
- *range*: Unknown deterministic scalar signal with a known range.  $\theta \in [-U, U]$ .
- *variance*: Random scalar signal with known variance.  $\mathbb{E}(\theta) = 0$ ,  $\mathbb{E}(\theta^2) = \eta^2$ .
- *density*: Random scalar signal with known probability density (equivalently distribution) function.  $f_{\Theta}(\theta)$ .  $\theta \sim f_{\Theta}(\cdot)$ .
- *Gaussian*: Random scalar signal that is Gaussian distributed.  $\theta \sim \mathcal{N}(0, \eta^2)$ .
- *polynomial*: Deterministic signal with known time dynamics.  $x_t = \sum_{p=0}^P \theta_p t^p + \epsilon_t$ ,  $\boldsymbol{\theta} \in \mathbb{R}^{P+1}$ .

## Signal models (time-varying/dynamic)

- *Ornstein-Uhlenbeck (OU) process*: Gaussian, temporally correlated, scalar signal.  
 $\theta_t \sim \mathcal{N}(0, \eta^2), \mathbb{E}[\theta_{t_1, t_2}] = \eta^2 \exp(-|t_1 - t_2|/\tau).$

## Noise Models

- *zero-noise*: Perfect observation model that observes a noiseless signal.  $\epsilon_n = 0$ .
- *iid-unknown*: Noise is independent and identically distributed but the noise distribution that is not known either to the FC or to the sensors themselves.  $\epsilon_n \sim f(w)$ , but pdf  $f(w)$  is unknown.
- *iid-M-moment*: Noise is independent and identically distributed but the noise distribution is known only in terms of the first  $M$  moments.  $\mathbb{E}(\epsilon_n) = 0$ ,  $\mathbb{E}(\epsilon_n^j) = \mu_j$  for  $1 \leq j \leq M$  and all  $n$ .
- *iid-density*: Noise is independent and identically distributed with known pdf.  $\epsilon_n \sim f(w)$ .
- *iid-range*: Noise is independent and identically distributed with known pdf and also range-limited.  $\epsilon_n \sim f(w)$  and  $\epsilon_n \in [-R, R]$ .
- *iid-Gaussian*: Noise is independent and identically distributed with Gaussian pdf.  $\epsilon_n \sim N(0, \sigma^2)$ .
- *Gaussian*: Noise vector  $\epsilon$  is jointly (multivariate) Gaussian and possibly correlated.  $\epsilon \sim \mathcal{N}(0, \Sigma)$ ,  $\Sigma$  is positive definite.
- *ind-variance-range*: Noise terms  $\epsilon_n$  are independent (but possibly heterogenous) and range-limited with known variance and range. The covariance matrix is hence diagonal.  $\mathbb{E}(\epsilon_n^2) = \sigma_n^2$  and  $\epsilon_n \in [-R, R]$ .
- *ind-variance*: Noise terms  $\epsilon_n$  are independent (but possibly heterogenous) with known variance.  $\mathbb{E}(\epsilon_n^2) = \sigma_n^2$ .

- *covariance*: Noise terms  $\epsilon_n$  are possibly correlated with known covariance matrix.  $\mathbb{E}(\epsilon\epsilon') = \Sigma$ .
- *covariance-range*: Noise terms  $\epsilon_n$  are possibly correlated and range-limited with known covariance matrix and range.  $\mathbb{E}(\epsilon\epsilon') = \Sigma$  and  $\epsilon_n \in [-R, R]$  for all  $n$ .
- *white Gaussian+OU process*: Noise terms  $\epsilon_{t,n}$  are spatially independent but temporally correlated.  $\epsilon_{t,n} = d_{t,n} + w_{t,n}$ , where  $w_{t,n}$  is temporally independent but  $d_{t,n}$  is a temporally correlated OU process.

### Joint Signal and Noise Models

- *joint-covariance*: Signal  $\theta$  and noise  $\epsilon_n$  are both zero-mean scalar random variables and are characterized by joint second-order moments.  $\mathbb{E}(\theta^2) = \eta^2$ ,  $\mathbb{E}(\epsilon\theta) = \Sigma_{\epsilon,\theta}$  and  $\mathbb{E}(\epsilon\epsilon') = \Sigma$ .

### 2.3.2 Quantization

As mentioned earlier, quantization/compression is an integral part of digital-communication based implementation of distributed estimation systems. An  $L$ -level quantization is a transformation  $\gamma(\cdot)$  whose domain is a continuous set and range is a discrete-set with cardinality  $L$ . The number of bits of information required to represent such a discrete variable is  $b = \log_2(L)$ . The case of  $L = 2$  or  $b = 1$  corresponds to binary quantization, and is widely researched because of its analytical tractability and also the fact that it represents the most-stringent constraint on bandwidth and serves as a worst-case analysis. When a universal and robust design is desired, the quantizer functions  $\gamma_n(\cdot)$  may be constrained to be identical, though it is well known that this comes at the cost of higher estimation error, especially for small magnitudes of  $L$ . For quantization procedures that incur unequal distortion for different points in the observation domain, addition of some noise prior to quantization *levels the playing field*, i.e., decreases the worst-case distortion at the cost of increasing the best-case distortion. Hence schemes like dithering or more generally

probabilistic quantization are often desirable for problems that minimize the maximum-distortion.

- *binary*: This refers to binary threshold-quantizers of the form

$$m_n = \gamma_n(x_n) = \begin{cases} 1 & x_n \geq \tau_n \\ 0 & x_n < \tau_n \end{cases}, \quad (2.1)$$

where each sensor can possibly have different thresholds  $\tau_n$ .

- *identical-binary*: All the sensors implement  $\tau_n = \tau, \forall n$ .
- *binary-dithering*: This is the binary non-subtractive dithering scheme. A random input  $v_n$  is generated from some pdf  $f_n^{(d)}(v_n)$  and is added to the observation  $x_n$  prior to *binary* quantization to reduce the worst case performance.
- *identical-binary-probabilistic*: This is the absolute generalization of the binary threshold and dithering quantizers. The quantizer is defined by the function  $\gamma_n(x_n) = P(m_n = 1|x_n)$  defined on the domain of the observation. Design of binary-probabilistic quantizers is more difficult than the threshold quantizers because this involves finding the function  $\gamma_n(\cdot)$  rather than a threshold  $\tau_n$ . Owing to this difficulty, research so far has only been able to address zero-noise and identical quantizer scenarios.
- *binary-equispaced*: Each sensor performs binary quantization with different thresholds  $\tau_n$ . These threshold values are uniformly spaced in the range of the observation, which is considered known.

- *multi-level*: A multi( $L$ )-level quantizer is of the form

$$m_n = \gamma_n(x_n) = \begin{cases} 1 & \tau_{n,1} \leq x_n \\ 2 & \tau_{n,2} \leq x_n < \tau_{n,1} \\ \dots & \dots\dots\dots \\ L & x_n < \tau_{n,(L-1)} \end{cases}, \quad (2.2)$$

As mentioned earlier, if  $L$  is large enough, having identical quantizers ( $\tau_{n,l} = \tau_l, \forall l$ ) does not adversely affect the performance. Hence multiple threshold quantizers are almost always assumed to be identical in most problem formulations.

- *uniform-probabilistic*: This refers to the case of non-identical *multi-level* quantizers with  $L_n = 2^{b_n}$  levels with uniform spacing between the thresholds. If observation  $x_n$  is assumed to be bounded to a finite interval  $[-U, U]$ , the quantization points  $\tau_{n,l} \in [-U, U], l = 1, \dots, 2^{b_n}$  are uniformly spaced such that  $\tau_{n,l} - \tau_{n,(l+1)} = 2U/(2^{b_n} - 1) \triangleq \Delta_n$ . Suppose that  $x_n \in [\tau_{n,(l+1)}, \tau_{n,l})$ . Then  $x_n$  is dither-quantized to either  $\tau_{n,(l+1)}$  or  $\tau_{n,l}$  according to

$$P(m_n = \tau_{n,(l+1)}) = q, \quad P(m_n = \tau_{n,l}) = 1 - q, \quad (2.3)$$

where  $m_n$  is the resulting message and  $q = (\tau_{n,l} - x_n)/\Delta_n$ .

### 2.3.3 Modulation for digital communication

There are usually two main choices to make while implementing a modulation scheme for digital communication - where to embed the information in the carrier signal (amplitude, phase or frequency) and what is the constellation size ( $2^{b_n}, b_n \geq 1$ ) [31]. Though frequency-based schemes have better resistance to channel impairments like fading and interference, amplitude/phase-based schemes are more bandwidth efficient. Also, modulations with large constellations have higher data rate for a given signal bandwidth, but are more susceptible to noise, fading and hardware imperfections. In Table 2.1, *BPSK* refers to the binary phase



shift keying scheme and *MQAM* refers to the quadrature amplitude modulation with  $M$  points in the signal constellation.

- *variable-const*: For an observation  $x_n$  quantized using  $b_n$  bits with  $b_n \geq 2$ , sensor  $n$  can choose to transmit all  $b_n$  bits in one symbol or choose to transmit  $b_n/b_0$  separate symbols, each consisting of  $b_0$  bits. Since the sensors can have potentially different values of  $b_n$ , the former case will be referred to as the *variable-const* strategy.
- *fixed-const*: On the other hand, if all the sensors choose to transmit with some pre-agreed  $b_0$ , we call it the *fixed-const* strategy. In such a case, the bit-allocation problem and power-allocation problem are the same.

### 2.3.4 Channel models

The digital information is transmitted to the FC over a communication channel using  $P_n$  units of power through a modulation scheme. The carrier-based modulation-demodulation process is often abstracted out while analyzing the error performance. For additive-white-gaussian (*AWGN*) channel, the error is represented as Gaussian noise  $v_n$  that is independent across time and space (sensors) and added to the unmodulated signal (in its constellation-form)  $z_n$ . For example, in binary pulse-amplitude-modulation (*BPAM*) or binary phase-shift-keying (*BPSK*), the received signal is modeled as  $y_n = z_n + v_n, z_n \in \{-\sqrt{P_n}, \sqrt{P_n}\}$  and  $v_n \sim \mathcal{N}(0, \xi_n^2)$  where  $\xi_n^2$  is related to the power spectral density of the underlying noise process. Sometimes the modulated signal incurs a path-loss before reaching the receiver. The effective power reaching the FC is often modeled as  $g_n P_n$ , where  $g_n$  is the channel gain. The channel-gain  $g_n$  may be known either completely or only through some statistical models. The design of optimum receivers and their error-analysis depends on the the nature of channel and type of modulation. Extensive studies have been done on these aspects for the additive Gaussian Noise channel and Rayleigh fading channel [72, 31].

- *path-loss*: In case of static channels  $g_n > 0$  may be estimated before use and may be assumed to be known. In a spatial setting, an often used model for path-loss is

$g_n = d_n^{-\kappa}$  where  $d_n$  is the distance of sensor  $n$  from the fusion center and  $\kappa$  is a known path-loss exponent, approximately in the range  $\kappa \in [2, 5]$ .

- *Rayleigh*: In case of slow-fading channels, the gain  $g_n$  is modeled as a random variable. For Rayleigh-fading channels,  $g_n$  is modeled as an exponential distribution with mean  $\mu_g$ .  $P(g_n = g) = \frac{1}{\mu_g} \exp(-g/\mu_g), g \in [0, \infty)$ .
- *identical-AWGN*: The noise process for all the channels are identical.  $\xi_n^2 = \xi^2, \forall n$ .
- *MAC-AWGN*: Sometimes in analog communication and also in digital communication scenario where sensors coherently uses the same channel, the received signal is modeled as  $y = \sum_{n=1}^N \sqrt{g_n} z_n + v$ , where the additive noise  $v$  is Gaussian distributed.  $v \sim \mathcal{N}(0, \xi^2)$ .

### 2.3.5 Fusion-rules and distortion metrics

The fusion center usually has access to received values  $y_n$ , which are the quantized and noisy versions of the original observation  $x_n$ . To estimate the parameter  $\hat{\theta}$  from  $\mathbf{y} = [y_1, y_2, \dots, y_n]$ , we need to know how the statistics of  $\mathbf{y}$  (pdf, mean, covariance, etc.) relate to the parameter of interest  $\theta$ . Depending on the amount of information about  $\theta$  contained in  $\mathbf{y}$  and the amount of computation that can be carried out in a realistic time-frame, the FC chooses to implement a particular fusion rule (or estimator). The topic of estimation theory is extensively discussed in the textbooks [49, 81].

- *MLE*: If  $\theta$  is a deterministic but unknown variable and the received signal vector has pdf  $\mathbf{y} \sim p(\mathbf{y}; \theta)$ , the maximum-likelihood estimator (MLE) is defined as

$$\hat{\theta}_{ML} \triangleq \arg \max_{\theta} p(\mathbf{y}; \theta). \quad (2.4)$$

MLE, if unique, is asymptotically unbiased and attains the Cramer-Rao Lower Bound (CRB), which is the inverse of the Fisher information of  $\theta$  in  $\mathbf{y}$ .

$$\begin{aligned} \mathbb{E}_{\mathbf{y};\theta}(\hat{\theta}_{ML} - \theta)^2 &\geq J_F^{-1}(\theta) \triangleq \text{CRB, where} \\ J_F &= -\mathbb{E}_{\mathbf{y};\theta} \left\{ \nabla_{\theta} [\nabla_{\theta} p(\mathbf{y}; \theta)]^T \right\}. \end{aligned} \quad (2.5)$$

For example, if  $\mathbf{y} \sim \mathcal{N}(\mathbf{H}\theta, \mathbf{\Sigma})$ , then

$$\begin{aligned} J_F &= \mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H}, \text{ and} \\ \hat{\theta}_{ML} &= J_F^{-1} \mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{y}. \end{aligned} \quad (2.6)$$

- *BLUE*: If  $\theta$  is a deterministic unknown random variable and the mean and covariance-matrix of the received signal vector is known to be  $\boldsymbol{\mu}_{\mathbf{y}}(\theta)$  and  $\mathbf{\Sigma}_{\mathbf{y}\mathbf{y}}(\theta)$ , then the best-linear-unbiased estimator (BLUE) can be thought of as the *MLE* of  $\theta$ , had  $\mathbf{y}$  actually followed a Gaussian distribution with the above mean and covariance. In general, BLUE may be quite suboptimal compared to MLE, but it can often be convenient to use and is easily computed. Of course, if the original pdf of  $\mathbf{y}$  is Gaussian, then MLE and BLUE are identical and optimal.
- *MMSE*: If  $\theta$  is an unknown random variable and combined with the received signal vector, the random vector  $[\theta, \mathbf{y}]$  has joint-pdf  $p(\mathbf{y}, \theta)$ , the minimum-mean-square estimator (MMSE) minimizes the Bayes risk for a quadratic cost function and is defined as

$$\hat{\theta}_{MSE} \triangleq \mathbb{E}(\theta|\mathbf{y}) = \int \theta p(\theta|\mathbf{y}) d\theta. \quad (2.7)$$

The performance of the MMSE estimator can be bounded by the Bayesian Cramér-Rao bound (BCRB), which is the inverse of the sum of prior and expected Fisher

information,

$$\begin{aligned}
\mathbb{E}_{\mathbf{y},\theta}((\hat{\theta}_{MSE} - \theta)^2) &\geq J_B^{-1} \triangleq \text{BCRB}, \text{ where} \\
J_B &= J_P + J_D, \text{ and} \\
J_P &= -\mathbb{E}_{\theta} \left\{ \nabla_{\theta} [\nabla_{\theta} p(\theta)]^T \right\}, \\
J_D &= -\mathbb{E}_{\theta} \left\{ \mathbb{E}_{\mathbf{y}|\theta} \left\{ \nabla_{\theta} [\nabla_{\theta} p(\mathbf{y}|\theta)]^T \right\} \right\}.
\end{aligned} \tag{2.8}$$

For example, if  $\theta \sim \mathcal{N}(\theta_0, \eta^2)$  and  $\mathbf{y}|\theta \sim \mathcal{N}(\mathbf{H}\theta, \Sigma)$ , then

$$\begin{aligned}
J_P &= \eta^{-2}, \\
J_D &= \mathbf{H}^T \Sigma^{-1} \mathbf{H}, \text{ and} \\
\hat{\theta}_{MSE} &= \theta_0 + J_B^{-1} \mathbf{H}^T \Sigma^{-1} (\mathbf{y} - \mathbf{H}\theta_0).
\end{aligned} \tag{2.9}$$

- *LMMSE*: If  $\theta$  is an unknown random variable and the mean and covariance-matrix of the vector  $[\theta, \mathbf{y}]$  are known, then the linear-minimum-mean-square estimator (LMMSE) can be thought of as the *MMSE* estimator of  $\theta$ , had  $[\theta, \mathbf{y}]$  actually followed a Gaussian distribution with the above mean and covariance. As with BLUE, in general, LMMSE may often be quite suboptimal compared to MMSE, but it is often convenient to use because of computational ease. Of course, if the original pdf of  $[\theta, \mathbf{y}]$  is Gaussian, then MMSE and LMMSE are identical and optimal.

Sometimes, for problems when the MLE (for deterministic  $\theta$ ) or MMSE (for random  $\theta$ ) estimator is difficult to compute and the performance of BLUE or LMMSE estimator is not satisfactory, estimators are derived based on comparing one or more easily available summary-statistics of  $\mathbf{y}$ . This is a rather broad class of estimators and sometimes called *moment-based* estimators.

Sometimes when  $\theta$  is deterministic and unknown, after quantization, the covariance-matrix of  $\mathbf{m}$  may depend on the actual value of  $\theta$  and hence evaluation of BLUE estimator becomes difficult. In those cases, a new covariance matrix is formed where each entry in the covariance matrix is replaced by the maximum possible value in the domain of  $\theta$ . The

BLUE corresponding to the actual mean and this new covariance matrix will be called *max-quant-BLUE* estimator. The BLUE corresponding to the actual mean of  $\mathbf{y}$  and the covariance of the full-precision observations  $\mathbf{x}$  will be called *unquantized-BLUE* estimator.

## 2.4 Major contributions

Having described the background on distributed estimation research, it is worthwhile to put in perspective our contributions on this topic that are presented in this thesis. Chapters 3 and 7 describe our contributions on distributed estimation. While Chapter 3 focuses on the design of identical binary quantizers for range-limited signal estimation, Chapter 7 provides the performance analysis for polynomial signal estimation using drift-corrupted (and possibly quantized) sensor data.

Our contributions on the collaborative estimation problem (which will be formalized later) appear in Chapters 4, 5 and 6. While Chapter 4 addresses the problem of bit-allocation in a digital communication based framework, Chapters 5 and 6 deal with energy/power allocation in an analog amplify-and-forward communication framework. In Chapter 5, we consider a single-snapshot estimation problem where the goal is to estimate a static parameter based on observations at only one particular instant. In Chapter 6, we consider a time-varying signal estimation problem where the goal is to estimate a random process (rather than a parameter) based on observations obtained over a period of time.

The aforementioned major contributions are summarized in Table 2.2.

Signal	Noise	Collaboration	Quantization/ Compression	Modulation	Channel	Fusion Rule	Objective	Ref.
Our Contributions								
range	density	No	identical-binary-probabilistic	noiseless		MLE	Quantizer design	Chapter 3 [42]
unknown	covariance	Yes	uniform+probabilistic	noiseless		Quasi-MLE	collaborative-strategy bit-allocation	Chapter 4 [46]
Gaussian	Gaussian	Yes	amplify-and-forward		coherent-MAC	MLE	collaboration strategy power-allocation	Chapter 5 [43]
OU process	Gaussian	Yes	amplify-and-forward		coherent-MAC	MLE	performance analysis	Chapter 6 [44]
polynomial	white Gaussian+ OU process	No	uniform-quantization	noiseless		Quasi-MLE	performance analysis	Chapter 7 [47]

Table 2.2: Summary of our contributions on distributed and collaborative estimation.

# Chapter 3

## Quantizer design for distributed estimation

### 3.1 Introduction

In this chapter, we address the problem of identical binary-probabilistic *quantizer design* for distributed estimation. We consider a distributed estimation problem with  $N$  sensors collecting noisy observations of an unknown but fixed scalar parameter  $\theta$  such that the local sensor observations  $\mathbf{X} = [X_1, X_2, \dots, X_N]'$  are independent and identically distributed (i.i.d.), i.e.,  $f(\mathbf{X}, \theta) = \prod_{n=1}^N f(X_n, \theta)$ , where  $f(\mathbf{X}, \theta)$  and  $f(X_n, \theta)$  are known probability density functions (pdf). One example of such a model is the location estimation problem with additive noise,

$$X_n = \theta + \epsilon_n, \quad 1 \leq n \leq N, \quad (3.1)$$

where the noise samples  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]'$  are zero-mean, additive, independent, and identically distributed with symmetric pdf  $f(w)$  and variance  $\sigma^2$ . In many practical applications, the dynamic range of  $\theta$  is often assumed to be known, such that  $\theta \in [\theta_0 - \Delta, \theta_0 + \Delta]$  where  $\theta_0$  and  $\Delta$  are known constants. Without loss of generality, we assume  $\theta_0 = 0$  and  $\Delta = 1$  and confine our attention to  $\theta \in [-1, 1]$  in the rest of this chapter.

As an application of this problem, one can consider an environmental monitoring system consisting of a central base station communicating with multiple thermal sensors with limited energy deployed over a region in a dense manner, so that they are more-or-less recording the same temperature at any given time. The redundancy in the number of sensors serves to increase robustness of the network, share the power resources and increase the lifetime of the monitoring system. The objective of the sensor network is to monitor the temperature in the region throughout the day, though the diurnal temperature variation (say  $10 - 40^\circ\text{C}$ ) is roughly known. The system diagram is provided in Figure 3.1.

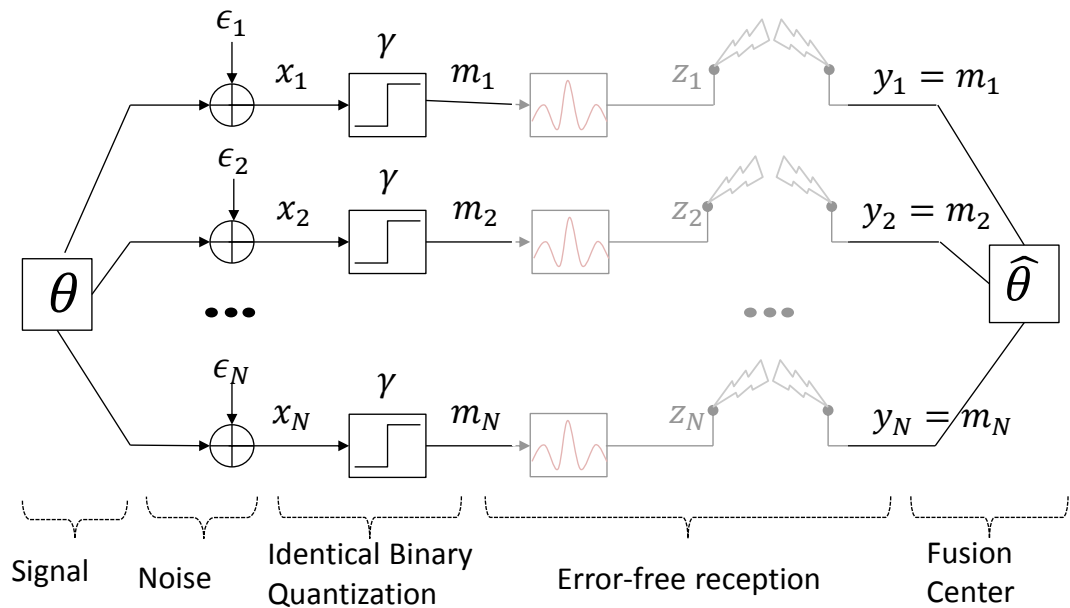


Figure 3.1: System diagram of identical binary quantization based distributed estimation.

### 3.1.1 Identical one-bit quantizers

Since the channel capacity of links between sensors and the fusion center and the energy resources for transmission in the battery-powered sensor nodes can be severely limited, we assume that each sensor performs a binary quantization and transmits only one-bit of information to the FC. With an appeal to symmetry, each sensor is designed to employ an identical quantization rule.



A one-bit quantizer can be defined as a mapping from the observation space  $\mathbb{R}$  to a symbol set of size 2, say  $\{\mathcal{S}_0, \mathcal{S}_1\}$ . Such a mapping can be expressed in two forms. In the often used *function-form* description, a quantizer explicitly maps its input  $X_n$  to the output  $Y_n$  through a function  $\varphi : \mathbb{R} \rightarrow \{\mathcal{S}_0, \mathcal{S}_1\}$ . For example, the function-form description of a zero-threshold quantizer is

$$Y_n = \varphi_T(X_n) \triangleq \begin{cases} \mathcal{S}_1, & X_n \geq 0 \\ \mathcal{S}_0, & X_n < 0 \end{cases}, \forall n. \quad (3.2)$$

Alternatively, in the *probability-form* description, a quantizer is defined as the conditional probability  $\gamma : \mathbb{R} \rightarrow [0, 1]$  of the output being a particular symbol (say  $\mathcal{S}_1$ ) given an input  $X_n$ ,

$$\gamma(X_n) \triangleq P(Y_n = \mathcal{S}_1 | X_n), \forall n. \quad (3.3)$$

For example, the equivalent probability-form description of the zero-threshold quantizer (3.2) is

$$\gamma_T(X_n) \triangleq \begin{cases} 1, & X_n \geq 0 \\ 0, & X_n < 0 \end{cases}, \forall n, \quad (3.4)$$

which we will refer to as the *Threshold Quantizer*. In this chapter, we will use the probability-form description (3.3) for analysis and subsequent design of quantizers. Here by allowing  $\gamma$  to take any value between 0 and 1, we consider all possible local quantization rules [10], i.e., the quantization rule can be either deterministic (e.g., single threshold quantizer [73],[89]) or probabilistic (e.g., dithered quantizer, i.e., some noise added to the signal before quantization [33],[69]).

### 3.1.2 Error-free reception

In this chapter, we assume that the stringent source-rate constraint (1-bit per observation) frees up resources so that adequate channel-coding is undertaken to counter noise/fading phenomena in the communication channel. As a result, the compressed information in  $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]$  is assumed to be obtained in a lossless fashion at the FC, an assumption that is consistent with several previous research contributions on this topic [73, 58]. However, in the more general scenario, absence of sufficient resources for adequate channel-coding may result in lossy transmission of  $\mathbf{Y}$  - an issue also considered by several researchers [3, 94]. The results in this chapter can be extended to the noisy channel scenario.

### 3.1.3 Performance metric

The goal of the fusion center is to use the quantized observations and obtain an estimate of the location parameter  $\hat{\theta}$  using an estimator  $h(\mathbf{Y}, \gamma)$ . The problem setup is summarized below by the Markov chain,

$$\theta \xrightarrow{f(w)} \mathbf{X} \xrightarrow{\gamma} \mathbf{Y} \xrightarrow{h} \hat{\theta}. \quad (3.5)$$

Let  $\hat{\theta}$  be an unbiased estimator of  $\theta$ . It is well known that the variance of any unbiased estimator is lower bounded by the Cramér-Rao lower bound (CRB) and that the CRB is asymptotically achieved by using the Maximum-Likelihood (ML) estimator (see [49]). Let  $g(\theta)$  denote the probability that the quantizer output is  $\mathcal{S}_1$  when the original parameter is  $\theta$ ,

$$\begin{aligned} g(\theta) &= P(Y_n = \mathcal{S}_1 | \theta) = \mathbb{E}_{\epsilon_n}(\gamma(\theta + \epsilon_n)) \\ &= \int_{-\infty}^{\infty} \gamma(x) f(x - \theta) dx. \end{aligned} \quad (3.6)$$

Then the sample mean  $\bar{\mathbf{Y}} = \frac{1}{N} \sum_{i=1}^N Y_i$  is the ML-estimate of  $g(\theta)$  and from the functional invariance property, we have

$$\hat{\theta}_{\text{ML}} = g^{-1}(\bar{\mathbf{Y}}). \quad (3.7)$$

For  $N$  independent observations, the variance of  $\hat{\theta}_{\text{ML}}$  satisfies

$$\mathbb{E}\{(\theta - \hat{\theta})^2\} \geq \frac{1}{N} \frac{1}{I(\theta)} \triangleq \frac{1}{N} \text{CRB}(\theta, \gamma, f), \quad (3.8)$$

where  $I(\theta) = -\mathbb{E}_W [\nabla_{\theta}^2 \ln p(Y_1; \theta)]$  is the Fisher Information (FI) for one sensor output and the equality can be achieved asymptotically [89]. In case of binary quantization, FI can be expressed as (see [10]),

$$I(\theta) = \frac{(g'(\theta))^2}{g(\theta)(1 - g(\theta))}. \quad (3.9)$$

In general,  $\text{CRB}(\theta, \gamma; f)$  is a function of the unknown parameter  $\theta$ , i.e., the quantizer  $\gamma$  may result in a high CRB for one  $\theta$  and a low CRB for another. To ensure accurate estimation over the entire parameter range, we use the maximum possible estimation variance or the *maximum-CRB*

$$\phi(\gamma, f) = \sup_{\theta \in (-1, 1)} \text{CRB}(\theta, \gamma, f), \quad (3.10)$$

as our performance metric.

Although it is relatively easy to obtain  $\phi(\gamma, f)$  for a given noise probability distribution function  $f(\cdot)$  and quantization rule  $\gamma$ , the problem of determining

$$\phi(f) = \inf_{\gamma} \sup_{\theta \in (-1, 1)} \text{CRB}(\theta, \gamma, f) \quad (3.11)$$

has been shown to be extremely difficult and remains unsolved [10],[89]. We refer to the minimizer of (3.11) as the *minimax-CRB* quantizer. Our goal in this chapter is to design minimax-CRB quantizers for arbitrary noise densities.

### 3.1.4 Previous work

The problem of quantizer design for minimax-CRB criterion has been addressed only in terms of some scattered results till now. It is well known that the Threshold Quantizer ( $\gamma_T$ , see (3.4)), though widely used in the literature [73],[89], is unsuitable for high-SNR situations because the maximum value of CRB, typically occurring at boundaries, may exponentially increase with decreasing variance  $\sigma^2$  [69]. This problem is often addressed by adding some additional noise (dithering) to the observation prior to threshold-quantization. We refer to this as a *Dithering Quantizer* ( $\gamma_D$ ). Dithering is often necessary only in the high-SNR situations, when the noise variance is below a critical magnitude, say  $\sigma_{\mathcal{F}}^2$ . The critical variance depends on the shape of the noise pdf and is determined by [69]

$$\sigma_{\mathcal{F}} = \arg \inf_{\sigma} \phi(\gamma_T, f(\sigma^2)). \quad (3.12)$$

By design, the Dithering Quantizer has the limitation that the maximum-CRB actually flattens out (does not decrease) below the critical variance, e.g., for Gaussian noise it was shown in [69] that  $\sigma_{\mathcal{F}} \approx 2/\pi$ .

*Zero-noise performance limit and Sine Quantizer* ( $\gamma_0$ ): The performance limit of  $\phi(f)$  for the noiseless situation, i.e., when  $f(w) = \delta(w)$ , was derived in [10]. For such a scenario, the optimum minimax-CRB quantizer and the corresponding performance were shown to

be

$$\gamma_0(x) \triangleq \begin{cases} 1, & x > 1 \\ \frac{1}{2} \left(1 + \sin \frac{\pi x}{2}\right), & x \in [-1, 1] \text{ , and} \\ 0, & x < -1 \end{cases} \quad (3.13)$$

$$\phi_0 \triangleq \frac{4}{\pi^2} \approx 0.4. \quad (3.14)$$

It must be noted here that analogous performance limits for finite-variance noise densities are extremely challenging and their derivation remains an open problem. While the quantizer given by (3.13) is insightful, it has limited applicability due to two reasons, (1) the noiseless scenario can only approximate high-SNR cases and (2) even for high-SNR cases,  $\gamma_0$  may be far from satisfying the minimax property, as we shall show later in this chapter.

In this chapter, we make some significant contributions towards the study of minimax-CRB quantizer design. We define antisymmetric quantizers and restrict our attention within that class. We determine certain conditions under which the shape of the optimal quantizer is greatly simplified, thereby enabling efficient implementation. We then identify a class of noise distributions for which the Threshold Quantizer is optimal. Lastly for other noise distributions, aided by some theoretical insights, we propose a class of piecewise-linear quantizers and formulate the quantizer design problem as one of numerical minimax optimization. The resulting quantizer is shown to perform significantly better compared to all three existing quantizers - namely the Threshold, Dithering and Sine quantizers.

## 3.2 Main results

Before presenting the results, we provide some definitions that will be needed for the subsequent discussion.

**Definition 3.2.1.** *A quantizer  $\gamma(x)$  is admissible if the resulting conditional probability distribution  $g(\theta)$  is monotonically increasing in  $\theta \in (-1, 1)$ .*

The monotonic property is desirable since it ensures that  $g^{-1}(\cdot)$  exists so the ML-estimator (3.7) is well-defined. The increasing property is without loss of generality, since, corresponding to every  $\gamma(x)$ , there is another valid quantizer  $\bar{\gamma}(x) \triangleq 1 - \gamma(x)$  such that  $\bar{g}(\theta) \triangleq \int_{-\infty}^{\infty} \bar{\gamma}(x) f(x - \theta) dx = 1 - g(\theta)$ . This reverses the increasing/decreasing property and yet has the same maximum-CRB, since by (3.9),  $\text{CRB}(\theta, \gamma, f) = \text{CRB}(\theta, \bar{\gamma}, f)$ . Hence it is sufficient that, in pursuit of a minimax-CRB quantizer, we restrict our attention to admissible quantizers. Alternatively, throughout the rest of the chapter, any reference to a minimax-CRB quantizer will imply that it is admissible.

**Definition 3.2.2.** *A quantizer  $\gamma(x)$  is antisymmetric if*

$$\gamma(x) + \gamma(-x) = 1, \quad \forall x. \quad (3.15)$$

It may be noted here that traditional quantizers like the Threshold, Dithering, and Sine quantizers are antisymmetric. It is easy to see that antisymmetric property of  $\gamma(x)$  together with the assumption of symmetric noise pdf  $f(w)$  implies that  $g(\theta)$  is also antisymmetric, i.e.,  $g(\theta) = 1 - g(-\theta)$ . This further means that,

$$\text{CRB}(\theta, \gamma; f) = \text{CRB}(-\theta, \gamma; f), \quad (3.16)$$

which implies that we can reduce the interval of interest in (3.10) by a factor of half, i.e., either  $\theta \in (-1, 0]$  or  $\theta \in [0, 1)$  is sufficient for analysis.

We note here that for an antisymmetric quantizer with symmetric noise pdf,  $g(\theta)$  can be simplified as,

$$g(\theta) = F(\theta) + \int_{-\infty}^0 \gamma(x) \xi(\theta, x) dx, \quad \text{where} \quad (3.17)$$

$$\xi(\theta, x) \triangleq f(x - \theta) - f(x + \theta), \quad (3.18)$$

and  $F(\theta) \triangleq \int_{-\infty}^{\theta} f(w) dw$  is the distribution function.

**Definition 3.2.3.** We call a quantizer  $\gamma_1(x)$  dominant over another quantizer  $\gamma_2(x)$  if

$$CRB(\theta, \gamma_1; f) \leq CRB(\theta, \gamma_2; f), \quad \forall \theta \in (-1, 1). \quad (3.19)$$

Clearly, a dominant quantizer is better in terms of performance, since it ensures a lesser maximum-CRB, i.e.,  $\phi(\gamma_1, f) \leq \phi(\gamma_2, f)$ . As a passing remark, it may be pointed here that the reverse is not necessarily true, i.e., lesser maximum-CRB does not necessarily imply dominance.

**Definition 3.2.4.** A probability density function  $f(w)$  is unimodal if it has only one maxima (at  $w = w_0$ , say), i.e.,  $f'(w) > 0$ , for  $w \in (-\infty, w_0)$  and  $f'(w) < 0$ , for  $w \in (w_0, \infty)$ . For example, commonly used Gaussian and Laplacian noise densities are unimodal.

In certain cases, the support of a minimax-CRB quantizer can be highly restricted. Lemma 3.2.5 lays out such a scenario.

**Lemma 3.2.5.** (Restricting the domain:) Assume the noise density  $f(w)$  to be zero-mean, symmetric and unimodal. Then an antisymmetric minimax-CRB quantizer is at most unit-support in the negative semi-axis, i.e.,

$$\gamma(x) = 0 \text{ for } x < -1. \quad (3.20)$$

To establish Lemma 3.2.5 we show that, for any antisymmetric  $\gamma(x)$ , there exists a unit-support quantizer  $\tilde{\gamma}(x)$  (namely, the trivially truncated quantizer),

$$\tilde{\gamma}(x) \triangleq \begin{cases} 1, & x > 1 \\ \gamma(x), & x \in [-1, 1] \\ 0, & x < -1 \end{cases}, \quad (3.21)$$

that is both antisymmetric and dominant over  $\gamma(x)$ . The full proof is provided in Appendix A.1.

The unit-support property helps make the quantizer structure simpler, which will be key in a subsequent theoretical result as well as our numerical design in Section 3.3. We note here that for an antisymmetric unit-support quantizer with symmetric noise pdf,  $g(\theta)$  and  $g'(\theta)$  can be simplified as,

$$g(\theta) = F(\theta) + \int_{-1}^0 \gamma(x) \xi(\theta, x) dx, \text{ and} \quad (3.22)$$

$$g'(\theta) = f(\theta) + \frac{d}{d\theta} \left\{ \int_{-1}^0 \gamma(x) \xi(\theta, x) dx \right\}. \quad (3.23)$$

In certain cases, the Threshold Quantizer ( $\gamma_T$ ) is also the minimax-CRB quantizer, an example of which is provided in Theorem 3.2.6.

**Theorem 3.2.6.** (Optimality of Threshold Quantizer:) *Assume the noise density  $f(w)$  to be zero-mean, symmetric, unimodal and such that*

$$f'(w - z) + f'(w + z) \leq 0, \text{ for } w \in [0, 1], z \in [0, 1]. \quad (3.24)$$

*Then, the Threshold Quantizer is dominant over all possible antisymmetric quantizers.*

The proof of Theorem 3.2.6 is given in Appendix A.2. This is an important result, since condition (3.24) is satisfied for a wide family of noise densities, including the following example.

**Example 3.2.7.** Gaussian density: *For Gaussian density with variance  $\sigma^2$ , it is easy to see that condition (3.24) holds for  $\sigma^2 \geq 1$  (derivation in Appendix A.3). Therefore, for Gaussian noise with variance  $\sigma^2 \geq 1$ , no probabilistic quantizer (within the antisymmetric class) can decrease the maximum-CRB beyond the Threshold Quantizer.*

We end this section by pointing out a deficiency of the Sine Quantizer that we alluded to in the introduction. We show that the CRB at the boundaries ( $\theta = \pm 1$ ) for vanishingly small variance ( $\sigma^2 \rightarrow 0$ ) is more than twice of that predicted for the noiseless case. The exact degree of sub-optimality depends on the shape of the noise density and is summarized in Proposition 3.2.8 below.



Laplacian	Gaussian	Noiseless case
$16/\pi^2 \approx 1.62$	$4/\pi \approx 1.27$	$4/\pi^2 \approx 0.41$

Table 3.1: Lower bound on maximum-CRB using Sine Quantizers.

**Proposition 3.2.8.** (High-SNR sub-optimality of Sine Quantizer.) *Let  $f(w; \sigma^2)$  denote a family of zero-mean, symmetric noise densities with  $\sigma^2$  signifying the variance. Assume that the moment condition  $\sigma^{-4} \int_{-\infty}^{\infty} w^4 f(w; \sigma^2) dw < \infty$  is satisfied. Then,*

$$\lim_{\sigma^2 \rightarrow 0} \phi(\gamma_0, f(w; \sigma^2)) \geq \lim_{\sigma^2 \rightarrow 0} CRB(\pm 1, \gamma_0; f(w; \sigma^2)) \quad (3.25)$$

$$= \frac{4}{\pi^2} \frac{1}{2\mu_1^2} \quad (3.26)$$

$$> \frac{8}{\pi^2}, \quad (3.27)$$

where  $\mu_1$  is the normalized one-sided mean,  $\mu_1 \triangleq \sigma^{-1} \int_0^{\infty} w f(w; \sigma^2) dw$ .

The proof of Proposition 3.2.8 is provided in Appendix A.4. The bound  $\frac{8}{\pi^2}$  in Proposition 3.2.8 can be compared directly with the theoretical limit  $\frac{4}{\pi^2}$  (3.14) to note that it is twice as large. For illustration, the specific limit in (3.26) for Gaussian and Laplacian pdf is tabulated in Table 3.1 (derivation in Appendix A.5). We will further substantiate these results numerically in Section 3.4. In terms of a low-noise sensing application with a pre-specified allowable distortion, Proposition 3.2.8 quantifies the scope of improvement over Sine Quantizer - by a *judicious* design of quantizer (detailed subsequently in Section 3.3), we can potentially reduce the required number of sensors to half.

Proposition 3.2.8 highlights the sub-optimality of the Sine Quantizer, which necessitates an alternative quantizer design in the high-SNR regime. Even in the moderate-SNR regime, in the absence of concrete analytical results for finite variance scenarios, it is not clear how one should design efficient minimax-CRB quantizers. In the following section, we describe a quantizer design method through direct numerical optimization.

### 3.3 Design of antisymmetric minimax-CRB quantizer

A general probabilistic quantizer  $\gamma(x)$  is any function that maps  $(-\infty, \infty) \rightarrow [0, 1]$ . But numerical search within such a functional space is extremely difficult and hence we make some additional assumptions.

First, the proposed quantizer  $\gamma_P(x)$  is assumed to be antisymmetric, and the noise density is assumed to be symmetric. From Lemma 3.2.5, this also means that it is unit-support. To further simplify the structure, we assume that  $\gamma_P(x)$  is piecewise linear. Hence, we divide the support interval  $[-1, 0]$  into several equally spaced intervals. We choose the observation grid-size  $\Delta_x$  or the number of grid intervals  $K$  so that  $K\Delta_x = 1$ . Define  $a_0 \equiv 0$  and for  $k = 1, 2, \dots, K$ , the following,

$$\begin{aligned}\mathcal{D}_k &\triangleq [x_{k-1}, x_k], \text{ where } x_i = -(K - i)\Delta_x \\ \gamma_P(x) &= a_{k-1} + m_k(x - x_{k-1}), \text{ for } x \in \mathcal{D}_k, \text{ and} \\ a_k &= a_{k-1} + m_k\Delta_x,\end{aligned}\tag{3.28}$$

where  $m_1, m_2, \dots, m_K$  are the slopes that need to be chosen.

*Notation:* Henceforth, we will refer to the quantizer  $\gamma_P(x)$  as the *Antisymmetric Unit-support Piecewise-Linear (AUPL)* quantizer. The AUPL quantizer is entirely specified in terms of the slope vector  $\mathbf{m}$ .

*Objective Function:* We characterize the objective function in terms of  $\mathbf{m}$ . For the piecewise linear quantizer  $\gamma_P$ , the expressions (3.22) and (3.23) reduce to linear functions

of  $\mathbf{m}$ , i.e.,

$$\begin{aligned}
g(\theta) &= [\mathbf{a}(\theta)]^T \mathbf{m} + F(\theta), \text{ and} \\
g'(\theta) &= [\mathbf{c}(\theta)]^T \mathbf{m} + f(\theta), \text{ where} \\
\mathbf{a}(\theta) &= J\mathbf{q}(\theta) + \mathbf{r}(\theta), \\
\mathbf{c}(\theta) &= J\mathbf{q}'(\theta) + \mathbf{r}'(\theta), \\
J &\triangleq \begin{bmatrix} K & 1 & \cdots & 1 \\ 0 & K-1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \\
[\mathbf{q}(\theta)]_k &= \Delta_x \int_{\mathcal{D}_k} \xi(\theta, x) dx, \text{ and} \\
[\mathbf{r}(\theta)]_k &= \int_{\mathcal{D}_k} x \xi(\theta, x) dx, \quad k = 1, 2, \dots, K.
\end{aligned} \tag{3.29}$$

Next, we discretize the parameter set. We note that the region of interest is only  $\theta \in [-1, 0]$ , with the other half taken care of through symmetry. We choose the parameter grid size  $\Delta_\theta$  or the number of grid partitions  $L$  so that  $L\Delta_\theta = 1$ . Let the discrete points be defined as

$$\theta_l = -\frac{l}{L}, \text{ for } l = 0, 1, \dots, L \tag{3.30}$$

Next, the maximum-CRB due to quantizer  $\gamma_P$  (see (3.10)) is approximated as

$$\phi(\mathbf{m}, f) = \max_l \frac{(\mathbf{a}_l^T \mathbf{m} + F_l)(1 - \mathbf{a}_l^T \mathbf{m} - F_l)}{(\mathbf{c}_l^T \mathbf{m} + f_l)^2}, \tag{3.31}$$

where  $\mathbf{a}_l \triangleq \mathbf{a}(\theta_l)$ ,  $\mathbf{c}_l \triangleq \mathbf{c}(\theta_l)$ ,  $F_l \triangleq F(\theta_l)$  and  $f_l \triangleq f(\theta_l)$ . In Equation (3.31),  $\phi(\cdot, f)$  is our objective function with  $\mathbf{m}$  as the variable.

*Constraints:* We identify two constraints. Firstly, the slopes  $m_k$  must be chosen so that the probability values for all observations  $x$  satisfy  $\gamma_P(x) \in [0, 1]$ . Since  $\gamma_P$  is piecewise

linear, this is ensured by placing inequality constraints at the boundary points. From (3.28), we obtain  $\gamma_P(x_k) = \Delta_x \sum_{j=1}^k m_j = \frac{1}{K} \sum_{j=1}^k m_j$ . Hence the probability constraint at point  $x_k$  can be expressed as  $0 \leq \sum_{j=1}^k m_j \leq K$ , for  $k = 1, 2, \dots, K$ . Secondly, from the antisymmetric property, assuming that  $\gamma_P(x)$  is continuous at 0, we have  $\gamma_P(0+) = \gamma_P(0-) = 1/2$ , and hence we need to ensure that  $\gamma_P(0) = \gamma_P(x_K) = 1/2$ , or equivalently,  $\sum_{j=1}^K m_j = K/2$ .

*Optimization Problem:* Finally, the minimax-CRB quantizer  $\phi(f)$  defined by (3.11) can be obtained as a solution to the following optimization problem in  $\mathbb{R}^K$ ,

$$\begin{aligned} & \underset{\mathbf{m}}{\text{minimize}} && \phi(\mathbf{m}, f), \\ & \text{s.t.} && \begin{bmatrix} -\mathbf{L} \\ \mathbf{L} \end{bmatrix} \mathbf{m} \leq \begin{bmatrix} \mathbf{0} \\ K\mathbf{i}_K \end{bmatrix}, \text{ and} \\ & && (\mathbf{i}_K)^T \mathbf{m} = K/2, \text{ where} \end{aligned} \tag{3.32}$$

$$\mathbf{L} \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \text{ and } \mathbf{i}_K \triangleq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.33}$$

*Implementation Notes:* It may be noted that  $\mathbf{a}_l, \mathbf{c}_l, F_l, f_l$  in  $\phi(\mathbf{m}, f)$  (see (3.31)) are all constants and may be pre-computed before running the optimizer. Also, once the noise pdf is known, the optimum quantizer  $\gamma_P$  can be computed offline and programmed into the sensor nodes. Choice of  $K$  and  $L$  essentially provides a tradeoff between discretization artifacts and numerical complexity. From numerical experiments,  $K = L \approx \lceil 10/\sigma \rceil$  was found to yield sufficiently convergent results. The problem given by (3.32) is not known to be convex (to the best of author's knowledge) and hence we require multiple and good starting points  $\mathbf{m}_0$  to obtain a satisfactory solution. In our implementation, we have chosen

two starting points for  $\mathbf{m}_0$ , namely the closest AUPL counterparts for the Threshold and Sine quantizers. We have used the MATLAB function FMINCON for optimization.

### 3.4 Illustrative examples

We illustrate some of the key ideas in this chapter through numerical results.

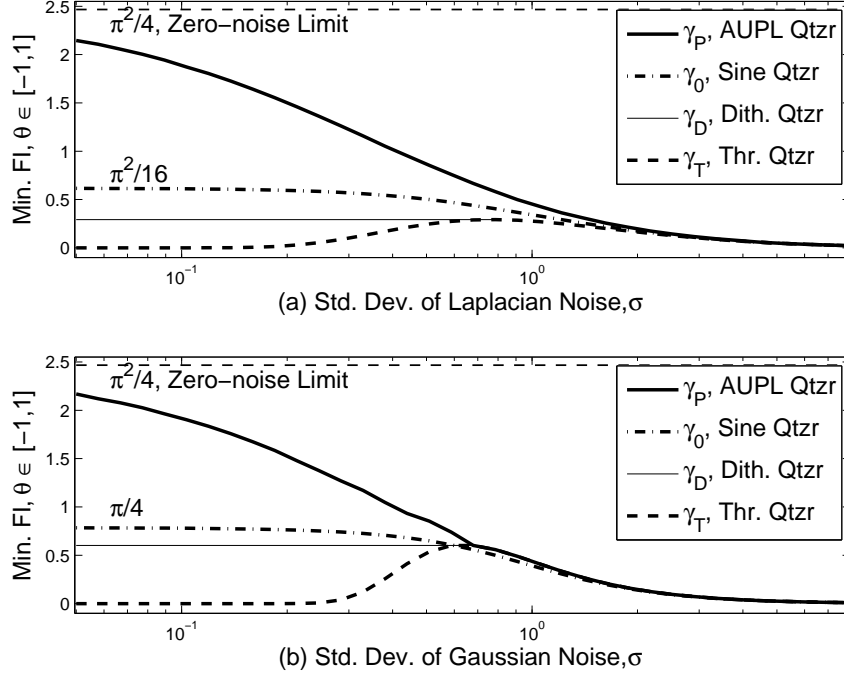


Figure 3.2: Minimum Fisher information  $(\phi(\gamma, f)^{-1})$  for Threshold, Sine, Dithering and AUPL quantizers for (a) Laplacian and (b) Gaussian noise.

*Sub-optimality of Sine Quantizer:* In Proposition 3.2.8, we showed that the Sine Quantizer given by (3.13), though optimum for zero-noise, is significantly sub-optimal when  $\sigma$  is small but finite (high-SNR). The results displayed in Figure 3.2 illustrate this phenomena. We display the minimum Fisher Information (inverse of Cramér-Rao bound) of the Sine Quantizer. As illustrative noise pdf-s, we consider Gaussian and Laplacian densities over a wide range of variance ( $0.05 \leq \sigma \leq 8$ ). The dotted line showing  $\phi_0^{-1} = \pi^2/4$  is the zero-noise limit. The dash-dotted lines corresponding to  $\pi^2/16$  and  $\pi/4$ , which are significantly less than the zero-noise limit, denote the performance of the Sine Quantizer. These results are consistent with the limits described in Table I.

*Performance of AUPL quantizer:* In Figure 3.2, we have also compared the AUPL quantizer  $\gamma_P$  with the Threshold  $\gamma_T$ , Dithering  $\gamma_D$  and Sine  $\gamma_0$  quantizers. The critical standard deviation for Dithering Quantizer corresponds to the maxima of the performance of  $\gamma_T$  (recall (3.12)). In Figure 3.2,  $\gamma_D$  corresponds to the unbroken horizontal lines connected to the maxima of  $\gamma_T$  performance curves. These critical variances are seen to approximately  $\sigma_{\mathcal{L}} \approx 0.79$  and  $\sigma_{\mathcal{N}} \approx 0.63$  for Laplacian and Gaussian noise respectively. We observe that the AUPL quantizer performs better than all three existing quantizers, and considerably so in the moderate to high-SNR regime.

*Minimax-optimality of Threshold Quantizer:* We showed in Example 3.2.7 that for Gaussian density with  $\sigma \geq 1$ , the Threshold Quantizer is also the antisymmetric minimax-CRB quantizer. We verify in Figure 3.2-(b) that the performance curves for AUPL and Threshold quantizers coincide for  $\sigma \geq 1$ . In fact, they seem to coincide somewhat earlier, around  $\sigma \geq 0.7$ . This is because dominance (see Theorem 3.2.6) is only a sufficient condition for minimax-CRB superiority. It may also be noted that no such coincidence is observed for the Laplacian case (Figure 3.2-(a)). Since the Laplacian density is not differentiable at the origin, Theorem 3.2.6 does not apply in this case.

*Shape of AUPL quantizer:* We display the shape of AUPL quantizer  $\gamma_P(x)$  and corresponding  $g(\theta)$  for various noise pdf-s in Figures 3.3 and 3.4 respectively. We consider Laplacian and Gaussian pdf-s for small ( $\sigma = 0.05$ ), medium ( $\sigma = 0.2$ ) and large ( $\sigma = 0.7$ ) variances. We note that for Gaussian noise, the AUPL quantizer displays a *damped oscillating* behavior, where the bumps get smaller but more in number, with decreasing variance. In the limit of small  $\sigma$ , the AUPL quantizer is seen to approach the shape of the Sine Quantizer, though not exactly. In the limit of large  $\sigma$ , for the Gaussian case, the AUPL quantizer is seen to approach the shape of the Threshold Quantizer. Figure 3.3 also shows that  $\gamma_P$  need not be monotonic. This is in contrast with commonly used Threshold, Dithering and Sine quantizers, all of which are monotonic. The AUPL quantizer relaxes this assumption and allows for non-monotone functions. The overall quantizer probability  $g(\theta)$ , however,

has to be monotonically increasing in  $\theta \in (-1, 1)$  to satisfy the admissibility property (see Definition 3.2.1). This can be verified in Figure 3.4.

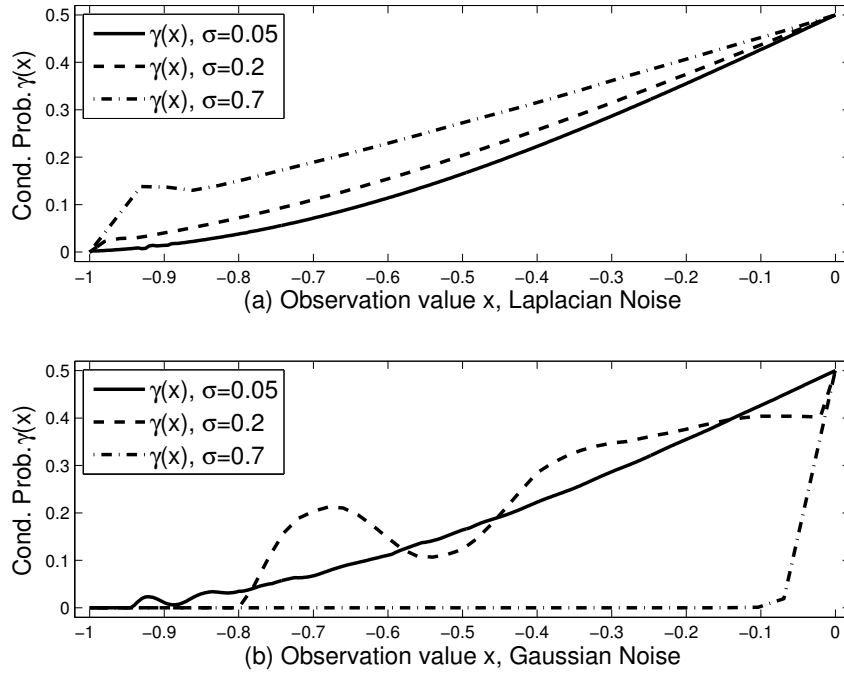


Figure 3.3: Designed probability  $\gamma_P(x) = P(Y = \mathcal{S}_1|x)$  for AUPL quantizer.

### 3.5 Summary

In this chapter, we studied the design of identical binary quantizers for distributed estimation using minimax Cramér-Rao lower bound as the performance criterion. Among other theoretical results, we have specified a broad family of distributions for which the Threshold Quantizer is optimal. Aided with some theoretical results, we formulated a numerical optimization problem to obtain the minimax-CRB quantizer within the antisymmetric and piecewise-linear class. We demonstrated the superior performance of the AUPL quantizer for a wide range of noise density functions. Though AUPL quantizers can demonstrably achieve better performance, they have some drawbacks that deserve mention. Firstly, AUPL quantizers are more difficult to implement because of the numerical complexity involved in the design process. Traditional quantizers like Sine, Threshold and Dithering

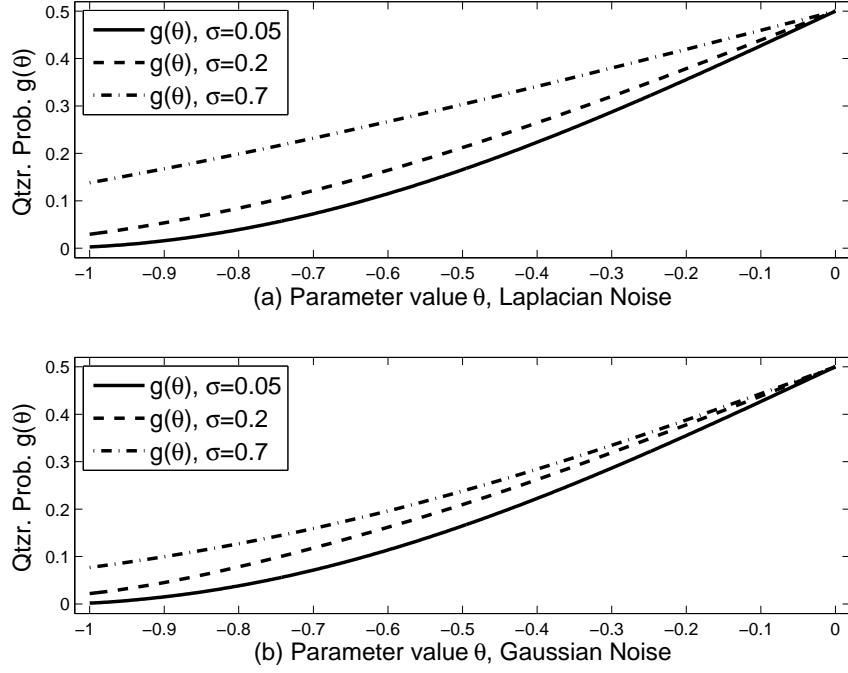


Figure 3.4: Overall probability  $g(\theta) = \mathbb{E}_W(\gamma_P(\theta + W))$  for AUPL quantizer.

quantizers are relatively simpler to design. Secondly, since AUPL quantizer is tailored to a particular noise density, it may not be suitable for applications where the ambient noise changes frequently. Lastly, the AUPL quantizer in Section 3.3 was derived under the assumption of noiseless channels. Extension of AUPL quantizer to noisy communication channels merits further investigation.



# Chapter 4

## Bit allocation for collaborative estimation using spatial whitening

### 4.1 Introduction

In this chapter, we demonstrate how spatial collaboration can result in more efficient resource allocation for estimation applications. Wireless sensor networks consist of spatially distributed noisy sensors that cooperatively monitor environmental conditions. Since the individual sensor nodes are characterized by limited energy, bandwidth and computational capability, the task of the fusion center (FC) is to make accurate inference about the phenomenon by requesting as little information from the sensor nodes as possible [73]. Depending on the particular application and set of constraints, the FC often has to adopt smart strategies to collect and process data [87]. While the design of optimum strategies in some cases is relatively easy under the assumption of *conditional independence*<sup>1</sup> across sensors, it is well known that the design gets harder and sometimes the optimum strategy is intractable when correlation has to be taken into account [51]. In particular, when the sensors are geographically close, they are expected to possess significant correlation among themselves and the optimum strategies derived for the independent case will no

---

<sup>1</sup>Here, ‘independence’ refers to the statistical independence of sensor data *conditioned* on the parameter of interest. For additive Gaussian observation noise, this is equivalent to the covariance matrix of noise being diagonal. The observations are still marginally dependent, since they are observing the same parameter.

longer be optimal. In this chapter, we introduce a framework called *spatial whitening* (to be formalized later) to deal with this problem.

Our framework stems from this idea: If two sensors in a network are highly correlated, they are also likely to be spatially close, which means that they should be able to communicate and exchange information among themselves in a relatively inexpensive manner (avoiding routing overheads and long distance communications). Each sensor in the network can now use the information from neighboring nodes to achieve a *local whitening transformation*. If each of such local transformations can be coordinated, one can aim to achieve *global whitening*, and the transformed observations can then be transmitted to the FC using optimum encoding strategies (for inference, resource allocation, etc) that were derived for conditionally independent scenarios. Hence, this two-stage (whitening followed by encoding) framework potentially enables the use of several earlier known results in the presence of correlated noise.

We introduce the log-determinant divergence based formulation of spatial whitening in Section 4.2. To illustrate the potential usage of this framework, we employ the problem of distributed parameter estimation [73], where several sensor nodes quantize their individual observations before sending them to FC. The goal is to minimize the expected distortion of the estimated parameter subject to a constraint on the total number of bits transmitted to the FC. We demonstrate that an optimal strategy for bit allocation (derived for independent scenario [51]) delivers increasingly better performance with increasing degree of whitening.

The whitening transformation described in this chapter requires local message passing which is certainly not without cost. However, in this chapter, we assign no cost to whitening, acknowledging fully that any actual implementation of a system would have to consider the tradeoff between the benefits of whitening and the cost of it. Investigations on this tradeoff is a worthy topic for future research.

The concept of whitening, in general, has mostly been addressed in a global framework till now. It is well known that the Karhunen-Loève Transform (KLT) [29] (also referred to as Principal Components Analysis, PCA or Singular Value Decomposition, SVD) of a

random vector with covariance matrix  $\Sigma = \mathbf{U}\Lambda\mathbf{U}^T$  provides the unique whitening transformation ( $\mathbf{U}^T$ ) that is also orthogonal. However, PCA is ill suited for our problem, since those whitening transformations are not local, while the orthogonality property serves no additional purpose. The Cholesky decomposition  $\Sigma = \mathbf{L}\mathbf{L}^T$ , which provides the unique lower triangular whitening transformation ( $\mathbf{L}^{-1}$ ), also requires non-local transformations. Moreover, the lower-triangular property imposes a tree-type dependence structure while in fact there is no natural ordering of spatially correlated data [100]. Other sparsity-inducing decompositions like Sparse-PCA [101] and vector Sparse-PCA [83] are *exploratory*<sup>2</sup> in nature, which means that the resulting transformations are not guaranteed to be local. In [88], a hardware-friendly technique was proposed to achieve generic spatial whitening transformations that were also global in scope. In distributed-KLT [29], individual nodes observe non-overlapping portions of a random vector and perform dimensionality-reduction (without collaboration with neighbors) for optimum reconstruction at the FC. Our primary contribution in this chapter is the formulation of a whitening framework that uses local communications among sensors and hence can be potentially used in sensor network scenarios.

## 4.2 Problem statement

In Figure 4.1, we consider  $N$  sensors in a network that is observing an unknown, deterministic, scalar parameter of interest  $\theta$  in the presence of zero-mean, correlated Gaussian noise with covariance  $\Sigma$ . Hence, the sensor observations  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  follow

$$\mathbf{x} \sim \mathcal{N}(\mathbf{1}\theta, \Sigma), \quad \mathbf{1} \triangleq [1 \cdots 1]^T \in \mathbb{R}^N. \quad (4.1)$$

Note that the sensor observations  $x_k$ -s are conditionally-independent when  $\Sigma$  is a diagonal matrix. Let the neighborhood structure among the various nodes be represented by the  $N \times N$  adjacency matrix  $\mathbf{A}$ ,  $A_{ij} \in \{0, 1\}$ , which is expected to be sparsely populated.

---

<sup>2</sup>The placeholders for non-zero coefficients are not known/specified beforehand.

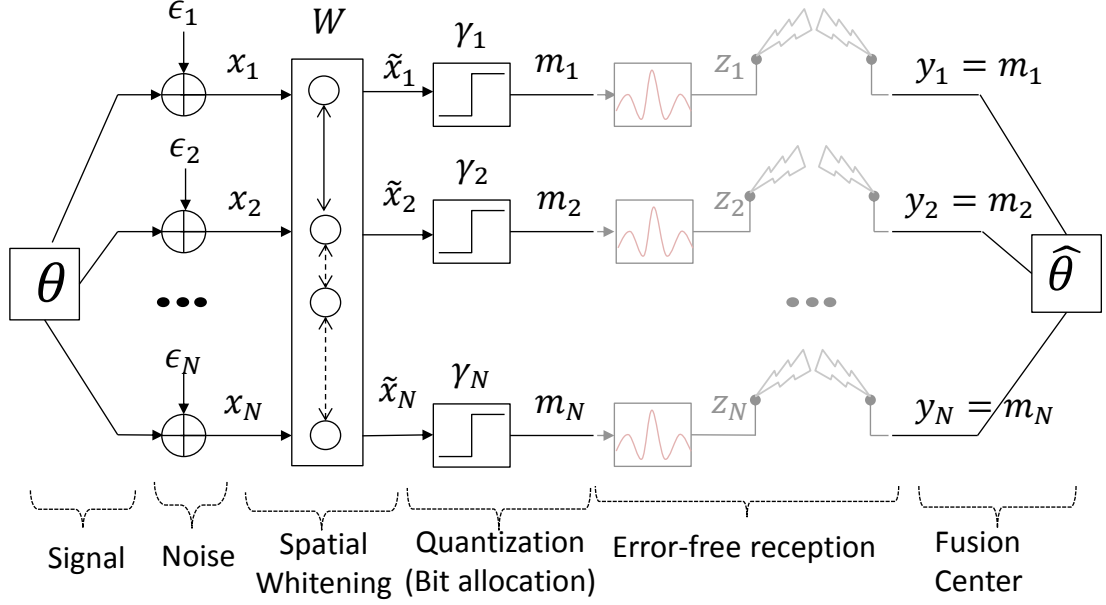


Figure 4.1: System diagram of spatial whitening based distributed estimation.

Entries  $A_{ij} = 1$  signify that node  $i$  is a neighbor of node  $j$ . A low-cost link for local communication is assumed to be available between two neighboring links. Since each node is trivially connected to itself,  $A_{ii} = 1$ . We denote the set of all  $\mathbf{A}$ -sparse matrices as

$$\mathcal{S}_A \triangleq \{\mathbf{W} \in \mathbb{R}^{N \times N} : W_{ij} = 0 \text{ if } A_{ij} = 0\}. \quad (4.2)$$

Note that because  $\mathbf{A}$  is the adjacency matrix, all linear transformations of the form

$$\tilde{\mathbf{x}} = \mathbf{W}\mathbf{x} \sim \mathcal{N}(\mathbf{W}\mathbf{1}\theta, \mathbf{W}\Sigma\mathbf{W}^T), \quad \mathbf{W} \in \mathcal{S}_A, \quad (4.3)$$

can be realized relatively inexpensively through local data transmissions, i.e., node  $k$  realizes the transformation  $\tilde{x}_k = \sum_{j \in \mathcal{N}_k} W_{kj}x_j$  by collaborating with its set of neighbors  $\mathcal{N}_k \triangleq \{j_1, j_2, \dots, j_{|\mathcal{N}_k|}\}$ , i.e., all the column-indices of  $\mathbf{A}$  such that  $A_{k,j_i} = 1$ .

The goal is to find the optimal mean-preserving, whitening transformation, i.e., one for which  $\mathbf{W}\mathbf{1} = \mathbf{1}$ , and  $\mathbf{W}\Sigma\mathbf{W}^T$  is as *near* to a diagonal matrix as possible. The mean-preserving condition ensures that the problem framework is preserved, i.e., any resource allocation algorithm previously designed for the observation domain  $\mathbf{x}$  is applicable to new

transformed domain  $\tilde{\mathbf{x}}$ . The whitening condition helps induce conditional independence across sensors (in some optimal sense). We choose the log-determinant divergence [17] as our metric for matrix-*nearness*, a point that we will elaborate later. The idea is that the nodes can use (optimally) whitened observations  $\tilde{x}_k$  (instead of original correlated observations  $x_k$ ) as the information to be encoded and relayed to the FC. This way an encoding strategy that was derived using conditional independence assumption across sensors can be used to enhance the performance of the system. We will consider the application of optimal encoding for distributed estimation in Section 4.4 and show the resulting improvement in performance due to the two-stage processing. But before that we describe our approach towards finding the optimum whitening transformation and comment on the computational aspects.

In the domain of symmetric positive-definite  $N \times N$  matrices, the log-determinant divergence of  $\mathbf{P}$  from  $\mathbf{Q}$  is defined [17] as

$$\mathcal{L}(\mathbf{P}; \mathbf{Q}) \triangleq \text{Tr } \mathbf{Q}^{-1} \mathbf{P} - \log \det \mathbf{P} - N + \log \det \mathbf{Q}. \quad (4.4)$$

It is well known that  $\mathcal{L}(\mathbf{P}; \mathbf{Q})$  is a Bregman-divergence [17] and hence convex in  $\mathbf{P}$  for any fixed  $\mathbf{Q}$ . Also  $\mathcal{L}(\mathbf{P}; \mathbf{Q}) \geq 0$  for all  $\mathbf{P}$  and  $\mathbf{Q}$  with equality if and only if  $\mathbf{P} = \mathbf{Q}$ . We formulate the spatial whitening problem as finding an  $\mathbf{A}$ -sparse, mean-preserving transformation  $\mathbf{W}$  and a diagonal matrix (with positive entries)  $\mathbf{D}$  such that the divergence  $\mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D})$  is minimized,

$$\min_{\mathbf{W}, \mathbf{D}} \mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D}) \quad \text{s.t.} \quad \mathbf{W} \in \mathcal{S}_A, \mathbf{W}\mathbf{1} = \mathbf{1}. \quad (4.5)$$

We note from definition (4.4) that

$$\mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D}) = \mathcal{L}(\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\Sigma\mathbf{W}^T\mathbf{D}^{-\frac{1}{2}}; \mathbf{I}), \quad (4.6)$$

where  $\mathbf{I}$  is the identity matrix. Using (4.6), we obtain an equivalent formulation of (4.5),

$$\min_{\mathbf{Z}} \quad \mathcal{L}(\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^T; \mathbf{I}) \quad \text{s.t.} \quad \mathbf{Z} \in \mathcal{S}_A, \quad (4.7)$$

$$\text{where} \quad \mathbf{W} = \delta^{-1}(\mathbf{Z}\mathbf{1})\mathbf{Z}, \quad \mathbf{D} = \delta^{-2}(\mathbf{Z}\mathbf{1}), \quad (4.8)$$

where  $\delta(\cdot)$  is the diagonalization<sup>3</sup> operator. We note that (4.7) is a significantly simplified re-formulation of (4.5). Using (4.4), we define the cost function w.r.t.  $\mathbf{Z}$  as

$$l(\mathbf{Z}) \triangleq \mathcal{L}(\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^T; \mathbf{I}) = \text{Tr } \mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^T - \log \det \mathbf{Z}\mathbf{Z}^T + c_0, \quad (4.9)$$

where  $c_0 \triangleq -N - \log \det \mathbf{\Sigma}$  is a constant. We refer to (4.7) as the *log-determinant divergence based spatial whitening* problem. If the cardinality of non-zero elements of  $\mathbf{A}$  is  $\text{nz}(\mathbf{A}) \leq N^2$ , then (4.7) is an optimization problem in  $\mathbb{R}^{\text{nz}(\mathbf{A})}$ .

A related problem is that of structured inverse-covariance selection (Example 2.5 in [86]) and its subsequent application for in-network inference using graphical models [8]. However, the problem we address in this chapter is different in two ways. Firstly, rather than selecting the inverse-covariance, we are interested in selecting the inverse of the Cholesky factor (modulo an orthogonal multiple). Secondly, we are addressing a fixed FC based problem as opposed to in-network inference problems like belief-propagation or consensus.

Since  $\mathbf{Z}$  is not restricted to the set of symmetric positive-definite matrices, our objective function (4.9) does not inherit the convexity property of well known *max-det* problems [86]. Neither does the first-order gradient condition, written in matrix-derivative notations [60],

$$\frac{\text{d}l(\mathbf{Z})}{\text{d}\mathbf{Z}} = 2(\mathbf{Z}\mathbf{\Sigma} - \mathbf{Z}^{-T}) \circ \mathbf{A} = 0, \quad (4.10)$$

where  $\circ$  denotes the element-wise (or Hadamard) product, lend itself to any known closed-form solution except in the trivial situation when  $\mathbf{A}$  is the all-1 matrix (in which case,  $\mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^T = \mathbf{I}$ , and any orthogonal multiple of the Cholesky factor  $\mathbf{L}^{-1}$  is a solution for

---

<sup>3</sup>Function  $\mathbf{X} = \delta(\mathbf{x})$  is defined as  $\delta : \mathbb{R}^N \rightarrow \mathbb{R}^{N \times N}$  such that  $\mathbf{x}$  corresponds to the diagonal elements of  $\mathbf{X}$ , other elements being zero.

$\mathbf{Z}$ ). In the next section, we provide an iterative algorithm that finds (locally) optimal solutions to problem (4.7). Multiple runs using different starting points must be used to mitigate the local-maxima problem and obtain a satisfactory solution. It may be noted here that in most of the existing literature, matrix factorization problems of this nature (involving sparsity/structure) are inherently non-convex and can only guarantee locally optimal solutions [29], [101], [83].

### 4.3 Iterative algorithm for spatial whitening

In our iterative approach to solving problem (4.7), we update each row of elements in  $\mathbf{Z}$  to achieve the optimum decrement in divergence, while keeping the rest of the matrix unchanged. This process is repeated until convergence. Each such iteration is a convex optimization problem and we obtain closed form expressions for the updates.

Optimizing (4.7) with respect to the row-vector  $\mathbf{z}_k \triangleq \mathbf{Z}_{k, \mathcal{N}_k} \in \mathbb{R}^{|\mathcal{N}_k|}$  while keeping all the other elements of  $\mathbf{Z}$  constant is equivalent to minimizing

$$g(\mathbf{z}_k) = \frac{1}{2} \mathbf{z}_k^T \boldsymbol{\Sigma}_k \mathbf{z}_k - \log(\mathbf{z}_k^T \mathbf{c}_k), \quad (4.11)$$

$$\boldsymbol{\Sigma}_k \in \mathbb{R}^{|\mathcal{N}_k| \times |\mathcal{N}_k|}, \mathbf{c}_k \in \mathbb{R}^{|\mathcal{N}_k|},$$

where  $\boldsymbol{\Sigma}_k$  denotes the  $\mathcal{N}_k$ -clique covariance matrix extracted from  $\boldsymbol{\Sigma}$ , and the elements of  $\mathbf{c}_k$  are defined by

$$(\mathbf{c}_k)_i \triangleq (-1)^{k+j_i} \det(\mathbf{Z}_{-k, j_i}), \quad i = 1, 2, \dots, |\mathcal{N}_k|, \quad (4.12)$$

with  $\mathbf{Z}_{-k, j_i}$  denoting the matrix obtained after truncating the  $k^{\text{th}}$  row and  $j_i^{\text{th}}$  column of  $\mathbf{Z}$ . The first-order gradient condition of (4.11) implies  $(\mathbf{z}_k^T \mathbf{c}_k) \boldsymbol{\Sigma}_k \mathbf{z}_k = \mathbf{c}_k$ , solving which

one obtains the unique extremum of (4.11),

$$\mathbf{z}_k^* = \frac{\boldsymbol{\Sigma}_k^{-1} \mathbf{c}_k}{\sqrt{\mathbf{c}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{c}_k}}. \quad (4.13)$$

That  $\mathbf{z}_k^*$  is the minimizer follows from the convexity of (4.11) (the Hessian is  $(\boldsymbol{\Sigma}_k + (\mathbf{z}_k^T \mathbf{c}_k)^{-2} \mathbf{c}_k \mathbf{c}_k^T)$ , which is positive definite).

Each rank-one update of the form (4.13) can be efficiently computed using the well-known Woodbury-formula. Since the overall divergence of (4.7) decreases at each of the iterations of (4.11), and the minimum divergence is lower bounded (see equation (4.10)) by

$$\sup_{\mathbf{Z} \in \mathcal{S}_A} l(\mathbf{Z}) \geq \sup_{\mathbf{Z} \in \mathbb{R}^{N \times N}} l(\mathbf{Z}) = l(\mathbf{L}^{-1}) = 0, \quad (4.14)$$

this iterative algorithm is guaranteed to converge. It may be noted that these kind of iterative techniques are sometimes called block-coordinate-descent or terminal-by-terminal optimization [29].

In the remainder of this chapter, we will focus on the application of spatial whitening to distributed estimation.

## 4.4 Bit-allocation for collaborative estimation

We consider the practical parameter-estimation problem where individual sensors in a network are required to quantize their real-valued local measurements to an appropriate length and send the resulting discrete message to the FC, while the latter combines all the received messages to produce a final estimate [73]. The critical resource that needs to be conserved is the bandwidth or equivalently, the rate of transmission. Assume that the network consisting of  $N$  nodes is allowed to transmit only  $B$  bits in totality for a one-shot estimation problem. The question then is how to judiciously allocate the  $B$  bits among the various sensors such that the resulting distortion of estimate is minimized at the FC [51],



[53]. For the sake of simplicity, we assume that each sensor incurs an equal per-bit cost for transmission.

We would use the quantization and bit allocation framework outlined in [51]. All observations  $x_k$ -s are assumed to be bounded to a finite interval  $[-U, U]$  and a uniform probabilistic quantization is performed. An observation is quantized with  $b_k$ -bits as follows. The quantization points  $a_j^{(k)} \in [-U, U], j = 1, \dots, 2^{b_k}$  are uniformly spaced such that  $a_{j+1}^{(k)} - a_j^{(k)} = 2U/(2^{b_k} - 1) \triangleq \Delta_k$ . Suppose that  $x_k \in [a_j^{(k)}, a_{j+1}^{(k)})$ . Then  $x_k$  is quantized to either  $a_{j+1}^{(k)}$  or  $a_j^{(k)}$  according to

$$P(m_k = a_j^{(k)}) = q, \quad P(m_k = a_{j+1}^{(k)}) = 1 - q, \quad (4.15)$$

where  $m_k$  is the resulting message and  $q = (a_{j+1}^{(k)} - x_k)/\Delta_k$ .

When the noise is Gaussian and independent across sensors, the subsequent near-optimal strategy [51] is particularly simple and allocates

$$b_k = \text{ROUND} \left[ \log_2 \left( 1 + \frac{1}{\lambda \sigma_k^2} \right) \right] \quad (4.16)$$

bits to the  $k^{\text{th}}$  sensor, where  $\sigma_k^2$  is the individual variance,  $\lambda > 0$  controls the overall sum of bits  $\sum_{k=1}^N b_k = B$  and the rounding is performed to the nearest integer. The idea is that FC broadcasts a lower value of  $\lambda$  when a more precise parameter estimate is needed. However, when the noise is correlated, strategy (4.16) is suboptimal and this is where spatial whitening can be of help. Once we perform a spatial whitening transformation in the observation space, the idea is that we effectively de-correlate the noise without losing any information and hence a strategy like (4.16) applied on the modified space can still deliver near-optimal performance.

Next, we state the distortion metric derived in [51] which we shall use for comparing the performance of various schemes. For a random variable  $\mathbf{y} \sim \mathcal{N}(\mathbf{1}\theta, \mathbf{C})$  that is effectively range limited in  $[-U, U]$ , the mean-square-error (MSE) for estimating  $\hat{\theta}$  at FC (when  $y_k$  is

quantized to  $m_k$  using  $b_k$  bits) following the scheme in (4.15), is given by

$$\text{MSE}(\hat{\theta}) \approx \frac{\mathbf{1}^T \mathbf{C}^{-1} (\mathbf{C} + \mathbf{Q}) \mathbf{C}^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2}, \quad (4.17)$$

where  $\mathbf{Q}$  is the diagonal matrix with elements  $Q_{kk} = (U^2)/(2^{b_k} - 1)^2$ . It is assumed that FC is using the optimally weighted fusion rule  $\hat{\theta} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{m}$  (see [49]) on the quantized observations.

Our simulation setup is as follows. The spatial placement and neighborhood structure is modeled as a Random Geometric Graph  $RGG(N, r)$  [24], where sensors are uniformly distributed over a unit square with communication links present only for pairwise distances of at most  $r$ . The noise is modeled as an exponentially correlated Gaussian covariance matrix  $\Sigma$ ,

$$\mathbf{x} \sim \mathcal{N}(\mathbf{1}\theta, \Sigma), \quad \Sigma_{i,j} = \sigma_i \sigma_j \alpha^{d_{i,j}}, \quad (4.18)$$

where  $\alpha \in (0, 1)$  is indicative of the degree of spatial correlation. A smaller value of  $\alpha$  indicates lower correlation with  $\alpha \rightarrow 0$  signifying completely independent observations.

We consider  $N = 50$  nodes and the particular RGG used for our simulation is depicted in Figure 4.2. The individual sensor variances  $\sigma_k^2$  are generated by uniform random numbers in the range  $[0.5, 1.5]$  and the correlation parameter  $\alpha = 0.02$ . The range-limit of observations is taken as  $U = 20$ .

In Figure 4.3, we compare the distortion performance  $\text{MSE}(\hat{\theta})$  (4.17) corresponding to the three scenarios when strategy (4.16) is applied to various transformations of the data. The line labeled *not whitened* corresponds to the naive case of strategy (4.16) being directly applied to the observation space  $\mathbf{x}$ . Expectedly, the performance of this scheme is suboptimal. In *spatially whitened* cases, we use the transformed variable (see (4.8))

$$\tilde{\mathbf{x}} = \mathbf{W}_r \mathbf{x}, \quad \mathbf{W}_r = \delta^{-1}(\mathbf{Z}_r \mathbf{1}) \mathbf{Z}_r, \mathbf{D}_r = \delta^{-2}(\mathbf{Z}_r \mathbf{1}), \quad (4.19)$$

where  $\mathbf{Z}_r$  is the minimum-divergence solution (4.7) subject to constraints that  $[\mathbf{Z}_r]_{ij} = 0$  if  $d_{ij} > r$ . We note that

$$\tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{1}\theta, \mathbf{W}_r \Sigma \mathbf{W}_r^T), \quad (4.20)$$

which implies that  $\tilde{x}_k$  possess the same mean as the signal, but corrupted only with *approximately* independent Gaussian noise with variance  $\gamma_k^2 \triangleq \text{Var}(\tilde{x}_k) \approx [\mathbf{D}_r]_{k,k} = ([\mathbf{Z}_r \mathbf{1}]_k)^{-2}$ . Strategy (4.16) is then applied on whitened space  $\tilde{\mathbf{x}}$  with  $\sigma_k^2$  replaced by  $\gamma_k^2$  in Equation (4.16). We have shown the performance for  $r = 0.1$  and  $r = 0.5$  in Figure 4.3. As the range  $r$  increases, we have more whitening and consequently the performance increases. Thirdly, we display the results for *orthogonally whitened (or PCA)* case, where we consider the well known eigenvalue decomposition  $\Sigma = \mathbf{U} \Lambda \mathbf{U}^T$  and consequently the whitening transformation

$$\tilde{\mathbf{x}} = \mathbf{D}^{-1} \mathbf{U}^T \mathbf{x}, \quad \mathbf{D} \triangleq \delta(\mathbf{U}^T \mathbf{1}). \quad (4.21)$$

Since PCA fully whitens  $\Sigma$  (by definition), its performance is expected to provide a lower bound on that of other schemes. This is confirmed by Figure 4.3. However, since the weights in PCA are not designed to be zero for sensors that are far apart, such a transformation may be impossible to realize in a power constrained network and hence not realistic. Finally, the Cramer-Rao lower bound  $\text{CRB} = 1/(\mathbf{1}^T \Sigma^{-1} \mathbf{1})$  is also displayed, which confirms that in the asymptotic regime with sufficient quantization bits per sensor, all these schemes perform identically.

## 4.5 Summary

In this chapter, we have considered a two-stage framework for distributed signal processing in the presence of spatially correlated data. The first stage is designed to whiten the observation space by communicating only with neighboring sensors. In the second stage,

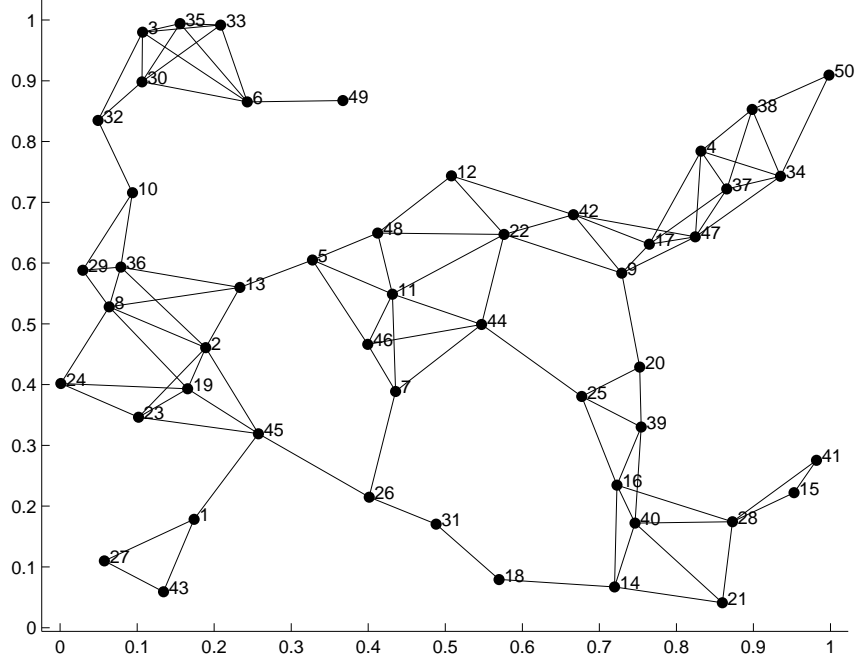


Figure 4.2: Random Geometric Graph used for example in Section 4.4. Edges are shown of pairwise distance less than 0.18.

each sensor encodes these whitened observations following well-known strategies derived using conditional independence assumption. We consider the example of bit-allocation for distributed estimation to demonstrate the potential applicability of this framework. Many research questions on whitening remain unaddressed in this chapter and are worthy topics of future research. Some of them are cost considerations for the whitening stage, theoretical analysis of whitening based distributed estimation and extension of this framework to vector parameter scenarios.

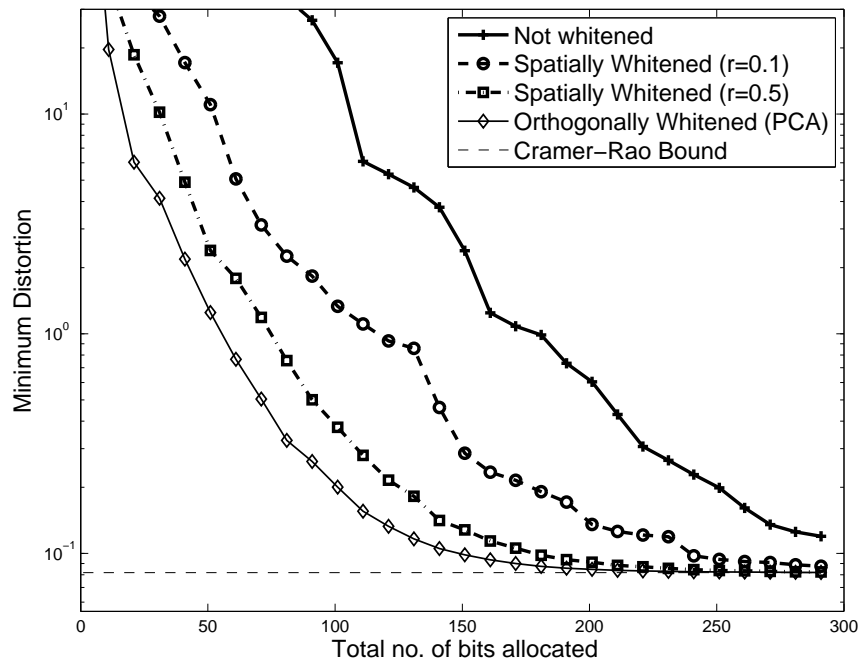


Figure 4.3: Distortion reduction achieved by spatial whitening.

# Chapter 5

## Power allocation for analog forwarding based collaborative estimation

### 5.1 Introduction

In Chapter 4, we have shown how spatial collaboration can lead to efficient resource allocation in quantization-based estimation applications. In this chapter, we demonstrate the same for analog amplify-and-forward based estimation applications. To recall, by *spatial collaboration*, we mean that the sensors can share their observations among other neighboring nodes prior to transmission to the FC. The observations from the neighbors are linearly combined using appropriate weights and then transmitted to the fusion center (FC) through a coherent multiple access channel (MAC). The FC receives the noise-corrupted signal and makes the final inference. The schematic diagram of such a system is shown in Figure 5.1 (we will introduce the notations and describe each block later in Section 5.2).

The individual sensor nodes are battery powered and hence the network is power limited. The power constraints can be described by the following two situations 1) *Cumulative*: Here the total power-usage in the network (summed across all the nodes) has to be be-

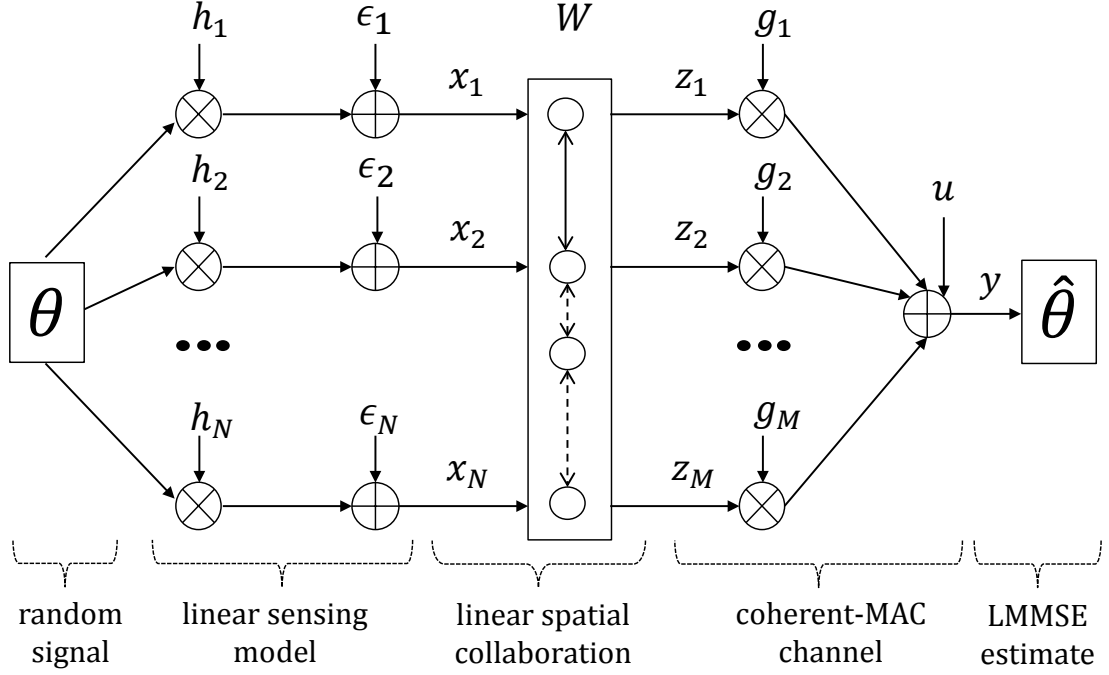


Figure 5.1: Sensor network performing collaborative estimation.

low a pre-specified limit. 2) *Individual*: Here each node has their own power constraint as dictated by the capacity of their batteries. The performance analysis of a *cumulative-constraint* problem is usually simpler and more insightful since we have only one constraint to take care of. The *individual-constraint* problem is more practical from an implementation perspective but also more difficult to analyze. In the absence of a power limit, the sensors could collaborate with all the other nodes, make the inference in the network, and transmit the estimated parameter to the FC without any further distortion (by using infinite transmission power). This is similar to the centralized inference situation, where the error in estimate is only due to the noisy observation process. However, with limited power availability, both collaboration and transmission have to be performed judiciously, so as to maximize the quality of inference at the FC. In this chapter, we study the following problems. For a cumulative power constraint, we study the optimal allocation of power resources among various nodes and tasks (namely collaboration and transmission) so as to achieve the best estimation performance at the FC. Regarding individual power constraints, the goal is to allocate power between collaboration and transmission at each node.

In the absence of collaboration, this problem reduces to the class of distributed inference/beamforming algorithms. In *distributed estimation*, the objective is to coordinate all the sensors so that without communicating with one-another, they collectively maximize the quality of inference at the FC. The quality of inference can be quantified by either the mean-square error (MSE) (in case of random signals) or Cramér-Rao lower bound (in case of deterministic unknown signals). Distributed estimation has been extensively researched both from analog [96],[14],[50] and digital [73],[54],[42] encoding perspectives. Examples of analog encoding include the amplify-and-forward (AF) scheme, where the nodes amplify the raw observations and transmit to the FC by either forming a coherent beam in a multiple access channel [96] or using their own dedicated links with the FC for transmission (sometimes referred to as orthogonal MAC) [14]. The AF framework appears extensively in the literature [96],[50],[12] due to its simplicity of implementation in complex networks and provable information-theoretic-optimality properties for simple networks [28]. In another research direction, quantization of the observations may be performed prior to transmission [58],[73]. The quantized observations are then communicated to the FC using digital communication, where further information may be lost due to channel errors [94]. Another closely related field of research in communication theory is *distributed beamforming in relay networks* [35], [41], [55], where the objective is to maximize the signal-to-noise ratio (SNR), rather than minimize the estimation error, at the FC. These two problems are sometimes related, as one might imagine. When the observation and channel gains are perfectly known, the SNR and MSE functions are monotonically related and the two problems are equivalent. However, in the presence of observation and channel gain uncertainties, the SNR and MSE functions are different.

Though distributed inference/beamforming has been widely studied, research regarding collaborative estimation is relatively nascent. When the transmission channels are orthogonal and cost-free collaboration is possible within a fully connected sensor network, the optimal strategy is to perform the inference in the network and use the best available channel to transmit the estimated parameter [20]. In the quantization based framework considered



in Chapter 4, we saw that spatial collaboration can be used to whiten the observation space, thereby enabling efficient resource allocation when the noise is correlated. In this chapter, we explore the benefits of spatial collaboration in the analog amplify-and-forward framework, where sensors are able to linearly combine the observations from neighboring nodes before transmitting to the FC. We obtain the optimal cumulative power-distortion tradeoff when a fixed but otherwise cost-free collaborative topology is used to transmit over a coherent MAC channel. Subsequently, we extend the problem formulation in three new directions, namely a) consideration of individual power constraints, b) consideration of imperfect information about observation and channel gains (the second order statistics are assumed to be known), and c) consideration of finite costs associated with collaboration. The primary contributions of this chapter are as follows

- Extending the amplify-and-forward framework to formulate and analyze the problem of estimation with spatial collaboration
- Defining a metric called *collaboration gain*, that quantifies the worthiness of spatial collaboration as a tool to enhance the estimation performance
- Demonstrating that for a fixed but otherwise cost-free (ideal) collaborative topology, the resulting optimization problem reduces to an eigen-decomposition problem for the cumulative-constraint case. For the individual-constraint case, accurate numerical solution can be obtained by solving several semi-definite feasibility problems. We investigate both cases further by deriving/analyzing the optimal achievable distortion and the corresponding weights for some special collaborative topologies like the distributed (no-connections), partially connected cycles, and fully connected cases. We also derive the explicit expression of collaboration gain for a homogeneous network with identical channel and observation gains and equicorrelated observation noise. In particular, we demonstrate that collaboration is particularly effective in a certain power regime that depends on various factors like the *skewness* (or variability) of

power-availability in the network, uncertainty in observation and channel gains, and the correlation of the measurement noise.

- Addressing the design of collaborative topologies where finite costs are involved in collaboration. We suggest an efficient algorithm that uses the results for the fixed-topology but cost-free case to find locally optimal solutions for the finite-cost case.

The rest of the chapter is organized as follows. In Section 5.2, we formulate the problem after describing each block of the system in Figure 5.1. We define “collaboration gain” (CG) that is normalized with respect to the operating region ( $\text{CG} \in [0, 1]$ ). It summarizes the efficacy of collaboration across various problem conditions. In Section 5.3, we solve the optimal transmission-power allocation problem for a fixed but otherwise cost-free collaborative topology, (i.e., extend the results of [45] to address points (a) and (b) in the previous paragraph). We also derive explicit expressions for collaboration gain for a homogeneous network with equicorrelated noise. In Section 5.4, we address the problem with finite collaboration costs and suggest a greedy algorithm to obtain a locally optimal solution in polynomial time. Concluding remarks are presented in Section 5.5.

## 5.2 Problem formulation

### 5.2.1 Linear sensing model

We consider the scenario where the parameter of interest is a scalar random variable with known statistics, specifically, Gaussian distributed with zero mean and variance  $\eta^2$ . The observations at the sensor nodes  $n = 1, 2, \dots, N$  are governed by the linear model  $x_n = \tilde{h}_n \theta + \epsilon_n$ , where  $\tilde{h}_n$  is the observation gain and  $\epsilon_n$  is the measurement noise. The second order statistics of the observation gain  $\tilde{\mathbf{h}} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_N]^T$  is assumed to be

$$\mathbb{E} \tilde{\mathbf{h}} = \mathbf{h}, \quad \text{var } \tilde{\mathbf{h}} = \Sigma_{\mathbf{h}}. \quad (5.1)$$

The measurement noise  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]^T$  is assumed to be zero-mean, Gaussian with (spatial) covariance  $\text{var } \boldsymbol{\epsilon} = \boldsymbol{\Sigma}$ . *Perfect knowledge of the observation model statistics  $\mathbf{h}$ ,  $\boldsymbol{\Sigma}_h$  and  $\boldsymbol{\Sigma}$  is assumed.* In vector notation, we have

$$\mathbf{x} = \tilde{\mathbf{h}}\theta + \boldsymbol{\epsilon}, \quad (5.2)$$

where  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$  denotes the observations.

### 5.2.2 Linear spatial collaboration

We consider an extension of the analog amplify-and-forward scheme as our encoding and modulation framework for communication to the fusion center. In the basic amplify-and-forward scheme, each node transmits a weighted version of its own observation, say  $W_n x_n$ , with resulting power  $W_n^2 \mathbb{E}[x_n^2]$ . Such a scheme is appealing and often-used [14],[96],[20] due to two reasons, 1) *Uncoded nature*: Does not require block coding across time and hence efficient for low-latency systems, 2) *Optimal in select cases*: For a memoryless Gaussian source transmitted through an additive white Gaussian noise (AWGN) channel (Figure 5.1 with  $N = 1$ ), an amplify-and-forward scheme helps achieve the optimal power-distortion tradeoff in an information-theoretic sense (see Example 2.2 in [26]). The optimality of linear coding has also been established [30] for distributed estimation over a coherent MAC (Figure 5.1 without spatial collaboration).

In general, all the  $N$  data-collecting sensor nodes in a network may not have the ability to communicate with the FC. In that case, they may still pass their information (through the act of collaboration) to another node which has a communication link with the FC (see the network in Figure 5.3, for example). We assume that the nodes are ordered in such a way that the first  $M$  nodes (where  $M \leq N$ ) are able to communicate with the FC. Let the availability of collaborative links among the various nodes be represented by the  $M \times N$  zero-one adjacency matrix (not necessarily symmetric)  $\mathbf{A}$ , where  $A_{mn} \in \{0, 1\}$ . An entry  $A_{mn} = 1$  signifies that node  $n$  shares its observation with node  $m$ . *Sharing of this*

observation is assumed to be realized through a reliable communication link that consumes power  $C_{mn}$ , regardless of the actual value of observation. The  $M \times N$  matrix  $\mathbf{C}$  describes all the costs of collaboration among various sensors and is assumed to be known. Since each node is trivially connected to itself,  $A_{mm} = 1$  and  $C_{mm} = 0$ . We denote the set of all  $\mathbf{A}$ -sparse matrices as

$$\mathcal{S}_A \triangleq \{\mathbf{W} \in \mathbb{R}^{M \times N} : W_{mn} = 0 \text{ if } A_{mn} = 0\}. \quad (5.3)$$

Corresponding to an adjacency matrix  $\mathbf{A}$  and an  $\mathbf{A}$ -sparse matrix  $\mathbf{W}$ , we define *collaboration* in the network as individual nodes being able to linearly combine local observations from other collaborating nodes,

$$z_m = \sum_{\substack{n=1, \dots, N \\ A_{mn}=1}} W_{mn} x_n, \quad m = 1, \dots, M. \quad (5.4)$$

In effect, the network is able to achieve a one-shot spatial transformation  $\mathbf{W} : \mathbf{x} \rightarrow \mathbf{z}$  of the form<sup>1</sup>

$$\mathbf{z} = \mathbf{W}\mathbf{x}, \quad \mathbf{W} \in \mathcal{S}_A. \quad (5.5)$$

We refer to  $\mathbf{W}$  as the matrix containing *collaboration weights*. It may be noted that, 1) *Particularization:* When  $\mathbf{W}$  is a diagonal matrix (equivalently,  $\mathbf{A}$  is the identity matrix  $\mathbf{I}_M$ ), our collaborative scheme simplifies to the basic amplify-and-forward relay strategy as in [50],[96], 2) *Collaboration cost:* Any collaboration involving  $\mathbf{W} \in \mathcal{S}_A$  is achieved at the

---

<sup>1</sup>It is worth emphasizing our assumption that though collaboration incurs a fixed cost (in terms of power consumed that could otherwise have been used for transmission), it is otherwise reliable, in the sense that the act of collaboration does not incur any errors. This can be implemented by, say communicating in a digital framework with sufficient precision and ensuring sufficient channel coding to counter the channel noise. We abstract this process by assigning a cost to the link when it is required to be used. An interesting problem which is worthy of research but beyond the scope of this chapter, is to assume possibly erroneous collaboration, where  $\mathbf{z} = \mathbf{W}\mathbf{x} + \boldsymbol{\zeta}$  (say) and errors incurred during collaboration ( $\boldsymbol{\zeta}$ ) decrease with collaboration power.

expense of power

$$Q_{\mathbf{A},m} \triangleq \sum_{n=1}^N C_{mn} A_{mn}, \quad (5.6)$$

at node  $m$ , and cumulatively

$$Q_{\mathbf{A}} \triangleq \sum_{m=1}^M Q_{\mathbf{A},m}, \quad (5.7)$$

for the entire network, and 3) *Transmission cost*: The power required for transmission of encoded message  $z_m$  at node  $m$  is,

$$P_{\mathbf{W},m} \triangleq \mathbb{E}_{\theta, \tilde{\mathbf{h}}, \epsilon} [z_m^2; \mathbf{W}] = [\mathbf{W} \mathbf{E}_x \mathbf{W}^T]_{m,m}, \text{ where} \quad (5.8)$$

$$\mathbf{E}_x \triangleq \mathbb{E}_{\theta, \tilde{\mathbf{h}}, \epsilon} [\mathbf{x} \mathbf{x}^T] = \mathbf{\Sigma} + \eta^2 (\mathbf{h} \mathbf{h}^T + \mathbf{\Sigma}_h). \quad (5.9)$$

Consequently, the cumulative transmission power in the network is

$$P_{\mathbf{W}} \triangleq \sum_{m=1}^M P_{\mathbf{W},m} = \text{Tr} [\mathbf{W} \mathbf{E}_x \mathbf{W}^T]. \quad (5.10)$$

### 5.2.3 Coherent multiple access channel

The transformed observations  $\mathbf{z}$  are assumed to be transmitted to the fusion center through a coherent MAC channel. In practice, a coherent MAC channel can be realized through *transmit beamforming* [62], where sensor nodes simultaneously transmit a common message (in our case, all  $z_m$ -s are scaled versions of a common  $\theta$ ) and the phases of their transmissions are controlled so that the signals constructively combine at the FC. Denote the channel gain at node  $m$  by  $\tilde{g}_m$ . The second order statistics of the channel  $\tilde{\mathbf{g}} \triangleq [\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_M]$  is assumed to be,

$$\mathbb{E} \tilde{\mathbf{g}} = \mathbf{g}, \quad \text{var} \tilde{\mathbf{g}} = \mathbf{\Sigma}_g, \quad (5.11)$$

and the noise of the coherent MAC channel  $u$  is assumed to be a zero-mean AWGN with variance  $\xi^2$ . *Perfect knowledge of the channel statistics  $\mathbf{g}$ ,  $\Sigma_{\mathbf{g}}$  and  $\xi^2$  is assumed.* The output of the coherent MAC channel (or the input to the fusion center) is

$$y = \tilde{\mathbf{g}}^T \mathbf{W} \mathbf{x} + u, \quad u \sim \mathcal{N}(0, \xi^2) \quad (5.12a)$$

$$= \underbrace{\tilde{\mathbf{g}}^T \mathbf{W} \tilde{\mathbf{h}}}_{\text{net gain}} \theta + \underbrace{\tilde{\mathbf{g}}^T \mathbf{W} \boldsymbol{\epsilon} + u}_{\text{net zero-mean noise}}. \quad (5.12b)$$

#### 5.2.4 Linear minimum mean square estimation

Having received  $y$ , the goal of the fusion center is to obtain an accurate estimate  $\hat{\theta}$  of the original random parameter  $\theta$ . We restrict our attention to linear estimators of the form  $\hat{\theta} = ay$ , where  $a$  is a fixed constant subject to design. We consider the mean square error (MSE) as the distortion metric

$$\mathcal{D}_{\mathbf{W}}(a) \triangleq \mathbb{E}_{\theta, \tilde{\mathbf{h}}, \boldsymbol{\epsilon}, \tilde{\mathbf{g}}, u} [(\theta - ay)^2; \mathbf{W}]. \quad (5.13)$$

From the theory of linear minimum mean square estimation (LMMSE, see [49], Chapter 12), we readily obtain that

$$a_{\text{LMMSE}} \triangleq \arg \min_a \mathcal{D}_{\mathbf{W}}(a) = \frac{\mathbb{E}[y\theta]}{\mathbb{E}[y^2]}, \text{ and} \quad (5.14a)$$

$$D_{\mathbf{W}} \triangleq \mathcal{D}_{\mathbf{W}}(a_{\text{LMMSE}}) = \eta^2 - \frac{(\mathbb{E}[y\theta])^2}{\mathbb{E}[y^2]}, \quad (5.14b)$$

where the above expectations are w.r.t. all random variables  $\{\theta, \tilde{\mathbf{h}}, \boldsymbol{\epsilon}, \tilde{\mathbf{g}}, u\}$ . From (5.12a) and (5.12b), we obtain

$$\mathbb{E}[y^2] = \text{Tr} [\mathbf{E}_{\mathbf{g}} \mathbf{W} \mathbf{E}_{\mathbf{x}} \mathbf{W}^T] + \xi^2, \text{ and} \quad (5.15)$$

$$\mathbb{E}[y\theta] = \eta^2 \tilde{\mathbf{g}}^T \mathbf{W} \tilde{\mathbf{h}}, \text{ where}$$

$$\mathbf{E}_{\mathbf{g}} \triangleq \mathbb{E} [\tilde{\mathbf{g}} \tilde{\mathbf{g}}^T] = \mathbf{g} \mathbf{g}^T + \Sigma_{\mathbf{g}}, \quad (5.16)$$

where  $\mathbf{E}_x$  is defined in (5.9).

*Remark: Perfect observation gain and channel state information (OGI and CSI):* When the observation and channel gains are precisely known, i.e.,  $\Sigma_h = \Sigma_g = 0$ , (5.12b) reduces to a linear Gaussian model conditioned on  $\theta$ ,

$$y|\theta \sim \mathcal{N}(\mathbf{g}^T \mathbf{W} \mathbf{h} \theta, \mathbf{g}^T \mathbf{W} \Sigma \mathbf{W}^T \mathbf{g} + \xi^2), \quad (5.17)$$

and hence the LMMSE estimator is also the minimum mean square estimator (MMSE) [49],

$$\hat{\theta}_{\text{LMMSE}} := a_{\text{LMMSE}} y = \mathbb{E}_{\theta, \epsilon, u}[\theta|y] =: \hat{\theta}_{\text{MMSE}}, \quad (5.18)$$

i.e.,  $\hat{\theta}_{\text{LMMSE}}$  minimizes the distortion over all possible estimators (not just within the linear class).

We know from the theory of MMSE estimation (see [49, 81]) that the optimal distortion  $D_{\text{MMSE}}$  is related to the Fisher Information (FI),

$$J \triangleq -\mathbb{E}_{\theta} \left\{ \mathbb{E}_{y|\theta} \left\{ \frac{d^2 \theta}{d\theta^2} p(\mathbf{y}|\theta) \right\} \right\}, \quad (5.19)$$

by  $D_{\text{MMSE}} \geq \left( \frac{1}{\eta^2} + J \right)^{-1}$  in general and that equality holds for linear Gaussian models of the form (5.17). Though the FI (Equation (5.19)) results in

$$J = \frac{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}{\mathbf{g}^T \mathbf{W} \Sigma \mathbf{W}^T \mathbf{g} + \xi^2} \quad (5.20)$$

for the case of perfect OGI and perfect CSI (the linear Gaussian model in (5.17)), the FI is difficult to derive for cases when the observation and channel gains are uncertain. In fact, this is the main reason why we consider LMMSE estimation (which is suboptimal in general but easier to compute) rather than MMSE estimation (which is optimal but difficult to compute).

For the purposes of notation in this chapter, for all cases (whether we have perfect OGI/CSI or not), we would find it convenient to work with the quantity

$$J_{\mathbf{W}} \triangleq \frac{1}{D_{\mathbf{W}}} - \frac{1}{\eta^2}. \quad (5.21)$$

as a surrogate for the LMMSE distortion  $D_{\mathbf{W}}$  (as in (5.14b)). Note that  $D_{\mathbf{W}}$  and  $J_{\mathbf{W}}$  are monotonically related and that minimizing  $D_{\mathbf{W}}$  is equivalent to maximizing  $J_{\mathbf{W}}$ . Motivated by the preceding discussions, we would refer to  $J_{\mathbf{W}}$  as the *equivalent Fisher Information*, or sometimes simply FI or even distortion, for the sake of brevity.

### 5.2.5 Problem statement

The design of the collaboration weights  $\mathbf{W}$  is critical since it affects both the power requirements and estimation performance of the entire application. Specifically, the following quantities depend on  $\mathbf{W}$ , 1) the resources required to collaborate, i.e.,  $Q_{\text{nz}(\mathbf{W}),m}^2$  for individual nodes or  $Q_{\text{nz}(\mathbf{W})}$  cumulatively for the network (see (5.6) and (5.7)), 2) the resources required to transmit, i.e.,  $P_{\mathbf{W},m}$  for individual nodes or  $P_{\mathbf{W}}$  cumulatively for the network (see (5.8) and (6.4)), and 3) the final distortion of the estimate at the FC,  $D_{\mathbf{W}}$ , provided by (5.14b). In this chapter, we address the problems of designing the collaboration matrix that minimizes the distortion subject to either 1) a system-wide cumulative power constraint,

$$\begin{aligned} & \underset{\mathbf{W}}{\text{minimize}} && D_{\mathbf{W}} \\ & \text{subject to} && P_{\mathbf{W}} + Q_{\text{nz}(\mathbf{W})} \leq P^C, \end{aligned} \quad (5.22)$$

---

<sup>2</sup>Definition of operators  $\text{nz}(\cdot)$ ,  $\text{zero}(\cdot)$ , and  $\text{nnz}(\cdot)$ : The operator  $\text{nz} : \mathbb{R}^{N \times N} \rightarrow \{0, 1\}^{N \times N}$  is used to specify the non-zero elements of a matrix. If  $W_{ij} \neq 0$ , then  $[\text{nz}(\mathbf{W})]_{ij} = 1$ , else  $[\text{nz}(\mathbf{W})]_{ij} = 0$ . Similarly, the operator  $\text{zero} : \mathbb{R}^{N \times N} \rightarrow \{0, 1\}^{N \times N}$  is used to specify the zero elements of a matrix,  $[\text{zero}(\mathbf{W})]_{ij} = 1 - [\text{nz}(\mathbf{W})]_{ij}$ . The operator  $\text{nnz} : \mathbb{R}^{N \times N} \rightarrow \mathbb{Z}_+$  is used to specify the number of non-zero elements of a matrix.



or 2) power constraints at individual sensor nodes,

$$\begin{aligned}
& \underset{\mathbf{W}}{\text{minimize}} && D_{\mathbf{W}} \\
& \text{subject to} && P_{\mathbf{W},m} + Q_{\text{nz}(\mathbf{W}),m} \leq P_m^C, \quad m = 1, \dots, M.
\end{aligned} \tag{5.23}$$

We note that problem (5.23) (with  $M$  individual power constraints) is more realistic from a deployment point of view, since various nodes in a network can possess significantly different power sources, based on age of deployment or make/type of the batteries. However, problem (5.23) is significantly more difficult than problem (5.22), which has only one cumulative power constraint. Problem (5.22) is more important from a system design and analysis point of view, since it is more tractable analytically and as a result, reveals significant insights on the various system level tradeoffs.

### 5.2.6 Solution methodology

Both problems (5.22) and (5.23), in general, have no known procedure that efficiently computes (in polynomial-time) the globally optimal solution(s). However, *for the special case when the entries of the collaboration cost matrix  $\mathbf{C}$  are either zero or infinity,  $C_{ij} \in \{0, \infty\}$ , we will show that globally optimal solutions for both the problems can be obtained using efficient numerical techniques, and for problem (5.22), even a closed-form solution can be derived.*

Physically, this special case corresponds to the situation when the topology of a network is fixed (and hence not subject to design) and communication among neighbors is relatively inexpensive compared to communication with the FC. Let  $\mathbf{A} = \text{zero}(\mathbf{C})$  denote the permitted adjacency matrix for such a situation. Hence, the collaboration costs vanish, and problems (5.22) and (5.23) simplify to

$$\begin{aligned}
& \underset{\mathbf{W} \in \mathcal{S}_{\mathbf{A}}}{\text{minimize}} && D_{\mathbf{W}} \\
& \text{subject to} && P_{\mathbf{W}} \leq P^C, \quad \text{and}
\end{aligned} \tag{5.24}$$

$$\begin{aligned}
& \underset{\mathbf{W} \in \mathcal{S}_{\mathbf{A}}}{\text{minimize}} && D_{\mathbf{W}} \\
& \text{subject to} && P_{\mathbf{W},m} \leq P_m^{\mathbf{C}}, m = 1, \dots, M.
\end{aligned} \tag{5.25}$$

respectively, which are optimization problems in  $\text{nnz}(\mathbf{A})$  variables. Since problems (5.24) and (5.25), which will be solved in Section 5.3, arise out of the assumption of zero-cost for collaboration, we would refer to them as *ideal-collaborative* problems.

For the general case, when the topology is flexible and collaboration incurs a finite cost, a polynomial-time sub-optimal algorithm is proposed in Section 5.4 where the bigger problem is broken into smaller sub-problems, where several ideal-collaborative problems of the form (5.24) or (5.25) are solved at each iteration. Specifically, we start from the distributed topology  $\mathbf{A} = [\mathbf{I}_M | \mathbf{0}]$ , and follow a greedy algorithm to augment the collaborative topology with the most power-efficient link at each iteration.

### 5.2.7 Performance metric - Collaboration gain

Let the optimal solutions to problems (5.22) and (5.23) be denoted by  $D_{\text{opt}}(\mathbf{P}^{\mathbf{C}})$ , where (note bold notation)  $\mathbf{P}^{\mathbf{C}} = P^{\mathbf{C}}$  for a cumulative-constraint (problem (5.22)) and  $\mathbf{P}^{\mathbf{C}} = [P_1^{\mathbf{C}}, P_2^{\mathbf{C}}, \dots, P_M^{\mathbf{C}}]^T$  for individual constraints (problem (5.23)). Generally, the distortion  $D_{\text{opt}}(\mathbf{P}^{\mathbf{C}})$  depends on a number of problem conditions other than  $\mathbf{P}^{\mathbf{C}}$ , which includes 1) the source variance  $\eta^2$ , 2) the noise variance  $\Sigma$ , 3) the observation gain statistics  $\{\mathbf{h}, \Sigma_{\mathbf{h}}\}$ , 4) the coherent channel gain statistics  $\{\mathbf{g}, \Sigma_{\mathbf{g}}\}$ , and 5) the power needed to collaborate (matrix  $\mathbf{C}$ ). However, to assess and compare the benefits of spatial collaboration for a wide-range of problem conditions, we seek a metric that is normalized with respect to the operational region. Towards that goal, we define a few quantities. 1) Let

$$D_0 \triangleq D_{\text{opt}}(\mathbf{P}^{\mathbf{C}} \rightarrow \infty; \mathbf{A} = \mathbf{1}\mathbf{1}^T), \tag{5.26}$$

denote the optimal distortion that can be obtained with arbitrary collaboration and without any power constraints (the actual value of  $D_0$  will be derived later, see (5.51) for an early

preview). 2) Let

$$D_{\text{opt}}^{\text{dist}}(\mathbf{P}^{\text{C}}) \triangleq D_{\text{opt}}(\mathbf{P}^{\text{C}}; \mathbf{A} = [\mathbf{I}_M | \mathbf{0}]) \quad (5.27)$$

denote the optimal distortion for the distributed scenario, i.e., transmission power is optimally allocated among sensors and there is no collaboration among them. 3) Also let

$$D_{\text{opt}}^{\text{conn}}(\mathbf{P}^{\text{C}}) \triangleq D_{\text{opt}}(\mathbf{P}^{\text{C}}; \mathbf{A} = \mathbf{1}\mathbf{1}^T) \quad (5.28)$$

denote the optimal distortion for the fully connected collaborative topology. Note that

$$\underbrace{D_0}_{\substack{\text{Infinite power} \\ \text{Full collab.}}} \leq \underbrace{D_{\text{opt}}^{\text{conn}}(\mathbf{P}^{\text{C}})}_{\substack{\text{Finite power} \\ \text{Full collab.}}} \leq \underbrace{D_{\text{opt}}^{\text{dist}}(\mathbf{P}^{\text{C}})}_{\substack{\text{Finite power} \\ \text{No collab.}}} \leq \underbrace{\eta^2}_{\substack{\text{Zero power} \\ \text{(prior only)}}}, \quad (5.29)$$

where  $\eta^2$  is the worst-case distortion that corresponds to the prior information only. Equation (5.29) is illustrated in Figure 5.2, where a typical operational region is depicted alongwith the power-distortion tradeoff for distributed and connected topologies for the cumulative-constraint problem<sup>3</sup>. The goal of any estimation application is to close as much of the performance gap ( $\eta^2 - D_0$ ) as possible using limited resources and spatial collaboration is a tool that enables efficient allocation of those resources. We are now in a position to define *Collaboration Gain* (CG), the following normalized (centered and scaled) metric,

$$\text{CG} = \frac{D_{\text{opt}}^{\text{dist}}(\mathbf{P}^{\text{C}}) - D_{\text{opt}}^{\text{conn}}(\mathbf{P}^{\text{C}})}{\eta^2 - D_0}. \quad (5.30)$$

Note that  $0 \leq \text{CG} \leq 1$ , which means that efficacy of collaboration can now be summarized for a wide range of problem conditions. For example, if for a problem  $\text{CG} = 0.01$  (say), we might conclude that collaboration is not sufficiently beneficial for that particular problem. On the other hand, if  $\text{CG} = 0.2$  (say), we would conclude that spatial collaboration closes the realizable performance gap by 20% and hence, may be worth considering.

---

<sup>3</sup>As we shall see later, for the cumulative-constraint problem, all the available power must be used at optimality, i.e.,  $P_{\text{opt}} = P^{\text{C}}$ , and hence the subscript (C) is dropped from  $P^{\text{C}}$  in Figure 5.2.

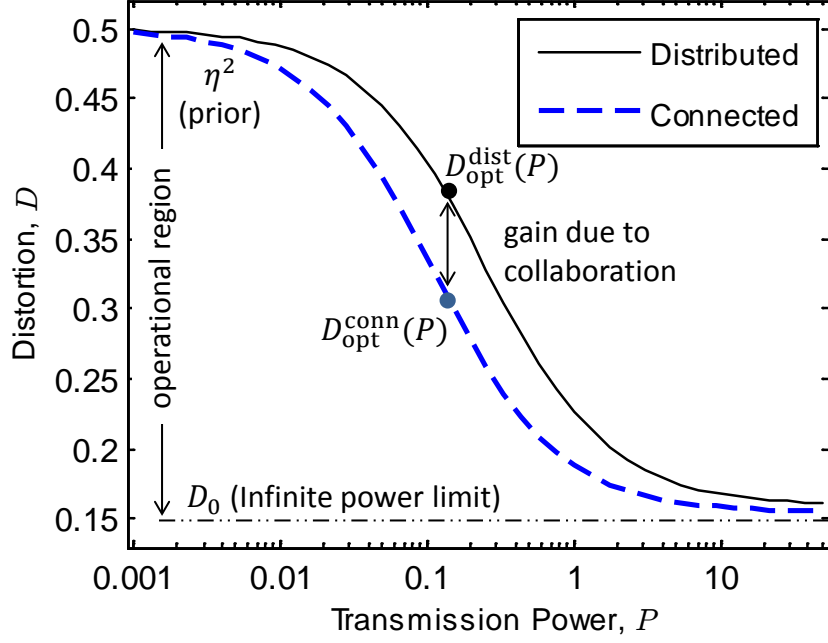


Figure 5.2: A typical power-distortion curve illustrating collaboration gain.

### 5.3 Main results - Ideal collaborative power allocation:

In this section, we consider the situation when the entries of the collaboration cost matrix  $\mathbf{C}$  are either zero or infinity,  $C_{ij} \in \{0, \infty\}$ , i.e., we will solve problems (5.24) and (5.25), where the topology  $\mathbf{A} = \text{zero}(\mathbf{C})$  is assumed to be fixed and not subject to design.

#### 5.3.1 Explicit formulation w.r.t. non-zero weights

From (5.14b) and (5.21), we note that minimizing the distortion  $D_{\mathbf{W}}$  is equivalent to maximizing the equivalent Fisher Information,

$$J_{\mathbf{W}} = \frac{(g^T \mathbf{W} \mathbf{h})^2}{\text{Tr} [\mathbf{E}_{\mathbf{g}} \mathbf{W} \mathbf{E}_{\mathbf{x}} \mathbf{W}^T] - \eta^2 (g^T \mathbf{W} \mathbf{h})^2 + \xi^2}, \quad (5.31)$$

where  $D_{\mathbf{W}}$  and  $J_{\mathbf{W}}$  are related by Equation (5.21). Part of the numerator and denominator of  $J_{\mathbf{W}}$  and also the expressions for power ((5.8) for individual and (6.4) for cumulative) are quadratic functions of the non-zero elements in  $\mathbf{W}$ . To see that explicitly, we concatenate

the elements of  $\mathbf{W}$  (column-wise, only those that are allowed to be non-zero), in  $\mathbf{w} = [w_1, w_2, \dots, w_L]^T$ . For  $l = 1, 2, \dots, L$ , define indices  $m_l$  and  $n_l$  such that  $w_l = \mathbf{W}_{m_l, n_l}$ . Further, we define  $L \times L$  matrices  $\mathbf{\Omega}_{\text{JN}}$ ,  $\mathbf{\Omega}_{\text{JD}}$ ,  $\mathbf{\Omega}_{\text{P}, m}$ ,  $\mathbf{\Omega}_{\text{P}} \triangleq \sum_{m=1}^M \mathbf{\Omega}_{\text{P}, m}$  and  $L \times N$  matrix  $\mathbf{G}$  such that the following identities,

$$\mathbf{g}^T \mathbf{W} = \mathbf{w}^T \mathbf{G}, \quad (5.32a)$$

$$J_{\mathbf{W}} = \frac{\overbrace{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}^{= \mathbf{w}^T \mathbf{\Omega}_{\text{JN}} \mathbf{w}}}{\underbrace{\text{Tr} [\mathbf{E}_{\text{g}} \mathbf{W} \mathbf{E}_{\text{x}} \mathbf{W}^T] - \eta^2 (\mathbf{g}^T \mathbf{W} \mathbf{h})^2}_{= \mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w}} + \xi^2}, \quad (5.32b)$$

$$P_{\mathbf{W}} = \underbrace{\text{Tr} [\mathbf{W} \mathbf{E}_{\text{x}} \mathbf{W}^T]}_{\mathbf{w}^T \mathbf{\Omega}_{\text{P}} \mathbf{w}} = \sum_{m=1}^M \underbrace{[\mathbf{W} \mathbf{E}_{\text{x}} \mathbf{W}^T]_{m, m}}_{\mathbf{w}^T \mathbf{\Omega}_{\text{P}, m} \mathbf{w}} \quad (5.32c)$$

are satisfied. Precisely, the elementwise descriptions for all the matrices are as follows,

$$\begin{aligned} [\mathbf{G}]_{l, n} &= \begin{cases} g_{m_l}, & n = n_l, \\ 0, & \text{otherwise} \end{cases}, \\ [\mathbf{\Omega}_{\text{JN}}]_{k, l} &= g_{m_k} g_{m_l} h_{n_k} h_{n_l} \Leftrightarrow \mathbf{\Omega}_{\text{JN}} = \mathbf{G} \mathbf{h} \mathbf{h}^T \mathbf{G}^T, \\ [\mathbf{\Omega}_{\text{JD}}]_{k, l} &= [\mathbf{E}_{\text{g}}]_{m_k, m_l} [\mathbf{E}_{\text{x}}]_{n_k, n_l} - \eta^2 [\mathbf{\Omega}_{\text{JN}}]_{k, l}, \text{ and} \\ [\mathbf{\Omega}_{\text{P}, m}]_{k, l} &= \begin{cases} [\mathbf{E}_{\text{x}}]_{n_k, n_l}, & m_k = m_l = m, \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \quad (5.33)$$

for  $k, l = 1, 2, \dots, L$ ,  $n = 1, 2, \dots, N$  and  $m = 1, 2, \dots, M$ .

Though  $\mathbf{\Omega}_{\text{JN}}$  is rank-1 (as described above), in general, there are no compact expressions for the matrices  $\mathbf{\Omega}_{\text{JD}}$  and  $\mathbf{\Omega}_{\text{P}, m}$ . We illustrate some relevant matrix definitions ( $\mathbf{\Omega}_{\text{P}, m}$  and  $\mathbf{G}$  in particular) through an example, in Figure 5.3, with  $N = 4$  data-collection nodes,  $M = 3$  communicating nodes and 3 collaborating links, resulting in a total of  $L = 6$  non-zero coefficients in the collaboration matrix  $\mathbf{W}$ .

For some special cases and regular topologies, more compact expressions for  $\mathbf{\Omega}_{\text{JD}}$  may be derived. For the special case when perfect channel state information is available ( $\Sigma_{\text{g}} = 0$ ),

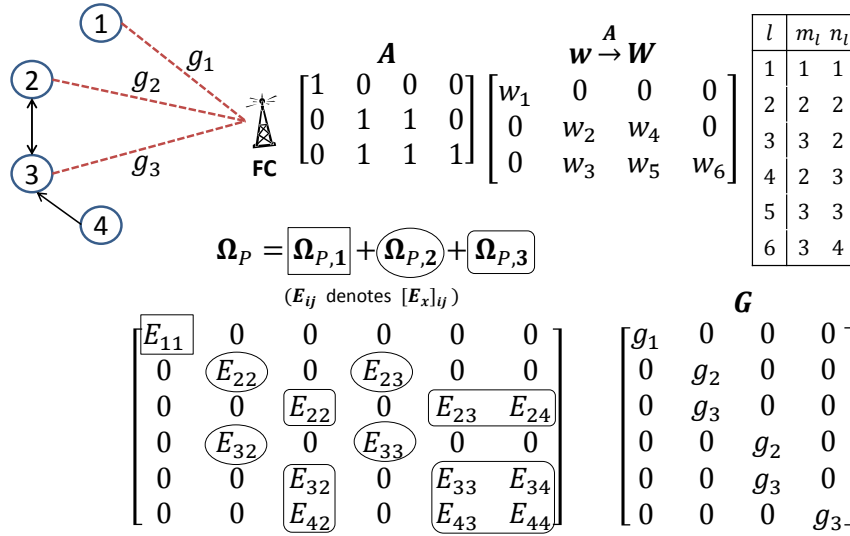


Figure 5.3: Matrices  $\mathbf{G}$  and  $\mathbf{\Omega}_P$  assuming columnwise ordering of elements in  $\mathbf{W}$

it is easy to see that

$$\begin{aligned}
 \mathbf{\Omega}_{\text{JD}} &= \mathbf{G} (\mathbf{E}_x - \eta^2 \mathbf{h} \mathbf{h}^T) \mathbf{G}^T \\
 &= \mathbf{G} \tilde{\mathbf{\Sigma}} \mathbf{G}^T, \text{ where } \tilde{\mathbf{\Sigma}} \triangleq \mathbf{\Sigma} + \eta^2 \mathbf{\Sigma}_h.
 \end{aligned} \tag{5.34}$$

This simplification will also be useful later in our discussion.

With the help of these definitions, the objective function (Fisher Information) is simplified as

$$J_w \triangleq \frac{\mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w}}{\mathbf{w}^T \mathbf{\Omega}_{\text{JD}} \mathbf{w} + \xi^2}, \tag{5.35}$$

and problems (5.24) and (5.25) are re-written as

$$\begin{aligned}
 &\underset{\mathbf{w}}{\text{maximize}} \quad J_w \\
 &\text{subject to} \quad \mathbf{w}^T \mathbf{\Omega}_P \mathbf{w} \leq P^c, \text{ and}
 \end{aligned} \tag{5.36}$$

$$\begin{aligned}
 &\underset{\mathbf{w}}{\text{maximize}} \quad J_w \\
 &\text{subject to} \quad \mathbf{w}^T \mathbf{\Omega}_{P,m} \mathbf{w} \leq P_m^c, \quad m = 1, \dots, M.
 \end{aligned} \tag{5.37}$$

respectively, which are both optimization problems with the same fractional-quadratic objective (5.35) and single (or multiple) quadratic constraint(s). Solution of problems (5.36) and (5.37) will be provided in Sections 5.3.2 and 5.3.3 respectively.

### 5.3.2 Cumulative power constraint

Since multiplying  $\mathbf{w}$  by a scalar  $\alpha > 1$  (strictly) increases both  $J_{\mathbf{w}}$  and power  $\mathbf{w}^T \mathbf{\Omega}_{\mathbf{P}} \mathbf{w}$  (and for  $\alpha < 1$ , strictly decreases them), problem (5.36) is equivalent to its converse formulation, where power is minimized subject to a maximum distortion constraint (represented by  $J^{\mathbf{C}}$ ),

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \mathbf{\Omega}_{\mathbf{P}} \mathbf{w} \\ & \text{subject to} && J_{\mathbf{w}} \geq J^{\mathbf{C}}, \end{aligned} \tag{5.38}$$

in the sense that the optimal solutions  $J_{\text{opt}}(P^{\mathbf{C}})$  (of (5.36)) and  $P_{\text{opt}}(J^{\mathbf{C}})$  (of (5.38)) are inverses of one another. Moreover, the optimal solutions hold with active constraints (satisfying equalities  $P = P^{\mathbf{C}}$  for (5.36) and  $J = J^{\mathbf{C}}$  for (5.38)). From (5.35), problem (5.38) is further equivalent to,

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \mathbf{\Omega}_{\mathbf{P}} \mathbf{w} \\ & \text{subject to} && \mathbf{w}^T (J \mathbf{\Omega}_{\mathbf{JD}} - \mathbf{\Omega}_{\mathbf{JN}}) \mathbf{w} + J \xi^2 \leq 0, \end{aligned} \tag{5.39}$$

which is a quadratically constrained quadratic program (QCQP) in  $L \triangleq \text{nnz}(\mathbf{A})$  variables. Note in (5.39) that, though  $\mathbf{\Omega}_{\mathbf{P}}$  is positive definite (it is composed of blocks of another positive definite matrix  $\mathbf{E}_{\mathbf{x}}$ , see Figure 5.3, for example), the matrix  $J \mathbf{\Omega}_{\mathbf{JD}} - \mathbf{\Omega}_{\mathbf{JN}}$  is not, and hence problem (5.39) is not convex. However, a QCQP with exactly one constraint (as in problem (5.39)) still satisfies strong duality (for a background, see Appendix B, [7]) and hence the optimal solution to (5.39) satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$(\mathbf{\Omega}_{\mathbf{P}} + \mu (J \mathbf{\Omega}_{\mathbf{JD}} - \mathbf{\Omega}_{\mathbf{JN}})) \mathbf{w} = 0. \tag{5.40}$$

Together with the following active constraint conditions at optimality

$$P = \mathbf{w}^T \mathbf{\Omega}_P \mathbf{w}, \quad \text{and} \quad \mathbf{w}^T (J \mathbf{\Omega}_{JD} - \mathbf{\Omega}_{JN}) \mathbf{w} + J \xi^2 = 0, \quad (5.41)$$

which implies  $\mu = \frac{P}{J\xi^2}$ , the solution to problem (5.39) (equivalently, problems (5.38), (5.36) and (5.24)) is summarized below.

**Theorem 5.3.1.** (Power-Distortion tradeoff for Linear Coherent Ideal-Collaborative Estimation) *For a given topology  $\mathbf{A}$ , let  $J \in (0, \lambda_{\mathcal{G}, \max}(\mathbf{\Omega}_{JN}, \mathbf{\Omega}_{JD}))^4$ . The optimal tradeoff between distortion (represented by  $J$ ) and cumulative transmission power  $P$  and also the optimal weights  $\mathbf{w}$  that achieve that tradeoff, are obtained through the solution of the generalized eigenvalue problem,*

$$\left( \frac{\mathbf{\Omega}_P}{P_\xi} - \frac{\mathbf{\Omega}_{JN}}{J} + \mathbf{\Omega}_{JD} \right) \mathbf{w} = 0, \quad \text{where } P_\xi \triangleq \frac{P}{\xi^2}. \quad (5.42)$$

*In particular, the function  $J_{\text{opt}}(P) : (0, \infty) \rightarrow (0, J_0^A)$  and its inverse  $P_{\text{opt}}(J)$  are*

$$J_{\text{opt}}(P) = \lambda_{\mathcal{G}, \max} \left( \mathbf{\Omega}_{JN}, \mathbf{\Omega}_{JD} + \frac{\mathbf{\Omega}_P}{P_\xi} \right), \quad \text{and} \quad (5.43a)$$

$$P_{\text{opt}}(J) = \lambda_{\mathcal{G}, \min}^{\text{pos}} \left( \mathbf{\Omega}_P, -\mathbf{\Omega}_{JD} + \frac{\mathbf{\Omega}_{JN}}{J} \right) \xi^2, \quad (5.43b)$$

*respectively. This optimal tradeoff is achieved when the weights of collaboration matrix is an appropriately scaled version of the (generalized) eigenvector corresponding to (5.43a) (or (5.43b), since they are equivalent), say  $\mathbf{v}_{\text{opt}}$ . That is,  $\mathbf{w}_{\text{opt}} = c \mathbf{v}_{\text{opt}}$ , where the scalar  $c$  is such that  $\mathbf{w}_{\text{opt}}^T \mathbf{\Omega}_P \mathbf{w}_{\text{opt}} = P$ .*

Theorem 5.3.1 is important since it helps to (numerically) compute the power-distortion tradeoff for arbitrary problem conditions (like topology, noise covariance, second-order

---

<sup>4</sup>Definitions of eigenvalue related operators: The operators  $\lambda(\mathbf{P})$  and  $\mathbf{v}(\mathbf{P})$  denote the solution(s) to the ordinary eigenvalue problem  $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$ . Operator  $\lambda_{\max}(\cdot)$  denotes the maximum among all real eigenvalues and  $\lambda_{\min}^{\text{pos}}(\cdot)$  denotes the minimum among all positive eigenvalues (i.e., the positive eigenvalue that is closest to, but different from, zero). The operators  $\lambda_{\mathcal{G}}(\mathbf{P}, \mathbf{Q})$  and  $\mathbf{v}_{\mathcal{G}}(\mathbf{P}, \mathbf{Q})$  denote the solution(s) to the generalized eigenvalue problem  $\mathbf{P}\mathbf{v} = \lambda\mathbf{Q}\mathbf{v}$ . Operators  $\lambda_{\mathcal{G}, \max}(\cdot, \cdot)$  and  $\lambda_{\mathcal{G}, \min}^{\text{pos}}(\cdot, \cdot)$  are similarly defined as  $\lambda_{\max}(\cdot)$  and  $\lambda_{\min}^{\text{pos}}(\cdot)$  respectively. Note that, when  $\mathbf{Q}$  is full-rank, then  $\lambda_{\mathcal{G}}(\mathbf{P}, \mathbf{Q}) = \lambda(\mathbf{Q}^{-1}\mathbf{P})$ .



statistics of the observation and channel gains). Since the numerical complexity for eigenvalue problems is roughly cubic in the size of the problem (see for example [68]), the complexity of computing the (cumulative) power-distortion tradeoff is  $\mathcal{O}(L^3)$ , where (recall that)  $L = \text{nnz}(\mathbf{A})$  is the number of non-zero collaboration weights.

Corresponding to the example topology in Figure 5.3 and randomly chosen system parameters  $\mathbf{h}, \mathbf{\Sigma}$  and  $\mathbf{g}$ , a typical power-distortion tradeoff curve is shown in Figure 5.2 (bold line). Theorem 5.3.1 can be simplified further for several specific scenarios, allowing deeper insight into the power-distortion tradeoff as it relates to the problem parameters. The first obvious simplification is because of the rank-1 property of  $\mathbf{\Omega}_{\text{JD}}$ . The only non-zero generalized eigenvalue is (provided the inverse exists)

$$J_{\text{opt}}(P) = \mathbf{h}^T \mathbf{G}^T \left( \mathbf{\Omega}_{\text{JD}} + \frac{\mathbf{\Omega}_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{G} \mathbf{h}, \quad (5.44a)$$

$$\text{with eigenvector } \mathbf{w}_{\text{opt}} \propto \left( \mathbf{\Omega}_{\text{JD}} + \frac{\mathbf{\Omega}_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{G} \mathbf{h}. \quad (5.44b)$$

Equation (5.44a) explicitly shows the effect of finite-power constraint  $P_{\xi}$  on the distortion. Some other insightful examples are discussed next.

*Example 1:* For the case of *perfect CSI* ( $\mathbf{\Sigma}_{\text{g}} = 0$ ), we note from (5.34) that  $\mathbf{\Omega}_{\text{JD}} = \mathbf{G} \tilde{\mathbf{\Sigma}} \mathbf{G}^T$ . It follows that

$$J_{\text{opt}}(P) = \lambda_{\mathcal{G}, \max} \left( \mathbf{G} \mathbf{h} \mathbf{h}^T \mathbf{G}^T, \mathbf{G} \tilde{\mathbf{\Sigma}} \mathbf{G}^T + \frac{\mathbf{\Omega}_{\text{P}}}{P_{\xi}} \right) \quad (5.45)$$

$$\stackrel{(a)}{=} \lambda_{\mathcal{G}, \max} \left( \mathbf{h} \mathbf{h}^T, \tilde{\mathbf{\Sigma}} + \frac{\mathbf{\Gamma}_{\text{P}}}{P_{\xi}} \right), \quad \mathbf{\Gamma}_{\text{P}} \triangleq (\mathbf{G}^T \mathbf{\Omega}_{\text{P}}^{-1} \mathbf{G})^{-1} \quad (5.46)$$

$$\stackrel{(b)}{=} \mathbf{h}^T \left( \tilde{\mathbf{\Sigma}} + \frac{\mathbf{\Gamma}_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}, \quad (5.47)$$

assuming all the inverses exist. Step (a) reduces the size of the eigenvalue problem from  $L$  to  $N$  yet preserving the non-zero eigenvalues. Note that the corresponding generalized eigenvectors of problems (5.45) (say  $\mathbf{v}_L$ ) and (5.46) (say  $\mathbf{v}_N$ ) are related by  $\mathbf{v}_L = \mathbf{\Omega}_{\text{P}}^{-1} \mathbf{G} \mathbf{\Gamma}_{\text{P}} \mathbf{v}_N$ . Step (b) describes the only non-zero generalized value of problem (5.46), since  $\mathbf{h} \mathbf{h}^T$  is rank-1. Note that optimal collaboration weights are provided by  $\mathbf{w}_{\text{opt}} \propto \mathbf{\Omega}_{\text{P}}^{-1} \mathbf{G} \tilde{\mathbf{\Sigma}} \left( \tilde{\mathbf{\Sigma}} + \frac{\mathbf{\Gamma}_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}$ .

When, in addition to *perfect CSI*, we also have *perfect OGI* ( $\Sigma_{\mathbf{h}} = 0$ ), Equation (5.47) further simplifies to

$$J_{\text{opt}}(P) = \mathbf{h}^T \left( \Sigma + \frac{\Gamma_{\mathbf{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}, \quad (5.48)$$

which is an insightful closed form expression. Equation (5.48) can be compared to the centralized case, where measurements  $x_n$  are directly observed through gains  $\mathbf{h}$  and measurement noise with variance  $\Sigma$ , for which the Fisher Information is  $J_{\text{cent}} \triangleq \mathbf{h}^T \Sigma^{-1} \mathbf{h}$ , which is also the infinite power limit of (5.48). One can think of the additional quantity  $\frac{\Gamma_{\mathbf{P}}}{P_{\xi}}$  in (5.48), which factors in the effect of channels, the collaboration topology and finite transmission power, as equivalent to the variance of an additional noise that is added to the measurement noise.

When, in addition to *perfect OGI and CSI*, we also have a *distributed* topology ( $\mathbf{A} = \mathbf{I}_M$ ) and the measurement noise is uncorrelated ( $\Sigma$  is diagonal), we can proceed as follows. We have  $\mathbf{w} = \text{diag}(\mathbf{W})^5$ . This means that  $\Omega_{\mathbf{P}} = \text{diag}(\text{diag}(\Sigma + \eta^2 \mathbf{h} \mathbf{h}^T))$  is a diagonal matrix with  $m^{\text{th}}$  element as  $\sigma_m^2 + \eta^2 h_m^2$ , and  $\mathbf{G} = \text{diag}(\mathbf{g})$ . Consequently,  $\Gamma_{\mathbf{P}}$  is a diagonal matrix with  $m^{\text{th}}$  element as  $\frac{\sigma_m^2 + \eta^2 h_m^2}{g_m^2}$ . Hence, the optimal power-distortion tradeoff in Equation (5.48) further simplifies to (define  $\sigma_m^2 \triangleq \Sigma_{m,m}$  and  $\gamma_m \triangleq \frac{\eta^2 h_m^2}{\sigma_m^2}$ )

$$J_{\text{opt}}(P) = \sum_{m=1}^M \frac{h_m^2}{\sigma_m^2} \left[ 1 + \frac{1 + \gamma_m}{P_{\xi} g_m^2} \right]^{-1}, \quad (5.49)$$

which was also obtained in [96]. Since for the centralized case, we have  $J_{\text{cent}} = \sum_{m=1}^M \frac{h_m^2}{\sigma_m^2}$ , Equation (5.49) indicates the exact fractions of individual Fisher Information that “reaches” the receiver. When subjected to a network-wide power constraint, information from the more informative (higher  $\gamma_m$ ) and less reliable (lower  $g_m$ ) sensor undergoes a higher degree of “attenuation”. While a higher observation gain  $h_m$  clearly carries more information, it

---

<sup>5</sup>Definition of operators  $\text{diag}(\cdot)$  and  $\text{vec}(\cdot)$ : While operating on a matrix,  $\text{diag} : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{\min(M,N)}$  is used to extract the diagonal elements. While operating on a vector,  $\text{diag} : \mathbb{R}^M \rightarrow \mathbb{R}^{M \times M}$  is used to construct a matrix by specifying only the diagonal elements, the other elements being zero. The vectorization operator  $\text{vec} : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{MN}$  stacks up all the elements of a matrix column-by-column.

also requires quadratically higher power to transmit in an amplify-and-forward framework such as ours (note that  $P_m = w_m^2(\sigma_m^2 + \eta^2 h_m^2)$ ). Similarly, a lower magnitude of channel gain  $g_m$  implies that quadratically higher transmission power is needed to compensate for the channel. Hence, in the optimal tradeoff (5.49), it turns out that information from higher- $h_m$  and lower- $g_m$  sensors are attenuated by a larger factor.

*Example 2:* When the network is fully *connected* ( $\mathbf{A} = \mathbf{1}\mathbf{1}^T$ ), we proceed from (5.44a) and obtain the following result.

**Proposition 5.3.2.** *The optimal solution for Example 2 is*

$$\begin{aligned} J_{\text{opt}}(P) &= \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{\mathcal{G}} \right]^{-1} \quad \text{and } W_{\text{opt}} \propto \mathbf{u}\mathbf{v}^T, \\ \text{where } \tilde{J} &\triangleq \mathbf{h}^T \tilde{\Sigma}^{-1} \mathbf{h}, \mathcal{G} \triangleq \mathbf{g}^T \tilde{\Sigma}_{\mathbf{g}}^{-1} \mathbf{g}, \mathbf{u} = \tilde{\Sigma}_{\mathbf{g}}^{-1} \mathbf{g}, \\ \mathbf{v} &= \tilde{\Sigma}^{-1} \mathbf{h}, \tilde{\Sigma} \triangleq \Sigma + \eta^2 \Sigma_h \text{ and } \tilde{\Sigma}_{\mathbf{g}} \triangleq \Sigma_{\mathbf{g}} + \frac{\mathbf{I}}{P_{\xi}}. \end{aligned} \quad (5.50)$$

Also, for the special case with perfect OGI and perfect CSI (when  $\Sigma_h = \Sigma_{\mathbf{g}} = 0$ ), the resulting distortion is information theoretically optimal.

*Proof:* See Appendix A.6. *QED.*

This example is important since a connected topology makes use of all possible collaboration links and hence Equation (5.50) gives the LMMSE performance limit for a network with cumulative transmission power constraint. From (5.50), we are now in a position to compute the optimal achievable Fisher Information  $J_0$  (and equivalently the distortion  $D_0$ ), by letting the power go to infinity,

$$J_0 = \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{\mathcal{G}_0} \right]^{-1}, \quad \text{where } \mathcal{G}_0 \triangleq \mathbf{g}^T \Sigma_{\mathbf{g}}^{-1} \mathbf{g}. \quad (5.51)$$

As discussed earlier,  $D_0$  (definition in (5.26)) forms a lower bound for distortion in a network and is useful to characterize the operational region and subsequently define the collaboration gain (Equation (5.30)).

From Equation (5.50), it can be explicitly seen that for a fully connected topology, more observation or channel uncertainty always deteriorate the estimation performance, a notion that is intuitive but still was not completely evident in the discussion so far. The following result formalizes this notion.

**Proposition 5.3.3.** *If  $\Sigma_{h,1} \succ \Sigma_{h,2}$  and  $\Sigma_{g,1} \succ \Sigma_{g,2}$  (here  $\mathbf{A} \succ \mathbf{B}$  implies that  $\mathbf{A} - \mathbf{B}$  is a symmetric positive definite matrix), then for a connected topology,*

$$J_{\text{opt},1}(P) < J_{\text{opt},2}(P). \quad (5.52)$$

*Proof:* This is established by re-writing Equation (5.50) as

$$J_{\text{opt}}(P) = \left[ \left( 1 + \frac{1}{\mathcal{G}} \right) \frac{1}{\tilde{J}} + \frac{\eta^2}{\mathcal{G}} \right]^{-1}, \quad (5.53)$$

which shows explicitly that  $J_{\text{opt}}(P)$  is a monotonically increasing function of  $\tilde{J}$  and  $\mathcal{G}$ . Recall that  $\tilde{J} = \mathbf{h}^T (\Sigma + \eta^2 \Sigma_h)^{-1} \mathbf{h}$  and  $\mathcal{G} = \mathbf{g}^T \tilde{\Sigma}_g^{-1} \mathbf{g}$ . Assuming  $\Sigma_{h,1} \succ \Sigma_{h,2}$  and  $\Sigma_{g,1} \succ \Sigma_{g,2}$ , it is sufficient to show that a)  $\tilde{J}_1 < \tilde{J}_2$  and b)  $\mathcal{G}_1 < \mathcal{G}_2$ , both of which are evident from the inequality involving positive definite matrices in Lemma A.7.1 (see Appendix A.7). *QED.*

The last two examples provided insight on the distributed and fully connected topologies respectively. The following example addresses a partially connected topology and shows how the distortion decreases with an increase in links available for collaboration.

*Partially connected cycle graphs:* In Figure 5.4, we display a class of graphs, namely the  $(K - 1)$  connected directed cycle, for  $K = 1, 2, \dots, M$ , in which each node shares its observations with the next  $K - 1$  nodes. The adjacency topology of such a graph will be denoted as  $\mathbf{A} = \mathcal{C}(K)$ . Note that  $K = 1$  denotes the distributed scenario while  $K = M$  denotes the fully connected scenario.

*Example 3:* We assume that the collaborative topology is a  $(K - 1)$  connected directed cycle ( $\mathbf{A} = \mathcal{C}(K)$ ) and the channel gain and uncertainties are such that the network is *homogeneous and equicorrelated*. In particular, we denote a) the *expected* observation and channel gains by  $h_0^2$  and  $g_0^2$ , b) the observation and channel gain uncertainties by  $\alpha_h$  and

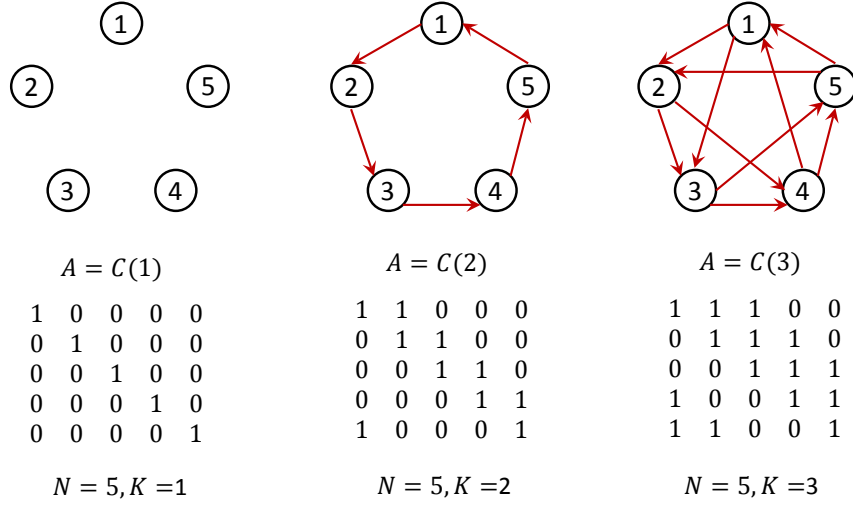


Figure 5.4: Directed cycle graphs,  $(K - 1)$  connected.

$\alpha_g$ , and c) measurement noise variance and correlation by  $\sigma^2$  and  $\rho$ , and thereby assume

$$\begin{aligned}
 \mathbf{h} &= h_0 \sqrt{\alpha_h} \mathbf{1}, \Sigma_h = h_0^2 (1 - \alpha_h) \mathbf{I}, \\
 \mathbf{g} &= g_0 \sqrt{\alpha_g} \mathbf{1}, \Sigma_g = g_0^2 (1 - \alpha_g) \mathbf{I}, \text{ and} \\
 \Sigma &= \sigma^2 ((1 - \rho) \mathbf{I} + \rho \mathbf{1} \mathbf{1}^T).
 \end{aligned} \tag{5.54}$$

These assumptions provide an analytically tractable example that is representative of a broad range of problem conditions and will also be used in the subsequent discussions. In addition to covering the partial collaboration ( $1 \leq K \leq N$ ) regime, note that  $\alpha_h = 1$  implies perfect OGI,  $\alpha_g = 1$  implies perfect CSI and  $\rho = 0$  implies uncorrelated measurement noise. Equation (5.44a) can be simplified further for this example, to obtain the following result.

**Proposition 5.3.4.** *The optimal solution for Example 3 is*

$$\begin{aligned}
 J_{opt}(P) &= \frac{h_0^2}{\sigma^2} \left[ \rho_N + \tilde{\alpha}_h \left( \rho_N + \frac{\gamma}{N} \right) \right. \\
 &\quad \left. + \frac{1}{N} \left( \tilde{\alpha}_g + \frac{1}{P_\xi g_0^2 \alpha_g} \right) \left\{ \gamma + \rho_K + \tilde{\alpha}_h \left( \rho_K + \frac{\gamma}{K} \right) \right\} \right]^{-1}, \\
 \text{where } \rho_t &\triangleq \rho + \frac{1 - \rho}{t}, \gamma \triangleq \frac{\eta^2 h_0^2}{\sigma^2}, \text{ and } \tilde{\alpha} \triangleq \frac{1}{\alpha} - 1.
 \end{aligned} \tag{5.55}$$

*Proof:* See Appendix A.8. *QED.*

Equation (5.55) helps us quantify the efficacy of collaboration in a partially connected network. For example, when the measurement noise is uncorrelated ( $\rho = 0$ ) and we have perfect OGI ( $\alpha_h = 1$ , equivalently  $\tilde{\alpha}_h = 0$ ) and perfect CSI ( $\alpha_g = 1$ , equivalently  $\tilde{\alpha}_g = 0$ ), Equation (5.55) reduces to

$$J = \frac{h_0^2 g_0^2}{\sigma^2 \xi^2} \frac{NP}{\gamma + \frac{1}{K}}. \quad (5.56)$$

For the high- $(NP)$  regime, (i.e., when  $J$  is large and  $D \approx J^{-1}$ ), we can now compare the power requirements of a distributed topology (say  $P^{\text{dist}}$ , for  $K = 1$ ) with that of a  $(K - 1)$  connected topology (say  $P^{\mathcal{C}(K)}$ ), provided an identical distortion performance is desired. Since  $\frac{P^{\text{dist}}}{\gamma + 1} = \frac{P^{\mathcal{C}(K)}}{\gamma + \frac{1}{K}}$ , this implies that the relative savings in power is

$$\frac{P^{\text{dist}} - P^{\mathcal{C}(K)}}{P^{\text{dist}}} = \frac{1 - \frac{1}{K}}{\gamma + 1}, \quad (5.57)$$

which depends on the local-SNR  $\gamma$ . For example, when the local-SNR is large, say  $\gamma = 100$ , then even a fully connected network (large  $K$ ) can provide only 1% power savings. On the other hand, if the local-SNR is small, say  $\gamma = 1$ , then even a 1-connected network (for which  $K = 2$ ) can provide 25% efficiency in power savings. The conclusion is that one needs to be judicious in the design of collaborative topologies, especially when there are overhead costs associated with it. The design of collaborative topologies with finite collaboration cost will be discussed later in Section 5.4.

### 5.3.3 Individual power constraints

In the previous subsection, we have discussed the solution to the ideal-collaborative power allocation problem with a cumulative transmission power constraint. In this subsection, we consider the case when the individual nodes have separate power constraints. We recall

problem (5.37),

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & J_{\mathbf{w}} = \frac{\mathbf{w}^T \boldsymbol{\Omega}_{\text{JN}} \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Omega}_{\text{JD}} \mathbf{w} + \xi^2} \\ \text{subject to} \quad & \mathbf{w}^T \boldsymbol{\Omega}_{\text{P},m} \mathbf{w} \leq P_m^C, \quad m = 1, \dots, M. \end{aligned} \quad (5.58)$$

Let  $J_{\text{opt}}$  be the optimal solution of problem (5.58), although it is not clear yet how this solution can be obtained. Unlike problem (5.36) for which a closed form solution was derived in Section 5.3.2, in general, there are no known closed form expressions of  $J_{\text{opt}}$ . However, as we shall show below,  $J_{\text{opt}}$  can still be precisely obtained using an efficient (polynomial time) numerical procedure. For some special cases though, closed form and insightful expressions for  $J_{\text{opt}}$  can be obtained, which will be discussed later.

## Numerical Solution

We would use the semi-definite relaxation (SDR) technique for quadratically constrained problems [35], [59] to solve problem (5.58). The SDR technique is widely used in the literature since it can reduce an otherwise intractable problem to one with polynomial time complexity. However, the main drawback of SDR technique is that, in general, it can guarantee only a sub-optimal solution. *But, in some special problems, which includes our problem at hand (as we shall establish later), the relaxation involved in SDR is exact and hence, the SDR technique becomes an efficient numerical tool to obtain the (precisely) optimal solution.* For more details on the SDR technique, including sub-optimality analysis for special classes of problems, the reader is referred to [59].

We proceed using arguments similar to [35], [98]. Define

$$\mathbf{X} \triangleq \mathbf{w} \mathbf{w}^T \in \mathbb{R}^{L \times L}, \quad (5.59)$$

so that problem (5.58) is equivalent to,

$$\begin{aligned}
& \underset{\mathbf{X}}{\text{maximize}} && J_{\mathbf{X}} = \frac{\text{Tr} [\boldsymbol{\Omega}_{\text{JN}} \mathbf{X}]}{\text{Tr} [\boldsymbol{\Omega}_{\text{JD}} \mathbf{X}] + \xi^2} \\
& \text{subject to} && \text{Tr} [\boldsymbol{\Omega}_{\text{P},m} \mathbf{X}] \leq P_m^{\text{C}}, m = 1, \dots, M, \\
& && \text{rank } \mathbf{X} = 1, \mathbf{X} \succeq 0,
\end{aligned} \tag{5.60}$$

Problem (5.60) is further equivalent to,

$$\begin{aligned}
& \text{find maximum} && J \\
& \text{such that} && \mathcal{X}(J) \text{ is nonempty,}
\end{aligned} \tag{5.61}$$

where  $\mathcal{X}(J)$  is defined as the following (feasible) set,

$$\mathcal{X}(J) \triangleq \left\{ \mathbf{X} \left| \begin{array}{l} \text{Tr} [(J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}) \mathbf{X}] + J\xi^2 \leq 0, \\ \text{Tr} [\boldsymbol{\Omega}_{\text{P},m} \mathbf{X}] \leq P_m^{\text{C}}, m = 1, \dots, M, \\ \text{rank } \mathbf{X} = 1, \mathbf{X} \succeq 0. \end{array} \right. \right\} \tag{5.62}$$

Note that by definition, any  $\mathbf{X} \in \mathcal{X}(J_{\text{opt}})$  will correspond to the optimal weights that maximize the Fisher Information. Also note that  $\mathcal{X}(J_1) \supset \mathcal{X}(J_2)$  when  $0 \leq J_1 < J_2$ . Therefore, assuming we can test the feasibility of  $\mathcal{X}(J)$  for some  $J$ , a simple bisection search over  $[0, J_0]$  can potentially yield  $J_{\text{opt}}$  with arbitrary accuracy<sup>6</sup> (since  $J_{\text{opt}} < J_0$  for finite power). However, testing the feasibility of  $\mathcal{X}(J)$  is a difficult problem.

Though the set of symmetric positive-semidefinite matrices is convex and the other  $M + 1$  inequalities in (5.62) are also convex,  $\mathcal{X}(J)$  is still not convex due to the rank

---

<sup>6</sup>Arbitrary accuracy is only of theoretical interest, since the solution to the feasibility problem (5.61) will have numerical errors. Hence, a more realistic stopping criterion is a fixed number of iterations, say 15.



constraint  $\text{rank } \mathbf{X} = 1$ . Relaxing this constraint, we define,

$$\mathcal{X}^{\text{R}}(J) \triangleq \left\{ \mathbf{X} \left| \begin{array}{l} \text{Tr} [(J\mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}) \mathbf{X}] + J\xi^2 \leq 0, \\ \text{Tr} [\mathbf{\Omega}_{\text{P},m} \mathbf{X}] \leq P_m^{\text{C}}, m = 1, \dots, M, \\ \mathbf{X} \succeq 0, \end{array} \right. \right\} \quad (5.63)$$

which is now a convex set (superscript R stands for *relaxation*). Because of this relaxation, we have  $\mathcal{X}(J) \subset \mathcal{X}^{\text{R}}(J)$  for all  $J$ . We denote the solution to the new problem,

$$\begin{aligned} & \text{find maximum } J \\ & \text{such that } \mathcal{X}^{\text{R}}(J) \text{ is nonempty,} \end{aligned} \quad (5.64)$$

as  $J_{\text{opt}}^{\text{R}}$ , so that  $J_{\text{opt}} \leq J_{\text{opt}}^{\text{R}}$ . Hence, in general, the solution of the relaxed problem (5.64) only provides an upper bound to the solution of the original problem (5.61). However, the following result establishes the fact that the relaxation is tight, i.e.,  $J_{\text{opt}} = J_{\text{opt}}^{\text{R}}$  for our specific problem.

**Proposition 5.3.5.** (Semidefinite relaxation is tight): *Assume  $\mathbf{\Omega}_{\text{P}}$  to be positive definite and note that  $\mathbf{\Omega}_{\text{JN}}$  is rank-1. Then, for any feasible  $J < J_{\text{opt}}$ ,  $\mathcal{X}^{\text{R}}(J)$  contains a rank-1 matrix.*

*Proof:* See Appendix A.9. *QED.*

The matrix  $\mathbf{\Omega}_{\text{P}}$  is usually positive definite since it is composed of blocks of the matrix  $\mathbf{E}_{\text{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ . Since  $\mathcal{X}^{\text{R}}(J)$  contains a rank-1 matrix,  $\mathcal{X}^{\text{R}}(J) \cap \mathcal{X}(J)$  is non-empty, and feasibility of  $\mathcal{X}^{\text{R}}(J)$  also implies the feasibility of  $\mathcal{X}(J)$ , thereby making the relaxation tight (in terms  $J_{\text{opt}}^{\text{R}}$  being equal to  $J_{\text{opt}}$ ). Note that the solution to (5.64) can also be obtained using a bisection search. However, feasibility test of  $\mathcal{X}^{\text{R}}(J)$  is a convex semi-definite programming problem and hence can be performed efficiently (in polynomial-time). The computational complexity<sup>7</sup> of such a feasibility problem is roughly  $\mathcal{O}(M^2 L^2)$ , where  $M$

---

<sup>7</sup>Generally, numerical techniques for semidefinite programming are iterative in nature, see [85] for a detailed discussion. In the dual formulation of the problem, each iteration solves a linear problem in  $M$  variables and  $\frac{L(L+1)}{2}$  equations, with the resulting complexity being  $\mathcal{O}(M^2 L^2)$ . The number of such

is the number of sensors and  $L$  is the number of non-zero collaboration weights. We have used the publicly available software **SeDuMi** as the optimization tool [77] for our numerical simulations.

### Closed Form solutions

Though numerical solution of the general problem (5.58) can be obtained using the procedure outlined in Section 5.3.3, closed form solutions can be obtained for some special cases. All the special cases discussed in this subsection will make use of the following optimization problem in its core,

$$\begin{aligned}
& \underset{t}{\text{maximize}} \quad F_t = \frac{\left(\sum_{m=1}^M a_m t_m\right)^2}{\sum_{m=1}^M b_m t_m^2 + \xi^2}, \\
& \text{subject to} \quad 0 \leq t_m \leq c_m, \quad m = 1, \dots, M, \\
& \text{given that} \quad a_m \text{ and } b_m \text{ are positive for all } m,
\end{aligned} \tag{5.65}$$

which is known to have the following solution.

**Proposition 5.3.6.** *(Solution of problem (5.65), see [41]): Order the sensors based on the parameter  $d_m \triangleq \frac{a_m}{b_m c_m}$  such that, without loss of generality,*

$$d_1 \geq d_2 \geq \dots \geq d_M. \tag{5.66}$$

*Define*

$$\Phi_k \triangleq \frac{\sum_{m=1}^k b_m c_m^2 + \xi^2}{\sum_{m=1}^k a_m c_m}. \tag{5.67}$$

*Also, define  $\tilde{m}$  algorithmically as follows - keep checking in the decreasing order  $\tilde{m} = \{M, M-1, \dots, 2\}$  whether  $\Phi_{\tilde{m}-1} \frac{a_{\tilde{m}}}{b_{\tilde{m}}} \geq c_{\tilde{m}}$ , and stop at the first instance this condition is satisfied. If  $\Phi_1 \frac{a_2}{b_2} < c_2$ , then  $\tilde{m} = 1$ . Then the solution to problem (5.65) is  $F_{\text{opt}}$  which is*

---

iterations is generally between 5 to 20 for many practical purposes (although theoretically, it is also a polynomial function).

achieved when  $\mathbf{t} = \mathbf{t}_{\text{opt}}$ , where

$$F_{\text{opt}} = \frac{\left(\sum_{m=1}^{\tilde{m}} a_m c_m\right)^2}{\sum_{m=1}^{\tilde{m}} b_m c_m^2 + \xi^2} + \sum_{m=\tilde{m}+1}^M \frac{a_m^2}{b_m}, \quad (5.68)$$

$$t_{\text{opt},m} = \begin{cases} c_m, & m = 1, \dots, \tilde{m} \\ \Phi_{\tilde{m}} \frac{a_m}{b_m} < c_m, & m = \tilde{m} + 1, \dots, M. \end{cases}$$

Note that all the constraints are active (i.e.,  $\tilde{m} = M$  or  $t_{\text{opt},m} = c_m$  for all  $m$ ) if and only if  $\Phi_{M-1} d_M \geq 1$ .

*Notation:* Corresponding to constants  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ , we would denote the optimal solution (5.68) of problem (5.65) as  $F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})$ . When mentioned in conjunction with  $F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})$ , the corresponding value of  $\mathbf{t}$  will be simply denoted as  $\mathbf{t}_{\text{opt}}$ , i.e., without the arguments, to avoid repetition.

Next, we provide some insightful examples. Some of these results will be used to derive the collaboration gain for homogeneous networks in Section 5.3.4.

*Example 4:* For the problem with a) *distributed topology*, b) *perfect information about observation and channel gains*, and c) *uncorrelated measurement noise*, problem (5.58) can be simplified as (note that  $\mathbf{w} = \text{diag}(\mathbf{W})$ ),

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & J_{\mathbf{w}} = \frac{\mathbf{w}^T \mathbf{a} \mathbf{a}^T \mathbf{w}}{\mathbf{w}^T \text{diag}(\mathbf{b}) \mathbf{w} + \xi^2} \\ \text{subject to} \quad & w_m^2 \sigma_{x,m}^2 \leq P_m^C, \quad m = 1, \dots, M, \end{aligned} \quad (5.69)$$

where

$$\begin{aligned} a_m &= g_m h_m, \quad b_m = g_m^2 \sigma_m^2, \quad \sigma_m^2 = [\boldsymbol{\Sigma}]_{m,m}, \\ \sigma_{x,m}^2 &= \sigma_m^2 + \eta^2 h_m^2. \end{aligned} \quad (5.70)$$

Note that since  $a_m$  and  $b_m$  are positive, the value of  $w_m$  at optimality has to be positive, hence the quadratic constraints in (5.69) reduce to the linear constraints as in (5.65). The rest of the problem is solved by defining  $c_m \triangleq \sqrt{\frac{P_m^C}{\sigma_m^2 + \eta^2 h_m^2}}$  and applying Proposition 5.3.6.

We obtain

$$J_{\text{opt}}(\mathbf{P}^C) = \frac{\left(\sum_{m=1}^{\tilde{m}} h_m g_m c_m\right)^2}{\sum_{m=1}^{\tilde{m}} \sigma_m^2 g_m^2 c_m^2 + \xi^2} + \sum_{m=\tilde{m}+1}^M \frac{h_m^2}{\sigma_m^2}. \quad (5.71)$$

Equation (5.71) is mathematically equivalent to the solution in [41], which was obtained in the context of maximum-SNR beamforming. This is because, for perfect OGI and perfect CSI, maximum-SNR also implies minimum-MSE. However, in the presence of observation and channel gain uncertainties, the two problems are different.

The optimal Fisher Information in (5.71) can also be compared to that in the cumulative-constraint case. With individual constraints, some sensors (those with lower “reliability-to-power ratio”  $d_m = \frac{h_m \sqrt{\sigma_m^2 + \eta^2 h_m^2}}{g_m \sigma_m^2 \sqrt{P_m^C}}$ , precisely sensors  $m = \tilde{m} + 1, \dots, M$ ) effectively “convey” the entirety of their individual Fisher Information to the FC while the other sensors ( $m = 1, \dots, \tilde{m}$ ) can convey only a fraction of their combined sum. In contrast, for the cumulative-constraint problem, different fractions of individual Fisher Information reach the FC. However, in the infinite-power limit, both the cases converge to the centralized Fisher Information  $J_{\text{cent}} = \sum_{m=1}^M \frac{h_m^2}{\sigma_m^2}$ .

The previous example assumed perfect CSI and perfect OGI. The following extension of Example 3 (homogeneous network with equicorrelated parameters) illustrates the deterioration of performance with observation and channel gain uncertainties.

*Example 5:* Consider a distributed topology with problem conditions similar to Example 3 (see (5.54)). To find the optimal distortion for this example, we proceed directly from (5.31), using the full-matrix notation

$$\begin{aligned} \underset{\substack{\mathbf{w} \\ \mathbf{W}=\text{diag}(\mathbf{w})}}{\text{maximize}} \quad & J_{\mathbf{w}} = \frac{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}{\text{Tr} [\mathbf{E}_g \mathbf{W} \mathbf{E}_x \mathbf{W}^T] - \eta^2 (\mathbf{g}^T \mathbf{W} \mathbf{h})^2 + \xi^2}, \\ \text{subject to} \quad & [\mathbf{W} \mathbf{E}_x \mathbf{W}^T]_{m,m} \leq P_m^C, \quad m = 1, \dots, M. \end{aligned} \quad (5.72)$$

Noting the following identities,

$$\begin{aligned}
(\mathbf{g}^T \mathbf{W} \mathbf{h})^2 &= \alpha_g \alpha_h g_0^2 h_0^2 (\mathbf{w}^T \mathbf{1})^2, \\
\text{Tr} [\mathbf{E}_g \mathbf{W} \mathbf{E}_x \mathbf{W}^T] &= g_0^2 \sigma_x^2 \{ (1 - \alpha_g \alpha_x) \mathbf{w}^T \mathbf{w} \\
&\quad + \alpha_g \alpha_x (\mathbf{w}^T \mathbf{1})^2 \}, \\
[\mathbf{W} \mathbf{E}_x \mathbf{W}^T]_{m,m} &= \sigma_x^2 w_m^2, \text{ where} \\
\sigma_x^2 &= \sigma^2 + \eta^2 h_0^2, \alpha_x = \frac{\rho \sigma^2 + \alpha_h \eta^2 h_0^2}{\sigma_x^2},
\end{aligned} \tag{5.73}$$

problem (5.72) is equivalent to,

$$\begin{aligned}
\underset{\mathbf{w}}{\text{maximize}} \quad & F_{\mathbf{w}} = \frac{\alpha_g \alpha_h g_0^2 h_0^2 (\mathbf{w}^T \mathbf{1})^2}{g_0^2 \sigma_x^2 (1 - \alpha_g \alpha_x) \mathbf{w}^T \mathbf{w} + \xi^2}, \\
\text{subject to} \quad & \sigma_x^2 w_m^2 \leq P_m^C, \quad m = 1, \dots, M,
\end{aligned} \tag{5.74}$$

in the sense that  $F_{\mathbf{w}}$  is monotonically related to  $J_{\mathbf{w}}$  through  $J_{\mathbf{w}} = \left[ \frac{1}{F_{\mathbf{w}}} + \frac{\rho \sigma^2}{\alpha_h h_0^2} \right]^{-1}$ . Problem (5.74) is similar in form to problem (5.65) and the solution can be obtained by applying Proposition 5.3.6. We have

$$\begin{aligned}
J_{\text{opt}}(\mathbf{P}) &= \left[ \frac{1}{F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})} + \frac{\rho \sigma^2}{\alpha_h h_0^2} \right]^{-1}, \text{ where} \\
a_m &= \sqrt{\alpha_g \alpha_h} g_0 h_0, \quad b_m = \sigma_x^2 g_0^2 (1 - \alpha_g \alpha_x), \quad c_m = \sqrt{\frac{P_m^C}{\sigma_x^2}}.
\end{aligned} \tag{5.75}$$

When all the constraints are active, the solution is

$$J_{\text{opt}}(\mathbf{P}) = \left[ \frac{g_0^2 (1 - \alpha_g \alpha_x) \sum P_m + \xi^2}{\alpha_g \alpha_h g_0^2 h_0^2 \sigma_x^{-2} (\sum \sqrt{P_m})^2} + \frac{\rho \sigma^2}{\alpha_h h_0^2} \right]^{-1}, \tag{5.76}$$

from which it is evident that the Fisher Information decreases with more uncertainty in observation and channel gains (lower values of  $\alpha_h$  and  $\alpha_g$ ). This result will be useful later to derive the collaboration gain for this example.

The following example considers a fully connected network.

*Example 6:* For the problem with a) *fully connected topology* ( $\mathbf{A} = \mathbf{1}\mathbf{1}^T$ ) and b) *uncorrelated channel gain uncertainty* ( $\mathbf{\Sigma}_g$  is diagonal), we can start from (5.72), where the variable of optimization is the entire matrix  $\mathbf{W}$ ,

$$\begin{aligned} \underset{\mathbf{W}}{\text{maximize}} \quad & J_{\mathbf{W}} = \frac{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}{\text{Tr} [\mathbf{E}_g \mathbf{W} \mathbf{E}_x \mathbf{W}^T] - \eta^2 (\mathbf{g}^T \mathbf{W} \mathbf{h})^2 + \xi^2}, \\ \text{subject to} \quad & [\mathbf{W} \mathbf{E}_x \mathbf{W}^T]_{m,m} \leq P_m^C, m = 1, \dots, M. \end{aligned} \quad (5.77)$$

Problem (5.77) can be simplified further based on the assumption of diagonal  $\mathbf{\Sigma}_g$ . However, the analysis is relegated to Appendix A.10 and we state just the result here.

**Proposition 5.3.7.** *The optimal solution for Example 6 is*

$$\begin{aligned} J_{\text{opt}}(\mathbf{P}^C) &= \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})} \right]^{-1}, \text{ where} \\ \mathbf{a} &= \mathbf{g}, \mathbf{b} = \text{diag}(\mathbf{\Sigma}_g), \text{ and } c_m = \sqrt{P_m^C}, \end{aligned} \quad (5.78)$$

and  $\tilde{J} = \mathbf{h}^T \tilde{\mathbf{\Sigma}}^{-1} \mathbf{h}$ ,  $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma} + \eta^2 \mathbf{\Sigma}_h$ , as defined in (5.50). Optimality is achieved when the corresponding weights are

$$\mathbf{W}_{\text{opt}} = \kappa \mathbf{t}_{\text{opt}} \mathbf{v}^T, \mathbf{v} = \tilde{\mathbf{\Sigma}}^{-1} \mathbf{h}, \kappa = \frac{1}{\sqrt{\tilde{J}(1 + \eta^2 \tilde{J})}}. \quad (5.79)$$

*Proof:* See Appendix A.10. *QED.*

From the formula for optimal weights  $\mathbf{W}_{\text{opt}}$  in (5.79), we note that all the sensors have identical fusion rules (precisely, the vector  $\mathbf{v}^T$ ) but they transmit using different transmission power (according to  $\kappa \mathbf{t}_{\text{opt}}$ ). It may be surprising to note that even though all the sensors are transmitting the same information coherently, they may still refrain from using the maximum power available ( $\mathbf{t}_{\text{opt},m} \leq c_m$ , in general). This is because of the uncertainty in the channel (which is captured by the diagonal entries of  $\mathbf{\Sigma}_g$ ). If a channel is too uncertain, allocating a large amount of power to the sensor (even if the power is locally available) may not be helpful for inference. Indeed, all other parameters being constant, a

higher magnitude of  $[\Sigma_g]_{m,m}$  results in a lower value of  $d_m$  (see problem (5.65)) and hence the power constraint is more likely to be inactive at optimality. However, for perfect CSI ( $\Sigma_g = 0$ ), all the sensors must transmit with maximum power available.

We are now in a position to explicitly derive the formula of collaboration gain for homogeneous networks. It must be mentioned here that explicit formulas of CG for arbitrary problems are difficult to derive and we only provide the example of homogeneous networks in this chapter. As mentioned in Section 5.2.7, collaboration gain is a useful metric that indicates the efficacy of spatial collaboration for estimation.

### 5.3.4 Collaboration gain for a homogeneous network

This analysis of homogeneous networks is particularly illustrative since it is actually possible to derive closed-form expressions of CG for a wide range of individual (and also cumulative) power constraints. From the closed form equation of CG, we can actually predict the conditions for which collaboration will be particularly effective. Specifically, we will derive the formula for collaboration gain and analyze how it depends on the problem parameters like size of the network, noise correlation, observation/channel gain uncertainties and power constraints.

#### Theoretical analysis for a homogeneous network

Let  $P$  be the total power used in the network. For the cumulative-constraint problem,  $P$  is a sufficient descriptor of the power constraints. However, for the individual-constraint problem, we have  $M$  different power constraints  $\{P_1, P_2, \dots, P_M\}$ , which makes it difficult to analyze the problem theoretically. Fortunately, for a large subset of those problems, only 2 summary-descriptors of  $\{P_1, P_2, \dots, P_M\}$  suffice to characterize the collaboration gain – namely, the cumulative power  $P$ , and a *skewness* parameter  $\kappa$ , defined as,

$$\kappa \triangleq \frac{P_{\text{sq}}}{P}, \quad P_{\text{sq}} \triangleq \left( \sum_{m=1}^M \sqrt{P_m} \right)^2, \quad P = \sum_{m=1}^M P_m. \quad (5.80)$$

As an example, we refer to Equation (5.76), where the distortion is seen to depend on only  $\sum P_m$  (which is  $P$ ) and  $(\sum \sqrt{P_m})^2$  (which is  $\kappa P$ ). It is easy to see (applying Cauchy-Schwartz inequality) that  $1 \leq \kappa \leq M$ , where the limiting values have distinct significance. The value  $\kappa = 1$  implies that only one of the sensors has all the power ( $P_1 = P$ , say) while other sensors have no power at all  $P_2 = \dots = P_M = 0$ . On the other hand, the value  $\kappa = M$  implies that all sensors have equal power allocated to them  $P_1 = \dots = P_M = \frac{P}{M}$ . Define  $\kappa_M$  (a normalized<sup>8</sup> version of  $\kappa$ ) as

$$\kappa_M = \frac{\kappa - 1}{M - 1} \in [0, 1]. \quad (5.81)$$

A stronger rationale behind naming the parameter  $\kappa$  as *skewness* is provided by the following result, which shows that  $\kappa$  is monotonically related to another quantity that explicitly enables power allocation in a skewed manner.

**Lemma 5.3.8** (Parameter  $\kappa$  is an indicator of skewness). *Assume that power available at the  $M$  nodes are*

$$P_1 = P_0, P_2 = \varrho P_0, \dots, P_M = \varrho^{M-1} P_0, \quad \varrho \in [0, 1], \quad (5.82)$$

*so that  $\varrho$  is an explicit measure of dissimilarity (or skewness) among the power available at various nodes. Then (recall definition of  $\kappa$  in (5.80)),*

$$\kappa(\varrho) = \frac{\left(1 + \sqrt{\varrho} + \dots + (\sqrt{\varrho})^{M-1}\right)^2}{1 + \varrho + \dots + \varrho^{M-1}} \quad (5.83)$$

*is a strictly monotonically increasing function of  $\varrho$ . Note that  $\kappa(\varrho = 0) = 1$  and  $\kappa(\varrho = 1) = M$ .*

*Proof:* See Appendix A.11. *QED.*

---

<sup>8</sup>The parameter  $\kappa_M$  is similar in structure to the Chiu-Jain fairness metric [11] used for congestion control in computer networks.



The following proposition provides an explicit formula for the collaboration gain in a homogeneous network with equicorrelated parameters for both the individual-constraint and cumulative-constraint problems. We would use Examples 3, 5 and 6 to establish this result, the derivation of which is relegated to Appendix A.12.

**Proposition 5.3.9.** (Collaboration gain for a homogeneous network): *For the individual-constraint problem, let the node indices be arranged (without loss of generality) in such a way so that  $P_1^C \leq \dots \leq P_M^C$ . Let  $P$  and  $\kappa$  denote the summary-descriptors of  $\{P_1^C, \dots, P_M^C\}$  as per Equation (5.80). Assume,*

$$P_M^C \leq \left( \frac{\sum_{m=1}^{M-1} P_m^C + \frac{\xi^2}{g_0^2(1-\alpha_g\alpha_x)}}{\sum_{m=1}^{M-1} \sqrt{P_m^C}} \right)^2. \quad (5.84)$$

*Then, all the power constraints are active at optimality,  $P_m = P_m^C, \forall m$ , for both distributed and connected topologies, and the collaboration gain is*

$$\text{CG} = \frac{\left(1 - \frac{1}{M}\right) \left(\frac{\kappa}{M} \frac{1+(M-1)\alpha_g}{1+(\kappa-1)\alpha_g}\right) \left(1 - \frac{\kappa-1}{M-1} \frac{\alpha_g}{1+P_g^{-1}}\right) (1 - \alpha_x)}{\left(1 + \frac{1}{P_g(1+(\kappa-1)\alpha_g)}\right) \left(1 + \frac{(\kappa-1)\alpha_g}{1+P_g^{-1}}\alpha_x\right)}, \quad (5.85)$$

where  $P_g = \frac{Pg_0^2}{\xi^2}, \alpha_x = \frac{\rho + \gamma\alpha_h}{1 + \gamma}, \gamma = \frac{\eta^2 h_0^2}{\sigma^2}.$

*For the cumulative-constraint problem, the collaboration gain is given by Equation (5.85) by setting  $\kappa = M$ .*

*Proof:* See Appendix A.12. *QED.*

Several remarks about Proposition (5.3.9) are in order. Condition (5.84) is basically an assumption that the power constraints are *not too skewed*. In Equation (5.85) note that all the quantities in parenthesis in the numerator denote quantities less than 1, while those in the denominator are greater than one, so this reaffirms the notion that collaboration gain is always less than 1. Furthermore,

a) *Dependence on local-SNR  $\gamma$ , noise correlation  $\rho$ , and observation gain uncertainty  $\alpha_h$ :* Collaboration gain increases with a decrease in  $\alpha_x$ , which means that CG increases as a)

Noise correlation decreases (smaller  $\rho$ ), b) observation gain uncertainty increases (smaller  $\alpha_h$ ), and c) local SNR decreases (smaller  $\gamma$ ), provided  $\rho < \alpha_h$  (which is typically true for a problem involving moderately correlated noise and sufficiently certain observation gain). Hence, collaboration is more effective when the local-SNR is small, measurement noise is uncorrelated and observation gains are uncertain.

b) *Dependence on (normalized) total power  $P_g$ , power skewness  $\kappa$  and channel gain uncertainty  $\alpha_g$* : To understand the effect of  $P_g, \kappa$  and  $M$  on collaboration gain, we simplify Equation (5.85) by considering the large- $M$  asymptotic regime. Note that in general, an infinite number of sensors implies that the Fisher Information is infinite and distortion is zero for both distributed and connected topologies, which is a trivial regime to consider. Towards the goal of analyzing regimes that incur only finite distortion in the asymptotic limit, we consider two cases, as listed below.

First, we would consider the *fixed- $\kappa$ -large- $P$  regime*, which signifies that the effective number of powered nodes are not increasing with  $M$  (recall that  $\kappa = 1$  implies that only one node has all the power, regardless of  $M$ ), i.e., most of the nodes are auxilliary nodes that provide their information to the powered nodes which then communicate with the fusion center. Hence, even with a large transmission power the distortion at the FC is finite, which makes this regime non-trivial. From (5.85), we note that

$$\lim_{\substack{M \rightarrow \infty \\ P_g \rightarrow \infty}} \text{CG} = \frac{(1 - \alpha_x)\alpha_g\kappa}{(1 + \alpha_g(\kappa - 1))(1 + \alpha_x\alpha_g(\kappa - 1))}. \quad (5.86)$$

With the additional technical assumption that  $\alpha_g > \frac{1}{1 + \alpha_x}$  (which basically means that the channel gains are sufficiently certain), the collaboration gain (Equation (5.86)) decreases as  $\kappa$  increases, and hence is maximum when  $\kappa = 1$ , at which point

$$\text{CG} = (1 - \alpha_x)\alpha_g. \quad (5.87)$$

That CG decreases with  $\alpha_x$  has already been discussed in the previous remark, and also applies for this regime. We conclude that in the fixed- $\kappa$ -large- $P$  regime, collaboration is

highly effective if there are only a few powered sensors (small  $\kappa$ ), and the channel gains (higher  $\alpha_g$ ) are fairly certain.

Next, we consider the *fixed-( $\kappa/M$ )-finite-( $P \times M$ ) regime*. Here, the normalized skewness parameter  $\kappa_M$  is kept constant, which means that effectively, the number of powered sensors  $\kappa$  increase linearly with  $M$  (note the contrast with previous regime). To keep the effective Fisher Information (and resulting distortion) finite, the total available power  $P$  must scale inversely proportional<sup>9</sup> to  $M$ . Let  $P_M \triangleq P_g M$  denote the normalized total power, which is a finite constant. From (5.85), we can derive that

$$\max_{P_M} \lim_{\substack{M \rightarrow \infty \\ P_g M = P_M}} \text{CG} = \frac{1 - \sqrt{\alpha_x}}{1 + \sqrt{\alpha_x}}, \text{ when } P_M^* = \frac{1}{\kappa_M \alpha_g \sqrt{\alpha_x}}. \quad (5.88)$$

This implies that collaboration is more effective for smaller  $\alpha_x$ , a fact that was established earlier as well. We conclude that in the fixed-( $\kappa/M$ )-finite-( $P \times M$ ) regime, collaboration is most effective when the normalized operating power is a particular finite quantity, precisely  $P_M^* = \frac{1}{\kappa_M \alpha_g \sqrt{\alpha_x}}$ . Moreover,  $P_M^*$  increases as the power constraints get more skewed (smaller  $\kappa_M$ ) or the channel gains get more uncertain (smaller  $\alpha_g$ ).

The assertions in Proposition 5.3.9 and the subsequent discussion are illustrated in Figure 5.5 for an example with the following problem parameters  $\eta^2 = \frac{1}{2}$ ,  $\sigma^2 = 1$ ,  $\rho = 0.1$ ,  $g_0 = h_0 = 1$  and  $\alpha_g = \alpha_h = 0.9$ . Contours of the actual collaboration gain are computed numerically using procedure outlined in Section 5.3.3 and displayed for a wide range of cumulative power  $P_g$ , number of nodes  $M$  and skewness of power constraints. To illustrate the fixed- $\kappa$ -large- $P$  regime, we have used  $\kappa = 1$  (only one sensor has all the power) in Figure 5.5(a). To illustrate the fixed-( $\kappa/M$ )-finite-( $P \times M$ ) regime, we have used  $\kappa_M = \{1/3, 2/3, 1\}$  in Figures 5.5(b), 5.5(c), and 5.5(d) respectively. To simulate a particular skewness  $\kappa$ , the local power constraints are generated in accordance with (5.82) by finding the corresponding value  $\varrho$  through bisection search (recall that  $\varrho$  and  $\kappa$  are monotonically

---

<sup>9</sup>Another asymptotic domain that is conceptually different but otherwise would yield identical results is to keep the power  $P$  constant and let the channel gains  $\mathbf{g}$  scale at the rate of  $\frac{1}{\sqrt{M}}$ . The argument here is that, though the network size increases with  $M$ , the channel capacity of the effective multiple-input-single output (MISO) channel induced by the  $M$  sensors, precisely  $\frac{1}{2} \log \left( 1 + \frac{P \sum_{m=1}^M g_m^2}{\xi^2} \right)$ , is held constant.

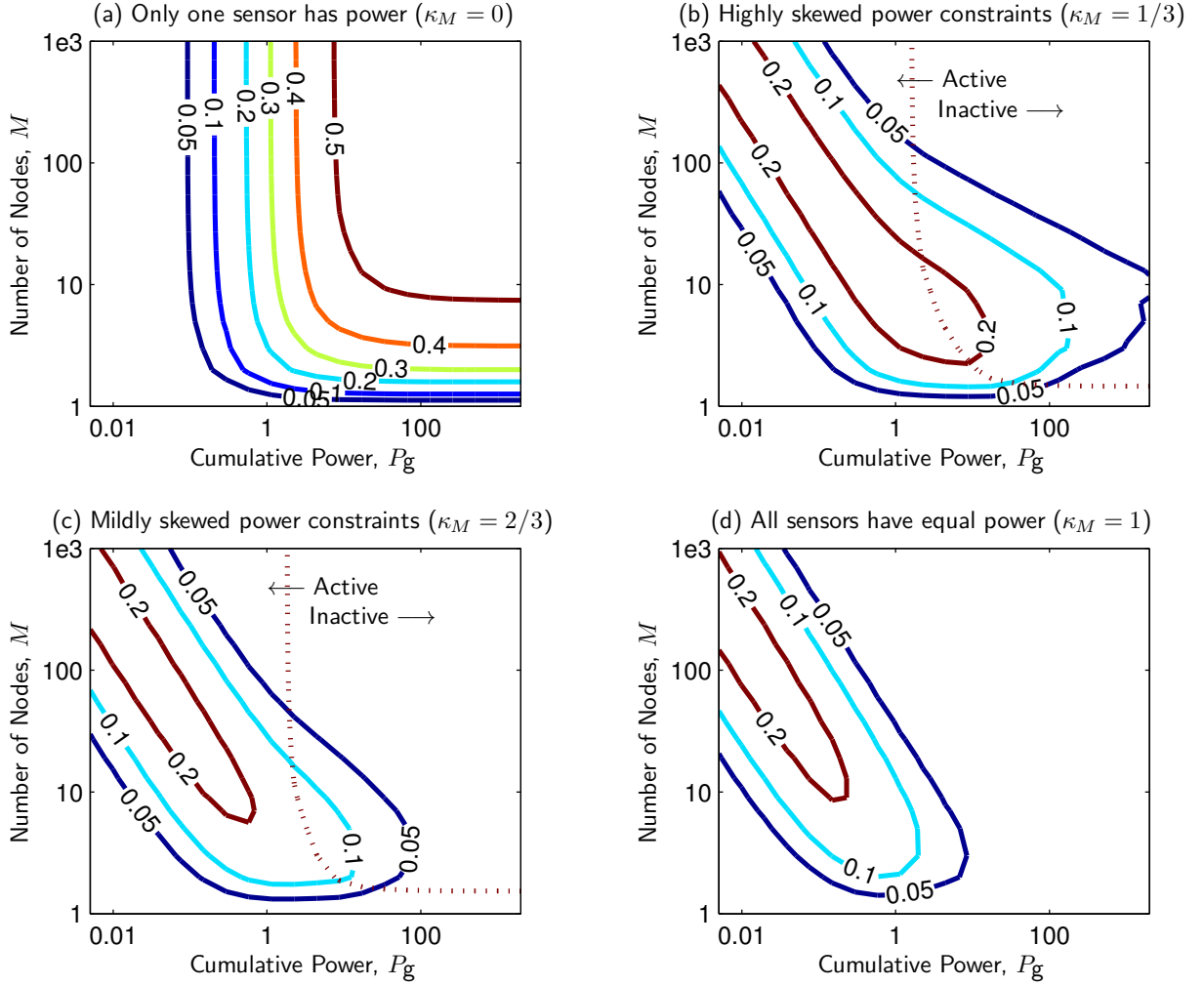


Figure 5.5: Collaboration gain in a homogeneous sensor network with skewed power constraints.

related, as per Lemma 5.3.8). The active-constraint condition of Equation (5.84) is depicted through the dotted line in Figures 5.5(b) and 5.5(c). The portion of the figure to the right of the dotted line suggests that one or more of the constraints are inactive, while the left half denotes the region where all the constraints are active and consequently Equation (5.85) is the accurate measure of collaboration gain. No dotted lines appear in Figure 5.5(a) and 5.5(d) because all the constraints are trivially active in both the cases, although for different reasons. For  $\kappa = 1$ , there is only one sensor with all the power and hence it must transmit with full power, which explains Figure 5.5(a). For  $\kappa = M$ , we have equal power allocation ( $P_m^C = P^C$ , say) and condition (5.84) reduces to  $P^C \leq \left( \sqrt{P^C} + \frac{\xi^2}{g_0^2(1-\alpha_g\alpha_x)(M-1)\sqrt{P^C}} \right)^2$ , which is trivially satisfied for all  $P^C$ .

For the problem parameters mentioned above, we calculate that  $\alpha_x \approx 0.37$ . For the fixed- $\kappa$ -large- $P$  regime, theoretical justifications predict that the maximum gain possible (across various problem conditions) is  $(1 - \alpha_x)\alpha_g \approx 0.57$ , which can be confirmed from the contours toward the top-right corner of Figure 5.5(a). For the fixed- $(\kappa/M)$ -finite- $(P \times M)$  regime, theoretical predictions yield that the maximum gain is  $\frac{1-\sqrt{\alpha_x}}{1+\sqrt{\alpha_x}} \approx 0.24$ , which can be validated from the innermost contours of Figures 5.5(b), 5.5(c) and 5.5(d). Note that the contours shifts to the left as  $\kappa_M$  increases. This is due to the fact that the normalized power required to achieve maximum collaboration gain decreases as  $\kappa_M$  become larger, a fact also discussed above. In conclusion, for this particular problem instance, upto 57% (and 24%) of the distortion performance can be recovered using collaboration in the fixed- $\kappa$ -large- $P$  regime (and fixed- $(\kappa/M)$ -finite- $(P \times M)$ ) regimes. We have established this fact using both numerical results and theoretical insights.

### Random geometric graphs

To demonstrate how the distortion decreases with increasing collaboration, we consider the following simulation setup. The spatial placement and neighborhood structure of the sensor network is modeled as a Random Geometric Graph,  $\text{RGG}(N, r)$  [24], where sensors are uniformly distributed over a unit square and bidirectional communication links are possible only for pairwise distances at most  $r$ , i.e., the adjacency matrix is  $\mathbf{A}$  such that  $A_{i,j} = \mathbb{1}_{[d_{i,j} \leq r]}$ . Correspondingly, the cost matrix is a  $\{0, \infty\}$  matrix with the  $(i, j)^{\text{th}}$  element being zero only if  $d_{i,j} \leq r$ , otherwise being infinity. We assume  $N = 20$  sensor nodes and gradually increase the radius of collaboration from  $r = 0$  (signifying distributed topology) to  $r = \sqrt{2}$  (signifying connected topology, since the sensors are placed in a unit square). The simulated sensor network is depicted in Figure 5.6, with collaboration radius  $r = 0.2$ . In general, for  $0 < r < \sqrt{2}$ , the network is only partially connected as in Figure 5.6.

We simulate a homogeneous network with the following parameters  $\sigma^2 = 1$ ,  $\rho = 0$  (independent noise),  $g_0 = h_0 = 1$  and  $\alpha_g = \alpha_h = 0.9$ . To contrast the effect of prior uncertainty on collaboration gain, we simulate two different variance of the prior,  $\eta^2 = 0.1$

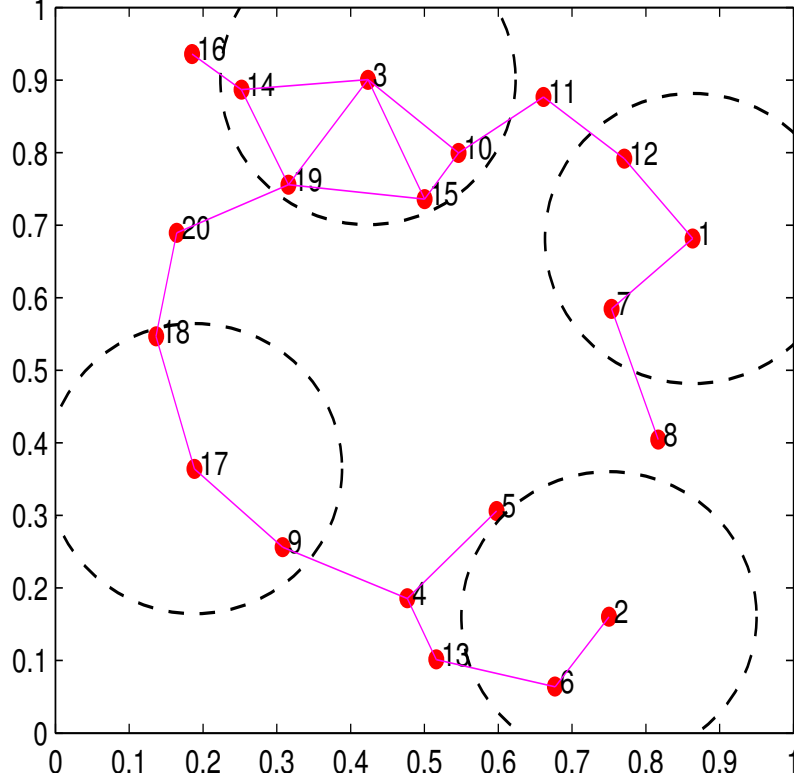


Figure 5.6: Random Geometric Graph with 20 nodes, used for example in Section 5.3.4. Edges are shown for pairwise distance less than 0.2. The radius of collaboration is depicted for sensors 1,2,3 and 17.

and 0.5. To illustrate the effect of power constraints, we simulate a wide range of both cumulative-power and skewness. We simulate three skewness conditions –  $\kappa_M = 0.5, 0.75$  and 1, the value of 1 implying equal power allocation. The three values of cumulative-power that were simulated were a)  $P_g = P^* = \frac{1}{0.75 \times 20} \frac{1}{\alpha_g \sqrt{\alpha_x}}$  (recall from the discussion in Section 5.3.4 that, for  $\kappa = 0.75$ , this is a high-CG operating region), b)  $P_g = P^*/4$  (depicting the low power regime), and c)  $P_g = 4P^*$  (depicting the high power regime).

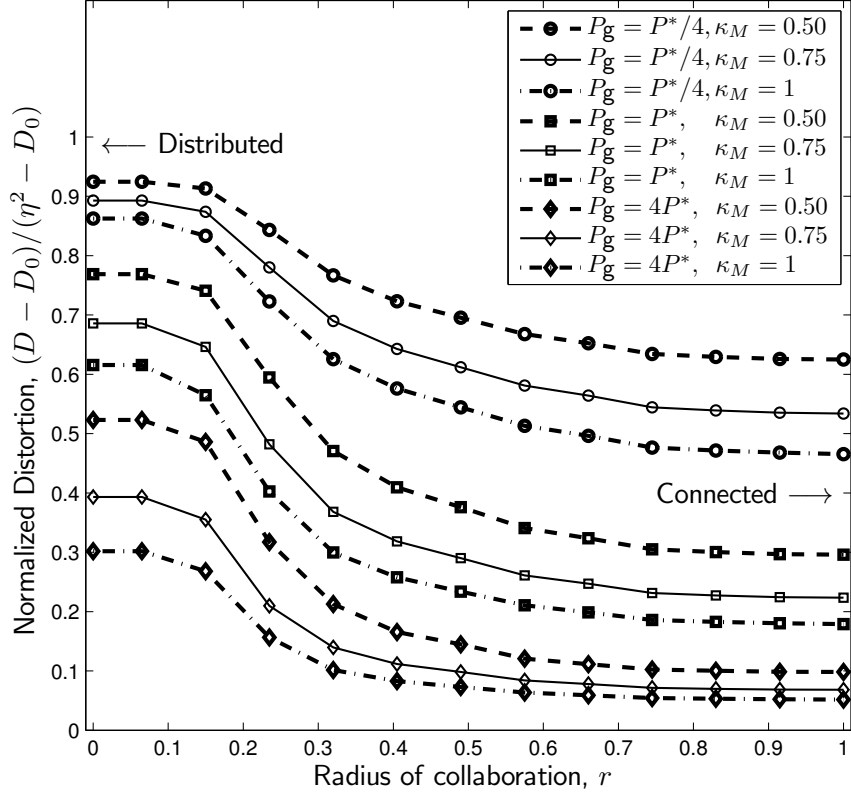
The simulation results are depicted in Figures 5.7-(a) and (b) for the two values of prior uncertainty  $\eta^2 = 0.1$  and  $\eta^2 = 0.5$  respectively. Corresponding to these values of  $\eta^2$ , the infinite-power distortion  $D_0$ , the maximum possible collaboration gain  $\text{CG}^*$  and the

corresponding operating power  $P^*$  for  $\kappa_M = 0.75$ , are calculated by using (5.55) and (5.88),

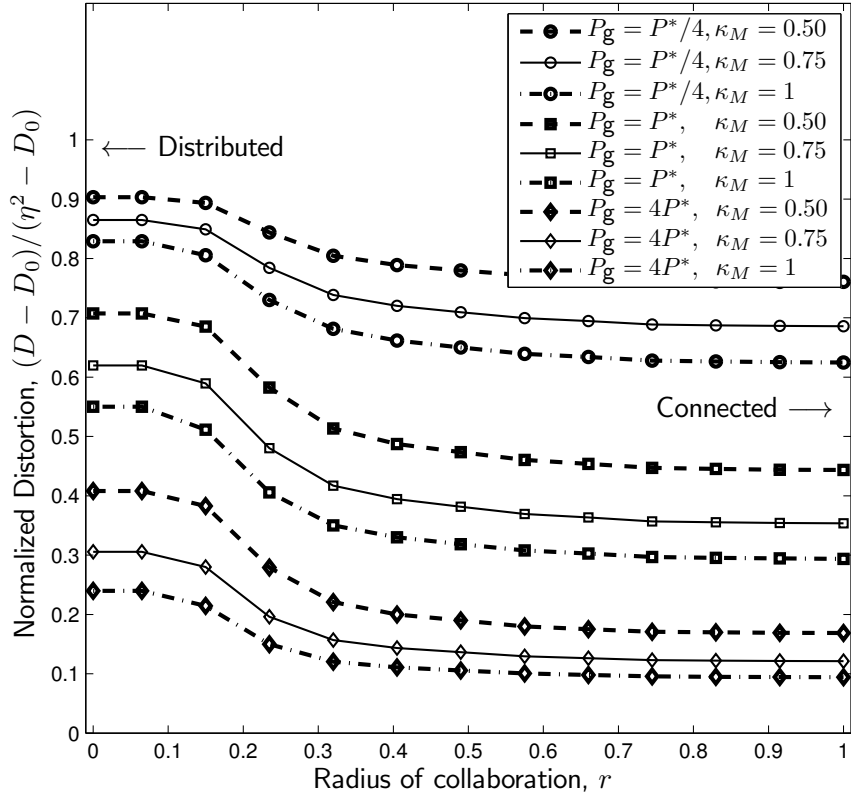
$\eta^2 \downarrow$	$D_0$	CG*	$P^*(\kappa_M = 0.75)$	
0.1	0.07	0.29	0.13	(5.89)
0.5	0.04	0.56	0.25.	

With varying power availability and varying extent of collaboration, the resulting distortion varies between  $D \in (D_0, \eta^2)$ , where the ranges are  $(0.07, 0.1)$  and  $(0.04, 0.5)$ , for the prior variance  $\eta^2 = 0.1$  and  $\eta^2 = 0.5$  respectively. To compare the effect of collaboration across the two different problem conditions, we depict the normalized distortion  $\frac{D-D_0}{\eta^2-D_0}$  in Figures 5.7-(a) and (b). Note that in this normalized scale, the collaboration gain is simply the difference between the right-most (distributed) and the left-most (connected) ends of a curve. The efficacy of collaboration, as indicated by the downward slope of the curves, is clearly demonstrated in Figures 5.7-(a) and (b), where it is noted that a large part of the overall gain is achieved only with partial collaboration, i.e., the distortion tends to saturate beyond a collaboration radius of  $r \approx 0.4$ . Hence, though the collaboration gain, as defined in (5.30), requires a fully connected topology, a large part of that gain can be realized with only a partially connected network. Also, we note that the efficacy of collaboration diminishes when the operating power is too small ( $P_g = P^*/4$ ) or too large ( $P_g = 4P^*$ ), indicating the fact that collaboration in a network should be used judiciously, especially when there are costs associated with it. Finally, we note that the efficacy of collaboration is higher when the prior has a lower variance (the curves in Figure 5.7-(a) has more downward slope than those in Figure 5.7-(b)). This observation is explained directly by the comments following Proposition 5.3.9, where we argued that collaboration gain increases with decreasing local-SNR.

The summary of all the main results in this Chapter are tabulated in Tables 5.1 and 5.2 for ease of reference.



(a)  $\eta^2 = 0.1$



(b)  $\eta^2 = 0.5$

Figure 5.7: Improvement in distortion due to increasing collaboration among 20 nodes in a random geometric graph.



Table 5.1: Main results: cumulative power constraint (Examples 1, 2 and 3)

Cases	Conditions	Optimal (equivalent) Fisher Information	Optimal weights are proportional to
(A) General		$J = \mathbf{h}^T \mathbf{G}^T \left( \Omega_{\text{JD}} + \frac{\Omega_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{G} \mathbf{h}$	$\mathbf{w} \propto \left( \Omega_{\text{JD}} + \frac{\Omega_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{G} \mathbf{h}$
(B) Perfect CSI	$\Sigma_{\text{g}} = 0$	$J = \mathbf{h}^T \left( \tilde{\Sigma} + \frac{\Gamma_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}$	$\mathbf{w} \propto \Omega_{\text{P}}^{-1} \mathbf{G} \Gamma_{\text{P}} \left( \tilde{\Sigma} + \frac{\Gamma_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}$
(C) Perfect OGI, Perfect CSI, [45]	$\Sigma_{\text{h}} = 0,$ $\Sigma_{\text{g}} = 0$	$J = \mathbf{h}^T \left( \Sigma + \frac{\Gamma_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}$	$\mathbf{w} \propto \Omega_{\text{P}}^{-1} \mathbf{G} \Gamma_{\text{P}} \left( \Sigma + \frac{\Gamma_{\text{P}}}{P_{\xi}} \right)^{-1} \mathbf{h}$
(D) Distributed, Uncorrelated, Perfect OGI, Perfect CSI, [96]	$\mathbf{A} = \mathbf{I},$ $\Sigma$ is diagonal $\Sigma_{\text{h}} = 0,$ $\Sigma_{\text{g}} = 0$	$J = \sum_{n=1}^N \frac{h_n^2}{\sigma_n^2} \left[ 1 + \frac{1 + \gamma_n}{P_{\xi} g_n^2} \right]^{-1},$ $\sigma_n^2 \triangleq \Sigma_{n,n}, \gamma_n \triangleq \frac{\eta^2 h_n^2}{\sigma_n^2}$	$\mathbf{W} \propto \text{diag}([v_1, v_2, \dots, v_N]),$ $v_n = \frac{h_n}{g_n \sigma_n^2} \left[ 1 + \frac{1 + \gamma_n}{P_{\xi} g_n^2} \right]^{-1}$
(E) Connected	$\mathbf{A} = \mathbf{1}\mathbf{1}^T$	$J = \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{\mathbf{g}^T \left( \Sigma_{\text{g}} + \frac{\mathbf{I}}{P_{\xi}} \right)^{-1} \mathbf{g}} \right]^{-1},$ $\tilde{J} \triangleq \mathbf{h}^T \tilde{\Sigma}^{-1} \mathbf{h}$	$\mathbf{W} \propto \mathbf{u} \mathbf{v}^T, \mathbf{u} = \left( \Sigma_{\text{g}} + \frac{\mathbf{I}}{P_{\xi}} \right)^{-1} \mathbf{g},$ $\mathbf{v} = \tilde{\Sigma}^{-1} \mathbf{h}$
(F) Cycle topology, Homogeneous, Equicorrelated- ( $\Sigma, \mathbf{E}_{\text{h}}, \mathbf{E}_{\text{g}}$ )	$\mathbf{A} = \mathcal{C}(K),$ $\mathbf{h} = h_0 \sqrt{\alpha_{\text{h}}} \mathbf{1},$ $\Sigma_{\text{h}} = h_0^2 (1 - \alpha_{\text{h}}) \mathbf{I},$ $\Sigma = \sigma^2 \mathbf{R}(\rho),$ $\mathbf{g} = g_0 \sqrt{\alpha_{\text{g}}} \mathbf{1},$ $\Sigma_{\text{g}} = g_0^2 (1 - \alpha_{\text{g}}) \mathbf{I}$	$J = \frac{h_0^2}{\sigma^2} \left[ \rho_N + \tilde{\alpha}_{\text{h}} \left( \rho_N + \frac{\gamma}{N} \right) \right. \\ \left. + \frac{1}{N} \left( \tilde{\alpha}_{\text{g}} + \frac{1}{P_{\xi} g_0^2 \alpha_{\text{g}}} \right) \left\{ \gamma + \rho_K + \tilde{\alpha}_{\text{h}} \left( \rho_K + \frac{\gamma}{K} \right) \right\} \right]^{-1},$ $\rho_t \triangleq \rho + \frac{1 - \rho}{t}, \gamma \triangleq \frac{\eta^2 h_0^2}{\sigma^2}, \tilde{\alpha} \triangleq \frac{1}{\alpha} - 1.$ Note: $K = 1 \Rightarrow$ Distributed, $K = N \Rightarrow$ Connected $\alpha_{\text{g}} = 1 \Rightarrow$ Perfect CSI, $\alpha_{\text{h}} = 1 \Rightarrow$ Perfect OGI	$\mathbf{w} \propto \mathbf{1}_L.$
Other definitions: $\Gamma_{\text{P}} \triangleq (\mathbf{G}^T \Omega_{\text{P}}^{-1} \mathbf{G})^{-1}, \tilde{\Sigma} \triangleq \Sigma + \eta^2 \Sigma_{\text{h}}, \mathbf{R}(\rho) \triangleq ((1 - \rho) \mathbf{I} + \rho \mathbf{1}\mathbf{1}^T).$			

Table 5.2: Main results: individual power constraint (Examples 4, 5 and 6)

Cases	Conditions	Optimal (equivalent) Fisher Information	Optimal weights
(A) Distributed, Uncorrelated, Perfect OGI, Perfect CSI, [41]	$\mathbf{A} = \mathbf{I},$ $\Sigma$ is diagonal $\Sigma_{\text{h}} = 0,$ $\Sigma_{\text{g}} = 0$	$J = F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c}),$ $a_m = g_m h_m,$ $b_m = g_m^2 \sigma_m^2, \sigma_m^2 = [\Sigma]_{m,m},$ $c_m = \sqrt{\frac{P_m^{\text{C}}}{\sigma_m^2 + \eta^2 h_m^2}}$	$\mathbf{W} = \text{diag}(\mathbf{t}_{\text{opt}})$
(B) Distributed, Homogeneous, Equicorrelated- ( $\Sigma, \mathbf{E}_{\text{h}}, \mathbf{E}_{\text{g}}$ )	$\mathbf{A} = \mathbf{I},$ $\mathbf{h} = h_0 \sqrt{\alpha_{\text{h}}} \mathbf{1},$ $\Sigma_{\text{h}} = h_0^2 (1 - \alpha_{\text{h}}) \mathbf{I},$ $\Sigma = \sigma^2 \mathbf{R}(\rho),$ $\mathbf{g} = g_0 \sqrt{\alpha_{\text{g}}} \mathbf{1},$ $\Sigma_{\text{g}} = g_0^2 (1 - \alpha_{\text{g}}) \mathbf{I}$	$J = \left[ \frac{1}{F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})} + \frac{\rho \sigma^2}{\alpha_{\text{h}} h_0^2} \right]^{-1},$ $\mathbf{a} = \sqrt{\alpha_{\text{g}} \alpha_{\text{h}}} g_0 h_0 \mathbf{1},$ $\mathbf{b} = \sigma_{\text{x}}^2 g_0^2 (1 - \alpha_{\text{g}} \alpha_{\text{x}}) \mathbf{1},$ $c_m = \sqrt{\frac{P_m^{\text{C}}}{\sigma_{\text{x}}^2}}$	$\mathbf{W} = \text{diag}(\mathbf{t}_{\text{opt}})$
(C) Connected, Uncorrelated- channel-gain	$\mathbf{A} = \mathbf{1}\mathbf{1}^T,$ $\Sigma_{\text{g}}$ is diagonal	$J = \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{F_{\text{opt}}(\mathbf{a}, \mathbf{b}, \mathbf{c})} \right]^{-1},$ $\tilde{J} \triangleq \mathbf{h}^T \tilde{\Sigma}^{-1} \mathbf{h},$ $\mathbf{a} = \mathbf{g}, \mathbf{b} = \text{diag}(\Sigma_{\text{g}}), c_m = \sqrt{P_m^{\text{C}}}$	$\mathbf{W} = \kappa \mathbf{t}_{\text{opt}} \mathbf{v}^T,$ $\mathbf{v} = \tilde{\Sigma}^{-1} \mathbf{h},$ $\kappa = \frac{1}{\sqrt{\tilde{J}(1 + \eta^2 \tilde{J})}}$
Other definitions: $\tilde{\Sigma} \triangleq \Sigma + \eta^2 \Sigma_{\text{h}}, \mathbf{R}(\rho) \triangleq ((1 - \rho) \mathbf{I} + \rho \mathbf{1}\mathbf{1}^T), \sigma_{\text{x}}^2 = \sigma^2 + \eta^2 h_0^2, \alpha_{\text{x}} = \frac{\rho \sigma^2 + \alpha_{\text{h}} \eta^2 h_0^2}{\sigma_{\text{x}}^2}$			

## 5.4 Collaboration with finite costs

In Section 5.3, we have solved the optimal collaboration problem for the situation when the cost of communication for each link is either zero or infinity, i.e.,  $C_{i,j} \in \{0, \infty\}$  (also termed as the ideal case). In this section, we address the general problem where communication may incur a non-zero but finite cost, i.e.,  $0 < C_{i,j} < \infty$  (also termed as the finite-cost case). Unlike the ideal case, finding the globally optimal solution for the finite-cost case is a difficult problem and there are no known numerical techniques that efficiently solve this problem. In this section, we outline an efficient (polynomial-time) numerical procedure that obtains a locally optimal solution to the finite-cost problem. We first describe our iterative solution for the individual-constraint problem. The cumulative-constraint problem follows from similar arguments and is described next. Lastly, solutions to both the problems are demonstrated using numerical simulations.

### 5.4.1 Individual power constraint

We propose an iterative solution as follows. Let the collaboration topology and transmission power availability (for all  $N$  nodes, in vector form) at iteration  $i$  be denoted by  $\mathbf{A}_i$  and  $\mathbf{P}_i^{\text{trans}}$  respectively. Note that the transmission power availability is the difference between the original power constraint and the collaboration cost due to the topology  $\mathbf{A}_i$ . For the  $n^{\text{th}}$  node, it means

$$[\mathbf{P}_i^{\text{trans}}]_n = [\mathbf{P}^C]_n - \sum_{m=1}^M [\mathbf{A}_i]_{mn} C_{mn}. \quad (5.90)$$

Note that for the auxiliary nodes  $n = M + 1, \dots, N$ , the transmission power availability  $[\mathbf{P}_i^{\text{trans}}]_n$  does not mean much, since they cannot transmit to the FC anyway. For those sensors, it is best if they use their entire resources for collaboration. Recall that the optimal distortion corresponding to any topology  $\mathbf{A}$  and transmission power constraint  $\mathbf{P}^{\text{trans}}$  can be obtained from the discussion in Section 5.3. Denote such a distortion by  $D_{\text{opt}}^{\mathbf{A}}(\mathbf{P}^{\text{trans}})$ . We start with a distributed topology, i.e.,  $\mathbf{A}_1 = [\mathbf{I}_M | \mathbf{0}]$  and follow a greedy

algorithm. At iteration  $i$ , we evaluate the distortion performance corresponding to all incremental topologies of the form  $\mathbf{A}_i + \mathbf{E}(m, n)$ , where  $\mathbf{E}(m, n)$  is an all-zero matrix except for the  $(m, n)^{\text{th}}$  element, which is 1 (signifying an incremental  $n \rightarrow m$  link). There are  $MN - \text{nnz}(\mathbf{A}_i)$  such possibilities for selecting  $\mathbf{E}(m, n)$ , each corresponding to a link that is current not being used (equivalently  $[\mathbf{A}_i]_{mn}$  is zero). The number of such possibilities may even be lesser if there is not enough power to make a link possible. For example, if  $C_{mn} > [\mathbf{P}_i^{\text{trans}}]_n$ , then the  $n^{\text{th}}$  node does not have sufficient power to use the  $n \rightarrow m$  link for collaboration. Among all such possible links, let  $n^* \rightarrow m^*$  denote the link that provides the best distortion performance. Then the iteration is concluded by augmenting the topology with the edge  $n^* \rightarrow m^*$ . Thus, in compact notations, each iteration is represented as,

$$(m^*, n^*) = \arg \min_{\substack{m, n \\ [A_i]_{mn}=0 \\ C_{mn} \leq [\mathbf{P}_i^{\text{trans}}]_n}} D_{\text{opt}}^{\mathbf{A}_i + \mathbf{E}(m, n)} (\mathbf{P}_i^{\text{trans}} - \mathbf{e}_n C_{mn}), \quad (5.91)$$

$$\mathbf{A}_{i+1} = \mathbf{A}_i + \mathbf{E}_{m^* n^*},$$

where  $\mathbf{e}_n$  is an all-zero vector with the exception of  $n^{\text{th}}$  element, which is 1. The iterations are terminated when one of the following conditions is satisfied, a) there is no feasible link, b) maximum number of iterations has exceeded a pre-specified limit or c) there is not enough increment in (relative) performance after a particular iteration, say, for a pre-specified  $\delta$ , if

$$\frac{D_{\text{opt}}^{\mathbf{A}_{i+1}} (\mathbf{P}_{i+1}^{\text{trans}}) - D_{\text{opt}}^{\mathbf{A}_i} (\mathbf{P}_i^{\text{trans}})}{\eta^2 - D_0} \leq \delta. \quad (5.92)$$

A rough estimate of computational complexity can be established as follows. Recall from Section 5.3.3 that computing  $D_{\text{opt}}^{\mathbf{A}}(\mathbf{P})$  is roughly  $\mathcal{O}(M^2 L^2)$ , where  $L = \text{nnz}(\mathbf{A})$ . Assume that the algorithm gets terminated after adding  $\mathcal{O}(N)$  links, which is to say that each node communicates with a fraction of its neighbors at optimality. Since we start with the distributed topology we have  $L = \mathcal{O}(N)$  for all the iterations. Since each of the iterations (as in (5.91)) require approximately  $MN$  function evaluations, the total complexity of the

finite-cost collaborative power allocation problem is

$$\underbrace{\mathcal{O}(M^2 N^2)}_{\text{evaluation of } D_{\text{opt}}} \times \underbrace{\mathcal{O}(MN)}_{\text{evaluations per iteration}} \times \underbrace{\mathcal{O}(N)}_{\text{number of iterations}} = \mathcal{O}(M^3 N^4). \quad (5.93)$$

It must be emphasized here that (5.93) is only a rough (but practical) measure of complexity and is not a rigorous bound. It assumes a fixed number of iterations to solve a semidefinite optimization problem (see discussion in Section 5.3.3) and also sufficiently high collaboration costs so that only  $\mathcal{O}(N)$  links are added starting from a distributed topology.

### 5.4.2 Cumulative power constraint

To solve the cumulative power-constraint problem, we proceed on similar lines. Let the cumulative power constraint be  $P^C$  and the cumulative transmission power availability at iteration  $i$  be denoted by  $P_i^{\text{trans}}$ , so that

$$P_i^{\text{trans}} = P^C - \sum_{m=1}^M \sum_{n=1}^N [\mathbf{A}_i]_{mn} \mathbf{C}_{mn}. \quad (5.94)$$

Denote the optimal distortion corresponding to any topology  $\mathbf{A}$  and transmission power constraint  $P^{\text{trans}}$  (see Section 5.3.2) as  $D_{\text{opt}}^{\mathbf{A}}(P^{\text{trans}})$ . Starting from a distributed topology, the best collaboration link  $n^* \rightarrow m^*$  is selected at each iteration according to

$$(m^*, n^*) = \arg \min_{\substack{m, n \\ [A_i]_{mn}=0 \\ C_{mn} \leq P_i^{\text{trans}}}} D_{\text{opt}}^{\mathbf{A}_i + \mathbf{E}(m, n)}(P_i^{\text{trans}} - C_{mn}), \quad (5.95)$$

$$\mathbf{A}_{i+1} = \mathbf{A}_i + \mathbf{E}_{m^* n^*},$$

with the stopping criteria being similar to that described in the previous subsection. As regards to computational complexity, recall from (5.3.2) that the complexity of computing  $D_{\text{opt}}^{\mathbf{A}}(\mathbf{P})$  is  $\mathcal{O}(L^3)$ , where  $L = \text{nnz}(\mathbf{A})$ . Assuming  $\mathcal{O}(N)$  iterations (as in the previous case) the overall complexity for the finite-cost-collaborative cumulative-constraint problem

is roughly

$$\underbrace{\mathcal{O}(N^3)}_{\text{evaluation of } D_{\text{opt}}} \times \underbrace{\mathcal{O}(MN)}_{\text{evaluations per iteration}} \times \underbrace{\mathcal{O}(N)}_{\text{number of iterations}} = \mathcal{O}(MN^5). \quad (5.96)$$

### 5.4.3 Numerical simulations

To demonstrate the efficacy of collaboration in finite-cost scenarios, we consider a random geometric graph of  $M = N = 10$  nodes. As in the previous examples, we consider a homogeneous network with the following parameters  $\eta^2 = 0.5$ ,  $\sigma^2 = 1$ ,  $\rho = 0$  (independent noise),  $g_0 = h_0 = 1$  and  $\alpha_g = \alpha_h = 0.9$ . The collaboration cost of link  $m \rightarrow n$  is assumed to increase quadratically with the distance between nodes  $m$  and  $n$ . This is because the gain of a wireless channel is often inversely proportional (upto a constant exponent) to the distance between a source and a receiver [31]. Consequently, to maintain a reliable communication link, the transmission power has to be scaled up accordingly. In particular, we assume

$$C_{m,n} = c_0 d_{m,n}^2, \quad (5.97)$$

where  $d_{m,n}$  denotes the distance between nodes  $m$  and  $n$  and  $c_0$  is a constant of proportionality. For our numerical simulations, we consider a wide range of  $c_0$ , specifically  $c_0 \in [10^{-4}, 10^4]$ , to depict the effect of collaboration cost on the distortion performance. A lower collaboration cost in effect allows the network to collaborate more and thereby reduces the distortion. We consider two magnitudes of total operating power (namely,  $P_g = 1$  and 3) for both the individual-constraint and cumulative-constraint cases. For the individual-constraint case, we consider two skewness conditions for the power-constraint, namely  $\kappa_M = 0.5$  and  $\kappa_M = 0.75$ . The corresponding distortion curves are shown in Figure 5.8. For very low values of the  $c_0$ , the distortion converges to that in a fully connected network. Similarly, for very high values of the  $c_0$ , no links are selected for collaboration, and the network operates in a distributed manner. Since in our example, the network is ho-

homogeneous, equal power allocation among nodes is also the optimum power allocation for a cumulative-power-constrained problem. Consequently, the performance of the cumulative-constraint problem (dash-dotted lines) is always better than that with individual power constraints (bold and dashed lines). In conclusion, Figure 5.8 shows that in a homogeneous network, the estimation performance improves with higher operating power, less skewed power constraints and lower collaboration cost among sensors, as expected.

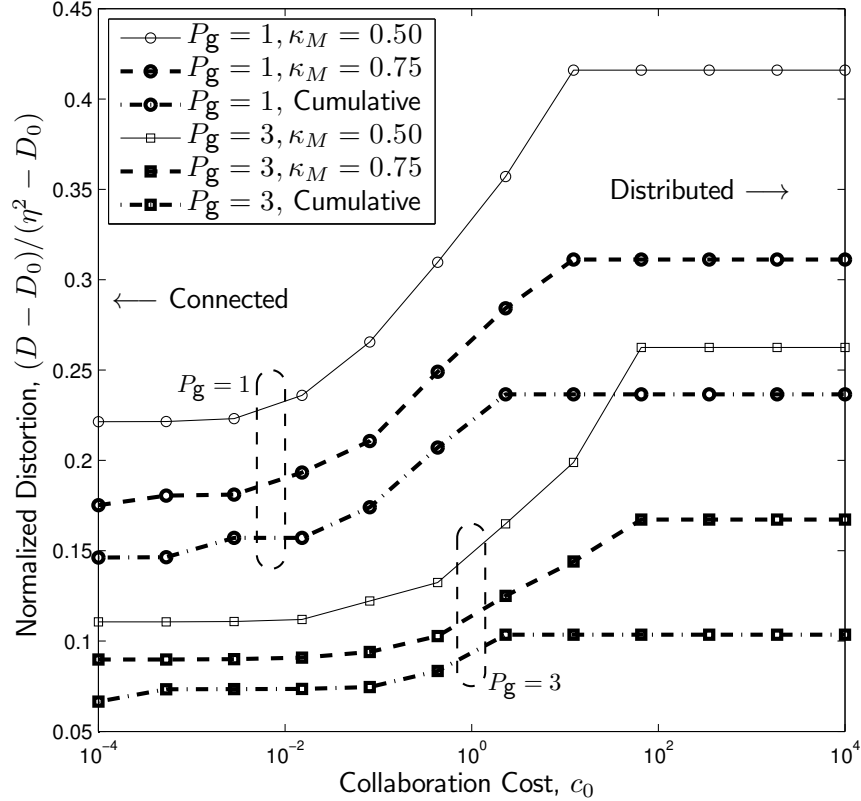


Figure 5.8: Estimation with finite collaboration cost - an example with 10 sensor nodes.

## 5.5 Summary

In this chapter, we addressed the problem of collaborative estimation in a sensor network where sensors communicate with the FC using a coherent MAC channel. For the scenario when the collaborative topology is fixed and collaboration is cost-free, we obtained the optimal (cumulative) power-distortion tradeoff in closed-form by solving a QCQP problem. With individual power constraints, we have shown that the semidefinite relaxation tech-

nique can be used to obtain precisely optimal numerical results. Several special cases are presented as examples to highlight the problem conditions for which collaboration is particularly effective. Through the use of both theoretical and numerical results, we established that collaboration helps to substantially reduce the distortion of the estimated parameter at the fusion center, especially in low local-SNR scenario. As future work, we wish to explore the collaborative estimation problem when the parameter to be estimated is a vector with correlated elements. The scenario when collaboration is erroneous, as mentioned earlier, is also important.



# Chapter 6

## Analog forwarding based collaborative estimation of dynamic parameters

### 6.1 Introduction

In Chapter 5, we presented an extension of the amplify-and-forward framework that allowed spatial collaboration in a partially connected network topology. It was observed that even a sparsely connected network was able to realize a performance which was very close to that of a fully connected network. This is due to the fact that in an amplify-and-forward framework, the observation noise is also amplified along with the signal, thereby significantly increasing the energy required for transmission. Spatial collaboration, in effect, smooths out the observation noise, thereby improving the quality of the signal that is transmitted to the FC using the same energy resources.

In this chapter, we explore the potential of collaborative estimation further by making two significant contributions. First, though it was observed earlier that even a moderately connected network performs almost as well as a fully connected network, no analytical results were presented. In this chapter, we extend our previous work by analyzing the es-

estimation performance for partially connected collaboration networks. Though the analysis of arbitrary network topologies is a difficult problem, we derive the estimation performance for a family of structured network topologies, namely the  $Q$ -cliques. We demonstrate that the insights obtained from the structured topology apply approximately to two practical topologies, namely the nearest neighbor and random geometric graphs, of similar connectivity. Given a particular network topology, we investigate two different energy allocation schemes for data transmission. In addition to the optimal energy-allocation (EA) scheme as derived in the previous chapter, we also consider the suboptimal but easy-to-implement equal energy-allocation scheme, where neighboring observations are simply averaged to mitigate the observation noise. For both of these schemes, we derive the performance in a closed form in the asymptotic domain where the number of nodes is large and the overall transmission capacity of the network is held constant. These results offer insights into the relationship between estimation performance and problem parameters like channel and observation gains, prior uncertainty and extent of spatial collaboration.

The collaborative estimation problem has so far been analyzed in the single-snapshot context, where energy-constrained spatial sampling is performed at one particular instant and the inference is made using those samples. In our second contribution in this chapter, we extend the problem formulation to consider power-constrained inference of a random process, where the goal is to estimate the process for *all* time instants. In contrast to the simple snapshot framework, this involves obtaining multiple samples in time and computing the filtered estimates for any desired time instants, including time instants where observation samples are not available. Since collection of each sample involves the expenditure of energy resources, the appropriate constraint in this situation is the energy spent per unit time (or power). A key parameter here is the sampling frequency, the choice of which affects the overall estimation performance. Note that a higher sampling frequency usually means that one can better capture the temporal variations. However, with a power constraint, less energy is available for the collection of each of those samples, which would result in more noisy samples. This trade-off is investigated in the context of a Gaussian

random process with exponential covariance, where it turns out that a higher sampling frequency *always* results in better estimates.

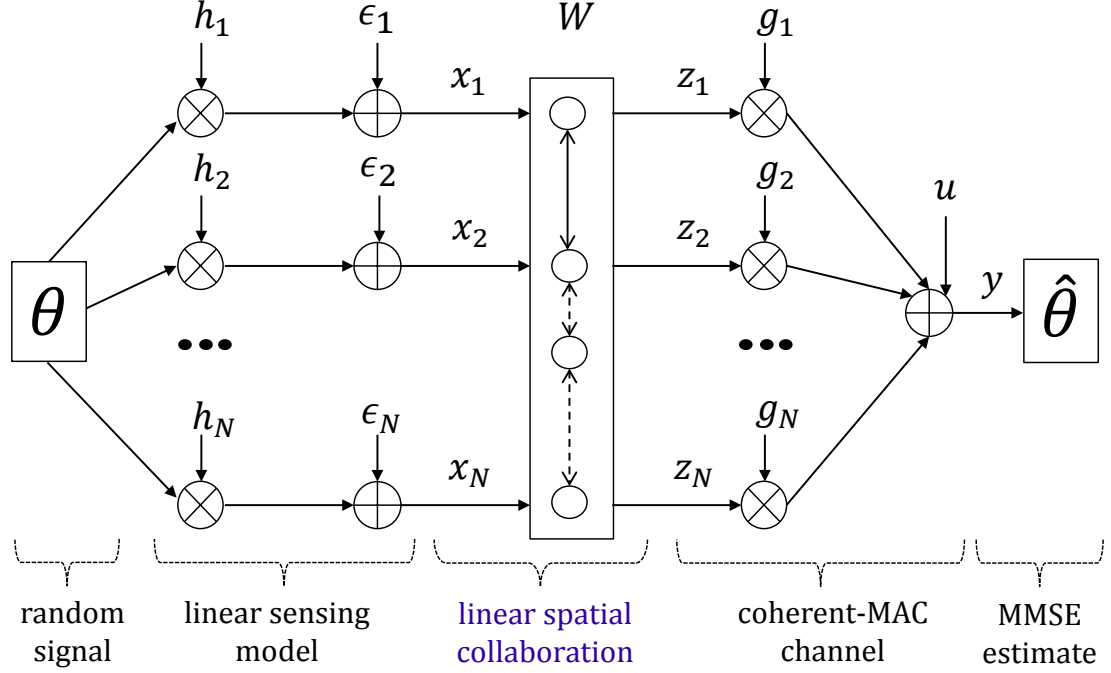


Figure 6.1: Wireless sensor network performing collaborative estimation.

## 6.2 Problem formulation

We first consider the single snapshot estimation problem, where the dimension of time is ignored. We will extend the discussion to time varying Gaussian process later in Section 6.2.3. The single snapshot framework is depicted in Figure 6.1. The parameter to be estimated,  $\theta$  (written without any time subscript to signify the single snapshot nature of the problem), is assumed to be a zero-mean Gaussian source with prior variance  $\eta^2$ . Different noisy versions of  $\theta$  are observed by  $N$  sensors. The observation vector is  $\mathbf{x} = [x_1, \dots, x_N]$  where  $x_n = h_n\theta + \epsilon_n$ , with  $h_n$  and  $\epsilon_n$  denoting the observation gain and measurement noise respectively. The measurement noise variables  $\{\epsilon_n\}_{n=1}^N$  are assumed to be independent and identically distributed (iid) Gaussian random variables with zero mean and variance  $\sigma^2$ .

Let the availability of collaboration links be represented by the adjacency matrix  $\mathbf{A}$ , where  $A_{nm} = 1$  (or  $A_{nm} = 0$ ) implies that node  $n$  has (or does not have) access to the

observation of node  $m$ . Define an  $\mathbf{A}$ -sparse matrix as one for which non-zero elements may appear only at locations  $(n, m)$  for which  $A_{nm} = 1$ . The set of all  $\mathbf{A}$ -sparse matrices is denoted by  $\mathcal{S}_A$ . Corresponding to an adjacency matrix  $\mathbf{A}$  and an  $\mathbf{A}$ -sparse matrix  $\mathbf{W}$ , we define *collaboration* in the network as individual nodes being able to linearly combine local observations from other collaborating nodes

$$z_n = \sum_{m \in \mathcal{A}(n)} \mathbf{W}_{nm} x_m, \quad (6.1)$$

where  $\mathcal{A}(n) \triangleq \{m : \mathbf{A}_{nm} = 1\}$ , without any further loss of information. In effect, the network is able to compute a one-shot spatial transformation of the form  $\mathbf{z} = \mathbf{W}\mathbf{x}$ . In practice, this transformation is realizable when any two neighboring sensors are close enough to ensure reliable information exchange. Note that, when  $\mathbf{W}$  is restricted to be diagonal (in other words, when  $\mathbf{A} = \mathbf{I}$ ), the problem reduces to the amplify-and-forward framework for distributed estimation, which is widely used in the literature [14],[96],[20] due to its simplicity in implementation and provably optimal information theoretic properties for simple networks [30].

The transformed observations  $\{z_n\}_{n=1}^N$  are transmitted to the FC through a coherent MAC channel, so that the received signal is  $y = \mathbf{g}^T \mathbf{z} + u$ , where  $\mathbf{g}$  and  $u$  describe the channel gains and the channel noise respectively. The channel noise  $u$  is assumed to be Gaussian distributed with zero mean and variance  $\xi^2$ . The FC receives the noise-corrupted signal  $y$  and computes an estimate of  $\theta$ . Since  $y$  is a linear Gaussian random variable conditioned on  $\theta$ ,

$$\begin{aligned} \theta &\sim \mathcal{N}(0, \eta^2), \text{ and} \\ y|\theta &\sim \mathcal{N} \left( \underbrace{\mathbf{g}^T \mathbf{W} \mathbf{h}}_{\triangleq \mu \text{ (net gain)}}, \theta, \underbrace{\mathbf{g}^T \mathbf{W} \Sigma \mathbf{W}^T \mathbf{g} + \xi^2}_{\triangleq \zeta^2 \text{ (net noise variance)}} \right), \end{aligned} \quad (6.2)$$

the minimum-mean-square-error (MMSE) estimator  $\hat{\theta} = \mathbb{E}[\theta|y]$  is the optimal fusion rule. From estimation theory (for details the reader is referred to [49]), the MMSE estimator

and resulting distortion  $D_{\mathbf{W}}$  is given by

$$\hat{\theta} = \frac{1}{1 + \frac{\zeta^2}{\eta^2 \mu^2}} \frac{y}{\mu}, \text{ and } \frac{1}{D_{\mathbf{W}}} = \frac{1}{\eta^2} + J_{\mathbf{W}}, \quad J_{\mathbf{W}} = \frac{\mu^2}{\zeta^2}, \quad (6.3)$$

where the quantity  $J_{\mathbf{W}}$  is the Fisher Information and  $\mu$  and  $\zeta^2$  are the net gain and net noise variance as defined in Equation (6.2). The cumulative transmission energy required to transmit the transformed observations  $\mathbf{z}$  is

$$\begin{aligned} \mathcal{E}_{\mathbf{W}} &= \mathbb{E}[\mathbf{z}^T \mathbf{z}] = \text{Tr} [\mathbf{W} \mathbf{E}_{\mathbf{x}} \mathbf{W}^T], \text{ where} \\ \mathbf{E}_{\mathbf{x}} &\triangleq \mathbb{E}[\mathbf{x} \mathbf{x}^T] = \eta^2 \mathbf{h} \mathbf{h}^T + \mathbf{\Sigma}. \end{aligned} \quad (6.4)$$

### 6.2.1 Collaboration strategies

Note that the quantities  $\mu$ ,  $\zeta^2$  and, therefore, the distortion  $D$  (equivalently  $J$ ) and also the transmission energy  $\mathcal{E}$  depend on the choice of the collaboration matrix  $\mathbf{W}$ . As indicated earlier, we explore two strategies to determine  $\mathbf{W}$ , namely 1) optimal and 2) equal energy-allocation (EA) strategies, that stem from two different engineering considerations.

In the optimal EA strategy, we assume that the FC knows the channel and observation gains and also the collaboration topology precisely. In such a situation, the FC can compute the optimal collaboration matrix subject to a cumulative transmission energy constraint

$$(\text{Optimal EA}) \quad \mathbf{W}_{\text{opt}} = \arg \min_{\mathbf{W} \in \mathcal{S}_A} D_{\mathbf{W}}, \quad \text{s.t. } \mathcal{E}_{\mathbf{W}} \leq \mathcal{E}, \quad (6.5)$$

and communicate the corresponding weights  $\mathbf{W}_{\text{opt}}$  to the sensor nodes via a separate and reliable control channel. The exact form of  $\mathbf{W}_{\text{opt}}$  and corresponding  $J_{\text{opt}}$  were derived in the previous chapter (see Equation 5.48) and briefly summarized below.

**Theorem 6.2.1** (Optimal single-snapshot estimation). *Let  $L$  be the cardinality of  $\mathbf{A}$ , which is also the number of non-zero collaboration weights. In an equivalent representation, construct  $\mathbf{w} \in \mathbb{R}^L$  by concatenating those elements of  $\mathbf{W}$  that are allowed to be non-zero.*

Accordingly, define the  $L \times L$  matrix  $\mathbf{\Omega}$  and  $L \times N$  matrix  $\mathbf{G}$  such that the identities

$$\text{Tr} [\mathbf{W} \mathbf{E}_x \mathbf{W}^T] = \mathbf{w}^T \mathbf{\Omega} \mathbf{w}, \text{ and } \mathbf{g}^T \mathbf{W} = \mathbf{w}^T \mathbf{G}, \quad (6.6)$$

are satisfied. Then the optimal Fisher Information is,

$$J_{\text{opt}} = \mathbf{h}^T (\mathbf{\Sigma} + \mathbf{\Gamma}/\mathcal{E}_\xi)^{-1} \mathbf{h}, \text{ where} \quad (6.7)$$

$$\mathcal{E}_\xi \triangleq \mathcal{E}/\xi^2, \text{ and } \mathbf{\Gamma} \triangleq (\mathbf{G}^T \mathbf{\Omega}^{-1} \mathbf{G})^{-1},$$

which is achieved when the collaboration weights are  $\mathbf{w}_{\text{opt}} = \kappa \mathbf{\Omega}^{-1} \mathbf{G} \mathbf{\Gamma} (\mathbf{\Sigma} + \mathbf{\Gamma}/\mathcal{E}_\xi)^{-1} \mathbf{h}$ , with the scalar  $\kappa$  chosen to satisfy  $\mathbf{w}_{\text{opt}}^T \mathbf{\Omega} \mathbf{w}_{\text{opt}} = \mathcal{E}$ .  $\mathbf{W}_{\text{opt}}$  is the matrix equivalent of  $\mathbf{w}_{\text{opt}}$ .

When either the FC is computationally limited or reliable control channels are not available, the optimal EA strategy cannot be implemented. In these situations, one reasonable way of assigning transmission energy and collaboration weights at each node may be the equal EA strategy, where all the sensors are allocated equal transmission energy (namely  $\frac{\mathcal{E}}{N}$ ). In addition, the  $n$ th sensor equally weighs all the observations from its neighbors where the weights (say  $\{d_n\}_{n=1}^N$ ) are chosen to satisfy  $\mathbb{E}[z_n^2] = \frac{\mathcal{E}}{N}$ . Note from (6.1) that  $z_n = d_n \sum_{m \in \mathcal{A}(n)} (h_m \theta + \epsilon_m)$ . Consequently,

$$\text{(Equal EA)} \quad [\mathbf{W}_{\text{eq}}]_{nm} = \begin{cases} d_n, & \text{if } m \in \mathcal{A}(n) \\ 0, & \text{else} \end{cases}, \quad (6.8)$$

$$d_n = \sqrt{\frac{\mathcal{E}/N}{\left(\sum_{m \in \mathcal{A}(n)} h_m\right)^2 \eta^2 + |\mathcal{A}(n)| \sigma^2}},$$

where  $|\mathcal{A}(n)|$  denotes the number of neighbors of  $n$ . The Fisher Information corresponding to the equal EA strategy is simply  $J_{\text{eq}} \triangleq J_{\mathbf{W}_{\text{eq}}}$ , which can be obtained by applying Equation (6.3).

Once the collaboration strategy (namely, either optimal or equal EA) is chosen and a cumulative operating energy  $\mathcal{E}$  is specified, the resulting distortion performance (FI-s  $J_{\text{opt}}$  or  $J_{\text{eq}}$ ) depends on the following problem parameters, 1) signal prior, measurement

noise and channel noise, which were assumed to be Gaussian distributed with zero mean and variances  $\eta^2$ ,  $\sigma^2 \mathbf{I}_N$  and  $\xi^2$  respectively, 2) observation and channel gains, and 3) the collaboration topology. To obtain analytical expressions for FI-s, it is clear that we need to make further simplifying assumptions on the observation/channel gains (which will be discussed in Section 6.3.1) and also the topology for collaboration.

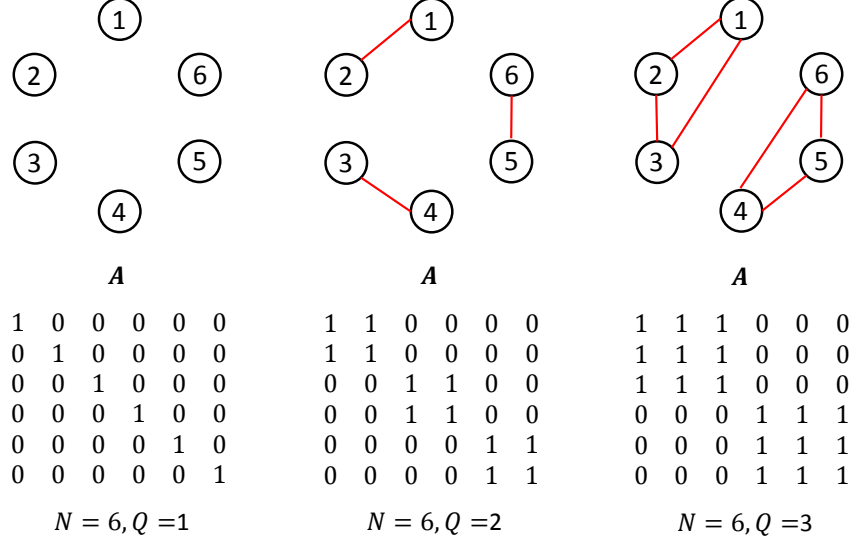


Figure 6.2: Example of  $Q$ -cliques.

### 6.2.2 Partially connected networks

With the goal to investigate partially connected collaboration topologies, we adopt the following methodology. Intuitively, we expect the distortion to decrease as the network becomes more connected (since it adds more degrees of freedom) and it is our aim to obtain asymptotic limits that explicitly reflect the effect of *connectedness*. Since the analysis of arbitrary topologies is difficult, we derive our analytical results for a structured collaboration topology that consists of several fully-connected clusters (or cliques) of finite size  $Q$ , as illustrated in Figure 6.2. To be precise, if  $N = KQ$ , we have  $\mathbf{A} = \mathbf{I}_K \otimes (\mathbf{1}_Q \mathbf{1}_Q^T)$ . Since all of the nodes in a  $Q$ -clique network are  $(Q - 1)$ -connected, the performance of this special topology may serve as an approximation to other topologies where the average number of neighbors per node is  $(Q - 1)$ . To demonstrate the efficacy of this approximation, we will compare the analytical results for  $Q$ -clique networks with numerical results

for two practical collaboration topologies, namely 1) nearest-neighbor (NN) topology and 2) random geometric graphs (RGG) [24]. For the NN-topology, a sensor receives the observations from its nearest  $Q - 1$  neighbors. For the RGG topology, a sensor collaborates with all other sensors that are located within a circle of radius  $r$  with the sensor at the center. The expected number of neighbors, which is a function of  $r$ , can be derived using geometric arguments, thereby enabling comparison with an equivalent  $Q$ -clique topology. Examples of these two topologies are illustrated in Figures 6.3(a) and 6.3(b) for a network with  $N = 20$  nodes. It may be noted that unlike the NN topology, all collaboration links of an RGG topology are bidirectional by definition.

### 6.2.3 Ornstein-Uhlenbeck process

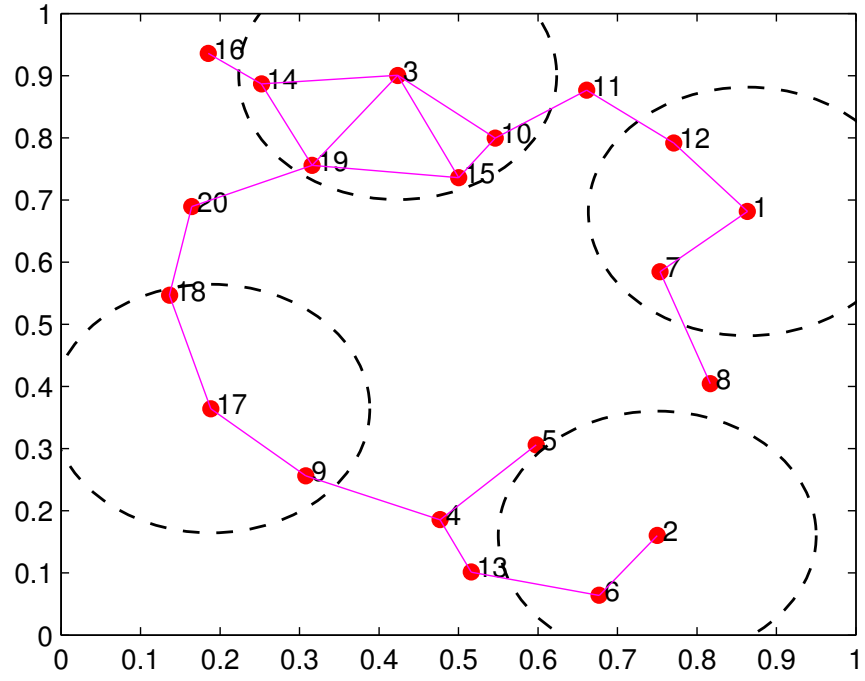
We now bring in the dimension of time, and consider the problem of power-constrained estimation of time-varying signals. In order to model the temporal dynamics, we assume that the signal of interest is a stationary zero-mean Gaussian random process  $\theta_t$  with exponential covariance function

$$\mathbb{E}[\theta_{t_1}, \theta_{t_2}] = \eta^2 e^{-(|t_1 - t_2|)/\tau}, \quad (6.9)$$

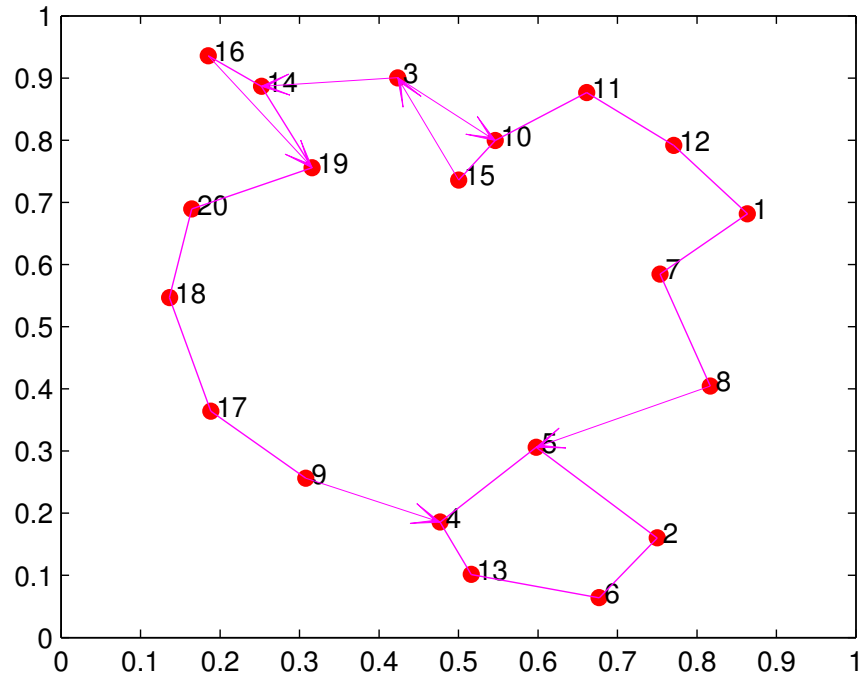
where  $\eta^2$  and  $\tau$  represent the magnitude and temporal variation of the parameter respectively. Note that  $\tau \rightarrow 0$  implies that the signal changes very rapidly, while  $\tau \rightarrow \infty$  means that the signal is constant over time. Such a process (with covariance parameterized by  $\eta^2$  and  $\tau$ ) is widely used in the literature [65] due to its ability to model a time varying Gaussian process while providing a relatively simple framework for analysis. This process is also sometimes known as the Ornstein-Uhlenbeck (OU) process.

Let  $P$  denote the power constraint in the network. We assume that the OU process is sampled periodically with the period  $T$  (see Figure 6.4), which implies that a total of  $\mathcal{E} = PT$  energy units is available for each sampling instant. Let the spatial sampling at each instant be performed in a manner similar to the single-snapshot framework discussed earlier,





(a) Random geometric graph with radius  $r = 0.2$  (total 44 links)



(b) Nearest neighbor topology with  $Q = 3$  (total 40 links)

Figure 6.3: Example of two practical collaboration topologies for a  $N = 20$ -node network. Bidirectional links are shown without arrows.

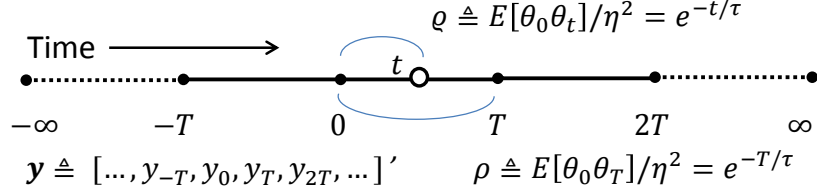


Figure 6.4: Periodically sampled Ornstein-Uhlenbeck process.

namely through collaboration and coherent amplify-and-forward beamforming. Let the signal received by the FC at instant  $t$  be denoted as  $y_t$ , so that the entire observed sequence can be represented by the following infinite-dimensional vector  $\mathbf{y} \triangleq [\dots, y_{-T}, y_0, y_T, y_{2T}, \dots]'$ . From the linear Gaussian model  $y|\theta$  in (6.2) and subsequent description of Fisher Information  $J$  in (6.3), the spatial sampling process can be abstracted (via an appropriate scaling) through the additive model  $y_t = \theta_t + v_t$ , where the *aggregate* noise  $v_t \sim \mathcal{N}(0, \frac{1}{J})$  summarizes the uncertainty due to  $\{\epsilon_{n,t}\}_{n=1}^N$  (measurement noise at sensors) and  $u_t$  (channel noise). We assume  $\epsilon_{n,t}$  and  $u_t$  to be temporally white, from which it follows that  $v_t$  is temporally white as well. In vector notations,

$$\mathbf{y} = \boldsymbol{\theta} + \mathbf{v}, \quad \mathbf{v} \sim \mathcal{N}\left(0, \frac{1}{J}\mathbf{I}\right), \quad (6.10)$$

where  $\boldsymbol{\theta} \triangleq [\dots, \theta_{-T}, \theta_0, \theta_T, \theta_{2T}, \dots]'$  and  $\mathbf{v} \triangleq [\dots, v_{-T}, v_0, v_T, v_{2T}, \dots]'$ . Having observed  $\mathbf{y}$ , the MMSE estimator of  $\theta_t$  (the value of OU process at any instant  $t$ ) is given by the conditional expectation (refer to [49] for details)

$$\hat{\theta}_t = \mathbb{E}[\theta_t|\mathbf{y}] = \mathbf{R}_{\theta_t\mathbf{y}'}\mathbf{R}_{\mathbf{y}\mathbf{y}'}^{-1}\mathbf{y}, \quad (6.11)$$

where  $\mathbf{R}_{\theta_t\mathbf{y}'} \triangleq \mathbb{E}[\theta_t\mathbf{y}']$  and  $\mathbf{R}_{\mathbf{y}\mathbf{y}'} \triangleq \mathbb{E}[\mathbf{y}\mathbf{y}']$ . Moreover, the variance of  $\hat{\theta}_t$  is given by

$$\text{Var}(\theta_t|\mathbf{y}) = \eta^2 - \mathbf{R}_{\theta_t\mathbf{y}'}\mathbf{R}_{\mathbf{y}\mathbf{y}'}^{-1}\mathbf{R}_{\theta_t\mathbf{y}}. \quad (6.12)$$

Since  $y_t$  is sampled periodically at instants  $t = kT$ ,  $k \in \mathbb{Z}$  and  $\mathbf{y}$  contains infinite elements in both time directions, the conditional variance  $\text{Var}(\theta_t|\mathbf{y})$  is also expected to be periodic in

time, i.e.,  $\text{Var}(\theta_t|\mathbf{y}) = \text{Var}(\theta_{t+kT}|\mathbf{y})$ ,  $\forall t, k$ . Hence any interval of length  $T$ , say  $t \in [0, T]$  is sufficient for analyzing the conditional variance (6.12). Since we are interested in estimating the OU process at *all* time instants, the quality of inference has to be summarized by a metric that is independent of time  $t$ . We consider two such performance metrics, first of which is the average variance

$$\text{Avar}(T) \triangleq \frac{1}{T} \int_0^T \text{Var}(\theta_t|\mathbf{y}) \, dt. \quad (6.13)$$

Note that average variance depends on the sampling period  $T$ , which is made explicit by the argument. However, there may be situations when the sampling period  $T$  is also subject to design. In this case, we need a metric that is independent of  $T$  as well. In this situation, we may use the performance metric to be the limiting value

$$\text{Var}_0 \triangleq \min_T \max_{t \in [0, T]} \text{Var}(\theta_t|\mathbf{y}), \quad (6.14)$$

which assumes that we select the sampling period  $T$  in a manner that minimizes the worst-case conditional variance for all time. We would consider both the metrics (6.13) and (6.14) in this chapter.

## 6.3 Main results

### 6.3.1 Single snapshot estimation

As motivated earlier, we consider an  $N$ -sensor network, the collaboration topology of which consists entirely of  $Q$ -cliques, where  $Q$  is a finite integer (see Figure 6.2). Let  $N = KQ$ , which ensures that there is an integral number of such cliques. Let the total energy available in the network be  $\mathcal{E}$ , which is finite. We consider the asymptotic limit when the network is large ( $N \rightarrow \infty$ ) but the transmission capacity of the equivalent Multiple-Input-Single-Output (MISO) channel is kept finite. We assume that the random variables  $\{\tilde{g}_n\}_{n=1}^N$  (which can be thought of as *unnormalized* channel gains) are iid realizations from the pdf  $f_{\tilde{g}}(\cdot)$  and

the channel gains are  $g_n = \frac{1}{\sqrt{N}}\tilde{g}_n$  so as to ensure that the transmission capacity remains the same even as the number of nodes increases<sup>1</sup>, thereby enabling a fair comparison of networks of various sizes. Without such a scaling, the transmission capacity would increase to infinity (and the resulting distortion would be driven down to zero) as the number of nodes increases, which is a trivial regime to consider.

Let  $J_{\text{opt}}$  and  $J_{\text{eq}}$  denote the asymptotic limits of the Fisher Information  $J_{\mathbf{W}}$  corresponding to the optimal (6.5) and equal energy-allocation strategies (6.8) respectively. The following results provide closed form expressions for these limits.

**Theorem 6.3.1** (Fisher Information for  $Q$ -clique topology).

$$(Optimal\ EA) \quad J_{\text{opt}} = \frac{\mathcal{E}}{\eta^2} \frac{\mathbb{E}[\tilde{g}^2]}{\xi^2} (1 - H_Q), \text{ and} \quad (6.15a)$$

$$(Equal\ EA) \quad J_{\text{eq}} = \frac{\mathcal{E}}{\eta^2} \frac{(\mathbb{E}[\tilde{g}])^2}{\xi^2} \frac{1}{1 + R_Q}, \quad (6.15b)$$

where  $H_Q$  and  $R_Q$  are defined as

$$H_Q = \mathbb{E} \left[ \frac{1}{1 + \frac{\eta^2}{\sigma^2} (h_1^2 + \dots + h_Q^2)} \right], \text{ and} \quad (6.16a)$$

$$R_Q = \frac{1}{Q (\mathbb{E}[h])^2} \left( \text{Var}[h] + \frac{\sigma^2}{\eta^2} \right), \quad (6.16b)$$

respectively.

The proof of Theorem 6.3.1 is relegated to Appendix A.13. A few remarks due to Theorem 6.3.1 are in order.

*Special Cases:* In general, the equal EA scheme is suboptimal, i.e.,  $J_{\text{opt}} \geq J_{\text{eq}}$ . However, the two energy allocation schemes are asymptotically equivalent when  $\text{Var}[h]$  and  $\text{Var}[\tilde{g}]$  are both zero, which is the case when the network is homogeneous, i.e.,  $\mathbf{h} = h_0 \mathbf{1}$  and  $\tilde{\mathbf{g}} = \tilde{g}_0 \mathbf{1}$

---

<sup>1</sup>Note that the channel capacity of the equivalent MISO channel is  $\frac{1}{2} \log \left( 1 + \frac{\mathcal{E} \|\mathbf{g}\|^2}{\xi^2} \right)$  and that  $\lim_{N \rightarrow \infty} \|\mathbf{g}\|^2 = \mathbb{E}[\tilde{g}^2]$  from the law of large numbers.

(say). For such a network,

$$J_{\text{opt}} = J_{\text{eq}} = \frac{\mathcal{E}\tilde{g}_0^2}{\xi^2\eta^2} \frac{1}{1 + \frac{1}{Q} \frac{\sigma^2}{\eta^2 h_0^2}}. \quad (6.17)$$

*Explicit expressions for Rayleigh distributed gains:* The evaluation of (6.15a) is, in general, hindered by the computation of  $H_Q$ , which involves the computation of a  $Q$ -dimensional integral. However, if the observation gains are Rayleigh<sup>2</sup> distributed  $f_h(h) = \text{Rayleigh}(h; \alpha_h)$ , we can show (derivation in Appendix A.14) that

$$H_Q = \frac{(-1)^{Q-1} \lambda^Q \exp(\lambda) \mathcal{E}i(\lambda) - \sum_{i=0}^{Q-2} i! (-\lambda)^{Q-1-i}}{(Q-1)!}, \quad (6.18)$$

where  $\lambda \triangleq \frac{\sigma^2}{2\alpha_h^2\eta^2}$  and  $\mathcal{E}i(z) \triangleq \int_z^\infty \exp(-t)/t dt$  is the exponential integral function. It immediately follows that  $H_1 = \lambda \exp(\lambda) \mathcal{E}i(\lambda)$ , which corresponds to the distributed case ( $Q = 1$ ), and  $H_Q \approx \frac{\lambda}{Q-1}$  for large values of  $Q$ . In our numerical simulations, we would consider Rayleigh distributed channel and observation gains, for which (6.18) will be useful.

## Simulation results

Theorem 6.3.1 is important since it provides a framework to evaluate the estimation performance for partially connected collaboration topologies. Though (6.15a) and (6.15b) are accurate indicators of performance for a structured network consisting only of  $Q$ -cliques, it is of interest to see how this insight applies for more complicated topologies. Towards that goal, we simulate the nearest-neighbor and random geometric graph topologies as described in Section 6.2.2. We consider a network with  $N = 10^4$  nodes, which is large enough to demonstrate convergent behavior. We consider  $\eta^2 = 1$ ,  $\xi^2 = 1$ ,  $f_h(h) = \text{Rayleigh}(h; 1)$ ,  $f_{\tilde{g}}(\tilde{g}) = \text{Rayleigh}(\tilde{g}; 1)$  and two values of observation noise variance, namely  $\sigma^2 = 1$  and  $\sigma^2 = 2$ . The operating energy is fixed at  $\mathcal{E} = 0.7$ . This choice of  $\mathcal{E}$  is made to reflect an

---

<sup>2</sup>A Rayleigh distributed random variable  $x$  with parameter  $\alpha$  has a probability density function  $\text{Rayleigh}(x; \alpha) = \frac{x}{\alpha^2} \exp\left(-\frac{x^2}{2\alpha^2}\right)$  for  $x \in [0, \infty)$ . The first two moments are  $\mathbb{E}[x] = \alpha\sqrt{\frac{\pi}{2}}$  and  $\mathbb{E}[x^2] = 2\alpha^2$  respectively.

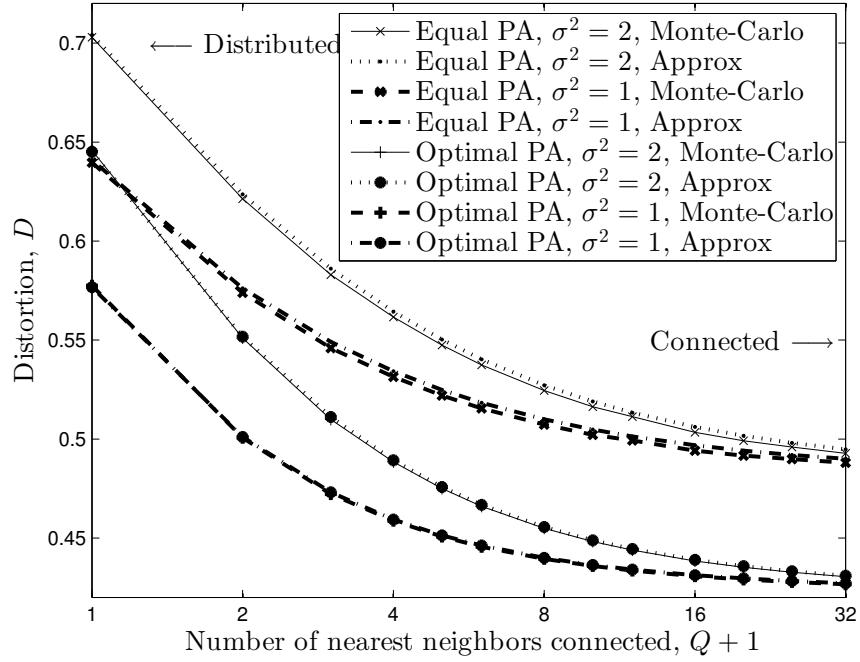
operating region where substantial performance gain is possible through spatial collaboration.

Numerical results obtained through Monte-Carlo simulations of both the optimal and equal EA strategies with varying degrees of spatial collaboration are illustrated in Figures 6.5(a) and 6.5(b) for the NN and RGG topologies respectively. For the  $(Q - 1)$ -nearest-neighbor case, the theoretical results corresponding to an equivalent problem with  $Q$ -cliques are juxtaposed. It is observed from Figure 6.5(a) that the performance of the two schemes are almost identical. For a random geometric graph, we consider that all the  $N$  sensors are randomly spread in a unit square. If the radius of collaboration is  $r$ , it follows that the expected number of neighbors is approximately  $\tilde{Q} = N\pi r^2$ , and that this approximation is more accurate for large values of  $r$ . Hence in Figure 6.5(b), the theoretical results corresponding to an equivalent problem with  $\tilde{Q}$ -cliques are juxtaposed. It is observed that the theoretical approximations compare favorably with the Monte-Carlo simulations, although they are less accurate compared to the  $(Q - 1)$ -nearest-neighbor case.

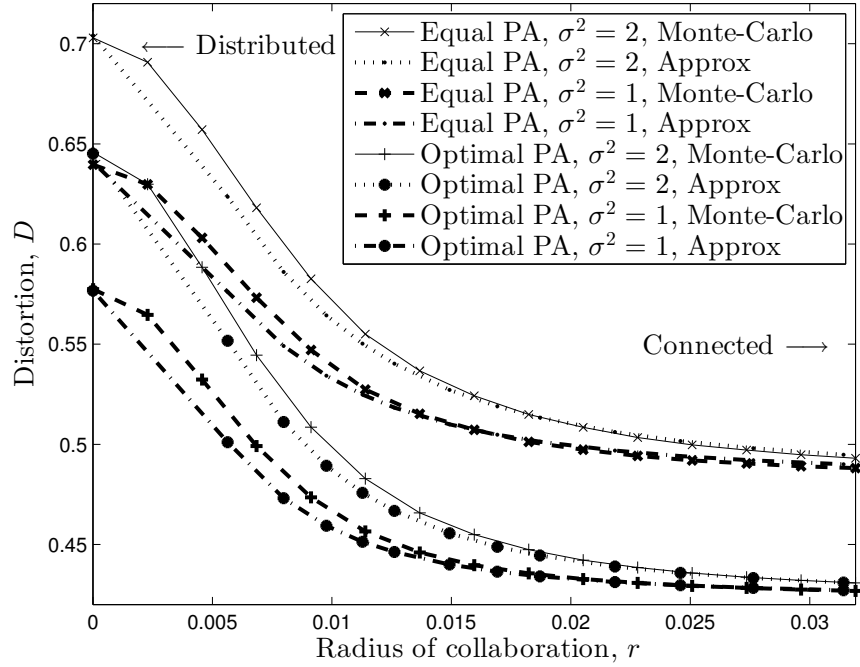
From the two examples given above, it is observed that only a small number of collaboration links are needed to achieve near-connected performance. In particular, the distortion performance is seen to saturate as early as  $Q \gtrapprox 20$ , though a fully connected network would imply  $Q = N = 10^4$  connections per node. This demonstrates the efficacy of spatial collaboration as an approach to enhance estimation performance beyond distributed networks.

### 6.3.2 Time varying process estimation

In this subsection, we compute the conditional variance of the OU process given the vector of periodically sampled observations. Towards computing (6.12), we begin by describing the matrix  $\mathbf{R}_{\mathbf{y}\mathbf{y}'}$  and vector  $\mathbf{R}_{\theta_i\mathbf{y}'}$ . The covariance matrix of the sampled parameter values



(a) Fixed number of nearest neighbors



(b) Random geometric graph

Figure 6.5: Energy-constrained estimation with single snapshot spatial sampling.

$\boldsymbol{\theta}$  takes the shape of the well known stationary matrix (e.g., [65]),

$$\mathbb{E}[\boldsymbol{\theta}\boldsymbol{\theta}'] = \eta^2 \mathbf{C}, \quad \mathbf{C} \triangleq \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \cdot \\ \rho & 1 & \ddots & \ddots & \vdots \\ \rho^2 & \ddots & \ddots & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & 1 & \rho \\ \cdot & \dots & \rho^2 & \rho & 1 \end{bmatrix}, \quad (6.19)$$

where  $\rho \triangleq e^{-T/\tau}$ . The structured matrix  $\mathbf{C}$  in Equation (6.19), is often referred to as the Kac–Murdock–Szegő matrix in the literature. From the additive model (6.10), it follows that

$$\mathbf{R}_{\mathbf{y}\mathbf{y}'} = (\mathbf{I} + \eta^2 J \mathbf{C}) / J. \quad (6.20)$$

Similarly, the following expression for  $\mathbf{R}_{\theta_t \mathbf{y}'}$  follows directly from the definition  $\varrho = e^{-t/\tau}$ , where  $t \in [0, T]$ ,

$$\mathbb{E}[\theta_t \mathbf{y}'] = \eta^2 [\dots, \rho^2 \varrho, \rho \varrho, \varrho, \rho / \varrho, \rho^2 / \varrho, \dots]. \quad (6.21)$$

With the help of the above descriptions of  $\mathbf{R}_{\mathbf{y}\mathbf{y}'}$  and  $\mathbf{R}_{\theta_t \mathbf{y}'}$ , computing (6.12) involves inverting the matrix  $\mathbf{I} + \eta^2 J \mathbf{C}$  (the asymptotic closed form expression for such an inverse can be found in the next chapter, see Proposition 7.3.1) followed by a quadratic product. The resulting algebra is involved but straightforward. We relegate details of the derivation to Appendix A.15 and state the result below.

**Theorem 6.3.2** (Variance of OU process estimates).

$$\text{Var}(\theta_t | \mathbf{y}) = \frac{\eta^2 \left[ 1 + \eta^2 J \rho' \left\{ 1 - \left( \frac{\varrho - \rho \varrho}{1 - \rho} \right)^2 \right\} \right]}{\sqrt{(\eta^2 J + \rho') (\eta^2 J + 1 / \rho')}}, \quad (6.22)$$

where  $\rho' \triangleq \frac{1 - \rho}{1 + \rho}$ ,  $\rho = e^{-T/\tau}$ ,  $\varrho = e^{-t/\tau}$ ,  $t \in [0, T]$ .



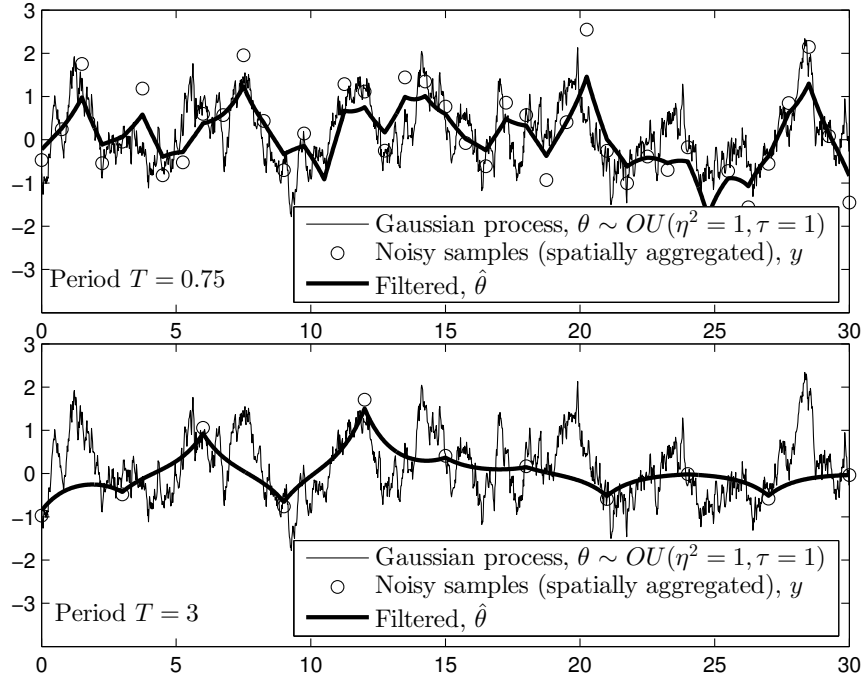
Equation (6.22) provides the closed form estimation variance of an OU process at any instant  $t \in [0, T]$ , provided that power constrained noisy samples are observed periodically with period  $T$ . In addition to  $\rho$ , the quantity  $J$  also depends on the sampling period  $T$  through the energy-FI (Fisher Information) relation  $J = \frac{cPT}{\eta^2}$ , which follows from Theorem 6.3.1 by using  $\mathcal{E} = PT$  and defining

$$c \triangleq \begin{cases} \frac{\mathbb{E}[\tilde{g}^2](1-H_Q)}{\xi^2} & \text{for Optimal EA,} \\ \frac{(E[\tilde{g}])^2}{\xi^2(1+R_Q)} & \text{for Equal EA.} \end{cases} \quad (6.23)$$

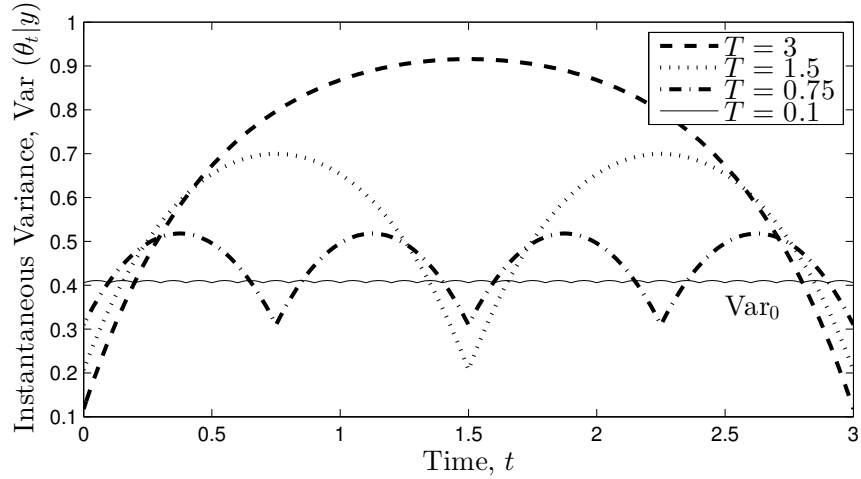
In the following discussions, we illustrate the power-constrained estimation of an OU process and show how the instantaneous variance (6.22) can be used to compute other performance measures described in Section 6.2.3, namely a) average variance,  $\text{Avar}(T)$  and b) min-max performance limit,  $\text{Var}_0$ .

## Simulation results

In Figure 6.6(a), we visualize the estimation of an OU process with stationary variance  $\eta^2 = 1$  and covariance drop-off parameter  $\tau = 1\text{s}$ . We simulate a total duration of  $T_{\text{obs}} = 30\text{s}$ , during which we consider sampling the same process using two different sampling periods,  $T = 0.75\text{s}$  (top) and  $T = 3\text{s}$  (bottom). The sampling noise sequence  $\{v_{kT}\}$ , which is an abstraction of the spatial data aggregation, is simulated as independent zero-mean Gaussian random variables with variance  $\frac{1}{2.5T}$  for the two different sampling periods. The inverse relation  $\text{Var}(v_{kT}) \propto \frac{1}{T}$  is due to the fact that  $v_{kT}$  represents a noise with variance  $\frac{1}{J}$  and the Fisher Information  $J = \frac{cPT}{\eta^2}$  as per Theorem 6.3.1. The constant 2.5 (which represents the quantity  $\frac{cP}{\eta^2}$ ) was chosen so as to produce a visible contrast between the sampling errors corresponding to the chosen sampling periods. As can be seen in Figure 6.6(a), the samples are obtained almost without any noise for  $T = 3\text{s}$  (bottom). The circles representing noisy samples align almost on top of the thin line that represents the path of the OU process. The samples are, however, significantly noisy for  $T = 0.75\text{s}$  (top), since less energy is available



(a) Effect of sampling period on aggregate noise and subsequent filtering



(b) Instantaneous variance for various sampling periods

Figure 6.6: Power-constrained estimation of OU process

per sampling duration. This is evidenced by the circles lying significantly distant from the OU process path. The filtered estimates  $\hat{\theta}_t = \mathbb{E}[\theta_t | \{y_{kT}\}]$  are obtained by applying (6.11) and are shown by the bold lines. When compared visually, the filtered estimates appear more accurate in the case of smaller sampling period (top). This observation is justified by plotting the steady state variance, as obtained from Theorem 6.3.2, in Figure 6.6(b) for various sampling periods  $T = \{0.1, 0.75, 1.5, 3\}$ . Though the best-case variance (occurring at  $t = kT$ ) is higher for smaller sampling periods, the worst-case variance (occurring at

$t = (k + 0.5)T$ ) goes down as the OU process is sampled more frequently. Because the power is kept constant, the variance converges to a finite value (rather than vanishing) for small values of  $T$ . Since the gap between the worst-case and best-case scenarios reduces with  $T$ , the limiting variance ( $\approx 0.4$ , annotated as  $\text{Var}_0$ ) is flat with respect to time. From Equation (6.22), the limiting variance can be derived precisely to be

$$\text{Var}_0 \triangleq \lim_{T \rightarrow 0} \text{Var}(\theta_t | \mathbf{y}) = \frac{\eta^2}{\sqrt{1 + 2P\tau c}}, \quad (6.24)$$

which also means that  $\text{Var}_0$  trivially satisfies

$$\min_T \max_{t \in [0, T]} \text{Var}(\theta_t | \mathbf{y}) = \text{Var}_0, \quad (6.25)$$

thereby answering the question of min-max performance limit as posed earlier in Equation (6.14). The following result is obtained by substituting in (6.24) the value of constant  $c$  (see (6.26)), thereby stating explicitly how the performance limit depends on channel conditions and collaboration topology.

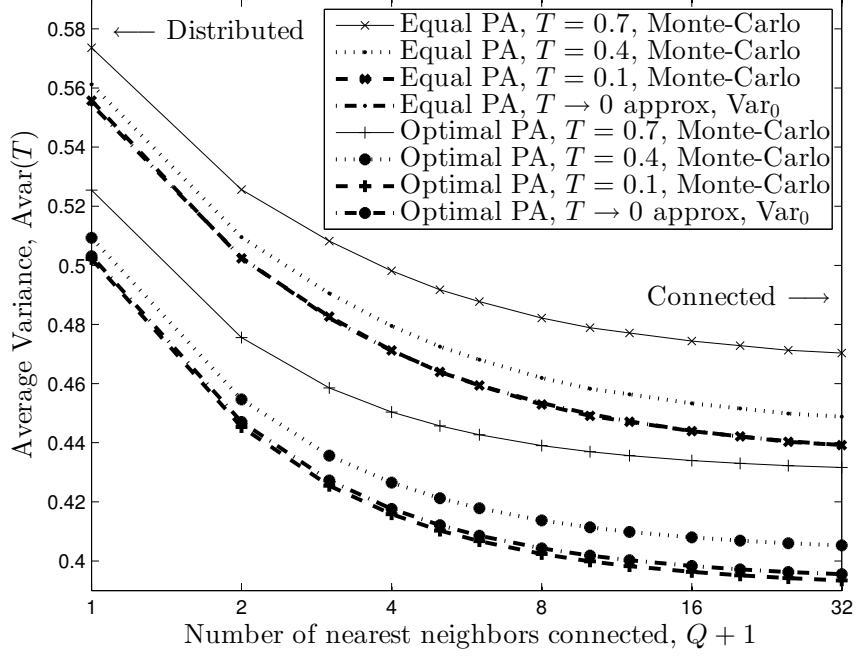
**Corollary 6.3.3** (Min-max performance limits).

$$\text{Var}_0 = \begin{cases} \eta^2 / \sqrt{1 + \frac{2P\tau \mathbb{E}[\tilde{g}^2](1-H_Q)}{\xi^2}} & \text{for Optimal EA,} \\ \eta^2 / \sqrt{1 + \frac{2P\tau (\mathbb{E}[\tilde{g}])^2}{\xi^2(1+R_Q)}} & \text{for Equal EA.} \end{cases} \quad (6.26)$$

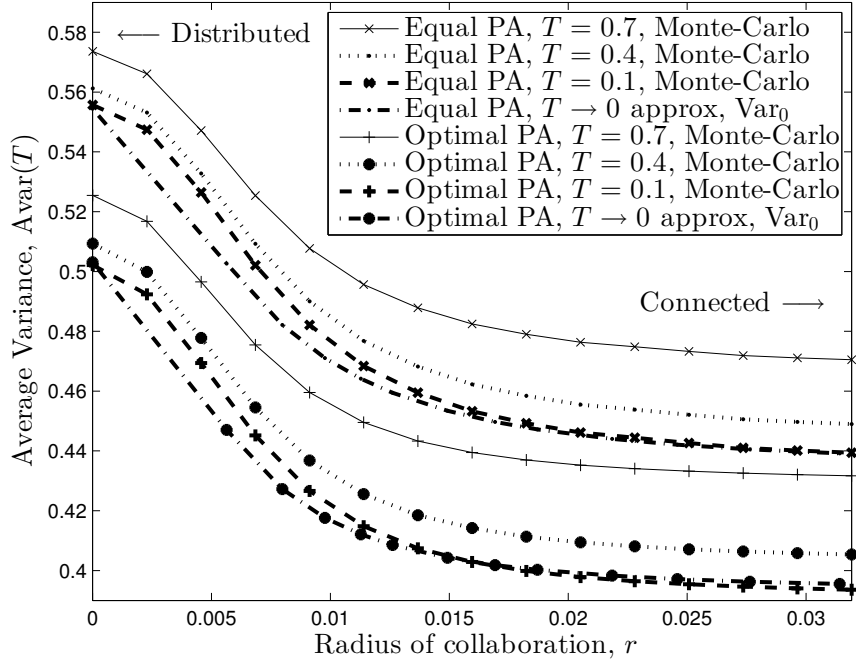
In addition to the instantaneous variance and min-max performance limits, one may also be interested in the average variance  $\text{Avar}(T)$  as defined by Equation (6.13) in Section 6.2.3. The average variance is obtained by integrating (6.22) over  $t \in [0, T]$ . Since  $\varrho = e^{-t/\tau}$  is the only variable in (6.22) that depends on  $t$ , we obtain

$$\text{Avar}(T) = \frac{\eta^2 \left[ 1 + \eta^2 J \rho' \left\{ 1 - \frac{\mathcal{I}_T}{(1-\rho)^2} \right\} \right]}{\sqrt{(\eta^2 J + \rho')(\eta^2 J + 1/\rho')}}, \quad (6.27)$$

where  $\mathcal{I}_T \triangleq \frac{1}{T} \int_0^T (\varrho - \rho/\varrho)^2 dt = \frac{1 - \rho^2}{T/\tau} - 2\rho$ .



(a) Fixed number of nearest neighbors



(b) Random geometric graph

Figure 6.7: Power-constrained estimation of OU process - Average variance

We use average variance as the performance metric in the following simulation, in which we consider both the aspects discussed in this chapter, namely the spatial aggregation procedure and the temporal dynamics. The simulation settings are similar to those considered in Section 6.3.1, which we repeat here for the sake of completeness. The sensor network

comprises of  $N = 10^4$  nodes. We consider  $\eta^2 = 1$ ,  $\xi^2 = 1$ ,  $\sigma^2 = 1$ ,  $f_h(h) = \text{Rayleigh}(h; 1)$  and  $f_g(g) = \text{Rayleigh}(g; 1)$ . The exponential drop-off parameter is set to  $\tau = 1\text{s}$  and an observation duration of  $T_{\text{obs}} = 30\text{s}$  is considered, which is large enough to demonstrate steady state behavior. The observation duration is discretized into  $M = 1600$  instants for generating the OU process sequence using a first-order autoregressive model. The average variance is obtained as the mean of the deviations from all  $M$  estimates. Three values of sampling period are considered for simulation, namely  $T = \{0.7, 0.4, 0.1\}$ . The limiting value when  $T \rightarrow 0$ ,  $\text{Var}_0$ , is also shown on all the graphs. The operating power is chosen as  $P = 1.4$  to reflect an operating region where substantial performance gain is possible through spatial collaboration. Both NN (Figure 6.7(a)) and RGG (Figure 6.7(b)) topologies are considered to show the applicability of the  $Q$ -clique results to practical collaboration scenarios. As earlier, both the equal-EA and optimal-EA spatial energy allocation strategies are simulated. The results in Figure 6.7 show that the average variance decreases with  $T$ . Though we have not rigorously proved that  $\text{Avar}(T)$  is monotonically decreasing in  $T$ , this assertion can be visually verified from Figure 6.6(b), by comparing the area under the curves for any two sampling periods ( $T = 3$  and  $T = 1.5$ , say). This observation coupled with the min-max property of  $\text{Var}_0$  leads to the conclusion that an OU process should be sampled as frequently as possible, even if that implies that less energy is available per sampling period (resulting in more noisy samples). However, this assertion is based on the assumption that the sampling noise is temporally white. In practical situations, the sampling errors may become correlated if the samples are obtained too frequently, and caution must be exercised to make sure that the temporal independence assumption is valid.

## 6.4 Summary

In this chapter, we have considered the linear coherent estimation problem in wireless sensor networks and investigated two key aspects. First, we have provided an asymptotic analysis of the single-snapshot estimation problem when the collaboration topology is only partially connected. We achieve this by obtaining the solutions for a family of structured

networks and then using those solutions to approximately predict the performance of more sophisticated networks using geometric arguments. Second, we have extended the problem formulation towards the estimation of a time varying signal. In particular, we have derived the instantaneous, average and worst case performance metrics when the signal is modeled as a Gaussian random process with exponential covariance. Both these aspects were investigated under the assumption of spatial and temporal independence among the measurement and channel noise samples. An important topic of research may be to relax this assumption and observe the effect of spatial and temporal correlation on the estimation performance.

# Chapter 7

## Signal estimation using sensors with baseline drift

### 7.1 Introduction

In all the discussions so far in this thesis, sensor inaccuracies are modeled as temporally independent measurement errors. In fact, this temporal independence property of noise is a convenient assumption in much of the sensor network literature, e.g., [73], [64] - one which greatly serves to bring down the complexity of the problem at hand. However, it has been widely reported that a variety of sensors used in various applications exhibit systematic errors or *drift* (to be formalized later), e.g., Ground Moving Target Indicator radar sensors [5], tilt sensors in bridge monitoring applications [92], CO<sub>2</sub> sensors used for air quality monitoring [23] and salinity sensors for ocean monitoring [66]. For example, gas sensors are known to drift due to other environmental parameters like temperature and humidity [38]. In this chapter, we drop the assumption of temporal independence of measurement noise and seek to explore the problem of estimation using sensors that exhibit baseline drift.

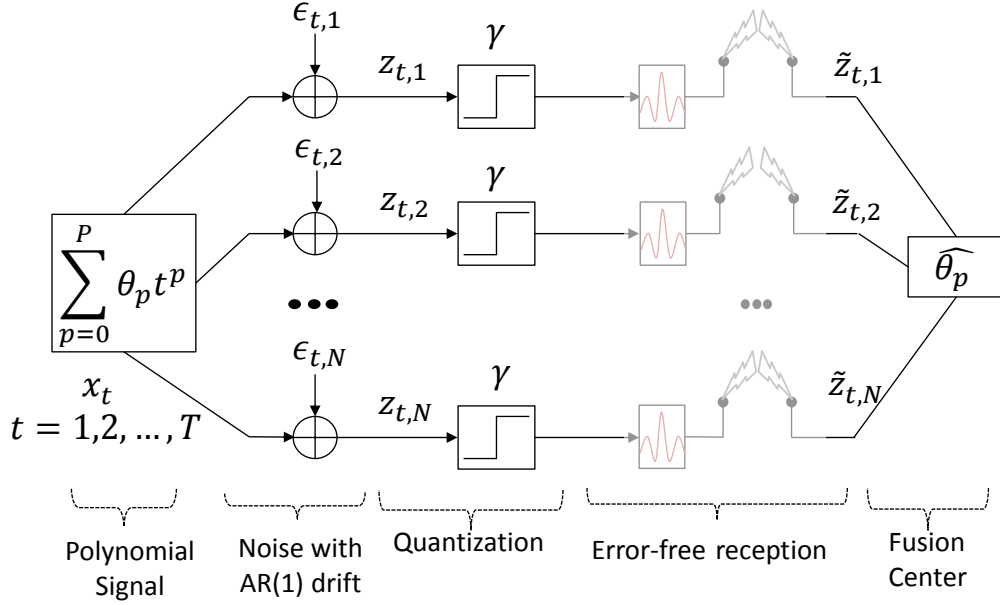


Figure 7.1: Estimation of a polynomial signal using sensors with baseline drift.

### 7.1.1 Inference with drifting sensors

When sensors in a network develop systematic errors with time, inferences based on these observations become increasingly inaccurate. For example, a scheme where each sensor can track their own drift in a collaborative manner is presented in [79] and such a system was shown to accumulate error over time. This necessitates periodic calibration/ registration of the sensors, a procedure that increases the operating cost for the network. Though the issue of drift is often acknowledged in the literature, performance analysis of a network consisting of drifting sensors has not been done before. Our goal in this chapter is to characterize the estimation performance of a sensor network in terms of the drift properties of the constituent sensors.

The schematic diagram of such a sensor network is illustrated in Figure 7.1. Multiple sensors  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_N$  are deployed in a field of interest to monitor a particular environment. The sensors observe the same phenomena over  $t = 1, 2, \dots, T$  time instants and their observations are corrupted by noise and drift. The noisy observations are relayed to a common sink, where the objective is to estimate the parameters governing the observed phenomena. Let  $z_{t,n}$  denote the noisy observation of the  $n^{\text{th}}$  sensor at the  $t^{\text{th}}$  instant and  $\mathbf{z}_n$



denote the vector  $[z_{1,n}, z_{2,n}, \dots, z_{T,n}]'$ . We assume a linear observation model of the form

$$\mathbf{z}_n = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}_n, \quad \boldsymbol{\epsilon}_n \sim \mathcal{N}(0, \boldsymbol{\Sigma}_n) \quad (7.1)$$

where  $\mathbf{x} \triangleq \mathbf{X}\boldsymbol{\theta}$  and  $\boldsymbol{\epsilon}_n$  are  $T \times 1$  vectors denoting the signal and error terms. Here  $\mathbf{X}$  is assumed to be a  $T \times (P + 1)$  known matrix (describing the temporal-dynamics) and  $\boldsymbol{\theta}$  is the  $(P + 1) \times 1$  vector denoting an unknown but deterministic signal. The error term  $\boldsymbol{\epsilon}_n$  consists of noise and drift components (to be described later) and is assumed to be Gaussian distributed with covariance matrix  $\boldsymbol{\Sigma}_n$ . The observations  $\mathbf{z}_n$  are possibly quantized and communicated to a sink, whose job is to obtain an accurate estimate of  $\boldsymbol{\theta}$ . From estimation theory (see, for example, [49] and [6]), it is well known that the variance of an unbiased estimator is lower bounded by the Cramér-Rao Lower Bound (CRB) and that the Maximum-Likelihood (ML) estimator asymptotically (for small errors) attains that bound. In this chapter, we derive the closed-form CRB (upto reasonable approximations) for estimating  $\boldsymbol{\theta}$  when  $\boldsymbol{\Sigma}_n$  corresponds to drift corrupted errors and  $\mathbf{X}$  corresponds to a polynomial signal. In the subsequent discussion, we provide the motivation and full descriptions of  $\boldsymbol{\Sigma}_n$  and  $\mathbf{X}$ .

### 7.1.2 Models of noise and drift

At instant  $t$ , let  $x_t$  and  $\epsilon_{t,n}$  denote the common signal magnitude and observation error of sensor  $n$ . When the effect due to drift is ignored, the measurement noise is often modeled as independent and identically distributed Gaussian noise  $w_{t,n}$ , (e.g., [73], [64]), i.e.,

$$z_{t,n} = x_t + w_{t,n}, \quad w_{t,n} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_n^2), \quad (7.2)$$

where  $\sigma_n^2$  is the noise variance. However, in the presence of drift, the observation error has two components: one due to baseline-drift  $d_{t,n}$  and the other due to random measurement noise  $w_{t,n}$ ,

$$z_{t,n} = x_t + d_{t,n} + w_{t,n}. \quad (7.3)$$

Drift is generally described as a gradual change of the bias of a sensor [38], [93]. Depending on the specific sensing methodology, various models have been proposed to characterize the drift sequence  $\{d_{t,n}\}_{t=1}^T$  in frequency and time domains (for a survey, see [38]). Below we describe three commonly used models of drift.

- *Frequency domain:* The phase-drift in an oscillator is often modeled in the frequency domain using several powers of frequency (power-law model) [93], i.e., the power spectral density (PSD) is of the form,  $S_n(f) = \sum_{i=-2}^2 h_{n,i} f^i$ , where  $h_{n,i}$  are appropriate constants.
- *Temporal domain - Deterministic:* Drift is also modeled sometimes as a linear movement of the sensor baseline, i.e.,  $d_{t,n} = a_n + (t - 1)b_n$ , where the intercept  $a_n$  is set to zero after each calibration and slope  $b_n$  is assumed to be an unknown constant that is often estimated later and compensated for. Example applications include odor identification using gas-sensor arrays [70] and air pollution monitoring using gas sensor networks [82].
- *Temporal domain - Random:* In the broader signal processing literature, autoregressive moving average (ARMA) processes are often used to describe serially correlated time series, an example of which is drift. As a trade-off between modeling-efficiency and analytical-complexity, drift is often modeled as a first-order autoregressive (AR(1)) process. Example applications include Ground Moving Target Indicator (GMTI) radars ([5], [48]), Ring-laser Gyroscopes [75], Liquid Chromatography [36] and sensor networks [79].

In this chapter, we will use the first-order autoregressive model to describe the statistical properties of drift. The AR(1) model characterizes the drift behavior at  $n^{\text{th}}$  sensor in terms of an auto-correlation parameter  $\rho_n$  (visually, a smaller value of  $\rho_n$  means that the baseline drift crosses the zero-line more often and looks more like white noise) and a strength

parameter  $\sigma_{\delta,n}^2$  (signifying the magnitude of drift),

$$\begin{aligned} d_{t+1,n} &= \rho_n d_{t,n} + \delta_{t,n}, \text{ where} \\ \rho_n &\in [0, 1] \text{ and } \delta_{t,n} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\delta,n}^2). \end{aligned} \tag{7.4}$$

Note that, when  $\rho = 1$ , the drift is modeled effectively as a non-stationary random walk, as in [79].

If we define  $\gamma_n$  such that  $\sigma_{\delta,n}^2 = \gamma_n \sigma_n^2$ , our “AR(1)+White noise” model for observation error is completely parameterized by  $\{\sigma_n^2, \rho_n, \gamma_n\}$ . These noise and drift parameters usually have to be identified from the empirical PSD for stationary noise (e.g., [36], [39]) or other time-domain features for non-stationary noise (e.g., [37]). In this chapter, we consider the characterization of the sensing uncertainty in terms of  $\{\sigma_n^2, \rho_n, \gamma_n\}$  as part of system identification that must be done prior to an observation cycle. Estimation of these parameters is beyond the scope of this chapter. Moreover, within an observation cycle, the drift-parameters are assumed to be constant. If the drift parameters change frequently, our proposed framework must be used in conjunction with periodic system re-identifications.

### 7.1.3 Deterministic signal model

Often in the sensor network literature, the signal of interest is assumed to be constant over the observation duration, e.g., [73],[64]. This means  $x_t = \theta, \forall t$ . However, such a signal model may be too simplistic for real applications and we consider a generalization of the form,

$$x_t = \sum_{p=0}^P \theta_p t^p, \quad \theta_p \in \mathbf{R}, \tag{7.5}$$

where  $\theta_p$ -s are the unknown constants that need to be estimated and  $P$  is the order of the polynomial time-series that is assumed to be known. In vector notations, the polynomial

signal  $\mathbf{x} \triangleq [x_1, x_2, \dots, x_T]'$  assumes a linear form  $\mathbf{x} = \mathbf{X}\boldsymbol{\theta}$ , where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^P \\ \vdots & \vdots & \ddots & \vdots \\ 1 & T & \cdots & T^P \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_P \end{bmatrix}. \quad (7.6)$$

It may be noted that time-varying signals in different applications are approximated by either polynomials or piecewise-polynomials [84],[90]. For example, a polynomial regression-based data gathering algorithm for environmental monitoring applications was suggested in [91]. Also, a polynomial spline approximation of stationary random processes, as applied to Clarke's model of multipath fading channels, was presented in [97].

This completes the description of the signal and observation noise considered in this chapter. We intend to derive the closed form CRB (for the estimation of  $\theta_p$ -s) in terms of the signal  $(P, T)$  and noise  $\{\sigma_n^2, \rho_n, \gamma_n\}$  parameters. This would help us characterize the performance of a sensor network and increase our understanding about its estimation capabilities.

#### 7.1.4 Related work

A related area of work is the study of systematic-bias or model-error estimation schemes using multiple, and sometimes collaborative sensors. In the radar signal processing literature, the process of model-based estimation and subsequent removal of systematic errors prior to target tracking is known as sensor-registration [15], [99]. In the weather research literature, the serially-correlated forecasting error arising due to modeling deficiency is often considered separately and tracked alongside the model parameters [16], [102]. In the sensor network literature, drift-aware networks perform learning-based collaborative bias estimation to enhance the effective lifetime of the network [79], [80]. However, in this chapter, we are focused on the quality of estimation in the presence of systematic errors, rather than techniques on mitigating systematic errors.

Several researchers have studied the Cramér-Rao bounds for polynomial (or polyphase) signal estimation in the presence of independent (or correlated) noise. The CRB is usually obtained from the inverse of the Fisher Information Matrix (FIM) [49]. For Additive White Gaussian Noise (AWGN), the large sample approximation of FIM is known to be a multiple of the Hilbert Matrix (e.g., [56]). The second order approximation was derived in [71] in the context of polynomial phase signals. For a mixture of additive and multiplicative white Gaussian noise, the large sample FIM was shown in [78] to be a scalar multiple of the AWGN case. *Our primary contribution in this chapter is the derivation of (approximate) closed-form CRB for polynomial signal estimation in the presence of a mixture of white (measurement noise) and AR(1) Gaussian noise (drift).* We also discuss the non-stationary case when the autoregressive parameter is equal to 1. To the best of author's knowledge, polynomial signal estimation in such a mixture of noise has not been considered earlier. As mentioned earlier, the study of estimating polynomial signals in AR(1)+White noise would help us characterize the capability of a sensor network to infer the parameters of a time-varying signal using sensors with drift.

The rest of the chapter is organized as follows. In Section 7.2, we formally list the assumptions and set up the problem for two scenarios based on the initial state of the sensor. In Section 7.3, we consider the single-sensor case and derive large sample approximations of the CRB for  $\rho < 1$  and  $\rho = 1$ . In Section 7.4, we extend the results to multiple sensors having different drift characteristics. In Section 7.5, we demonstrate the application of the results to a bandwidth limited sensor network that communicates only quantized observations. Concluding remarks are provided in Section 7.6.

## 7.2 Problem formulation

To be more specific about our problem framework, we formally state the assumptions below.

1. *Same phenomenon:* Each sensor is observing the same physical phenomenon (e.g., temperature), which is modeled as a time-polynomial signal. In our framework, if

multiple signals are to be sensed (e.g., temperature and humidity), the observations have to be transmitted separately and the parameters independently estimated.

2. *Known drift statistics:* The drift in each sensor is modeled as AR(1) Gaussian time-series with known statistics - namely the autoregressive and strength parameters,  $\rho_n$  and  $\gamma_n$  respectively. The drift parameters are assumed to be accurately estimated from previous experiments and remain unchanged during the measurement duration.
3. *Spatially uncorrelated noise and drift:* We assume that the observations at all the sensors at a particular instant are independent, conditioned on the signal magnitude at that instant. In other words, though the observation noise samples  $d_{t,n} + w_{t,n}$  are temporally correlated (due to drift), there is no spatial correlation among them.
4. *Synchronized observations:* The clocks of the sensors are synchronized and they have identical sampling intervals. Hence they collect the observations at the same time instants. It may be noted that clock synchronization in sensor networks can be achieved through periodic gossip among neighboring sensors [18].
5. *Parallel sensor network with fusion center (Figure 7.1):* We assume that the sensors do not collaborate among themselves and rather communicate their observations only to the sink. The sink, having collected all the observations from the individual sensor nodes, performs inference and takes appropriate action. In the distributed inference literature, this topology is known as a parallel sensor network [87]. It may be noted in passing that this is in contrast to other frameworks where inference is performed in-network and without a fixed fusion center, e.g., [40].
6. *Reliable signal transmission:* In Sections 7.3 and 7.4, we assume that the noisy observations are communicated perfectly to the sink without any further distortion. We call this the *full-precision* case which helps obtain a benchmark performance. However, there may be cases when, due to power and bandwidth limitations at the sensor nodes, reliable communication of full-precision observations may be impossible. In such cases, a digital communication based framework in conjunction with efficient

channel coding can be used to reliably transmit only a finite number of bits [73], [58]. In Section 7.5, we will discuss inference using digitized observations where all sensor nodes perform quantization with identical fidelity. It may be noted in passing that, for sensor networks where there is a constraint on system-wide bandwidth (summed across all nodes), different sensors nodes may be assigned different fidelity of quantization based on their observation quality [51]. Though outside the scope of this chapter, optimal fidelity assignment for sensors in the presence of drift is a challenging topic worthy of future research.

With the above-mentioned assumptions, our goal in this chapter is to derive approximate closed form expressions for the Cramér-Rao bounds.

### 7.2.1 Sensor calibration and noise covariance

Consider a single sensor with noise and drift properties denoted by  $\{\sigma^2, \gamma, \rho\}$  (the subscript  $n$  is dropped for most of this subsection). Assume that inference has to be performed from noisy observations  $z_t$  at instants  $t = 1, \dots, T$ . Let  $\tau > 0$  denote the time-instants elapsed since the sensor was last calibrated, i.e., when the drift component was set to zero by correcting the baseline. Since the duration of inference starts from instant 1, it follows that the drift sequence was set to zero at instant  $1 - \tau$  (which is a non-positive index, a slight notational inconvenience). Therefore, the drift sequence proceeds as follows,

$$d_{1-\tau} = 0, d_{2-\tau} = \delta_1, d_{3-\tau} = \rho\delta_1 + \delta_2 \dots \text{etc..} \quad (7.7)$$

Since each of the drift innovations  $\delta_i \sim \mathcal{N}(0, \gamma\sigma^2)$ , we have

$$\text{Var}(d_t) = \gamma\sigma^2(1 + \rho^2 + \dots + \rho^{2(\tau+t-1)}) =: \gamma\sigma^2 S_t^\tau, \quad (7.8)$$

where the summation within the parenthesis is defined as  $S_t^\tau$ . Though the drift sequence  $\{d_t\}_{t=1}^T$  is stationary for large magnitudes of  $\tau$ ,

$$\lim_{\tau \rightarrow \infty} S_t^\tau = \frac{1}{1 - \rho^2}, \quad \rho < 1, \quad (7.9)$$

it is not stationary in the transient stage, for which,

$$S_1^\tau < S_2^\tau < \dots < S_T^\tau, \quad \tau < \infty. \quad (7.10)$$

In this chapter, though we would solve the estimation problem for a general  $\tau$ , we would refer to the limiting cases  $\tau = 1$  as *calibrated* (C) and  $\tau \rightarrow \infty$  as *uncalibrated* (U) respectively.

We define  $\mathbf{R}$  such that  $\gamma\sigma^2\mathbf{R}$  is the covariance matrix of the drift vector  $\mathbf{d} \triangleq [d_1, d_2, \dots, d_T]'$

$$\begin{aligned} \mathbb{E}[\mathbf{d}'\mathbf{d}] &= \gamma\sigma^2 \begin{bmatrix} S_1^\tau & \rho S_1^\tau & \rho^2 S_1^\tau & \dots & \rho^{T-1} S_1^\tau \\ \rho S_1^\tau & S_2^\tau & \rho S_2^\tau & \dots & \rho^{T-2} S_2^\tau \\ \rho^2 S_1^\tau & \rho S_2^\tau & S_3^\tau & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \rho^{T-1} S_1^\tau & \rho^{T-2} S_2^\tau & \dots & \ddots & S_T^\tau \end{bmatrix} \\ &=: \gamma\sigma^2 \mathbf{R}. \end{aligned} \quad (7.11)$$

For the uncalibrated case, the diagonal elements are the same ( $S_t^\tau = \frac{1}{1-\rho^2}$ ), and  $\mathbf{R}$  takes the shape of the well known stationary matrix (e.g., [65]),

$$\mathbb{E}[\mathbf{d}'\mathbf{d}] = \frac{\gamma\sigma^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & 1 \end{bmatrix}, \quad (7.12)$$



sometimes referred to as the Kac–Murdock–Szegő matrix.

For the  $n^{\text{th}}$  sensor, we denote the covariance of the drift sequence by  $\mathbf{R}_n$ . Let  $\mathbf{\Sigma}_n$  denote the covariance matrix of the total error (AR(1)+White noise), i.e.,  $\boldsymbol{\epsilon}_n = \mathbf{w}_n + \mathbf{d}_n$ , so that,

$$\mathbf{\Sigma}_n = \sigma_n^2 (\mathbf{I} + \gamma_n \mathbf{R}_n), \quad (7.13)$$

where  $\mathbf{I}$  is the  $T \times T$  identity matrix.

### 7.2.2 Maximum-Likelihood estimation and Cramér-Rao bound

Given the linear observation model (7.1) and noise covariance described in Section 7.2.1, the Maximum-Likelihood (ML) estimator of  $\boldsymbol{\theta}$  at the sink is of the form (for a reference on estimation theory, see [49], [6]),

$$\hat{\boldsymbol{\theta}}^{\text{ML}} = \mathbf{J}^{-1} \mathbf{X}' \sum_{n=1}^N \mathbf{\Sigma}_n^{-1} \mathbf{z}_n, \quad (7.14)$$

where  $\mathbf{J}$  is known as the Fisher Information matrix,

$$\mathbf{J} \triangleq \mathbf{X}' \left( \sum_{n=1}^N \mathbf{\Sigma}_n^{-1} \right) \mathbf{X}, \quad (7.15)$$

It is well known in estimation theory [49] that within the class of unbiased estimators, the ML estimator of a linear model is optimal in terms of estimation variance [49]. Also, the least possible estimation variance is provided by the Cramér-Rao lower bound,  $\mathbf{V} \triangleq \mathbf{J}^{-1}$ , so that

$$\mathbb{E} \left[ (\theta_p - \hat{\theta}_p^{\text{ML}})^2 \right] \geq \mathbf{V}_{p,p} = [\mathbf{J}^{-1}]_{p,p}, \quad (7.16)$$

for  $0 \leq p \leq P$ . Since Equation (7.16) holds with a strict equality for linear models, the CRB is an appropriate performance metric for our problem.

It is unclear from (7.15) and (7.16) exactly how the estimation variance  $\mathbf{V}_{p,p}$  depends on the drift parameters  $\sigma_n^2, \gamma_n, \rho_n$ , signal parameters  $P$  and the sample size  $T$ . *Our goal in this chapter is to derive large sample approximations for these Cramér-Rao bounds and thereby provide insight into the behavior of estimation performance as it relates to the drift properties of the sensors.*

## 7.3 Main result: single sensor case

Towards obtaining the CRB for the general case of multiple sensors, we first consider the single-sensor scenario and thereby obtain the core results of this chapter. Subsequent applications of these core results to the multi-sensor scenario (Section 7.4) and bandwidth limited networks (Section 7.5), as we shall see later, will be somewhat straightforward extensions.

While analyzing the single-sensor scenario, for notational brevity, we drop the sensor-index subscript  $n$  from  $\sigma_n^2, \rho_n, \gamma_n, \mathbf{R}_n$  and  $\Sigma_n$  for the remainder of this section. Note from (7.15) and (7.16) that, for a single sensor, the Fisher Information matrix is  $\mathbf{J} = \mathbf{X}'\Sigma^{-1}\mathbf{X}$  and the Cramér-Rao bound is  $[\mathbf{V}]_{p,p} = [\mathbf{J}^{-1}]_{p,p}$ . In the following discussion, we first compute the inverses of the disturbance covariance matrices. Next, we use those results to derive the CRB.

### 7.3.1 Inverse of disturbance covariance

Our goal is to approximate  $\Sigma^{-1}$  in a form that are analytically tractable. The exact closed form expression for  $\Sigma_U^{-1}$  (for  $\tau \rightarrow \infty$ ) is known to be quite complicated (e.g., p-53 of [76], [25]) and we will not use it. The author have not found a closed form expression for either  $\Sigma_C^{-1}$  (the  $\tau = 1$  case) or  $\Sigma^{-1}$  (for general  $\tau$ ) in the literature.

In this chapter, we propose an approximation to  $\Sigma^{-1}$  that is novel to the best knowledge of the author. We suggest the following form for the inverse

$$(\mathbf{I} + \gamma \mathbf{R})^{-1} = \mathbf{I} - \nu \mathbf{A} + \mathcal{O}(y^T), \quad (7.17)$$

where  $\nu$  is a constant,  $\mathbf{A}$  is a matrix with certain structure (to be described later), and  $y$  is a quantity less than 1, so that  $\mathcal{O}(y^T)$  represents a term that vanishes exponentially with sample size  $T$ . The idea is that, for large  $T$ , we should be able to use the approximation

$$(\mathbf{I} + \gamma \mathbf{R})^{-1} \approx \mathbf{I} - \nu \mathbf{A} \quad (7.18)$$

towards computing the CRB. The specifics of this approximation are laid out in Proposition 7.3.1.

**Proposition 7.3.1.** *Based on drift parameters  $\gamma$  and  $\rho$ , define the following constants,*

$$\begin{aligned} y &\triangleq \frac{1}{2} \left[ \frac{\gamma + 1}{\rho} + \rho - \sqrt{\left( \frac{\gamma + 1}{\rho} + \rho \right)^2 - 4} \right], \\ \nu &\triangleq \frac{y\gamma}{\rho(1 - y^2)}, \text{ and } \kappa \triangleq \frac{y(\rho - y)}{1 - \rho y}. \end{aligned} \quad (7.19)$$

Then  $(\mathbf{I} + \gamma \mathbf{R})^{-1} = \mathbf{I} - \nu \mathbf{A} + \mathcal{O}(y^T)$ , where  $\mathbf{A}$  has the structure shown in Equation (7.21), with

$$\begin{aligned} a_i &\triangleq 1 + y^{2(i-1)} \eta_\tau, \quad b_j \triangleq 1 + y^{2(j-1)} \kappa, \\ \eta_\tau &\triangleq \frac{1 - y^2}{1 - \rho y + \varrho_\tau y / \rho} - 1, \quad \varrho_\tau \triangleq \frac{\rho^{2\tau}}{1 + \rho^2 + \dots + \rho^{2\tau-2}}. \end{aligned} \quad (7.20)$$

$$\mathbf{A} = \begin{bmatrix} a_1 & ya_1 & y^2a_1 & \cdots & \cdots & \cdots & y^{T-2}a_1 & y^{T-1}b_1 \\ ya_1 & a_2 & ya_2 & \cdots & \cdots & y^{T-4}a_2 & y^{T-3}b_2 & \vdots \\ y^2a_1 & ya_2 & a_3 & & & y^{T-5}b_3 & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & y^{T-4}a_2 & y^{T-5}b_3 & \cdots & \cdots & b_3 & yb_2 & y^2b_1 \\ y^{T-2}a_1 & y^{T-3}b_2 & \cdots & \cdots & \cdots & yb_2 & b_2 & yb_1 \\ y^{T-1}b_1 & \cdots & \cdots & \cdots & \cdots & y^2b_1 & yb_1 & b_1 \end{bmatrix}, \quad \begin{aligned} a_i &= 1 + y^{2(i-1)}\eta_\tau, \\ b_j &= 1 + y^{2(j-1)}\kappa. \end{aligned} \quad (7.21)$$

The constant  $y$  can perhaps be more conveniently described as the smaller of the roots of the quadratic equation (both of which are positive)

$$y^2 - \left( \frac{\gamma + 1}{\rho} + \rho \right) + 1 = 0. \quad (7.22)$$

It is easy to establish that  $y < 1$  for  $\rho \leq 1$  and for all  $\gamma$ , so that  $y^T \rightarrow 0$  for large sample size  $T$ . The assertion of Proposition 7.3.1 can be verified by matrix multiplication, which we show in Appendix A.16. Also note that for the two special cases of  $\tau = 1$  and  $\tau = \infty$ ,  $\varrho_C = \rho^2, \eta_C = -y^2$  and  $\varrho_U = 0, \eta_U = \kappa$ . Next, a few remarks are in order.

*Remark 7.3.1.1: Identity:* The following identity will be used in several places in this chapter and follows from definition (7.19) after some algebraic manipulations,

$$\gamma y = (\rho - y)(1 - \rho y). \quad (7.23)$$

*Remark 7.3.1.2:* It may be noted that  $\mathbf{A}$  is defined entirely by the diagonal elements  $a_1, a_2, \dots, a_{\lfloor \frac{T}{2} \rfloor}$  (counted from top) and  $b_1, b_2, \dots, b_{\lceil \frac{T}{2} \rceil}$  (counted from bottom). From the expressions for  $a_i$  and  $b_j$  in Equation (7.21), we note that for large  $T$ , both  $a_{\lfloor \frac{T}{2} \rfloor}$  and  $b_{\lceil \frac{T}{2} \rceil}$  converge to the same value, namely 1. That means that it should not matter whether the anti-diagonal elements are expressed in terms of  $a_i$ -s or  $b_j$ -s. In Proposition 7.3.1, we describe the antidiagonal elements in terms of  $b_j$ -s just for the sake of being definitive.

### 7.3.2 Computation of Fisher Information Matrix

We use Proposition 7.3.1 in (7.13) and (7.15) to compute the Fisher Information matrix,

$$\mathbf{J} \approx \frac{1}{\sigma^2} \mathbf{X}'(\mathbf{I} - \nu \mathbf{A})\mathbf{X}, \quad (7.24)$$

which would be inverted later to obtain the CRB-s. We consider the matrices  $\mathbf{X}'\mathbf{X}$  and  $\mathbf{X}'\mathbf{A}\mathbf{X}$  individually before summing them up. We would approximate all the elements of these matrices as polynomials in  $T$  (i.e., the sample size). This will help derive approximations of the CRB that are correct upto the second order for  $\rho \leq 1$ .

We note from (7.6) that  $\mathbf{X}'\mathbf{X}$  is a Hankel (equal skew-diagonal elements) matrix,

$$[\mathbf{X}'\mathbf{X}]_{k,l} = \sum_{t=1}^T t^{k+l}, \quad 0 \leq k, l \leq P. \quad (7.25)$$

It is a well known result, e.g., [71], [78] that summations of the form (7.25) can be written as

$$\begin{aligned} \sum_{t=1}^T t^q &= \sum_{i=0}^q \mathcal{B}_{q,i} T^{q+1-i}, \quad q \geq 0, \quad \text{where} \\ \mathcal{B}_{q,0} &\triangleq \frac{1}{q+1}, \mathcal{B}_{q,1} \triangleq \frac{1}{2}, \mathcal{B}_{q,2} \triangleq \frac{q}{12}, \mathcal{B}_{q,3} \triangleq 0, \text{ etc.} \end{aligned} \quad (7.26)$$

A general form for  $\mathcal{B}_{q,i}$  can be found, for example, on p-1, [32]. However, all the results in this chapter will be established using the first four terms which we have enumerated in (7.26).

Similar polynomial expressions can be obtained for  $\mathbf{X}'\mathbf{A}\mathbf{X}$  and are derived in Appendix A.17. We state the result in Proposition 7.3.2.

**Proposition 7.3.2.**

$$\begin{aligned} [\mathbf{X}'\mathbf{A}\mathbf{X}]_{k,l} &= \sum_{i=0}^{k+l} \mathcal{A}_{k,l,i} T^{k+l+1-i} + \alpha_{k,l}^{(\tau)} \\ &\quad + \mathcal{O}(T^{k+l+1} y^T), \quad 0 \leq k, l \leq P, \end{aligned} \quad (7.27)$$

where some leading constants  $\mathcal{A}_{k,l,i}$  and  $\alpha_{k,l}^{(\tau)}$  are

$$\begin{aligned}\mathcal{A}_{k,l,0} &\triangleq \frac{1+y}{1-y} \frac{1}{k+l+1}, \quad \forall k, l, \\ \mathcal{A}_{k,l,1} &\triangleq \frac{1+2\kappa-2y-y^2}{2(1-y)^2}, \quad k+l \geq 1, \\ \alpha_{0,0}^{(\tau)} &\triangleq \frac{-2y+\eta_\tau+\kappa}{(1-y)^2}, \text{ etc.}\end{aligned}\tag{7.28}$$

The exact form of  $\mathcal{A}_{k,l,i}$  and  $\alpha_{k,l}^{(\tau)}$  are provided in Appendix A.17. From (7.27), it is clear that the Fisher information for calibrated and uncalibrated cases differ only by a constant (dependence on  $\tau$  is explicitly indicated). This means that when sample size is large, estimation accuracy will be similar for both the cases. We will elaborate this point later. Next, we use the result in Proposition 7.3.2 to derive the CRBs for the cases when  $\rho < 1$  and  $\rho = 1$ .

### 7.3.3 Cramér-Rao bounds for $\rho < 1$

Using (7.25), (7.26) and (7.27) in (7.24), we obtain the compact second order expression for the Fisher Information Matrix,

$$\mathbf{J} = T\mathbf{E} \left[ \xi_0 \mathbf{H} + \frac{\xi_1 \mathbf{e}\mathbf{e}' + \xi_2^{(\tau)} \mathbf{f}\mathbf{f}'}{T} + \mathcal{O}\left(\frac{1}{T^2}\right) \right] \mathbf{E}, \tag{7.29}$$

where  $\mathbf{E}, \mathbf{H}$  are  $(P+1) \times (P+1)$  matrices,  $\mathbf{e}, \mathbf{f}$  are  $(P+1)$ -dimensional vectors and  $\xi_0, \xi_1, \xi_2^{(\tau)}$  are constants defined by

$$\begin{aligned}\mathbf{E} &\triangleq \text{diag}\{1, T, \dots, T^P\}, \\ \mathbf{H}_{k,l} &\triangleq \frac{1}{k+l+1}, \text{ for } 0 \leq k, l \leq P. \\ \mathbf{e} &\triangleq [1, 1, \dots, 1]', \quad \mathbf{f} \triangleq [1, 0, \dots, 0]', \\ \xi_0 &\triangleq \frac{1}{\sigma^2} \left[ 1 - \nu \frac{1+y}{1-y} \right], \quad \xi_1 \triangleq \frac{1}{\sigma^2} \left[ \frac{1}{2} - \nu \mathcal{A}_1 \right], \text{ and} \\ \xi_2^{(\tau)} &\triangleq -\xi_1 - \frac{1}{\sigma^2} \nu \alpha_{0,0}^{(\tau)},\end{aligned}\tag{7.30}$$

respectively. The notation  $\mathcal{A}_1$  is actually  $\mathcal{A}_{k,l,1}$  with the subscripts dropped, since  $\mathcal{A}_{k,l,1}$  does not depend on  $k$  and  $l$  (see (7.28)).  $\mathbf{H}$  is the well known Hilbert matrix [71].

The leading constant  $\xi_0$  can be simplified further based on definitions in (7.19) and the identity (7.23),

$$\xi_0 = \frac{1}{\sigma^2} \left[ 1 + \frac{\gamma}{(1-\rho)^2} \right]^{-1}. \quad (7.31)$$

The CRB will be obtained as the inverse of the FIM described in (7.29). Since  $\rho < 1$ , we ensure from (7.31) that  $\xi_0 > 0$ . Therefore, the FIM of the form (7.29) can be inverted assuming  $\xi_0 \mathbf{H}$  as the dominant term. Such an inversion was performed in [71] in the context of polynomial-phase signals. We summarize the analysis in [71] in the form of Lemma 7.3.2.1.

**Lemma 7.3.2.1.** [71]: *Let  $\mathbf{H}$  be the Hilbert matrix and  $\mathbf{e}, \mathbf{f}$  vectors as defined in (7.30). Let  $c_0, c_1, c_2$  be constants such that  $c_0 \neq 0$ . Then, for  $0 \leq p \leq P$ , the diagonal elements are*

$$\begin{aligned} & \left[ \left[ c_0 \mathbf{H} + \frac{1}{T} (c_1 \mathbf{e} \mathbf{e}' + c_2 \mathbf{f} \mathbf{f}') \right]^{-1} \right]_{p,p} = \\ & \frac{K_{P,p}}{c_0} \left[ \frac{1}{2p+1} - \frac{1}{T c_0} (c_1 + c_2 L_{P,p}) + \mathcal{O} \left( \frac{1}{T^2} \right) \right], \end{aligned} \quad (7.32)$$

where  $K_{P,p}$  and  $L_{P,p}$  are defined by

$$\begin{aligned} K_{P,p} & \triangleq \left[ (P+p+1) \binom{P+p}{p} \binom{P}{p} \right]^2, \text{ and} \\ L_{P,p} & \triangleq \left[ \frac{P+1}{p+1} \right]^2, \end{aligned} \quad (7.33)$$

respectively.

The idea behind Lemma 7.3.2.1 is that, for sufficiently large  $T$ , terms of order  $\mathcal{O}(\frac{1}{T^2})$  in (7.32) can be ignored and we obtain approximate closed form expressions for the diagonal elements. The approximation in (7.32) is accurate only for small relative magnitude of the second order term, i.e., small magnitudes of  $\frac{(2p+1)(c_1+c_2 L_{P,p})}{c_0 T}$ . Lemma 7.3.2.1 can be used

now to invert (7.29). The Cramér-Rao bound (diagonal terms) obtained in such a manner are summarized in Proposition 7.3.3.

**Proposition 7.3.3.** *For  $\rho < 1$  and sufficiently large  $T$ , the CRB for estimating  $\theta_p$  is*

$$[\mathbf{V}]_{p,p} \approx \frac{K_{P,p}}{T^{2p+1}\xi_0} \left[ \frac{1}{2p+1} - \frac{1}{T\xi_0} \left( \xi_1 + \xi_2^{(\tau)} L_{P,p} \right) \right], \quad (7.34)$$

for  $0 \leq p \leq P$ .

A few remarks due to Proposition 7.3.3 are in order.

*Remark 7.3.3.1: Asymptotic performance:* As expected, the estimation performance does not depend on transient information ( $\tau$ ) when the number of samples used for inference is sufficiently large. From (7.31) and (7.34), the first order approximation of CRB is

$$[\mathbf{V}]_{p,p} \approx \left[ 1 + \frac{\gamma}{(1-\rho)^2} \right] \frac{\sigma^2 K_{P,p}}{T^{2p+1}(2p+1)}, \quad (7.35)$$

which is true for any value of  $\tau$ .

*Remark 7.3.3.2: Equivalent AWGN noise:* For estimation of polynomial signals, the effect of drift is asymptotically (upto first order) equivalent to scaling the measurement noise by a factor of  $1 + \gamma/(1-\rho)^2$ , where  $\gamma, \rho$  characterizes the drift properties of a sensor.

*Remark 7.3.3.3: Constant signal:*  $x_t = \theta_0, \forall t$ : Simpler and more precise expressions for CRB may be obtained for a constant signal. When  $P = 0$ , we can directly proceed from (7.29). In this case, all of the matrices  $\mathbf{e}\mathbf{e}' = \mathbf{f}\mathbf{f}' = \mathbf{E} = \mathbf{H} = 1$ , and hence we obtain

$$\mathbf{V} \approx \frac{1}{\xi_0 T + \xi_1 + \xi_2^{(\tau)}} = \frac{1}{\xi_0 T - \nu \alpha_{0,0}^{(\tau)} / \sigma^2}, \quad (7.36)$$

where the scalar  $\mathbf{V}$  denotes the Cramér-Rao bound.

*Remark 7.3.3.4: Dependence on  $\tau$ :* In terms of estimation of a signal, intuitively, the transient state  $\tau < \infty$  should be more informative, since the drift starts from a known point, rather than an unknown point. This intuition is corroborated by equation (7.34), from which it is easy to derive that  $[\mathbf{V}]_{p,p}$  increases with  $\tau$  (shown below). In conjunction with



Remark 7.3.3.1, we can conclude that, though calibration always results in better estimates, the relative gain in performance  $([\mathbf{V}_U]_{p,p} - [\mathbf{V}_C]_{p,p})/[\mathbf{V}_U]_{p,p}$  diminishes for higher sample sizes.

*Proof that  $[\mathbf{V}]_{p,p}$  increases with  $\tau$ :* Since  $\xi_0 > 0$  (from (7.31)), it suffices to show that  $\xi_2^{(\tau)}$  decreases with  $\tau$  (see (7.34)), or equivalently,  $\alpha_{0,0}^{(\tau)}$  increases with  $\tau$  (see (7.30), note that  $\nu > 0$ ), which is further equivalent to showing that  $\eta_\tau$  increases with  $\tau$  (see (7.28)). The last condition is verified from (7.20), since  $\rho_\tau$  is a decreasing function of  $\tau$  and  $0 < \rho, y < 1$ .

*Remark 7.3.3.5: Approximation region:* The CRB as expressed by (7.34) is accurate only for cases when the relative magnitudes of the second order terms,

$$\epsilon_p = \frac{(2p+1)(\xi_1 + \xi_2^{(\tau)} L_{P,p})}{\xi_0 T}, \quad 0 \leq p \leq P, \quad (7.37)$$

are small. This condition helps specify the operating values of  $\sigma^2, \gamma, \rho, P$  and  $T$  for which the closed-form CRB (7.34) can be used for performance analysis.

### 7.3.4 Cramér-Rao bounds for $\rho = 1$

For the case when the AR parameter  $\rho = 1$ , the drift phenomenon is a non-stationary even in the limit of  $\tau \rightarrow \infty$ , since it is a random walk (e.g., [79]) where the bias builds up with time and is unbounded. Hence, the CRB for the uncalibrated scenario is infinity and we consider only the transient case  $\tau < \infty$  here.

For  $\rho = 1$ , the leading term of the Fisher information matrix  $\mathbf{J}$  (7.29) vanishes since (7.31) implies that  $\xi_0 = 0$ . Hence, terms of order  $\mathcal{O}(T^{-2})$  need to be considered. In particular, using definitions (7.19), (7.28) and (7.30), the constants can further be simplified as

$$\begin{aligned} \xi_0 &= 0, \quad \xi_1 = 0, \quad \varrho_\tau = \frac{1}{\tau}, \text{ and} \\ \xi_2^{(\tau)} &= \frac{1}{\sigma^2} \left[ \frac{2}{\tilde{\gamma} + 1} \left( 1 + \frac{2\tau}{\tilde{\gamma} - 1} \right) \right]^{-1}, \quad \tilde{\gamma} \triangleq \sqrt{1 + 4\gamma^{-1}}, \end{aligned} \quad (7.38)$$

and we subsequently note (see Appendix A.18) that the FIM is of the structure depicted in equation (7.39),

$$\mathbf{J} = \mathbf{E} \begin{bmatrix} \xi_2^{(\tau)} & \frac{1}{T} \xi_3^{(\tau)} \mathbf{f}'_P + \mathcal{O}\left(\frac{1}{T^2}\right) \\ * & \frac{1}{T} \mathbf{D} \left[ \xi_4 \mathbf{H}_P + \frac{1}{T} \left( \xi_5 \mathbf{e}_P \mathbf{e}'_P + \xi_6^{(\tau)} \mathbf{f}_P \mathbf{f}'_P \right) + \mathcal{O}\left(\frac{1}{T^2}\right) \right] \mathbf{D} \end{bmatrix} \mathbf{E}, \quad (7.39)$$

where the constants  $\xi_3^{(\tau)}, \xi_4, \xi_5, \xi_6^{(\tau)}$ , vectors  $\mathbf{e}_P, \mathbf{f}_P$  and matrices  $\mathbf{H}_P, \mathbf{D}$  are defined by

$$\begin{aligned} \xi_3^{(\tau)} &\triangleq -\frac{1}{\sigma^2} \nu \alpha_{1,0}^{(\tau)}, \quad \xi_4 \triangleq \frac{1}{\sigma^2 \gamma}, \\ \xi_5 &\triangleq -\frac{1}{\sigma^2} \frac{y^2}{(1-y)^3}, \quad \xi_6^{(\tau)} \triangleq -\xi_5 - \frac{1}{\sigma^2} \nu \alpha_{1,1}^{(\tau)}, \\ \mathbf{f}_P &\triangleq [1, 0, \dots, 0]', \quad \mathbf{e}_P \triangleq [1, 1, \dots, 1]', \\ \mathbf{D} &\triangleq \text{diag}\{1, 2, \dots, P\}, \text{ and} \\ [H_P]_{k,l} &\triangleq \frac{1}{k+l+1}, \text{ for } 0 \leq k, l \leq P-1. \end{aligned} \quad (7.40)$$

Here  $\mathbf{e}_P, \mathbf{f}_P$  and  $\mathbf{H}_P$  are the  $P$ -dimensional equivalents of  $\mathbf{e}, \mathbf{f}$  and  $\mathbf{H}$  (defined in (7.30)) respectively. The CRB is obtained from the FIM (7.39) by using block-inversion and Lemma 7.3.2.1 (see Appendix A.18). The results are summarized in Proposition 7.3.4.

**Proposition 7.3.4.** *For  $\rho = 1$  and sufficiently large  $T$ , the CRB for estimating  $\theta_p$ ,  $[\mathbf{V}]_{p,p}$  is equal to*

$$[\mathbf{V}]_{p,p} \approx \begin{cases} \frac{1}{\xi_2^{(\tau)}} \left[ 1 + \frac{1}{T \xi_2^{(\tau)}} \frac{P^2 (\xi_3^{(\tau)})^2}{\xi_4} \right], & p = 0, \\ \frac{K_{P-1,p-1}}{T^{2p-1} p^2 \xi_4} \left[ \frac{1}{2p-1} - \frac{1}{T \xi_4} \left( \xi_5 + \left( \xi_6^{(\tau)} - \frac{(\xi_3^{(\tau)})^2}{\xi_2^{(\tau)}} \right) \frac{P^2}{p^2} \right) \right], & 1 \leq p \leq P. \end{cases} \quad (7.41)$$

A few remarks due to Proposition 7.3.4 are noted below.

*Remark 7.3.4.1: Constant Signal:*  $x_t = \theta_0, \forall t$ : For a random walk drift, a constant signal can only be estimated inconsistently with asymptotic variance

$$\mathbf{V} \approx \frac{1}{\xi_2^{(\tau)}} = \frac{2\sigma^2}{\tilde{\gamma} + 1} \left( 1 + \frac{2\tau}{\tilde{\gamma} - 1} \right). \quad (7.42)$$

It is explicitly seen from (7.42) that the variance increases with white noise variance  $\sigma^2$ , the drift strength  $\gamma$  (note that  $\tilde{\gamma}$  decreases with  $\gamma$ , see (7.38)), and the time since last calibration,  $\tau$ . This phenomenon of inconsistent estimation can be intuitively understood as follows. Even without the white noise component in (7.3), the observation never captures independent readings of  $\theta_0$ . Rather, the samples are

$$\begin{aligned} z_1 &= \theta_0 + d_\tau + w_1, \\ z_2 &= \theta_0 + d_{\tau+1} + w_2 = (\theta_0 + d_\tau) + \delta_{\tau+1} + w_2, \text{ etc.}, \end{aligned} \quad (7.43)$$

which means that  $\theta_0 + d_\tau$  appears together in all subsequent observations, thereby making  $\theta_0$  indistinguishable from  $d_\tau$ . Since  $d_\tau$  has a finite variance,  $\theta_0$  can only be estimated inconsistently.

*Remark 7.3.4.2: Time Features:* For other parameters whose effect on the signal vary with time, i.e.,  $\theta_p$  for  $p \geq 1$ , the CRB is (to first order) equivalent to the case of estimating the derivative of the signal in drift-only noise. We note that the derivative of the signal is  $x'(t) = \sum_{p=1}^P p\theta_p t^{p-1}$  and the forward difference (see (7.3)),

$$\begin{aligned} z_{t+1} - z_t &= x_{t+1} - x_t + d_{t+1} - d_t + w_{t+1} - w_t \\ &= x'(t)(\Delta T) + \delta_t + w_{t+1} - w_t, \end{aligned} \quad (7.44)$$

which is an estimator of the derivative, contains independent drift innovations  $\delta_t$  with variance  $\gamma\sigma^2$ . Hence in Proposition 7.3.4, the variance of estimating  $\theta_p$  is scaled down by a factor of  $p^2$  compared to the equivalent case with white noise  $\gamma\sigma^2$ .

*Remark 7.3.4.3: Approximation region:* The CRB as expressed by (7.41) is accurate only for cases when the relative magnitude of the second order terms,

$$\epsilon_p = \begin{cases} \frac{1}{T\xi_2^{(\tau)}} \frac{P^2 (\xi_3^{(\tau)})^2}{\xi_4}, & p = 0, \\ \frac{2p-1}{T\xi_4} \left( \xi_5 + \left( \xi_6^{(\tau)} - \frac{(\xi_3^{(\tau)})^2}{\xi_2^{(\tau)}} \right) \frac{P^2}{p^2} \right), & 1 \leq p \leq P. \end{cases} \quad (7.45)$$

are small. This condition helps specify the operating values of  $\sigma^2, \gamma, P$  and  $T$  for which the closed-form CRB (7.41) can be used for performance analysis.

### 7.3.5 Numerical results - Maximum relative error

In the remainder of this section, we demonstrate the accuracy of the closed-form performance approximations using representative examples. We shall consider the approximations expressed in both the forms of Fisher Information matrix (Equations (7.29) and (7.39), with  $\mathcal{O}(\frac{1}{T^2})$  terms truncated) and Cramér-Rao bounds (Equations (7.34) and (7.41)).

We use the metric called Maximum Relative Error (MRE) which was also used in [71]. This involves computing both the exact (7.16) and approximate CRB-s. The relative deviations of the approximations are then calculated for all  $0 \leq p \leq P$  and the largest one is called the MRE. We denote

1.  $[\mathbf{V}^{\text{TH}}]_{p,p}$  as the theoretical CRB as in (7.16), and
2.  $[\mathbf{V}^{\text{AP}}]_{p,p}$  as the approximate CRB derived either through
  - *FIM Approx*: inversion of the intermediate FIM, i.e.,  $[\mathbf{V}^{\text{AP}}]_{p,p} = [\mathbf{J}^{-1}]_{p,p}$ , in Equations (7.29) and (7.39) with  $\mathcal{O}(\frac{1}{T^2})$  terms truncated, or
  - *CRB Approx*: final Cramér-Rao bounds in Equations (7.34) and (7.41).

Then the maximum relative error is defined by

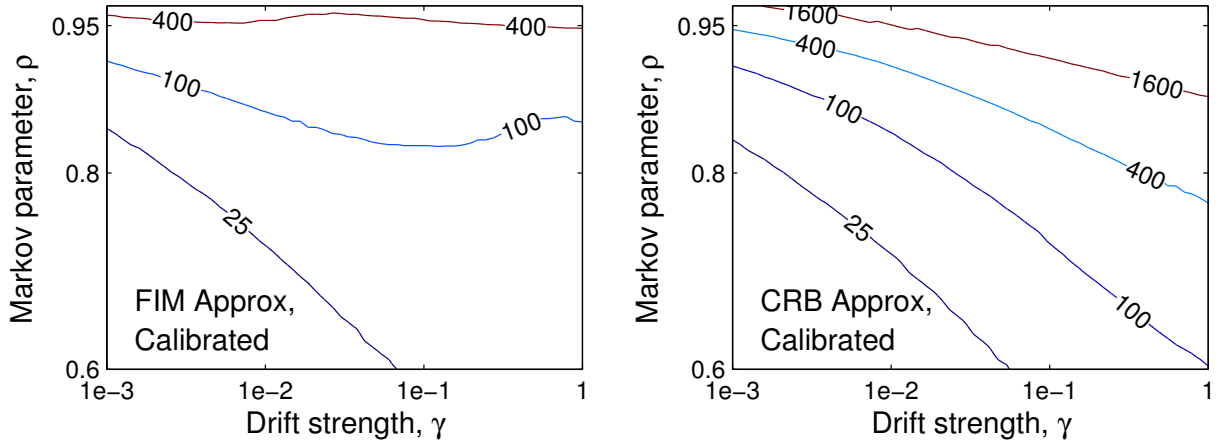
$$\text{MRE} = \max_{p=0,1,\dots,P} \frac{|[\mathbf{V}^{\text{TH}}]_{p,p} - [\mathbf{V}^{\text{AP}}]_{p,p}|}{[\mathbf{V}^{\text{TH}}]_{p,p}}. \quad (7.46)$$

In short, the MRE summarizes the approximation error over all components of the parameter vector.

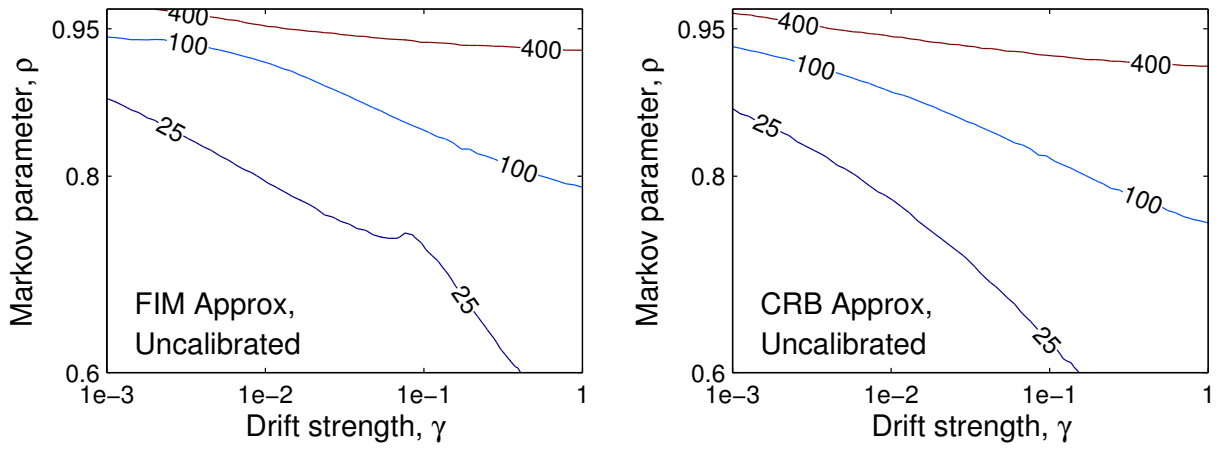
As an example, we consider the estimation of a cubic-polynomial signal, i.e.,  $P = 2$ . Since our performance bounds are asymptotically accurate, it is expected that the MRE will decrease with increasing sample size  $T$ . In Figure 7.2, we display the sample size  $T_\epsilon$  required for 95% accuracy, or in other words,  $\text{MRE} < \epsilon = 0.05$ . Since MRE is a ratio of variances, it does not depend on the measurement noise variance  $\sigma^2$ . While considering the  $\rho < 1$  case, we have displayed the wide parameter region  $\gamma \in [10^{-3}, 1]$ ,  $\rho \in [0.6, 0.97]$  in Figures 7.2(a) and 7.2(b) - which demonstrate the uncalibrated ( $\tau \rightarrow \infty$ ) and calibrated ( $\tau = 1$ ) scenarios respectively. The  $\rho = 1, \tau = 1$  case is demonstrated in 7.2(c), where we have displayed the parameter region  $\gamma \in [10^{-3}, 1]$ .

Figure 7.2 provides useful guidance on the applicability and limitations of the performance bounds derived in this chapter. Firstly, the performance bounds are found to be reasonably accurate over a wide range of possible parameter values and for a moderate number (tens and hundreds) of samples. For accurate performance prediction in stationary drift (Figures 7.2(a) and 7.2(b)), higher values of drift-autocorrelation ( $\rho$ ) and drift-strength ( $\gamma$ ) generally requires larger observation durations ( $T_\epsilon$ ). This is predicted by the approximation-region condition in (7.37), since the denominator  $\xi_0$  (see (7.31)) is inversely proportional to  $\rho$  and  $\gamma$ . For the random-walk drift scenario (Figure 7.2(c)), higher value of drift-strength  $\gamma$  requires smaller observation durations ( $T_\epsilon$ ) for accurate performance predictions. This too, can be explained from approximation-region condition in (7.45), where, with the aid of definitions (7.38) and (7.40), it can be established that  $\epsilon_p = \mathcal{O}(\gamma^{-1/2})$  for small  $\gamma$ . This also partially explains the log-linear relation between  $T_\epsilon$  and  $\gamma$  in Figure 7.2(c).

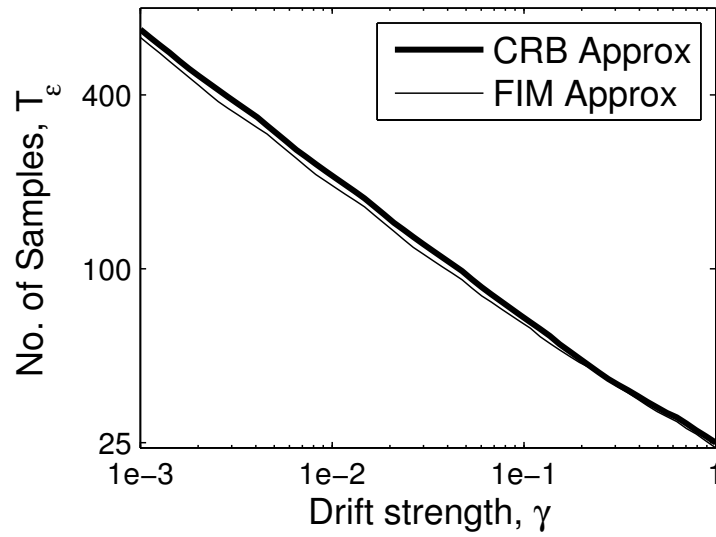
We have so far only considered the estimation performance of a single sensor that is able to reliably communicate its observations to the sink without any distortion. In Sections 7.4 and 7.5, we discuss extensions to the multiple sensor framework and bandwidth limited networks.



(a) Stationary drift ( $\rho < 1$ ) and calibrated sensors.



(b) Stationary drift ( $\rho < 1$ ) and uncalibrated sensors.



(c) Non-stationary drift ( $\rho = 1$ ).

Figure 7.2: Sample size  $T_\epsilon$  required for 95% accurate performance prediction

## 7.4 Multiple sensors

In this section, we consider the application of the results in Section 7.3 to multiple sensors with different noise and drift parameters  $\{\sigma_n^2, \gamma_n, \rho_n\}$ . Since the sensor noise and drift are independent across the sensors (Assumption 3), the Fisher Information for  $\boldsymbol{\theta}$  is equal to the sum of individual FIM-s (see equation (7.15)),

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 + \cdots + \mathbf{J}_N. \quad (7.47)$$

Hence the expressions for FIM (given by (7.29) and (7.39)) and subsequently CRB (given by (7.34) and (7.41)) can be extended (in a rather straightforward manner) with the following change in definitions,

$$\xi_{\#, \text{eff}} \triangleq \sum_{n=1}^N \xi_{\#}(\sigma_n^2, \gamma_n, \rho_n), \text{ for } [\#] = 0, 1, 2, 3, 4, \quad (7.48)$$

where  $\xi_{\#}(\sigma^2, \gamma, \rho)$  can be thought of as a function of its arguments as defined in (7.30) and (7.40) for  $[\#] = 0, 1, 2, 3, 4$ , and  $\xi_{\#, \text{eff}}$  denote the effective value of the constant.

As an illustration of this extension, we consider an example where noise and drift parameters are uniformly (randomly) distributed over a given range, say,

$$\rho_n \in [\rho_l, \rho_u], \sigma_n^2 \in [\sigma_l^2, \sigma_u^2] \text{ and } \gamma_n \in [\gamma_l, \gamma_u]. \quad (7.49)$$

When the number of sensors  $N$  is large, (7.48) can be further approximated by substituting the summations by integrals,

$$\begin{aligned} \xi_{\#, \text{eff}} &= N \mathbb{E} [\xi_{\#}] \\ &= \frac{N}{\Delta \rho \Delta \sigma^2 \Delta \gamma} \int_{\sigma_l^2}^{\sigma_u^2} \int_{\gamma_l}^{\gamma_u} \int_{\rho_l}^{\rho_u} \xi_{\#}(\sigma^2, \gamma, \rho) \, d\rho \, d\gamma \, d\sigma^2. \end{aligned} \quad (7.50)$$

We would refer to the CRB derived using the constants  $\xi_{\#, \text{eff}}$  (arising out of integration) as the *Average-CRB*. We expect the Average-CRB to be an effective indicator of system

performance for large number of sensors  $N$ . We show some numerical results below to demonstrate the accuracy of Average-CRB.

### 7.4.1 Numerical results

We demonstrate the results for both  $\rho < 1$  and  $\rho = 1$  cases. The simulation setup is described below:

- We assume a linear signal, i.e.,  $P = 1$  and assume all the sensors are observing the same phenomenon. The parameters to be estimated are the constant term ( $\theta_0$ ) and the slope term ( $\theta_1$ ).
- We consider multi-sensor systems with the number of sensors,  $N$ , starting from 25 and going upto 100.
- For the  $\rho < 1$  scenario, for each sensor, the drift parameters are randomly selected by choosing the parameters from the range,

$$\rho_n \in [.85, .95], \sigma_n^2 \in [72, 288] \text{ and } \gamma_n \in [.6, 2.4], \quad (7.51)$$

which is close to an estimated spectrum in [36].

- Within the  $\rho < 1$  scenario, we simulate both calibrated ( $\tau_n = 1, \forall n$ ) and uncalibrated ( $\tau_n = \infty, \forall n$ ) cases. In one simulation, we assume that all  $N$  sensors are calibrated while in another simulation we assume that none of them are calibrated.
- For the  $\rho = 1$  scenario, the drift parameters are randomly selected by choosing the parameters from the range,

$$\sigma_n^2 \in [2, 12] \text{ and } \gamma_n \in [.05, .3]. \quad (7.52)$$



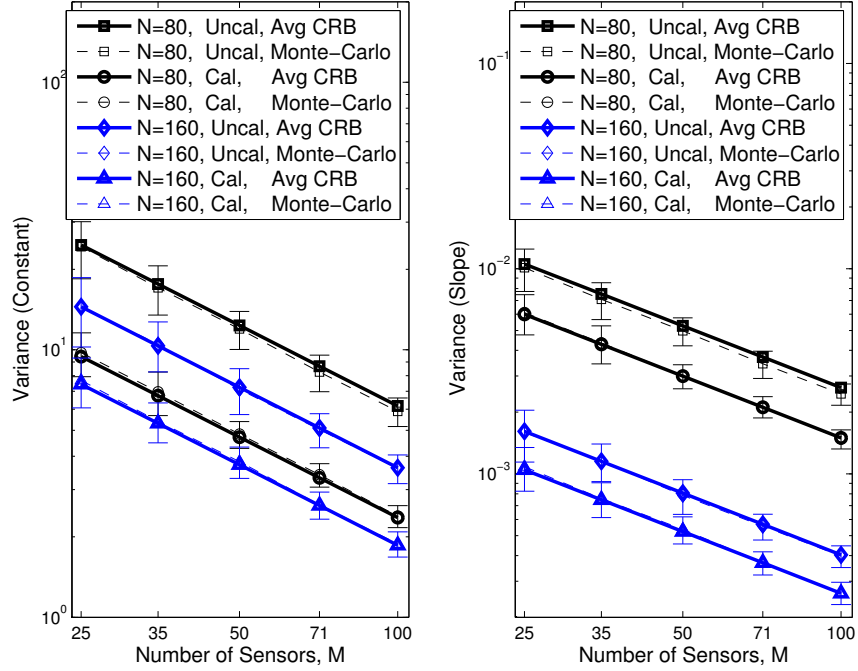


Figure 7.3: Average performance for multi-sensor systems for  $\rho \in [0.85, 0.95]$ .

- For each realization of an  $N$ -sensor network, ML-estimation of the constant and slope parameters were performed and the error variances were averaged over  $10^5$  Monte-Carlo trials (realizing the measurement noise and sensor drift).
- Several ( $10^3$ ) Monte-carlo trials (realizing  $N$ -sensor networks with  $\sigma^2, \rho_n, \gamma_n$  chosen from above parameter range) are performed and the average and 95% confidence interval of the error variances are observed.
- We repeat the experiments for sample sizes  $T = 80$  and  $160$ . These samples sizes were chosen to ensure moderate computational (Monte-Carlo) effort.

Since we have only a small number of samples, we have used the expressions for Fisher Information Matrix (Equations (7.29) and (7.39)) to predict the estimation error variance of the constant ( $\theta_0$ ) and slope ( $\theta_1$ ) portion of a signal. The results are displayed in Figures 7.3 and 7.4, of which we make some comments below.

Firstly, an approximation to the system performance using (7.50) is seen to be fairly accurate, as depicted in Figures 7.3 and 7.4. In all cases, the variance is inversely pro-

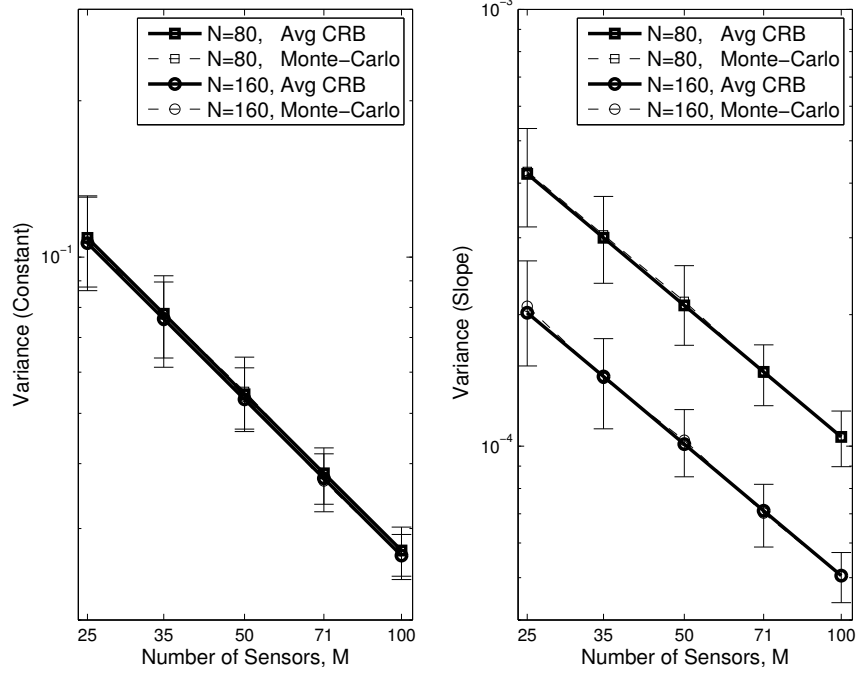


Figure 7.4: Average performance for multi-sensor systems for  $\rho = 1$ .

portional to the number of sensors, as depicted by the log-linearity of all the curves. The (average) performance prediction improves (error-bars becomes shorter) for higher sample sizes, since the summation (7.48) is better approximated by the integral (7.50) for large  $N$ . This means that just by knowing the range of the parameter values and the number of sensors, we can have a good understanding about the estimation performance of the entire sensor network. Secondly, from Figure 7.3, we note the intuitive phenomenon that the performance using calibrated sensors is better than that using uncalibrated sensors. Also, though for  $T = 80$ , the performance gap is quite large, the gap between calibrated and uncalibrated cases narrows down for  $T = 160$ . This corroborates Remark 7.3.3.4, where we noted that the relative performance gain diminishes for higher sample sizes. For the  $\rho = 1$  case, from Figure 7.4, we note that increasing the sample size does not help in estimation of the constant portion of the signal. This is due the inconsistency property of the non-stationary drift model, as described in Remark 7.3.4.1.

## 7.5 Quantized observations

As another application of the results in Section 7.3, we consider the performance characterization of a sensor network where resource (bandwidth, power, etc.) available for communication is limited [73], [58]. Bandwidth constraints preclude the reliable communication of full-precision observations from the individual sensor nodes to the sink. In such situations, the observations must be compressed and digitized prior to transmission [31]. In this section, we digitize each observation by using uniform quantization, as will be discussed shortly. Since the reliability of the transmission of digitized observations may be further enhanced through the use of efficient channel coding (error-control codes), it is reasonable to assume that once the quantization is performed, there is no further deterioration of the observation-quality [73]. In other words, the quantized observations are assumed to be available error-free at the sink. The schematic diagram of such a system is depicted in Figure 7.1. *Our goal in this section is to predict the estimation performance of the sensor network composed of sensors with drift, when only quantized observations are available for inference at the sink.*

We use uniform quantization [51], in which each of the observations  $z_{t,n}$  are quantized uniformly with  $l$ -bits. Assuming that the observations are bounded  $z_{t,n} \in [U_0, U_1]$  and the quantization thresholds  $a_j \in [U_0, U_1]$ ,  $j = 0, \dots, 2^l$  are uniformly spaced such that  $a_{j+1} - a_j = (U_1 - U_0)/(2^l - 1) \triangleq \Delta$ , the expected distortion due to the quantization process is

$$\sigma_Q^2 \triangleq \mathbb{E}[(\tilde{z}_{t,n} - z_{t,n})^2] = \frac{\Delta^2}{12}, \quad \forall t, n. \quad (7.53)$$

Since the quantization noise  $\tilde{z}_{t,n} - z_{t,n}$ , in general, is neither independent across time nor Gaussian distributed, the Maximum-Likelihood estimator is difficult to design. However, the quasi-ML estimator is easier to implement, which is designed on the assumption that the quantization noise is i.i.d. Gaussian with variance  $\sigma_Q^2$ . We will use the quasi-ML estimator, described below, to perform inference using quantized observations.

Effectively, the quasi-ML estimator assumes that the noise at individual sensor nodes has an added component  $\sigma_Q^2 \mathbf{I}$ , with the total covariance being (compare with (7.13))

$$\begin{aligned}\tilde{\Sigma}_{*,n} &\triangleq \sigma_Q^2 \mathbf{I} + \Sigma_{*,n} \\ &= \sigma_Q^2 \mathbf{I} + \sigma_n^2 (\mathbf{I} + \gamma_n \mathbf{R}_{*,n}) \\ &\triangleq \tilde{\sigma}_n^2 (\mathbf{I} + \tilde{\gamma}_n \mathbf{R}_{*,n}),\end{aligned}\tag{7.54}$$

where  $\tilde{\sigma}_n, \tilde{\gamma}_n$  satisfy

$$\tilde{\sigma}_n^2 \triangleq \sigma_n^2 + \sigma_Q^2, \text{ and } \tilde{\gamma}_n \triangleq \frac{\gamma_n}{1 + \sigma_Q^2/\sigma_n^2}.\tag{7.55}$$

Accordingly, the quasi-ML estimator is defined as (compare to (7.14)),

$$\begin{aligned}\tilde{\boldsymbol{\theta}}^{\text{ML}} &= \tilde{J}^{-1} \mathbf{X}' \sum_{n=1}^N \tilde{\Sigma}_n^{-1} \tilde{\mathbf{z}}_n, \text{ where} \\ \tilde{J} &\triangleq \mathbf{X}' \left( \sum_{n=1}^N \tilde{\Sigma}_n^{-1} \right) \mathbf{X}.\end{aligned}\tag{7.56}$$

Note that, in the limiting case when quantization errors are small (equivalently, large  $l$ ), we have  $\tilde{\sigma}_n^2 \rightarrow \sigma_n^2, \tilde{\gamma}_n \rightarrow \gamma_n$  and the quasi-ML-estimator is identical to the ML-estimator.

We would refer to the CRB derived using modified noise parameters  $\{\tilde{\sigma}_n^2, \tilde{\gamma}_n, \rho_n\}$  as the *Modified-CRB*. We expect the Modified-CRB to be an effective indicator of system performance for moderate to large number of quantization levels  $2^l$ . Similar to Section 7.4, with certain change in definitions, the expressions for FIM (given by (7.29) and (7.39)) and subsequently CRB (given by (7.34) and (7.41)) can be extended to obtain the Modified-CRB,

$$\xi_{\#, \text{eff}} \triangleq \sum_{n=1}^N \xi_{\#}(\rho_n, \tilde{\sigma}_n^2, \tilde{\gamma}_n),\tag{7.57}$$

where  $\xi_{\#}(\rho, \sigma^2, \gamma)$  can be thought of as a function of its arguments as defined in (7.30) and (7.40) for  $[\#] = 0, 1, 2, 3, 4$ . We show some numerical results below to corroborate the effectiveness of Modified-CRB as an efficient performance predictor.

### 7.5.1 Numerical results

The simulation setup is described as follows.

- A sample size of  $T = 400$  and network size of  $N = 25$  nodes was considered.
- The noise and drift parameters of the sensor nodes are chosen by uniformly spacing them in the range,

$$72 = \sigma_1^2 < \sigma_2^2 < \cdots < \sigma_N^2 = 288,$$

$$0.6 = \gamma_1 < \gamma_2 < \cdots < \gamma_N = 2.4, \text{ and}$$

$$0.85 = \rho_1 < \rho_2 < \cdots < \rho_N = 0.95.$$

Both calibrated ( $\tau_n = 1, \forall n$ ) and uncalibrated ( $\tau_n = \infty, \forall n$ ) cases were considered.

- We assume a linear signal, i.e.,  $P = 1$  with the constant term ( $\theta_0 = 400$ ) and the slope term  $\theta_1 = 0.9$ . The range of the observations to be quantized was chosen to be  $U_0 = 0, U_1 = 1200$ , beyond which the observations were clipped. The range was deliberately chosen large so that clipping (which is another source of distortion which we have not modeled) does not occur frequently.
- Uniform quantization is performed using  $l = 5, 6, \dots, 9$  bits per observation.
- Quasi-ML estimation of the constant and slope parameters was performed and the error variances were averaged over  $10^8$  Monte-Carlo trials (realizing the measurement noise and sensor drift). The 95% confidence interval of the error variances were also observed.

The results of Monte-Carlo simulations are compared to theoretical predictions from Modified-CRB and displayed in Figure 7.5. The full-precision CRB is also displayed in the figure (labeled w/o Quant), marking the convergence of Modified-CRB in the large- $l$  regime. The actual estimation variance of both the constant and slope parameters (of the linear signal) seem to agree, with reasonable accuracy, to the Modified-CRB.

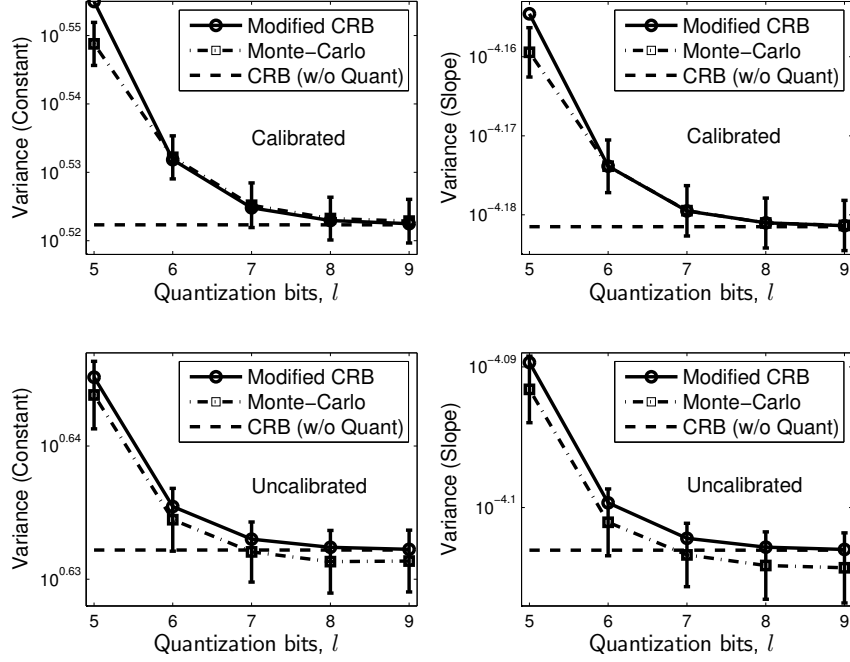


Figure 7.5: Cramér-Rao bounds with quantized observations.

## 7.6 Summary

In this chapter, we have derived approximate bounds for the estimation accuracy of polynomial signals using sensors that exhibit drift in addition to having measurement errors. This is important since sensor drift, or loss of calibration with time, is a major problem in many applications. The theoretical closed form expressions are validated through numerical results. As future work, it may be worthwhile to analyze the performance for other signal models, e.g., stochastic models rather than deterministic. Additionally, it may be interesting to investigate the impact of spatial correlation (i.e, relaxing Assumption 3) on estimation performance. More bit-efficient quantization schemes (other than uniform quantization) that optimally allocate bandwidth to each of the sensors is also another topic of interest. Finally, in-network inference where only the summary of estimated parameters are communicated to the sink, rather than entire observations, will be another framework that one might consider.

# Chapter 8

## Conclusions and suggestions for future work

We have explored the problem of distributed parameter estimation in resource constrained wireless sensor networks. This problem is significantly more difficult compared to centralized estimation because of the complexities involved in choosing the *optimum* compression/transformation scheme for communication with the fusion center. Adding to the complexity is the fact that the sensor nodes may have access to the correlated observations from neighboring nodes, using which the sensors can potentially improve the *overall effectiveness* of the information it relays to the fusion center using finite energy resources. Determining the optimal *collaboration* strategies in such scenarios is a non-trivial problem. Owing to these challenges, the understanding of the performance of distributed and collaborative estimation problems is often coupled to particular kinds of observation (signal and noise) characteristics. One particular observation characteristic that lends tractability to many estimation problems is the assumption of time-independent measurement noise. However, in some practical applications (e.g., gas sensing applications), the sensors usually exhibit baseline *drift*, that leads to time-correlated measurement noise. Analysis of such scenarios is extremely difficult. In this thesis, we addressed some of the aforementioned challenges and thereby improved our understanding of distributed parameter estimation problems. Specifically, our contributions are listed below.

In Chapter 3, we studied distributed parameter estimation using identical binary quantizers. The problem involved finding an optimum probabilistic mapping from the observation space to the binary symbol space. We made the analysis tractable by defining and focusing our attention to the subset of antisymmetric quantizers. Among other theoretical results, we have specified a broad family of distributions for which the commonly-used threshold quantizer is optimal. Aided with some theoretical results, we formulated a numerical optimization problem to obtain the minimax-CRB quantizer within the antisymmetric and piecewise-linear class.

In Chapter 4, we studied the collaborative parameter estimation problem in a quantization based framework. We addressed this problem by using an earlier result on optimal bit-allocation that was derived based on the assumption of spatially independent measurement noise. We used spatial collaboration as a preprocessing tool that maximally whitens the noise components, i.e., the covariance matrix of the transformed noise is as close to a diagonal matrix as possible. Such a whitening transformation allowed us to improve the efficiency of aforementioned bit-allocation scheme, as evidenced by numerical simulations.

In Chapter 5, we studied the collaborative parameter estimation problem in an analog-forwarding based framework. We restricted our attention to coherent multiple access channels. For the scenario when the collaborative topology is fixed and collaboration is cost-free, we obtained the optimal (cumulative) power-distortion tradeoff in closed-form by solving a quadratically-constrained quadratic-programming problem. With individual power constraints, we showed that the semidefinite relaxation technique can be used to obtain precisely optimal numerical results. Several special cases were presented as examples to highlight the problem conditions for which collaboration is particularly effective. Through the use of both theoretical and numerical results, we established that collaboration helps to substantially reduce the distortion of the estimated parameter at the fusion center, especially in low local-SNR scenario.

In Chapter 6, we built on the results of Chapter 5 and investigated two key aspects. First, we provided an asymptotic analysis of the single-snapshot estimation problem when



the collaboration topology is only partially connected. We achieved this by obtaining the solutions for a family of structured networks and then using those solutions to approximately predict the performance of more sophisticated networks using geometric arguments. Second, we extended the problem formulation towards the estimation of a time varying signal. In particular, we derived the instantaneous, average and worst case performance metrics when the signal is modeled as a Gaussian random process with exponential covariance. Both these aspects were investigated under the assumption of spatial and temporal independence among the measurement and channel noise samples.

In Chapter 7, we analyzed how time-correlated measurement noise affects the distributed parameter estimation problem. We have derived approximate Cramér-Rao bounds for the estimation accuracy when the inference has to be performed using sensors that exhibit drift. We assumed a polynomial time-series as the representative signal and an autoregressive process model for the drift. For stationary drift, we showed that the first-order effect of drift is asymptotically equivalent to scaling the measurement noise by an appropriate factor. When the drift is non-stationary, we showed that the constant part of a signal can only be estimated inconsistently (non-zero asymptotic variance).

To summarize, the main contributions of the work reported in this dissertation are as follows.

- Designed optimal probabilistic antisymmetric quantizers for distributed estimation
- Introduced the spatial whitening transformation as a tool that enables efficient bit-allocation in distributed estimation
- Derived the optimal strategy for collaboration when analog-forwarding is performed through a coherent multiple access channel
- Analyzed the collaborative estimation problem for the Ornstein-Uhlenbeck process
- Analyzed the distributed estimation problem for sensors with baseline drift

Despite our contributions in this thesis, the understanding of distributed parameter estimation problem remains incomplete. However, the line of questioning, the techniques

used and the results achieved in this thesis provide new tools for further exploration on this topic. Below are some examples of well-formulated research questions that follow directly from this thesis.

- *Identical binary quantizer design:* We know that in some cases, the threshold quantizer is also the optimal antisymmetric quantizer, assuming that the binary observations are correctly received by the fusion center. What happens when the channel is noisy, as in the case of a binary symmetric channel? What is the noise variance of the channel below which the threshold quantizer is also the optimal antisymmetric quantizer?
- *Spatial whitening:* Through the bit-allocation example in Chapter 4, we have seen that spatial whitening transformations can lead to vastly efficient usage of network resources. Though the concept of whitening is intuitive, it is not trivial to actually show that log-determinant divergence based and adjacency-restricted spatial whitening always improves the efficiency of resource allocation. Theoretical studies on this line can lead to potentially interesting results. For example, can we find the class of problems for which spatial whitening will certainly benefit resource allocation?
- *Analog-forwarding based collaborative estimation:* Since this area was largely unexplored prior to this thesis, several extensions may lead to tractable formulation and interesting observations. One may ask, what is the optimal collaboration strategy when
  - the parameter to be estimated is a vector with possibly correlated elements?
  - the channel to the fusion center is an orthogonal multiple access channel?
  - the act of collaboration incurs some loss of information?

Regarding the estimation of time-varying processes, we have shown that when an Ornstein-Uhlenbeck process is to be estimated in the presence of time-independent measurement noise, the observations should be sampled as frequently as possible. How does this strategy change when the noise is time-correlated?

- *Inference using sensors with drift:* We know that while estimating a deterministic signal, the effect of drift is equivalent to scaling the white noise component by an appropriate multiplicative factor. What is the multiplicative factor, if any, when
  - the signal of interest is stochastic?
  - the drift process is modeled as a general autoregressive moving average (ARMA) process?

# Appendix A

## Appendix - Proofs of various results

### A.1 Proof of Lemma 3.2.5

Let  $\tilde{g}(\theta)$  be the conditional probability corresponding to  $\tilde{\gamma}(x)$ . We need to show that  $\tilde{\gamma}(x)$  dominates  $\gamma(x)$  for the (half-range)  $\theta \in [0, 1)$ . Refer to the expression of  $I(\theta)$  in (3.9). It suffices to show that the numerator and denominator terms satisfy, for  $\theta \in [0, 1)$ , the inequalities (N)  $(g')^2 \leq (\tilde{g}')^2$  and (D)  $g(1 - g) \geq \tilde{g}(1 - \tilde{g})$ . Since admissibility (Definition 3.2.1) implies  $g' > 0$  and  $g \geq 1/2$  for  $\theta \in [0, 1)$ , it suffices to show that (N1)  $g' \leq \tilde{g}'$  and (D1)  $g \leq \tilde{g}$ . From (3.17), we obtain

$$g - \tilde{g} = \int_{-\infty}^{-1} \gamma(x)(f(x - \theta) - f(x + \theta)) dx. \quad (\text{A.1})$$

With unimodality of  $f(w)$  implying  $f(x - \theta) \leq f(x) \leq f(x + \theta)$ , for  $x \in (-\infty, -1)$ ,  $\theta \in [0, 1)$  and  $\gamma(x)$  being positive by definition, (D1) is established from (A.1). Since  $f'(w) > 0$  in  $(-\infty, 0)$  (see Definition 3.2.4), we can interchange the order of integration and derivative in (A.1), to obtain,

$$g' - \tilde{g}' = - \int_{-\infty}^{-1} \gamma(x)(f'(x - \theta) + f'(x + \theta)) dx \leq 0,$$

thereby establishing (N1).

## A.2 Proof of Theorem 3.2.6

Note that conditions for Lemma 3.2.5 are satisfied, hence it suffices to show that the Threshold Quantizer dominates any admissible antisymmetric unit-support quantizer  $\gamma(x)$ . Refer to expression of  $I(\theta)$  in (3.9). It suffices to show that the numerator and denominator terms satisfy, for  $\theta \in [0, 1)$ , the inequalities (N)  $(g')^2 \leq (f)^2$  and (D)  $g(1 - g) \geq F(1 - F)$ . Since admissibility (Definition 3.2.1) implies  $g' > 0$  and  $g \geq 1/2$  for  $\theta \in [0, 1)$ , it suffices to show that (N1)  $g' \leq f$  and (D1)  $g \leq F$ . From the definition of  $\xi(\theta, x)$  in (3.18) and condition (3.24), we have for  $x \in [-1, 0], \theta \in [0, 1)$

$$\frac{d}{d\theta}\xi(\theta, x) = -(f'(x - \theta) + f'(x + \theta)) \leq 0. \quad (\text{A.2})$$

Since  $\gamma(x)$  is always positive, Equations (3.23) (with interchanged order of integration and differentiation) and (A.2) yield (N1). By integrating Equation (A.2) along  $\theta \in [0, \theta_1]$  with the boundary condition  $\xi(0, x) = 0$  (which is true by definition), we obtain

$$\xi(\theta_1, x) \leq 0, \text{ for } x \in [-1, 0], \theta_1 \in [0, 1). \quad (\text{A.3})$$

Once again, since  $\gamma(x)$  is always positive, Equations (3.22) and (A.3) yield (D1).

## A.3 Derivation for Example 3.2.7

We will show that (3.24) holds for Gaussian distribution with  $\sigma^2 \geq 1$ . Since unimodality ensures that  $f'(w) < 0$  for  $w > 0$ , it suffices to show that (3.24) hold in the (restricted) domain  $0 \leq z \leq w \leq 1$ . Noting that  $f'(w - z) = -f'(z - w)$  (from symmetric property of  $f(w)$ ) and defining  $\alpha \triangleq w/z$ , it suffices to establish

$$f'(z(1 + \alpha)) \leq f'(z(1 - \alpha)) \quad (\text{A.4})$$

for the domain  $\alpha \in [0, 1]$ ,  $z \in [0, 1]$  and  $\sigma^2 \geq 1$ . Substituting  $f'(w) = -(\sqrt{2\pi}\sigma^3)^{-1}w \exp(-w^2/(2\sigma^2))$  and rearranging terms, condition (A.4) is equivalent to showing

$$\log \frac{1+\alpha}{1-\alpha} \geq \frac{2\alpha z^2}{\sigma^2}. \quad (\text{A.5})$$

The following identity can be ascertained easily for  $0 \leq \alpha \leq 1$

$$h(\alpha) \triangleq \log \frac{1+\alpha}{1-\alpha} - 2\alpha \geq 0, \quad (\text{A.6})$$

by noting that  $h(0) = 0$  and  $h'(\alpha) \geq 0$ . The additional conditions  $0 \leq z \leq 1$  and  $\sigma^2 \geq 1$  imply (A.5), thereby completing the derivation.

## A.4 Proof of Proposition 3.2.8

We would prove Equation (3.26) and (3.27). Starting from (3.22), for  $\theta = -1$  and small  $\sigma$  we proceed from (3.22) as follows

$$\begin{aligned} g(-1) &= F(-1) + \int_{-1}^0 \gamma_0(x) (f(x+1) - f(x-1)) \, dx \\ &\stackrel{(a)}{=} F(-1) + \int_0^1 \gamma_0(z-1) (f(z) - f(z-2)) \, dz \\ &\stackrel{(b)}{=} \int_0^\infty \gamma_0(z-1) f(z) \, dz + \mathcal{O}(\sigma^4) \\ &\stackrel{(c)}{=} \int_0^\infty \left( \frac{\pi^2}{16} z^2 + \mathcal{O}(z^4) \right) f(z) \, dz + \mathcal{O}(\sigma^4) \\ &\stackrel{(d)}{=} \frac{\pi^2 \sigma^2}{32} + \mathcal{O}(\sigma^4), \end{aligned} \quad (\text{A.7})$$

where (a) is due to change in variables, (b) is due to bounding three distinct terms. First,  $F(-1) = \int_{-\infty}^{-1} f(z) \, dz$  can be bounded to  $\mathcal{O}(\sigma^4)$ <sup>1</sup> assuming that the normalized fourth-moment is bounded and applying an inequality (precisely, no. 26.1.41) in [1]. The other two terms are themselves bounded by  $F(-1)$ , i.e.,  $\int_1^\infty \gamma_0(z-1) f(z) \, dz < \int_1^\infty f(z) \, dz = F(-1)$

---

<sup>1</sup>Recall the Landau or "big O" notation: a function  $f$  is asymptotically bounded above by  $g$ , written  $f(n) = \mathcal{O}(g(n))$ , if there exist constants  $N > 0$  and  $c > 0$  such that  $f(n) \leq cg(n)$  for all  $n > N$ .

and  $\int_0^1 \gamma_0(z-1)f(z-2) dz < \int_0^1 f(z-2) dz < F(-1)$  and hence are of the order  $\mathcal{O}(\sigma^4)$ .  
(c) follows from the Taylor-series expansion of (3.13) in the vicinity of  $z = 0$ ,

$$\gamma_0(z-1) = \frac{1}{2} \left[ 1 - \cos\left(\frac{\pi z}{2}\right) \right] = \frac{\pi^2 z^2}{16} + \mathcal{O}(z^4),$$

and (d) follows from the partial moment relation  $\int_0^\infty z^n f(z) dz = \mathcal{O}(\sigma^n)$ . Similarly, for  $g'(\theta)$  in (3.23),

$$\begin{aligned} g'(-1) &= -\frac{\pi^2}{16} \int_0^\infty z^2 f'(z) dz + \mathcal{O}(\sigma^3) \\ &\stackrel{(a)}{=} \frac{\pi^2 \sigma \mu_1}{8} + \mathcal{O}(\sigma^3), \end{aligned} \tag{A.8}$$

where (a) follows from integration by parts and the fact that  $z^2 f(z)|_0^\infty = 0$ . Applying (A.7) and (A.8) in (3.9) we obtain (3.26).

The normalized one-sided mean  $\mu_1$  depends on the shape of the noise density and the inequality in (3.27) is due to the fact that  $\mu_1 < 1/2$  for any zero-mean, symmetric noise density  $f(w)$ . Consider the function

$$f_0(w) = \begin{cases} 2f(w) & w \geq 0 \\ 0 & w < 0 \end{cases}, \tag{A.9}$$

which is also a density function since  $\int_{-\infty}^\infty f_0(w) dw = 1$  and hence must have a positive variance. Thus  $\text{Var}_{f_0}(w) = \mathbb{E}_{f_0}(w^2) - (\mathbb{E}_{f_0}(w))^2 = \sigma^2(1 - 4\mu_1^2) > 0$ . Hence  $\mu_1 < 1/2$ .

## A.5 Derivation of limits in Table 3.1

To derive the limits in Table 3.1, we consider the *generalized Gaussian* density [63], specified in terms of the shape parameter  $\beta$  and variance  $\sigma^2$  as  $f(w; \beta, \sigma^2) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-(|w|/\alpha)^\beta\right)$ , where  $\alpha$  is related to variance by  $\alpha^2 = \sigma^2 \frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}$  and the one-sided mean is  $\int_0^\infty wf(w) dw = \frac{\alpha\Gamma(2/\beta)}{2\Gamma(1/\beta)}$ . Here  $\Gamma(b) \triangleq \int_0^\infty t^{b-1} e^{-t} dt$  denotes the Gamma function. Common densities like Laplacian ( $\beta = 1$ ) and Gaussian ( $\beta = 2$ ) pdf-s are specific examples of this family.

From [63], the normalized fourth-moment is  $\sigma^{-4} \int_{-\infty}^{\infty} w^4 f(w; \sigma^2) = \frac{\Gamma(5/\beta)}{\Gamma(1/\beta)}$ , which is clearly bounded for finite  $\beta$ . Hence Proposition 3.2.8 applies, and we have from (3.26),

$$\lim_{\sigma^2 \rightarrow 0} \text{CRB}(\pm 1, \gamma_0; f(w; \beta, \sigma^2)) = \frac{8}{\pi^2} \frac{\Gamma(1/\beta)\Gamma(3/\beta)}{(\Gamma(2/\beta))^2}. \quad (\text{A.10})$$

Specific instances of this result  $\beta = 1$  and  $\beta = 2$  are shown in Table 3.1. Note that  $\Gamma(1) = 1$ , and  $\Gamma(b) = (b-1)\Gamma(b-1)$  for all  $b > 1$ , which simplifies to  $\Gamma(b) = (b-1)!$  for integer  $b > 1$ . Furthermore,  $\Gamma(1/2) = \sqrt{\pi}$ .

## A.6 Proof of Proposition 5.3.2

Since  $\mathbf{w} = \text{vec}(\mathbf{W})$  for a connected topology, we obtain  $\mathbf{\Omega}_P = \mathbf{E}_x \otimes \mathbf{I}$ ,  $\mathbf{G} = \mathbf{I} \otimes \mathbf{g}$  and

$$\begin{aligned} \mathbf{\Omega}_{\text{JD}} &= (\mathbf{E}_x \otimes \mathbf{E}_g) - \eta^2 \mathbf{G} \mathbf{h} \mathbf{h}^T \mathbf{G}^T \\ &= (\mathbf{E}_x \otimes \mathbf{\Sigma}_g) + \mathbf{G} \tilde{\mathbf{\Sigma}} \mathbf{G}^T. \end{aligned} \quad (\text{A.11})$$

Substituting the appropriate values in (5.44a), we have (assuming all the inverses exist)

$$\begin{aligned} J_{\text{opt}}(P) &\stackrel{(a)}{=} \mathbf{h}^T \mathbf{G}^T \left( \mathbf{G} \tilde{\mathbf{\Sigma}} \mathbf{G}^T + \tilde{\mathbf{\Omega}}_P \right)^{-1} \mathbf{G} \mathbf{h} \\ &\stackrel{(b)}{=} \mathbf{h}^T \left( \tilde{\mathbf{\Sigma}} + \tilde{\mathbf{\Gamma}}_P \right)^{-1} \mathbf{h} \\ &\stackrel{(c)}{=} \mathbf{h}^T \left( \left( 1 + \frac{1}{\mathcal{G}} \right) \tilde{\mathbf{\Sigma}} + \frac{1}{\mathcal{G}} \eta^2 \mathbf{h} \mathbf{h}^T \right)^{-1} \mathbf{h} \\ &\stackrel{(d)}{=} \tilde{J} \left[ 1 + \frac{1 + \eta^2 \tilde{J}}{\mathcal{G}} \right]^{-1}, \end{aligned} \quad (\text{A.12})$$

where step (a) follows by defining  $\tilde{\mathbf{\Omega}}_P \triangleq \mathbf{E}_x \otimes \tilde{\mathbf{\Sigma}}_g$  where  $\tilde{\mathbf{\Sigma}}_g$  is already defined in (5.50). Step (b) follows from arguments similar to those used in (5.46) and by defining (and subsequently



simplifying)  $\tilde{\Gamma}_{\mathbf{p}}$  as

$$\begin{aligned}\tilde{\Gamma}_{\mathbf{p}} &\triangleq \left( \mathbf{G}^T \tilde{\Omega}_{\mathbf{p}}^{-1} \mathbf{G} \right)^{-1} \\ &= \left( (\mathbf{I} \otimes \mathbf{g}^T) \left( \mathbf{E}_{\mathbf{x}}^{-1} \otimes \tilde{\Sigma}_{\mathbf{g}}^{-1} \right) (\mathbf{I} \otimes \mathbf{g}) \right)^{-1} \\ &= \frac{\mathbf{E}_{\mathbf{x}}}{\mathcal{G}}, \quad \text{where } \mathcal{G} \triangleq \mathbf{g}^T \tilde{\Sigma}_{\mathbf{g}}^{-1} \mathbf{g},\end{aligned}\tag{A.13}$$

step (c) follows from the fact that  $\mathbf{E}_{\mathbf{x}} = \tilde{\Sigma} + \eta^2 \mathbf{h} \mathbf{h}^T$ , and step (d) follows from the definition  $\tilde{\mathbf{J}} \triangleq \mathbf{h}^T \tilde{\Sigma}^{-1} \mathbf{h}$  and following identities involving rank-1 updated matrix inverses. For any scalars  $\alpha \neq 0$  and  $\beta$ , vector  $\mathbf{p}$  and invertible matrix  $\mathbf{Q}$ ,

$$(\alpha \mathbf{Q} + \beta \mathbf{p} \mathbf{p}^T)^{-1} = \frac{\mathbf{Q}^{-1}}{\alpha} - \frac{\beta \mathbf{Q}^{-1} \mathbf{p} \mathbf{p}^T \mathbf{Q}^{-1}}{\alpha(\alpha + \beta Q_p)}, \quad \text{and}\tag{A.14}$$

$$\mathbf{p}^T (\alpha \mathbf{Q} + \beta \mathbf{p} \mathbf{p}^T)^{-1} \mathbf{p} = \frac{Q_p}{\alpha + \beta Q_p}, \quad Q_p \triangleq \mathbf{p}^T \mathbf{Q}^{-1} \mathbf{p}.\tag{A.15}$$

From equations (5.44b) and discussion in Example 1, the optimal weights are,

$$\begin{aligned}\mathbf{w}_{\text{opt}} &\propto \tilde{\Omega}_{\mathbf{p}}^{-1} \mathbf{G} \tilde{\Sigma} \left( \tilde{\Sigma} + \tilde{\Gamma}_{\mathbf{p}} \right)^{-1} \mathbf{h}, \quad (\text{from (5.44b)}) \\ &\propto \tilde{\Omega}_{\mathbf{p}}^{-1} \mathbf{G} \mathbf{h} \\ &= \left( \mathbf{E}_{\mathbf{x}}^{-1} \otimes \tilde{\Sigma}_{\mathbf{g}}^{-1} \right) (\mathbf{h} \otimes \mathbf{g}) \\ &= (\mathbf{E}_{\mathbf{x}}^{-1} \mathbf{h}) \otimes \left( \tilde{\Sigma}_{\mathbf{g}}^{-1} \mathbf{g} \right)\end{aligned}\tag{A.16}$$

which implies that  $\mathbf{W}_{\text{opt}} \propto \tilde{\Sigma}_{\mathbf{g}}^{-1} \mathbf{g} \mathbf{h}^T \mathbf{E}_{\mathbf{x}}^{-1}$ . Step (a) follows the fact that  $\left( \tilde{\Sigma} + \tilde{\Gamma}_{\mathbf{p}} \right)^{-1} \mathbf{h} \propto \tilde{\Sigma}^{-1} \mathbf{h}$  (see (A.14)).

From Corollary 2.3.5 of [27], the sum-rate required to encode a single-dimensional real-valued Gaussian source with variance  $\eta^2$ , observed through the vector  $\mathbf{h}$  and Gaussian observation noise with covariance  $\Sigma$ , in such a way that reconstruction incurs an average distortion of at most  $D$ , satisfies

$$R_{\text{tot}} \geq \frac{1}{2} \log \frac{\lambda}{D - D_0}, \quad \text{where } \lambda = \frac{\eta^4 J_0}{1 + \eta^2 J_0}.\tag{A.17}$$

Since, for a fixed sum-power  $P$ , the sum-rate has to be lesser than the (centralized) capacity of the coherent MAC channel, i.e.,  $R_{\text{tot}} \leq C$ , where  $C = \frac{1}{2} \log(1 + \|\mathbf{g}\|^2 P_\xi)$ , we obtain

$$1 + \|\mathbf{g}\|^2 P_\xi \geq \frac{\eta^4 J_0}{(D - D_0)(1 + \eta^2 J_0)}. \quad (\text{A.18})$$

Replacing  $D$  by  $J$  (recall,  $J = \frac{1}{D} - \frac{1}{\eta^2}$ ) and after some algebra, we obtain,

$$J \leq J_0 \left[ 1 + \frac{1 + \eta^2 J_0}{\|\mathbf{g}\|^2 P_\xi} \right]^{-1}. \quad (\text{A.19})$$

But the right hand side is precisely the distortion achieved by a connected network (see (5.50) and note that  $\mathcal{G} = \|\mathbf{g}\|^2 P_\xi$  for  $\Sigma_{\mathbf{g}} = 0$ ). This establishes the information theoretic optimality.

## A.7 An equality involving positive definite matrices

**Lemma A.7.1** (An inequality). *For any  $N$ -dimensional vector  $\mathbf{p}$  and  $N \times N$  symmetric positive definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,*

$$\frac{1}{\mathbf{p}^T (\mathbf{A} + \mathbf{B})^{-1} \mathbf{p}} \geq \frac{1}{\mathbf{p}^T \mathbf{A}^{-1} \mathbf{p}} + \frac{1}{\mathbf{p}^T \mathbf{B}^{-1} \mathbf{p}}. \quad (\text{A.20})$$

*Proof:* Since  $\mathbf{A}, \mathbf{B} \in \mathcal{S}^{++}$ ,  $\mathbf{A}^{-\frac{1}{2}} \mathbf{B} \mathbf{A}^{-\frac{1}{2}} \in \mathcal{S}^{++}$ . Define by  $\mathbf{U}$  and  $\mathbf{\Lambda}$  the following eigendecomposition  $\mathbf{A}^{-\frac{1}{2}} \mathbf{B} \mathbf{A}^{-\frac{1}{2}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ . Hence  $\lambda_n > 0, \forall n$ . Define  $\mathbf{q} = \mathbf{U}^T \mathbf{A}^{-\frac{1}{2}} \mathbf{h}$ . Note that

$$\begin{aligned} \mathbf{q}^T \mathbf{q} &= \mathbf{p}^T \mathbf{A}^{-1} \mathbf{p}, \\ \mathbf{q}^T \mathbf{\Lambda}^{-1} \mathbf{q} &= \mathbf{p}^T \mathbf{B}^{-1} \mathbf{p}, \text{ and} \\ \mathbf{q}^T (\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{q} &= \mathbf{p}^T (\mathbf{A} + \mathbf{B})^{-1} \mathbf{p}. \end{aligned} \quad (\text{A.21})$$

Hence, to prove (A.20), it suffices to show that

$$\frac{1}{\sum_{n=1}^N \frac{q_n^2}{1+\lambda_n}} \geq \frac{1}{\sum_{n=1}^N q_n^2} + \frac{1}{\sum_{n=1}^N \frac{q_n^2}{\lambda_n}},$$

or equivalently, with  $a_n \triangleq \frac{1}{1+\lambda_n}$  and  $b_n \triangleq \frac{1+\lambda_n}{\lambda_n}$ ,

$$\sum_{n=1}^N q_n^2 \sum_{n=1}^N q_n^2 a_n b_n \geq \sum_{n=1}^N q_n^2 a_n \sum_{n=1}^N q_n^2 b_n. \quad (\text{A.22})$$

Since  $\lambda_n > 0, \forall n$ , both  $a_n$  and  $b_n$  are decreasing functions of  $\lambda_n$ . Hence inequality (A.22) follows from the Chebyshev's (sum) inequality (page 240, Equation 1.4, [61]). Equality holds if and only if, for all indices  $k$  for which  $q_k \neq 0$  (denote such a set by  $\text{ixnz}(\mathbf{q})$ ), the eigenvalues are similar. That is, iff  $\lambda_k = \lambda, \forall k \in \text{ixnz}(\mathbf{q})$ . *QED*.

## A.8 Proof of Proposition 5.3.4

We would use Equation (5.44a). To start with, we note that  $\mathbf{G}\mathbf{h}$  is a multiple of  $\mathbf{1}$  and both matrices  $\mathbf{\Omega}_{\text{JD}}$  and  $\mathbf{\Omega}_{\text{P}}$  has an eigenvector as  $\mathbf{1}$ , where  $\mathbf{1}$  has dimension  $L = MK$ . In particular, careful inspection of (5.33) (the elements of matrices  $\mathbf{E}_{\text{g}}$  and  $\mathbf{E}_x$  take two distinct values, diagonal and otherwise) yields

$$\begin{aligned} \mathbf{G}\mathbf{h} &= g_0 h_0 \sqrt{\alpha_{\text{g}} \alpha_{\text{h}}} \mathbf{1}, \\ \mathbf{\Omega}_{\text{P}} \mathbf{1} &= \sigma_{\text{x}}^2 (1 + (K-1)\alpha_{\text{x}}) \mathbf{1}, \text{ and} \\ \mathbf{\Omega}_{\text{JD}} \mathbf{1} &= [\sigma_{\text{x}}^2 g_0^2 \{1 + (K-1)(\alpha_{\text{x}} + \alpha_{\text{g}}) \\ &\quad + (MK - 2K + 1)\alpha_{\text{x}} \alpha_{\text{g}}\} - \eta^2 g_0^2 h_0^2 \alpha_{\text{g}} \alpha_{\text{h}} MK] \mathbf{1}, \end{aligned} \quad (\text{A.23})$$

where  $\sigma_{\text{x}}^2 = \sigma^2(1 + \gamma)$ ,  $\alpha_{\text{x}} = \frac{\rho + \gamma \alpha_{\text{h}}}{1 + \gamma}$  and  $\gamma = \frac{\eta^2 h_0^2}{\sigma^2}$ . Based on the above equations, define scalars  $\phi, \mu, \nu$  be such that  $\mathbf{G}\mathbf{h}\mathbf{h}^T \mathbf{G}^T \mathbf{1} = \phi \mathbf{1}$ ,  $\mathbf{\Omega}_{\text{P}} \mathbf{1} = \mu \mathbf{1}$  and  $\mathbf{\Omega}_{\text{JD}} \mathbf{1} = (\nu - \eta^2 \phi) \mathbf{1}$ , in

particular,

$$\begin{aligned}
\phi &= g_0^2 h_0^2 \alpha_g \alpha_h MK, \\
\mu &= \sigma_x^2 (1 + (K - 1)\alpha_x), \text{ and} \\
\nu &= \sigma_x^2 g_0^2 \{1 + (K - 1)(\alpha_x + \alpha_g) + (MK - 2K + 1)\alpha_x \alpha_g\},
\end{aligned} \tag{A.24}$$

From (5.44a), we therefore obtain

$$J_{\text{opt}}(P) = \frac{\phi}{\nu + \frac{\mu}{P_\xi} - \eta^2 \phi}, \tag{A.25}$$

which when simplified further leads to (5.55). Since  $\mathbf{1}$  is the corresponding eigenvector, we also have  $\mathbf{w}_{\text{opt}} \propto \mathbf{1}$ , i.e., the sensors just average all the observations.

## A.9 Proof of Proposition 5.3.5

Our goal is to show that for any feasible  $J < J_{\text{opt}}$ ,  $\mathcal{X}^R(J)$  contains a rank-1 matrix. Specifically, we will show that

$$\widetilde{\mathbf{X}} \triangleq \arg \max_{\mathbf{X} \in \mathcal{X}^R(J)} \text{Tr} [\mathbf{\Omega}_P \mathbf{X}], \tag{A.26}$$

which is the global optimizer to the (convex) semi-definite optimization problem

$$\begin{aligned}
&\underset{\mathbf{X}}{\text{minimize}} && \text{Tr} [\mathbf{\Omega}_P \mathbf{X}] \\
&\text{subject to} && \text{Tr} [(J\mathbf{\Omega}_{\text{JD}} - \mathbf{\Omega}_{\text{JN}}) \mathbf{X}] + J\xi^2 \leq 0, \\
&&& \text{Tr} [\mathbf{\Omega}_{P,m} \mathbf{X}] \leq P_m^c, \ m = 1, \dots, M, \\
&&& -\mathbf{X} \preceq 0,
\end{aligned} \tag{A.27}$$

is rank-1. The Lagrangian of (A.27) is given by,

$$\begin{aligned}
\mathcal{L}(\mathbf{X}, \alpha, \boldsymbol{\beta}, \mathbf{Z}) = & \text{Tr} [\boldsymbol{\Omega}_{\text{P}} \mathbf{X}] \\
& + \alpha \left( \text{Tr} [(J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}) \mathbf{X}] + J\xi^2 \right) \\
& + \sum_{m=1}^M \beta_m \left( \text{Tr} [\boldsymbol{\Omega}_{\text{P},m} \mathbf{X}] - P_m^{\text{C}} \right) \\
& - \text{Tr} [\mathbf{X} \mathbf{Z}],
\end{aligned} \tag{A.28}$$

with the dual problem being,

$$\begin{aligned}
& \underset{\alpha, \boldsymbol{\beta}}{\text{maximize}} \quad \alpha J\xi^2 - \sum_{m=1}^M \beta_m P_m^{\text{C}} \\
& \text{subject to} \quad \tilde{\mathbf{Z}} \triangleq \boldsymbol{\Omega}_{\text{P}} + \alpha (J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}) + \sum_{m=1}^M \beta_m \boldsymbol{\Omega}_{\text{P},m} \succeq 0, \\
& \quad \alpha \geq 0, \boldsymbol{\beta} \geq 0.
\end{aligned} \tag{A.29}$$

Also, we have the following complementary conditions (let *tilde* denote respective values at optimality),

$$\tilde{\alpha} \left( \text{Tr} [(J\boldsymbol{\Omega}_{\text{JD}} - \boldsymbol{\Omega}_{\text{JN}}) \tilde{\mathbf{X}}] + J\xi^2 \right) = 0, \tag{A.30a}$$

$$\tilde{\beta}_m \left( \text{Tr} [\boldsymbol{\Omega}_{\text{P},m} \tilde{\mathbf{X}}] - P_m^{\text{C}} \right) = 0, \quad m = 1, \dots, M, \tag{A.30b}$$

$$\text{Tr} [\tilde{\mathbf{X}} \tilde{\mathbf{Z}}] = 0. \tag{A.30c}$$

Without loss of generality, let  $\tilde{\mathbf{X}} = \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T$  (such a decomposition is possible since  $\tilde{\mathbf{X}}$  is symmetric positive semidefinite). Also denote the columns of  $\tilde{\mathbf{Y}}$  as  $\tilde{\mathbf{w}}_l \in \mathbb{R}^L$  for  $l = 1, 2, \dots, L$ , so that  $\tilde{\mathbf{Y}} = [\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \dots, \tilde{\mathbf{w}}_L]$ . From (A.30c), we have

$$\text{Tr} [\tilde{\mathbf{X}} \tilde{\mathbf{Z}}] = \sum_{l=1}^L \tilde{\mathbf{w}}_l^T \tilde{\mathbf{Z}} \tilde{\mathbf{w}}_l = 0, \tag{A.31}$$

which coupled with the fact that  $\tilde{\mathbf{Z}} \succeq 0$  implies that

$$\tilde{\mathbf{Z}}\tilde{\mathbf{w}}_l = 0, \text{ for all } l = 1, \dots, L. \quad (\text{A.32})$$

Thus, from definition of  $\tilde{\mathbf{Z}}$  in (A.29), we conclude that  $\tilde{\mathbf{w}}_l$ -s are the generalized eigenvectors that satisfy

$$\begin{aligned} \left( \tilde{\mathbf{\Omega}} + \tilde{\alpha} (J\mathbf{\Omega}_{\text{JD}} - \mathbf{t}\mathbf{t}^T) \right) \tilde{\mathbf{w}}_l &= \mathbf{0}_L, \quad \forall l, \text{ with} \\ \tilde{\mathbf{\Omega}} &\triangleq \mathbf{\Omega}_{\text{P}} + \sum_{m=1}^M \tilde{\beta}_m \mathbf{\Omega}_{\text{P},m}, \\ \mathbf{t} &\triangleq \mathbf{G}\mathbf{h} \text{ (note } \mathbf{\Omega}_{\text{JN}} \text{ in (5.33))}. \end{aligned} \quad (\text{A.33})$$

Hence it follows that for all  $l$ ,

$$\begin{aligned} &\left( \left( \tilde{\mathbf{\Omega}} + \tilde{\alpha} J\mathbf{\Omega}_{\text{JD}} \right) - \tilde{\alpha} \mathbf{t}\mathbf{t}^T \right) \tilde{\mathbf{w}}_l = \mathbf{0}_L, \\ \Leftrightarrow &\left( \mathbf{I}_L - \left( \tilde{\mathbf{\Omega}} + \tilde{\alpha} J\mathbf{\Omega}_{\text{JD}} \right)^{-1} \tilde{\alpha} \mathbf{t}\mathbf{t}^T \right) \tilde{\mathbf{w}}_l = \mathbf{0}_L, \end{aligned} \quad (\text{A.34})$$

$$\Rightarrow \tilde{\mathbf{w}}_l \propto \left( \tilde{\mathbf{\Omega}} + \tilde{\alpha} J\mathbf{\Omega}_{\text{JD}} \right)^{-1} \mathbf{t}, \quad (\text{A.35})$$

where (A.34) is because  $\mathbf{\Omega}_{\text{P}}$  is positive definite and hence the matrix  $\left( \tilde{\mathbf{\Omega}} + \tilde{\alpha} J\mathbf{\Omega}_{\text{JD}} \right)$  is also positive definite and invertible, and (A.35) follows from the fact that both  $\tilde{\alpha}$  and  $\mathbf{t}^T \tilde{\mathbf{w}}_l$  has to be non-zero to satisfy (A.33). We thus conclude that

$$\begin{aligned} &\tilde{\mathbf{w}}_l \text{ is unique upto its norm, } \forall l, \\ \Rightarrow &\tilde{\mathbf{X}} \text{ is a rank-1 matrix,} \end{aligned} \quad (\text{A.36})$$

thereby establishing Proposition 5.3.5.

## A.10 Proof of Proposition 5.3.7

Problem (5.77) is equivalent to

$$\begin{aligned} \underset{\mathbf{W}}{\text{maximize}} \quad & \mathcal{J}_{\mathbf{W}} = \frac{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}{\text{Tr} [\mathbf{E}_{\mathbf{g}} \mathbf{W} \mathbf{E}_{\mathbf{x}} \mathbf{W}^T] + \xi^2}, \\ \text{subject to} \quad & [\mathbf{W} \mathbf{E}_{\mathbf{x}} \mathbf{W}^T]_{m,m} \leq P_m^{\mathbf{C}}, \quad m = 1, \dots, M, \end{aligned} \quad (\text{A.37})$$

in the sense that  $J_{\mathbf{W}}$  and  $\mathcal{J}_{\mathbf{W}}$  are monotonically related through  $J_{\mathbf{W}} = \frac{\mathcal{J}_{\mathbf{W}}}{1 - \eta^2 \mathcal{J}_{\mathbf{W}}}$  and  $\mathbf{W}_{\text{opt}}$  is the same for both problems. This is further equivalent to

$$\begin{aligned} \underset{\mathbf{V}}{\text{maximize}} \quad & \mathcal{J}_{\mathbf{V}} = \frac{(\mathbf{g}^T \mathbf{V} \mathbf{h}_{\mathbf{x}})^2}{\text{Tr} [\mathbf{E}_{\mathbf{g}} \mathbf{V} \mathbf{V}^T] + \xi^2}, \\ \text{subject to} \quad & \|\mathbf{v}_m\|^2 \leq P_m^{\mathbf{C}}, \quad m = 1, \dots, M, \end{aligned} \quad (\text{A.38})$$

by defining  $\mathbf{V}$ ,  $\mathbf{v}_m$  and  $\mathbf{h}_{\mathbf{x}}$  such that

$$\mathbf{V} \triangleq \mathbf{W} \mathbf{E}_{\mathbf{x}}^{\frac{1}{2}} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_M^T \end{bmatrix}, \quad \text{and } \mathbf{h}_{\mathbf{x}} \triangleq \mathbf{E}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{h}. \quad (\text{A.39})$$

With a goal to reduce the number of optimization variables from  $MN$  to  $M$ , we define the matrix transformation  $\mathbf{V} \rightarrow \mathbf{V}_{\mathbf{x}}$  as one that retains the norm of its individual row vectors but otherwise aligns the rows to  $\mathbf{h}_{\mathbf{x}}^T$ , i.e.,

$$\begin{aligned} \mathbf{V}_{\mathbf{x}} &\triangleq \mathbf{t} \frac{\mathbf{h}_{\mathbf{x}}^T}{\|\mathbf{h}_{\mathbf{x}}\|}, \quad t_m \triangleq \|\mathbf{v}_m\|, \quad \text{so that} \\ \mathcal{J}_{\mathbf{V}_{\mathbf{x}}} &= \|\mathbf{h}_{\mathbf{x}}\|^2 \frac{(\mathbf{g}^T \mathbf{t})^2}{\mathbf{t}^T \mathbf{E}_{\mathbf{g}} \mathbf{t} + \xi^2}. \end{aligned} \quad (\text{A.40})$$

We would need the following result to proceed further.

**Lemma A.10.1.** *When  $\Sigma_{\mathbf{g}}$  is diagonal,*

$$\mathcal{J}_{\mathbf{V}} \leq \mathcal{J}_{\mathbf{V}_{\mathbf{x}}}, \quad \text{for any } \mathbf{V}. \quad (\text{A.41})$$

*Proof:* To prove Lemma (A.10.1), we will show that

$$\frac{\mathcal{J}_{\mathbf{V}}}{\|\mathbf{h}_x\|^2} \stackrel{(a)}{\leq} \frac{\|\mathbf{g}^T \mathbf{V}\|^2}{\text{Tr} [\mathbf{E}_g \mathbf{V} \mathbf{V}^T] + \xi^2} \stackrel{(b)}{\leq} \frac{(\mathbf{g}^T \mathbf{t})^2}{\mathbf{t}^T \mathbf{E}_g \mathbf{t} + \xi^2} = \frac{\mathcal{J}_{\mathbf{V}_x}}{\|\mathbf{h}_x\|^2}, \quad (\text{A.42})$$

where (a) follows from definition of  $\mathcal{J}_{\mathbf{V}}$  in (5.77) and Cauchy-Schwartz inequality implying  $(\mathbf{g}^T \mathbf{V} \mathbf{h}_x)^2 \leq \|\mathbf{g}^T \mathbf{V}\|^2 \|\mathbf{h}_x\|^2$ , and the last equality is due to definition of  $\mathcal{J}_{\mathbf{V}_x}$  in (A.40). Hence it remains to prove (b), which can be established by showing

$$\text{Tr} \left[ \left( (\mathbf{t}^T \Sigma_g \mathbf{t}) \mathbf{g} \mathbf{g}^T - (\mathbf{g}^T \mathbf{t})^2 \Sigma_g \right) \mathbf{V} \mathbf{V}^T \right] \leq 0, \quad (\text{A.43a})$$

$$\text{and } \xi^2 \left( \|\mathbf{g}^T \mathbf{V}\|^2 - (\mathbf{g}^T \mathbf{t})^2 \right) \leq 0. \quad (\text{A.43b})$$

Define  $\bar{\mathbf{g}}$ ,  $\bar{\mathbf{V}}$  and its aligned equivalent  $\bar{\mathbf{V}}_x$  as

$$\begin{aligned} \bar{\mathbf{g}} &\triangleq \Sigma_g^{-\frac{1}{2}} \mathbf{g}, \quad \bar{\mathbf{V}} \triangleq \Sigma_g^{\frac{1}{2}} \mathbf{V}, \text{ so that} \\ \bar{\mathbf{V}}_x &= \bar{\mathbf{t}} \frac{\mathbf{h}_x^T}{\|\mathbf{h}_x\|}, \text{ where } \bar{t}_m = \|\bar{\mathbf{v}}_m\|, \text{ or} \\ \bar{\mathbf{t}} &= \Sigma_g^{\frac{1}{2}} \mathbf{t} \text{ (since } \Sigma_g \text{ is diagonal).} \end{aligned} \quad (\text{A.44})$$

Note that  $\text{Tr} [\bar{\mathbf{V}} \bar{\mathbf{V}}^T] = \|\bar{\mathbf{t}}\|^2$ . Condition (A.43a) is therefore equivalent to showing

$$\begin{aligned} \text{Tr} \left[ \left( \|\bar{\mathbf{t}}\|^2 \bar{\mathbf{g}} \bar{\mathbf{g}}^T - (\bar{\mathbf{g}}^T \bar{\mathbf{t}})^2 \right) \bar{\mathbf{V}} \bar{\mathbf{V}}^T \right] &\leq 0, \\ \text{or equivalently } \|\bar{\mathbf{t}}\|^2 \left( \|\bar{\mathbf{g}}^T \bar{\mathbf{V}}\|^2 - (\bar{\mathbf{g}}^T \bar{\mathbf{t}})^2 \right) &\leq 0, \end{aligned} \quad (\text{A.45})$$

which is similar to condition (A.43b). Thus it remains to establish (A.43b), which is true because

$$\|\mathbf{g}^T \mathbf{V}\|^2 \stackrel{(a)}{=} \left\| \sum_{m=1}^M g_m \mathbf{v}_m \right\|^2 \stackrel{(b)}{\leq} \left\| \sum_{m=1}^M g_m \|\mathbf{v}_m\| \right\|^2 \stackrel{(c)}{=} (\mathbf{g}^T \mathbf{t})^2, \quad (\text{A.46})$$



where (a) and (c) are due to definitions of  $\mathbf{v}_m$  and  $\mathbf{t}$  respectively, and (b) is due to Cauchy-Schwartz inequality,  $\|\mathbf{v}_m^T \mathbf{v}_n\| \leq \|\mathbf{v}_m\| \|\mathbf{v}_n\|$  for all  $1 \leq m, n \leq M$ . This completes the proof. *QED.*

Note that  $\mathcal{J}_{\mathbf{V}_x}$  is a function of the vector  $\mathbf{t}$ , whose elements are non-negative (since  $t_m$  is a norm). Therefore problem (A.38), in conjunction with Lemma A.10.1, is equivalent to

$$\begin{aligned} & \underset{\mathbf{t}}{\text{maximize}} && \mathcal{J}_{\mathbf{V}_x}(\mathbf{t}), \\ & \text{subject to} && t_m^2 \leq P_m^C, m = 1, \dots, M, \end{aligned} \tag{A.47}$$

through the relations  $\mathbf{V}_{\text{opt}} = \mathbf{t}_{\text{opt}} \frac{\mathbf{h}_x^T}{\|\mathbf{h}_x\|}$  and  $\mathcal{J}_{\mathbf{V}_{\text{opt}}} = \mathcal{J}_{\mathbf{V}_x}(\mathbf{t}_{\text{opt}})$ . Problem (A.47) is further equivalent to

$$\begin{aligned} & \underset{\mathbf{t}}{\text{maximize}} && F_{\mathbf{t}} = \frac{(\mathbf{g}^T \mathbf{t})^2}{\mathbf{t}^T \Sigma_{\mathbf{g}} \mathbf{t} + \xi^2}, \\ & \text{subject to} && t_m^2 \leq P_m^C, m = 1, \dots, M, \end{aligned} \tag{A.48}$$

through the following relations between variables  $\mathcal{J}_{\mathbf{V}_x}(\mathbf{t}) = \|\mathbf{h}_x\|^2 \frac{F_{\mathbf{t}}}{1+F_{\mathbf{t}}}$ , which proves Proposition 5.3.7

## A.11 Proof of Lemma 5.3.8

We start by noting that for  $\varrho \in (0, 1)$

$$\kappa(\varrho) = \left( \frac{1 - \varrho^{\frac{M}{2}}}{1 - \varrho^{\frac{1}{2}}} \right)^2 \frac{1 - \varrho}{1 - \varrho^M} = \frac{1 - \varrho^{\frac{M}{2}}}{1 + \varrho^{\frac{M}{2}}} \frac{1 + \varrho^{\frac{1}{2}}}{1 - \varrho^{\frac{1}{2}}}. \tag{A.49}$$

We would show that  $\frac{d\kappa(\varrho)}{d\varrho} > 0$  for  $\varrho \in (0, 1)$ . From (A.49),

$$\frac{d\kappa(\varrho)}{d\varrho} = \frac{\varrho^{-\frac{1}{2}} - M\varrho^{\frac{M}{2}-1} - \varrho^{M-\frac{1}{2}} + M\varrho^{\frac{M}{2}}}{\left(1 + \varrho^{\frac{M}{2}}\right)^2 \left(1 - \varrho^{\frac{1}{2}}\right)^2}, \tag{A.50}$$

the numerator of which can be rearranged as

$$\begin{aligned}
& \varrho^{-\frac{1}{2}} (1 - \varrho^M) - M \varrho^{\frac{M}{2}-1} (1 - \varrho) \\
&= \varrho^{-\frac{1}{2}} (1 - \varrho) \left( 1 + \varrho + \cdots + \varrho^{M-1} - M \varrho^{\frac{M-1}{2}} \right) \\
&= \varrho^{-\frac{1}{2}} (1 - \varrho) \sum_{m=1}^{\lfloor \frac{M}{2} \rfloor} \left( \varrho^{\frac{m-1}{2}} - \varrho^{\frac{M-m}{2}} \right)^2,
\end{aligned} \tag{A.51}$$

which is evidently a positive quantity. This completes the proof.

## A.12 Proof of Proposition 5.3.9

We start with the cumulative-constraint case, for which we will use the results from Example 3. From Equation (A.25) and corresponding to the distributed ( $K = 1$ ) and connected cases ( $K = M$ ), we denote the constants  $\{\mu, \nu\}$  with subscripts 1, 2, as  $\{\mu_1, \nu_1\}$  and  $\{\mu_2, \nu_2\}$  respectively, i.e.,

$$\begin{aligned}
J_{\text{opt}}^{\text{dist}}(P) &= \frac{\phi}{\nu_1 + \frac{\mu_1}{P_\xi} - \eta^2 \phi}, \text{ and} \\
J_{\text{opt}}^{\text{conn}}(P) &= \frac{\phi M}{\nu_2 + \frac{\mu_2}{P_\xi} - \eta^2 \phi M},
\end{aligned} \tag{A.52}$$

where  $\phi = g_0^2 h_0^2 \alpha_g \alpha_h M$ ,  $\nu_1 = \sigma_x^2 g_0^2 (1 + (M-1) \alpha_g \alpha_x)$ ,  $\mu_1 = \sigma_x^2$ ,  $\nu_2 = \sigma_x^2 g_0^2 (1 + (M-1) \alpha_g) (1 + (M-1) \alpha_x)$  and  $\mu_2 = \sigma_x^2 (1 + (M-1) \alpha_x)$ . Applying the corresponding distortion terms (note  $D = \left( \frac{1}{\eta^2} + J \right)^{-1}$ ) in (5.30), (note that  $J_0 = \frac{\phi M}{\nu_2 - \eta^2 \phi M}$  therefore the denominator term of (5.30) is  $\eta^2 - D_0 = \frac{\eta^2 \phi}{\nu_2}$ ) the collaboration gain can be simplified as

$$\text{CG} = \frac{\frac{M\mu_1 - \mu_2}{M\mu_1} + P_\xi \frac{M\nu_1 - \nu_2}{M\mu_1}}{\left( 1 + \frac{1}{P_\xi} \frac{\mu_2}{\nu_2} \right) \left( 1 + P_\xi \frac{\nu_1}{\mu_1} \right)}. \tag{A.53}$$

Each of the fragments can be simplified further,  $\frac{M\mu_1 - \mu_2}{M\mu_1} = \frac{1}{M}(M-1)(1 - \alpha_x)$ ,  $\frac{M\nu_1 - \nu_2}{M\mu_1} = \frac{g_0^2}{M}(M-1)(1 - \alpha_g)(1 - \alpha_x)$ ,  $\frac{\mu_2}{\nu_2} = \frac{1}{g_0^2(1+(M-1)\alpha_g)}$  and  $\frac{\nu_1}{\mu_1} = g_0^2(1 + (M-1)\alpha_g \alpha_x)$ . Replacing

these fragments in (A.53), defining  $P_g = P_\xi g_0^2$  and dividing both numerator and denominator by  $(1 + P_g)$  leads to Equation (5.85) (with  $\kappa$  replaced by  $M$ ).

For the individual-constraint case, we would use Examples 5 (distributed) and 6 (connected) to compute the collaboration gain. First we show that all the constraints are active for both the distributed and connected cases if condition (5.84) is satisfied. If we apply Proposition 5.3.6 to a homogeneous problem with  $a_1 = \dots = a_M$  and  $b_1 = \dots = b_M$ , the active constraint condition  $\Phi_{M-1} d_M \geq 1$  simplifies to

$$\frac{\sum_{m=1}^{M-1} c_m^2 + \frac{\xi^2}{b_m}}{\sum_{m=1}^{M-1} c_m} \geq c_M. \quad (\text{A.54})$$

For the distributed case (Example 5), we refer to problem (5.75) to find that  $b_m = \sigma_x^2 g_0^2 (1 - \alpha_g \alpha_x)$  and  $c_m = \frac{\sqrt{P_m^C}}{\sigma_x}$ , so that inequality (A.54) explicitly evaluates to condition (5.84). For the connected case (specialized version of Example 6 for homogeneous parameters), we refer to Equation (5.78) to note that  $b_m = g_0^2 (1 - \alpha_g)$  and  $c_m = \sqrt{P_m^C}$ , so that inequality (A.54) evaluates to

$$\frac{\sum_{m=1}^{M-1} P_m^C + \frac{\xi^2}{g_0^2 (1 - \alpha_g)}}{\sum_{m=1}^{M-1} \sqrt{P_m^C}} \geq \sqrt{P_M^C}, \quad (\text{A.55})$$

which is clearly true if condition (5.84) holds (since  $\alpha_x \in [0, 1]$ ).

As regards collaboration gain, we proceed as we did in the cumulative case. For the distributed (Equation (5.76)) and connected (Equation (5.78) with homogeneous parameters) cases, we can rearrange the terms of  $J_{\text{opt}}(P)$  to express them in the form of (A.52), where the various constants are now  $\phi = g_0^2 h_0^2 \alpha_g \alpha_h \kappa$ ,  $\nu_1 = \sigma_x^2 g_0^2 (1 + (\kappa - 1) \alpha_g \alpha_x)$ ,  $\mu_1 = \sigma_x^2$ ,  $\nu_2 = \sigma_x^2 g_0^2 (1 + (\kappa - 1) \alpha_g) (1 + (M - 1) \alpha_x)$  and  $\mu_2 = \sigma_x^2 (1 + (M - 1) \alpha_x)$ . However, in this case  $J_0$  is obtained not just by letting  $P_\xi \rightarrow \infty$  (thereby letting  $\mu$  vanish), but also by setting  $\kappa = M$  in both  $\phi$  and  $\nu$ , i.e.,  $J_0 = \frac{\phi(\kappa=M)M}{\nu_2(\kappa=M) - \eta^2 \phi(\kappa=M)M}$ . Therefore the denominator term of (5.30) is  $\eta^2 - D_0 = \frac{\eta^2 \phi(\kappa=M)}{\nu_2(\kappa=M)} = \frac{\eta^2 \phi}{\nu_2} \frac{M}{\kappa} \frac{(1 + (\kappa - 1) \alpha_g)}{(1 + (M - 1) \alpha_g)}$ , which is just a scaled version of  $\frac{\eta^2 \phi}{\nu_2}$ . Adjusting (A.53) for this scaling and rest of the derivation remaining similar, we

obtain

$$\text{CG} = \frac{\frac{\kappa}{M} \frac{(1+(M-1)\alpha_g)}{(1+(\kappa-1)\alpha_g)} \left( \frac{M\mu_1 - \mu_2}{M\mu_1} + P_\xi \frac{M\nu_1 - \nu_2}{M\mu_1} \right)}{\left( 1 + \frac{1}{P_\xi} \frac{\mu_2}{\nu_2} \right) \left( 1 + P_\xi \frac{\nu_1}{\mu_1} \right)}, \quad (\text{A.56})$$

which is precisely Equation (5.85), thereby completing the proof.

## A.13 Proof of Theorem 6.3.1

We will derive  $J_{\text{opt}}$  (Equation (6.15a)) first and  $J_{\text{eq}}$  (Equation (6.15b)) later, for the case when  $\mathbf{A} = \mathbf{I}_K \otimes (\mathbf{1}_Q \mathbf{1}_Q^T)$  (note that  $N = KQ$ ). From Theorem 6.2.1, we have

$$J_{\text{opt}} = \mathbf{h}^T (\boldsymbol{\Sigma} + \boldsymbol{\Gamma} / \mathcal{E}_\xi)^{-1} \mathbf{h}, \quad \mathcal{E}_\xi \triangleq \mathcal{E} / \xi^2, \quad \boldsymbol{\Gamma} \triangleq (\mathbf{G}^T \boldsymbol{\Omega}^{-1} \mathbf{G})^{-1} \quad (\text{A.57})$$

The construction of matrices  $\boldsymbol{\Omega}$  and  $\mathbf{G}$  are shown in Figure A.1 for the example of  $N = 4$ ,  $K = 2$ ,  $Q = 2$ . Since the asymptotic framework implies  $\mathbf{g}_k = \frac{1}{N} \widetilde{\mathbf{g}}_k$ , the norm of  $\boldsymbol{\Gamma}$  increases with  $N$  (in contrast to  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$ ), implying

$$\begin{aligned} J_{\text{opt}} &\rightarrow \mathbf{h}^T (\boldsymbol{\Gamma} / \mathcal{E}_\xi)^{-1} \mathbf{h} \\ &\stackrel{(a)}{=} \mathcal{E}_\xi \mathbf{h}^T \mathbf{G}^T \boldsymbol{\Omega}^{-1} \mathbf{G} \mathbf{h} \\ &\stackrel{(b)}{=} \mathcal{E}_\xi \sum_{k=1}^K (\mathbf{g}_k^T \mathbf{g}_k) \mathbf{h}_k^T \mathbf{E}_{\mathbf{x},k}^{-1} \mathbf{h}_k \\ &\stackrel{(c)}{=} \mathcal{E}_\xi \sum_{k=1}^K (\mathbf{g}_k^T \mathbf{g}_k) \mathbf{h}_k^T (\sigma^2 \mathbf{I} + \eta^2 \mathbf{h}_k \mathbf{h}_k^T)^{-1} \mathbf{h}_k \\ &\stackrel{(d)}{=} \frac{\mathcal{E}_\xi}{\eta^2} \sum_{k=1}^K \left( \sum_{q=1}^Q g_{kq}^2 \right) \left( 1 - \frac{1}{1 + \frac{\eta^2}{\sigma^2} \left( \sum_{q=1}^Q h_{kq}^2 \right)} \right), \end{aligned} \quad (\text{A.59})$$

where (a) follows from the definition of  $\boldsymbol{\Gamma}$  (see (A.57)), (b) follows from the structure of  $\mathbf{G}$  and  $\boldsymbol{\Omega}$  by defining  $Q$ -dimensional vectors  $\mathbf{g}_k \triangleq [g_{k1}, \dots, g_{kQ}]^T$ ,  $\mathbf{h}_k \triangleq [h_{k1}, \dots, h_{kQ}]^T$  and  $Q \times Q$  matrix  $\mathbf{E}_{\mathbf{x},k} \triangleq [\mathbb{E}_{\mathbf{x}}]_{k1, \dots, kQ}$  (see example in Figure A.1 for a particular ordering of the elements of  $\mathbf{W}$  in which the clusters  $E_{\mathbf{x},1}$  and  $E_{\mathbf{x},2}$  appear twice to from the diagonal



and the resulting Fisher Information,

$$\begin{aligned}
J_{\mathbf{W}(d)} &= \frac{(\mathbf{g}^T \mathbf{W} \mathbf{h})^2}{\sigma^2 \mathbf{g}^T \mathbf{W} \mathbf{W}^T \mathbf{g} + \xi^2} \\
&= \frac{\left( \frac{d}{Q} \sum_{k=1}^K \left( \sum_{q=1}^Q g_{kq} \right) \left( \sum_{q=1}^Q h_{kq} \right) \right)^2}{\sigma^2 \frac{d^2}{Q^2} \sum_{k=1}^K Q \left( \sum_{q=1}^Q g_{kq} \right)^2 + \xi^2} \\
&\rightarrow \frac{d^2}{\xi^2} \mathbb{E}^2[\tilde{g}] \mathbb{E}^2[h] \quad \left( \text{Note: } \sum_{q=1}^Q g_{kq} = \mathcal{O}\left(\frac{1}{K}\right) \right). \tag{A.61}
\end{aligned}$$

Eliminating the scalar  $d$  between (A.60) and (A.61) yields the energy-distortion tradeoff in equation (6.15b).

## A.14 Explicit expressions for Rayleigh distributed gains

To derive (6.18), we first note that the square of a Rayleigh distributed random variable is exponential distributed,

$$\begin{aligned}
f_h(h) &= \frac{h}{\alpha^2} \exp\left(-\frac{h^2}{2\alpha^2}\right) \implies f_t(t) = \exp(-t), \\
&\text{for } t \triangleq \frac{h^2}{2\alpha^2}, \quad h > 0, t > 0, \tag{A.62}
\end{aligned}$$

consequently, it suffices to show that for  $T_i \stackrel{\text{i.i.d.}}{\sim} f_t(t)$ ,

$$\mathbb{E} \left[ \frac{1}{1 + \frac{1}{\lambda} (T_1 + \dots + T_Q)} \right] = H_Q \text{ as given by (6.18),} \tag{A.63}$$

where  $\lambda = \frac{\sigma^2}{2\alpha^2\eta^2}$ . We further note that the finite-sum of i.i.d. exponential distributed random variables is Gamma distributed,

$$\begin{aligned}
f_t(t) = \exp(-t), \quad T_i \stackrel{\text{i.i.d.}}{\sim} f_t(t) &\implies f_{s;Q}(s) = \frac{1}{(Q-1)!} \exp(-s) s^{Q-1}, \\
&\text{for } s \triangleq t_1 + \dots + t_Q, \quad t_i > 0, s > 0. \tag{A.64}
\end{aligned}$$

Hence it further suffices to establish that for  $S \sim f_{s;Q}(s)$ ,

$$\mathbb{E} \left[ \frac{1}{1 + \frac{1}{\lambda} S} \right] = H_Q \text{ as given by (6.18).} \quad (\text{A.65})$$

We establish (6.18) through recursion. First note that for  $Q = 1$ ,  $f_{s;1}(s) = \exp(-s)$  and  $H_1$  evaluates to

$$\begin{aligned} H_1 &= \int_0^\infty \frac{\exp(-s)}{1 + \frac{1}{\lambda} s} ds \\ &= \lambda \exp(\lambda) \int_\lambda^\infty \frac{\exp(-s')}{s'} ds' \quad (\text{by substitution } s' = s + \lambda) \\ &= \lambda \exp(\lambda) \mathcal{E}i(\lambda), \quad (\text{by definition of } \mathcal{E}i(\lambda) \text{ in (6.18)}) \end{aligned} \quad (\text{A.66})$$

For  $Q \geq 2$ , we have the following recursion

$$\begin{aligned} H_Q &= \int_0^\infty f_{s;Q}(s) \frac{1}{1 + \frac{s}{\lambda}} ds \\ &= \int_0^\infty \frac{1}{(Q-1)!} \exp(-s) s^{Q-1} \frac{1}{1 + \frac{s}{\lambda}} ds \\ &= \frac{\lambda}{Q-1} \int_0^\infty \frac{1}{(Q-2)!} \exp(-s) s^{Q-2} \frac{\frac{s}{\lambda}}{1 + \frac{s}{\lambda}} ds \\ &= \frac{\lambda}{Q-1} \int_0^\infty f_{s;Q-1}(s) \left( 1 - \frac{1}{1 + \frac{s}{\lambda}} \right) ds \\ &= \frac{\lambda}{Q-1} (1 - H_{Q-1}), \end{aligned} \quad (\text{A.67})$$

which can be further verified to result in (6.18). The first few terms are given below for ease of reference

$$\begin{aligned} H_2 &= \lambda - \lambda H_1, \\ H_3 &= \frac{1}{2} (\lambda - \lambda^2 + \lambda^2 H_1), \\ H_4 &= \frac{1}{6} (2\lambda - \lambda^2 + \lambda^3 - \lambda^3 H_1). \end{aligned} \quad (\text{A.68})$$

## A.15 Proof of Theorem 6.3.2

We will establish equation (6.22) using a result in matrix inverse approximation derived in Chapter 7. Comparing the closely related stationary matrices in Chapter 7 denoted by  $\mathbf{R}$  (see (7.12), this is not to be confused with  $\mathbf{R}_{yy'}$  denoting covariance of  $\mathbf{y}$  in this chapter) with that of  $\mathbf{C}$  given by (6.19), we note that  $\mathbf{R} = \frac{1}{1-\rho^2}\mathbf{C}$ . Hence, combining expressions for  $\mathbf{R}_{yy'}$  from (6.20),  $\mathbb{E}[\theta_t \mathbf{y}']$  from (6.21) and  $\text{Var}(\theta_t | \mathbf{y})$  from (6.12), we write

$$\text{Var}(\theta_t | \mathbf{y}) = \eta^2 - \eta^4 J \underbrace{\begin{bmatrix} \cdots & \rho^2 \varrho & \varrho & \rho/\varrho & \rho^2/\varrho & \cdots \end{bmatrix}}_{\triangleq \mathbf{p}^T \text{ (say)}} \begin{bmatrix} \mathbf{I} + \gamma \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \vdots \\ \rho^2 \varrho \\ \varrho \\ \rho/\varrho \\ \rho^2/\varrho \\ \vdots \end{bmatrix}, \quad (\text{A.69})$$

where  $\gamma \triangleq \eta^2 J (1 - \rho^2)$ ,

and vector  $\mathbf{p}$  is defined as above. The inverse  $(\mathbf{I} + \gamma \mathbf{R})^{-1}$  has an elegant asymptotic approximation (precisely summarized and proved in Proposition 7.3.1) of the form

$$\begin{aligned} (\mathbf{I} + \gamma \mathbf{R})^{-1} &= \mathbf{I} - \nu \mathbf{A} \\ &= \mathbf{I} - \nu (\mathbf{I} + \mathbf{A}_1 + \kappa \mathbf{A}_2 + \kappa \mathbf{A}_3) \end{aligned} \quad (\text{A.70})$$

where  $\nu, \kappa$  are scalars that depends on  $\gamma$  and  $\rho$ , precisely

$$\begin{aligned} y &\triangleq \frac{1}{2} \left[ \frac{\gamma + 1}{\rho} + \rho - \sqrt{\left( \frac{\gamma + 1}{\rho} + \rho \right)^2 - 4} \right], \\ \nu &\triangleq \frac{y\gamma}{\rho(1 - y^2)}, \text{ and } \kappa \triangleq \frac{y(\rho - y)}{1 - \rho y}. \end{aligned} \quad (\text{A.71})$$

and  $\mathbf{A}$  is a matrix composed of several structured matrices  $\mathbf{A}_i$ -s (see Appendix A.17 for the precise structures, we refrain from repeating here). Using (A.69), the rest of the problem



involves finding some quadratic products,

$$\text{Var}(\theta_t|\mathbf{y}) = \eta^2 - \eta^4 J \left\{ (1 - \nu) \mathbf{p}^T \mathbf{p} - \nu \mathbf{p}^T \mathbf{A}_1 \mathbf{p} - \nu \kappa \underbrace{\mathbf{p}^T (\mathbf{A}_2 + \mathbf{A}_3) \mathbf{p}}_{\rightarrow 0} \right\} \quad (\text{A.72})$$

among whom the last term is negligible, due to the following explanation. We illustrate the infinite-dimensional matrix  $\mathbf{p}\mathbf{p}^T$  in Figure A.2, which has finite components in the “center” of the matrix (defined by quadrants  $Q_1, Q_2, Q_3, Q_4$ ) and the elements (exponentially) decrease in magnitude towards the “periphery”. On the other hand, matrices  $\mathbf{A}_2, \mathbf{A}_3$  (see

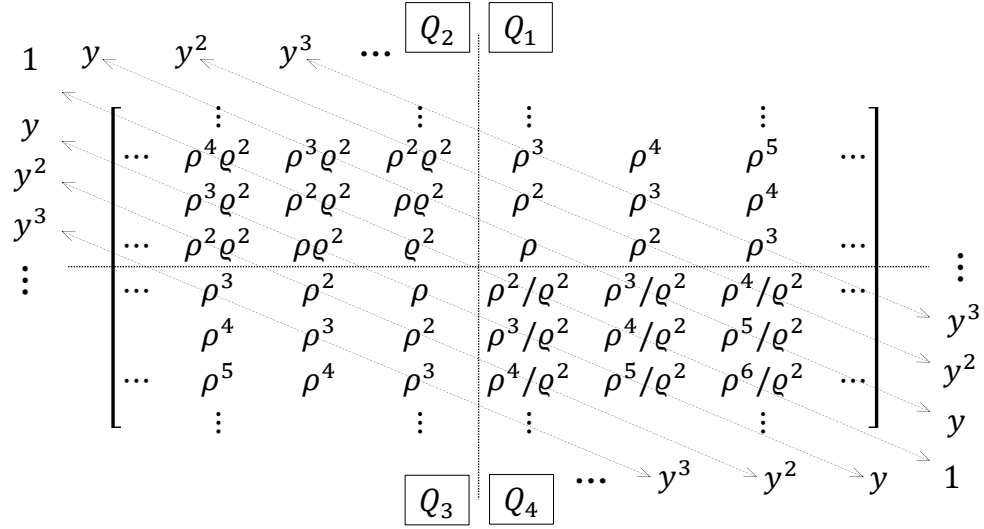


Figure A.2: Illustrating  $\mathbf{p}\mathbf{p}^T$  and the summations along diagonals in various quadrants

Appendix A.17) has finite elements only in the “periphery” and vanishingly small elements towards the “center”. Hence  $\mathbf{p}^T (\mathbf{A}_2 + \mathbf{A}_3) \mathbf{p}$  represents a sum that is vanishingly small and can be ignored. It remains to compute  $\mathbf{p}^T \mathbf{p}$  and  $\mathbf{p}^T \mathbf{A}_1 \mathbf{p}$  to finish the derivation. The product  $\mathbf{p}^T \mathbf{p}$  is obtained by simply adding the diagonal terms in Figure A.2,

$$\mathbf{p}^T \mathbf{p} = \left( \underbrace{\varrho^2}_{Q_2 \text{ diag}} + \underbrace{\frac{\rho^2}{\varrho^2}}_{Q_4 \text{ diag}} \right) \frac{1}{1 - \rho^2}, \quad (\text{A.73})$$

where two separate summations for quadrants  $Q_2$  and  $Q_4$  are indicated above. As regards  $\mathbf{p}^T \mathbf{A}_1 \mathbf{p}$ , we note that  $\mathbf{A}_1$  (Appendix A.17) is a matrix with  $y^i$  along the  $i^{\text{th}}$  off-diagonal

lines. Working out the algebra, we have

$$\mathbf{p}^T \mathbf{A}_1 \mathbf{p} = 2 \times \underbrace{\frac{y\rho}{(1-y\rho)^2}}_{\text{both } Q_1 \text{ and } Q_3 \text{ sums}} + \frac{2y\rho}{1-y\rho} \left( \underbrace{\varrho^2}_{Q_2 \text{ sum}} + \underbrace{\frac{\rho^2}{\varrho^2}}_{Q_4 \text{ sum}} \right) \frac{1}{1-\rho^2}, \quad (\text{A.74})$$

which completes the derivation. It is easy to verify that the sum computed by inserting (A.73) and (A.74) in (A.72), with scalars  $y$  (not to be confused with observation vector  $\mathbf{y}$ ) and  $\nu$  as in (A.71), is algebraically equivalent to equation (6.22).

## A.16 Proof of Proposition 7.3.1

In Proposition 7.3.1, we need to show that

$$(\mathbf{I} + \gamma \mathbf{R})^{-1} = \mathbf{I} - \nu \mathbf{A} + \mathcal{O}(y^T). \quad (\text{A.75})$$

Let  $\tilde{\mathbf{A}}$  be such that

$$(\mathbf{I} + \gamma \mathbf{R})^{-1} = \mathbf{I} - \nu \tilde{\mathbf{A}}, \quad (\text{A.76})$$

i.e., the exact form of Equation (A.75). Then it suffices to prove that

$$\begin{aligned} \mathbf{A} &= \tilde{\mathbf{A}} + \mathcal{O}(y^T), \text{ or equivalently,} \\ \tilde{\mathbf{A}}^{-1} \mathbf{A} &= \mathbf{I} + \mathcal{O}(y^T). \end{aligned} \quad (\text{A.77})$$

To prove (A.77), we need  $\tilde{\mathbf{A}}^{-1}$ . Note that (A.76) implies

$$\tilde{\mathbf{A}}^{-1} = \nu(\mathbf{I} + \mathbf{R}^{-1}/\gamma). \quad (\text{A.78})$$

The inverse of  $\mathbf{R}_U$  (given by (7.12)) is well known [65]. The inverse of  $\mathbf{R}$  is very similar to that of  $\mathbf{R}_U$  except for the top-left element. Specifically,

$$\mathbf{R}^{-1} = \begin{bmatrix} 1 + \varrho_\tau & -\rho & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 + \rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \quad (\text{A.79})$$

From (A.78) and with the help of some identities that follow directly from (7.19), namely,

$$\begin{aligned} \nu(1 + 1/\gamma) &= (1 - \rho y + y^2)/(1 - y^2), \text{ and} \\ \nu(1 + (1 + \rho^2)/\gamma) &= (1 + y^2)/(1 - y^2), \end{aligned} \quad (\text{A.80})$$

we express  $\tilde{\mathbf{A}}^{-1}$  as,

$$\tilde{\mathbf{A}}^{-1} = \frac{1}{1 - y^2} \times \begin{bmatrix} m_\tau & -y & \cdots & 0 & 0 \\ -y & 1 + y^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 + y^2 & -y \\ 0 & 0 & \cdots & -y & 1 - \rho y + y^2 \end{bmatrix}, \quad (\text{A.81})$$

where  $m_\tau \triangleq 1 - \rho y + y^2 + \varrho_\tau y / \rho$ . Constructing  $\mathbf{A}$  from Equation (7.21), we can now directly verify (A.77). Specifically, the residual is of the following structure

$$\tilde{\mathbf{A}}^{-1} \mathbf{A} - \mathbf{I} = \frac{y^T}{r_0} \begin{bmatrix} 0 & 0 & \cdots & r_1 & r_3 \\ 0 & 0 & \ddots & r_2 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ r_1 & r_2 & \ddots & 0 & 0 \\ r_2 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad (\text{A.82})$$

where constants  $r_0, r_1, r_2, r_3$  are defined as

$$\begin{aligned} r_0 &\triangleq (1 - \rho y)(\rho - \rho^2 y + \varrho_\tau y), \quad r_1 \triangleq -\varrho_\tau y \\ r_2 &\triangleq -\rho(\rho - y)(1 - \rho y) + \varrho_\tau(1 + y^2 - \rho y), \\ r_3 &\triangleq r_2 + \varrho_\tau y(\varrho_\tau - \rho^2)/\rho, \end{aligned} \quad (\text{A.83})$$

respectively. Since (A.82) are terms of order  $\mathcal{O}(y^T)$ , this completes the proof of (A.77) and hence Proposition 7.3.1.

## A.17 Polynomial Approximation of $\mathbf{X}'\mathbf{A}\mathbf{X}$

First, we would decompose  $\mathbf{A}$  (see (7.21)) to aid analytical calculations. Note that

$$\mathbf{A} = \mathbf{I} + \mathbf{A}_1 + \eta_\tau \mathbf{A}_2 + \kappa \mathbf{A}_3, \quad (\text{A.84})$$

and matrices  $\mathbf{A}_1, \mathbf{A}_2$  and  $\mathbf{A}_3$  are defined by

$$\mathbf{A}_1 \triangleq \begin{bmatrix} 0 & y & \cdots & y^{T-2} & y^{T-1} \\ y & 0 & \ddots & y^{T-3} & y^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ y^{T-2} & y^{T-3} & \ddots & 0 & y \\ y^{T-1} & y^{T-2} & \cdots & y & 0 \end{bmatrix}, \quad (\text{A.85})$$

$$\mathbf{A}_2 \triangleq \begin{bmatrix} 1 & y & y^2 & \cdots & y^{T-2} & 0 \\ y & y^2 & \ddots & \ddots & 0 & 0 \\ y^2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ y^{T-2} & 0 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, \text{ and} \quad (\text{A.86})$$

$$\mathbf{A}_3 \triangleq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & y^{T-1} \\ 0 & 0 & \ddots & \ddots & y^{T-1} & y^{T-2} \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & y^2 \\ 0 & y^{T-1} & \ddots & \ddots & y^2 & y \\ y^{T-1} & y^{T-2} & \cdots & y^2 & y & 1 \end{bmatrix}, \quad (\text{A.87})$$

respectively. To calculate  $\mathbf{X}'\mathbf{A}\mathbf{X}$ , we compute each of terms in (A.84).  $\mathbf{X}'\mathbf{X}$  is given by (7.25) and (7.26). For  $\mathbf{X}'\mathbf{A}_1\mathbf{X}$ , we refer to (A.85) and collect the identical powers of  $y$  from either side of the principal diagonal,

$$\begin{aligned} [\mathbf{X}'\mathbf{A}_1\mathbf{X}]_{k,l} &= \sum_{i=1}^{T-1} y^i \sum_{j=1}^{T-i} (j^k(i+j)^l + j^l(i+j)^k) \\ &\stackrel{(a)}{=} \sum_{i=1}^{T-1} y^i \sum_{j=1}^{T-i} \sum_{r=0}^{k+l} A_r i^r j^{k+l-r} \\ &= \sum_{i=1}^{T-1} y^i \sum_{r=0}^{k+l} A_r i^r \sum_{j=1}^{T-i} j^{k+l-r} \\ &\stackrel{(b)}{=} \sum_{i=1}^{T-1} y^i \sum_{r=0}^{k+l} A_r i^r \sum_{s=0}^{k+l-r} B_{s,r} (T-i)^{k+l-r+1-s} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(c)}{=} \sum_{i=1}^{T-1} y^i \sum_{r=0}^{q-1} A_r i^r \sum_{s=0}^{q-r-1} B_{s,r} \sum_{t=0}^{q-r-s} C_{t,s,r} T^{q-r-s-t} (-i)^t \\
&\stackrel{(d)}{=} \sum_{i=1}^{T-1} y^i \sum_{r=0}^{q-1} \sum_{s=0}^{q-r-1} \sum_{t=0}^{q-r-s} X_{r,s,t} \\
&\stackrel{(e)}{=} \sum_{i=1}^{T-1} y^i \left( \sum_{u=0}^{q-1} \sum_{v=0}^u \sum_{r=0}^v X_{r,u-v,v-r} + \sum_{v=1}^q \sum_{r=0}^{v-1} X_{r,q-v,v-r} \right) \\
&\stackrel{(f)}{=} \sum_{i=1}^{T-1} y^i \left( \sum_{u=0}^{q-1} T^{q-u} \sum_{v=0}^u i^v \sum_{r=0}^v D_{r,v,u} + \sum_{v=1}^q i^v \sum_{r=0}^{v-1} D_{r,v,q} \right) \\
&\stackrel{(g)}{=} \sum_{u=0}^{q-1} T^{q-u} \underbrace{\sum_{v=0}^u Y_v \sum_{r=0}^v D_{r,v,u}}_{\triangleq \mathcal{A}_{u,M_1}, u \leq k+l} + \underbrace{\sum_{v=1}^q Y_v \sum_{r=0}^{v-1} D_{r,v,q}}_{\triangleq \alpha_{M_1}} + \mathcal{O}(T^q y^T), \tag{A.88}
\end{aligned}$$

where in step (a) we collect all like powers of  $i$  and  $j$  and define  $A_r \triangleq \binom{l}{r} + \binom{k}{r}$ , (b) follows from the summation formula in (7.26) and the definition  $B_{s,r} \triangleq \mathcal{B}_{k+l-r,s}$ , (c) is due to binomial expansion and definitions  $q \triangleq k + l + 1$  and  $C_{t,s,r} \triangleq \binom{q-r-s}{t}$ . Step (d) follows from the definition  $X_{r,s,t} \triangleq i^{r+t} T^{q-r-s-t} A_r B_{s,r} C_{t,s,r} (-1)^t$ , step (e) is an identity involving series rearrangement (i.e., true for any  $X_{r,s,t}$ , by defining  $u \triangleq r + s + t$  and  $v \triangleq r + t$ ), step (f) follows from defining  $D_{r,v,u} \triangleq A_r B_{u-v,r} C_{v-r,u-v,r} (-1)^{v-r}$ . Step (g) follows from the definition  $Y_v \triangleq \sum_{i=1}^{\infty} i^v y^i$  (which converges for finite  $v$  since  $y < 1$  and the error term is of the order  $T^v y^T$ ). Quantities of the form  $Y_v$  are also known as poly-logarithms. In the last step, we denote the polynomial coefficients by  $\mathcal{A}_{u,M_1}$  and the constant term by  $\alpha_{M_1}$ .

For  $\mathbf{X}' \mathbf{A}_2 \mathbf{X}$ , we start with (A.86) and collect the identical powers of  $y$  from the top-left half,

$$\begin{aligned}
[\mathbf{X}' \mathbf{A}_2 \mathbf{X}]_{k,l} &= \sum_{i=1}^{T-1} y^{i-1} \sum_{j=1}^i j^k (i+1-j)^l \\
&= \underbrace{\sum_{i=1}^{\infty} y^{i-1} \sum_{j=1}^i j^k (i+1-j)^l}_{\triangleq \alpha_{M_2}} + \mathcal{O}(T^q y^T), \tag{A.89}
\end{aligned}$$

where  $\alpha_{M_2}$  is a constant.

Similarly, for  $\mathbf{X}'\mathbf{A}_3\mathbf{X}$ , we start with (A.87) and collect the identical powers of  $y$  from the bottom-right half,

$$\begin{aligned}
& [\mathbf{X}'\mathbf{A}_3\mathbf{X}]_{k,l} \\
&= \sum_{i=1}^{T-1} y^{i-1} \sum_{j=1}^i (T-j+1)^k (T+j-i)^l \\
&\stackrel{(a)}{=} \sum_{i=1}^{T-1} y^{i-1} \sum_{j=1}^i \sum_{r=0}^k G_r T^{k-r} (-j+1)^r \sum_{s=0}^l H_s T^{l-s} (j-i)^s \\
&\stackrel{(b)}{=} \sum_{i=1}^{T-1} y^{i-1} \sum_{j=1}^i \sum_{t=0}^{k+l} T^{k+l-t} \sum_{s=0}^t K_{t,s} \\
&\stackrel{(c)}{=} \sum_{t=0}^{k+l} T^{k+l-t} \mathcal{A}_{t,M_3} + \mathcal{O}(T^q y^T), \tag{A.90}
\end{aligned}$$

where (a) follows from binomial expansion and definitions  $G_r \triangleq \binom{k}{r}$ ,  $H_s \triangleq \binom{l}{s}$ . In (b), we have rearranged the sum to group similar exponents of  $T$  with the transformation  $t \triangleq r+s$  and defined  $K_{t,s} \triangleq G_{t-s}(-j+1)^{t-s} H_s(j-i)^s$ , (c) follows from similar arguments while deriving (A.88) and (A.89) and the definition

$$\mathcal{A}_{t,M_3} \triangleq \sum_{i=1}^{\infty} y^{i-1} \sum_{j=1}^i \sum_{s=0}^t K_{t,s}, \quad t \leq k+l. \tag{A.91}$$

We note that all the constants  $\mathcal{A}_{u,M_1}, \alpha_{M_1}, \alpha_{M_2}, \mathcal{A}_{t,M_3}$  depend on  $k$  and  $l$ . We can now prove Proposition 7.3.2 by composing  $\mathbf{X}'\mathbf{A}\mathbf{X}$  with the help of (A.88), (A.89) and (A.90) in (A.84). Define  $\mathcal{A}_{-1,M_3} = 0$ . The constants  $\mathcal{A}_{k,l,i}$  and  $\alpha_{k,l}^{(\tau)}$  in Proposition 7.3.2 are given by

$$\begin{aligned}
\mathcal{A}_{k,l,i} &= \mathcal{B}_{k+l,i} + \mathcal{A}_{i,M_1} + \kappa \mathcal{A}_{i-1,M_3}, \quad 0 \leq i \leq k+l, \\
\alpha_{k,l}^{(\tau)} &= \alpha_{M_1} + \eta_{\tau} \alpha_{M_2} + \kappa \mathcal{A}_{k+l,M_3},
\end{aligned} \tag{A.92}$$

respectively. Some of these constants are enumerated in Table A.1. This completes the proof. For example, to compute  $\mathcal{A}_{k,l,0}$ , we note that  $\mathcal{B}_{k+l,0} = \frac{1}{k+l+1}$ ,  $\mathcal{A}_{0,M_1} = \frac{2Y_0}{k+l+1}$ ,  $Y_0 = \frac{y}{1-y}$

$i \downarrow$	$\mathcal{B}_{k+l,i}$	$\mathcal{A}_{i,M_1}$	$\mathcal{A}_{i-1,M_3}$	$Y_i$
0	$\frac{1}{k+l+1}$	$\frac{2}{k+l+1}Y_0$	0	$\frac{y}{1-y}$
1	$\frac{1}{2}$	$Y_0 - Y_1$	$\frac{1}{(1-y)^2}$	$\frac{y}{(1-y)^2}$
2	$\frac{k+l}{12}$	$\frac{k+l}{6}Y_0 - \frac{k+l}{2}Y_1 + \frac{\binom{k}{2}+\binom{l}{2}}{k+l-1}Y_2$	$-\frac{(k+l)y}{(1-y)^3}$	$\frac{y(1+y)}{(1-y)^3}$
3	0	$-\frac{\binom{k+l}{2}}{6}Y_1 + \frac{\binom{k}{2}+\binom{l}{2}}{2}Y_2 - \frac{\binom{k}{2}+\binom{l}{2}-\frac{kl}{2}}{3}Y_3$	$\frac{y((1+y)(\binom{k}{2}+\binom{l}{2})+kly)}{(1-y)^4}$	$\frac{y(1+4y+y^2)}{(1-y)^4}$

$(k, l) \downarrow$	$\alpha_{M_1}$	$\alpha_{M_2}$	$\mathcal{A}_{k+l,M_3}$
(0, 0)	$-\frac{2y}{(1-y)^2}$	$\frac{1}{(1-y)^2}$	$\frac{1}{(1-y)^2}$
(0, 1), (1, 0)	$-\frac{y}{(1-y)^2}$	$\frac{1}{(1-y)^3}$	$-\frac{y}{(1-y)^3}$
(1, 1)	$\frac{2y^2}{(1-y)^4}$	$\frac{1}{(1-y)^4}$	$\frac{y^2}{(1-y)^4}$

Table A.1: Values of several constants helpful to derive  $\mathcal{A}_{k,l,i}$  and  $\alpha_{k,l}^{(\tau)}$ .

which means

$$\mathcal{A}_{k,l,0} = \mathcal{B}_{k+l,0} + Y_0 \mathcal{A}_{0,M_1} = \frac{1+y}{1-y} \frac{1}{k+l+1}, \quad (\text{A.93})$$

which agrees with (7.28).

## A.18 FIM and CRB for $\rho = 1$

First we derive the FIM given by Equation (7.39). Let the Taylor Series of the FIM be of the form

$$\mathbf{J} = \frac{1}{\sigma^2} \mathbf{E} \left[ \mathbf{Q}_0 + \frac{1}{T} \mathbf{Q}_1 + \frac{1}{T^2} \mathbf{Q}_2 + \mathcal{O}\left(\frac{1}{T^3}\right) \right] \mathbf{E}. \quad (\text{A.94})$$

The first term  $\mathbf{Q}_0$  is equal to  $\xi_2 \mathbf{f} \mathbf{f}'$  (follows from (7.29) and the fact that  $\xi_0 = \xi_1 = 0$  for  $\rho = 1$ ). The second order term  $\mathbf{Q}_1$  follows from collecting third-order terms in (7.25) and (7.27) and adding them according to (7.24), i.e.,

$$[\mathbf{Q}_1]_{k,l} = \begin{cases} 0, & k = l = 0, \\ -\nu \alpha_{k,l}^{(\tau)}, & k + l = 1, \\ \mathcal{B}_{k+l,2} - \nu \mathcal{A}_{k,l,2}, & k + l \geq 2. \end{cases} \quad (\text{A.95})$$



From definitions (7.19) and identity (7.23), we obtain the following simplifications for  $\rho = 1$ ,

$$\nu = \frac{1-y}{1+y}, \quad \kappa = y, \quad \text{and} \quad \frac{1}{\gamma} = \frac{y}{(1-y)^2}, \quad (\text{A.96})$$

which when applied to (A.92) (and using values from Table A.1) leads to,

$$[\mathbf{Q}_1]_{k,l} = \frac{kl}{k+l-1} \frac{1}{\gamma}, \quad k+l \geq 2, \quad (\text{A.97})$$

which explains the term  $\xi_4$  in (7.40). Similarly,  $\mathbf{Q}_2$  follows from collecting fourth-order terms,

$$[\mathbf{Q}_2]_{k,l} = \begin{cases} 0, & k+l \leq 1, \\ -\nu \alpha_{k,l}^{(\tau)}, & k+l = 2, \\ \mathcal{B}_{k+l,3} - \nu \mathcal{A}_{k,l,3}, & k+l \geq 3, \end{cases} \quad (\text{A.98})$$

the last term of which when simplified yields

$$[\mathbf{Q}_2]_{k,l} = -kl \frac{y^2}{(1-y)^3}, \quad k+l \geq 3. \quad (\text{A.99})$$

that explains the term  $\xi_5$  in (7.40).

The derivation of the CRLB in Equation (7.41) follows from the block inversion of  $\mathbf{J}$  and subsequent application of Lemma 7.3.2.1. The block inversion formula is

$$\begin{aligned} \begin{bmatrix} a & \mathbf{b}' \\ \mathbf{b} & \mathbf{C} \end{bmatrix}^{-1} &= \begin{bmatrix} \frac{1}{a} + \frac{\mathbf{b}'[\mathbf{C} - \frac{1}{a}\mathbf{b}\mathbf{b}']^{-1}\mathbf{b}}{a^2} & * \\ * & [\mathbf{C} - \frac{1}{a}\mathbf{b}\mathbf{b}']^{-1} \end{bmatrix} \\ &\triangleq \begin{bmatrix} m & * \\ * & \mathbf{A} \end{bmatrix}, \text{ say.} \end{aligned} \quad (\text{A.100})$$

From (7.39), we have

$$\begin{aligned} a &= \xi_2^{(\tau)}, \quad \mathbf{b} = \frac{1}{T} \xi_3^{(\tau)} \mathbf{f}' + \mathcal{O}\left(\frac{1}{T^2}\right), \\ \mathbf{C} &= \frac{1}{T} \mathbf{D} \left[ \xi_4 \mathbf{H} + \frac{1}{T} \left( \xi_5 \mathbf{e} \mathbf{e}' + \xi_6^{(\tau)} \mathbf{f} \mathbf{f}' \right) + \mathcal{O}\left(\frac{1}{T^2}\right) \right] \mathbf{D}, \end{aligned} \quad (\text{A.101})$$

where we have dropped the subscript  $P$  from  $\mathbf{H}_P, \mathbf{e}_P$  and  $\mathbf{f}_P$ . Next we apply Lemma 7.3.2.1 to obtain the diagonal elements  $[\mathbf{A}]_{q,q}$  of block  $\mathbf{A}$ . By substituting  $c_0 = \xi_4$ ,  $c_1 = \xi_5$ ,  $c_2 = \xi_6^{(\tau)} - \left(\xi_3^{(\tau)}\right)^2 / \xi_2^{(\tau)}$ , and observing the facts that  $[\mathbf{D}^{-1}]_{p,p} = \frac{1}{p}$  and  $L_{P-1,p-1} = P^2/p^2$ , we obtain  $[\mathbf{V}]_{p,p}$  for  $p \geq 1$  as in (7.41). For the top-left element,  $m$ , the second order term is

$$\frac{\mathbf{b}' \mathbf{A} \mathbf{b}}{a^2} = \frac{\left(\xi_3^{(\tau)}\right)^2}{\left(\xi_2^{(\tau)}\right)^2} \mathbf{A}_{1,1}, \quad \text{but } \mathbf{A}_{1,1} = \frac{K_{P-1,0}}{T \xi_4} + \mathcal{O}\left(\frac{1}{T^2}\right), \quad (\text{A.102})$$

which completes the derivation of (7.41).

# Bibliography

- [1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, fifth edition, 1964.
- [2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *Communications Magazine, IEEE*, 40(8):102–114, 2002.
- [3] T.C. Aysal and K.E. Barner. Constrained decentralized estimation over noisy channels for sensor networks. *Signal Processing, IEEE Transactions on*, 56(4):1398–1410, April 2008.
- [4] I. Bahceci and A. Khandani. Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation. *Communications, IEEE Transactions on*, 56(7):1146–1156, July 2008.
- [5] Y. Bar-Shalom. Airborne GMTI radar position bias estimation using static-rotator targets of opportunity. *Aerospace and Electronic Systems, IEEE Transactions on*, 37(2):695–699, Apr. 2001.
- [6] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley-Interscience, 2001.
- [7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [8] M. Cetin, L. Chen, J.W. Fisher III, A.T. Ihler, R.L. Moses, M.J. Wainwright, and A.S. Willsky. Distributed fusion in sensor networks - A graphical models perspective. *Signal Processing Magazine, IEEE*, 23(4):42–55, July 2006.
- [9] H. Chen and P. K. Varshney. Nonparametric one-bit quantizers for distributed estimation. *Signal Processing, IEEE Transactions on*, 58(7):3777–3787, July 2010.
- [10] H. Chen and P. K. Varshney. Performance limit for distributed estimation systems with identical one-bit quantizers. *Signal Processing, IEEE Transactions on*, 58(1):466–471, Jan. 2010.
- [11] D.-M. Chiu and R. Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN Systems*, 17(1):1– 14, 1989.
- [12] J. Choi. Distributed beamforming using a consensus algorithm for cooperative relay networks. *Communications Letters, IEEE*, 15(4):368–370, April 2011.

- [13] C.Y. Chong and S.P. Kumar. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91(8):1247–1256, Aug. 2003.
- [14] S. Cui, J.-J. Xiao, A.J. Goldsmith, Z.-Q. Luo, and H.V. Poor. Estimation diversity and energy efficiency in distributed sensing. *Signal Processing, IEEE Transactions on*, 55(9):4683–4695, Sept. 2007.
- [15] M. P. Dana. Registration: A prerequisite for multiple sensor tracking. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*, chapter 5, pages 155–185. Artech House, Norwood, MA, 1990.
- [16] D. P. Dee and R. Todling. Data assimilation in the presence of forecast bias: The GEOS moisture analysis. *Monthly Weather Review*, 128(9):3268–3282, Sept. 2000.
- [17] I. S. Dhillon and J. A. Tropp. Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1120–1146, 2007.
- [18] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, Nov. 2010.
- [19] D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world with wireless sensor networks. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 4, pages 2033–2036, 2001.
- [20] J. Fang and H. Li. Power constrained distributed estimation with cluster-based sensor collaboration. *Wireless Communications, IEEE Transactions on*, 8(7):3822–3832, July 2009.
- [21] J. Fang and H. Li. Distributed estimation of Gauss-Markov random fields with one-bit quantized data. *Signal Processing Letters, IEEE*, 17(5):449–452, May 2010.
- [22] J. Fang and H. Li. Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios. *Signal Processing, IEEE Transactions on*, 58(8):4339–4353, Aug. 2010.
- [23] W. J. Fisk. A pilot study of the accuracy of CO<sub>2</sub> sensors in commercial buildings. *Proceedings of IAQ, Healthy and Sustainable Buildings, Baltimore, Maryland*, 2007.
- [24] N. M. Freris, H. Kowshik, and P. R. Kumar. Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation. *Proceedings of the IEEE*, 98(11):1828–1846, Nov. 2010.
- [25] R. F. Galbraith and J. I. Galbraith. On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability*, 11(1):63–71, 1974.
- [26] M. Gastpar. To code or not to code. *Thèse EPFL, No 2687*, 2002. IC School of Computer and Communication Sciences, École polytechnique fédérale de Lausanne EPFL, Ph.D.

- [27] M. Gastpar. Information-theoretic bounds on sensor network performance. In Ananthram Swami, Qing Zhao, Yao-Win Hong, and Lang Tong, editors, *Wireless Sensor Networks: Signal Processing and Communications Perspectives*, chapter 2, pages 9–41. John Wiley & Sons Ltd, 2007.
- [28] M. Gastpar. Uncoded transmission is exactly optimal for a simple gaussian sensor network. *Information Theory, IEEE Transactions on*, 54(11):5247–5251, Nov. 2008.
- [29] M. Gastpar, P.L. Dragotti, and M. Vetterli. The distributed Karhunen-Loève transform. *Information Theory, IEEE Transactions on*, 52(12):5177–5196, Dec. 2006.
- [30] M. Gastpar, B. Rimoldi, and M. Vetterli. To code, or not to code: Lossy source-channel communication revisited. *Information Theory, IEEE Transactions on*, 49(5):1147–1158, May 2003.
- [31] A. Goldsmith. *Wireless Communications*. Cambridge Univ. Press., Cambridge [u.a.], 2005.
- [32] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger. *Table of Integrals, Series and Products*. Elsevier, Amsterdam, 2007.
- [33] R.M. Gray and Jr. Stockham, T.G. Dithered quantizers. *Information Theory, IEEE Transactions on*, 39(3):805–812, May 1993.
- [34] M. Guerriero, S. Marano, V. Matta, and P. Willett. Optimal relay function in the low-power regime for distributed estimation over a MAC. *Signal Processing, IEEE Transactions on*, 58(5):2907–2911, May 2010.
- [35] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Zhi-Quan Luo. Distributed beamforming for relay networks based on second-order statistics of the channel state information. *Signal Processing, IEEE Transactions on*, 56(9):4306–4316, Sept. 2008.
- [36] Y. Hayashi and R. Matsuda. Deductive prediction of measurement precision from signal and noise in liquid chromatography. *Analytical Chemistry*, 66(18):2874–2881, 1994.
- [37] G. Hazel, F. Bucholtz, and I. D. Aggarwal. Characterization and modeling of drift noise in fourier transform spectroscopy: Implications for signal processing and detection limits. *Applied Optics*, 36(27):6751–6759, 1997.
- [38] M. Holmberg and T. Artursson. Drift compensation, standards, and calibration methods. In T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, editors, *Handbook of Machine Olfaction: Electronic Nose Technology*, chapter 13, pages 325–346. Wiley, Weinheim, FRG, 2004.
- [39] C. G. Jakobson, M. Feinsod, and Y. Nemirovsky. Low frequency noise and drift in Ion Sensitive Field Effect Transistors. *Sensors and Actuators B: Chemical*, 68(1-3):134–139, 2000.
- [40] S. K. Jayaweera and C. Mosquera. Distributed sequential estimation with noisy, correlated observations. *Signal Processing Letters, IEEE*, 15:741–744, 2008.

- [41] Yindi Jing and H. Jafarkhani. Network beamforming using relays with perfect channel information. *Information Theory, IEEE Transactions on*, 55(6):2499–2517, June 2009.
- [42] S. Kar, H. Chen, and P. K. Varshney. Optimal identical binary quantizer design for distributed estimation. *Signal Processing, IEEE Transactions on*, 60(7):3896–3901, July 2012.
- [43] S. Kar and P. K. Varshney. Linear coherent estimation with spatial collaboration. *to appear in IEEE Transactions on Information Theory*.
- [44] S. Kar and P. K. Varshney. Controlled collaboration for linear coherent estimation in wireless sensor networks. In *Proc. 50th Annual Allerton Conference on Communication, Control and Computing 2012*, Monticello, IL, Oct. 2012.
- [45] S. Kar and P. K. Varshney. On linear coherent estimation with spatial collaboration. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1448–1452, July 2012.
- [46] S. Kar, P. K. Varshney, and H. Chen. Spatial whitening framework for distributed estimation. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 Proc. 4th IEEE International Workshop on*, pages 293–296, Dec. 2011.
- [47] S. Kar, P. K. Varshney, and M. Palaniswami. Cramér-Rao bounds for polynomial signal estimation using sensors with AR(1) drift. *Signal Processing, IEEE Transactions on*, 60(10):5494–5507, Oct. 2012.
- [48] K. Kastella, B. Yeary, T. Zadra, R. Brouillard, and E. Frangione. Bias modelling and estimation for GMTI applications. *Proc. 3rd Intl. Conf. Information Fusion, Paris, France*, 2000.
- [49] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [50] N. Khajehnouri and A. H. Sayed. Distributed MMSE relay strategies for wireless sensor networks. *Signal Processing, IEEE Transactions on*, 55(7):3336–3348, July 2007.
- [51] A. Krasnoperov, J.-J. Xiao, and Z.-Q. Luo. Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises. *EURASIP J. on Wireless Communications and Networking*, 4:473–482, 2005.
- [52] J. Li. Distributed estimation in resource-constrained wireless sensor networks. *ProQuest Dissertations and Theses*, 2008. Electrical engineering, Georgia Institute of Technology, Ph.D.
- [53] J. Li and G. AlRegib. Rate-constrained distributed estimation in wireless sensor networks. *Signal Processing, IEEE Transactions on*, 55(5):1634–1643, May 2007.
- [54] J. Li and G. AlRegib. Distributed estimation in energy-constrained wireless sensor networks. *Signal Processing, IEEE Transactions on*, 57(10):3746–3758, Oct. 2009.

- [55] J. Li, A. P. Petropulu, and H. V. Poor. Cooperative transmission for relay networks based on second-order statistics of channel state information. *Signal Processing, IEEE Transactions on*, 59(3):1280–1291, March 2011.
- [56] T. Li, H. Zhou, and O. M. Collins. Estimating the polynomial channel. *International Symposium on Information Theory and its Applications, Parma, Italy*, 2004.
- [57] Z.-Q. Luo. An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network. *Selected Areas in Communications, IEEE Journal on*, 23(4):735–744, April 2005.
- [58] Z.-Q. Luo. Universal decentralized estimation in a bandwidth constrained sensor network. *Information Theory, IEEE Transactions on*, 51(6):2210–2219, June 2005.
- [59] Z.-Q. Luo and Tsung-Hui Chang. SDP relaxation of homogeneous quadratic optimization: Approximation bounds and applications. In Daniel P. Palomar and Yonina C. Eldar, editors, *Convex Optimization in Signal Processing and Communications*, chapter 4, pages 117–165. Cambridge University Press, 2009.
- [60] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley, New York, 1999.
- [61] D. S. Mitrinović, J. E. Pečarić, and A. M. Fink. *Classical and New Inequalities in Analysis*. Dordrecht, Kluwer Academic Publishers, 1993.
- [62] R. Mudumbai, D. R. Brown, U. Madhow, and H. V. Poor. Distributed transmit beamforming: Challenges and recent progress. *Communications Magazine, IEEE*, 47(2):102–110, Feb. 2009.
- [63] S. Nadarajah. A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694, 2005.
- [64] R. Niu and P. K. Varshney. Target location estimation in sensor networks with quantized data. *Signal Processing, IEEE Transactions on*, 54(12):4519–4528, Dec. 2006.
- [65] Ruixin Niu and P. K. Varshney. Sampling schemes for sequential detection with dependent observations. *Signal Processing, IEEE Transactions on*, 58(3):1469–1481, Mar. 2010.
- [66] E. Oka. Long-term sensor drift found in recovered argo profiling floats. *Journal of Oceanography*, 61(4):775–781, 2005.
- [67] O. Ozdemir. Distributed estimation in wireless sensor networks: Physical layer considerations. *ProQuest Dissertations and Theses*, 2009. Electrical engineering, Syracuse University, Ph.D.
- [68] V. Y. Pan and Z. Q. Chen. The complexity of the matrix eigenproblem. In *Proc. ACM symposium on Theory of computing (STOC 99)*, pages 507–516, Atlanta, Georgia, 1999.

- [69] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim. Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *Information Theory, IEEE Transactions on*, 47(3):978–1002, Mar. 2001.
- [70] T. C. Pearce and J. W. Gardner. Predicting organoleptic scores of sub-ppm flavour notes: Part 2. computational analysis and results. *Analyst*, 123:2057–2066, 1998.
- [71] S. Peleg and B. Porat. The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. *Signal Processing, IEEE Transactions on*, 39(3):749–752, Mar 1991.
- [72] J. G. Proakis. *Digital Communications*. Boston, McGraw-Hill, 2001.
- [73] A. Ribeiro and G. B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case. *Signal Processing, IEEE Transactions on*, 54(3):1131–1143, 2006.
- [74] A. Ribeiro and G. B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-Part II: unknown probability density function. *Signal Processing, IEEE Transactions on*, 54(7):2784–2796, July 2006.
- [75] S. M. Seong, J. G. Lee, and C. G. Park. Equivalent ARMA model representation for RLG random errors. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(1):286–290, Jan. 2000.
- [76] L. J. Skidmore. Discrete and continuous estimation in correlated noise with finite observation time. *Dissertation (Ph.D.), California Institute of Technology.*, 1964.
- [77] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999. Version 1.3 available from <http://sedumi.ie.lehigh.edu/>.
- [78] A. Swami. Cramer-Rao bounds for deterministic signals in additive and multiplicative noise. *Signal Processing*, 53(2-3):231–244, 1996.
- [79] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, and M. Palaniswami. Online drift correction in wireless sensor networks using spatio-temporal modeling. *Proc. 11th Intl. Conf. Information Fusion, Cologne, Germany*, pages 158–165, 2008.
- [80] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, and M. Palaniswami. Spatio-temporal modelling-based drift-aware wireless sensor networks. *Wireless Sensor Systems, IET*, 1(2):110–122, June 2011.
- [81] H. L. V. Trees and K. L. Bell. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering / Tracking (ed.)*. Piscataway, N.J., IEEE ; Wiley-Interscience, 2007.
- [82] W. Tsujita, H. Ishida, and T. Moriizumi. Dynamic gas sensor network for air pollution monitoring and its auto-calibration. In *Sensors, 2004. Proceedings of IEEE*, pages 56–59, Oct. 2004.



- [83] M. O. Ulfarsson and V. Solo. Vector  $l_0$  sparse variable PCA. *Signal Processing, IEEE Transactions on*, 59(5):1949–1958, May 2011.
- [84] M. Unser. Splines: A perfect fit for signal and image processing. *Signal Processing Magazine, IEEE*, 16(6):22–38, Nov. 1999.
- [85] L. Vandenberghe and S. Boyd. Semidefinite Programming. *SIAM Review*, 38(1):49–95, 1996.
- [86] L. Vandenberghe, S. Boyd, and S. P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.
- [87] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer, New York, 1997.
- [88] S. Venkatesan, L. Mailaender, and J. Salz. An iterative algorithm for computing a spatial whitening filter. In *Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on*, July 2004.
- [89] P. Venkitasubramaniam, L. Tong, and A. Swami. Quantization for maximin ARE in distributed estimation. *Signal Processing, IEEE Transactions on*, 55(7):3596–3605, July 2007.
- [90] M. Vetterli, P. Marziliano, and T. Blu. A sampling theorem for periodic piecewise polynomial signals. In *IEEE Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing, 2001. (ICASSP '01)*, pages 3893–3896, 2001.
- [91] G. Wang, J. Cao, H. Wang, and M. Guo. Polynomial regression for data gathering in environmental monitoring applications. In *Global Telecommunications Conf., IEEE Globecom*, pages 1307–1311, 2007.
- [92] G. A. Washer, P. A. Fuchs, M. K. Masterson, and C. Phillips. Long-term structural health monitoring for bridge piers. *ASNT Research Symposium and Spring Conference, Anaheim, CA*, Mar 31–Apr 3, 2008.
- [93] G. Wei. Estimations of frequency and its drift rate. *Instrumentation and Measurement, IEEE Transactions on*, 46(1):79–82, Feb. 1997.
- [94] T. Wu and Q. Cheng. Distributed estimation over fading channels using one-bit quantization. *Wireless Communications, IEEE Transactions on*, 8(12):5779–5784, Dec. 2009.
- [95] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A.J. Goldsmith. Power scheduling of universal decentralized estimation in sensor networks. *Signal Processing, IEEE Transactions on*, 54(2):413–422, Feb. 2006.
- [96] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A.J. Goldsmith. Linear coherent decentralized estimation. *Signal Processing, IEEE Transactions on*, 56(2):757–770, Feb. 2008.
- [97] Y. V. Zakharov, T. C. Tozer, and J. F. Adlard. Polynomial spline-approximation of Clarke’s model. *Signal Processing, IEEE Transactions on*, 52(5):1198–1208, May 2004.

- [98] G. Zheng, K.-K. Wong, A. Paulraj, and B. Ottersten. Collaborative-relay beamforming with perfect CSI: Optimum and distributed implementation. *Signal Processing Letters, IEEE*, 16(4):257–260, April 2009.
- [99] Y. Zhou, H. Leung, and P. C. Yip. An exact maximum likelihood registration algorithm for data fusion. *Signal Processing, IEEE Transactions on*, 45(6):1560–1573, June 1997.
- [100] Z. Zhu and Y. Liu. Estimating spatial covariance using penalised likelihood with weighted  $L_1$  penalty. *Journal of Nonparametric Statistics*, 21(7):925–942, 2009.
- [101] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.
- [102] D. Zupanski and M. Zupanski. Model error estimation employing an ensemble data assimilation approach. *Monthly Weather Review*, 134(5):1337–1354, May 2006.

# Vita

NAME OF AUTHOR: Swarnendu Kar

PLACE OF BIRTH: Midnapore, West Bengal, India

DATE OF BIRTH: April 03, 1982

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, USA, 2006–2013

Indian Institute of Technology, Kharagpur, India, 2000–2004

DEGREES AWARDED:

M.S. Mathematics, 2009, Syracuse University, USA

B.Tech. Electronics and Electrical Comm. Engg, 2004, IIT Kharagpur, India

PROFESSIONAL EXPERIENCE:

- DSP Engineer (Video codecs), Ittiam Systems, Bangalore, 2004 – 2006
- Smartphone Engg Intern, Intel Corporation, Hillsboro OR, Sept 2012 – April 2013