Distributed Inference using Bounded Transmissions

by

Sivaraman Dasarathan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2013 by the
Graduate Supervisory Committee:

Cihan Tepedelenlioğlu, Chair
Antonia Papandreou-Suppappola
Martin Reisslein
Michael Goryll

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Distributed inference has applications in a wide range of fields such as source localization, target detection, environment monitoring, and healthcare. In this dissertation, distributed inference schemes which use bounded transmit power are considered. The performance of the proposed schemes are studied for a variety of inference problems.

In the first part of the dissertation, a distributed detection scheme where the sensors transmit with constant modulus signals over a Gaussian multiple access channel is considered. The deflection coefficient of the proposed scheme is shown to depend on the characteristic function of the sensing noise, and the error exponent for the system is derived using large deviation theory. Optimization of the deflection coefficient and error exponent are considered with respect to a transmission phase parameter for a variety of sensing noise distributions including impulsive ones. The proposed scheme is also favorably compared with existing amplify-and-forward (AF) and detect-and-forward (DF) schemes. The effect of fading is shown to be detrimental to the detection performance and simulations are provided to corroborate the analytical results.

The second part of the dissertation studies a distributed inference scheme which uses bounded transmission functions over a Gaussian multiple access channel. The conditions on the transmission functions under which consistent estimation and reliable detection are possible is characterized. For the distributed estimation problem, an estimation scheme that uses bounded transmission functions is proved to be strongly consistent provided that the variance of the noise samples are bounded and that the transmission function

is one-to-one. The proposed estimation scheme is compared with the amplify and forward technique and its robustness to impulsive sensing noise distributions is highlighted. It is also shown that bounded transmissions suffer from inconsistent estimates if the sensing noise variance goes to infinity. For the distributed detection problem, similar results are obtained by studying the deflection coefficient. Simulations corroborate our analytical results.

In the third part of this dissertation, the problem of estimating the average of samples distributed at the nodes of a sensor network is considered. A distributed average consensus algorithm in which every sensor transmits with bounded peak power is proposed. In the presence of communication noise, it is shown that the nodes reach consensus asymptotically to a finite random variable whose expectation is the desired sample average of the initial observations with a variance that depends on the step size of the algorithm and the variance of the communication noise. The asymptotic performance is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. It is shown that using bounded transmissions results in slower convergence compared to the linear consensus algorithm based on the Laplacian heuristic. Simulations corroborate our analytical findings.

Finally, a robust distributed average consensus algorithm in which every sensor performs a nonlinear processing at the receiver is proposed. It is shown that non-linearity at the receiver nodes makes the algorithm robust to a wide range of channel noise distributions including the impulsive ones. It is shown that the nodes reach consensus asymptotically and similar results are obtained as in the case of transmit non-linearity. Simulations corroborate our analytical findings and highlight the robustness of the proposed algorithm.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere thanks and appreciation to my advisor, Professor Cihan Tepedelenlioğlu, whose valuable guidance and support were essential for me to accomplish this work. In addition to academics, interaction with him has helped me develop several other skills which will be useful throughout my life.

I would like to extend my thanks to all members of my committee, Professors Antonia Papandreou-Suppappola, Martin Reisslein and Michael Goryll. I would also take this opportunity to thank all the faculty members from whom I have learned, including, but not limited to Professors Tolga Duman, Douglas Cochran and Junshan Zhang. This work would not have been possible without those course-works as the building blocks for my research.

I am also thankful to Professor Joseph Palais, and Clayton Javurek for offering me financial assistantship throughout my graduate study. Thanks to all the staff members of Electrical Engineering department; Darleen Mandt, Esther Korner, Jenna Marturano, and Donna Rosenlof, to name just a few, for their extraordinary kindness and infinite patience, they helped me every time I visited them.

I express my thanks and appreciations to Mr.Mark Rentz, Jessica Jensen, and Adin Armey of American English and Culture Program, ASU for providing me financial assistantship. Thanks to Senh Luu, Venkatesh Mandalapa, and Sashi Gangaraju of University Technology Office, ASU for their constant help and support I received from them at work.

I am grateful to all my friends and colleagues in the Signal Processing and Communication group. Thanks to Adithya Rajan, Yuan Zhang,

Junghoon Lee, Ruochen Zeng, Mahesh Banavar and Adarsh Narasimhamurthy for their help and several useful discussions. Thanks to Professor Govindarajulu, Vijayavel Mohan, Jothi Sundaram, Ashok Kannaiyan, SenthilKumar Kadirvelu, Sharada Ramesh, Ganesh Balasubramanian, Gokula Thulasingam, Balaji Kadambi, Rajender Rajasekar, Hussain Mohammed, Irfan Ahmed, Kabeer Ahmed, Sabil Ahmed, Malarvizhi Velappan and many other friends who have supported and encouraged me during the most difficult times of my graduate study.

Finally, my deepest gratitude goes to my parents, brother Muralidharan Dasarathan and family members whose contributions cannot be mentioned in words. I also offer my deepest gratitude to my Guru Bhagavan Sri Ramana Maharishi whose Consciousness made a revolutionary change in the very nature of my perception about life.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Sensor Networks

Sensor networks (SNs) are designed to observe and collect information about a phenomenon of interest by using sensor nodes deployed in space. The configuration of the network depends on the application requirements such as health, military and home applications [1–3]. The nodes in a sensor network typically have sensing, processing and communication capabilities, that can help make intelligent decisions. Depending on the nature of the application and the types of sensors, a sensor network can be designed to observe a single physical phenomenon, or a single network can be designed to collect information from various physical conditions. Recent advancements in electronics and hardware technology have enabled the development of small, low cost, low power sensors. These devices have the capability of wireless communications and some of them are capable of locomotion. These features make the sensor networks suitable for a variety of applications discussed in Section 1.1.1.

The size of sensor nodes can range from being extremely small (of the order of cubic-millimetre) called as the smart-dust [4] to large platforms collecting telemetry information in a aircraft. Depending on how the sensor nodes are used and deployed, their capabilities may vary widely. Extremely small sensors have limited memory capacity [1], they may be able to do sensing but may not have sophisticated signal processing capabilities. Larger sensor nodes that are supported by a more complex infrastructure can have more sophisticated and different types of sensors. These can also be supported by larger computers with better computing capacities [5].

Wireless sensor networks (WSNs) differ fundamentally from general data networks such as the internet, and there are challenging problems in designing such networks. Some of the issues briefly follow. The local sensor observations need to be processed (compressed) before they are sent to the fusion center for joint processing. This arises for example, if the range of the observed data at local sensors is large, or the channel capacity between sensors and the fusion center is limited. If, on the other hand, the raw data observed at local sensors are accessible in their entirety at the fusion center, the problem of studying the physical phenomenon could be solved by one of the techniques of classical statistical inference [6, 7]. The system should operate with the stringent power constraints imposed by the WSN. The communication among the sensors and between the sensors and the fusion center should happen through the unreliable wireless channels.

1.1.1 Applications of Sensor Networks

Numerous applications of SNs have been discussed in detail in [2]. Due to the advances in the past decade in microelectronics, sensing, analog and digital signal processing, wireless communications, and networking, the design challenges are being tackled and WSNs are expected to have significant impact on lives of people in the twenty-first century. WSNs can be an integral part of military command, control, communications, computing, intelligence, surveillance, reconnaissance and targeting systems. There are many environmental applications such as forest fire detection, bio-complexity mapping of the environment, flood detection and precision agriculture etc. WSNs have significant number of health applications such as tele-monitoring of human physiological data, tracking and monitoring doctors and patients inside a hospital drug administration in hospitals. These applications involve identification of certain

signal sources that may be characteristics of hazardous material, monitoring chemicals near a volcano, temperatures in a furnace, shifts in undersea tectonic plates, or explosives in the air, to mention a few. Thus, SNs provide a safe and low-cost inference alternative. Interesting commercial applications include environmental control in office buildings, interactive museums, detecting and monitoring car thefts, managing inventory control and vehicle tracking and detection. Sensor networks can also be used for traffic control [8]. They can be used to inform drivers about the areas of congestion, and to divert the traffic to increase the efficiency of the roadways. They can be used to monitor roads for accidents and stoppages. SNs can be deployed to manage parking areas and to detect illegal use of parking areas. There are applications in manufacturing, transportation, and home appliances in which multiple decision makers arise naturally. Therefore, the study and design of distributed methods for distributed inference in WSNs becomes an important subject.

1.1.2 Architecture of Sensor Networks

There are three different architectures for sensor networks; 1). Ad-hoc Networks, 2). Hierarchical Networks and 3). Conventional Sensor Networks. These are briefly discussed here.

1.1.2.1 Ad-hoc Networks

In network literature, ad-hoc networks (Figure 1.1) refer to devices placed to form a network without a controlling base station. These devices discover each other and cooperate intelligently in order to function as a network. The ad-hoc sensor networks are constructed in the same manner. Low-power sensors are placed in an observation field and the sensor network exists without a fusion center. Algorithms are developed for diverse applications such as data routing, collaborative inference and distributed signal processing, all subject

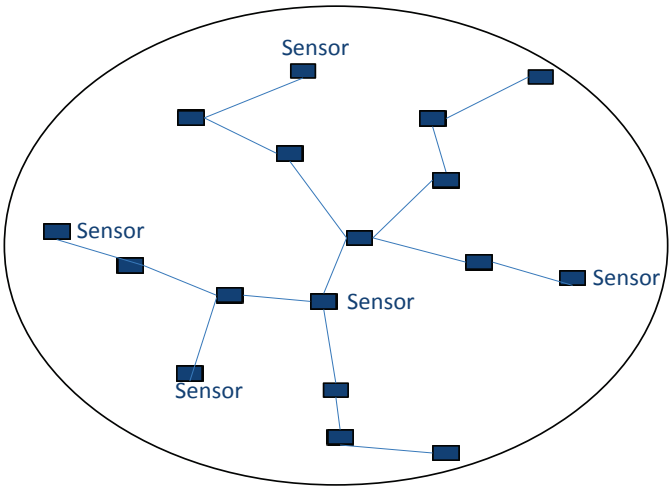

Figure 1.1: Ad-hoc sensor network without fusion center.

to a strict power constraint. Data-transmission between sensors in an ad-hoc network is typically achieved using multi-hop routing, i.e., sensors in between the source and destination are used to route the data between the transmitter and the receiver. These sensors behave as relays in addition to their functions as sensors [5].

Connectivity between sensors is a design issue in ad-hoc networks. An ad-hoc network consists of nodes which share a common wireless medium. Signals which are intended for a given receiver node can cause interference to the other receiver nodes. This can potentially reduce the signal to noise ratio (SNR) of the other receiver nodes. At the same time, each transmitter's (sensor node's) power needs to be sufficiently high enough to reach the intended receivers, while causing minimum interference on other receivers (nodes) sharing the same channel [9–13]. For instance, in an ad-hoc WSN, the nodes in the network are assumed to co-operate in routing each other's packets, each sensor node should transmit with sufficient power to ensure connectivity in the network. Considering this problem, [10] derives an expression for the critical

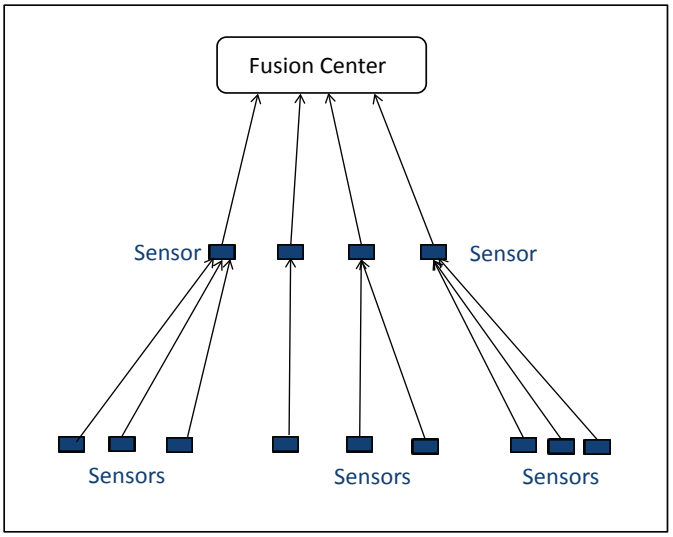

Figure 1.2: Hierarchical model: Data passes through multiple sensors.

power that is necessary for a node to transmit in order to guarantee that the network is connected almost surely when the number of nodes is large. For a unit bandwidth, the maximum rate at which reliable data transfer can happen between a given set of transmitters and receivers of an ad-hoc network is called as the network. Capacity of wireless ad-hoc networks for different conditions are analyzed in [14–19].

1.1.2.2 Hierarchical Networks

The second type of configuration used in SNs is called the hierarchical configuration (See Figure 1.2). In this setting, in addition to observing data, sensors collect decisions from other sensors [20, 20, 21]. They jointly process all this information to arrive at their own decisions and pass along their decision to subsequent sensors. This type of architecture is used in sequential detection and sequential estimation [22, 23].

1.1.2.3 Conventional Sensor Networks

In conventional sensor networks, sensors observe data and transmit them to a fusion center (in Figure 1.3). Instead of transmitting the raw observations sensed, the sensors use their signal processing abilities to locally carry out simple computations and transmit only the required and thus partially processed data to the FC which performs the final task of detection or estimation [2]. The transmissions between sensors and the FC may happen over orthogonal or multiple-access channels [5]. When the transmissions are orthogonal, the transmissions from each sensor reach the FC separately corrupted by an additive noise. There is no interference between transmitted signals. Therefore, the fusion center can process the transmitted signals independently. On the other hand, when the communications happen over multiple-access channels, the transmitted signals from the sensors are added (incoherently when there is time delay involved) by the channel and the FC can not have access to the individual data from the sensors [5]. The disadvantage of orthogonal channels is that the bandwidth scales linearly with the number of sensors, whereas, when the channels are multiple-access, transmissions are simultaneous and in the same frequency band, keeping the utilized bandwidth independent of the number of sensors in the sensor network. For this multiple access channel model, it has been shown in [24] that a simple amplify-and-forward (AF) scheme for analog signals is asymptotically optimal over AWGN channels. It has also been shown in a distributed estimation context, that if the fading channels are zero-mean, having no channel state information at the sensors results in poor performance [25]. Sensor networks that use this architecture are typically used for collaborative signal processing applications like joint estimation, distributed detection, histogram estimation, etc. Due to the presence of

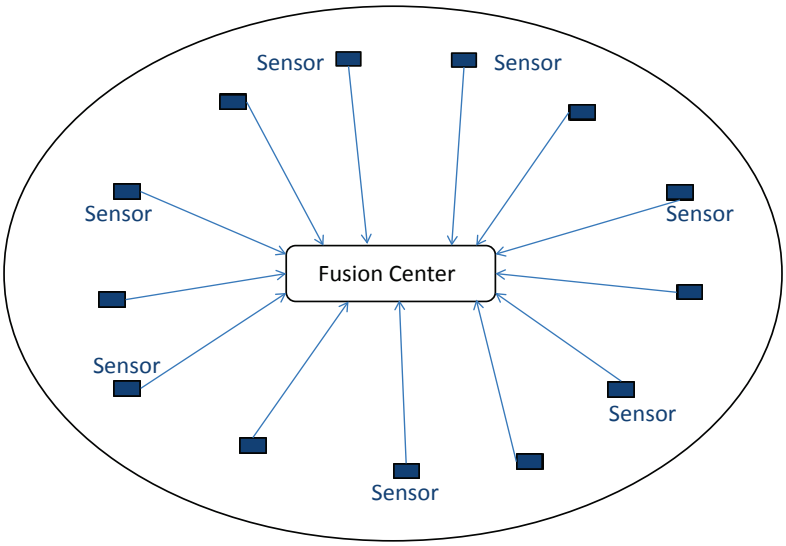

Figure 1.3: Sensor network with a fusion center.

multiple sensors, statistical methods perform very well since the number of observations can be very large. Histogram estimation using type based multiple access (TBMA) is introduced in [26].

Transmissions from the sensors to the FC can be analog or digital [5]. The digital method consists of quantizing the sensed data and then transmitting the data digitally over a rate-constrained channel [27]. In this case, the required channel bandwidth is quantified by the number of bits being transmitted between the sensors and the fusion center. When the number of quantization levels is high to reduce the loss of information, the bandwidth requirements increases for transmitting the information digitally. If the bandwidth is limited, analog methods of transmissions may be used. One such analog method consists of amplifying and then forwarding the sensed data to the FC, while imposing a power constraint [25]. The transmissions can be appropriately pulse-shaped and amplitude modulated to consume finite bandwidth. The major drawback of the amplify-and-forward scheme is that the transmit power depends on the sensing noise realizations and therefore may

not be bounded. One solution to this problem is the use of phase modulation techniques with constant modulus transmissions from the sensors. Another solution is to map the sensed observations through a bounded function before transmission so that the transmit power is always constrained.

1.1.3 Design Challenges in WSNs

The three major design principles for energy-constrained WSNs are summarized below (please see [1, 28–32] for detailed discussions).

1. Low Power Constraints: Exploit low power hardware, and external assets, to the greatest extent possible.

2. Communication Constraints: Optimize distributed detection and estimation network tasks while minimizing the use of communications.

3. Network Constraints: Support network specific goals while minimizing idle listening, network set-up, and network maintenance.

One of the most important constraints that is faced when dealing with autonomous nodes is that these nodes are severely power limited [1,2]. In most cases, the nodes are supplied with power with batteries when deployed and these batteries cannot be recharged or replaced. Typically energy consumption depends on the state of the sensor node, such as transmit, receive, idle etc. For instance, it is shown in [33] that the transceiver of the sensor node consumes more power in receive mode than in transmit mode. Since these nodes may be deployed in remote regions, it is preferred that they stay powered for large periods of time before the nodes are replaced. Therefore, whenever nodes are autonomously deployed, the algorithms used on the nodes have to be designed to consume minimal power so that the battery life can be maximized. It should be noted here that while computing operations do consume power, maximum power is consumed by the transceivers on the nodes. Hence, there is a need

for efficient multi-access schemes to maximize transfer of information subject to strict power constraints. When the sensors are deployed individually, they can perform only simple computations and perform poorly at sensing. However, when deployed in large numbers, they can collaborate among themselves to form intelligent networks and complicated tasks using statistical inference could be accomplished. A number of useful references for various other design issues such time synchronization, node localization, medium access control, hardware and routing in an energy constrained WSN, are enumerated in [1].

In this work, we consider several distributed inference problems with bounded transmit power from the sensors. Statistical inference can be broadly classified into hypothesis testing and estimation. Accordingly, there are two areas of distributed signal processing: Distributed detection and Distributed estimation. In what follows, a brief overview of these two areas with some of the relevant literature is described.

1.2 Distributed Detection

In a classical centralized detection scheme, the local sensors are assumed to communicate their observations to a central processor that performs the task of optimal detection using conventional statistical techniques. This is an ideal scenario without loss of information and noise free communication is assumed between the sensors and the central processor. In a distributed detection system, generally, the raw observations are processed (quantized and channel coded in case of digital transmissions) at the local sensors before they are transmitted to the fusion center. This results in loss of information (often referred to as lossy compression) at the local sensors and further loss could occur if the transmission is over a fading/noisy channel. For these reasons, the performance of a distributed detection system will always be suboptimal

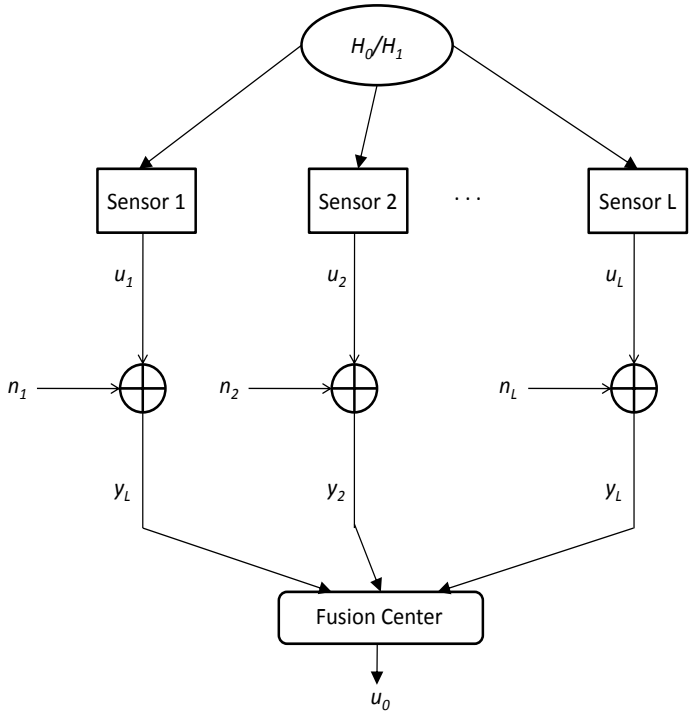

Figure 1.4: Distributed detection: Parallel topology with the fusion center

compared to the centralized detection system. However, there are significant advantages of the distributed detection such as reduced bandwidth requirement (due to lossy compression), increased reliability, and reduced cost. In addition, due to the relatively low cost of sensors, the availability of high speed communication networks, and increased computational capability, distributed detection has become a topic of great research interest [34] in the last two decades. There are four different topologies used for distributed detection: 1). Parallel topology, 2). Serial topology, 3). Tree topology, and 4). Multiple access topology [22, 35–38].

Figure 1.4 shows a simple parallel topology with a fusion center. In this study, we are concerned with the binary hypothesis testing problem for the detection problem. There are basically L sensors in this system. Each sensor observes the phenomenon and independently takes its decision based on its local decision rule. Here, $x_1, \ldots, x_L$ are the observations of the L sensors and $u_1, \ldots, u_L$ are their corresponding decisions. Each sensor then transmits

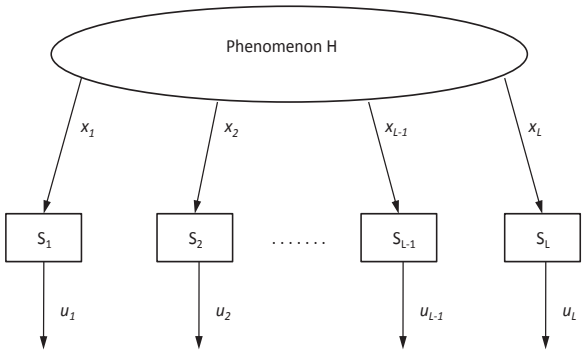

Figure 1.5: Parallel topology without the fusion center

its decision to the fusion center over a dedicated orthogonal channel and n_1, ..., n_L are the additive noise samples associated with the reception of u_1, ..., u_L respectively. Here it is clear that L parallel channels are needed. So, as the number of sensors increase, the bandwidth requirements for the successful operation of the network would increase greatly. The fusion center's task is to decide which one of the hypothesis is true by jointly processing the noisy versions of the decisions u_1, ..., u_L received across the L branches. In a more general setting, u_1, ..., u_L could be functions of x_1, ..., x_L respectively instead of the binary decisions of the individual sensors. Note that there is information loss in this general setting as well, and therefore the performance at the FC will be sub-optimal compared to the centralized set-up. Figure 1.5 shows a DD system without a fusion center. All the sensors observe a common phenomenon and make local decisions about the phenomenon. There is no task of fusion in this setting, whereas costs of decision making at different sensors are assumed to be coupled and a system wide optimization is performed based on the coupled cost function [22].

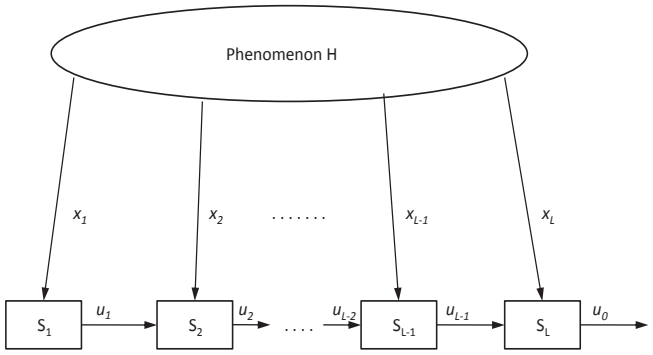

Figure 1.6: Serial topology

Figure 1.6 shows the Serial Topology for a distributed detection system. In this setting, the decisions of the individual sensors are combined in a serial manner. For instance, consider the case of only two sensors. First sensor 1 observes x_1 and takes a decision u_1. This decision is given to sensor 2 which takes the final decision by combining its own observation x_2 and the decision made by x_1. Figure 1.7 shows the tree topology for a DD system. The functioning of a tree network is similar to a serial network, but a given sensor receives decisions from more than one sensor. A DD system in which sensors communicate their processed observations to the FC over a multiple access channel is shown in Figure 1.8.

For detailed discussions and design of optimal decision rules for the first three topologies please see [22, 35, 39–41] and for the multiple access topology please see [36–38, 42–44].

When the prior probabilities of the hypothesis are known, Bayesian formulation is employed for detection. Bayesian formulation is a strategy to minimize the probability of error at the fusion center. When the prior

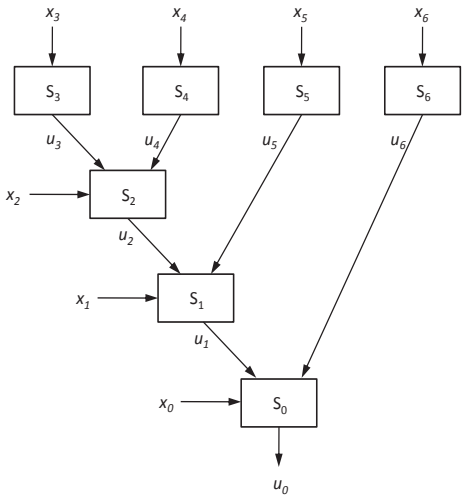

Figure 1.7: Tree topology

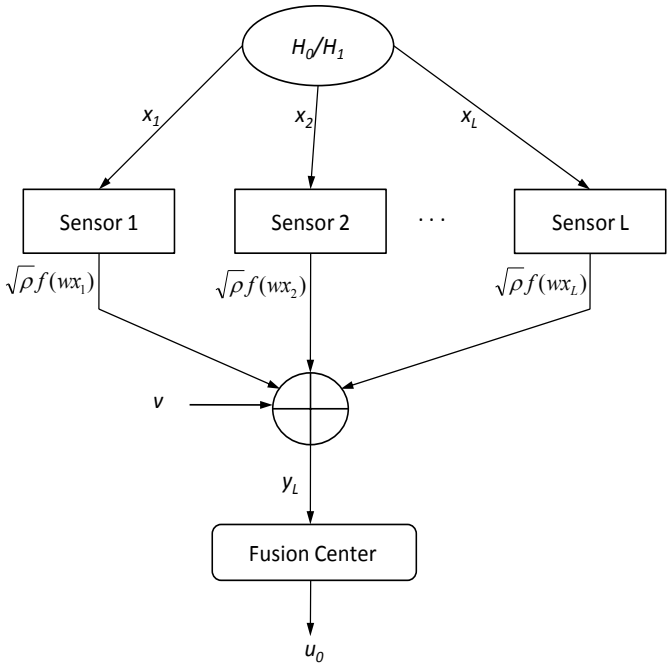

Figure 1.8: Distributed detection: Multiple access topology with the fusion center

probabilities are not known, for instance in case of a target detection, the Neyman-Pearson formulation is employed. Neyman-Pearson is a strategy to maximize the probability of detection (minimize probability of miss) subject to a false alarm probability constraint. The probability of false alarm is also

called as the Type I error probability and probability of miss is also called as Type II error probability [6].

Some useful metrics to measure the performance of a DD system are described next. Under Bayesian hypothesis testing, the overall probability of error at the fusion center is an important performance metric, and it generally depends on fusion rule, channel SNR, the performance indices of the individual sensors, noise variance at the fusion center, the channel parameters between sensors and the fusion center. Under the Neyman-Pearson formulation, for a given probability of false alarm at the fusion center, the probability of detection (or the probability of miss P_M) characterizes the over all detection capability of the system. In a sensor network with a large number of sensors, for any reasonable collection of transmission strategies and a fusion rule, the probability of error P_e at the fusion center goes to zero exponentially fast as L tends to infinity [35]. For example, with conditional independence of sensor observations, using identical transmission mappings for all the sensor nodes is asymptotically optimal [45]. For such a network, an important characteristic of interest is how fast the probability of error at the fusion center goes to zero. This is called as the error exponent of the sensor network [35, 46]. In fact the probability of error can be approximated using the error exponent [47, pp. 10].

1.3 Distributed Estimation

Distributed estimation with multiple sensors is required in areas such as environmental monitoring and remote sensing. For distributed estimation, there are two major types of topologies used: 1). Parallel topology, 2). Multiple access topology [22,38,48] as shown in Figures 1.9 and 1.10. In a distributed estimation set up, sensors observe an unknown physical phenomenon (indicated as θ) embedded in noise. The sensors send the processed versions (indicated

as $f(x_i)$) of these observations to the FC over orthogonal channels as shown in Figure 1.9 or over a multiple access channel as shown in Figure 1.10. Here ρ is the scale factor which controls the per-sensor transmit power. The FC has to estimate the unknown phenomenon θ by jointly processing the signals received from the sensors.

The authors in [49, 50] considered the estimation of a random parameter with a large number of local sensor observations. Basically, the distributed sensors observe the data, processes them, (quantize and channel code in case of digital transmissions) and then transmit their processed observations to a fusion center over the fading wireless channels. The task of the fusion center is to combine the data from the sensors to give an estimate of the parameter under consideration. For the case of i.i.d. observations with different noise variances at different sensors and with i.i.d. fading channels, between the sensors and the fusion center, it is shown in [51] that the full estimation diversity (on the order of the number of sensor nodes) can be achieved even with simple equal power transmission strategies.

Strict power constraints must be placed on the sensor nodes to ensure the maximum lifetime of the batteries. As mentioned earlier, one method is to use amplify-and-forward with a total power constraint and it is asymptotically optimal [24]. In [52], a power scheduling strategy is developed for decentralized estimation in wireless sensor networks, when sensor nodes adopt a uniform randomized quantization scheme with an uncoded quadrature amplitude modulation scheme. The optimal quantization level and transmission is shown to be a function of channel path loss and the local observation noise level.

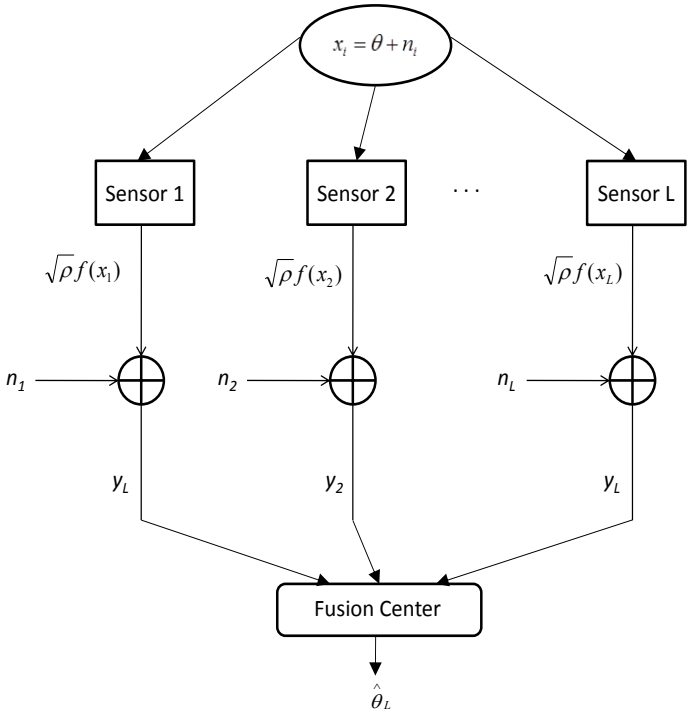

Figure 1.9: Distributed estimation: Parallel topology with the fusion center

For a bandwidth constrained sensor network, the reference [53] studies the scenario where the sensor observes a noisy version of the true parameter and transmits one bit of information to the fusion center. A new type of estimation technique called the type based estimation is introduced in [26]. It is shown in [26] with the assumptions that the channels between sensors and fusion center are zero mean and sensors have no channel state information (CSI), the performance of estimator is poor when the histogram of a finite alphabet source is being estimated. In this case, if the fusion center sends CSI to the sensors, the performance is improved significantly. For the TBMA scheme, practical issues such as the impact of the interference between the orthogonal waveforms, and channel estimation error on the histogram and parameter estimation is studied in [54]. The authors in [55] consider a strategy of selecting the sensor gains to minimize the the variance of the estimator with perfect channel feedback.

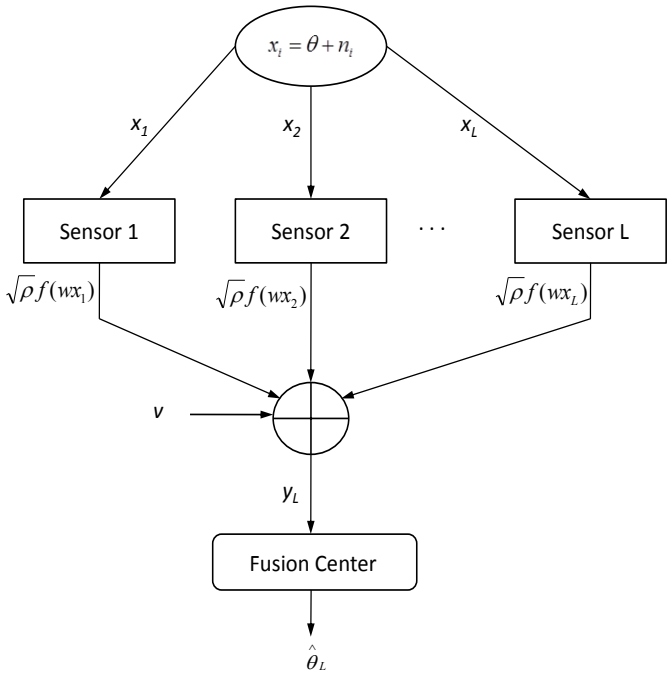

Figure 1.10: Distributed estimation: Multiple access topology with the fusion center

Distributed systems without a fusion center (fully distributed) have the advantages of robustness to node failures and being able to function autonomously without a central node controlling the entire network [2]. In a fully distributed set up, the sensors can collaborate with their neighbours by exchanging information locally to achieve a desired global objective. In what follows we describe one such objective called distributed consensus.

1.4 Distributed Consensus

Consensus literally means a group of agents coming to an agreement on a certain quantity of interest. In a distributed consensus problem, multiple nodes (sensors) which are distributed across a network (wired or wireless) agree on some desired parameter. For example, the nodes in a WSN can use the consensus value to perform useful actions such as detecting a signal, estimating an unknown parameter or controlling a process. Distributed consensus algorithms have attracted significant interest in the recent past and has found a

wide range of applications [56–58]. Here we introduce the basic idea behind distributed consensus algorithm in a WSN.

Consider a WSN with N sensors deployed with sensing, communication and signal processing capabilities. Let each sensor hold an initial measurement $x_i(0) \in \mathbb{R}$, $i = 1, \ldots, N$, measured by each of the N sensors. The measurements could contain information about the unknown parameters of a physical phenomenon such as temperature or strength of an unknown signal. Let the average of the initial measurements $\bar{x} = N^{-1} \sum_{i=1}^{N} x_i(0)$ be the parameter to be estimated by a distributed algorithm, in which each node communicates only with its neighbours. If the states of all the sensor nodes converge asymptotically with time to $\bar{x}$, then the network is said to have reached *consensus* on the sample average.

Distributed average consensus algorithms have been considered in the literature (please see [58–67] and references therein). In most of these papers, it is assumed that a given node can obtain exact information of the state values of its neighbours through local communications. This essentially means that there is unlimited energy and/or bandwidth. However, as mentioned in Section 1.1.3, practical WSNs are severely power limited and the available bandwidth is finite. Therefore, there is a need for consensus algorithms which could work under strict resource constraints of power and bandwidth imposed by the WSNs.

1.5 Contributions of the Dissertation

In this work, we address the first two design challenges discussed in Section 1.1.3 by proposing distributed detection and estimation schemes which require bounded transmit power and finite bandwidth. The proposed schemes will

be shown to outperform the existing schemes under the stringent power and bandwidth constraints imposed by WSNs.

A distributed detection scheme relying on constant modulus transmissions from the sensors is proposed over a Gaussian multiple access channel. The instantaneous transmit power does not depend on the random sensing noise, which is a desirable feature for low-power sensors with limited peak power capabilities. In addition to the desirable constant-power feature, the proposed detector is robust to impulsive noise, and performs well even when the moments of the sensing noise do not exist as in the case of the Cauchy distribution. It is shown that over Gaussian multiple access channels, the proposed detector outperforms AF, DF and modified DF schemes consistently, and the modified AF scheme when the sensing SNR is greater than 4 dB. The proposed detector is shown to perform well even when the channel noise is non-Gaussian. The error exponent is also derived for the proposed scheme and large deviation theory is used to approximate the probability of error for large L.

A distributed inference scheme which uses bounded transmission functions over a Gaussian multiple access channel is considered. The conditions on the transmission functions under which consistent estimation and reliable detection are possible is characterized. For the distributed estimation problem, an estimation scheme that uses bounded transmission functions is proved to be strongly consistent provided that the variance of the noise samples are bounded and that the transmission function is one-to-one. The asymptotic variance is derived, and shown to depend on the derivative of the transmission function and the sensing noise statistics and channel noise variance. The proposed estimation scheme is compared with the amplify and forward technique

and its robustness to impulsive sensing noise distributions is highlighted. It is also shown that bounded transmissions suffer from inconsistent estimates if the sensing noise variance goes to infinity. For the distributed detection problem, similar results are obtained by studying the deflection coefficient.

A distributed average consensus algorithm in which every sensor transmits with bounded peak power is proposed. In the presence of communication noise, it is shown that the nodes reach consensus asymptotically to a finite random variable whose expectation is the desired sample average of the initial observations with a variance that depends on the step size of the algorithm and the variance of the communication noise. The asymptotic performance is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. It is shown that using bounded transmissions results in slower convergence compared to the linear consensus algorithm based on the Laplacian heuristic.

A distributed average consensus algorithm in which every sensor performs a nonlinear processing at the receiver is proposed. We prove that nonlinearity at the receiver nodes makes the algorithm robust to a wide range of channel noise distributions including the impulsive ones. This work is the first of its kind in the literature to propose a consensus algorithm which relaxes the requirement of finite moments on the communication noise. When the communication noise samples are i.i.d., it is shown that the nodes reach consensus asymptotically to a finite random variable whose expectation is the desired sample average of the initial observations with a variance that depends on the step size of the algorithm and the receiver nonlinear function. The asymptotic performance is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. It is shown that scaling

the receiver nonlinear function does not affect the convergence speed of the algorithm. An interesting relationship between the Fisher information and the asymptotic covariance matrix is shown.

1.6 Outline of the Dissertation

The rest of the Dissertation is organized as follows. Chapter 2 describes a distributed detection scheme relying on constant modulus transmissions from the sensors over a Gaussian multiple access channel and analyses the performance of the proposed detection scheme by deriving the deflection coefficient and the error exponent for several cases. Chapter 3 considers a general problem of distributed inference using bounded transmission functions and establish regimes under which estimation and detection will be possible and discuss regimes under which estimation will fail and reliable detection will be impossible. Chapter 4 studies the merits and demerits of bounded transmissions in distributed consensus problems and characterizes the performance for a variety of bounded transmission functions. Chapter 5 proposes a robust consensus algorithm and characterizes its asymptotic performance. Finally the conclusions are presented in Chapter 6.

Chapter 2

Distributed Detection with Constant Modulus Signaling

2.1 Literature Survey and Motivation

In Chapter 1 we learnt that, in inference-based wireless sensor networks, low-power sensors with limited battery and peak power capabilities transmit their observations to a fusion center (FC) for detection of events or estimation of parameters. For distributed detection, much of the literature has focused on the parallel topology where each sensor uses a dedicated channel to transmit to a fusion center. Multiple access channels offer bandwidth efficiency since the sensors transmit over the same time/frequency slot.

In [68], the distributed detection over a multiple access channel is studied where arbitrary number of quantization levels at the local sensors are allowed, and transmission from the sensors to the fusion center is subject to both noise and inter-channel interference. References [69–72] discuss distributed detection over Gaussian multiple access channels. In [69], detection of a deterministic signal in correlated Gaussian noise and detection of a first-order autoregressive signal in independent Gaussian noise are studied using an amplify-and-forward scheme where the performance of different fusion rules is analyzed. In [70], a type-based multiple access scheme is considered in which the local mapping rule encodes a waveform according to the type [73, pp. 347] of the sensor observation and its performance under both the per-sensor and total power constraints is investigated. This scheme is extended to the case of fading between the sensors and the FC in [71] and its performance is analyzed using large deviation theory. In the presence of non-coherent fading over a Gaussian multi-access channel, type-based random access is proposed and analyzed in [72]. In [74], the optimal distributed detection scheme in a

clustered multi-hop sensor network is considered where a large number of distributed sensor nodes quantize their observations to make local hard decisions about an event. The optimal decision rule at the cluster head is shown to be a threshold test on the weighted sum of the local decisions and its performance is analysed.

Two schemes called modified amplify-and-forward (MAF) and the modified detect-and-forward (MDF) are developed in [37] which generalize and outperform the classic amplify-and-forward (AF) and detect-and-forward (DF) approaches to distributed detection. It is shown that MAF outperforms MDF when the number of sensors is large and the opposite conclusion is true when the number of sensors is smaller. For the MDF scheme with identical sensors, the optimal decision rule is proved to be a threshold test in [36]. Decision fusion with a non-coherent fading Gaussian multiple access channel is considered in [75] where the optimal fusion rule is shown to be a threshold test on the received signal power and on-off keying is proved to be the optimal modulation scheme. A distributed detection system where sensors transmit their observations over a fading Gaussian multiple-access channel to a FC with multiple antennas using amplify-and-forward is studied in [76]. In all these cases, the sensing noise distribution is assumed to be Gaussian. Even though the Gaussian assumption is widely used, sensor networks which operate in adverse conditions require detectors which are robust to non-Gaussian scenarios. Moreover, in the literature there has been little emphasis on distributed schemes with the desirable feature of using constant modulus signals with fixed instantaneous power.

A distributed estimation scheme where the sensor transmissions have constant modulus signals is considered in [38]. Distributed estimation in a

bandwidth-constrained sensor network with a noisy channel is investigated in [77] and distributed estimation of a vector signal in a sensor network with power and bandwidth constraints is studied in [78]. The estimator proposed in [38] is shown to be strongly consistent for any sensing noise distribution in the iid case. Inspired by the robustness of this estimation scheme, in this work, a distributed detection scheme where the sensors transmit with constant modulus signals over a Gaussian multiple access channel is proposed for a binary hypothesis testing problem. The sensors transmit with constant modulus transmissions whose phase is linear with the sensed data. The output-signal-to-noise-ratio, also called as the deflection coefficient (DC) of the system, is derived and expressed in terms of the characteristic function (CF) of the sensing noise. The optimization of the DC with respect to the transmit phase parameter is considered for different distributions on the sensing noise including impulsive ones. The error exponent is also derived and shown to depend on the CF of the sensing noise. It is shown that both the DC and the error exponent can be used as accurate predictors of the phase parameter that minimizes the detection error rate. The proposed detector is favorably compared with MAF and the MDF schemes developed in [36,37] for the Gaussian sensing noise and its robustness in the presence of other sensing noise distributions is highlighted. The effect of fading between the sensors and the fusion center is shown to be detrimental to the detection performance through a reduction in the DC depending on the fading statistics. Different than [38] where the asymptotic variance of an estimator is analyzed, the emphasis herein is on derivation, analysis, and optimization of detection-theoretic metrics such as the DC and error exponent. Our aim in this chapter is to develop a distributed detection scheme where the instantaneous transmit power is not influenced by possibly unbounded sensor measurement noise.

This chapter is organized as follows. In Section 2.2, the system model is described with per-sensor power constraint and total power constraint. In Section 2.3, the detection problem is described and a linear detector is proposed. The probability of error performance of the detector is analyzed in Section 2.4. The DC is defined and its optimization for several cases is studied in Section 2.5. The presence of fading between the sensors and the fusion center is discussed in Section 2.6. The error exponent of the proposed detector is analyzed in Section 2.7. Non-Gaussian channel noises are discussed in Section 2.8. Simulation results are provided in Section 2.9 which support the theoretical results.

2.2 System Model

Consider a binary hypothesis testing problem with two hypotheses H_0, H_1 where P_0, P_1 are their respective prior probabilities. Let the sensed signal at the i^{th} sensor be,

$$x_i = \begin{cases} \theta + n_i & \text{under } H_1 \\ n_i & \text{under } H_0 \end{cases} \tag{2.1}$$

$i = 1, \ldots, L$, $\theta > 0$ [1] is a known parameter whose presence or absence has to be detected, L is the total number of sensors in the system, and n_i is the noise sample at the i^{th} sensor. The sensing noise samples are independent, have zero median and an absolutely continuous distribution but they need not be identically distributed or have any finite moments. We consider a setting where the i^{th} sensor transmits its measurement using a constant modulus signal $\sqrt{\rho}e^{j\omega x_i}$ over a Gaussian multiple access channel so that the received signal at

[1] the proposed scheme will work without any difference for $\theta < 0$ due to symmetry if we substitute $-\theta$ in the place of θ in all the equations.

the FC is given by

$$y_L = \sqrt{\rho} \sum_{i=1}^{L} e^{j\omega x_i} + v \tag{2.2}$$

where ρ is the power at each sensor, $\omega > 0$ is a design parameter to be optimized and $v \sim \mathcal{CN}(0, \sigma_v^2)$ is the additive channel noise. We consider two types of power constraints: Per-sensor power constraint and total power constraint. In the former case, each sensor has a fixed power ρ so that the total power $P_{\mathrm{T}} = \rho L$, and as $L \to \infty$, $P_{\mathrm{T}} \to \infty$; in the later case, the total power P_{T} is fixed for the entire system and does not depend on L, so that the per-sensor power $\rho = P_{\mathrm{T}}/L \to 0$ as $L \to \infty$.

2.3 The Detection Problem

The received signal y_L under the total power constraint can be written as

$$y_L = \sqrt{\frac{P_{\mathrm{T}}}{L}} \sum_{i=1}^{L} e^{j\omega x_i} + v. \tag{2.3}$$

We assume throughout that $P_0 = P_1 = 0.5$ for convenience even though other choices can be easily incorporated. With the received signal in (2.3), the FC has to decide which hypothesis is true. It is well known that the optimal fusion rule under the Bayesian formulation is given by:

$$\frac{f(y_L|H_1)}{f(y_L|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} = 1 \tag{2.4}$$

where $f(y_L|H_i)$, is the conditional probability density function of y_L when H_i is true. The equation (2.3) can be rewritten as follows:

$$y_L = \sqrt{\frac{P_{\mathrm{T}}}{L}} \left(\sum_{i=1}^{L} \cos(\omega x_i) \right) + j \sqrt{\frac{P_{\mathrm{T}}}{L}} \left(\sum_{i=1}^{L} \sin(\omega x_i) \right) + v.$$

Since there are L terms in the first summation involving the cosine function, we need to do L fold convolutions with the PDFs of $\cos(\omega x_i)$ and another set of L fold convolutions with the PDFs of $\sin(\omega x_i)$. Then we need to find the joint

distribution of the PDFs obtained thus for the cosine and sine counterparts. This joint PDF will need to be convolved with the PDF of v. It is not possible to obtain a closed form expression for these $(2L+1)$ fold convolutions. Hence, $f(y_L|H_i)$ is not tractable. Therefore, we consider the following linear detector which is argued next to be optimal for large L:

$$\Re[y_L e^{-j\omega\theta}] - \Re[y_L] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \,, \tag{2.5}$$

where we define $\Re[y]$ as the real part, and $\Im[y]$ as the imaginary part of y. Note that the detector in (2.5) would be optimal if y_L were Gaussian. Clearly due to central limit theorem y_L in (2.3) is asymptotically Gaussian, which indicates that (2.5) approximates (2.4) for large L. With the Gaussian assumption, the variances of y_L in (2.3) under the two hypotheses are the same and given by $\mathrm{Var}(y_L|H_0) = \mathrm{Var}(y_L|H_1) = [P_{\mathrm{T}}(1-\varphi_n^2(\omega)) + \sigma_v^2]$, where $\varphi_n(\omega)$ is the characteristic function of n_i. Hence, the optimal likelihood ratio simplifies to the detector in (2.5) which is linear in y_L, when y_L is assumed Gaussian which holds for large L. However as will be seen in Section 2.4, we do not assume that y_L is Gaussian for any fixed L when we analyze the performance of the detector in (2.5) or in finding the associated error exponent in Section 2.7. We proceed by expressing the probability of error.

2.4 Probability of Error

The detector in (2.5) depends on the design parameter ω and this means that the probability of error will in turn depend on ω. Let $P_{\mathrm{e}}(\omega)$ be the probability of error at the FC:

$$P_{\mathrm{e}}(\omega) = \frac{1}{2}\Pr\left[error|H_0\right] + \frac{1}{2}\Pr\left[error|H_1\right] = \Pr\left[error|H_0\right] \tag{2.6}$$

where $\Pr\left[error|H_i\right]$ is the error probability when H_i, $i \in \{0,1\}$, is true and the last equality holds due to symmetry between the two hypotheses which is

explained as follows. From the detection rule (2.5), the probability of error under H_0 is given by

$$\Pr\left[error|H_0\right] = \Pr\left[\Re[y_L] < \Re[y_L e^{-j\omega\theta}]|H_0\right] \ , \tag{2.7}$$

where the received signal in (2.3) under H_0 is given by

$$y_L = \sqrt{\frac{P_{\mathrm{T}}}{L}} \sum_{i=1}^{L} e^{j\omega n_i} + v. \tag{2.8}$$

Substituting (2.8) for y_L in (2.7) and doing some algebraic simplifications, $\Pr\left[error|H_0\right]$ can be written as

$$\Pr\left[\underbrace{\sum_{i=1}^{L} 2\sin\left(\frac{\omega\theta}{2}\right)\cos\left(\omega n_i - \frac{\omega\theta}{2} + \frac{\pi}{2}\right) + \sqrt{\frac{L}{P_{\mathrm{T}}}}v^{\mathrm{T}}}_{Z_L(\omega):=} < 0\right] \tag{2.9}$$

where $v^{\mathrm{T}} := \Re[v](1-\cos(\omega\theta)) - \Im[v]\sin(\omega\theta)$. Similarly, $\Pr\left[error|H_1\right]$ is same as that of (2.9) except the argument of the cosine function is replaced by $(\omega n_i + \omega\theta/2 - \pi/2)$. To see the symmetry between the two hypotheses asserted in (2.6), let $\zeta := (\omega\theta/2 - \pi/2)$ for convenience, so that $\cos(\omega n_i \mp \zeta) = [\cos(\omega n_i)\cos\zeta + \sin(\pm\omega n_i)\sin\zeta]$. Since n_i is symmetric, ωn_i and $-\omega n_i$ have the same distribution which implies that the random variables $\cos(\omega n_i - \zeta)$ and $\cos(\omega n_i + \zeta)$ have the same distribution establishing that $\Pr\left[error|H_1\right] = \Pr\left[error|H_0\right]$. Therefore, the probability of error in (2.6) is given by (2.9). We are interested in using (2.9) to find the ω that minimizes the probability of error at FC. Since $P_{\mathrm{e}}(\omega)$ is not straightforward to evaluate, we optimize two surrogate metrics to select ω. These are the error exponent and the DC. The error exponent is an asymptotic measure of how fast the $P_{\mathrm{e}}(\omega)$ decreases as $L \to \infty$, and is specific to the detector used in (2.5) and will be considered in

Section 2.7. The DC, on the other hand, is specific to the model in (2.3), and does not depend on any detector.

2.5 Deflection Coefficient and its Optimization

We will now define and use the deflection coefficient which reflects the output-signal-to-noise-ratio and widely used in optimizing detectors [79–82]. The DC is mathematically defined as,

$$D(\omega) := \frac{1}{L}\frac{|\mathrm{E}[y_L|H_1] - \mathrm{E}[y_L|H_0]|^2}{\mathrm{var}[y_L|H_0]}. \tag{2.10}$$

By calculating the expectations in (2.10), it can be easily verified that the DC for the signal model in (2.2) is given by:

$$D(\omega) = \frac{2\varphi_n^2(\omega)[1-\cos(\omega\theta)]}{\left[1-\varphi_n^2(\omega)+\frac{\sigma_v^2}{P_{\mathrm{T}}}\right]} \tag{2.11}$$

where $\varphi_n(\omega) = \mathrm{E}[e^{j\omega n_i}]$ is the CF of n_i. The CF $\varphi_n(\omega)$ does not depend on the sensor index i, since we will be initially assuming that n_i are iid. We will consider the non-identically distributed case in Section 2.5.4. Note that $D(\omega) \geq 0$ and that $\varphi_n(\omega)$ is real-valued since n_i is a symmetric random variable. Moreover, $\varphi_n(\omega) = \varphi_n(-\omega)$ so that $D(\omega) = D(-\omega)$ which justifies why we will focus on $\omega > 0$ throughout. The factor $(1/L)$ introduced in (2.10) does not appear in conventional definitions of the DC but included here for simplicity since it does not affect the optimal ω.

2.5.1 Optimizing $D(\omega)$

We are now interested in finding ω by optimizing $D(\omega)$:

$$\omega^* := \arg\max_{\omega>0} D(\omega). \tag{2.12}$$

Since $\varphi_n(\omega) \leq 1$, when $\sigma_v^2 > 0$, $D(\omega)$ is bounded, and achieves its smallest value of $D(\omega) = 0$ as $\omega \to 0$. On the other hand, as $\omega \to \infty$, we

have $\lim_{\omega\to\infty} D(\omega) = 0$. This implies that the maximum in (2.12) cannot be achieved by $\omega = 0$ or $\omega = \infty$ and establishes that there must be a finite $\omega^* \in (0, \infty)$ which attains the maximum in (2.12).

In what follows, we will further characterize ω^* by assuming that $\varphi_n(\omega)$ ¿ 0 and $\varphi_n^{'}(\omega) < 0$ for all $\omega > 0$. Many distributions including the Laplace, Gaussian and Cauchy have CFs that satisfy this assumption. Indeed all symmetric alpha-stable distributions [83, pp. 20] of which the latter two is a special case, satisfy this assumption. We now have the following theorem which restricts ω^* in (2.12) to a finite interval.

Theorem 2.5.1. *If $\varphi_n(\omega)$ is decreasing and differentiable over $\omega > 0$, then $\omega^* \in (0, \pi/\theta)$.*

Proof. First, note that $\varphi_n(\omega) \geq 0$ which is implied by the assumption that $\varphi_n(\omega)$ is decreasing and that $\varphi_n(\omega) \to 0$ as $\omega \to \infty$. Let $D(\omega) = C(\omega)[1 - \cos(\omega\theta)]$ with $C(\omega) := 2\varphi_n^2(\omega)/[1 - \varphi_n^2(\omega) + \sigma_v^2/P_{\mathrm{T}}]$ for brevity. Since $\varphi_n(\omega)$ is decreasing on $\omega > 0$ and $\varphi_n(\omega) \geq 0$, $C(\omega)$ is also decreasing. Because $[1 - \cos(\omega\theta)]$ is periodic in ω with period $2\pi/\theta$,

$$D\left(\omega + \frac{2\pi}{\theta}\right) = [1 - \cos(\omega\theta)]C\left(\omega + \frac{2\pi}{\theta}\right) < [1 - \cos(\omega\theta)]C(\omega) = D(\omega). \tag{2.13}$$

Noticing that $D(2\pi/\theta) = 0$ which rules out $\omega^* = 2\pi/\theta$, we have $\omega^* \in (0, 2\pi/\theta)$. To further reduce the range of ω^* by half, consider the fact that $D(0) = D(2\pi/\theta) = 0$, which combined with $D(\omega) > 0$ for $\omega \in (0, 2\pi/\theta)$ implies that $\omega^* \in (0, 2\pi/\theta)$ satisfies $D^{'}(\omega^*) = 0$. Writing $D^{'}(\omega^*) = 0$ we obtain:

$$\frac{[\theta \sin(\omega^*\theta)]}{[\cos(\omega^*\theta) - 1]} = \frac{C^{'}(\omega^*)}{C(\omega^*)}. \tag{2.14}$$

Since $C(\omega) > 0$ is decreasing, the right hand side (rhs) of (2.14) is negative and it follows that $\omega^* \in (0, \pi/\theta)$ as required. □

By the definition of ω^*, it is clearly a function of θ. We showed in Theorem 2.5.1 that $0 < \omega^* < \pi/\theta$ if $\varphi_n^{'}(\omega) < 0$ for $\omega > 0$. Note that when $\omega = 0$, there is no phase modulation done, and what is transmitted is a constant signal which actually contains no information about x_i. Therefore the boundary value $\omega = 0$ is not a valid choice. When $\omega = \pi/\theta$, the detector in (2.5) actually simplifies to: $\Re[y_L] \underset{H_1}{\overset{H_0}{\gtrless}} 0$. While $\omega = \pi/\theta$ is a valid choice, it is optimal only when θ is large as will be proved in Theorem 2. We now investigate the behavior of ω^* when θ is large without assuming anything about $\varphi_n(\omega)$ except the absolute continuity of its distribution, and show that $\omega^* \approx \pi/\theta$ for large θ in the sense that $\omega^*\theta \to \pi$, as $\theta \to \infty$.

Theorem 2.5.2. *If $\sigma_v^2 > 0$, and n_i are iid and have absolutely continuous distributions,*

$$\lim_{\theta\to\infty} \omega^*\theta = \pi. \tag{2.15}$$

Proof. We have

$$D\left(\frac{\pi}{\theta}\right) \leq D(\omega^*) \leq \sup_{\omega>0}[1-\cos(\omega\theta)] \sup_{\omega>0} C(\omega) = \frac{4P_{\mathrm{T}}}{\sigma_v^2}\,, \tag{2.16}$$

where the first inequality is because ω^* maximizes $D(\omega)$, and the second inequality follows from $D(\omega) = C(\omega)[1-\cos(\omega\theta)]$. Recalling that $\lim_{\omega\to 0}\varphi_n(\omega) = 1$ we take the limit as $\theta \to \infty$ in (2.11) and obtain $\lim_{\theta\to\infty} D(\pi/\theta) = 4P_{\mathrm{T}}/\sigma_v^2$, which using (2.16) shows that $\lim_{\theta\to\infty} D(\omega^*) = 4P_{\mathrm{T}}/\sigma_v^2$. Since $\varphi_n(0) > \varphi_n(\omega)$ and because $D(\omega)$ is an increasing function of $\varphi_n^2(\omega)$, from (2.11) it is clear that the only way $\lim_{\theta\to\infty} D(\omega^*) = 4P_{\mathrm{T}}/\sigma_v^2$ holds is if $\omega^* \to 0$ and $\omega^*\theta \to \pi$, as $\theta \to \infty$. □

Theorem 2.5.2 establishes that when θ is large we have an approximate closed-form solution for $\omega^* \approx \pi/\theta$ for any absolutely continuous sensing noise distribution.

2.5.2 Finding the Optimum ω for Specific Noise Distributions

Theorem 2.5.1 showed that $\omega^* \in (0, \pi/\theta)$ for a general class of distributions. Under more general conditions, Theorem 2.5.2 establishes that $\omega^* \approx \pi/\theta$ when θ is large. To find ω^* exactly, we need to specify the sensing noise distribution through its CF, $\varphi_n(\omega)$. In what follows we describe how to find ω^* for several specific but widely used sensing noise distributions. We will assume throughout that the assumptions of Theorem 2.5.1 ($\varphi_n'(\omega) < 0$ for $\omega > 0$) are satisfied so that $\omega^* \in (0, \pi/\theta)$, which holds for Gaussian, Cauchy and Laplacian distributions, among others. We will assume $\sigma_v^2 > 0$ throughout this subsection.

2.5.2.1 Gaussian Sensing Noise

In this case, we have $\varphi_n(\omega) = e^{-\omega^2\sigma_n^2/2}$ so that $\varphi_n^2(\omega) = e^{-\omega^2\sigma_n^2}$, where σ_n^2 is the variance of n_i. To simplify (2.11) we substitute $\beta = \omega\theta$. Since $\omega \in (0, \pi/\theta)$ we have $\beta \in (0, \pi)$. Note that the value of ω that maximizes (2.11) over ω is related to the β that maximizes $D(\beta/\theta)$ through the relation $\omega = \beta/\theta$. Differentiating $D(\beta/\theta)$ with respect to β, equating to 0 and simplifying we obtain,

$$G_{\mathrm{G}}(\beta) := \alpha - e^{-\frac{\sigma_n^2}{\theta^2}\beta^2} - \frac{2\alpha\sigma_n^2}{\theta^2}\beta \tan\left(\frac{\beta}{2}\right) = 0 \tag{2.17}$$

with $\alpha := [1 + (\sigma_v^2/P_{\mathrm{T}})]$. Equation (2.17) can not be solved in closed-form. However it does have a unique solution in $\beta \in (0, \pi)$ as shown below.

First we note that $G_{\mathrm{G}}(0) = (\alpha - 1) > 0$ since $\sigma_v^2 > 0$ and $G_{\mathrm{G}}(\pi) = -\infty$. Since $G_{\mathrm{G}}(\beta)$ is continuous, (2.17) has at least one solution. To show that this solution is unique, consider the first derivative:

$$G_{\mathrm{G}}'(\beta) = \frac{\sigma_n^2}{\theta^2}\left[2\beta e^{-\frac{\sigma_n^2}{\theta^2}\beta^2} - 2\alpha\left(\frac{\beta}{2}\sec^2\left(\frac{\beta}{2}\right) + \tan\left(\frac{\beta}{2}\right)\right)\right]. \tag{2.18}$$

Now, using $\tan(\beta/2) \geq \beta/2$ and $\sec^2(\beta/2) \geq 1 + (\beta^2/4)$ for $\beta \in (0, \pi)$, we get the following upper bound:

$$G'_{\mathrm{G}}(\beta) \leq \frac{\sigma_n^2}{\theta^2} \left[2\beta e^{-\frac{\sigma_n^2}{\theta^2}\beta^2} - \alpha\beta \left(1 + \frac{\beta^2}{4}\right) - \alpha\beta \right]. \tag{2.19}$$

Since $\sigma_v^2 > 0$ we have $\alpha > 1$. Recall that $\beta \in (0, \pi)$, and the rhs of (2.19) is always negative. It follows that $G_{\mathrm{G}}(\beta)$ is monotonically decreasing over $\beta \in (0, \pi)$ and (2.17) has a unique solution which corresponds to the global maximum of $D(\beta/\theta)$. The solution to (2.17), β^*_{G}, can be found numerically and the optimum ω for the Gaussian case is $\omega^*_{\mathrm{G}} = \beta^*_{\mathrm{G}}/\theta$.

2.5.2.2 Cauchy Sensing Noise

In this case, $\varphi_n(\omega) = e^{-\gamma\omega}$ so that $\varphi_n^2(\omega) = e^{-2\gamma\omega}$ where γ is the scale parameter of the Cauchy distribution. It is well known that no moments of this distribution exists. Substituting $\varphi_n(\omega)$ in $D(\omega)$ and letting $\beta = \omega\theta$ we have,

$$D\left(\frac{\beta}{\theta}\right) = \frac{[1 - \cos(\beta)]}{[\alpha e^{\frac{2\gamma}{\theta}\beta} - 1]} \tag{2.20}$$

with $\alpha := [1 + (\sigma_v^2/P_{\mathrm{T}})]$ and $\beta \in (0, \pi)$. It can be verified that the equation (2.20) has a unique maximum over $\beta \in (0, \pi)$ as shown below.

The first derivative of $D(\beta/\theta)$ is given by,

$$D'\left(\frac{\beta}{\theta}\right) = \left[\frac{\sin(\beta) e^{\frac{2\gamma}{\theta}\beta}}{(\alpha e^{\frac{2\gamma}{\theta}\beta} - 1)^2}\right] \left[\alpha - e^{-\frac{2\gamma}{\theta}\beta} - \frac{2\gamma}{\theta}\alpha \tan\left(\frac{\beta}{2}\right)\right]. \tag{2.21}$$

Since the first term on the rhs of (2.21) is non-zero for $\beta \in (0, \pi)$, we need to solve

$$G_{\mathrm{C}}(\beta) := \alpha - e^{-\frac{2\gamma}{\theta}\beta} - \frac{2\gamma}{\theta}\alpha \tan\left(\frac{\beta}{2}\right) = 0. \tag{2.22}$$

First we see that $G_{\mathrm{C}}(0) = (\alpha - 1) > 0$ and $G_{\mathrm{C}}(\pi) = -\infty$ which implies that there is at least one solution to (2.22) in $\beta \in (0, \pi)$ as $G_{\mathrm{C}}(\beta)$ is continuous. The second derivative of $G_{\mathrm{C}}(\beta)$ is given by

$$G''_{\mathrm{C}}(\beta) = -\left[\left(\frac{4\gamma^2}{\theta^2} e^{-\frac{2\gamma}{\theta}\beta}\right) + \frac{\gamma\alpha}{\theta} \sec^2\left(\frac{\beta}{2}\right) \tan\left(\frac{\beta}{2}\right)\right]. \tag{2.23}$$

Clearly, $G''_{\mathrm{C}}(\beta) < 0$ for $\beta \in (0, \pi)$ which establishes that $G_{\mathrm{C}}(\beta)$ is concave. Therefore, (2.22) has a unique solution which corresponds to the global maximum of $D(\beta/\theta)$.The β^*_{C} that maximizes (2.20) can be found numerically and $\omega^*_{\mathrm{C}} = \beta^*_{\mathrm{C}}/\theta$.

When $\sigma_v^2/P_{\mathrm{T}}$ is sufficiently large (i.e., the low channel SNR regime) compared to $[1 - \varphi_n^2(\omega)]$ in $D(\omega)$, the problem in (2.11) can be transformed into maximizing $\varphi_n^2(\omega)[1 - \cos(\omega\theta)]$ over $\omega \in (0, \pi/\theta)$. In this low channel SNR regime, we have a closed form solution for the Cauchy case:

$$\omega^*_{\mathrm{C}} = \frac{2}{\theta} \tan^{-1} \frac{\theta}{2\gamma}. \tag{2.24}$$

If we let $\theta \to \infty$ in (2.24), we get $\omega^*_{\mathrm{C}} = \pi/\theta$ which agrees with Theorem 2.5.2.

2.5.2.3 Laplace Sensing Noise

In this case, we have $\varphi_n(\omega) = 1/(1 + b^2\omega^2)$ and $b^2 := \sigma_n^2/2$. Substituting this in $D(\omega)$ and letting $\beta = \omega\theta$, and differentiating $D(\beta/\theta)$ with respect to β, equating to 0 and simplifying we get,

$$G_{\mathrm{L}}(\beta) := \left[1 + \frac{b^2}{\theta^2}\beta^2\right]^2 - \frac{4b^2}{\theta^2}\beta\left[1 + \frac{b^2}{\theta^2}\beta^2\right]\tan\left(\frac{\beta}{2}\right) - \frac{1}{\alpha} = 0 \tag{2.25}$$

with $\alpha := [1 + (\sigma_v^2/P_{\mathrm{T}})]$. It can be easily verified that equation (2.25) has a unique solution in $\beta \in (0, \pi)$ as shown below.

First we note that $G_{\mathrm{L}}(0) = (1-(1/\alpha)) > 0$ if $\sigma_v^2 > 0$ and $G_{\mathrm{L}}(\pi) = -\infty$. This means that (2.25) has at least one solution. The first derivative of $G_{\mathrm{L}}(\beta)$ is given by,

$$G'_{\mathrm{L}}(\beta) = \frac{2b^2}{\theta^2}\left[2\beta\left(1 + \frac{b^2}{\theta^2}\beta^2\right) - \left(\beta + \frac{b^2}{\theta^2}\beta^3\right)\sec^2\left(\frac{\beta}{2}\right) + 2\left(1 + 3\frac{b^2}{\theta^2}\beta^2\right)\tan\left(\frac{\beta}{2}\right)\right]. \tag{2.26}$$

Now, using $\tan(\beta/2) \geq \beta/2$ and $\sec^2(\beta/2) \geq 1 + (\beta^2/4)$ over $\beta \in (0, \pi)$ in (2.26) and simplifying, we get the following upper bound:

$$G_{\mathrm{L}}^{'}(\beta) \leq -\frac{b^2}{2\theta^4}\left[(\theta^2 + 8b^2)\beta^3 + b^2\beta^5\right] \tag{2.27}$$

Clearly, for $\beta \in (0, \pi)$, the rhs of (2.27) is always negative which implies $G_{\mathrm{L}}^{'}(\beta) < 0$. It follows that $G_{\mathrm{L}}(\beta)$ is monotonically decreasing over $\beta \in (0, \pi)$ and (2.25) has a unique solution which corresponds to the global maximum of $D(\beta/\theta)$.The β_{L}^* that solves (2.25) can be found numerically and $\omega_{\mathrm{L}}^* = \beta_{\mathrm{L}}^*/\theta$.

2.5.2.4 Uniform Sensing Noise

For the uniform sensing noise, we have $\varphi_n(\omega) = \sin(\omega a)/\omega a$, where $\sigma_n^2 = a^2/3$. Substituting $\varphi_n(\omega)$ in (2.11) and letting $\beta = \omega a$ for convenience we have

$$D(\beta) = \frac{\left[1 - \cos\left(\frac{\beta\theta}{a}\right)\right]}{\left[\alpha\beta^2 \csc^2(\beta) - 1\right]} = C(\beta)\left[1 - \cos\left(\frac{\beta\theta}{a}\right)\right] \tag{2.28}$$

where $C(\beta) := 1/[\alpha\beta^2 \csc^2(\beta) - 1]$. Writing $D^{'}(\beta) = 0$ gives

$$\left[\alpha\beta^2 \csc^2(\beta) - 1\right] - \frac{2\alpha a\beta}{\theta \sin^2(\beta)} \tan\left(\frac{\theta\beta}{2a}\right)\left[1 - \beta\cot(\beta)\right] = 0 \tag{2.29}$$

with $\alpha := [1 + (\sigma_v^2/P_{\mathrm{T}})]$. Theorem 2.5.1 does not apply for the uniform sensing noise. However if $\theta/a \geq 2$, then using $C(\beta) \geq C(\beta + k\pi)$, $k = 1, 2, \ldots$, and using the periodicity of $[1 - \cos(\beta\theta/a)]$, we can show that $\beta_{\mathrm{U}}^* \in (0, \pi a/\theta]$. Following similar arguments to the Laplacian noise case, it can be shown that there is only one stationary point in $(0, \pi a/\theta]$ which corresponds to the global maximum. The β_{U}^* that solves (2.29) can be found numerically and therefore, $\omega_{\mathrm{U}}^* = \beta_{\mathrm{U}}^*/a$. On the other hand if $\theta/a < 2$, multiple local maxima are possible in $\beta \in (0, \pi a/\theta]$ and (2.29) can have multiple solutions. In this case, that β_{U}^* which yields the largest value for $D(\beta)$ in (2.28) should be chosen.

2.5.3 Per-sensor Power Constraint or high Channel SNR

We now consider the DC under the per-sensor power constraint. In this setting, as $L \to \infty$, $P_{\mathrm{T}} \to \infty$ which makes $(\sigma_v^2/P_{\mathrm{T}}) \to 0$. Therefore the DC for the per-sensor constraint when L is large becomes:

$$D_{\mathrm{p}spc}(\omega) = \frac{2\varphi_n^2(\omega)[1-\cos(\omega\theta)]}{[1-\varphi_n^2(\omega)]}. \tag{2.30}$$

Equation (2.30) can also be interpreted as the DC when $\sigma_v^2 = 0$ for any finite L. In what follows, we characterize ω^* in this per-sensor constraint regime, which effectively amounts to the removal of $(\sigma_v^2/P_{\mathrm{T}})$ from (2.11). In this case there is not necessarily a ω^* that attains the maximum in (2.12). Our first result reveals that (2.30) can be made large by choosing ω sufficiently close to zero when n_i are Gaussian, and yields an interesting relationship between the DC and the Fisher information.

Theorem 2.5.3. *When n_i are Gaussian,*

$$\sup_{\omega>0} D_{pspc}(\omega) = \frac{\theta^2}{\sigma_n^2} = \lim_{\omega\to 0} D_{pspc}(\omega) \tag{2.31}$$

Proof. We begin with the inequality $[1-\cos(\omega\theta)] \leq \omega^2\theta^2/2$. Consider [84, eqn (1)], which using the fact that $\varphi_n(\omega)$ is real-valued, reveals $\varphi_n^2(\omega) \leq (1+\varphi_n(2\omega))/2$. Using these two inequalities we can write the following:

$$\frac{1}{D_{\mathrm{p}spc}(\omega)} \geq \frac{[1-\varphi_n(2\omega)]}{2\omega^2\varphi_n^2(\omega)\theta^2}. \tag{2.32}$$

Now from [84, eqn (2)] with the fact that $\varphi_n(\omega)$ is real-valued, we have:

$$\frac{[1-\varphi_n(2\omega)]}{2\omega^2\varphi_n^2(\omega)} \geq \frac{1}{J} \tag{2.33}$$

where J is the Fisher information of n_i with respect to a location parameter [85, eqn (8)] (i.e., the Fisher information in x_i about θ). Combining (2.32) and

(2.33) we have:

$$\frac{D_{\mathrm{p}spc}(\omega)}{\theta^2} \leq J = \frac{1}{\sigma_n^2} \tag{2.34}$$

where the equality follows from the fact that for Gaussian random variables the Fisher information is given by the inverse of the variance. Now, we also see that using l'Hôspital's rule on (2.30), $\lim_{\omega\to 0} D_{\mathrm{p}spc}(\omega) = \theta^2/\sigma_n^2$, which shows that the inequality in (2.34) can be made arbitrarily tight establishing $\sup_{\omega>0} D_{\mathrm{p}spc}(\omega) = \theta^2/\sigma_n^2$. □

The proof of Theorem 2.5.3 also reveals an interesting inequality between the DC and the Fisher information, which of course is related to the Cramér-Rao bound for unbiased estimators. So for the per-sensor power constraint case with Gaussian noise, ω should be chosen as small as possible for the best performance and it does not depend on the value of θ.

For the Laplacian case, the solution is similar to the Gaussian case. It can be easily verified that, with $(\sigma_v^2/P_{\mathrm{T}}) = 0$, $D'_{\mathrm{p}spc}(\omega) < 0$ over $\omega \in (0, \pi/\theta)$. This means that $D_{\mathrm{p}spc}(\omega)$ is monotonically decreasing with ω which implies that ω should be chosen arbitrarily small.

On the other hand, when n_i are Cauchy distributed, then $\varphi_n(\omega) = e^{-\gamma\omega}$. Substituting in (2.30) and using l'Hôspital's rule we observe that $\lim_{\omega\to 0} D_{\mathrm{p}spc}(\omega) = 0$ for Cauchy sensing noise. This implies that, for the Cauchy sensing noise with per-sensor power constraint, smaller values of ω should be avoided for reliable detection to be possible.

2.5.4 Analysis of the DC for Non-homogeneous Sensors

Consider now the case where n_i are independent with non-identical distributions. This could occur if n_i have the same type of distribution (e.g. Gaussian)

with different variances. Letting $\varphi_{n_i}(\omega) = \mathrm{E}[e^{j\omega n_i}]$, the DC in (2.10) becomes

$$D_L(\omega) = \frac{2[1-\cos(\omega\theta)]\left(L^{-1}\sum_{i=1}^{L}\varphi_{n_i}(\omega)\right)^2}{\left[1-L^{-1}\sum_{i=1}^{L}\varphi_{n_i}^2(\omega)+\frac{\sigma_v^2}{P_{\mathrm{T}}}\right]} \tag{2.35}$$

which is now a function of L unlike in (2.11), and reduces to (2.11) if $\varphi_{n_i}(\omega) = \varphi_n(\omega)$, as in the iid case. We now study the conditions on the variances $\sigma_i^2 := \mathrm{var}(n_i)$ for $\lim_{L\to\infty} D_L(\omega) = 0$ for all $\omega > 0$. When this asymptotic DC is zero for all $\omega > 0$, the interpretation is that there is no suitable choice for $\omega > 0$. The following result establishes that if the sensing noise variances are going to infinity, the asymptotic DC is zero for all $\omega > 0$, indicating a regime where reliable detection is not possible.

Theorem 2.5.4. *Let $\varphi_{n_i}(\omega) = \varphi_n(\sigma_i\omega)$ for some CF $\varphi_n(\omega)$ where n has an absolutely continuous distribution. Suppose also that $\lim_{i\to\infty}\sigma_i = \infty$. Then $\lim_{L\to\infty} D_L(\omega) = 0$ for all $\omega > 0$.*

Proof. Clearly the denominator of (2.35) is bounded between $(\sigma_v^2/P_{\mathrm{T}})$ and $(1+\sigma_v^2/P_{\mathrm{T}})$. Therefore, it suffices to show that $L^{-1}\sum_{i=1}^{L}\varphi_{n_i}(\omega) = L^{-1}\sum_{i=1}^{L}\varphi_n(\sigma_i\omega) \to 0$ as $L\to\infty$. Since n has an absolutely continuous distribution, $\lim_{x\to\infty}\varphi_n(x) = 0$, and because $\lim_{i\to\infty}\sigma_i = \infty$, it follows that $\lim_{i\to\infty}\varphi_n(\sigma_i\omega) = 0$ for $\omega > 0$. From [86, pp. 411] we know that if a sequence satisfies $\lim_{i\to\infty} a_i = 0$ then its partial sums also satisfy $\lim_{L\to\infty} L^{-1}\sum_{i=1}^{L} a_i = 0$, which gives us the proof when applied to the sequence $\varphi_n(\sigma_i\omega)$. □

If, instead of $\sigma_i^2 \to \infty$ as $i\to\infty$, the variances σ_i^2 are bounded, we can show the existence of an $\omega > 0$ for which $\lim_{L\to\infty} D_L(\omega) > 0$ which is done next.

Theorem 2.5.5. *Let $var(n_i)$ exist for all i and $\sigma_{max} := \sup_i (var(n_i))^{1/2}$ be finite. Then any $0 < \omega < \sqrt{2}/\sigma_{max}$ satisfies $\lim_{L\to\infty} D_L(\omega) > 0$.*

Proof. To show $\lim_{L\to\infty} D_L(\omega) > 0$ for $\omega > 0$, it suffices to show that $L^{-1}\sum_{i=1}^{L} \varphi_{n_i}(\omega) > 0$ for $\omega > 0$. From [87, pp. 89] we have $\varphi_{n_i}(\omega) \geq 1 - \sigma_i^2\omega^2/2$ for any CF with finite variance. Therefore, $L^{-1}\sum_{i=1}^{L} \varphi_{n_i}(\omega) \geq 1 - (L^{-1}\sum_{i=1}^{L} \sigma_i^2)\omega^2/2 \geq 1 - \sigma_{max}^2\omega^2/2 > 0$ where the last inequality holds provided that $\omega < \sqrt{2}/\sigma_{max}$. □

This shows that if the noise variances are bounded, there exists (a small enough) ω that yields a strictly positive asymptotic DC, establishing that there is a choice of ω that enables reliable detection.

2.6 Fading Channels

Suppose that the channel connecting the i^{th} sensor and the FC has a fading coefficient $h_i := |h_i|e^{j\phi_i}$ normalized to satisfy $\mathrm{E}[|h_i|^2] = 1$. If the sensors do not know or utilize their local channel information, and the fading has zero-mean ($\mathrm{E}[h_i] = 0$), then the performance over fading channels is poor because the DC in (2.10) becomes zero due to law of large numbers and reliable detection is not possible. On the other hand, if the i^{th} sensor corrects for the channel phase before transmission, using local channel phase information, the received signal under the TPC becomes

$$y_L = e^{j\omega\theta}\sqrt{\frac{P_{\mathrm{T}}}{L}}\sum_{i=1}^{L} |h_i|e^{j\omega n_i} + v \ , \tag{2.36}$$

where we focus on the iid sensing noise case to highlight the effect of fading even though the non-homogeneous case can also be easily pursued. The phase correction does not change the constant power nature of the transmission. By calculating the expectations in (2.10), for the signal model in (2.36), the DC

in the presence of fading is given by:

$$D(\omega) = \frac{2(\mathrm{E}[|h_i|])^2\varphi_n^2(\omega)[1-\cos(\omega\theta)]}{\left[1-(\mathrm{E}[|h_i|])^2\varphi_n^2(\omega)+\frac{\sigma_v^2}{P_{\mathrm{T}}}\right]}. \tag{2.37}$$

We see that in case of fading, the term $\varphi_n^2(\omega)$ is scaled by the factor $(\mathrm{E}[|h_i|])^2$ in the DC expression. Since $\mathrm{E}[|h_i|^2] = 1$, using Jensen's inequality, the factor $(\mathrm{E}[|h_i|])^2 < 1$ unless $|h_i|$ is deterministic in which case it is one. Comparing (2.11) and (2.37) we have, with $(\mathrm{E}[|h_i|])^2 < 1$, the numerator of (2.37) is decreased and the denominator of (2.37) is increased, leading to a reduction in DC and thus fading has a detrimental effect on the detection performance, as expected.

Note that if the optimization of the DC is desired in the fading case, the factor $(\mathrm{E}[|h_i|])^2$ in the denomenator of (2.37) affects the optimum ω value. Theorem 2.5.1 can be proved for the fading case as well with $C(\omega) := 2(\mathrm{E}[|h_i|])^2\varphi_n^2(\omega)/[1-(\mathrm{E}[|h_i|])^2\varphi_n^2(\omega)+\sigma_v^2/P_{\mathrm{T}}]$ which is still decreasing with ω if $\varphi_n(\omega)$ is. Therefore the conclusion of Theorem 2.5.1, namely, $\omega^* \in (0, \pi/\theta)$, does not change. The procedure to find the ω^* under the TPC for Gaussian, Cauchy and Laplacian is the same as described in Sections 2.5.2.1, 2.5.2.2 and 2.5.2.3 respectively. The equations (2.17), (2.18), (2.19) and (2.22) remain valid with the exponentials in these equations scaled by the factor $(\mathrm{E}[|h_i|])^2$. The equations (2.25), (2.26) and (2.27) for the Laplacian case also remain valid except the term $1/\alpha$ in (2.25) scaled by $(\mathrm{E}[|h_i|])^2$.

We note that if sensors have imperfect knowledge of the phase, $|h_i|$ will be replaced by $|h_i|e^{j\widetilde{\phi}_i}$ where $\widetilde{\phi}_i$ is the phase error. Clearly this error can also be subsumed in (2.36) as replacing ωn_i with $\omega n_i + \widetilde{\phi}_i$ which changes the sensing noise by a term independent of ω. This establishes the interesting fact

that phase error over fading channels can be treated as a change in sensing noise distribution.

2.7 Asymptotic Performance and Optimization of ω based on error exponent

The error exponent in a distributed detection system is a measure of how fast the probability of error goes to zero as $L \to \infty$. Mathematically error exponent is defined as:

$$-\lim_{L\to\infty} \frac{\log P_{\mathrm{e}}(\omega)}{L}. \tag{2.38}$$

Large deviation theory [47, 88] provides a systematic procedure to calculate the error exponent which is briefly reviewed next. Let Y_L be a sequence of random variables without any assumptions on their dependency structure and let $M(t) = \lim_{L\to\infty}(1/L)\log \mathrm{E}\{e^{tY_L}\}$ exist and is finite for all $t \in R$. Define

$$\varepsilon(z) = -\lim_{L\to\infty} \frac{1}{L}\log \Pr\left[Y_L < z\right] \ , \tag{2.39}$$

where z is the threshold and Y_L is the test statistic of a detector. Gärtner-Ellis Theorem [88, pp. 14] states that $\varepsilon(z)$ in (2.39) can be calculated using,

$$\varepsilon(z) = \sup_{t\in R}[tz - M(t)] \ , \tag{2.40}$$

where

$$M(t) = \lim_{L\to\infty} \frac{1}{L}\log \mathrm{E}\{e^{tY_L}\}. \tag{2.41}$$

We will now use the Gärtner-Ellis Theorem with Y_L replaced by $Z_L(\omega)$ in (2.9) and $z = 0$. Letting $M_\omega(t) := \lim_{L\to\infty}(1/L)\log \mathrm{E}\{e^{tZ_L(\omega)}\}$, and $\varepsilon_\omega(z) = \sup_{t\in R}[tz - M_\omega(t)]$ we have the following theorem which relates the error exponent to the CF $\varphi_n(\omega)$ of the sensing noise distribution.

Theorem 2.7.1. *For the detector in (2.5), the error exponent in* (2.38) *is* $\varepsilon_\omega(0) = -\inf_{t\in R} M_\omega(t)$ *where* $M_\omega(t)$ *is given by*

$$\log\left[I_0(mt) + 2\sum_{k=1}^{\infty} I_k(mt)\varphi_n(k\omega)\cos\left(k\left(\frac{\pi}{2} - \frac{\omega\theta}{2}\right)\right)\right] + \left[\frac{t^2\sigma_v^2(1-\cos(\omega\theta))}{2P_T}\right] \tag{2.42}$$

where $I_k(t)$ *is the modified Bessel function of the first kind and* $m := 2\sin(\omega\theta/2)$.

Proof. We use the Gärtner-Ellis theorem from large deviation theory [88, pp. 14] to calculate the error exponent. To this end, we need to calculate $M_\omega(t)$ in (2.41) and substitute into (2.40).

$$\begin{aligned} M_\omega(t) &= \lim_{L\to\infty}\frac{1}{L}\log \mathrm{E}\{\exp[tZ_L]\} \\ &= \lim_{L\to\infty}\frac{1}{L}\log \mathrm{E}\left\{\exp\left[t\left(\sum_{i=1}^{L} 2\sin\left(\frac{\omega\theta}{2}\right)\right.\right.\right. \\ &\qquad \left.\left.\left.\cos\left(\omega n_i - \frac{\omega\theta}{2} + \frac{\pi}{2}\right) + \sqrt{\frac{L}{P_{\mathrm{T}}}}v^{\mathrm{T}}\right)\right]\right\} \\ &= \log \mathrm{E}\left\{\exp\left[2t\sin\left(\frac{\omega\theta}{2}\right)\cos\left(\omega n_i - \frac{\omega\theta}{2} + \frac{\pi}{2}\right)\right]\right\} \\ &\quad + \left[\frac{t^2\sigma_v^2(1-\cos(\omega\theta))}{2P_{\mathrm{T}}}\right] \end{aligned} \tag{2.43}$$

From [89, pp. 376], we have the Fourier series expansion of the periodic function $e^{p\cos(u)}$ as,

$$e^{p\cos(u)} = I_0(p) + 2\sum_{k=1}^{\infty} I_k(p)\cos(ku) \tag{2.44}$$

Using the equation (2.44) in (2.43) with $p = 2t\sin(\omega\theta/2)$ and $u = (\omega n_i - \omega\theta/2 + \pi/2)$ and then applying the expectation on the resulting summation, we get $M_\omega(t)$ as in (2.42). □

It is well known that the function $M_\omega(t)$ is convex in t [88]. Therefore the supremum in (2.40) can be found efficiently for $z = 0$. The t^* that maxi-

mizes (2.40) satisfies $M_{\omega}^{'}(t^*) = 0$ which can be found by convex methods with geometric convergence [90].

In addition to the error exponent, it is also possible to approximate $P_{\mathrm{e}}(\omega)$ using the function $\varepsilon_{\omega}(z)$. In fact Bahadur and Rao [47, pp. 10] have proved that this probability can be approximated using the error exponent and is given by:

$$P_{\mathrm{e}}(\omega) = \frac{1}{\sqrt{2\pi\hat{\sigma}_{\omega}^2}} e^{-L\varepsilon_{\omega}(0)(1+o(1))} , \tag{2.45}$$

as $L \to \infty$ and $\hat{\sigma}_{\omega}^2 := [\varepsilon_{\omega}^{'}(0)]^2/[\varepsilon_{\omega}^{''}(0)]$. The quantities $\varepsilon_{\omega}^{'}(0)$ and $\varepsilon_{\omega}^{''}(0)$ are the first and second derivatives of $\varepsilon_{\omega}(z)$ at $z = 0$ respectively, and can be calculated from the following equations [90, pp. 121]: $\varepsilon_{\omega}^{'}(0) = t^*$, and $\varepsilon_{\omega}^{''}(0) = 1/M_{\omega}^{''}(t^*)$.

The error exponent given in Theorem 2.7.1 is a function of ω and let us denote it by ε_{ω} for convenience. It will be illustrated in Section 2.9 that the values of ω that minimizes $P_{\mathrm{e}}(\omega)$ is closely predicted by the value obtained by maximizing $D(\omega)$ or ε_{ω}. We will also examine in the simulations in Section 2.9 how accurately (2.45) can be used to approximate $P_{\mathrm{e}}(\omega)$ for finite L.

2.8 Non-Gaussian Channel Noise

We have so far assumed that the channel noise as Gaussian. However, we verified that the detector in (2.5) works well even if the channel noise is mixed Gaussian, uniform or Laplacian. The channel noise distribution will only affect the error exponent through the second term in (2.42). Using this, the effect of different channel noise distributions we considered are briefly sketched below.

We considered the case of mixed Gaussian having two different variances drawn from a Bernoulli distribution. Let p_0 be the probability that the samples drawn from the mixture have variance $\sigma_{v_0}^2$ and $p_1 = 1 - p_0$ be the probability corresponding to $\sigma_{v_1}^2$ and let $\sigma_{v_1}^2 > \sigma_{v_0}^2$. In this case, we found that

the error exponent is affected only by the larger variance in the mixture. While using Gärtner-Ellis Theorem to calculate $M_\omega(t)$, the second term in (2.42) for the mixed Gaussian becomes $\lim_{L\to\infty} L^{-1}\log[p_0 \exp\left(t^2\sigma_{v_0}^2(1-\cos(\omega\theta))/2P_{\mathrm{T}}\right) + p_1 \exp\left(t^2\sigma_{v_1}^2(1-\cos(\omega\theta))/2P_{\mathrm{T}}\right)]$ and this limit in fact evaluates to the quantity $[t^2\sigma_{v_1}^2(1-\cos(\omega\theta))/2P_{\mathrm{T}}]$ which proves that only the larger variance $\sigma_{v_1}^2$ in the mixture affects the error exponent.

For the uniform channel case, interestingly we found that the second term in (2.42) evaluates to 0 and thus proving that the error exponent is not impacted by the uniform channel noise. We do not include the straightforward derivation due to lack of space. We will discuss the performance of the mixed Gaussian and Laplacian cases in Section 2.9.6.

2.9 Simulations

We define the sensing and channel SNRs as $\rho_s := \theta^2/\sigma_n^2$, $\rho_c := P_{\mathrm{T}}/\sigma_v^2$ and assume $P_1 = P_0 = 0.5$ throughout. Note also that $\rho = P_{\mathrm{T}}/L$ is the power at each sensor as defined in Section 2.2.

2.9.1 Effect of ω on Performance

We begin by comparing the optimized ω values using $D(\omega)$, ε_ω and $P_{\mathrm{e}}(\omega)$ for the TPC. The values of $\omega^* > 0$ obtained by maximizing the error exponent ε_ω and the DC $D(\omega)$ were found to be very close over the entire range of P_{T}. Figure 2.1 shows the plots of $D(\omega), \varepsilon_\omega$, and $P_{\mathrm{e}}(\omega)$ vs ω for Gaussian and Cauchy sensing noise distributions where the $P_{\mathrm{e}}(\omega)$ plot is obtained using Monte-Carlo simulations. The different ω^* values in Figure 2.1 correspond to the best ω values obtained by optimizing $D(\omega)$, ε_ω and $P_{\mathrm{e}}(\omega)$ respectively. It is interesting to see that the ω^* that minimizes $P_{\mathrm{e}}(\omega)$ is very close to that which maximizes $D(\omega)$ and ε_ω. For Laplacian and Uniform sensing noises (not shown), the same trends were observed.

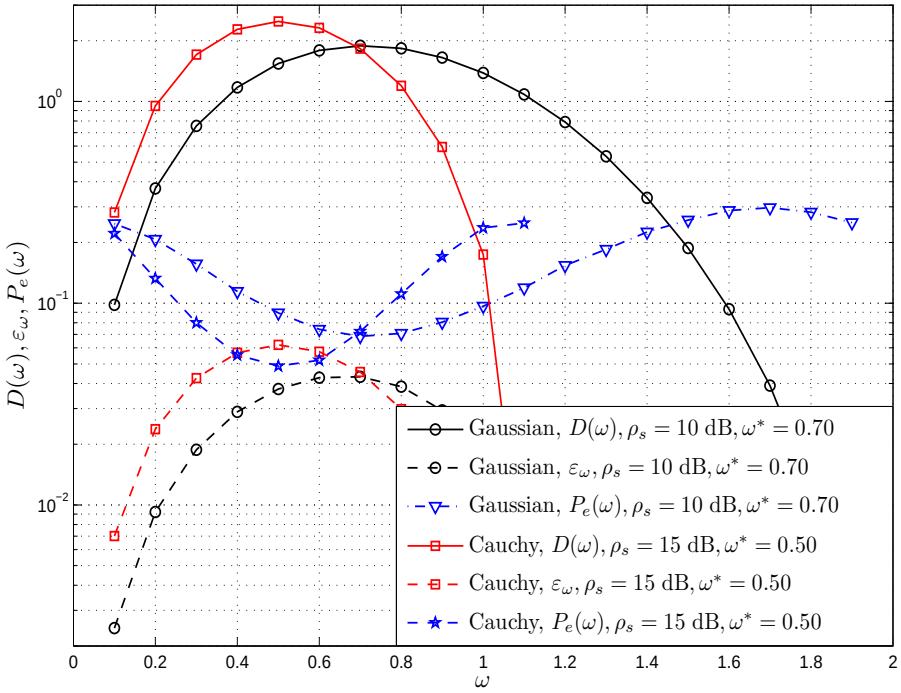

Figure 2.1: Total Power Constraint, $D(\omega), \varepsilon_\omega, P_{\mathrm{e}}(\omega)$ vs ω: ρ_s=10, 15 dB, ρ_c=-10 dB, L=20

Figure 2.2 shows the performance under per-sensor power constraint with large L. It is observed that smaller ω yields better error probability. This agrees with our findings in Section 2.5.3 where it was shown that $D_{\mathrm{p}spc}(\omega)$ can be made larger by choosing $\omega > 0$ arbitrarily small. Since both Figures 2.1 and 2.2 verify that the choice of ω based on minimizing $P_{\mathrm{e}}(\omega)$ can be closely approximated by that which maximizes $D(\omega)$, in all subsequent simulations, we have used the ω^* values obtained by maximizing $D(\omega)$.

2.9.2 Comparison against MAF and MDF Schemes

In Figure 2.3, the proposed scheme is compared under the TPC with the MAF and MDF schemes which have been shown in [37] to outperform conventional amplify-and-forward (AF) and detect-and-forward (DF) schemes. We observe that the proposed scheme outperforms MAF when $\rho_s > 4$ dB, and MDF for the

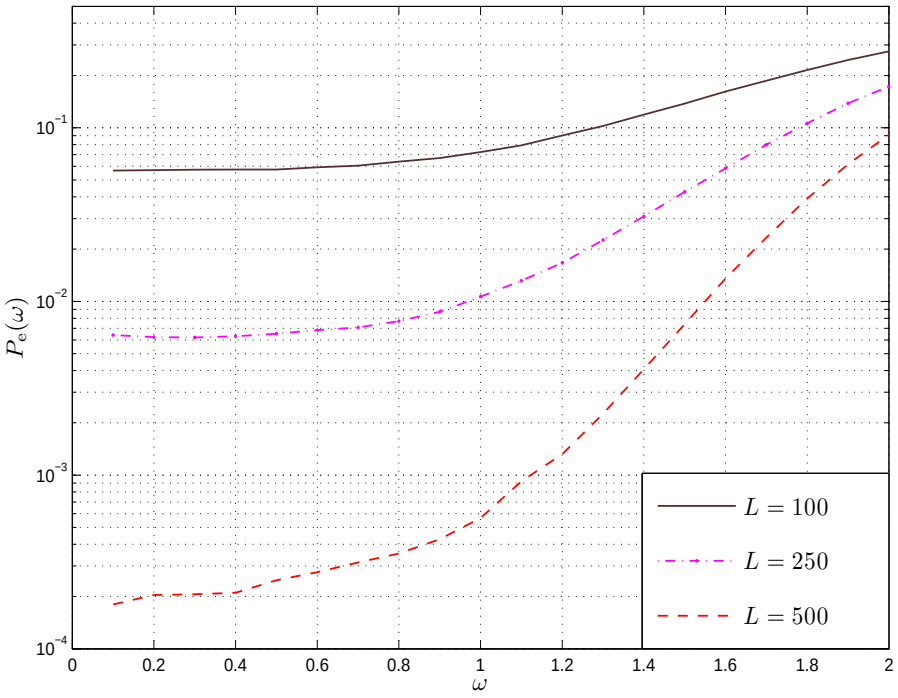

Figure 2.2: Per-sensor Power Constraint, Gaussian, $P_e(\omega)$ vs ω: ρ_s=-10 dB, ρ=10 dB

entire range of ρ_s. The same trend was observed when L is increased to 90 with an improvement in the detection error probability. The ML performance shown was obtained by the Monte-Carlo implementation of the ML detector and is computationally complex, but serves as a performance benchmark. Figure 2.4 shows the P_e performance versus L under the TPC. Clearly the proposed scheme outperforms the AF, DF, MAF and MDF schemes consistently since $\rho_s = 15$ dB.

The proposed scheme requires the fine tuning of the transmission phase parameter ω either through optimizing the deflection coefficient or the error exponent. However, it should be noted that similar type of fine tuning is also required in the competing schemes such as the MAF or the MDF. We note that the proposed scheme is inferior to MAF at low sensing SNRs ($\rho_s < 4$

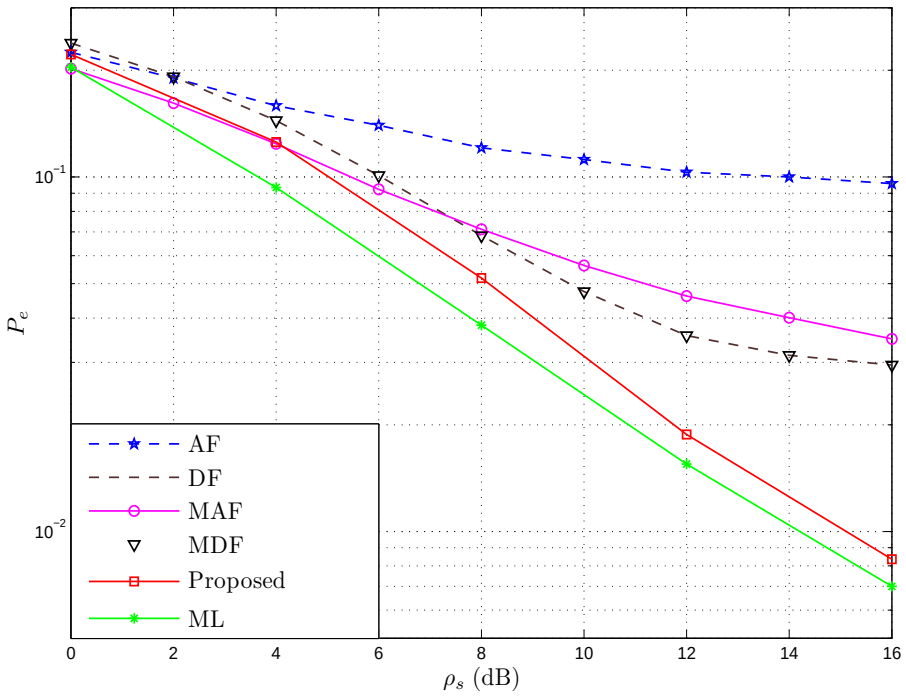

Figure 2.3: Total Power Constraint, P_{e} vs ρ_s: ρ=-30 dB, L=60

dB). On the whole, the benefits of constant modulus signaling and improved performance at higher sensing SNRs make the proposed approach a viable alternative.

2.9.3 Total Power Constraint: Different Noise Distributions

For the Total Power Constraint, Figure 2.5 shows that Cauchy sensing noise results in better performance when ρ_s is low, and worse when ρ_s is high compared with other sensing noise distributions. This agrees with the fact that $D(\omega^*)$ is smaller for Cauchy sensing noise when ρ_s is high than other distributions and vice versa when ρ_s is low. When ρ_s is moderately high, we observe that Gaussian, Laplacian and Uniform distributions have identical performance if ρ_c is very low for a wide range of L as illustrated in the Figure 2.5. We found numerically that the similarity of the $P_{\mathrm{e}}(\omega)$ curves under dif-

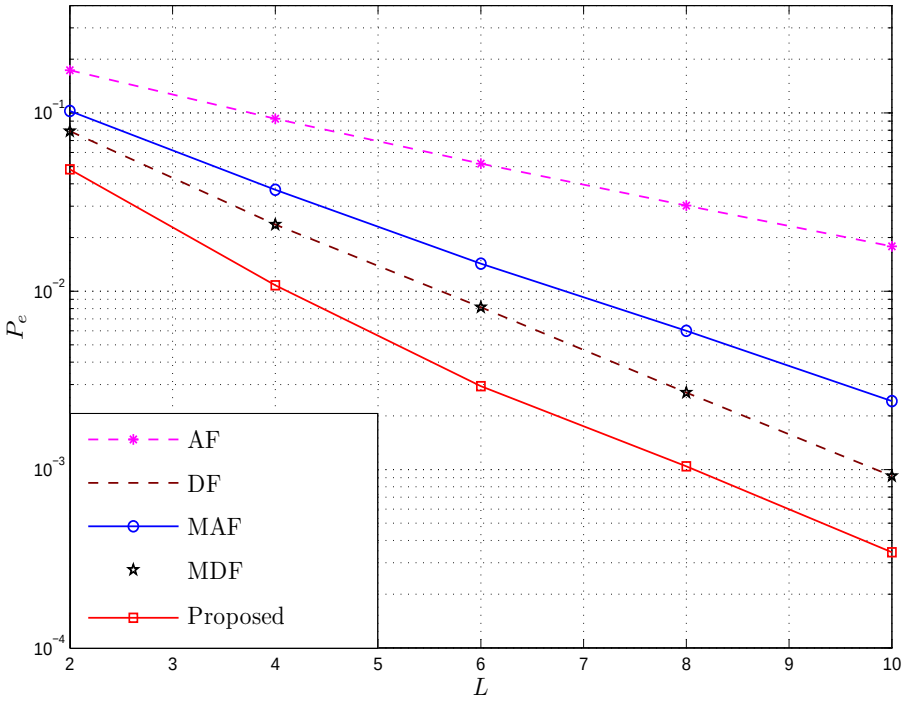

Figure 2.4: Total Power Constraint, P_{e} vs L: ρ_s=15 dB, ρ_c=0 dB

ferent sensing noise distributions was also reflected in the corresponding $D(\omega)$ values where they were also verified to be similar.

Figure 2.6 compares the performance of the proposed scheme in the presence of Rayleigh flat fading between the sensors and the FC against without fading with the Gaussian sensing noise. Clearly, fading has a detrimental effect on the detection performance as argued in Section 2.6. It is also observed that, in the presence of fading, P_{e} is not as sensitive to the increase in ρ_s as that of the no fading case.

2.9.4 Error Exponent

Figure 2.7 depicts the error exponent of the proposed scheme under the PSPC and illustrates its improvement with increase in ρ_s for all the sensing noise distributions. Recall that σ_v^2 has no effect on the error exponent for the PSPC case since $(\sigma_v^2/P_{\mathrm{T}}) \to 0$ in (2.42). It is interesting to see that Cauchy sens-

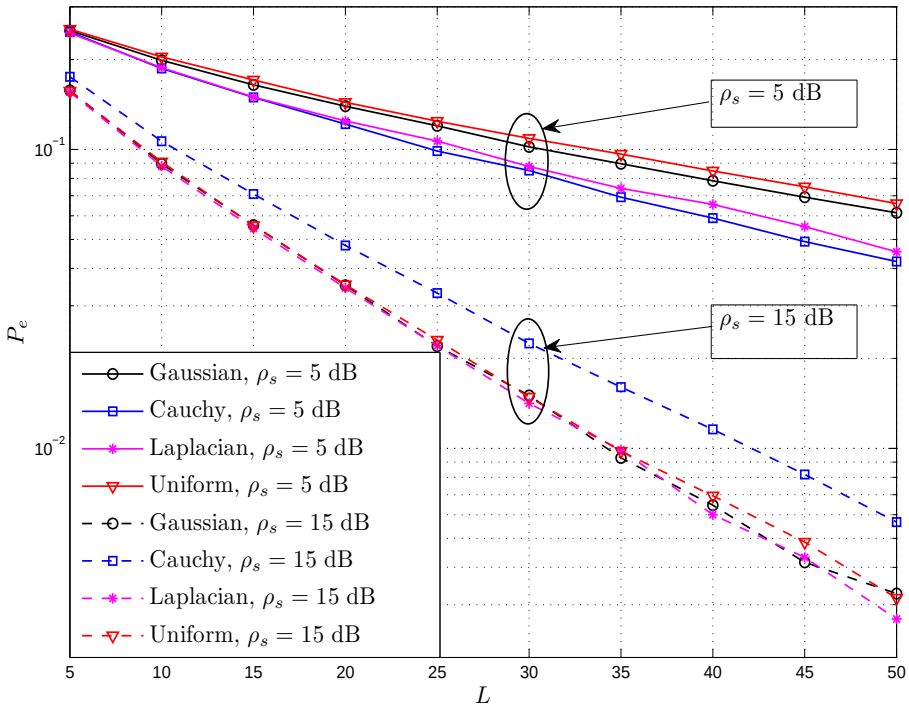

Figure 2.5: Total Power Constraint, P_{e} vs L: ρ_s=5, 15 dB, ρ_c=-10 dB

ing noise has a better error exponent than Gaussian, Laplacian and Uniform sensing noise distributions when $\rho_s \leq 4$ dB while it is worse when $\rho_s > 4$ dB. The error exponent with Gaussian sensing noise is better than that of Laplacian noise when when $\rho_s > 7.5$ dB and the uniform distribution has a better error exponent than other sensing noise distributions when $\rho_s > 4$ dB. The error exponent of the proposed scheme is compared with the error exponents of MAF and MDF schemes which were only derived for the Gaussian case (please see equations (24) and (25) in [37]). It is seen that, for the PSPC case, the MAF scheme (whose error exponent is $\rho_s/8$) and the proposed scheme with optimum ω have identical error exponents leading us to conjecture that $\sup_\omega[-\inf_{t\in R} M_\omega(t)] = \rho_s/8$ when n_i are Gaussian. The MDF error exponent is inferior compared to MAF and the proposed scheme.

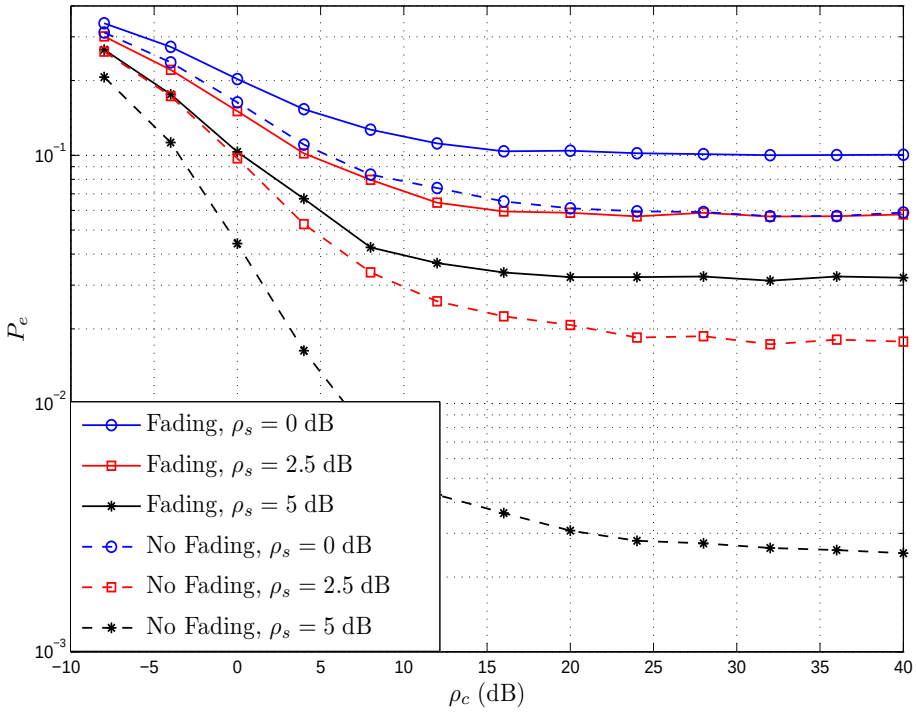

Figure 2.6: Rayleigh flat fading, P_{e} vs ρ_c, n_i Gaussian, L=10

Figure 2.8 shows the error exponent under the TPC with $\rho_s = 0$ dB. In this scenario, Cauchy sensing noise has the best error exponent since ρ_s is low. This concurs with the fact illustrated in Section 2.9.3 that the DC of Cauchy is better at lower values of ρ_s than other distributions and this was justified by the simulation results as shown in Figure 2.5. We found that when ρ_s is increased, Cauchy becomes inferior to other noise distributions. For all the distributions, increasing ρ_c results in an increase in the error exponent which becomes a constant beyond $\rho_c = 15$ dB. This is because, for a given ρ_s, increasing ρ_c combats the effect of channel noise, thereby improving the error exponent. However, the effect of sensing noise can not be overcome by increasing ρ_c indefinitely. This can be seen from (2.42) as well where the second term vanishes while the first term remains even for large P_{T}. For the Gaussian case, we derived the error exponent of the MAF scheme under the

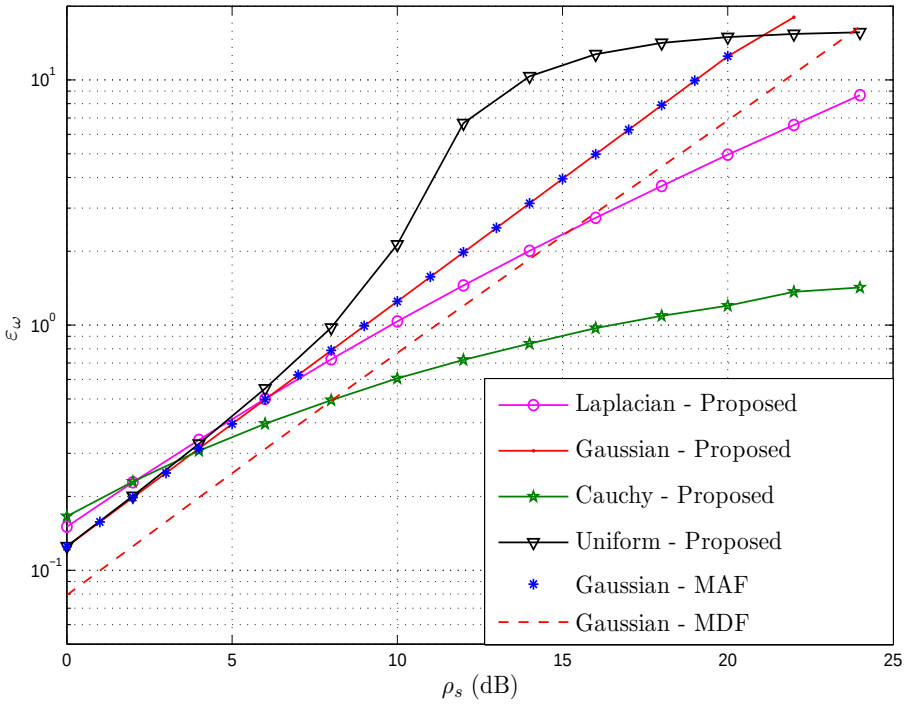

Figure 2.7: Per-sensor Power Constraint, ε_ω vs ρ_s

TPC as $\varepsilon_{MAF} = \theta^2/8[\sigma_n^2 + (\sigma_v^2(\sigma_n^2 + P_0 P_1 \theta^2)/P_{\mathrm{T}})]$. If $P_{\mathrm{T}} \to \infty$, this reduces to $\rho_s/8$ for the PSPC case. It is seen that with $\rho_s = 0$ dB, the MAF scheme is better than the proposed method when $\rho_c < 15$ dB. However, under the TPC, the error exponent of the proposed scheme was found to beat the MAF scheme when $\rho_s > 4.5$ dB and an example plot is shown in Figure 2.8 for $\rho_s = 10$ dB. This crossover between the MAF and the proposed schemes is also reflected in their respective P_{e} performance curves approximately around the same ρ_s value (please see Figure 2.3). However, if ρ_c is increased beyond 15 dB, we see that the error exponents of both the schemes become very close.

2.9.5 Approximations of $P_e(\omega)$ through $\varepsilon_\omega(z)$

Equation (2.45) provides an approximation of $P_{\mathrm{e}}(\omega)$ based on the error exponent. The expression in (2.45) is found to match with the simulations when $\rho_c > 0$ dB and $\rho_s >$ -5 dB. Figures 2.9 and 2.10 elucidate this behavior for

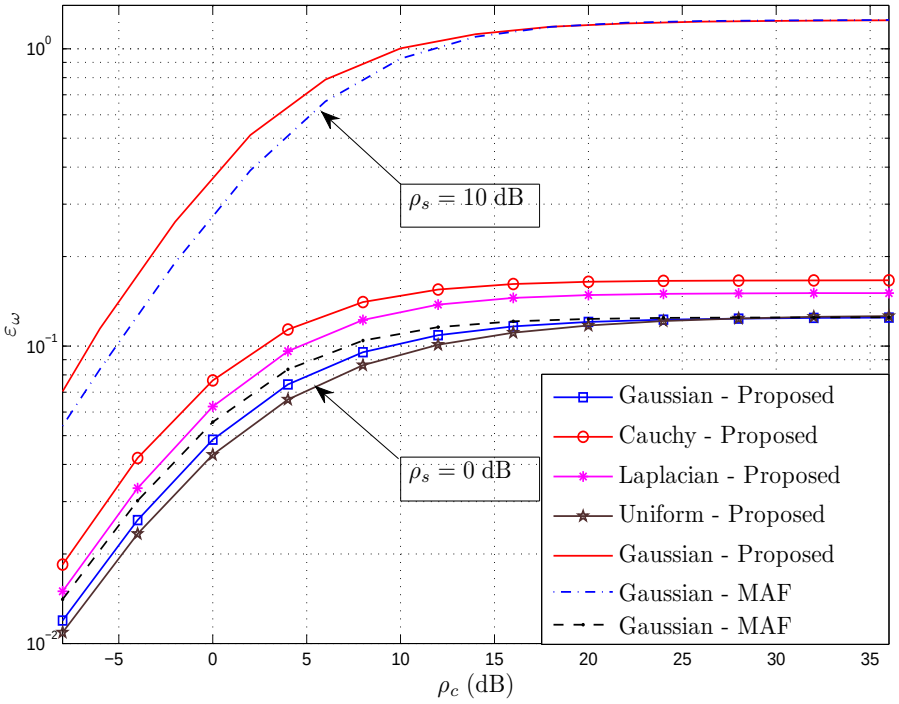

Figure 2.8: Total Power Constraint, ε_ω vs ρ_c: ρ_s=0, 10 dB, σ_v^2=1

Gaussian sensing noise distribution. Similar trends were observed for the other sensing noise distributions as well but are not shown due to space constraints. When L is small, the gap between theory and simulation is significant as shown in Figure 2.9. This can be explained by the $o(1)$ term in (2.45). Accordingly, when L is increased to about 40, we see the theory and simulation curves merging as shown in Figure 2.9. Figure 2.10 shows that when ρ_s is moderately high, smaller L is required to get the performance match between theory and simulation.

From the various simulation plots in Figures 2.1, 2.5, 2.7, and 2.8, we see that the proposed scheme is robust in the sense that it works very well for a variety of sensing noise distributions including the impulsive Laplacian distribution and the Cauchy distribution which has no finite moments.

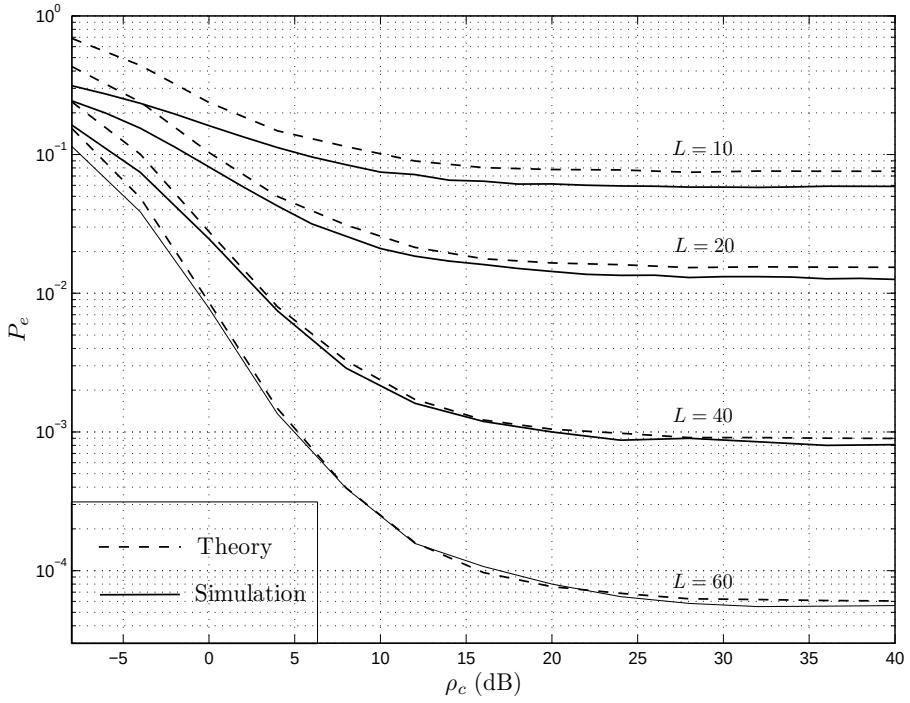

Figure 2.9: Gaussian Sensing Noise: P_{e} vs ρ_c: ρ_s=0 dB, L=10, 20, 40, 60

2.9.6 Non-Gaussian Channel Noise

Figure 2.11 shows the error exponent plot for the case where $\sigma^2_{v_0} = 0.25, p_0 = 0.80, \sigma^2_{v_1} = 4, p_1 = 0.20$ (note that the effective channel noise variance is: $\sigma^2_{v_{eff}} = p_0\sigma^2_{v_0} + p_1\sigma^2_{v_1} = 1$). We see that the error exponent of mixed Gaussian with $\sigma^2_{v_{eff}} = 1$ is worse compared to that of the Gaussian with $\sigma^2_v = 1$ case. This is because, in the mixed Gaussian case, the error exponent is a function of the larger variance of $\sigma^2_{v_1} = 4$.

Figure 2.12 shows the performance of the proposed detector with Laplacian channel noise against the Gaussian channel noise when the sensing noise is Gaussian. We note that when sensing SNR ρ_s is moderately high, the impulsive Laplacian channel noise is worse compared to Gaussian channel noise.

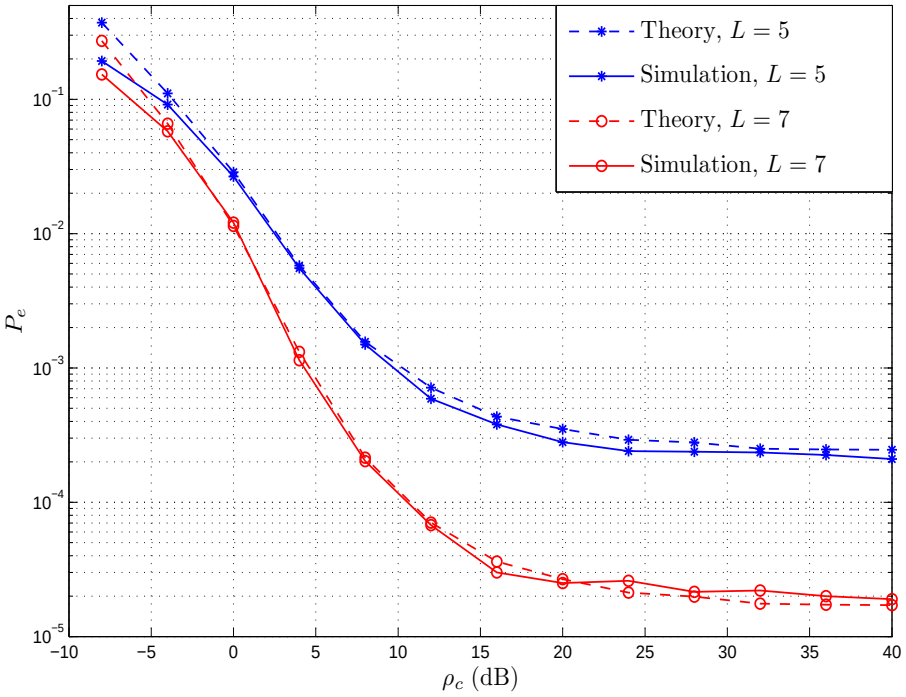

Figure 2.10: Gaussian Sensing Noise: P_{e} vs ρ_c: ρ_s=10 dB, L=5, 7

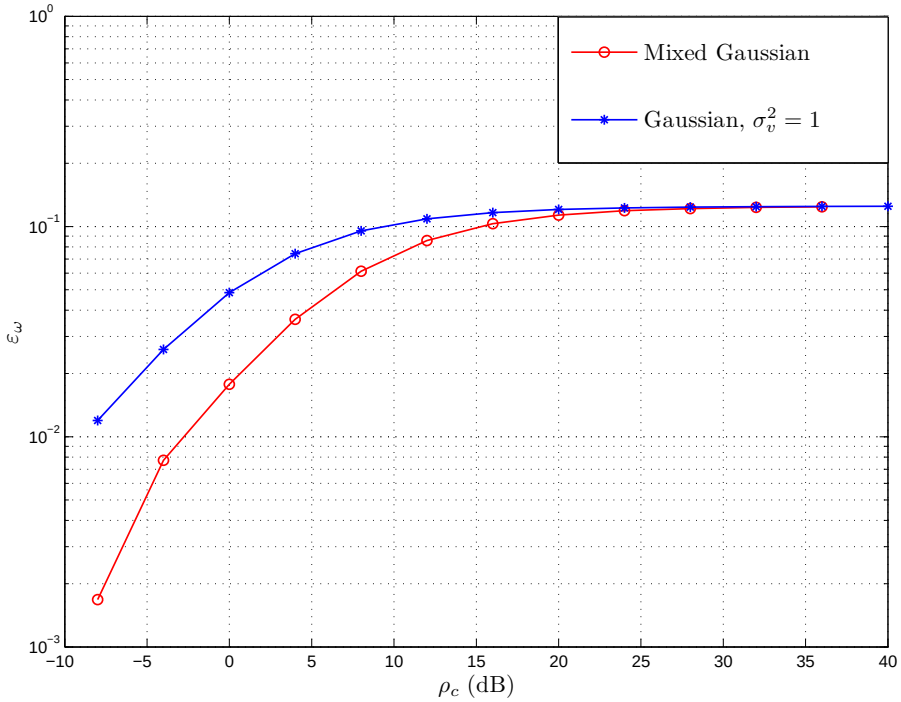

Figure 2.11: Mixed Gaussian channel noise: ε_ω vs ρ_c: $\sigma^2_{v_0}$=0.25, $p_0 = 0.8$, $\sigma^2_{v_1}$=4, $p_1 = 0.2$, $\sigma^2_{v_{eff}} = 1$, ρ_s=0 dB.

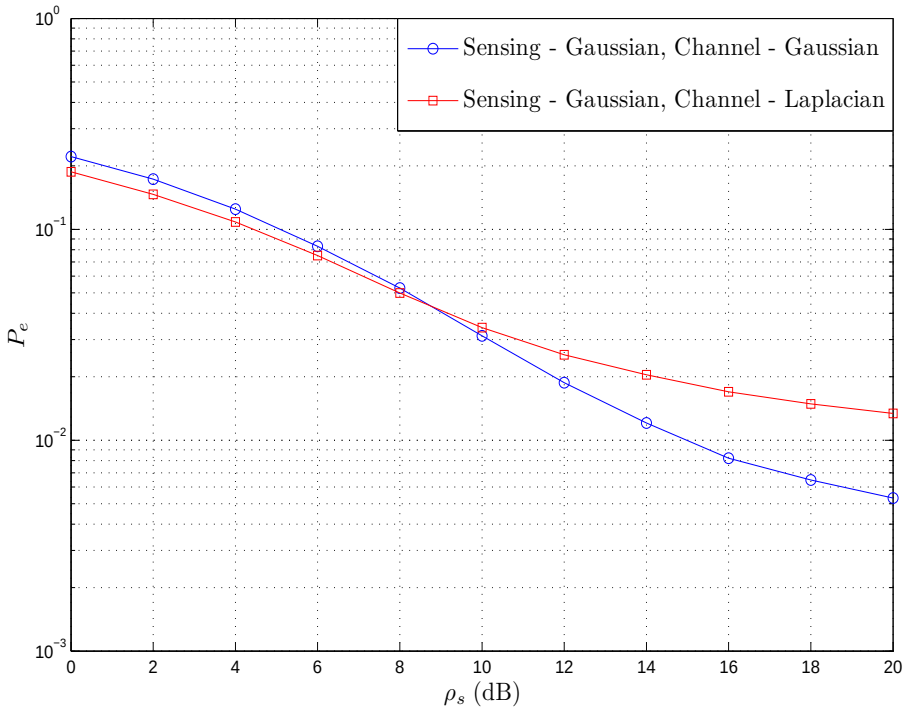

Figure 2.12: Gaussian Sensing Noise: P_{e} vs ρ_s: P_{T}=-12.22 dB, L=60

Chapter 3

Distributed Inference with Bounded Transmissions

3.1 Literature Survey and Motivation

In this chapter we consider the general problem of distributed inference using bounded transmission functions and establish regimes under which estimation and detection will be possible and discuss regimes under which estimation will fail and reliable detection will be impossible.

In inference-based wireless sensor networks (WSNs), low-power sensors with limited battery and peak-power capabilities transmit their observations to a fusion center (FC) for detection of events or estimation of parameters. For distributed estimation and distributed detection, much of the literature has focused on a set of orthogonal (parallel) channels between the sensors and the FC (please see [22,91] and the references therein). The bandwidth requirements of such an orthogonal WSN scale linearly with the number of sensors. In contrast, over multiple access channels where the sensor transmissions are simultaneous and in the same frequency band, the utilized bandwidth does not depend on the number of sensors.

Sensors may adopt either a digital or analog method for relaying the sensed information to the FC. The digital method consists of quantizing the sensed data and transmitting with digital modulation over a rate-constrained channel. In this case, the required channel bandwidth is proportional to the number of bits at the output of the quantizer which are transmitted after pulse shaping and digital modulation. The analog method consists of transmitting unquantized data by appropriately pulse shaping and amplitude or phase modulating to consume finite bandwidth. One such method is the amplify-and-forward (AF) scheme in which sensors send scaled versions of their measure-

ments to the FC. However, using the AF technique is not a viable option for WSNs because it requires high transmission power when the values to be transmitted are large [92]. Moreover, the linear transmit amplifier characteristics required for AF are often very power-inefficient [93], requiring the study of the effect of nonlinear transmissions on performance. Distributed systems which employ the AF technique for transmission of the sensed data often assume that the power amplifiers used are perfectly linear over the entire range of the sensed observations. In practice, the amplifiers exhibit nonlinear behaviour when the amplitude of the sensed data is relatively high [93–95]. Wireless sensor networks have stringent power and bandwidth constraints, therefore distributed schemes which use bounded instantaneous transmit power over multiple access channels are highly desirable.

References [38,42–44,70,96,97] discuss distributed estimation over Gaussian multiple access channels. In [38, 42], a distributed estimation scheme where the sensor transmissions are phase-modulated to make constant modulus transmissions is considered. The estimator proposed in [38] is shown to be strongly consistent for any symmetric sensing noise distribution when the noise samples are i.i.d.. In [43, 44], the mean and variance of a signal embedded in noise (not necessarily Gaussian) are estimated which are then used to estimate the SNR of the signal. In all the above cases of [38, 42–44], the desired constant modulus property is achieved by phase modulating the sensed data before transmission. The authors in [98] discuss the effect of nonlinear transmissions on the convergence speed of a consensus algorithm proposed for a distributed average consensus problem.

References [36, 37, 99, 100] discuss distributed detection using constant modulus transmissions over Gaussian multiple access channels for a binary

hypothesis testing problem. Inspired by the robustness of the estimation scheme in [38], the authors in [99] and [100] proposed a distributed detection scheme where the sensors transmit with constant modulus signals over a Gaussian multiple access channel. Here again, the sensors transmit with constant modulus transmissions whose phase varies linearly with the sensed data and the performance is analysed using deflection coefficient and error exponent. In [36] and [37], two schemes called modified amplify-and-forward (MAF) and the modified detect-and-forward (MDF) are developed which generalize and outperform the classic amplify-and-forward (AF) and detect-and-forward (DF) approaches to distributed detection. It is shown that MAF outperforms MDF when the number of sensors is large and the opposite conclusion is true when the number of sensors is smaller. In both the DF and MDF schemes, the sensors individually take a decision by quantizing the sensed measurement and transmit the one bit information to the FC by BPSK modulation and therefore the transmit power is always constant. Bounded transmission schemes are highly desirable and practically viable for the power constrained WSNs. In addition, bounded transmissions are robust to impulsive measurements [38,42–44] which could happen for WSNs deployed in adverse conditions.

In this chapter, we are interested in studying the effect of general nonlinear transmissions (as opposed to the linear AF scheme) from the sensors to the FC in a distributed inference framework. The sensors map their observations using a bounded function before transmission to constrain the transmit power and these observations are transmitted to the FC over a Gaussian multiple access channel. Our emphasis in this Chapter is not so much to propose a specific estimator or a detector, rather we want to focus on studying the implications of bounded transmission schemes on distributed inference in resource

constrained WSNs. Moreover, this work also studies the merits and demerits of distributed schemes involving realistic, nonlinear amplifier characteristics. We characterize the general conditions on the sensing noise statistics and the nonlinear function under which consistent estimation and reliable detection are possible. We also compare the bounded transmissions scheme with the AF scheme and highlight the advantages and disadvantages of each other. We show that if the measurement accuracy degrades progressively in the sense that the sensing noise variance goes to infinity, bounded transmission is not useful for distributed inference. On the other hand, it is shown that AF scheme does not suffer from this issue. These conclusions are drawn by studying the fundamental metrics such the asymptotic variance and the deflection coefficient.

This chapter is organized as follows. In Section 3.2, the system model is described with the total power constraint. In Section 3.3, the estimation problem is described and the properties of the proposed estimator is studied. Optimization of $AsV(\omega)$ is considered in this section. In Section 3.3, the detection problem is described and a quadratic detector is proposed. The probability of error performance of the detector is analyzed and the optimization of DC is studied in this section. The proposed scheme is compared against the AF scheme in Section 3.2.4. Simulation results are provided in Section 3.4 which support the theoretical results.

3.2 Distributed Estimation with Bounded Transmissions

3.2.1 System Model

Consider the sensing model, with L sensors,

$$x_i = \theta + \sigma_i n_i \qquad\qquad i = 1, \ldots, L \tag{3.1}$$

where θ is an unknown real-valued parameter, n_i is symmetric real-valued noise with zero median (i.e., its probability density function (PDF) is sym-

metric about zero), and x_i is the measurement at the i^{th} sensor. The noise samples n_i are assumed to be independent identically distributed (i.i.d.) but not necessarily with finite mean or variance. We consider a setting where the i^{th} sensor transmits its measurement using a bounded function $\sqrt{\rho}f(x_i)$ over a Gaussian multiple access channel (please see Figure 3.1) so that the received signal at the FC is given by

$$y_{_L} = \sqrt{\rho}\sum_{i=1}^{L} f\left(x_i\right) + v \tag{3.2}$$

where ρ is a power scale factor and v is the additive Gaussian noise with zero mean and variance σ_v^2. Parameter σ_i is a deterministic scale parameter which makes the variance (when it exists) of the noise samples different for each sensor depending on how they are distributed in space and how accurate their measurements are. For instance, if the phenomenon quantified by θ happens near a sensor, it is reasonable to expect that the variances of the sensing noise would be higher compared to those that are farther. Moreover, in case of WSNs operating in adverse conditions, the sensing noise n_i could be impulsive characterized by heavy tailed distributions [101]. We also point out that the received signal at the FC as modeled in (3.2) is realistic if the transmit amplifiers at the local sensors are nonlinear.

In this chapter, we study the consequences of the boundedness of $f(\cdot)$ on performance. In particular, we assume that the transmit function $f(x)$ satisfies the following conditions.

Assumptions:

(A1): $f(x)$ is differentiable such that $0 < f^{'}(x) \leq d$, $\forall x \in \mathbb{R}$.

(A2): $f(x)$ is bounded, $\sup_{x\in\mathbb{R}} |f(x)| = c$.

Note that the transmitted signal at the i^{th} sensor has the instantaneous power $\rho f^2(\omega x_i)$ and it is always constrained within ρc^2, which does not suffer from

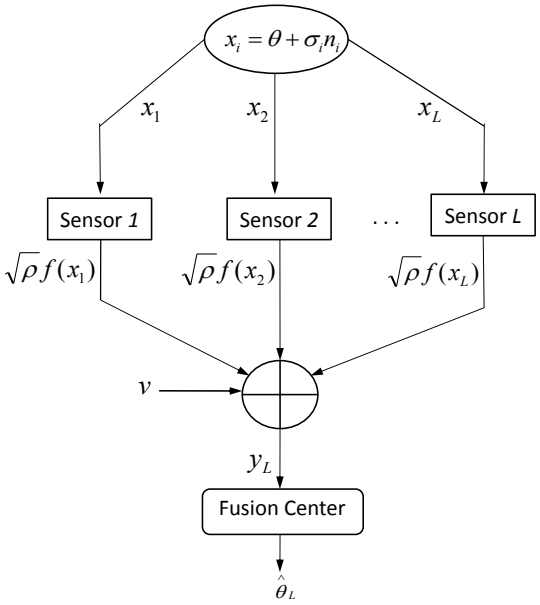

Figure 3.1: Bounded transmissions over Gaussian multiple access channel

the problems of unbounded transmit power seen in AF schemes for which $f(x) = \alpha x$. The total transmit power from all the sensors in (3.2) is upper bounded by $\rho c^2 L$. We begin by considering a fixed total power constraint P_{T} for the entire network implying that the per-sensor power is less than or equal to P_{T}/L. Clearly the per-sensor power is a function of L when P_{T} is fixed.

3.2.2 The Estimation Problem

First we consider estimating θ from $y_{_L}$. Let σ_i be a deterministic sequence capturing the reliability of the i^{th} sensor's measurement. The received signal $y_{_L}$ under the total power constraint is given by

$$y_{_L} = \sqrt{\frac{P_{\mathrm{T}}}{L}} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) + v. \tag{3.3}$$

Let $z_{_L}$ denote the normalized received signal:

$$z_{_L} := \frac{y_L}{\sqrt{L}} = \sqrt{P_{\mathrm{T}}} \frac{1}{L} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) + \frac{v}{\sqrt{L}}\ , \tag{3.4}$$

and define $h_\omega(\theta) := L^{-1}\sum_{i=1}^{L} \mathrm{E}_{n_i}\left[f(\theta+\sigma_i n_i)\right]$ where $\mathrm{E}(\cdot)$ denotes expectation. We now state an important result known as the Kolmogorov's strong law of large numbers [102, pp. 259] which handles the case of independent non-identically distributed RVs, due to the fact that the σ_i are different.

Theorem 3.2.1. *Let X_1, X_2, ..., X_L be a sequence of independent and not necessarily identically distributed RVs. Let var$[X_k]$ denote the variance of X_k and $\bar{X}_L = L^{-1}\sum_{k=1}^{L} X_k$ denote the partial sum of the sequence. If $\sum_{k=1}^{\infty} var[X_k]/k^2 < \infty$, then, $\bar{X}_L - E[\bar{X}_L] \to 0$ almost surely as $L \to \infty$.*

Due to the law of large numbers in Theorem 3.2.1 we have

$$\lim_{L\to\infty} \frac{1}{L}\sum_{i=1}^{L} f(\theta+\sigma_i n_i) = h_\omega(\theta) \tag{3.5}$$

where we use the fact that the variances $\mathrm{var}[f(\theta+\sigma_i n_i)] \leq c^2$ are bounded. Therefore, we have $\lim_{L\to\infty} z_{_L} = \sqrt{P_{\mathrm{T}}}h_\omega(\theta)$. Due to the boundedness of $f(\cdot)$, (3.5) holds regardless of the sensing noise distributions. Consider estimating θ from,

$$\widehat{\theta}_L = h_\omega^{-1}\left(\frac{z_{_L}}{\sqrt{P_{\mathrm{T}}}}\right), \tag{3.6}$$

where $z_{_L}$ is as given in (3.4). To recover θ uniquely from $h_\omega^{-1}(\cdot)$, we need $h_\omega(\theta)$ to be one-to-one in θ for which **(A1)** and **(A2)** are sufficient as shown in Lemma 3.2.1.

Lemma 3.2.1. *Let $g_{\sigma_i}(\theta) := E_{n_i}\left[f(\theta+\sigma_i n_i)\right]$ and suppose that the assumptions **(A1)** and **(A2)** hold. Then, $h_\omega(\theta)$ is one-to-one in θ.*

Proof. Differentiating $g_{\sigma_i}(\theta)$ with respect to θ, we have

$$g_{\sigma_i}(\theta) = \int_{-\infty}^{\infty} f(\theta + \sigma_i n_i) p(n_i) dn_i \ ,$$

$$\frac{\partial g_{\sigma_i}(\theta)}{\partial \theta} = \int_{-\infty}^{\infty} \frac{\partial f(\theta + \sigma_i n_i)}{\partial \theta} p(n_i) dn_i \ , \tag{3.7}$$

$$> 0 \ , \tag{3.8}$$

where we have applied Corollary 5.9 in [103, pp. 46] using assumptions **(A1)** and **(A2)** to move the derivative inside the integral in (3.7). The last inequality follows from the fact that convex combination of positive valued functions is positive. Therefore, $g_{\sigma_i}(\theta)$ is strictly an increasing function of θ. Since $h_\omega(\theta)$ is a convex combination of strictly increasing and differentiable functions, we have $h^{'}(\theta) > 0, \theta > 0$. Therefore, $h_\omega(\theta)$ is a strictly increasing function and thus it is one-to-one in θ. □

We now state a Lemma about a convergent sequence which will be used in the sequel.

Lemma 3.2.2. *Let a_i be a converging sequence such that $\lim_{i\to\infty} a_i = a$. Then, $\lim_{L\to\infty} L^{-1} \sum_{i=1}^{L} a_i = a$.*

Proof. Please see [86, pp. 411]. □

An estimator $\widehat{\theta}_L$ is strongly consistent if $\widehat{\theta}_L$ converges to the true value θ almost surely as $L \to \infty$ [86]. Now we establish the strong consistency of the class of estimators $\widehat{\theta}_L$ in (3.6) In Theorem 3.2.2.

Theorem 3.2.2. *Let the assumptions **(A1)** and **(A2)** hold. Let $\sigma_{max} := \max_i \sigma_i$ be finite. Then, the estimator $\widehat{\theta}_L$ in (3.6) is strongly consistent.*

Proof. Since $f(x)$ is a bounded function by assumption **(A2)**, the Kolmogorov's condition $\sum_{i=1}^{\infty} \text{var}[f((\theta + \sigma_i n_i)]/i^2 \leq \rho^2 c^2 \pi^2/6 < \infty$ is satisfied, therefore the strong law of large numbers for the non-identically distributed random variables (RVs) is applicable and $z_{_L} \to \sqrt{P_{\text{T}}} h_\omega(\theta)$ almost surely. Since $f^{'}(x) > 0$ by assumption **(A1)**, it follows from Lemma 3.2.1 that $h(\theta)$ is one-to-one in θ. Due to the fact that $\widehat{\theta}_L$ is a continuous function of $z_{_L}$, $\widehat{\theta}_L \to \theta$ almost surely [86, Thm 3.14] proving that the estimator in (3.6) is strongly consistent. □

On the other hand, if $\sigma_i \to \infty$ as $i \to \infty$, then the estimator in (3.6) is not consistent and θ can not be estimated from $z_{_L}$. A more formal statement is presented next as a theorem.

Theorem 3.2.3. *Let the assumptions **(A1)** and **(A2)** hold and σ_i be a deterministic sequence such that $\sigma_i \to \infty$ as $i \to \infty$, then $h_\omega(\theta)$ is independent of θ.*

Proof. First we note that due to assumption **(A2)**, the variances $\text{var}[f((\theta + \sigma_i n_i)]$ are bounded. According to Kolmogorov's strong law of large numbers

for non-identically distributed random variables, we have

$$h_\omega(\theta) = \lim_{L\to\infty} z_{_L} = \sqrt{P_{\mathrm{T}}} \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} \mathrm{E}_{n_i} f(\theta + \sigma_i n_i) \tag{3.9}$$

$$= \sqrt{P_{\mathrm{T}}} \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} \int_{-\infty}^{\infty} f(\theta + \sigma_i n_i) p(n_i) dn_i \tag{3.10}$$

$$= \sqrt{P_{\mathrm{T}}} \lim_{L\to\infty} \int_{-\infty}^{\infty} \frac{1}{L} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) p(n_i) dn_i \tag{3.11}$$

$$= \sqrt{P_{\mathrm{T}}} \int_{-\infty}^{\infty} \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) p(n_i) dn_i \tag{3.12}$$

$$= \sqrt{P_{\mathrm{T}}} \int_{-\infty}^{0} \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) p(n_i) dn_i$$

$$+ \sqrt{P_{\mathrm{T}}} \int_{0}^{\infty} \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} f(\theta + \sigma_i n_i) p(n_i) dn_i \tag{3.13}$$

$$= \sqrt{P_{\mathrm{T}}} \left(-\frac{c_2}{2} + \frac{c_1}{2} \right) = \frac{c_1 - c_2}{2} \tag{3.14}$$

for some $c_1 \leq c$, $c_2 \leq c$. We have exchanged the summation and expectation in (3.10) to get (3.11). We have used assumption **(A2)** to apply bounded convergence theorem [104, pp. 288] to move the limit in (3.11) inside the integral as in (3.12). In (3.13), we have used Lemma 3.2.2 for the sequence $f(\theta + \sigma_i n_i)$ along with the fact that $f(x)$ converges to some constant as $|x| \to \infty$ by the virtue of assumptions **(A1)** and **(A2)**. Thus if $\sigma_i \to \infty$, then $z_{_L} \to (c_1 - c_2)/2$ almost surely so that $h_\omega(\theta)$ is independent of θ and therefore θ can not be recovered from $h_\omega(\theta)$ and the theorem is proved. □

Theorem 3.2.3 indicates that if sensors use a bounded function to transmit their measurements to the FC, there is a penalty incurred when the variance of the noise samples are going to infinity. When the noise samples are very high in magnitude, the sensors will be transmitting the boundary values

(c_1 or $-c_2$) most of the time. These boundary values do not contain any information on the quantity of interest θ, therefore we can not construct any useful estimator of θ from $z_{_L}$.

We like to point out that the assumption **(A2)** is not necessary for Theorems 3.2.2 and 3.2.3 to hold. It is sufficient if $f(x)$ is just an increasing function such that the variances $\mathrm{var}[f((\theta+\sigma_i n_i))]$ are bounded (boundedness of $f(x)$ is not necessary) to prove Theorems 3.2.2 and 3.2.3. For instance, the function $f(x)=\mathrm{sign}(x)|x|^p, 0<p<1/2$ is not bounded, but $\mathrm{var}[\mathrm{sign}((\theta+\sigma_i n_i))|(\theta+\sigma_i n_i)|^p]$ is bounded. Therefore, Kolmogorov's strong law of large numbers is still applicable and it is possible to estimate θ from $z_{_L}$ in (3.4).

3.2.3 Asymptotic Normality of the Estimator

We now investigate the asymptotic normality of the estimator in (3.6). For the sake of simplicity we assume that n_i are i.i.d. and $\sigma_i=1, i=1,\ldots,L$.

Theorem 3.2.4. *Let the assumption **(A1)** hold and suppose that $\sigma_i=1, i=1,\ldots,L$. Let n_i be i.i.d. and $v\sim\mathcal{N}(0,\sigma_v^2)$, then $\sqrt{L}\left(\widehat{\theta}_L-\theta\right)$ is asymptotically normal with zero mean and variance given by*

$$AsV=\frac{\int\limits_{-\infty}^{\infty} f^2(\theta+n_i)\;p(n_i)dn_i-h_\omega^2(\theta)+\frac{\sigma_v^2}{P_T}}{\left(\int\limits_{-\infty}^{\infty} f'(\theta+n_i)\;p(n_i)dn_i\right)^2}. \tag{3.15}$$

Proof. Due to the central limit theorem, we see that $\sqrt{L}\,[z_{_L}-h_\omega(\theta)]$ is asymptotically normal with zero mean and variance $\tilde{\sigma}^2$ given by

$$\tilde{\sigma}^2=P_{\mathrm{T}}\left[\int\limits_{-\infty}^{\infty} f^2(\theta+n_i)\;p(n_i)dn_i-h_\omega^2(\theta)\right]+\sigma_v^2. \tag{3.16}$$

Applying [86, Thm 3.16] the asymptotic variance of the estimator in (3.6) is given by

$$AsV=G^2\tilde{\sigma}^2 \tag{3.17}$$

where

$$G := \frac{\partial h_\omega^{-1}(\frac{z_L}{\sqrt{P_{\mathrm{T}}}})}{\partial z_{_L}}\bigg|_{z_L=\sqrt{P_{\mathrm{T}}}h_\omega(\theta)} = \frac{1}{h'_\omega\left(h_\omega^{-1}\left(\frac{z_L}{\sqrt{P_{\mathrm{T}}}}\right)\right)}\bigg|_{z_L=\sqrt{P_{\mathrm{T}}}h_\omega(\theta)}$$
$$= \frac{1}{\sqrt{P_{\mathrm{T}}}h'_\omega(\theta)} \tag{3.18}$$

Substituting G in (3.17) and simplifying we obtain the theorem. □

3.2.4 Comparison with Amplify and Forward Scheme

For the AF scheme, the transmitted signal at the i^{th} sensor is given by $\alpha_L x_i$ where α_L depends on the number of sensors L to ensure the total power constraint, but is independent of x_i [38], [51], [105]. To begin with, we focus on the case when n_i are i.i.d., and choosing α_L identical across sensors. In what follows, we will show that the scheme in (3.6) is superior to AF when the sensing noise has a heavy-tailed density.

The received signal for the AF scheme is given by

$$y_{_L} = \alpha_L \sum_{i=1}^{L}(\theta + \sigma_i n_i) + v\ . \tag{3.19}$$

We have already seen that the per-sensor power $\alpha_L^2(\theta+\sigma_i n_i)^2$ is an unbounded RV, when the PDF of the sensing noise has support over the entire real line. This is undesirable especially for low-power sensor networks with limited peak-power capabilities. We reiterate that using a bounded transmission function is preferable to AF, with respect to the management of the instantaneous transmit power of sensors.

Since the total instantaneous power is random for AF, the total power is defined as an average $P_{\mathrm{T}} = \alpha_L^2 \sum_{i=1}^{L} \mathrm{E}[(\theta + \sigma_i n_i)^2]$, where the expectation is taken with respect to the sensing noise distribution. We will consider a total power constraint case where P_{T} is not a function of L so that $\alpha_L =$

$\sqrt{\frac{P_{\mathrm{T}}}{\sum_{i=1}^{L}(\theta^2+\sigma_i^2\sigma_n^2)}}$. For the AF scheme the estimator is given by $\widehat{\theta}_{AF} = y_{_L}/(L\alpha_L)$ so that

$$(\widehat{\theta}_{AF} - \theta) = \frac{1}{L}\sum_{i=1}^{L}\sigma_i n_i + \frac{1}{L}\sqrt{\frac{\sum_{i=1}^{L}(\theta^2+\sigma_i^2\sigma_n^2)}{P_{\mathrm{T}}}}\, v\ . \tag{3.20}$$

The normalized multiple access channel output for the AF scheme is proportional to the sample mean, which is not a good estimator of θ when the sensing noise is heavy-tailed. As a specific example, consider the case when n_i is Cauchy distributed. From (3.20) it is clear that $(\widehat{\theta}_{AF} - \theta) \to 0$ is not possible since the sample mean $L^{-1}\sum_{i=0}^{L}\sigma_i n_i$ is Cauchy distributed for any value of L. Since the sample mean is not a consistent estimator for Cauchy noise, the AF approach over multiple access channels fails for such a heavy-tailed distribution. On the other hand, the estimator proposed in (3.6) is strongly consistent in the presence of any sensing noise distribution, including Cauchy distribution. This example illustrates that the inherent robustness of using the bounded transmission function in the presence of heavy-tailed sensing noise distributions. The sample mean, "computed" by the multiple access channel in the AF approach, is highly suboptimal, and sometimes not consistent like in the Cauchy case, whereas in the proposed approach the channel computes a noisy and normalized version of the function of the sensed samples, from which a consistent estimator can be constructed for any sensing noise distribution.

We saw that bounded transmissions are more robust to impulsive sensing noise compared to AF. On the other hand, AF can be superior to bounded transmissions if $\sigma_i \to \infty$. Recall Theorem 3.2.3 which says that if $\sigma_i \to \infty$, then the estimator in (3.6) is not consistent. It is clear from (3.20) that AF is strongly consistent provided that $L^{-1}\sum_{i=0}^{L}\sigma_i n_i$ converges to zero. A sufficient condition for this is given by Theorem 3.2.1 which is given by $\sum_{i=1}^{\infty}\sigma_i^2/i^2 < \infty$ in this case. It is possible for $\sigma_i \to \infty$ while $\sum_{i=1}^{\infty}\sigma_i^2/i^2 < \infty$, when the vari-

ances of n_i exist. For example, if $\sigma_i = \sqrt{i}\sigma$ for some $\sigma > 0$, then $\sigma_i \to \infty$ as $i \to \infty$. However, $\sum_{i=1}^{\infty} \sigma_i^2/i^2 = \sigma^2 \sum_{i=1}^{\infty} i^{\frac{-3}{2}} < \infty$. Therefore, in this case the strong law of large numbers holds, and the AF scheme is consistent. Whereas the proposed scheme fails to be consistent as was proved in Theorem 3.2.3 irrespective of at what rate σ_i goes to ∞. Thus, AF is consistent over a less strict set of conditions on σ_i.

3.3 Distributed Detection with Bounded Transmissions

For the distributed estimation problem, we saw that consistency requires that $f(\cdot)$ is one-to-one. For distributed detection this is not necessary, since we do not seek to estimate θ but to distinguish between two hypothesis. Indeed, conventionally, $f(\cdot)$ is chosen as a quantizer in distributed detection. In this section, we want to address the choice of $f(\cdot)$ whether it is a quantizer, or an invertible bounded function. We also want to study the consequences of boundedness for $f(\cdot)$ through the deflection coefficient.

3.3.1 System Model

Consider a binary hypothesis testing problem with two hypotheses H_0, H_1 where P_0, P_1 are their respective prior probabilities. Let the sensed signal at the i^{th} sensor be,

$$x_i = \begin{cases} \theta + \sigma_i n_i & \text{under } H_1 \\ \sigma_i n_i & \text{under } H_0 \end{cases} \tag{3.21}$$

$i = 1, \ldots, L$, $\theta > 0$ is a known parameter whose presence or absence has to be detected, L is the total number of sensors in the system, and n_i is the noise sample at the i^{th} sensor. As explained in Section 3.2.1, $\sigma_i > 0$ is a deterministic scale parameter. The sensing noise samples are i.i.d, have zero median but they need not be bounded or have any finite moments. We consider a setting where the i^{th} sensor transmits its measurement using a bounded function $\sqrt{\rho} f(x_i)$

over a Gaussian multiple access channel so that the received signal at the FC is given by (3.2) where ρ is a power scale factor and $f(x)$ satisfies the same conditions as in Section 3.2.1, and $v \sim \mathcal{N}(0, \sigma_v^2)$ is the additive channel noise. Note that the power at each sensor is upper bounded by ρc^2. We also assume that the total power $\rho c^2 L$ for the entire network is constrained to P_{T}.

3.3.2 The Detection Problem

The received signal $y_{_L}$ under the total power constraint can be written as

$$y_{_L} = \sqrt{\frac{P_{\mathrm{T}}}{L}} \sum_{i=1}^{L} f\left(x_i\right) + v \ . \tag{3.22}$$

With the received signal in (3.22), the FC has to decide which hypothesis is true. It is well known that the optimal decision rule under the Bayesian formulation is given by:

$$\frac{p(y_{_L}|H_1)}{p(y_{_L}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} \tag{3.23}$$

where $p(y_{_L}|H_i)$, is the conditional probability density function of $y_{_L}$ when the hypothesis H_i, $i \in \{0, 1\}$, is true.

3.3.3 Probability of Error

The PDFs of $y_{_L}$ in (3.23) under the hypothesis H_i involve $(L+1)$ convolutions and are not tractable in general. Let P_{e} be the probability of error at the FC:

$$P_{\mathrm{e}} = P_0 \Pr\left[error|H_0\right] + P_1 \Pr\left[error|H_1\right] \tag{3.24}$$

where $\Pr\left[error|H_i\right]$ is the error probability when H_i is true. Since P_{e} is not straightforward to evaluate, we will study a surrogate metric called the deflection coefficient (DC) to identify regimes where reliable detection is possible. The DC, depends only on the system model in (3.22), and does not depend on any detector. As we are considering a general transmission scheme at the local sensors, and P_{e} is not tractable, it is more insightful to study the DC.

3.3.4 Deflection Coefficient and its Optimization

We will now define and use the deflection coefficient which reflects the output-signal-to-noise-ratio and widely used in optimizing detectors [79–82]. The DC is defined as,

$$D := \frac{1}{L}\frac{|\mathrm{E}[y_{_L}|H_1] - \mathrm{E}[y_{_L}|H_0]|^2}{\mathrm{var}[y_{_L}|H_0]}. \tag{3.25}$$

When σ_i is a deterministic sequence, the DC for the system in (3.22) is given by

$$D_L = \frac{\left(L^{-1}\sum_{i=1}^{L}\int_{-\infty}^{\infty}[f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i\right)^2}{L^{-1}\sum_{i=1}^{L}\left[\int_{-\infty}^{\infty} f^2(\sigma_i n_i)p(n_i)dn_i - \left(\int_{-\infty}^{\infty} f(\sigma_i n_i)p(n_i)dn_i\right)^2\right] + \frac{\sigma_v^2}{P_{\mathrm{T}}}}. \tag{3.26}$$

We now study the conditions on the sequence σ_i for $\lim_{L\to\infty} D_{_L} = 0$. When this asymptotic DC is zero, the interpretation is that reliable detection is not possible. The following result establishes that if σ_i goes to infinity, the asymptotic DC is zero.

Theorem 3.3.1. *Let σ_i be a deterministic sequence such that $\lim_{i\to\infty}\sigma_i = \infty$, suppose that the assumptions **(A1)** and **(A2)** hold. Then, $\lim_{L\to\infty} D_{_L} = 0$.*

Proof. Clearly the denominator of (3.26) is bounded between $(\sigma_v^2/P_{\mathrm{T}})$ and $(c^2 + \sigma_v^2/P_{\mathrm{T}})$. Therefore, it suffices to show that the numerator goes to 0 as

$L \to \infty$. Consider

$$\lim_{L\to\infty} D_{_L} = \lim_{L\to\infty} \frac{1}{L}\sum_{i=1}^{L} \int_{-\infty}^{\infty} [f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i \tag{3.27}$$

$$= \int_{-\infty}^{\infty} \lim_{L\to\infty} \frac{1}{L}\sum_{i=1}^{L} [f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i \tag{3.28}$$

$$= \int_{-\infty}^{0} \lim_{L\to\infty} \frac{1}{L}\sum_{i=1}^{L} [f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i$$

$$+ \int_{0}^{\infty} \lim_{L\to\infty} \frac{1}{L}\sum_{i=1}^{L} [f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i \tag{3.29}$$

$$= \left(-\frac{c_2}{2} + \frac{c_2}{2}\right) + \left(\frac{c_1}{2} - \frac{c_1}{2}\right) = 0 \tag{3.30}$$

for some $c_1 \leq c$, $c_2 \leq c$ and we have used assumption **(A2)** to apply bounded convergence theorem [104, pp. 288] to move the limit in (3.27) inside the integral as in (3.28). In (3.29), we have used Lemma 3.2.2 for the sequences $f(\theta+\sigma_i n_i)$ and $f(\sigma_i n_i)$ along with the fact that $f(x)$ converges to some constant as $|x| \to \infty$ by the virtue of assumptions **(A1)** and **(A2)**. Thus if $\sigma_i \to \infty$, then $\lim_{L\to\infty} D_{_L} = 0$. □

Theorem 3.3.1 indicates that if sensors use a bounded function to transmit their measurements to the FC, there is a penalty incurred when the variance of the noise samples are very high. When the noise samples are very high in magnitude, the sensors will be transmitting the boundary values of $f(x)$, i.e., c_1 or $-c_2$ most of the time. These boundary values do not contain any information about the signal θ to be detected when H_1 is true. Hence it is not possible to distinguish between the hypothesis H_1 and H_0 and accordingly we have the asymptotic DC equal to 0.

However, if σ_i are bounded, then we can show that $\lim_{L\to\infty} D_{_L} > 0$ which is done next.

Theorem 3.3.2. *Let $\sigma_{max} := \max_i \sigma_i$ be finite and suppose that the assumptions **(A1)** and **(A2)** hold. Then,* $\lim_{L\to\infty} D_{_L} > 0$.

Proof. Let $g_{\sigma_i}(\theta) := \int_{-\infty}^{\infty} [f(\theta+\sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i$. To show $\lim_{L\to\infty} D_{_L} > 0$, it suffices to show that $g_{\sigma_i}(\theta) > 0$, $\forall\theta > 0$ for some i. Using the assumption **(A1)** we have,

$$g_{\sigma_i}(\theta) = \int_{-\infty}^{\infty} [f(\theta + \sigma_i n_i) - f(\sigma_i n_i)]p(n_i)dn_i \ ,$$

$$\frac{\partial g_{\sigma_i}(\theta)}{\partial\theta} = \int_{-\infty}^{\infty} \frac{\partial f(\theta + \sigma_i n_i)}{\partial\theta} p(n_i)dn_i \ , \tag{3.31}$$

$$> 0 \ , \tag{3.32}$$

where we have applied Corollary 5.9 in [103, pp. 46] using assumptions **(A1)** and **(A2)** to move the derivative in (3.31) inside the integral. The last inequality follows from the fact that convex combination of positive valued functions is positive. Therefore, $g_{\sigma_i}(\theta)$ is strictly an increasing function of θ. When $\theta = 0$, clearly $g_{\sigma_i}(0) = 0$ and together with the fact that $\partial g_{\sigma_i}(\theta)/\partial\theta > 0$, $\forall\theta > 0$, we have $g_{\sigma_i}(\theta) > 0$, $\forall\theta > 0$. □

Theorem 3.3.2 says that if the deterministic σ_i are bounded, then the asymptotic DC is strictly positive which means that reliable detection is possible in this regime.

Next we will prove that for the DC to be greater than zero, we do not need $f(x)$ to be a differentiable or strictly increasing. In the following theorem we prove that $D_{_L} > 0$ for an uniform quantizer with bounded number of quantization levels.

Theorem 3.3.3. *Let $\sigma_{max} := \max_i \sigma_i$ be finite and suppose that $f(x)$ is a uniform quantizer with M levels such that*

$$f(x) = \begin{cases} k\Delta\ , & (k-\frac{1}{2})\Delta \leq x < (k+\frac{1}{2})\Delta\ , \\ K\Delta\ , & x \geq (K+\frac{1}{2})\Delta\ , \\ -K\Delta\ , & x \leq -(K+\frac{1}{2})\Delta \end{cases} \tag{3.33}$$

where $k = -K, -(K-1), \ldots, 0, \ldots, (K-1), K$, $M = 2K+1$, $\Delta = 2x_{max}/M$ and x_{max} is the saturation point of the finite level quantizer. Suppose that n_i has infinite support. Then, $D_{_L} > 0$.

Proof. Let $g_{\sigma_i}(\theta) := \int_{-\infty}^{\infty} [f(\theta + \sigma_i n_i) - f(\sigma_i n_i)] p(n_i) dn_i$. To show $D_{_L} > 0$, it suffices to show that $g_{\sigma_i}(\theta) > 0$, $\forall \theta > 0$. Note that the function $f(x)$ in (3.33)

is non-decreasing, i.e., $f(x) - f(y) \geq 0, \forall x \geq y$. Consider

$$g_{\sigma_i}(\theta) = \int_{-\infty}^{\infty} [f(\theta + \sigma_i n_i) - f(\sigma_i n_i)] p(n_i) dn_i \tag{3.34}$$

$$= \frac{1}{\sigma_i} \int_{-\infty}^{\infty} [f(\theta + v_i) - f(v_i)] p(v_i) dv_i \tag{3.35}$$

$$= \frac{1}{\sigma_i} \int_{-\infty}^{-[(K+\frac{1}{2})\Delta+\theta]} [f(\theta + v_i) - f(v_i)] p(v_i) dv_i$$

$$+ \frac{1}{\sigma_i} \int_{-[(K+\frac{1}{2})\Delta+\theta]}^{(K+\frac{1}{2})\Delta} [f(\theta + v_i) - f(v_i)] p(v_i) dv_i$$

$$+ \frac{1}{\sigma_i} \int_{(K+\frac{1}{2})\Delta}^{\infty} [f(\theta + v_i) - f(v_i)] p(v_i) dv_i \tag{3.36}$$

$$\geq \frac{1}{\sigma_i} \int_{-[(K+\frac{1}{2})\Delta+\theta]}^{(K+\frac{1}{2})\Delta} [f(\theta + v_i) - f(v_i)] p(v_i) dv_i \tag{3.37}$$

$$= \frac{1}{\sigma_i} \sum_{k=-K}^{K} \int_{[(k-\frac{1}{2})\Delta-\theta]}^{(k+\frac{1}{2})\Delta} \Delta p(v_i) dv_i \tag{3.38}$$

$$> 0 \tag{3.39}$$

where in (3.34) we substituted $v_i = \sigma_i n_i$ to get (3.35). The inequality in (3.39) follows from the fact that $\Delta > 0$ and v_i has infinite support (since n_i has infinite support so that $v_i = \sigma_i n_i$ has infinite support as well). When $\theta = 0$, clearly $g_{\sigma_i}(0) = 0$ and therefore, we have $D_{_L} > 0$, $\forall \theta > 0$. □

Theorem 3.3.3 can in fact be proved for non-uniform quantizer as long as $M \geq 2$ and n_i has infinite support.

We would ideally like to find the $f(x)$ that maximizes the DC in (3.26) but this is not tractable. However, when θ is small, and channel noise is neg-

ligible, we have a closed form expression for $f(x)$ through the locally optimal detection strategy. We now briefly discuss the use of nonlinear functions in the context of locally optimal detection.

3.3.5 Locally Optimal Detection

A detector is said to be locally optimal (most powerful) if it is better than any other detector in the sense of minimizing the probability of error for very small values of θ (please see [101] for more details). The problem of designing optimum detectors in the presence of additive noise has a long history in the statistical signal processing literature [101]. Usually the sensing noise corrupting the signal is assumed to be Gaussian. However there are situations when the noise is impulsive, which are characterized by symmetric alpha stable distributions [101]. In such scenarios, linear detector is not necessarily optimal, and therefore nonlinear functions are applied on the sensed observations to minimize the impact of impulsive sensing noise distributions with heavy tails.

In [101], it is shown that for a given sensing noise distribution $p(n)$, the nonlinear function $f(x)$ that would be locally optimal is given by

$$f(x) = -\frac{p^{'}(x)}{p(x)}. \tag{3.40}$$

One may be interested in the inverse problem that given a nonlinear function $f(x)$, for which sensing noise distribution, it would be locally optimal. From (3.40) it is easy to answer this question. We have,

$$p(x) = Ce^{-\int_{-\infty}^{x} f(y)dy}. \tag{3.41}$$

Here the $p(x)$ obtained from (3.41) should be a valid PDF satisfying $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x)dx = 1$. For example, if $f(x) = \tanh(x)$, we get $p(x) = \pi\text{sech}(x) = 2\pi e^{-x}/(1+e^{-2x})$. The $\text{sech}(x)$ distribution behaves like the heavy-tailed Laplacian distribution when x is relatively high. It is interesting to note that $\tanh(x)$

behaves like the hard clipper non-linearity [101] which is a bounded function and is locally optimal for Laplacian noise distribution. In fact, a closer look at (3.40) reveals that if $p(x)$ behaves like an exponential density (for relatively large x), then the $f(x)$ that would be locally optimal would behave like a constant (for relatively large x). This shows that the family of increasing bounded functions are locally optimal for the family of heavy tailed sensing noise distributions. When n is Gaussian, bounded $f(x)$ is no longer optimal as it is well known that $f(x) = x$ is optimal for Gaussian sensing noise. We will illustrate this in the Simulations section.

3.4 Simulations

In this section, we corroborate our analytical results through Monte Carlo simulations for both the distributed estimation and distributed detection problems. In all of the simulations we have assumed $\sigma_i = 1, i = 1, \ldots, L$.

3.4.1 Distributed Estimation Performance

In Figure 3.2 we chose $f(x) = \tanh(\omega x)$, $\omega > 0$ is a scale parameter. Here we compare $AsV(\omega)$ and $L\text{var}(\hat{\theta}_{_L} - \theta)$ versus ω under the total power constraint for various distributions on the sensing noise n_i. We observe that the variance of the asymptotic distribution, $AsV(\omega)$ and the normalized limiting variance $L\text{var}(\hat{\theta}_{_L} - \theta)$ are closer to each other when L is sufficiently large. However if L is smaller, we see that there is significant difference between $AsV(\omega)$ and $L\text{var}(\hat{\theta}_{_L} - \theta)$ as illustrated in Figure 3.3. This is due to the finite sample effect, and when L is increased, $L\text{var}(\hat{\theta}_{_L} - \theta)$ decreases to converge its limiting value of $AsV(\omega)$. In Figure 3.4, we compare $AsV(\omega)$ and $L\text{var}(\hat{\theta}_{_L} - \theta)$ versus L. Clearly in all cases, as L increases the $L\text{var}(\hat{\theta}_{_L} - \theta)$ approaches its limiting values of $AsV(\omega)$.

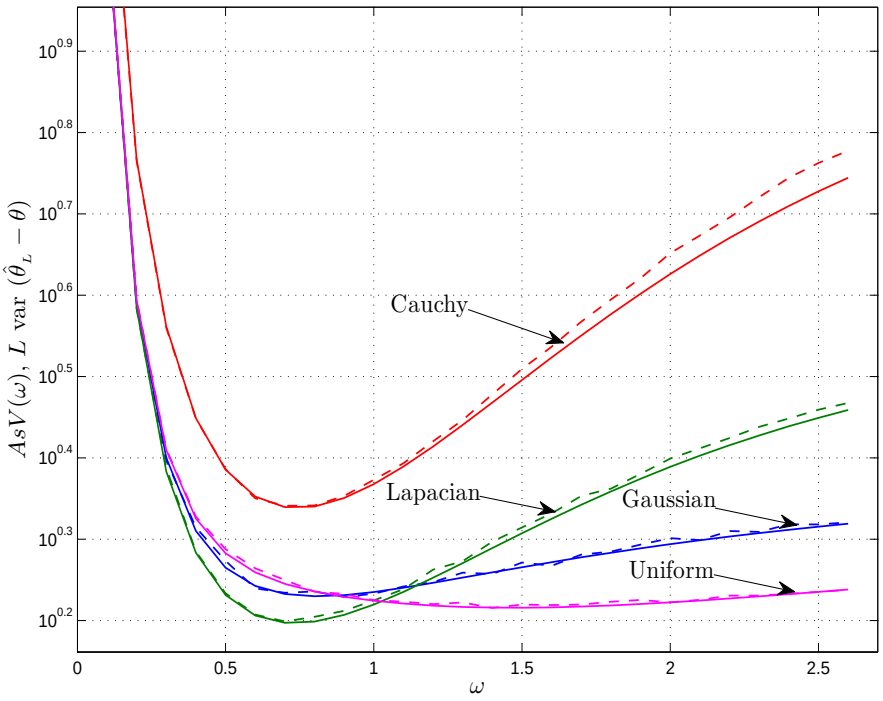

Figure 3.2: Total Power Constraint: $f(x) = \tanh(\omega x)$, σ_n^2=1, σ_v^2=1, P_{T}=10, L=500

In Figure 3.5, we compare the performance among different bounded transmission functions when n_i is Gaussian. All the functions used in this plot are appropriately normalized so that $-1 \leq f(x) \leq 1$. Here $\mathrm{gd}(x)$ is the Gudermannian function given by $\mathrm{gd}(x) = \arctan(\sinh(\omega x))$. We note that $\tanh(\omega x)$ has the lowest asymptotic variance compared to other functions. Intuitively, this is due to the fact that for a given ω, $\tanh(\omega x)$ is closest to the linear function among the other functions considered here. For the Gaussian sensing noise, since linear estimator is optimal, $\tanh(\omega x)$ performs better than other functions.

3.4.2 Distributed Detection Performance

We define the sensing and channel SNRs as $\rho_s := \theta^2/\sigma_n^2$, $\rho_c := P_{\mathrm{T}}/\sigma_v^2$ and assume $P_1 = P_0 = 0.5$. Note also that $\rho = P_{\mathrm{T}}/L$ is the power at each

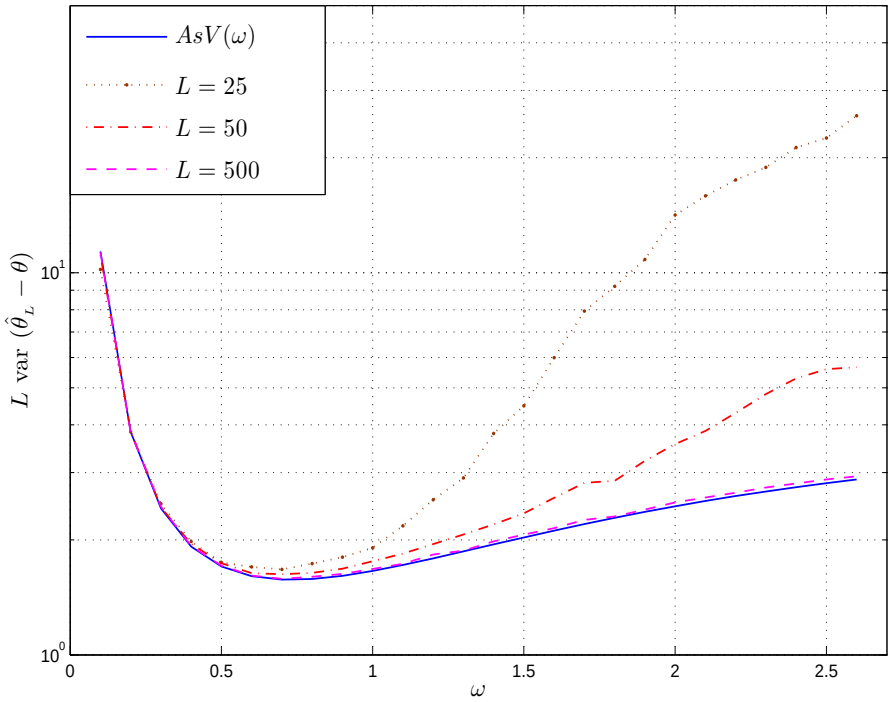

Figure 3.3: Total Power Constraint, n_i Laplacian: $f(x) = \tanh(\omega x)$, σ_n^2=1, σ_v^2=1, P_{T}=10, L=25, 50, 500

sensor as defined in Section 3.2.1. We used a quadratic detector based on the assumption that $y_{_L}$ in (3.22) is Gaussian under both hypotheses in the simulations provided here.

In Figure 3.6, we chose $f(x) = \tanh(\omega x)$, $\omega > 0$ is a scale parameter and show that maximizing the DC approximately results in minimizing the probability of error. Figure 3.6 shows the plots of $D(\omega)$ and $P_{\mathrm{e}}(\omega)$ vs ω for Gaussian, Laplacian and Cauchy sensing noise distributions where the $P_{\mathrm{e}}(\omega)$ plot is obtained using Monte-Carlo simulations. The different ω^* values in Figure 3.6 correspond to the best ω values obtained by optimizing $D(\omega)$ and $P_{\mathrm{e}}(\omega)$ respectively. It is interesting to see that the ω^* that minimizes $P_{\mathrm{e}}(\omega)$ is very close to that which maximizes $D(\omega)$ and thus DC is justified as a performance metric.

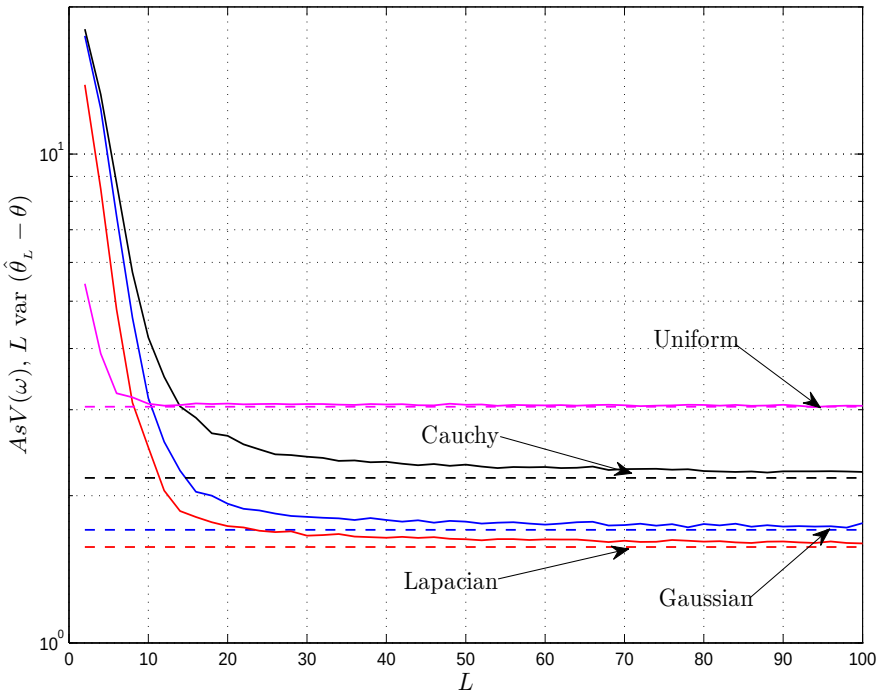

Figure 3.4: Total Power Constraint: $f(x) = \tanh(\omega x)$, σ_n^2=1, σ_v^2=1, ρ=1, ω=0.75

Finally we depict the P_{e} performance versus L for different bounded functions in Figure 3.7. In each of these cases, ω^* that maximized the deflection coefficient were used. We note that AF outperforms all other functions since for the AF scheme, the detector is a linear function of observations which is optimal when n_i is Gaussian. The function $\omega x/(1 + |\omega x|)$ exhibits the worst performance as it has the largest deviation from the linear function compared to the other candidate functions considered in this simulation.

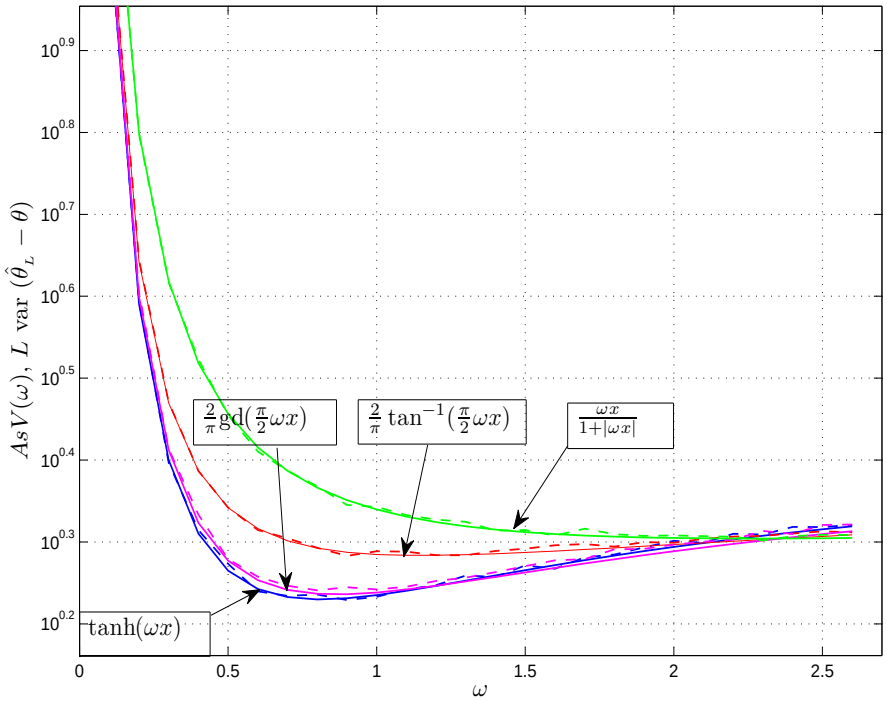

Figure 3.5: Total Power Constraint, Different bounded functions: σ_n^2=1, σ_v^2=1, P_{T}=10, L=500

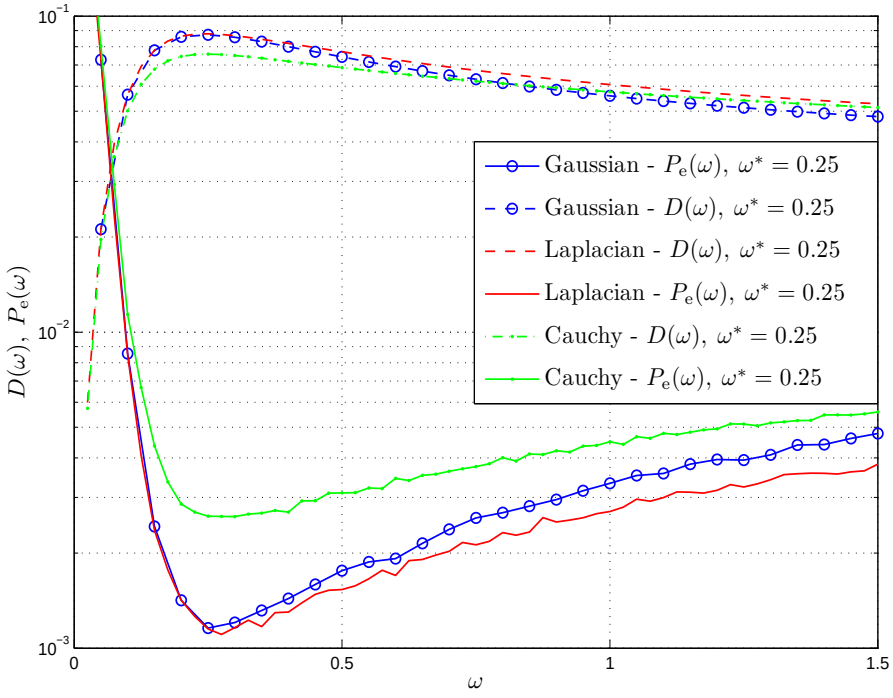

Figure 3.6: Total Power Constraint, $f(x) = \tanh(\omega x)$, $D(\omega)$ $P_{\mathrm{e}}(\omega)$ versus ω, $\rho_s = 10$ dB, $\rho_c = 3$ dB, L=20

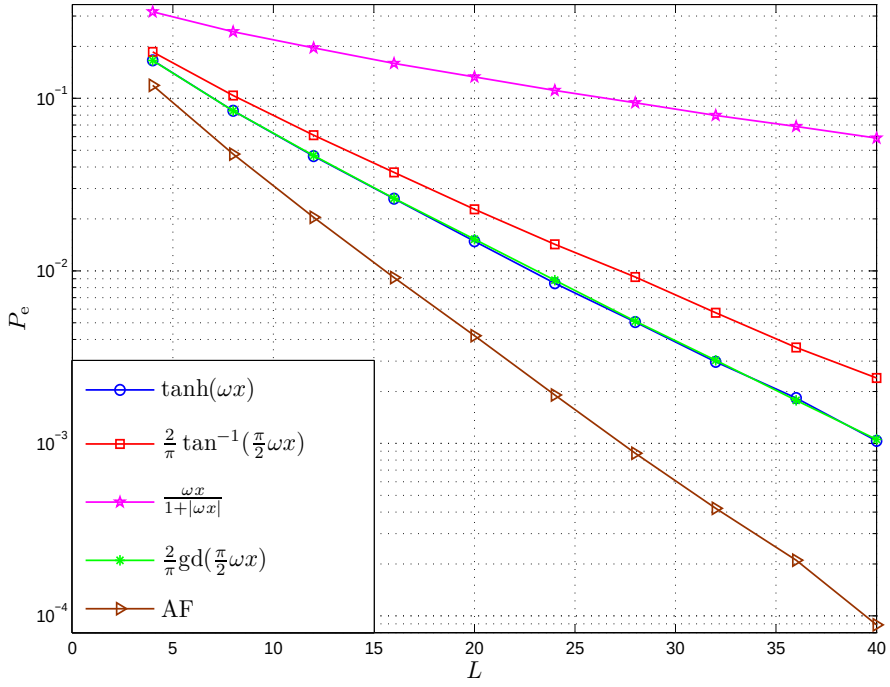

Figure 3.7: Total Power Constraint, n_i Gaussian, P_e versus L, $\rho_s = 10$ dB, $\rho_c = 0$ dB

Chapter 4

Distributed Consensus with Bounded Transmissions

4.1 Literature Survey and Motivation

In this chapter, we consider the problem of estimating the average of samples measured at the nodes of a sensor network without a fusion center (fully distributed network). Distributed computing has been a subject of extensive research in the last two decades with a wide range of applications (see for example [1, 2]). Originally, the purpose of distributed algorithms was to reduce the burden on a single processor and perform computationally large tasks in a distributed manner on multiple processors which were physically wired. With the recent interest in wireless sensor networks (WSNs), the focus has got shifted to performing global computations with the data available locally at each node of a WSN in a distributed manner. For example, the sensor nodes in a WSN could measure information about a physical phenomenon of interest. These measurements can be used for estimating unknown parameters of the physical phenomenon (distributed estimation) or they can be used to make a decision about the state of the physical phenomenon (distributed detection).

Wireless sensor networks (WSNs) without a fusion center have the advantages of robustness to node failures and they can function autonomously without a central node controlling the entire network [2]. In such fully distributed networks, sensors collaborate with their neighbours by repeatedly exchanging information locally to achieve a desired global objective. For example, the sensors could come to an agreement on the sample average (or on a global function) of initial measurements. This is called distributed consensus. Distributed consensus algorithms have attracted significant interest in the recent past and have found several applications in areas such as healthcare, envi-

ronmental monitoring, military and home appliances (please see [58–61,63–65] and references therein). In this body of literature, it is often assumed that a given node can obtain exact information of the state values of its neighbours through local communications. This essentially means that the system consumes theoretically unlimited energy and bandwidth. However, practical WSNs are severely power limited and the available bandwidth is finite. Moreover, the main source of power consumption in a sensor is its transceiver [106]. Therefore, there is a need for consensus algorithms which work under strict resource constraints of power and bandwidth imposed by the WSNs.

Sensors may adopt either a digital or analog method for transmitting their information to their neighbours. Digital methods of transmissions may be using low transmit power but require increased bandwidth especially when the number of quantization levels is high. Distributed consensus algorithms using quantized transmissions have been studied in [107–111]. The analog method consists of transmitting unquantized data by appropriately pulse shaping and amplitude or phase modulating to consume finite bandwidth. One such method is the amplify-and-forward (AF) scheme in which sensors send scaled versions of their measurements to their neighbours. However, using the AF technique is not a viable option for WSNs because it requires high transmission power when the values to be transmitted are large [92]. Moreover, the linear transmit amplifier characteristics required for AF are often very power-inefficient [93], requiring the study of the effect of non-linear transmissions on performance. In distributed systems which employ the AF technique for transmission of the sensed data, it is often assumed that the power amplifiers used are perfectly linear over the entire range of the sensed observations. In

practice, the amplifiers exhibit non-linear behaviour when the amplitude of the sensed data is relatively high [93–95].

In this chapter, we propose a non-linear distributed consensus (NLC) algorithm in which every sensor maps its state value through a bounded function before transmission to constrain the peak transmit power. Therefore the magnitude of the transmitted signal at every node in every iteration is always bounded, making it ideal for resource-constrained WSNs. In the presence of communication noise, we prove that all the sensors employing the NLC algorithm reach consensus to a finite random variable whose mean is the desired sample average. We characterize the asymptotic performance by deriving the asymptotic covariance matrix using results from stochastic approximation theory. We show that using the NLC algorithm results in larger asymptotic covariance compared to the linear consensus algorithm. Finally we explore the performance of the proposed algorithm employing various bounded transmission functions. Different from [65] which also considered consensus in the presence of noisy transmissions, herein we analyse non-linear transmissions and study the asymptotic covariance matrix and its dependence on the non-linearity. Our work in this chapter also studies the merits and demerits of distributed schemes involving realistic amplifier models with non-linear characteristics such as the ones discussed in [93, 94].

The rest of this chapter is organized as follows. We begin by reviewing some basics of network graph theory in Section 4.2. In Section 4.3, we describe the system model and review the previous work on non-linear consensus. We consider the NLC algorithm in the presence of noise in Section 4.4, and prove that the sensors reach consensus to a random variable. In Section 4.5, we

present several simulation examples to study the performance of the proposed algorithm.

Notations and Conventions

Vectors are denoted by boldface upper-case or lower-case letters and matrices are denoted by boldface upper-case letters. $\max\{a_1, a_2\}$ denotes the maximum of a_1 and a_2. $\text{diag}[a_1,\ a_2,\ \ldots,\ a_N]$ denotes an $N \times N$ diagonal matrix whose diagonal elements are given by $a_1, a_2, \ldots, a_N$. $\text{E}[\cdot]$ denotes the expectation operator and $\mathbf{I}$ denotes the identity matrix. The symbol $\|\cdot\|$ denotes the l_2 norm for real vectors and spectral norm for symmetric matrices. For a matrix $\mathbf{M}$, $\lambda_i(\mathbf{M})$ denotes the i^{th} smallest eigenvalue. The vector $\mathbf{1}$ denotes an $N \times 1$ column vector of all ones, $\mathbf{1} = [1\ 1 \ldots 1]^{\text{T}}$.

4.2 Review of Network Graph Theory

In this chapter, we model a sensor network as an undirected graph. In this section, we provide a brief background on network graph theory which we will use to derive our results. Consider an undirected graph $\mathbb{G} = (\mathbb{N}, \mathbb{E})$ containing a set of nodes $\mathbb{N} = \{1, \ldots, N\}$ and a set of edges $\mathbb{E}$. Nodes that communicate with each other have an edge between them. We denote the set of neighbours of node i by $\mathbb{N}_i$, $\mathbb{N}_i = \{\ j | \{i, j\} \in \mathbb{E}\}$ where $\{i, j\}$ indicates an edge between the nodes i and j [112]. A graph is connected if there exists at least one path between every pair of nodes. We denote the number of neighbours of a node i by d_i and $d_{max} = \max_i d_i$. The graph structure is described by an $N \times N$ symmetric matrix called the adjacency matrix $\mathbf{A} = \{a_{ij}\}$, $a_{ij} = 1$ if $\{i, j\} \in \mathbb{E}$. The diagonal matrix $\mathbf{D} = \text{diag}[d_1,\ d_2,\ \ldots,\ d_N]$ captures the degrees of all the nodes in the network. The Laplacian matrix of the graph is given by $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

The graph Laplacian characterises a number of useful properties of the graph. The eigenvalues of $\mathbf{L}$ are non-negative and the number of zero eigenvalues denotes the number of distinct components of the graph. When the graph is connected, $\lambda_1(\mathbf{L}) = 0$, and $\lambda_i(\mathbf{L}) > 0, i \geq 2$, so that the rank of $\mathbf{L}$ for a connected graph is $N-1$. The vector $\mathbf{1}$ is the eigenvector of $\mathbf{L}$ associated with the eigenvalue 0, i.e, $\mathbf{L1} = \mathbf{0}$. The eigenvalue $\lambda_2(\mathbf{L})$ characterizes how densely the graph is connected and the performance of consensus algorithms depend on this eigenvalue [62].

4.3 System Model and Previous Work

4.3.1 System Model

Consider a WSN with N sensor nodes each with an initial measurement $x_i(0) \in \mathbb{R}$. Measurements made at the sensor nodes are modeled as

$$x_i(0) = \theta + n_i \ , \quad i = 1, \ldots, N \tag{4.1}$$

where θ is an unknown real-valued parameter and n_i is the sensing noise at the i^{th} sensor. The sample mean of these initial measurements in (4.1) is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i(0) \ . \tag{4.2}$$

Let $\bar{x}$ be the estimate of the parameter θ to be computed by an iterative distributed algorithm, in which each sensor communicates only with its neighbours. If the states of all the sensor nodes converge to $\bar{x}$, then the network is said to have reached *consensus* on the sample average.

4.3.2 Previous Work

A commonly used iterative algorithm for distributed consensus can be written as

$$x_i(t+1) = x_i(t) - \alpha \sum_{j \in \mathbb{N}_i} h(x_i(t) - x_{ij}(t)) \,, \tag{4.3}$$

where $i = 1, \ldots, N$, $t = 0, 1, 2, \ldots$, is the time index, $x_i(t+1)$ is the updated state value of sensor node i at time $t+1$, $\mathbb{N}_i$ is the set of neighbours of sensor node i, $x_{ij}(t), j \in \mathbb{N}_i$ are the state values of the neighbours of sensor node i at time t, and α is a constant step size. If $h(\cdot)$ is linear, then (4.3) is a linear distributed average-consensus (LDAC) algorithm [58, 61, 62]. In [58], it is proved that if $0 < \alpha < 2/\lambda_N(\mathbf{L})$, then $x_i(t)$ converges to $\bar{x}$ exponentially and (4.3) is then called as the LDAC algorithm based on the Laplacian heuristic. If $h(\cdot)$ is non-linear then the algorithm belongs to the class of non-linear distributed average-consensus algorithms [60, 113–117]. In [60], the average consensus problem is solved when $h(x)$ in (4.3) is differentiable and odd. In [113], it is illustrated that when $h(x)$ in (4.3) is $\sin(x)$, faster convergence is possible compared to the LDAC algorithm based on the Laplacian heuristic. In all of these cases, $x_{ij}(t)$ has to be transmitted to node i before it can apply the function $h(\cdot)$ to get the new updated state value. Therefore, the transmit peak power in (4.3) is determined by $x_i(t)$ and not necessarily bounded, even if $h(\cdot)$ is bounded. Moreover, there is no communication noise assumed in all the previous work on non-linear consensus.

4.4 Consensus with Bounded Transmissions and Communication Noise

In this work, we propose a distributed non-linear average consensus algorithm in which every sensor maps its state value through a bounded function before transmission to constrain the transmit power. Therefore the magnitude of the transmitted signal at every node in every iteration is always bounded making it ideal for resource-constrained WSNs.

In this section, we will study the NLC algorithm with communication noise when sensors exchange information. Our approach is similar to, but more general than [65] in that we analyse non-linear transmissions. Moreover, unlike [65] we study the asymptotic covariance matrix of the state vector and its dependence on the non-linearity. Unlike [113] and [60], we assume transmit non-linearity which allows for bounded transmissions. Moreover, we consider the presence of communication noise.

4.4.1 The NLC Algorithm with Communication Noise

Let each sensor map its state value at time t through the function $h(x)$ before transmission, and consider the following NLC algorithm with communication noise:

$$x_i(t+1) = x_i(t) - \alpha(t) \sum_{j \in \mathbb{N}_i} \left[h(x_i(t)) - h(x_{ij}(t)) + n_{ij}(t) \right] , \qquad (4.4)$$

where $i = 1, \ldots, N, t = 0, 1, 2, \ldots$, is the time index. The value $x_i(t+1)$ is the state update of node i at time $t+1$, $x_{ij}(t)$ is the state value of the j^{th} neighbour of node i at time t and $\alpha(t)$ is a positive step size which will further be assumed to satisfy assumption **(A4)** in the sequel. The node j transmits its information $x_{ij}(t)$ by mapping it through the function $h(x)$, node i receives a noisy version of $h(x_{ij}(t))$ and $n_{ij}(t)$ is the noise associated with the reception of $h(x_{ij}(t))$.

Note that the proposed scheme (4.4) is different from (4.3) in the following aspects. Firstly, in (4.3), $x_{ij}(t)$ has to be transmitted which could exhibit variation over a wide range of values if $x_i(0)$ has a large dynamic range and hence (4.3) does not guarantee bounded transmission power. In contrast, in the proposed scheme the non-linearity is applied before the state value is transmitted so that the magnitude of the transmitted state value is always

constrained within the maximum value of $h(x)$ irrespective of the range of $x_i(t)$ and the realizations of noise $n_{ij}(t)$. Finally, (4.4) involves communication noise while (4.3) does not. Thus the proposed scheme is more suited to resource constrained WSNs when compared to (4.3).

The recursion in (4.4) can be written in vector form as

$$\mathbf{X}(t+1) = \mathbf{X}(t) - \alpha(t)\left[\mathbf{L}\mathbf{h}(\mathbf{X}(t)) + \mathbf{n}(t)\right] \ , \tag{4.5}$$

where $\mathbf{X}(t)$ is state vector at time t given by $\mathbf{X}(t) = [x_1(t)\ x_2(t) \dots\ x_N(t)]^{\mathrm{T}}$, and $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^N$ such that $\mathbf{h}(\mathbf{X}(t)) = [h(x_1(t))\ h(x_2(t))\ \dots\ h(x_N(t))]^{\mathrm{T}}$. The vector $\mathbf{n}(t)$ captures the additive noise at N nodes contributed by their respective neighbours and its i^{th} component is given by

$$\mathbf{n}_i(t) = -\sum_{j \in \mathbb{N}_i} n_{ij}(t) \ , 1 \leq i \leq N \ . \tag{4.6}$$

Our model in (4.5) is more general than the linear consensus algorithm considered in [65] which is a special case of $\mathbf{h}(\mathbf{x})$ when it is linear. We make the following assumptions on $h(x)$, $n_{ij}(t)$, $\alpha(t)$ and the graph:

Assumptions

(A1): The graph $\mathbb{G}$ is connected so that $\lambda_2(\mathbf{L}) > 0$.

(A2): The function $h(\cdot)$ is differentiable, and has a bounded derivative such that $0 < h^{'}(x) \leq c$, for some $c > 0$.

(A3) Independent Noise Sequence: The channel noise $\{n_{ij}(t)\}_{t \geq 0, 1 \leq i,j \leq N}$ is an independent sequence across time and space. It also satisfies

$$\mathrm{E}[n_{ij}(t)] = 0, 1 \leq i, j \leq N, t \geq 0, \ \sup_{i,j,t} \mathrm{E}[n_{ij}^2(t)] \leq \sigma^2 < \infty. \tag{4.7}$$

From (4.6) we have

$$\mathrm{E}[\mathbf{n}(t)] = \mathbf{0} \ , \forall t \ , \ \ \mu := \sup_t \mathrm{E}[\|\mathbf{n}(t)\|^2] \leq N d_{\max} \sigma^2 < \infty. \tag{4.8}$$

Note that (4.8) is because of the fact that the number of neighbours of a given node is upper bounded by $d_{\max}$.

(A4) Decreasing Weight Sequence: The channel noise in (4.5) could make the algorithm diverge. In order to control the variance growth rate of the noise we need the following conditions on the sequence $\alpha(t)$:

$$\alpha(t) > 0 \ , \ \sum_{t=0}^{\infty} \alpha(t) = \infty \ , \ \sum_{t=0}^{\infty} \alpha^2(t) < \infty \ . \tag{4.9}$$

Our primary motivation for considering non-linear transmissions is to impose the realistic assumption of bounded peak per-sensor power by ensuring that $h(\cdot)$ is bounded. However, as seen in **(A2)** this assumption is not needed for our subsequent development as long as $h^{'}(\cdot)$ is bounded.

We will prove convergence and asymptotic normality result of the NLC algorithm in (4.5). For the sake of clarity, we now present a result on the convergence of a discrete time Markov process which will be used in establishing convergence of the NLC algorithm in (4.5).

4.4.2 A Result on the Convergence of Discrete time Markov Processes

Let $\mathcal{X} = \{\mathbf{X}(t)\}_{t\geq 0}$ be a discrete time vector Markov process on $\mathbb{R}^N$. The generating operator $\mathcal{L}$ of $\mathcal{X}$ is defined as

$$\mathcal{L}V(\mathbf{x}) = \mathrm{E}\left[V(\mathbf{X}(t+1))|\mathbf{X}(t) = \mathbf{x}\right] - V(\mathbf{x}) \tag{4.10}$$

for functions $V(\mathbf{x}), \mathbf{x} \in \mathbb{R}^N$, provided that the conditional expectation exists. Let $\mathbb{B} \subset \mathbb{R}^N$ and its complement be $\mathbb{B}^{'} = \mathbb{R}^N \setminus \mathbb{B}$. We now state the desired result as a simplification of Theorem 2.7.1 in [118] (see also Theorem 1 in [65]).

Theorem 4.4.1. *Let $\mathcal{X}$ be a discrete time vector Markov process with the generator operator $\mathcal{L}$ as in* (4.10)*. If there exists a potential function $V(\mathbf{x})$:*

$\mathbb{R}^N \to \mathbb{R}^+$, *and* $\mathbb{B} \subset \mathbb{R}^N$ *with the following properties*

$$V(\mathbf{x}) > 0, \mathbf{x} \in \mathbb{B}^{'}, \;\; V(\mathbf{x}) = 0, \; \mathbf{x} \in \mathbb{B} \; , \tag{4.11}$$

$$\mathcal{L}V(\mathbf{x}) \leq -\gamma(t)\varphi(\mathbf{x}) + mg(t)[1 + V(\mathbf{x})] \tag{4.12}$$

where $m > 0$, $\varphi(\mathbf{x})$ *is such that*

$$\varphi(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B}, \; \varphi(\mathbf{x}) > 0, \mathbf{x} \in \mathbb{B}^{'} \; , \tag{4.13}$$

and

$$\gamma(t) > 0, g(t) > 0, \; \sum_{t=0}^{\infty} \gamma(t) = \infty, \; \sum_{t=0}^{\infty} g(t) < \infty \; , \tag{4.14}$$

then, the discrete time vector Markov process $\mathcal{X} = \{\mathbf{X}(t)\}_{t \geq 0}$ *with arbitrary initial distribution converges almost surely (a.s.) to the set* $\mathbb{B}$ *as* $t \to \infty$. *That is,*

$$Pr\left[\lim_{t \to \infty} \inf_{\mathbf{Y} \in \mathbb{B}} \; \|\mathbf{X}(t) - \mathbf{Y}\| = 0\right] = 1. \tag{4.15}$$

Intuitively, Theorem 4.4.1 indicates that if the one-step prediction error of the Markov process evaluated at the potential function in (4.10) is bounded as in (4.12) then it is possible to establish convergence of $\mathbf{X}(t)$.

To prove the a.s. convergence of the consensus algorithm in (4.5) using Theorem 4.4.1, we define the consensus subspace $\mathbb{B}$, the set of all vectors whose entries are of equal value as,

$$\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^N | \mathbf{x} = a\mathbf{1} \; , a \in \mathbb{R}\} \; . \tag{4.16}$$

We are now ready to state the main result of Section 4.4.

Theorem 4.4.2. *Let the assumptions **(A1)**, **(A3)** and **(A4)** hold, and assume $h(x)$ is strictly increasing. Consider the NLC algorithm in* (4.5) *with the initial state vector $\mathbf{X}(0) \in \mathbb{R}^N$. Then, the state vector $\mathbf{X}(t)$ in* (4.5) *approaches the consensus subspace $\mathbb{B}$ a.s., i.e.,*

$$Pr\left[\lim_{t\to\infty} \inf_{\mathbf{Y}\in\mathbb{B}} \|\mathbf{X}(t) - \mathbf{Y}\| = 0\right] = 1. \tag{4.17}$$

Proof. We will make use of Theorem 4.4.1 to prove (4.17). We will choose an appropriate potential function $V(\mathbf{x})$ that is non-negative which satisfies equation (4.11). We will then prove that the generating operator $\mathcal{L}$ applied on $V(\mathbf{x})$ as in (4.10) can be upper bounded as in (4.12) with $\gamma(t) = \alpha(t)$, and a $\varphi(\mathbf{x})$ can be found that satisfies (4.13).

First we see that under the assumptions **(A1)**, **(A2)** and the assumption on $h(x)$, the discrete time vector process $\{\mathbf{X}(t)\}_{t\geq 0}$ is Markov. Since $\mathbf{L}$ is a positive semi-definite matrix, it has an eigenvalue decomposition (EVD) given by $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{\Sigma}$ is the diagonal matrix containing the eigenvalues of $\mathbf{L}$ in the increasing order, and $\mathbf{U}$ is a unitary matrix with $\mathbf{1}$ as its first column vector which corresponds to the 0 eigenvalue. Define a positive semi-definite matrix $\mathbf{M}$ as a function of $\mathbf{U}$ such that $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$ and $\mathbf{\Lambda} = \mathrm{d}iag[0,\ 1,\ 1\ , \ldots,\ 1]$. Let $V(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}$, then the function $V(\mathbf{x})$ is non-negative since $\mathbf{M}$ is a positive semi-definite matrix by construction. Note that $\mathbf{x} \in \mathbb{B}$ is an eigenvector of $\mathbf{M}$ associated with the zero eigenvalue, therefore we have

$$V(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B}\ . \tag{4.18}$$

Let $\mathbf{x} = \mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp}$ where $\mathbf{x}_{\mathbb{B}}$ is the orthogonal projection of $\mathbf{x}$ on $\mathbb{B}$. When $\mathbf{x} \in \mathbb{B}^{'}$, we have $\|\mathbf{x}_{\mathbb{B}\perp}\| > 0$. Let $\mathbf{x} \in \mathbb{B}^{'}$ and $\mathbf{h}(\mathbf{x})$ be as defined in (4.5). Then,

$\mathbf{h}(\mathbf{x}) = \mathbf{h}_{\mathbb{B}}(\mathbf{x}) + \mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})$, where $\mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})$ is non-zero, i.e., $\|\mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})\| > 0$. Define $\beta := \|\mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})\|^2/\|\mathbf{x}_{\mathbb{B}\perp}\|^2$, then $\beta > 0$, $\mathbf{x} \in \mathbb{B}'$. Therefore, for any $\mathbf{x} \in \mathbb{B}'$,

$$\begin{aligned} V(\mathbf{x}) &= \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} = V(\mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp}) = V(\mathbf{x}_{\mathbb{B}\perp}) \\ &\geq \min_{\mathbf{x}_{\mathbb{B}\perp}\neq 0} \mathbf{x}_{\mathbb{B}\perp}^{\mathrm{T}}\mathbf{M}\mathbf{x}_{\mathbb{B}\perp} = \lambda_2(\mathbf{M})\|\mathbf{x}_{\mathbb{B}\perp}\|^2 > 0\,, \end{aligned} \tag{4.19}$$

where the last inequality is due to $\lambda_2(\mathbf{L}) > 0$ by assumption **(A1)**. The equations (4.18) and (4.19) establish that the conditions in (4.11) in Theorem 4.4.1 are satisfied.

Now we will prove that (4.12) is satisfied as well. Towards this end, consider $\mathcal{L}V(\mathbf{x})$ defined in (4.10),

$$\mathcal{L}V(\mathbf{x}) = \mathrm{E}\left[\mathbf{X}(t+1)^{\mathrm{T}}\mathbf{M}\mathbf{X}(t+1)|\mathbf{X}(t) = \mathbf{x}\right] - V(\mathbf{x}), \tag{4.20}$$

$$\begin{aligned} &= \mathrm{E}\left[\left(\mathbf{x}^{\mathrm{T}} - \alpha(t)\left(\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} + \mathbf{n}(t)^{\mathrm{T}}\right)\right)\cdot\right. \\ &\quad \left.(\mathbf{M}\mathbf{x} - \alpha(t)\left(\mathbf{M}\mathbf{L}\mathbf{h}(\mathbf{x}) + \mathbf{M}\mathbf{n}(t)\right))\right] - V(\mathbf{x}) \end{aligned} \tag{4.21}$$

$$\begin{aligned} &= -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{L}\mathbf{h}(\mathbf{x})\right] \\ &\quad + \alpha^2(t)\left[\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{M}\mathbf{L}\mathbf{h}(\mathbf{x}) + \mathrm{E}\left[\mathbf{n}(t)^{\mathrm{T}}\mathbf{M}\mathbf{n}(t)\right]\right]. \end{aligned} \tag{4.22}$$

We get (4.22) by expanding (4.21) and taking the expectations and using the fact that $\mathrm{E}[\mathbf{n}(t)] = \mathbf{0}$. Recall the EVDs of $\mathbf{L}$ and $\mathbf{M}$ from which we have

$$\mathbf{L}\mathbf{M} = \mathbf{M}\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}} = \mathbf{L}\,. \tag{4.23}$$

Since $\lambda_2(\mathbf{M}) = \lambda_N(\mathbf{M}) = 1$, we have

$$\mathrm{E}\left[\mathbf{n}(t)^{\mathrm{T}}\mathbf{M}\mathbf{n}(t)\right] \leq \mathrm{E}\left[\lambda_N(\mathbf{M})\|\mathbf{n}(t)\|^2\right] \leq \mu, \tag{4.24}$$

where the second inequality follows from (4.8) and the fact that $\lambda_N(\mathbf{M}) = 1$. Using (4.23) and (4.24) in (4.22), we get the following bound

$$\mathcal{L}V(\mathbf{x}) \leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x})\right] + \alpha^2(t)\left[\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{L}^2\mathbf{h}(\mathbf{x}) + \mu\right] \tag{4.25}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x})\right] + \alpha^2(t)\left[\lambda_N^2(\mathbf{L})\beta\|\mathbf{x}_{\mathbb{B}\perp}\|^2 + \mu\right] \tag{4.26}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x})\right] + \alpha^2(t)\left[\beta\frac{\lambda_N^2(\mathbf{L})}{\lambda_2(\mathbf{M})}\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} + \mu\right] \tag{4.27}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x})\right] + m\alpha^2(t)\left[1 + \beta_2\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}\right] \tag{4.28}$$

$$\leq -\alpha(t)\varphi(\mathbf{x}) + m\alpha^2(t)\left[1 + V(\mathbf{x})\right] \ , \tag{4.29}$$

where $\varphi(\mathbf{x}) := 2\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x})$, $m := \max\{\beta\lambda_N^2(\mathbf{L})/\lambda_2(\mathbf{M}), \mu\}$, $\beta_2 := \mu/m$ and $\beta_2 \in (0, 1]$. In (4.26), we have used the fact $\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{L}^2\mathbf{h}(\mathbf{x}) \leq \lambda_N^2(\mathbf{L})\|\mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})\|^2$ and $\|\mathbf{h}_{\mathbb{B}\perp}(\mathbf{x})\|^2 = \beta\|\mathbf{x}_{\mathbb{B}\perp}\|^2$. In (4.27), we have used the fact that $\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} \geq \lambda_2(\mathbf{M})\|\mathbf{x}_{\mathbb{B}\perp}\|^2$ due to (4.19). We will now prove that $\varphi(\mathbf{x})$ in (4.29) satisfies the equation (4.13) of Theorem 4.4.1.

Recall that $\mathbf{L}$ is the Laplacian matrix of the graph and that $\mathbf{1}$ is in its null space, that is, $\mathbf{L}\mathbf{1} = \mathbf{0}$. Whenever $\mathbf{x} \in \mathbb{B}$, i.e., $\mathbf{x} = a\mathbf{1}, a \in \mathbb{R}$, then $\mathbf{h}(\mathbf{x}) = b\mathbf{1}$ for some $b \in \mathbb{R}$. This implies $\mathbf{L}\mathbf{h}(a\mathbf{1}) = \mathbf{L}b\mathbf{1} = \mathbf{0}$. Therefore we have $\varphi(\mathbf{x}) = 2\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{B}$.

To prove $\varphi(\mathbf{x}) > 0$ when $\mathbf{x} \in \mathbb{B}'$, consider $\varphi(\mathbf{x})$ for a connected graph with $\mathbf{L}$ of dimension $N \times N$,

$$\varphi(\mathbf{x}) = 2\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{x}) \tag{4.30}$$

$$= 2\left[\sum_{j\in\mathbb{N}_1}(x_1 - x_j)h(x_1) + \sum_{j\in\mathbb{N}_2}(x_2 - x_j)h(x_2) + \ldots + \sum_{j\in\mathbb{N}_N}(x_N - x_j)h(x_N)\right] \ , \tag{4.31}$$

where (4.31) follows from the structure of the symmetric matrix $\mathbf{L}$ (recall $\mathbf{L} = \mathbf{D} - \mathbf{A}$). Note that the i^{th} summation in (4.31) corresponds to the i^{th}

node. Now suppose that node i is connected to node j. Then there exists a term $(x_i - x_j)h(x_i)$ in the summation corresponding to the i^{th} node in (4.31), and a term $(x_j - x_i)h(x_j)$ in the summation corresponding to the j^{th} node in (4.31). Both of these terms can be combined as $(x_i - x_j)(h(x_i) - h(x_j))$ and this corresponds to the edge $\{i, j\} \in \mathbb{E}$. Thus equation (4.31) can be written as pairwise products enumerated over all the edges in the graph as follows

$$\varphi(\mathbf{x}) = 2 \sum_{\{i,j\} \in \mathbb{E}} (x_i - x_j)(h(x_i) - h(x_j)) \; . \tag{4.32}$$

Since $\mathbf{x} \in \mathbb{B}'$, $\varphi(\mathbf{x})$ in (4.32) is positive due to the fact that $h(x)$ is strictly increasing so that there is at least one term in the sum which is strictly greater than zero. Letting $\gamma(t) = \alpha(t), g(t) = \alpha^2(t)$ and by assumption **(A4)**, we see that the sequence $\alpha(t)$ in (4.29) satisfies (4.14). Thus all the conditions of Theorem 4.4.1 are satisfied to yield (4.17). □

Theorem 4.4.2 states that the sample paths of $\mathbf{X}(t)$ approach the consensus subspace almost surely. We note that the assumption **(A2)** is not necessary for Theorem 4.4.2 to hold. Instead we assumed $h(x)$ is strictly increasing (not necessarily differentiable) to prove Theorem 4.4.2. Now, like in [65], we will prove the convergence of $\mathbf{X}(t)$ to a finite point in $\mathbb{B}$ in Theorem 4.4.3.

Theorem 4.4.3. *Let the assumptions of Theorem 4.4.2 hold. Consider the NLC algorithm in* (4.5) *with the initial state* $\mathbf{X}(0) \in \mathbb{R}^N$. *Then, there exists a finite real random variable* θ^* *such that*

$$Pr\left[\lim_{t \to \infty} \mathbf{X}(t) = \theta^* \mathbf{1}\right] = 1. \tag{4.33}$$

Proof. Let the average of $\mathbf{X}(t)$ be $\bar{x}(t) = \mathbf{1}^{\mathrm{T}}\mathbf{X}(t)/N$. Since $\mathbf{1}\bar{x}(t) \in \mathbb{B}$, Theorem 4.4.2 implies,

$$Pr\left[\lim_{t\to\infty} \|\mathbf{X}(t) - \bar{x}(t)\mathbf{1}\| = 0\right] = 1\ , \tag{4.34}$$

where (4.34) follows from (4.17) since the infimum in (4.17) is achieved by $\mathbf{Y} = \bar{x}(t)\mathbf{1}$. Pre-multiplying (4.5) by $\mathbf{1}^{\mathrm{T}}/N$ on both sides and noting that $\mathbf{1}^{\mathrm{T}}\mathbf{L}\mathbf{h}(\mathbf{X}(t)) = 0$ we get,

$$\bar{x}(t+1) = \bar{x}(t) - \tilde{v}(t) \tag{4.35}$$

$$= \bar{x}(0) - \sum_{0\le k\le t} \tilde{v}(k) \tag{4.36}$$

where $\tilde{v}(t) = \alpha(t)\mathbf{1}^{\mathrm{T}}\mathbf{n}(t)/N$. From assumption **(A3)**, it follows that

$$\mathrm{E}[\tilde{v}(t)] = 0,$$

$$\sum_{t\ge 0} \mathrm{E}[\tilde{v}(t)]^2 = \sum_{t\ge 0} \frac{\alpha^2(t)}{N^2}\mathrm{E}\|\mathbf{n}(t)\|^2 \le \frac{\mu}{N^2}\sum_{t\ge 0}\alpha^2(t) < \infty$$

which implies

$$\mathrm{E}[\bar{x}(t+1)]^2 \le \bar{x}^2(0) + \frac{\mu}{N^2}\sum_{t\ge 0}\alpha^2(t)\ , \forall t\ . \tag{4.37}$$

Equation (4.37) implies that the sequence $\{\bar{x}(t)\}_{t\ge 0}$ is an $\mathcal{L}_2$ bounded martingale [1] and hence converges a.s. and in $\mathcal{L}_2$ to a finite random variable θ^* (see [118, Theorem 2.6.1]). Therefore the theorem follows from (4.34). □

It should be noted that the results in Theorems 4.4.2 and 4.4.3 are similar to the results in [65], but we have proved it for a more general case of which [65] is a special case when $\mathbf{h}(\mathbf{x}) = \mathbf{x}$. In what follows, we present the properties of the limiting random variable θ^*.

[1] A sequence of random variables $\{y(t)\}_{t\ge 0}$ is called as a martingale if for all $t \ge 0$, $\mathrm{E}[|y(t)|] < \infty$ and $\mathrm{E}[y(t+1) \mid y(1)\, y(2) \ldots y(t)] = y(t)$. The sequence $\{y(t)\}_{t\ge 0}$ is an $\mathcal{L}_2$ bounded martingale if $\sup_t \mathrm{E}\left[y^2(t)\right] < \infty$ (see [119, pp. 110]).

4.4.3 Mean Square Error of NLC Algorithm

The Theorems 4.4.2 and 4.4.3 establish that the sensors reach consensus asymptotically and converge a.s. to a finite random variable θ^*. We can view θ^* as an estimate of $\bar{x}$. In the following theorem we characterize the unbiasedness and means squared error (MSE) properties of θ^*. We define the MSE of θ^* as $\xi_{_N} = \mathrm{E}[(\theta^* - \bar{x})^2]$.

Theorem 4.4.4. *Let θ^* be the limiting random variable as in Theorem 4.4.3. Then θ^* is unbiased, $E[\theta^*] = \bar{x}$, and its MSE is bounded, $\xi_{_N} \leq \mu N^{-2} \sum_{t\geq 0} \alpha^2(t)$.*

The proof is obtained by following the same steps of the Lemma 5 in [65].

We point out that with non-linear transmissions, we have obtained the same bound on the MSE $\xi_{_N}$ as that of the linear consensus algorithm in [65]. It should be noted that $\mu \leq N d_{\max} \sigma^2$ from (4.8) which implies that $\xi_{_N} \leq d_{\max} N^{-1} \sum_{t\geq 0} \alpha^2(t) \sigma^2$. Therefore, if $d_{\max}$ is finite for a large connected network, we have $\lim_{N\to\infty} \xi_{_N} = 0$ and this means that θ^* converges to $\bar{x}$ as the variance of θ^* approaches 0. If the graph is densely connected, then $d_{\max}$ is relatively high which increases the worst-case MSE. On the other hand, when the graph is densely connected, $\lambda_2(\mathbf{L})$ is larger which aids in the speed of convergence to θ^*, as quantified through the covariance matrix in Section 4.4.4.

For any connected graph with N nodes, if $\sigma^2 = 0$ then $\lim_{t\to\infty} \mathbf{X}(t) = \bar{x}\mathbf{1}$, which means all the sensor states asymptotically converge to the desired sample average. In fact, in the absence of communication noise, under assumptions **(A1)** and **(A2)**, we believe that it is possible to prove exponential

convergence of $\mathbf{X}(t)$ to $\bar{x}\mathbf{1}$ by letting $\alpha(t) = \alpha$ such that $0 < \alpha < 2/(c\lambda_N(\mathbf{L}))$ and by following a similar approach as in [113].

Similar results as in Theorems 4.4.2 and 4.4.3 could be easily proved under more general assumptions. For example, the graph can be randomly varying over time due to link failures. As long as the graph is connected on an average, it can be easily proved that the Theorems 4.4.2 and 4.4.3 hold. The independent assumption on the noise sequence can also relaxed and the noise sequence can be allowed to depend on $\mathbf{X}(t)$. For detailed discussions on these assumptions and its variations, please see Section III-A in [65]. We do not pursue these extensions herein since our focus is on studying the effect of non-linear transmissions on performance.

4.4.4 Asymptotic Normality of NLC Algorithm

The NLC algorithm in (4.5) belongs to the class of stochastic approximation algorithms. The convergence speed of these algorithms is an important issue from a practical perspective. There are various criteria for determining the rate of convergence. For instance, one can try to estimate $\mathrm{E}\left[\|\mathbf{X}(t) - \theta^*\mathbf{1}\|^2\right]$ or $\mathrm{P}r\left[\|\mathbf{X}(t) - \theta^*\mathbf{1}\| \leq \epsilon(t)\right]$ [120]. Estimating these parameters may be difficult in practice. However, it is usually possible to establish that $\sqrt{t}(\mathbf{X}(t) - \theta^*\mathbf{1})$ is asymptotically normal with zero mean and some covariance matrix. Asymptotic normality of stochastic approximation algorithms have been established under some general conditions in [118] and for the linear consensus algorithms in [63].

In this section, we establish the asymptotic normality of the NLC algorithm in (4.5). Our approach here is similar to the one in [63]. Basically, we decompose the NLC algorithm in $\mathbb{R}^N$ into a scalar recursion and a recursion in $\mathbb{R}^{(N-1)}$. In this section, for the sake of simplicity we assume that the noise

sequence $\{\mathbf{n}(t), t \geq 0\}$ are i.i.d. random vectors with zero mean and finite covariance. We now formally state and prove the result as a theorem.

Theorem 4.4.5. *Let $\alpha(t) = a/(t+1), a > 0$, then the NLC algorithm in* (4.5) *becomes*

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t}\left[-\mathbf{L}\mathbf{h}(\mathbf{X}(t)) + \mathbf{n}(t)\right]. \tag{4.38}$$

*Suppose that the assumptions **(A1)**, **(A2)**, **(A3)** and **(A4)** hold and that the noise sequence $\{\mathbf{n}(t), t \geq 0\}$ are i.i.d. across time and space with zero mean and covariance $\sigma_v^2\mathbf{I}$. Let the EVD of $\mathbf{L}$ be given by $\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{U}$ is a unitary matrix whose columns are the eigenvectors of $\mathbf{L}$ such that*

$$\mathbf{U} = \left[\frac{\mathbf{1}}{\sqrt{N}} \;\; \mathbf{\Phi}\right], \mathbf{\Phi} \in \mathbb{R}^{N\times(N-1)}\ ,\ -\mathbf{\Sigma} = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{B} \end{bmatrix}\ , \tag{4.39}$$

where $\mathbf{B} \in \mathbb{R}^{(N-1)\times(N-1)}$ is a diagonal matrix containing the $N-1$ negative eigenvalues of $-\mathbf{L}$ (this means that $\mathbf{B}$ is a stable matrix). In addition, let θ_0 be a realization of the random variable θ^* and $2a\lambda_2(\mathbf{L})h'(\theta_0) > 1$ so that the matrix $\left[ah'(\theta_0)\mathbf{B} + \mathbf{I}/2\right], \theta_0 \in \mathbb{R}$ is stable. Define $[\tilde{n}(t) \;\; \tilde{\mathbf{n}}(t)^T]^T := N^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{n}(t)$, $\tilde{\mathbf{n}}(t) \in \mathbb{R}^{(N-1)}$, so that $\tilde{n}(t) = N^{-1}\mathbf{1}^{\mathrm{T}}\mathbf{n}(t)$ and $\tilde{\mathbf{n}}(t) = N^{-1/2}\mathbf{\Phi}^{\mathrm{T}}\mathbf{n}(t)$. Let $\mathbf{C} = E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T]$, $\mathbf{C} \in \mathbb{R}^{(N-1)\times(N-1)}$. Then, as $t \to \infty$,

$$\sqrt{t}(\mathbf{X}(t) - \theta^*\mathbf{1}|\theta^* = \theta_0) \sim \mathcal{N}\left(0, N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}} + N^{-1}\mathbf{\Phi}\mathbf{S}^{\theta_0}\mathbf{\Phi}^{\mathrm{T}}\right)\ , \tag{4.40}$$

where

$$\mathbf{S}^{\theta_0} = a^2\int_0^\infty e^{\left[ah'(\theta_0)\mathbf{B}+\frac{\mathbf{I}}{2}\right]t}\,\mathbf{C}\,e^{\left[ah'(\theta_0)\mathbf{B}+\frac{\mathbf{I}}{2}\right]t}dt\ . \tag{4.41}$$

Proof. Define $[\tilde{x}(t) \;\; \tilde{\mathbf{X}}(t)^{\mathrm{T}}]^{\mathrm{T}} := N^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{X}(t), \tilde{\mathbf{X}}(t) \in \mathbb{R}^{(N-1)}$. From Theorem 4.4.3, we have $\mathbf{X}(t) \to \theta^*\mathbf{1}$ a.s. as $t \to \infty$ which implies that $[\tilde{x}(t) \;\; \tilde{\mathbf{X}}(t)]^{\mathrm{T}} \to$

$[\theta^* \quad \mathbf{0}]^{\mathrm{T}}$ a.s. as $t \to \infty$, and therefore $\tilde{\mathbf{X}}(t) \to \mathbf{0}$ a.s. as $t \to \infty$. The error $[\mathbf{X}(t) - \theta_0 \mathbf{1}]$ can be written as the sum of two error components (see also Section VI in [63]) as given below

$$[\mathbf{X}(t) - \theta_0 \mathbf{1}] = [\tilde{x}(t) - \theta_0]\mathbf{1} + \frac{1}{\sqrt{N}} \mathbf{\Phi} \tilde{\mathbf{X}}(t) \ , \tag{4.42}$$

$$= \mathbf{e}_1 + \mathbf{e}_2 \ , \tag{4.43}$$

where $\mathbf{e}_1 = [\tilde{x}(t) - \theta_0]\mathbf{1}$ and $\mathbf{e}_2 = N^{-1/2} \mathbf{\Phi} \tilde{\mathbf{X}}(t)$. By calculating the covariance matrix between $\mathbf{e}_1$ and $\mathbf{e}_2$, it can be proved that they are asymptotically uncorrelated as $t \to \infty$, and that asymptotically $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, N^{-1} a^2 \sigma_v^2 \mathbf{1}\mathbf{1}^{\mathrm{T}})$ (see Theorem 12 in [63]). To show that $\sqrt{t}\mathbf{e}_2$ is asymptotically normal, it suffices to show that $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal. To this end, express $h(x)$ in (4.38) around $x = \theta_0$ using Taylor's series expansion,

$$h(x) = h(\theta_0) + h^{'}(\theta_0)(x - \theta_0) + o(|x - \theta_0|) \ , \mathrm{a}s \ x \to \theta_0 \ . \tag{4.44}$$

Using (4.44) in (4.38) we get

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t+1} \left[-\mathbf{L} \left(h(\theta_0)\mathbf{1} + h^{'}(\theta_0)[\mathbf{X}(t) - \theta_0 \mathbf{1}] \right) + \delta(\mathbf{X}(t)) + \mathbf{n}(t) \right] \tag{4.45}$$

$$= \mathbf{X}(t) + \frac{a}{t+1} \left[h^{'}(\theta_0) \left(-\mathbf{L}\mathbf{X}(t) \right) + \delta(\mathbf{X}(t)) + \mathbf{n}(t) \right] \ , \ \mathrm{a}s \ t \to \infty \ , \tag{4.46}$$

where $\|\delta(\mathbf{X}(t))\| \to 0$ as $t \to \infty$. Pre-multiplying (4.46) on both sides by $N^{-1/2}\mathbf{U}^{\mathrm{T}}$ and using (4.39) we get the following recursions

$$\tilde{x}(t+1) = \tilde{x}(t) + \frac{a}{t+1} \tilde{n}(t) \ , \tag{4.47}$$

$$\tilde{\mathbf{X}}(t+1) = \tilde{\mathbf{X}}(t) + \frac{a}{t+1} \left[h^{'}(\theta_0) \mathbf{B} \tilde{\mathbf{X}}(t) + \tilde{\delta}(\mathbf{X}(t)) + \tilde{\mathbf{n}}(t) \right] \ , \mathrm{a}s \ t \to \infty \ , \tag{4.48}$$

where $\tilde{\delta}(\mathbf{X}(t)) = N^{-1/2}\mathbf{\Phi}^{\mathrm{T}}\delta(\mathbf{X}(t))$. With the assumption that $\left[ah^{'}(\theta_0)\mathbf{B} + \mathbf{I}/2\right]$, $\theta_0 \in \mathbb{R}$ is a stable matrix, it can be verified that all the conditions of Theorem 6.6.1 in [118, p. 147] are satisfied for the process $\tilde{\mathbf{X}}(t)$ in (4.48). Therefore, for a given θ_0, the process $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal with zero mean and covariance matrix given by (4.41). Since $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}})$ and using (4.41) together with the fact that $\mathbf{e}_1$ and $\mathbf{e}_2$ are asymptotically independent as $t \to \infty$, we get (4.40) which completes the proof. □

Equation (4.40) indicates how fast the process $\sqrt{t}(\mathbf{X}(t) - \theta_0\mathbf{1})$ will converge to $\theta_0\mathbf{1}$ for a given θ_0 as $t \to \infty$. The convergence speed clearly depends on $h^{'}(\theta_0)$. We note that if $h(x) = x$, then $h^{'}(\theta_0) = 1, \forall\theta_0 \in \mathbb{R}$, and substituting this in (4.41), we get the results for the linear case as in Theorem 12 of [63].

Let the asymptotic covariance in (4.40) be denoted by $\mathbf{C}_{\mathrm{nlc}}$. Since $\mathbf{n}(t)$ are i.i.d., $\mathbf{C}$ in (4.41) becomes $\mathbf{C} = \sigma_v^2\mathbf{I}$ and thus we have $\mathbf{C}_{\mathrm{nlc}} = N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}} + N^{-1}\mathbf{\Phi}\mathbf{S}^{\theta_0}\mathbf{\Phi}^{\mathrm{T}}$ where $\mathbf{S}^{\theta_0}$ is a diagonal matrix whose diagonal elements are given by $\mathbf{S}_{ii}^{\theta_0} = a^2\sigma_v^2/[2ah^{'}(\theta_0)\lambda_{i+1}(\mathbf{L}) - 1]$. A reasonable quantitative measure of largeness [120] of the asymptotic covariance matrix is $\|\mathbf{C}_{\mathrm{nlc}}\|$ which is the maximum eigenvalue of the symmetric matrix $\mathbf{C}_{\mathrm{nlc}}$. Further, $\|\mathbf{C}_{\mathrm{nlc}}\|$ can be minimized with respect to the parameter a. This can be formulated as the following optimization problem,

$$\min_{\{a|2ah^{'}(\theta_0)\lambda_2(\mathbf{L})>1\}} \max_{\{\mathbf{x}|\mathbf{x}\in\mathbb{R}^N, \|\mathbf{x}\|^2\leq 1\}} \mathbf{x}^{\mathrm{T}}\mathbf{C}_{\mathrm{nlc}}\mathbf{x}\,, \tag{4.49}$$

which can be solved analytically by using the KKT conditions [90]. The value of a that optimizes (4.49) is $a_{nlc}^{*} = (N+1)/[2N\lambda_2(\mathbf{L})h^{'}(\theta_0)]$ and the corresponding optimal value of the $\|\mathbf{C}_{\mathrm{nlc}}\|$is given by

$$\|\mathbf{C}_{\mathrm{nlc}}^{*}\| = \left(\frac{N+1}{2N}\right)^2 \left(\frac{\sigma_v^2}{\lambda_2^2(\mathbf{L})}\right) \left(\frac{1}{h^{'}(\theta_0)}\right)^2 . \tag{4.50}$$

The size of the asymptotic covariance matrix in (4.50) is inversely proportional to the square of the smallest non-zero eigenvalue $\lambda_2(\mathbf{L})$ which quantifies how densely a graph is connected. We also note that (4.50) is directly proportional to the channel noise variance σ_v^2.

Equation (4.50) also gives some useful insights to design the transmission function $h(x)$. If we choose two functions $h_1(x)$ and $h_2(x)$ such that $h_1^{'}(x) > h_2^{'}(x), \forall x \in \mathbb{R}$, it is easy to see from (4.50) that $\|\mathbf{C}^*_{\text{nlc1}}\| < \|\mathbf{C}^*_{\text{nlc2}}\|, \forall \theta_0 \in \mathbb{R}$. This means that the convergence will be faster when $h_1(x)$ is employed in the NLC algorithm (4.5) than when $h_2(x)$ is employed. However, it should be noted that if $h_1^{'}(x) > h_2^{'}(x), \forall x \in \mathbb{R}$ and suppose $h_1(0) = h_2(0) = 0$ then we have $h_1^2(x) > h_2^2(x), \forall x$ which implies that on an average the transmit power is greater when $h_1(x)$ is employed compared to $h_2(x)$. Thus, optimization of asymptotic covariance with respect to the parameter a helps us to do a comparative study among different $h(x)$ functions without the knowledge of the limit point θ_0. We will illustrate these findings in the simulations in Section 4.5. Comparing the $\|\mathbf{C}^*_{\text{nlc}}\|$ against the special case of $h(x) = cx$ yields $\|\mathbf{C}^*_{\text{nlc}}\| = \|\mathbf{C}^*_{\text{lin}}\|(c/h^{'}(\theta_0))^2$. Clearly $c/h^{'}(\theta_0) \leq 1$ and therefore if $h(x)$ is bounded, appropriately normalized by letting $c = 1$, so that $0 < h^{'}(x) \leq 1$, we conclude that the best case linear algorithm outperforms the best case NLC algorithm in terms of speed of convergence. However, the improved asymptotic covariance matrix in the former is achieved at the cost of increased peak and average transmit power compared to the latter.

4.5 Simulations

In this section, we corroborate our analytical findings through various simulations. In all the simulations presented, the initial samples $x_i(0) \in \mathbb{R}, i = 1, 2, \ldots, N$, were generated randomly using Gaussian distribution with a stan-

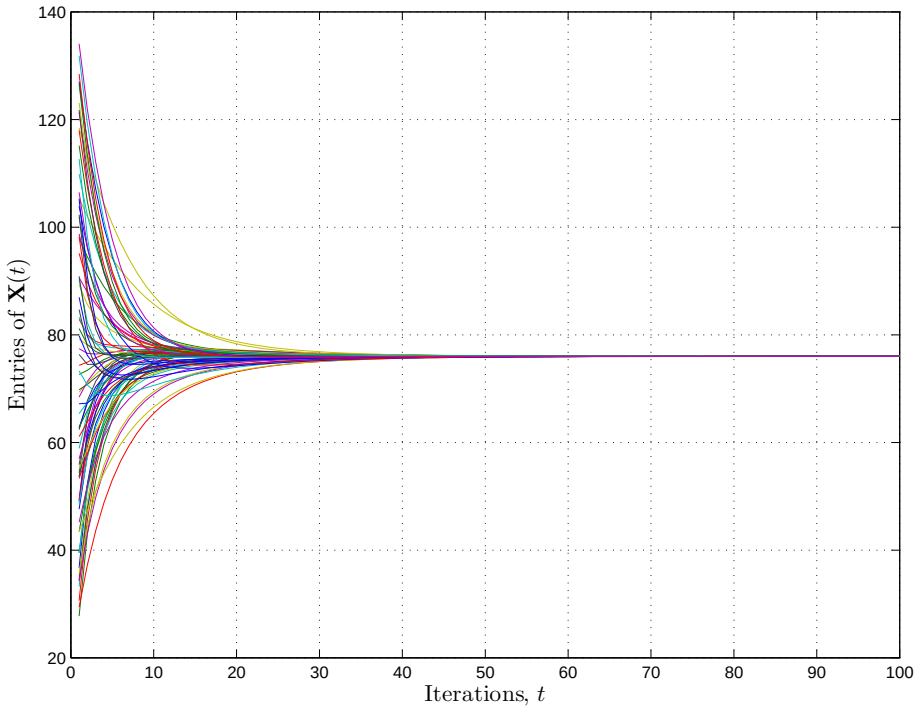

Figure 4.1: Entries of $\mathbf{X}(t)$ versus Iterations t: $\alpha = 1.5$, $\omega = 0.01$, $N = 75$, $h(x) = \tanh(\omega x)$, $\bar{x} = 76$.

dard deviation equal to 10. The desired global average value is indicated in each of the simulations. We focus here on bounded transmission functions to study their performance. Please note that our results are valid for a broader class of increasing functions (see Section 4.4.1) than the ones considered in this section.

4.5.1 Performance of NLC Algorithm without Channel Noise

Our focus in this chapter is on non-linear transmissions in the presence of noise. However, we would also like to illustrate the convergence behavior on the absence of noise. Figures 4.1, 4.2 and 4.3 depict the performance of the proposed NLC algorithm in the absence of channel noise for a large network with $N = 75$. In all the cases, we have used α values such that $0 < \alpha < 2/(c\lambda_N(\mathbf{L}))$ as mentioned in Section 4.4.3.

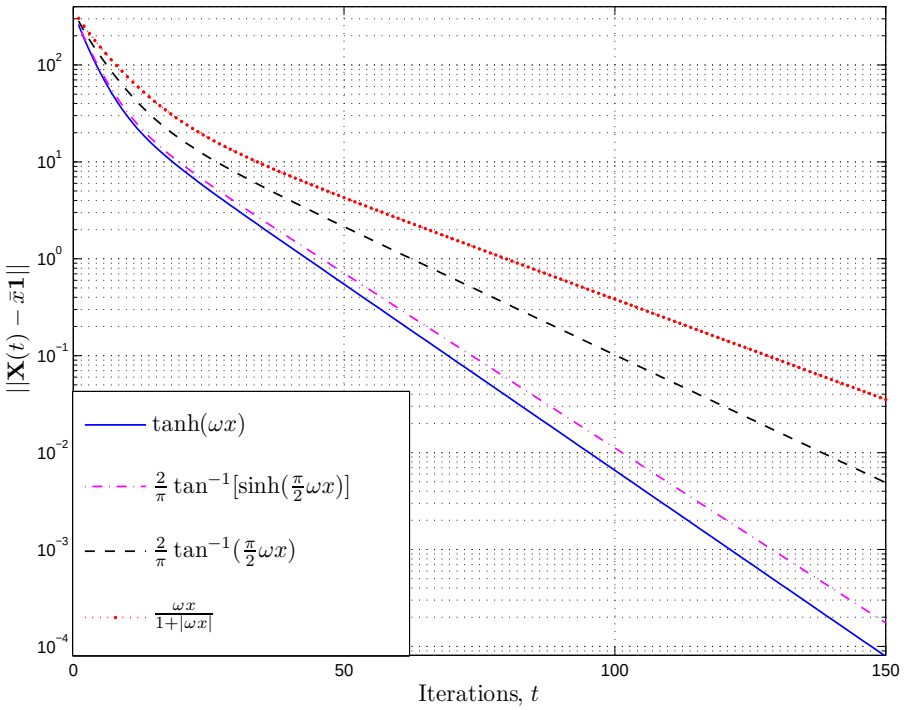

Figure 4.2: Evolution of error $||\mathbf{X}(t) - \bar{x}\mathbf{1}||$ versus Iterations t: $\alpha = 1.5$, $\omega = 0.01$, $N = 75$, $\bar{x} = 76$.

From Figure 4.1, we infer that in about 50 iterations, all the nodes reach consensus on the desired global average of $\bar{x} = 76$. Figure 4.2 shows evolution of error norm $||\mathbf{X}(t) - \bar{x}\mathbf{1}||$ for various bounded functions. We see that the convergence is exponential in all cases as noted in Section 4.4.3. Figure 4.3 illustrates the performance of the NLC algorithm when α is varied. Interestingly, by adjusting the step size α it is indeed possible to achieve the same convergence speed using the NLC algorithm as that of optimal linear consensus algorithm using the Laplacian heuristic [58].

4.5.2 Performance of NLC Algorithm with Channel Noise

Figures 4.4 - 4.8 illustrate the performance of NLC algorithm in the presence of communication noise. As explained in the assumption **(A4)** in Section 4.4.1, we chose the decreasing step sequence to be $\alpha(t) = 1/(t+1), t \geq 0$, in

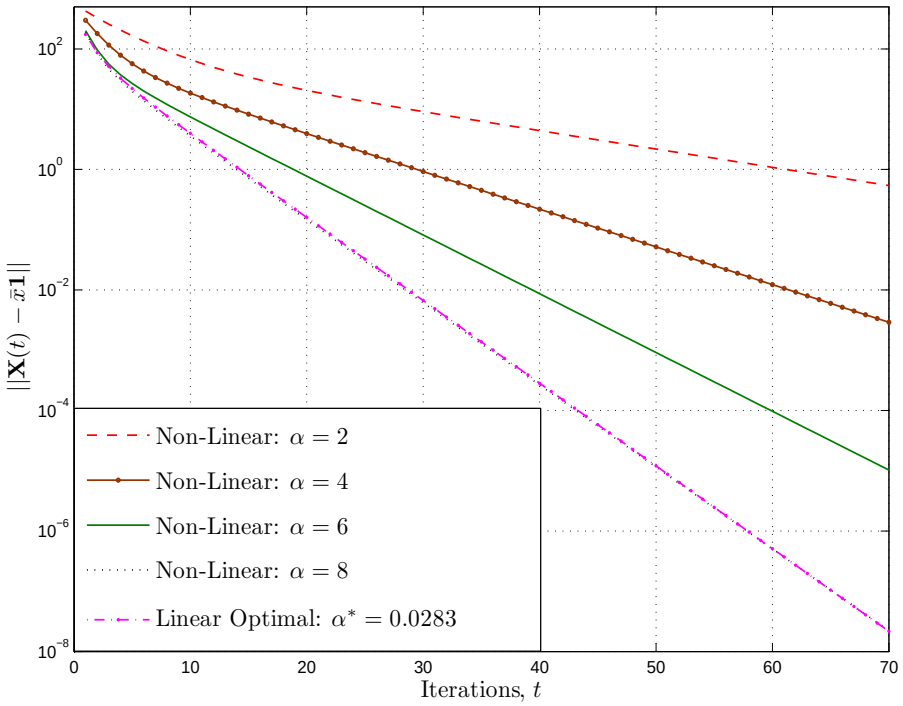

Figure 4.3: Evolution of error $||\mathbf{X}(t) - \bar{x}\mathbf{1}||$ versus Iterations t: $\alpha = 2, 4, 6, 8$, $\omega = 0.005$, $N = 75$, $h(x) = \frac{2}{\pi}\tan^{-1}[\sinh(\frac{\pi}{2}\omega x)]$, $\bar{x} = 114$.

all simulations. Here we assumed that $\rho = \max_x h^2(x)$ is the maximum power available at each sensor to transmit its state value. Figure 4.4 shows that the nodes employing the NLC algorithm reach consensus for a small network with $N = 10$. Figure 4.5 shows the transmit power $h^2(x_i(t)), i = 1, 2, \ldots, N$, per-neighbour versus iterations for a large network. Clearly, the transmit power is always constrained within the upper bound of ρ (indicated by the dashed line) making the proposed scheme practically viable for the power constrained WSNs.

In Figures 4.6, 4.7 and 4.8, we show the convergence speed performance of the proposed NLC algorithm by plotting $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus iterations t. These plots indicate how fast the mean of the process $\mathbf{X}(t)$ converges towards the desired global mean vector $\bar{x}\mathbf{1}$.

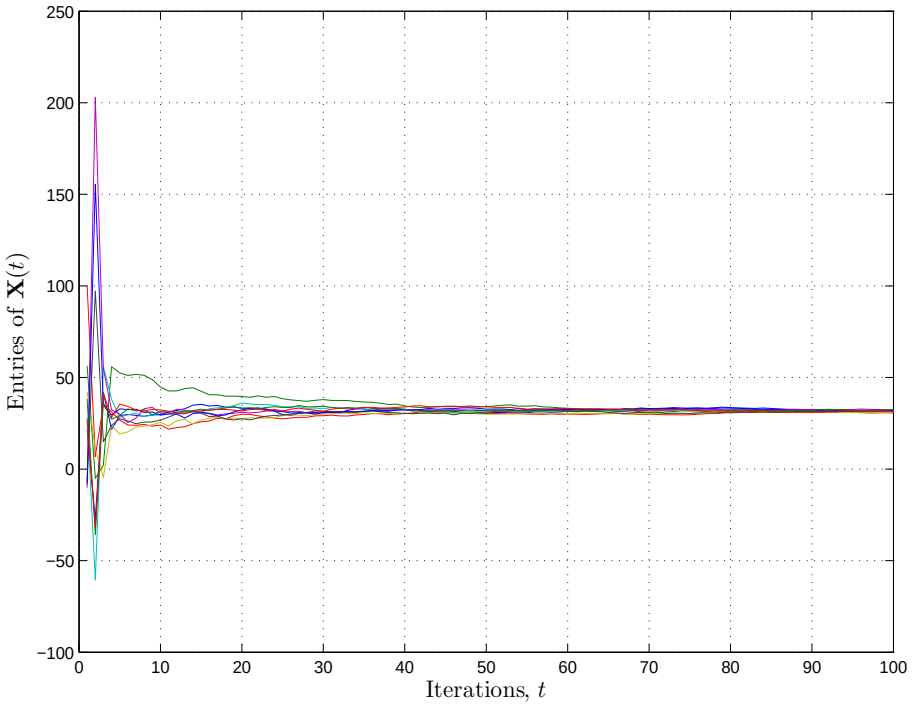

Figure 4.4: Entries of $\mathbf{X}(t)$ versus Iterations t: $h(x) = \sqrt{\rho}\tanh(\omega x)$, $\omega = 0.05$, $N = 10$, $\bar{x} = 36.24$, $\rho = 10$ dB, $\sigma_v^2 = 1$.

In Theorem 4.4.5, we saw that if two functions $h_1(x)$ and $h_2(x)$ such that $h_1^{'}(x) > h_2^{'}(x), \forall x \in \mathbb{R}$, are employed in the NLC algorithm then the convergence will be faster for $h_1(x)$ compared to that of $h_2(x)$. This is illustrated in Figure 4.6 where we have chosen $h_1(x) = \sqrt{\rho}\tan^{-1}(\omega x)$ and $h_2(x) = \sqrt{\rho}\tanh(\omega x)$. The performance gain of $h_1(x)$ obtained over $h_2(x)$ can be understood intuitively by observing that on an average the transmit power will be more when $h_1(x)$ is employed than when $h_2(x)$ is employed. The speed of convergence for various transmit functions appropriately normalized to have the same peak power ρ is shown in Figure 4.7. Here, we see that the transmit function $h_1(x)$ has the best performance and $h_4(x)$ has the worst performance. Intuitively this is due to the fact that $h_1^{'}(\bar{x}) > h_2^{'}(\bar{x}) > h_3^{'}(\bar{x}) > h_4^{'}(\bar{x})$. Finally, we depict the convergence speed versus the power scaling constant ρ, the upper bound on the transmit power, in Figure 4.8. For a given transmit

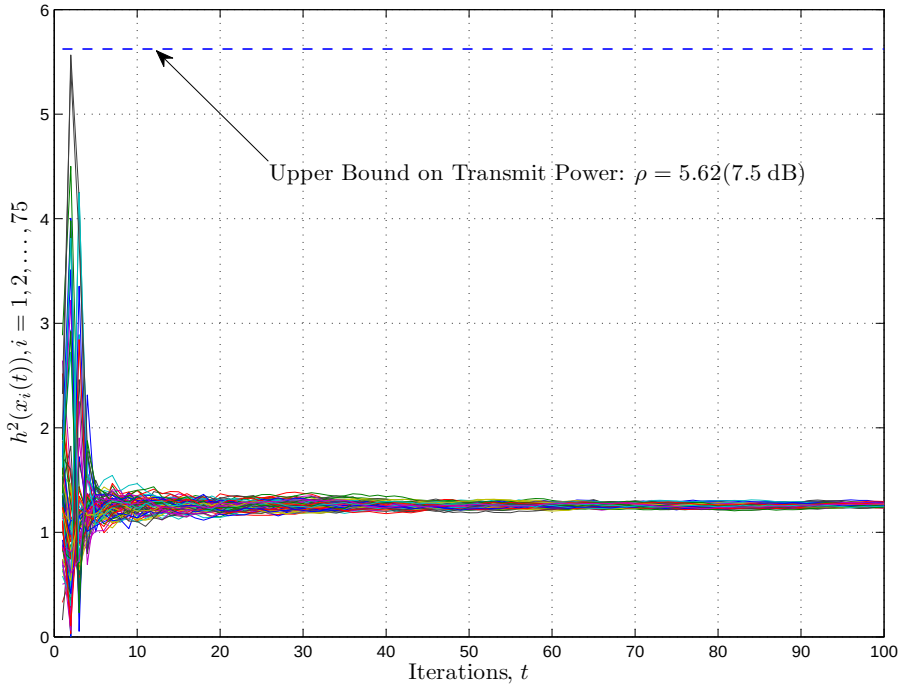

Figure 4.5: Transmit power $h^2(x_i(t))$ per-neighbour versus Iterations t: $h(x) = \sqrt{\rho}\tanh(\omega x)$, $\omega = 0.005$, $N = 75$, $\bar{x} = 102$, $\rho = 7.5$ dB, $\sigma_v^2 = 0.1$.

function, increased power leads to faster convergence as would be expected, and we also observe that when the consensus iterations were increased, speed of convergence improves.

4.6 Distributed Consensus on other Functions using NLC Algorithm

In the consensus literature, so far the focus has been mostly on the computation of average of the samples measured at $t = 0$. There has been little emphasis on the actual sensing model. The estimation of other statistics such as variance and SNR are equally important in statistical inference problems just like the mean. Moreover, when the sensing noise is Cauchy distributed, the sample mean is not a consistent estimator of a location parameter any more [38, 42–44]. To overcome these bottlenecks in the consensus set up, we propose a scheme which is robust to impulsive sensing noise distributions and using which joint estimation of mean and variance would be possible.

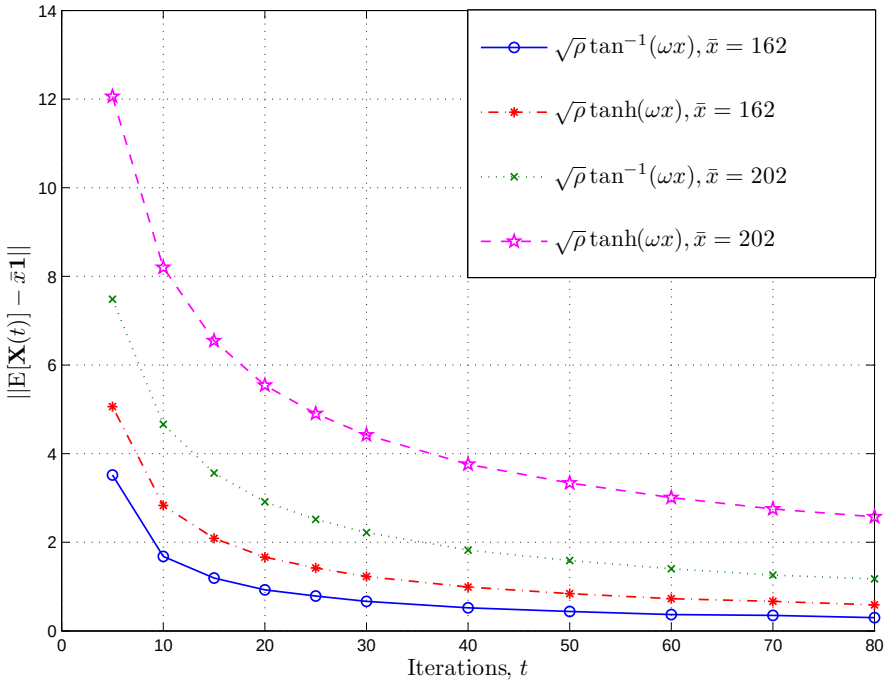

Figure 4.6: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: $h_1(x) = \sqrt{\rho}\tan^{-1}(\omega x)$, $h_2(x) = \sqrt{\rho}\tanh(\omega x)$, $\omega = 0.005$, $N = 75$, $\bar{x} = 162, 202$, $\rho = 7.5$ dB, $\sigma_v^2 = 1$.

4.6.1 Distributed Variance and SNR Estimation

Consider the following sensing model

$$x_i(0) = \theta + \sigma n_i \qquad\qquad i = 1, \ldots, N\ , \tag{4.51}$$

where θ is an unknown real-valued parameter in a bounded interval $[0,\ \theta_R]$ of known length, $\theta_R < \infty$, n_i are a mutually independent, symmetric real-valued noise with zero median (i.e., its PDF, when it exists, is symmetric about zero), and x_i is the measurement at the i^{th} sensor. Note that n_i are not necessarily identically distributed, bounded, and need not have finite moments and σ is the scale parameter which measures the standard deviation when the standard deviation of n_i exists. The sensing SNR is defined as $\gamma_{\mathrm{s}} := \theta^2/\sigma^2$.

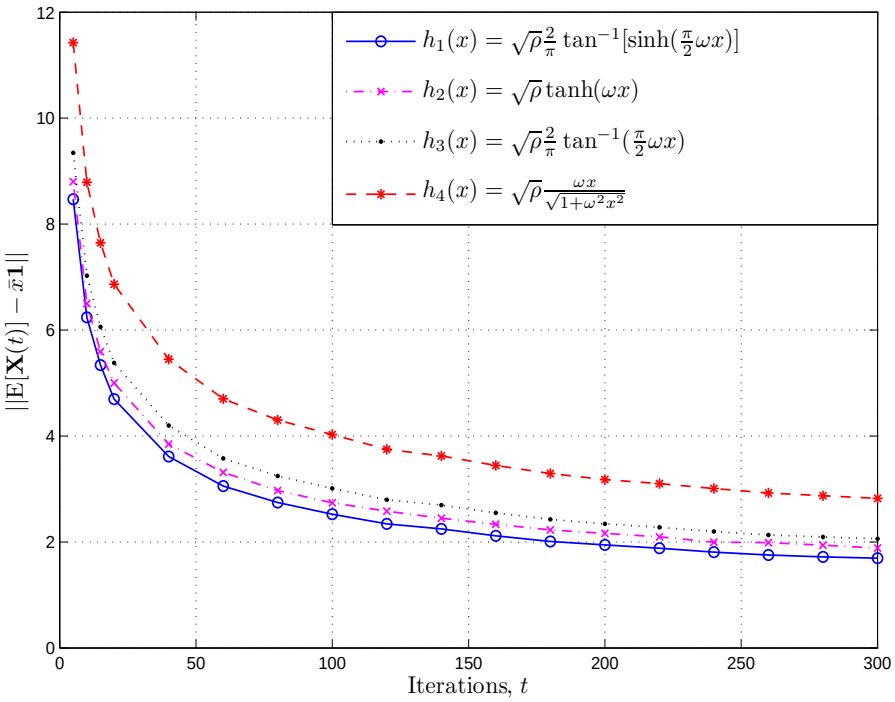

Figure 4.7: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: $\omega = 0.04$, $N = 10$, $\bar{x} = 36.24$, $\rho = 5$ dB, $\sigma_v^2 = 1$.

4.6.1.1 Pre-processing the Sensing Measurements

Let n_i are Cauchy distributed and suppose we employ the NLC algorithm in (4.5), the state values of the nodes will converge closer to $N^{-1}\sum_{i=0}^{N}(\theta + \sigma n_i)$ which is not a consistent estimator of θ for Cauchy noise. In order to solve this problem, we introduce the pre-processing of the initial sensing measurements. Let

$$y_i(0) = e^{jx_i(0)} = [y_i^{\mathrm{R}}(0) + jy_i^{\mathrm{I}}(0)], \quad i = 1, \ldots, N\ , \tag{4.52}$$

where $y_i(0)$ are the pre-processed measurements. We now treat $y_i(0)$ to be the initial measurements. We can apply either the linear consensus algorithm or the NLC algorithm in (4.5). Consider the following recursive consensus

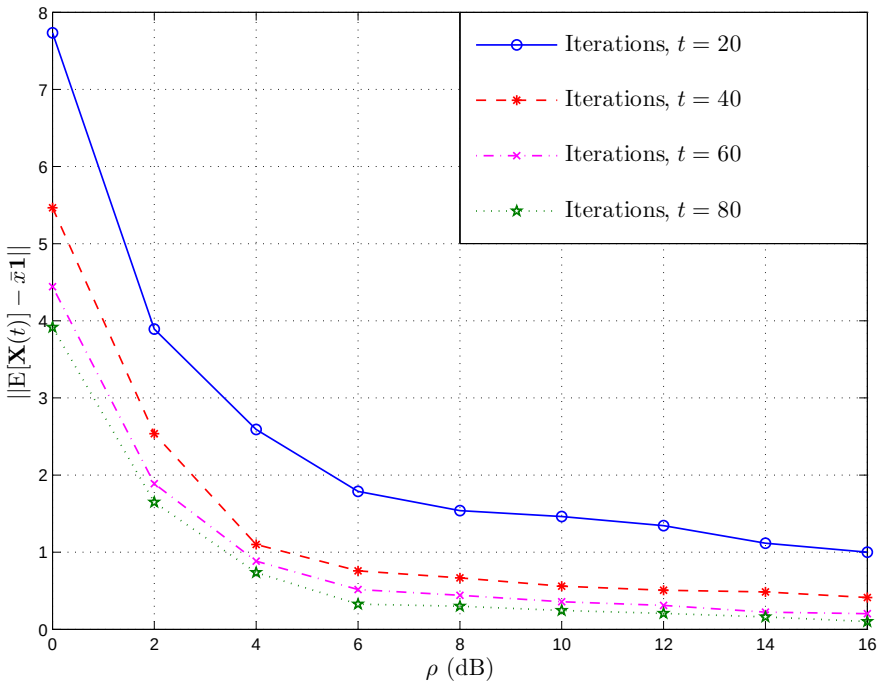

Figure 4.8: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus ρ: $h(x) = \sqrt{\rho}\frac{\omega x}{\sqrt{1+\omega^2 x^2}}$, $\omega = 0.006$, $N = 75$, $\bar{x} = 77$, Iterations $t = 20, 40, 60, 80$, $\sigma_v^2 = 1$.

algorithms

$$y_i^{\mathrm{R}}(t+1) = y_i^{\mathrm{R}}(t) - \alpha(t)\left[\sum_{j\in\mathbb{N}_i} h(y_i^{\mathrm{R}}(t)) - h(y_{ij}^{\mathrm{R}}(t)) + n_{ij}^{\mathrm{R}}(t)\right], \tag{4.53}$$

$$y_i^{\mathrm{I}}(t+1) = y_i^{\mathrm{I}}(t) - \alpha(t)\left[\sum_{j\in\mathbb{N}_i} h(y_i^{\mathrm{I}}(t)) - h(y_{ij}^{\mathrm{I}}(t)) + n_{ij}^{\mathrm{I}}(t)\right], \tag{4.54}$$

where $i = 1, \ldots, N, t = 0, 1, 2, \ldots$, indicates recursion in time. We assume that the function $h(x)$ is differentiable, and is increasing in x. In equation (4.53), the value $y_i^{\mathrm{R}}(t+1)$ is the state update of node i at time $t+1$, $y_{ij}^{\mathrm{R}}(t)$ is the state value of the j^{th} neighbour of node i at time t and $\alpha(t)$ is a positive decreasing step size. The node j transmits its information $y_{ij}^{\mathrm{R}}(t)$ by mapping it through the function $h(x)$, node i receives a noisy version of $h(y_{ij}^{\mathrm{R}}(t))$ and $n_{ij}^{\mathrm{R}}(t)$ is the zero mean additive noise with variance $\sigma^2/2$, and similarly for (4.54).

From Theorems 4.4.2 and 4.4.3, we have

$$\lim_{t\to\infty}[y_i^{\mathrm{R}}(t)+jy_i^{\mathrm{I}}(t)]=\theta^*=\theta^{\mathrm{R}}+j\theta^{\mathrm{I}},\ \forall i, \tag{4.55}$$

and

$$\mathrm{E}[\theta^*]=e^{j\theta}\left[\frac{1}{N}\sum_{i=1}^{N}e^{jn_i}\right]\ . \tag{4.56}$$

Now from (4.56), when N is sufficiently large, we have

$$\lim_{N\to\infty}\mathrm{E}[\theta^*]=e^{j\theta}\varphi(\sigma)\ , \tag{4.57}$$

where $\varphi(\sigma)$ is the characteristic function of n_i in (4.51).

From (4.55) it is possible to estimate θ,σ and γ_{s} using the same approach in [38, 42–44].

In a consensus algorithm, an important metric is the average power consumption for a given number of consensus iterations T. This can be defined as $P_{\mathrm{T}}=T^{-1}\sum_{t=0}^{T}\mathrm{E}\left[\|\mathbf{X}(t)\|^2\right]$, where $\mathbf{X}(t)$ is the state vector of the consensus algorithm. We now reiterate the advantages of proposed algorithm in ((4.53) and (4.54)) against the linear consensus algorithm in literature. 1). The proposed approach is robust to a wide range of sensing noise distributions (as long as n_i is symmetric the proposed approach will work). 2). The nonlinear pre-processing of the observations as in (4.52) results in a significant saving of total power compared to the case without pre-processing, and we will illustrate this in the following simulations section.

4.6.1.2 Simulations

In this section, we compare the total power consumption for the approach in ((4.53), (4.54)) with and without pre-processing using Monte Carlo simulations. In all the simulations, we fixed $T=500$ and $\theta=2$. Figure 4.9 shows

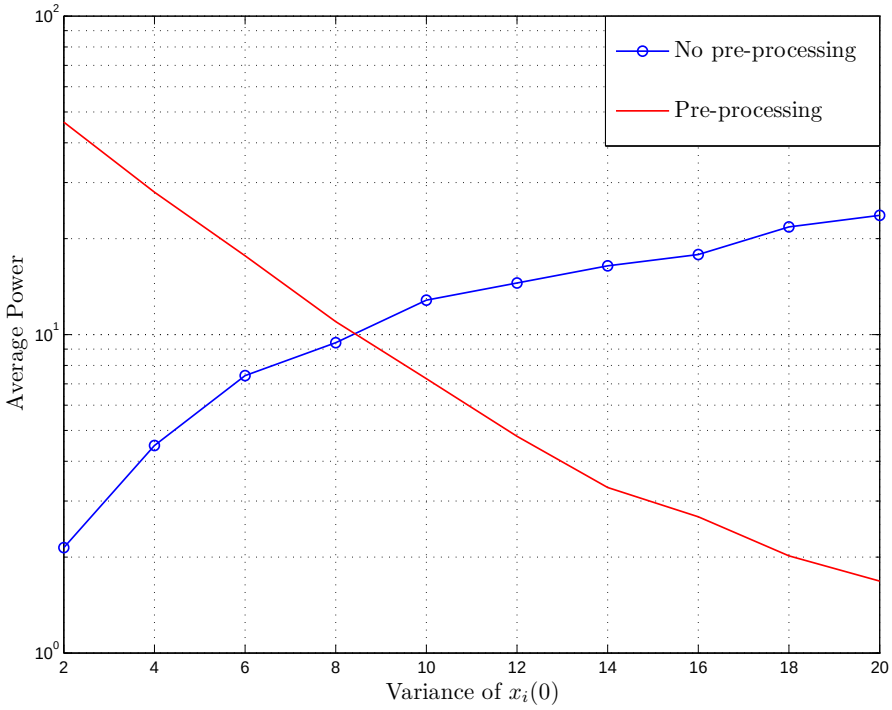

Figure 4.9: Total power versus Variance of $x_i(0)$: n_i is Gaussian, $\theta = 2$, $N = 75$, $h(x) = x$, $\sigma^2 = 0.01$.

that the total power required versus variance of $x_i(0)$ when n_i is Gaussian. Clearly, pre-processing the initial measurements significantly reduces the total transmit power. This is because, the state vector converges to θ^* such that $\mathrm{E}[\theta^*] = e^{2j}\varphi(\sigma)$ and when n_i is Gaussian, we have $\varphi(\sigma) = e^{-\frac{\sigma^2}{2}}$ which is inversely proportional to the variance of $x_i(0)$. Similar trend is observed when n_i is Laplacian as seen in Figure 4.10.

4.6.2 Consensus on Arbitrary Functions

The general problem of consensus on arbitrary functions of initial measurements has been considered in [121, 122]. In [121], the authors obtain necessary and sufficient conditions under which an algorithm would asymptotically achieve consensus on a desired function of initial measurements under some mild smoothness assumptions. Using this result, the authors discuss max and

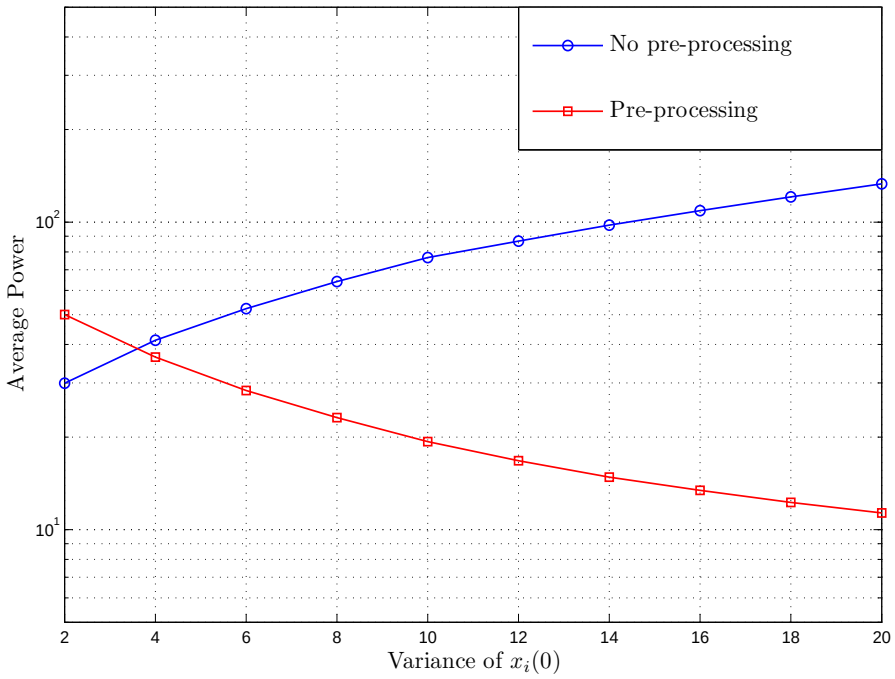

Figure 4.10: Total power versus Variance of $x_i(0)$: n_i is Laplacian, $\theta = 2$, $N = 75$, $h(x) = x$, $\sigma^2 = 0.01$.

min consensus algorithms. The same problem is considered in [122] but in the presence of malicious agents. In both of these papers, the linear iterations is assumed for achieving consensus.

We saw how to achieve consensus on the sample mean using bounded transmissions. If we want to get an estimate of the mean of the square of the initial samples (second moment), all we need to do is to square the measurements (pre-processing) initially and employ the NLC algorithm. The same technique could be employed to get consensus on the moment of any order. An interesting future direction is to design consensus algorithms for arbitrary functions using the NLC algorithm which we have proved to be ideal for resource constrained WSNs.

Chapter 5

Robust Consensus with Receiver Non-Linearity

5.1 Literature Survey and Motivation

In Chapter 4, we saw that the non-linear mapping was done at the transmit node before transmitting the state information and we proved convergence and convergence rate results. In existing literature on consensus (please see [63, 65, 67, 123–129] and references therein) which discuss consensus in the presence of communication noise, the additive noise is assumed to have finite moments. All the existing consensus algorithms will fail if the communication noise does not have finite second moment such as in the case of Cauchy noise. Sensor networks which operate in adverse conditions require algorithms which are robust to impulsive channel noise distributions. Therefore there is a need to develop consensus algorithms which are impervious to impulsive channel noise (like Cauchy) by performing some nonlinear operation at the receiver node. Moreover, consensus with nonlinear combining at the receiver has been considered in [60,113] only in the absence of inter-sensor communication noise. Therefore, it is of interest to solve the problem of distributed consensus with receiver non-linearity when there is communication noise.

In this chapter, we propose a robust nonlinear distributed consensus (RNLC) algorithm which is robust to impulsive communication noise by performing nonlinear processing at the receiver sensor node. We consider a more general class of channel noise distributions than the very restricted class of noise distributions considered in the literature. We do not require the channel noise to have finite moments as is assumed in all the previous work on distributed average consensus algorithms, please see [63, 65, 67, 123–129]. We only need the channel noise to be a zero median symmetric random variable,

i.e., the probability density function (PDF) is symmetric about zero. In addition, we also assume that every sensor maps its state value through a bounded function before transmission so that the transmit power at every node in every iteration is always bounded, making it ideal for resource-constrained WSNs. The analysis of the RNLC algorithm with both the transmit and receiver non-linearities in the presence of channel noise is non-trivial. We prove that all the sensors employing the RNLC algorithm reach consensus to a finite random variable whose mean is the desired sample average. We characterize the asymptotic performance by deriving the asymptotic covariance matrix using results from stochastic approximation theory. Finally we explore the performance of the proposed algorithm employing various functions for the transmit and receiver non-linearities. Different from [63, 65, 67] which also considered consensus in the presence of noisy transmissions, herein we analyse nonlinear processing both at the transmit and receiver nodes and study the asymptotic covariance matrix and its dependence on both the non-linearities. Our work in this chapter also shows an interesting relationship between the Fisher information and the asymptotic covariance matrix which comes out as a result of having nonlinear processing at the receiver sensor node. To the best of our knowledge, we are the first ones in the consensus literature to propose a consensus algorithm which relaxes the requirement of finite moments on the communication noise.

The rest of this chapter is organized as follows. In Section 5.2, we describe the system model and review the previous work on non-linear consensus. We consider the RNLC algorithm in the presence of noise in Section 5.3, and prove that the sensors reach consensus to a random variable. In Sec-

tion 5.4, we present several simulation examples to study the performance of the proposed algorithm.

We will use the same notations used in the previous chapter. We note that we will drop the index of some variables for convenience and the intention will be clear from the context.

5.2 System Model and Previous Work

5.2.1 System Model

Consider a WSN with N sensor nodes each with an initial measurement $x_i(0) \in \mathbb{R}$. Measurements made at the sensor nodes are modeled as

$$x_i(0) = \theta + n_i \ , \quad i = 1, \ldots, N \tag{5.1}$$

where θ is an unknown real-valued parameter and n_i is the sensing noise at the i^{th} sensor. The sample mean of these initial measurements in (5.1) is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i(0) \ . \tag{5.2}$$

Let $\bar{x}$ be the estimate of the parameter θ to be computed by an iterative distributed algorithm, in which each sensor communicates only with its neighbours. If the states of all the sensor nodes converge to $\bar{x}$, then the network is said to have reached *consensus* on the sample average.

5.2.2 Previous Work

A commonly used iterative algorithm for distributed consensus can be written as

$$x_i(t+1) = x_i(t) - \alpha \sum_{j \in \mathbb{N}_i} h(x_i(t) - x_{ij}(t)) \ , \tag{5.3}$$

where $i = 1, \ldots, N$, $t = 0, 1, 2, \ldots$, is the time index, $x_i(t+1)$ is the updated state value of sensor node i at time $t+1$, $\mathbb{N}_i$ is the set of neighbours of sensor

node i, $x_{ij}(t), j \in \mathbb{N}_i$ are the state values of the neighbours of sensor node i at time t, and α is a constant step size. If $h(\cdot)$ is linear, then (5.3) is a linear distributed average-consensus (LDAC) algorithm [58,61,62]. In [58], it is proved that if $0 < \alpha < 2/\lambda_N(\mathbf{L})$, then $x_i(t)$ converges to $\bar{x}$ exponentially and (5.3) is then called as the LDAC algorithm based on the Laplacian heuristic. If $h(\cdot)$ is non-linear then the algorithm belongs to the class of non-linear distributed average-consensus algorithms [60,113]. In [60], the average consensus problem is solved when $h(x)$ in (5.3) is differentiable and odd. In [113], it is illustrated that when $h(x)$ in (5.3) is $\sin(x)$, faster convergence is possible compared to the LDAC algorithm based on the Laplacian heuristic. In all of these cases, $x_{ij}(t)$ has to be transmitted to node i before it can apply the function $h(\cdot)$ to get the new updated state value. Therefore, the transmit peak power in (5.3) is determined by $x_i(t)$ and not necessarily bounded, even if $h(\cdot)$ is bounded. Moreover, there is no communication noise assumed in all the previous work on non-linear consensus.

5.3 Robust Consensus with Impulsive Communication Noise

In this work, we propose a robust distributed nonlinear average consensus algorithm in which every node also performs a nonlinear operation upon the receiving the state values from its neighbours. Every sensor also maps its state value through a bounded function before transmission to constrain the transmit power. Therefore the magnitude of the transmitted signal at every node in every iteration is always bounded making it ideal for resource-constrained WSNs and the receiver non-linearity makes the algorithm robust to a wide range of channel noise distributions.

In this section, we will study the RNLC algorithm with communication noise when sensors exchange information. Our approach is similar to,

but more general than [65] and [98] in that we analyse non-linearity at both the receiver and transmit sensor nodes. Moreover, unlike [65] we study the asymptotic covariance matrix of the state vector and its dependence on the non-linearity. Unlike [113] and [60], we assume receiver non-linearity in the presence of communication noise which brings us the benefit of robustness to a wide range of noise distributions.

5.3.1 The RNLC Algorithm with Communication Noise

Let each sensor map its state value at time t through the function $h(x)$ before transmission, and combines the received state values through a nonlinear function according to the following RNLC algorithm:

$$x_i(t+1) = x_i(t) - \alpha(t) \sum_{j \in \mathbb{N}_i} \left[f\left(h(x_i(t)) - h(x_{ij}(t)) + n_{ij}(t) \right) \right] , \tag{5.4}$$

where $i = 1, \ldots, N, t = 0, 1, 2, \ldots$, is the time index. The value $x_i(t+1)$ is the state update of node i at time $t+1$, $x_{ij}(t)$ is the state value of the j^{th} neighbour of node i at time t and $\alpha(t)$ is a positive step size which will further be assumed to satisfy assumption **(A5)** in the sequel. The node j transmits its information $x_{ij}(t)$ by mapping it through the function $h(x)$, node i receives a noisy version of $h(x_{ij}(t))$ and $n_{ij}(t)$ is the noise associated with the reception of $h(x_{ij}(t))$. The function $f(x)$ is applied at the receiver side to combat the effect of impulsive channel noise and will be further assumed to satisfy **(A2)** in the sequel.

Note that the proposed scheme (5.4) is different from (5.3) in the following aspects. Firstly, in (5.3), $x_{ij}(t)$ has to be transmitted which could exhibit variation over a wide range of values if $x_i(0)$ has a large dynamic range and hence (5.3) does not guarantee bounded transmission power. In contrast, in the proposed scheme the non-linearity is applied before the state value is

transmitted so that the magnitude of the transmitted state value is always constrained within the maximum value of $h(x)$ irrespective of the range of $x_i(t)$ and the realizations of noise $n_{ij}(t)$. Finally, (5.4) involves communication noise while (5.3) does not. Thus the proposed scheme is more suited to resource constrained WSNs when compared to (5.3), and it is practically viable for WSNs operating in adverse conditions.

Our model in (5.4) is more general than the linear consensus algorithm considered in [65] which is a special case when $f(x)$ and $h(x)$ are linear. We make the following assumptions on $f(x)$, $h(x)$, $n_{ij}(t)$, $\alpha(t)$ and the graph:

Assumptions

(A1): The graph $\mathbb{G}$ is connected so that $\lambda_2(\mathbf{L}) > 0$.

(A2): The function $f(x)$ is strictly increasing, odd such that for a given x, the variance $\text{var}[f(x+n)] \leq \sigma^2$ is bounded (we have dropped the indices of the noise samples $n_{ij}(t)$ for convenience).

(A3): The function $h(x)$ is differentiable, and has a bounded derivative such that $0 < h^{'}(x) \leq c$, for some $c > 0$.

(A4) Independent Noise Sequence: The noise samples $n_{ij}(t)$ are mutually independent identically distributed (i.i.d.), symmetric real-valued with zero median (i.e., its PDF symmetric about zero). Note that $n_{ij}(t)$ are not necessarily bounded, and need not have finite moments. We make no distributional assumptions on the communication noise PDF. Note that this assumption is more general than [63, 65, 67, 123–129] in that they require the communication noise to have finite moments whereas we do not require finite mean or variance for the noise distribution.

(A5) Decreasing Weight Sequence: Even though $\text{var}[f(x+n)]$ is bounded, the recursive nature of (5.6) could make the algorithm diverge. In order to

control the variance growth rate of the noise we need the following conditions on the sequence $\alpha(t)$:

$$\alpha(t) > 0 \ , \ \sum_{t=0}^{\infty} \alpha(t) = \infty \ , \ \sum_{t=0}^{\infty} \alpha^2(t) < \infty \ . \tag{5.5}$$

Let $g(x) := \mathrm{E}_n \left[f(x+n) \right]$ where $\mathrm{E}(\cdot)$ denotes the expectation so that $f(x+n) = g(x) + \tilde{n}(x)$. It can be easily proved that $g(x)$ is an odd function under the assumptions **(A2)**, **(A3)** and **(A4)**, and hence $g(0) = 0$. Clearly $\mathrm{E}[\tilde{n}(x)] = 0$ and $\mathrm{E}[\tilde{n}^2(x)] \leq \sigma^2$. Using $g(x)$ the recursion in (5.4) can be written in vector form as

$$\mathbf{X}(t+1) = \mathbf{X}(t) - \alpha(t) \left[\boldsymbol{\mu}(\mathbf{X}(t)) + \mathbf{n}(t, \mathbf{X}(t)) \right] \ . \tag{5.6}$$

where $\mathbf{X}(t)$ is state vector at time t given by $\mathbf{X}(t) = [x_1(t) \ x_2(t) \dots \ x_N(t)]^{\mathrm{T}}$, and $\boldsymbol{\mu} : \mathbb{R}^N \to \mathbb{R}^N$ such that its i^{th} element is given by

$$\boldsymbol{\mu}_i(\mathbf{X}(t)) = \sum_{j \in \mathbb{N}_i} g(h(x_i(t)) - h(x_{ij}(t))) \ , 1 \leq i \leq N \ . \tag{5.7}$$

The vector $\mathbf{n}(t, \mathbf{X}(t))$ captures the additive noise at N nodes contributed by their respective neighbours and their state values and its i^{th} component is given by

$$\mathbf{n}_i(t, \mathbf{X}(t)) = - \sum_{j \in \mathbb{N}_i} \tilde{n}(t, x_i(t), x_{ij}(t)) \ , 1 \leq i \leq N \ . \tag{5.8}$$

Clearly, conditioned on $\mathbf{X}(t)$, the channel noise $\{\tilde{n}(t, x_i(t), x_{ij}(t))\}_{t \geq 0, 1 \leq i,j \leq N}$ is an independent sequence across time and space. It also satisfies

$$\mathrm{E}[\mathbf{n}(t, \mathbf{X}(t))] = \mathbf{0} \ , \forall t \ , \ \ \mu := \sup_t \mathrm{E}[\|\mathbf{n}(t, \mathbf{X}(t))\|^2] \leq N d_{\max} \sigma^2 < \infty. \tag{5.9}$$

Note that (5.9) is because of the fact that the number of neighbours of a given node is upper bounded by $d_{\max}$.

Our primary motivation for considering non-linear processing is two fold: 1). We want to develop a consensus algorithm that is robust to the

impulsive communication noise which do not necessarily have finite moments. We accomplish this by the receiver nonlinear function $f(\cdot)$ as it will be shown in Theorem 5.3.1; 2). We want to impose the realistic assumption of bounded peak per-sensor power by ensuring that the transmit nonlinear function $h(\cdot)$ is bounded.

We will prove convergence and asymptotic normality result of the RNLC algorithm in (5.6) using the same approach used in Chapter 4. Recall the result on the convergence of a discrete time Markov process which will now be used in establishing convergence of the RNLC algorithm in (5.6).

To prove the a.s. convergence of the consensus algorithm in (5.6) using Theorem 4.4.1, we define the consensus subspace $\mathbb{B}$, the set of all vectors whose entries are of equal value as,

$$\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^N | \mathbf{x} = a\mathbf{1} \, , a \in \mathbb{R}\} \, . \tag{5.10}$$

We are now ready to state the main result of Section 5.3.

Theorem 5.3.1. *Let the assumptions **(A1)**, **(A2)**, **(A4)** and **(A5)** hold, and assume $h(x)$ is strictly increasing. Consider the RNLC algorithm in* (5.6) *with the initial state vector $\mathbf{X}(0) \in \mathbb{R}^N$. Then, the state vector $\mathbf{X}(t)$ in* (5.6) *approaches the consensus subspace $\mathbb{B}$ a.s., i.e.,*

$$Pr\left[\lim_{t\to\infty} \inf_{\mathbf{Y}\in\mathbb{B}} \|\mathbf{X}(t) - \mathbf{Y}\| = 0\right] = 1. \tag{5.11}$$

Proof. We will make use of Theorem 4.4.1 to prove (5.11). We will choose an appropriate potential function $V(\mathbf{x})$ that is non-negative which satisfies equation (4.11). We will then prove that the generating operator $\mathcal{L}$ applied

on $V(\mathbf{x})$ as in (4.10) can be upper bounded as in (4.12) with $\gamma(t) = \alpha(t)$, and a $\varphi(\mathbf{x})$ can be found that satisfies (4.13).

First we see that under the assumptions **(A1)**, **(A2)**, **(A4)** and the assumption on $h(x)$, the discrete time vector process $\{\mathbf{X}(t)\}_{t\geq 0}$ is Markov. Define a positive semi-definite matrix $\mathbf{M}$ such that $m_{ij} = -1, i \neq j$, $m_{ii} = N-1$. Let $V(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}$, then the function $V(\mathbf{x})$ is non-negative since $\mathbf{M}$ is a positive semi-definite matrix by construction. Note that $\mathbf{x} \in \mathbb{B}$ is an eigenvector of $\mathbf{M}$ associated with the zero eigenvalue, therefore we have

$$V(\mathbf{x}) = 0, \mathbf{x} \in \mathbb{B} \, . \tag{5.12}$$

Let $\mathbf{x} = \mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp}$ where $\mathbf{x}_{\mathbb{B}}$ is the orthogonal projection of $\mathbf{x}$ on $\mathbb{B}$. When $\mathbf{x} \in \mathbb{B}^{'}$, we have $\|\mathbf{x}_{\mathbb{B}\perp}\| > 0$. Let $\mathbf{x} \in \mathbb{B}^{'}$ and $\boldsymbol{\mu}(\mathbf{x})$ be as defined in (5.6). Then, $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\mu}_{\mathbb{B}}(\mathbf{x}) + \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$, where $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ is non-zero, i.e., $\|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\| > 0$ which is proved now. First we note that $\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x}) > 0$ (please see equations (5.23) through (5.29). This means $(\mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp})\mathbf{M}(\boldsymbol{\mu}_{\mathbb{B}}(\mathbf{x}) + \boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})) = \mathbf{x}_{\mathbb{B}\perp}\mathbf{M}\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) > 0$. If $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ were zero, then $\mathbf{x}_{\mathbb{B}\perp}\mathbf{M}\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) = 0$ which contradicts with the fact that $\mathbf{x}_{\mathbb{B}\perp}\mathbf{M}\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x}) > 0$. Therefore, $\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})$ is non-zero. Define $\beta := \|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2 / \|\mathbf{x}_{\mathbb{B}\perp}\|^2$, then $\beta > 0$, $\mathbf{x} \in \mathbb{B}^{'}$. Therefore, for any $\mathbf{x} \in \mathbb{B}^{'}$,

$$\begin{aligned} V(\mathbf{x}) &= \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} = V(\mathbf{x}_{\mathbb{B}} + \mathbf{x}_{\mathbb{B}\perp}) = V(\mathbf{x}_{\mathbb{B}\perp}) \\ &\geq \min_{\mathbf{x}_{\mathbb{B}\perp} \neq 0} \mathbf{x}_{\mathbb{B}\perp}^{\mathrm{T}}\mathbf{M}\mathbf{x}_{\mathbb{B}\perp} = \lambda_2(\mathbf{M})\|\mathbf{x}_{\mathbb{B}\perp}\|^2 > 0 \, , \end{aligned} \tag{5.13}$$

where the last inequality is due to $\lambda_2(\mathbf{L}) > 0$ by assumption **(A1)**. The equations (5.12) and (5.13) establish that the conditions in (4.11) in Theorem 4.4.1 are satisfied.

Now we will prove that (4.12) is satisfied as well. Towards this end, consider $\mathcal{L}V(\mathbf{x})$ defined in (4.10),

$$\mathcal{L}V(\mathbf{x}) = \mathrm{E}\left[\mathbf{X}(t+1)^{\mathrm{T}}\mathbf{M}\mathbf{X}(t+1)|\mathbf{X}(t)=\mathbf{x}\right] - V(\mathbf{x}) \,, \tag{5.14}$$

$$= \mathrm{E}\left[\left(\mathbf{x}^{\mathrm{T}} - \alpha(t)\left(\boldsymbol{\mu}(\mathbf{x})^{\mathrm{T}} + \mathbf{n}^{\mathrm{T}}(t,\mathbf{x})\right)\right)\cdot \right.$$
$$\left.\left(\mathbf{M}\mathbf{x} - \alpha(t)\left(\mathbf{M}\boldsymbol{\mu}(\mathbf{x}) + \mathbf{M}\mathbf{n}(t,\mathbf{x})\right)\right)\right] - V(\mathbf{x}) \,, \tag{5.15}$$

$$= -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})\right] + \alpha^2(t)\left[\boldsymbol{\mu}(\mathbf{x})^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x}) + \mathrm{E}\left[\mathbf{n}^{\mathrm{T}}(t,\mathbf{x})\mathbf{M}\mathbf{n}(t,\mathbf{x})\right]\right] . \tag{5.16}$$

We get (5.16) by expanding (5.15) and taking the expectations and using the fact that $\mathrm{E}[\mathbf{n}(t)] = \mathbf{0}$. We have

$$\mathrm{E}\left[\mathbf{n}^{\mathrm{T}}(t,\mathbf{x})\mathbf{M}\mathbf{n}(t,\mathbf{x})\right] \leq \mathrm{E}\left[\lambda_N(\mathbf{M})\|\mathbf{n}^{\mathrm{T}}(t,\mathbf{x})\|^2\right] \leq \lambda_N(\mathbf{M})\mu, \tag{5.17}$$

where the second inequality follows from (5.9). Using (5.17) in (5.16), we get the following bound

$$\mathcal{L}V(\mathbf{x}) \leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})\right] + \alpha^2(t)\left[\boldsymbol{\mu}(\mathbf{x})^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x}) + \mu\lambda_N(\mathbf{M})\right] \,, \tag{5.18}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})\right] + \alpha^2(t)\left[\lambda_N(\mathbf{M})\beta\|\mathbf{x}_{\mathbb{B}\perp}\|^2 + \mu\lambda_N(\mathbf{M})\right] \,, \tag{5.19}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})\right] + \alpha^2(t)\left[\beta\frac{\lambda_N(\mathbf{M})}{\lambda_2(\mathbf{M})}\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} + \mu\lambda_N(\mathbf{M})\right] \,, \tag{5.20}$$

$$\leq -2\alpha(t)\left[\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})\right] + m\alpha^2(t)\left[1 + \beta_2\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}\right] \,, \tag{5.21}$$

$$\leq -\alpha(t)\varphi(\mathbf{x}) + m\alpha^2(t)\left[1 + V(\mathbf{x})\right] \,, \tag{5.22}$$

where $\varphi(\mathbf{x}) := 2\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})$, $m := \max\{\beta\lambda_N(\mathbf{M})/\lambda_2(\mathbf{M}), \lambda_N(\mathbf{M})\mu\}$, $\beta_2 := \mu\lambda_N(\mathbf{M})/m$ and $\beta_2 \in (0,1]$. In (5.19), we have used the fact $\boldsymbol{\mu}(\mathbf{x})^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x}) \leq \lambda_N(\mathbf{M})\|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2$ and $\|\boldsymbol{\mu}_{\mathbb{B}\perp}(\mathbf{x})\|^2 = \beta\|\mathbf{x}_{\mathbb{B}\perp}\|^2$. In (5.20), we have used the fact that $\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} \geq \lambda_2(\mathbf{M})\|\mathbf{x}_{\mathbb{B}\perp}\|^2$ due to (5.13). We will now prove that $\varphi(\mathbf{x})$ in (5.22) satisfies equation (4.13) of Theorem 4.4.1.

Whenever $\mathbf{x} \in \mathbb{B}$, i.e., $\mathbf{x} = a\mathbf{1}, a \in \mathbb{R}$, then $x_i = x_j, \forall i, j$, which means $g(h(x_i) - h(x_j)) = 0, \forall i, j$, and hence $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{0}$. This implies that $\varphi(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{B}$.

To prove $\varphi(\mathbf{x}) > 0$ when $\mathbf{x} \in \mathbb{B}'$, consider $\varphi(\mathbf{x}) = 2\mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\mu}(\mathbf{x})$ with $\mathbf{M}$ of dimension $N \times N$,

$$\varphi(\mathbf{x}) = 2\begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_N \end{bmatrix}\begin{bmatrix} N-1 & -1 & -1 & \dots & -1 \\ -1 & N-1 & -1 & \dots & -1 \\ -1 & -1 & N-1 & \dots & -1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -1 & -1 & -1 & \dots & N-1 \end{bmatrix}\boldsymbol{\mu}(\mathbf{x}) \tag{5.23}$$

$$\begin{aligned} = 2\sum_{k=1}^{N-1}\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) - 2\Bigg\{ & \begin{bmatrix} x_2 & x_3 & x_4 & \dots & x_N & x_1 \end{bmatrix} + \\ & \begin{bmatrix} x_3 & x_4 & x_5 & \dots & x_1 & x_2 \end{bmatrix} + \\ & \dots \\ & \begin{bmatrix} x_N & x_1 & x_2 & \dots & x_{N-2} & x_{N-1} \end{bmatrix} + \\ & \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_{N-1} & x_N \end{bmatrix} - \\ & \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_{N-1} & x_N \end{bmatrix} \Bigg\}\boldsymbol{\mu}(\mathbf{x}) \end{aligned} \tag{5.24}$$

$$= 2\left[(N-1)\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) - \left[N(x_1 + x_2 + \ldots + x_N)\mathbf{1}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x})\right] + \mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x})\right] \tag{5.25}$$

$$= 2N\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) \,, \tag{5.26}$$

where we get (5.25) by expanding the terms in the curly braces in (5.24) and simplifying and we have used the fact that $\mathbf{1}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) = 0$ in (5.25) to get (5.26).

Now consider

$$\varphi(\mathbf{x}) = 2N\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) \tag{5.27}$$

$$= 2N\left[\sum_{j\in\mathbb{N}_1} g(h(x_1)-h(x_j))x_1 + \sum_{j\in\mathbb{N}_2} g(h(x_2)-h(x_j))x_2 + \ldots + \sum_{j\in\mathbb{N}_N} g(h(x_N)-h(x_j))x_N\right], \tag{5.28}$$

where (5.28) follows from the symmetric structure of the graph. Note that the i^{th} summation in (5.28) corresponds to the i^{th} node. Now suppose that node i is connected to node j. Then there exists a term $g(h(x_i)-h(x_j))x_i$ in the summation corresponding to the i^{th} node in (5.28), and a term $g(h(x_j)-h(x_i))x_j$ in the summation corresponding to the j^{th} node in (5.28). Both of these terms can be combined as $(x_i-x_j)g(h(x_i)-h(x_j))$ and this corresponds to the edge $\{i,j\}\in\mathbb{E}$. Thus equation (5.28) can be written as pairwise products enumerated over all the edges in the graph as follows

$$\varphi(\mathbf{x}) = 2N\sum_{\{i,j\}\in\mathbb{E}} (x_i-x_j)\, g(h(x_i)-h(x_j))\,. \tag{5.29}$$

Since $\mathbf{x}\in\mathbb{B}'$, $\varphi(\mathbf{x})$ in (5.29) is positive due to the facts that $h(x)$ is strictly increasing and $g(x)$ is strictly an increasing odd function so that there is at least one term in the sum which is strictly greater than zero. Letting $\gamma(t)=\alpha(t), g(t)=\alpha^2(t)$ and by assumption **(A5)**, we see that the sequence $\alpha(t)$ in (5.22) satisfies (4.14). Thus all the conditions of Theorem 4.4.1 are satisfied to yield (5.11). □

Theorem 5.3.1 states that the sample paths of $\mathbf{X}(t)$ approach the consensus subspace almost surely. We note that the assumption **(A3)** is not necessary for Theorem 5.3.1 to hold. Instead we assumed $h(x)$ is strictly increasing (not necessarily differentiable) to prove Theorem 5.3.1. Now, like

in [65], we will prove the convergence of $\mathbf{X}(t)$ to a finite point in $\mathbb{B}$ in Theorem 5.3.2.

Theorem 5.3.2. *Let the assumptions of Theorem 5.3.1 hold. Consider the RNLC algorithm in* (5.6) *with the initial state* $\mathbf{X}(0) \in \mathbb{R}^N$. *Then, there exists a finite real random variable* θ^* *such that*

$$Pr\left[\lim_{t\to\infty} \mathbf{X}(t) = \theta^* \mathbf{1}\right] = 1. \tag{5.30}$$

Proof. Let the average of $\mathbf{X}(t)$ be $\bar{x}(t) = \mathbf{1}^{\mathrm{T}}\mathbf{X}(t)/N$. Since $\mathbf{1}\bar{x}(t) \in \mathbb{B}$, Theorem 5.3.1 implies,

$$\mathrm{P}r\left[\lim_{t\to\infty} \|\mathbf{X}(t) - \bar{x}(t)\mathbf{1}\| = 0\right] = 1\ , \tag{5.31}$$

where (5.31) follows from (5.11) since the infimum in (5.11) is achieved by $\mathbf{Y} = \bar{x}(t)\mathbf{1}$. Pre-multiplying (5.6) by $\mathbf{1}^{\mathrm{T}}/N$ on both sides and noting that $\mathbf{1}^{\mathrm{T}}\boldsymbol{\mu}(\mathbf{x}) = 0$ due to the symmetric structure of the graph we get,

$$\bar{x}(t+1) = \bar{x}(t) - \tilde{v}(t) \tag{5.32}$$

$$= \bar{x}(0) - \sum_{0\le k\le t} \tilde{v}(k) \tag{5.33}$$

where $\tilde{v}(t) = \alpha(t)\mathbf{1}^{\mathrm{T}}\mathbf{n}(t, \mathbf{X}(t))/N$. From (5.9) it follows that

$$\mathrm{E}[\tilde{v}(t)] = 0,$$

$$\sum_{t\ge 0} \mathrm{E}[\tilde{v}(t)]^2 = \sum_{t\ge 0} \frac{\alpha^2(t)}{N^2} \mathrm{E}\|\mathbf{n}(t, \mathbf{X}(t))\|^2 \le \frac{\mu}{N^2} \sum_{t\ge 0} \alpha^2(t) < \infty$$

which implies

$$\mathrm{E}[\bar{x}(t+1)]^2 \le \bar{x}^2(0) + \frac{\mu}{N^2} \sum_{t\ge 0} \alpha^2(t)\ , \forall t\ . \tag{5.34}$$

Equation (5.34) implies that the sequence $\{\bar{x}(t)\}_{t\geq 0}$ is an $\mathcal{L}_2$ bounded martingale [1] and hence converges a.s. and in $\mathcal{L}_2$ to a finite random variable θ^* (see [118, Theorem 2.6.1]). Therefore the theorem follows from (5.31). □

It should be noted that the results in Theorems 5.3.1 and 5.3.2 are similar to the results in [65], but we have proved it for a more general case of which [65] is a special case when $f(x) = x$ and $h(x) = x$, and [98] is a special case of (5.6) when $f(x) = x$. In what follows, we present the properties of the limiting random variable θ^*.

5.3.2 Mean Square Error of RNLC Algorithm

The Theorems 5.3.1 and 5.3.2 establish that the sensors reach consensus asymptotically and converge a.s. to a finite random variable θ^*. We can view θ^* as an estimate of $\bar{x}$. In the following theorem we characterize the unbiasedness and means squared error (MSE) properties of θ^*. We define the MSE of θ^* as $\xi_{_N} = \mathrm{E}[(\theta^* - \bar{x})^2]$.

Theorem 5.3.3. *Let θ^* be the limiting random variable as in Theorem 5.3.2. Then θ^* is unbiased, $E[\theta^*] = \bar{x}$, and its MSE is bounded, $\xi_{_N} \leq \mu N^{-2} \sum_{t\geq 0} \alpha^2(t)$.*

The proof is obtained by following the same steps of the Lemma 5 in [65].

We point out that with non-linear processing, we have obtained a similar bound on the MSE $\xi_{_N}$ as that of the linear consensus algorithm in [65] but in our case the bound depends on the functions $f(x)$ and $h(x)$. It should be noted that $\mu \leq N d_{\max}\sigma^2$ from (5.9) which implies that $\xi_{_N} \leq d_{\max} N^{-1} \sum_{t\geq 0} \alpha^2(t)\sigma^2$.

[1] A sequence of random variables $\{y(t)\}_{t\geq 0}$ is called as a martingale if for all $t \geq 0$, $\mathrm{E}\left[|y(t)|\right] < \infty$ and $\mathrm{E}\left[y(t+1) \mid y(1)\, y(2) \ldots y(t)\right] = y(t)$. The sequence $\{y(t)\}_{t\geq 0}$ is an $\mathcal{L}_2$ bounded martingale if $\sup_t \mathrm{E}\left[y^2(t)\right] < \infty$ (see [119, pp. 110]).

Therefore, if $d_{\max}$ is finite for a large connected network, we have $\lim_{N\to\infty} \xi_N = 0$ and this means that θ^* converges to $\bar{x}$ as the variance of θ^* approaches 0. If the graph is densely connected, then $d_{\max}$ is relatively high which increases the worst-case MSE. On the other hand, when the graph is densely connected, $\lambda_2(\mathbf{L})$ is larger which aids in the speed of convergence to θ^*, as quantified through the covariance matrix in Section 5.3.3.

For any connected graph with N nodes, if $n_{ij}(t) = 0$ then $\lim_{t\to\infty} \mathbf{X}(t) = \bar{x}\mathbf{1}$, which means all the sensor states asymptotically converge to the desired sample average. In fact, in the absence of communication noise, under assumptions **(A1)**, **(A2)** and **(A3)**, we believe that it is possible to prove exponential convergence of $\mathbf{X}(t)$ to $\bar{x}\mathbf{1}$ by letting $\alpha(t) = \alpha$ such that $0 < \alpha < 2/(c\lambda_N(\mathbf{L}))$ and by following a similar approach as in [113].

5.3.3 Asymptotic Normality of RNLC Algorithm

The RNLC algorithm in (5.6) belongs to the class of stochastic approximation algorithms. The convergence speed of these algorithms is an important issue from a practical perspective. There are various criteria for determining the rate of convergence. For instance, one can try to estimate $\mathrm{E}\left[\|\mathbf{X}(t) - \theta^*\mathbf{1}\|^2\right]$ or $\mathrm{P}r\left[\|\mathbf{X}(t) - \theta^*\mathbf{1}\| \leq \epsilon(t)\right]$ [120]. Estimating these parameters may be difficult in practice. However, it is usually possible to establish that $\sqrt{t}(\mathbf{X}(t) - \theta^*\mathbf{1})$ is asymptotically normal with zero mean and some covariance matrix. Asymptotic normality of stochastic approximation algorithms have been established under some general conditions in [118] and for the linear consensus algorithms in [63].

In this section, we establish the asymptotic normality of the RNLC algorithm in (5.6). Our approach here is similar to the one in [63]. Basically, we decompose the RNLC algorithm in $\mathbb{R}^N$ into a scalar recursion and a recursion

in $\mathbb{R}^{(N-1)}$. In this section, for the sake of simplicity we assume that the noise sequence $n_{ij}(t)$ are i.i.d. random variables with zero median and the graph is k regular. We now formally state and prove the result as a theorem.

Theorem 5.3.4. *Let $\alpha(t) = a/(t+1), a > 0$, then the RNLC algorithm in* (5.6) *becomes*

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t+1}\left[-\boldsymbol{\mu}(\mathbf{X}(t)) + \mathbf{n}(t, \mathbf{X}(t))\right]. \tag{5.35}$$

*Suppose that the assumptions **(A1)**, **(A3)**, **(A4)** and **(A5)** hold, the function $f(x)$ is differentiable and suppose also that graph is a k regular connected graph. Let the EVD of $\mathbf{L}$ be given by $\mathbf{L} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{U}$ is a unitary matrix whose columns are the eigenvectors of $\mathbf{L}$ such that*

$$\mathbf{U} = \left[\frac{\mathbf{1}}{\sqrt{N}} \;\; \boldsymbol{\Phi}\right], \boldsymbol{\Phi} \in \mathbb{R}^{N\times(N-1)} \;,\; -\boldsymbol{\Sigma} = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \;, \tag{5.36}$$

where $\mathbf{B} \in \mathbb{R}^{(N-1)\times(N-1)}$ is a diagonal matrix containing the $N-1$ negative eigenvalues of $-\mathbf{L}$ (this means that $\mathbf{B}$ is a stable matrix). In addition, let θ_0 be a realization of the random variable θ^* and $2a\lambda_2(\mathbf{L})g^{'}(0)h^{'}(\theta_0) > 1$ so that the matrix $\left[ag^{'}(0)h^{'}(\theta_0)\mathbf{B} + \mathbf{I}/2\right], \theta_0 \in \mathbb{R}$ is stable. Define $[\tilde{n}(t) \;\; \tilde{\mathbf{n}}(t)^T]^T := N^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{n}(t, \mathbf{X}(t))$, $\tilde{\mathbf{n}}(t) \in \mathbb{R}^{(N-1)}$, so that $\tilde{n}(t) = N^{-1}\mathbf{1}^{\mathrm{T}}\mathbf{n}(t, \mathbf{X}(t))$ and $\tilde{\mathbf{n}}(t) = N^{-1/2}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{n}(t, \mathbf{X}(t))$. Let $\mathbf{C} = E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T]$, $\mathbf{C} \in \mathbb{R}^{(N-1)\times(N-1)}$. Then, as $t \to \infty$,

$$\sqrt{t}(\mathbf{X}(t) - \theta^*\mathbf{1}|\theta^* = \theta_0) \sim \mathcal{N}\left(0, N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}} + N^{-1}\boldsymbol{\Phi}\mathbf{S}^{\theta_0}\boldsymbol{\Phi}^{\mathrm{T}}\right) \;, \tag{5.37}$$

where $\sigma_v^2 = kE[f^2(n)]$ and

$$\mathbf{S}^{\theta_0} = a^2 \int_0^\infty e^{\left[ag^{'}(0)h^{'}(\theta_0)\mathbf{B} + \frac{\mathbf{I}}{2}\right]t}\, \mathbf{C}\, e^{\left[ag^{'}(0)h^{'}(\theta_0)\mathbf{B} + \frac{\mathbf{I}}{2}\right]t} dt \;. \tag{5.38}$$

Proof. Define $[\tilde{x}(t) \;\; \tilde{\mathbf{X}}(t)^{\mathrm{T}}]^{\mathrm{T}} := N^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{X}(t), \tilde{\mathbf{X}}(t) \in \mathbb{R}^{(N-1)}$. From Theorem 5.3.2, we have $\mathbf{X}(t) \to \theta^*\mathbf{1}$ a.s. as $t \to \infty$ which implies that $[\tilde{x}(t) \;\; \tilde{\mathbf{X}}(t)]^{\mathrm{T}} \to [\theta^* \;\; \mathbf{0}]^{\mathrm{T}}$ a.s. as $t \to \infty$, and therefore $\tilde{\mathbf{X}}(t) \to \mathbf{0}$ a.s. as $t \to \infty$. The error $[\mathbf{X}(t) - \theta_0\mathbf{1}]$ can be written as the sum of two error components (see also Section VI in [63]) as given below

$$[\mathbf{X}(t) - \theta_0\mathbf{1}] = [\tilde{x}(t) - \theta_0]\mathbf{1} + \frac{1}{\sqrt{N}}\mathbf{\Phi}\tilde{\mathbf{X}}(t) \;, \tag{5.39}$$

$$= \mathbf{e}_1 + \mathbf{e}_2 \;, \tag{5.40}$$

where $\mathbf{e}_1 = [\tilde{x}(t) - \theta_0]\mathbf{1}$ and $\mathbf{e}_2 = N^{-1/2}\mathbf{\Phi}\tilde{\mathbf{X}}(t)$. By calculating the covariance matrix between $\mathbf{e}_1$ and $\mathbf{e}_2$, it can be proved that they are asymptotically uncorrelated as $t \to \infty$, and that asymptotically $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}})$ (see Theorem 12 in [63]). To show that $\sqrt{t}\mathbf{e}_2$ is asymptotically normal, it suffices to show that $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal. To this end, express $\boldsymbol{\mu}(\mathbf{x})$ in (5.35) around $\mathbf{x} = \theta_0\mathbf{1}$ using Taylor's series expansion,

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\mu}(\theta_0\mathbf{1}) + \left.\frac{\partial\boldsymbol{\mu}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\theta_0\mathbf{1}} + o(\|\mathbf{x} - \theta_0\mathbf{1}\|) \;, \mathrm{a}s\; \mathbf{x} \to \theta_0\mathbf{1} \;, \tag{5.41}$$

$$= g^{'}(0)h^{'}(\theta_0)\mathbf{L}\mathbf{x} + o(\|\mathbf{x} - \theta_0\mathbf{1}\|) \;, \mathrm{a}s\; \mathbf{x} \to \theta_0\mathbf{1} \;. \tag{5.42}$$

Using (5.42) in (5.35) we get, as $t \to \infty$,

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \frac{a}{t+1}\left[g^{'}(0)h^{'}(\theta_0)\left(-\mathbf{L}\mathbf{X}(t)\right) + \delta(\mathbf{X}(t)) + \mathbf{n}(t, \mathbf{X}(t))\right] \;, \tag{5.43}$$

where $\|\delta(\mathbf{X}(t))\| \to 0$ as $t \to \infty$. Pre-multiplying (5.43) on both sides by $N^{-1/2}\mathbf{U}^{\mathrm{T}}$ and using (5.36) we get the following recursions

$$\tilde{x}(t+1) = \tilde{x}(t) + \frac{a}{t+1}\tilde{n}(t)\ , \tag{5.44}$$

$$\tilde{\mathbf{X}}(t+1) = \tilde{\mathbf{X}}(t) + \frac{a}{t+1}\left[g^{'}(0)h^{'}(\theta_0)\mathbf{B}\tilde{\mathbf{X}}(t) + \tilde{\delta}(\mathbf{X}(t)) + \tilde{\mathbf{n}}(t)\right]\ , \mathrm{as}\ t \to \infty\ , \tag{5.45}$$

where $\tilde{\delta}(\mathbf{X}(t)) = N^{-1/2}\mathbf{\Phi}^{\mathrm{T}}\delta(\mathbf{X}(t))$. With the assumption $\left[ag^{'}(0)h^{'}(\theta_0)\mathbf{B} + \mathbf{I}/2\right]$, $\theta_0 \in \mathbb{R}$ is a stable matrix, it can be verified that all the conditions of Theorem 6.6.1 in [118, p. 147] are satisfied for the process $\tilde{\mathbf{X}}(t)$ in (5.45). Therefore, for a given θ_0, the process $\sqrt{t}\tilde{\mathbf{X}}(t)$ is asymptotically normal with zero mean and covariance matrix given by (5.38). Since $\sqrt{t}\mathbf{e}_1 \sim \mathcal{N}(0, N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}})$ and using (5.38) together with the fact that $\mathbf{e}_1$ and $\mathbf{e}_2$ are asymptotically independent as $t \to \infty$, we get (5.37) which completes the proof. □

Equation (5.37) indicates how fast the process $\sqrt{t}(\mathbf{X}(t) - \theta_0\mathbf{1})$ will converge to $\theta_0\mathbf{1}$ for a given θ_0 as $t \to \infty$. The convergence speed clearly depends on $g^{'}(0)$ and $h^{'}(\theta_0)$ which captures the effect of receiver and transmit non-linearities respectively.

Let the asymptotic covariance in (5.37) be denoted by $\mathbf{C}_{\mathrm{rnlc}}$. Since $\mathbf{n}(t, \mathbf{X}(t))$ are asymptotically i.i.d., $\mathbf{C}$ in (5.38) becomes $\mathbf{C} = \sigma_v^2\mathbf{I}$ and thus we have $\mathbf{C}_{\mathrm{rnlc}} = N^{-1}a^2\sigma_v^2\mathbf{1}\mathbf{1}^{\mathrm{T}} + N^{-1}\mathbf{\Phi}\mathbf{S}^{\theta_0}\mathbf{\Phi}^{\mathrm{T}}$ where $\mathbf{S}^{\theta_0}$ is a diagonal matrix whose diagonal elements are given by $\mathbf{S}_{ii}^{\theta_0} = a^2\sigma_v^2/[2ag^{'}(0)h^{'}(\theta_0)\lambda_{i+1}(\mathbf{L}) - 1]$. A reasonable quantitative measure of largeness [120] of the asymptotic covariance matrix is $\|\mathbf{C}_{\mathrm{rnlc}}\|$ which is the maximum eigenvalue of the symmetric matrix $\mathbf{C}_{\mathrm{rnlc}}$. Further, $\|\mathbf{C}_{\mathrm{rnlc}}\|$ can be minimized with respect to the parameter a. This can be formulated as the following optimization problem,

$$\min_{\{a|2ag^{'}(0)h^{'}(\theta_0)\lambda_2(\mathbf{L})>1\}} \max_{\{\mathbf{x}|\mathbf{x}\in\mathbb{R}^N, \|\mathbf{x}\|^2\leq 1\}} \mathbf{x}^{\mathrm{T}}\mathbf{C}_{\mathrm{rnlc}}\mathbf{x}\ , \tag{5.46}$$

which can be solved analytically by using the KKT conditions [90]. The value of a that optimizes (5.46) is $a^*_{nlc} = (N+1)/[2N\lambda_2(\mathbf{L})g^{'}(0)h^{'}(\theta_0)]$ and the corresponding optimal value of the $\|\mathbf{C}_{\text{rnlc}}\|$is given by

$$\|\mathbf{C}^*_{\text{rnlc}}\| = k\left(\frac{N+1}{2N}\right)^2\left(\frac{\mathrm{E}[f^2(n)]}{(\mathrm{E}[f^{'}(n)])^2}\right)\left(\frac{1}{\lambda_2^2(\mathbf{L})}\right)\left(\frac{1}{h^{'}(\theta_0)}\right)^2 . \tag{5.47}$$

The size of the asymptotic covariance matrix in (5.47) is inversely proportional to the square of the smallest non-zero eigenvalue $\lambda_2(\mathbf{L})$ which quantifies how densely a graph is connected. We also note that (5.47) depends on the receiver nonlinear function used.

Equation (5.47) gives some useful insights to design the transmission function $h(x)$ as discussed in [98] and all the conclusions drawn in [98] are applicable here as well. For a fixed $f(x)$, if we choose two functions $h_1(x)$ and $h_2(x)$ such that $h_1^{'}(x) > h_2^{'}(x), \forall x \in \mathbb{R}$, it is easy to see from (5.47) that $\|\mathbf{C}^*_{\text{rnlc}1}\| < \|\mathbf{C}^*_{\text{rnlc}2}\|, \forall\theta_0 \in \mathbb{R}$. This means that the convergence will be faster when $h_1(x)$ is employed in the RNLC algorithm (5.6) than when $h_2(x)$ is employed.

When $f(x)$ is a bounded function, from equation (8) in [84] we have

$$\frac{\mathrm{E}[f^2(n)]}{(\mathrm{E}[f^{'}(n)])^2} \geq \frac{1}{J} , \tag{5.48}$$

where J is the Fisher information of n with respect to a location parameter [85, eqn (8)] and thus we see an interesting relationship between the maximum eigenvalue of the asymptotic covariance and the Fisher information. For a given $h(x)$, the best choice of $f(x)$ is the one that achieves equality in (5.48). For instance, when n is Gaussian, $f(x) = x$ achieves equality in (5.48) in which case we have var$[f(n)]$ equals the inverse of Fisher information. In addition, when n has finite moments, if we let $f(x) = x$, we get the same result as in [98], and together with this if we also let $h(x) = x$, we get the results for

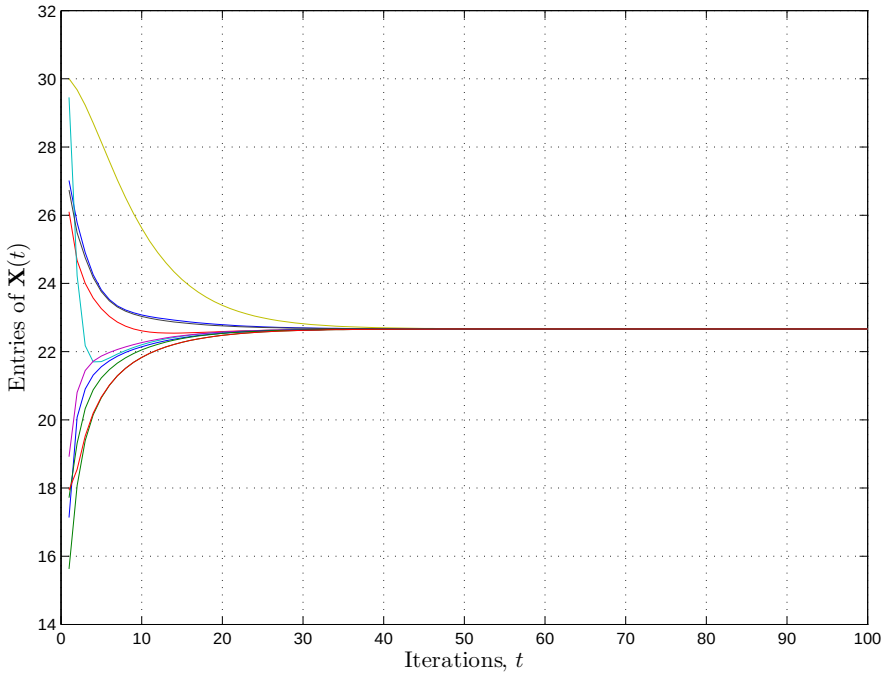

Figure 5.1: Entries of $\mathbf{X}(t)$ versus Iterations t: $\alpha = 100$, $N = 15$, $h(x) = \tanh(0.05x)$, $f(x) = \frac{2}{\pi}\tan^{-1}(\frac{\pi}{2}0.05x)$, $\bar{x} = 22.67$.

the linear case as in Theorem 12 of [63]. Equation (5.47) also indicates when $h(x)$ is fixed, scaling $f(x)$ does not improve the speed of convergence. We will illustrate these findings in the simulations in Section 5.4.

5.4 Simulations

In this section, we corroborate our analytical findings through various simulations. In all the simulations presented, the initial samples $x_i(0) \in \mathbb{R}, i = 1, 2, \ldots, N$, were generated randomly using Gaussian distribution with a standard deviation equal to 10. The desired global average value is indicated in each of the simulations. We focus here on bounded functions for both the transmit and receiver non-linearities to study their performance. Please note that our results are valid for a broader class of increasing functions (see Section 5.3.1) than the ones considered in this section.

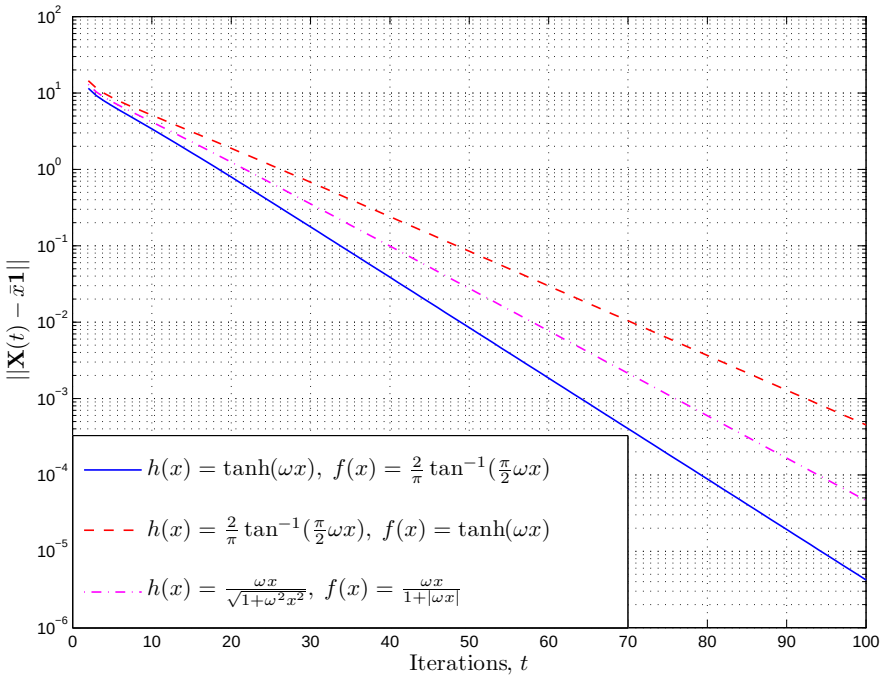

Figure 5.2: Evolution of error $||\mathbf{X}(t) - \bar{x}\mathbf{1}||$ versus Iterations t: $\alpha = 100$, $\omega = 0.05$, $N = 15$, $\bar{x} = 22.67$.

5.4.1 Performance of RNLC Algorithm Without Channel Noise

Our focus in this paper is on non-linear processing in the presence of noise. However, we would also like to illustrate the convergence behavior on the absence of noise. Figures 5.1 and 5.2 depict the performance of the proposed RNLC algorithm in the absence of channel noise for a large network with $N = 15$. In all the cases, we have used α values such that $0 < \alpha < 2/(c\lambda_N(\mathbf{L}))$ as mentioned in Section 5.3.2. From Figure 5.1, we infer that in about 40 iterations, all the nodes reach consensus on the desired global average of $\bar{x} = 22.67$. Figure 5.2 shows evolution of error norm $||\mathbf{X}(t) - \bar{x}\mathbf{1}||$ for various bounded functions. We see that the convergence is exponential in all cases as noted in Section 5.3.2.

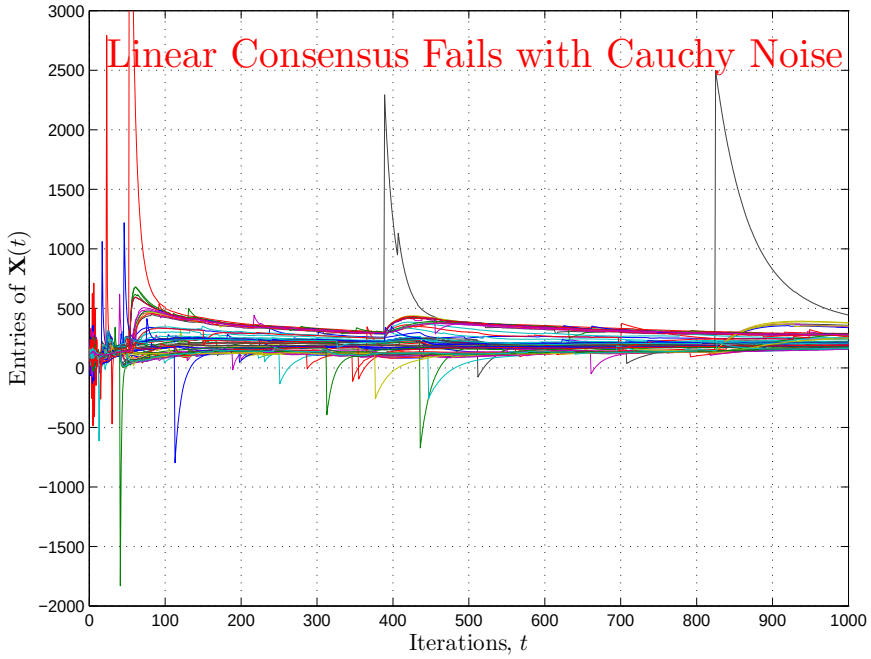

Figure 5.3: Entries of $\mathbf{X}(t)$ versus Iterations t: Cauchy noise, $h(x) = x, f(x) = x$, $N = 75$, $\bar{x} = 134.31$, $\gamma = 1$.

5.4.2 Performance of RNLC Algorithm with Channel Noise

First, we highlight that the linear consensus algorithms in [63,65,67,123–129] fail to achieve consensus when the channel noise does not have finite variance. An example plot is shown in Figure 5.3 for the case when the channel noise is Cauchy distributed with $\gamma = 1$. Clearly, the sensors do not reach consensus. Whereas the proposed RNLC algorithm will work when we choose $f(x)$ as a nonlinear function as shown next.

Figures 5.4 - 5.10 illustrate the performance of RNLC algorithm in the presence of communication noise. As explained in the assumption **(A5)** in Section 5.3.1, we chose the decreasing step sequence to be $\alpha(t) = 1/(t + 1), t \geq 0$, in all simulations. Here we assumed that $\rho = \max_x h^2(x)$ is the maximum power available at each sensor to transmit its state value. The

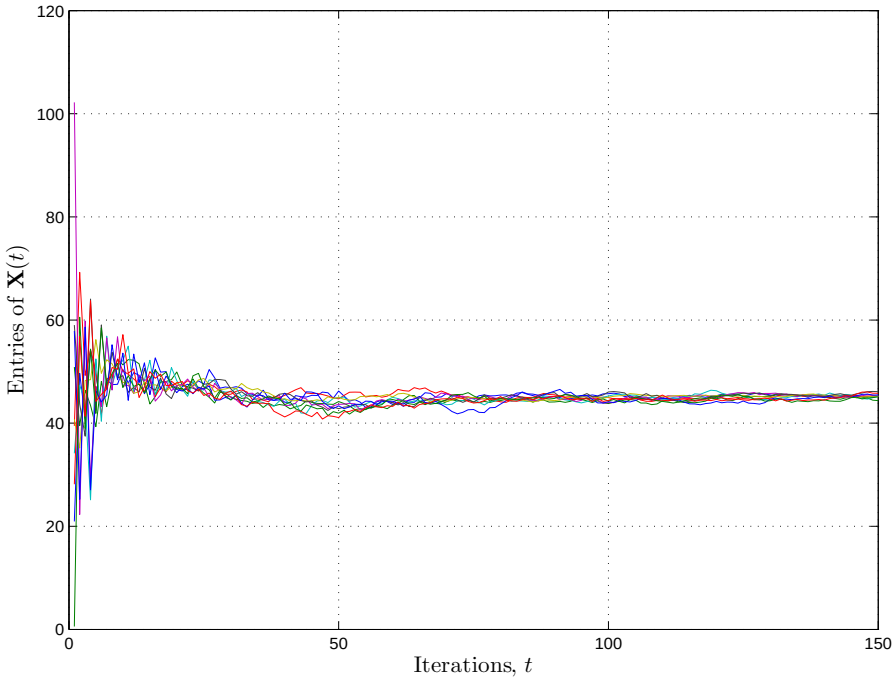

Figure 5.4: Entries of $\mathbf{X}(t)$ versus t: Cauchy noise, $h(x) = \sqrt{\rho}\frac{2}{\pi}\tan^{-1}(\frac{\pi}{2}0.01x)$, $f(x) = \tanh(5x)$, $N = 10$, $\bar{x} = 43.96$, $\rho = 15$ dB, $\gamma = 0.1$.

receiver nonlinear function $f(x)$ is indicated in each case. Figure 5.4 shows that the nodes employing the RNLC algorithm reach consensus for a small network with $N = 10$ in about 100 iterations and Figure 5.5 shows convergence for a large network with $N = 75$ in about 40 iterations. Note that the transmit power $h^2(x_i(t)), i = 1, 2, \ldots, N$, is always constrained within the upper bound of ρ (not shown) making the proposed scheme practically viable for the power constrained WSNs.

In Figures 5.6, 5.7, 5.8 and 5.9 we show the convergence speed performance of the proposed RNLC algorithm by plotting $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus iterations t. These plots indicate how fast the mean of the process $\mathbf{X}(t)$ converges towards the desired global mean vector $\bar{x}\mathbf{1}$.

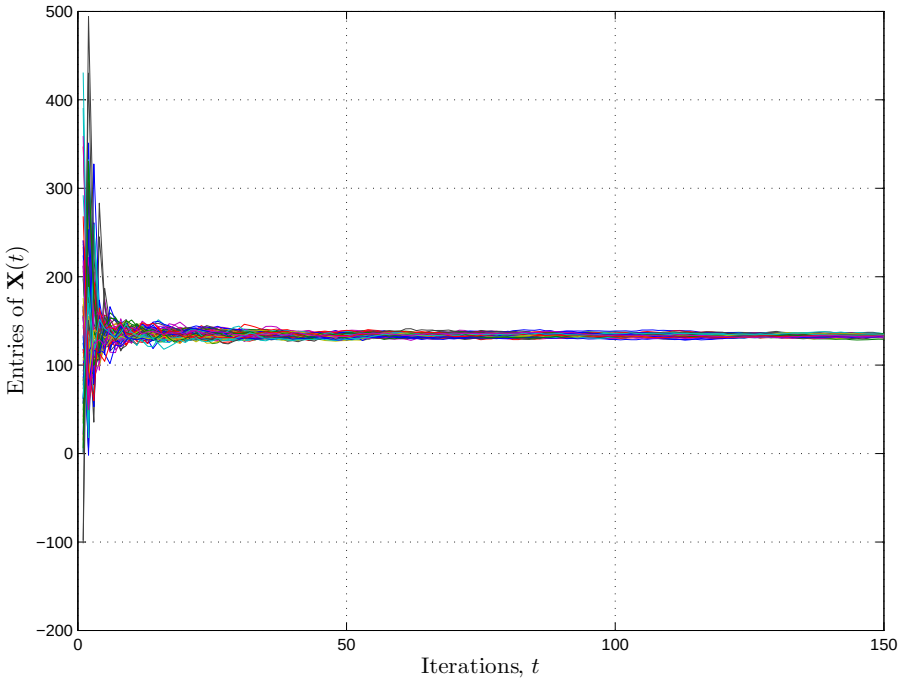

Figure 5.5: Entries of $\mathbf{X}(t)$ versus t: Cauchy noise, $h(x) = \sqrt{\rho}\frac{2}{\pi}\tan^{-1}(\frac{\pi}{2}0.01x)$, $f(x) = \tanh(5x)$, $N = 75$, $\bar{x} = 134.31$, $\rho = 5$ dB, $\gamma = 0.1$.

In Theorem 5.3.4, we saw that for a fixed $f(x)$, if two functions $h_1(x)$ and $h_2(x)$ such that $h_1^{'}(x) > h_2^{'}(x), \forall x \in \mathbb{R}$, are employed in the RNLC algorithm then the convergence will be faster for $h_1(x)$ compared to that of $h_2(x)$. This is illustrated in Figure 5.6 where we have chosen $f(x) = \frac{1.5x}{1+|1.5x|}$ and $h_1(x) = \sqrt{\rho}\tan^{-1}(\omega x)$, $h_2(x) = \sqrt{\rho}\frac{2}{\pi}\tan^{-1}(\frac{\pi}{2}\omega x)$ and $h_3(x) = \sqrt{\rho}\frac{\omega x}{\sqrt{1+\omega^2 x^2}}$. The performance gain of $h_1(x)$ obtained over $h_2(x)$ and $h_3(x)$ can be understood intuitively by observing that on an average the transmit power will be more when $h_1(x)$ is employed than when $h_2(x)$ or $h_3(x)$ is employed. The speed of convergence for two graphs with different algebraic connectivity is illustrated in Figure 5.7. We see that the graph with smaller connectivity (smaller $\lambda_2(\mathbf{L})$) converges slower than the one with large connectivity as dictated by (5.47). In Theorem 5.3.4, we also saw that scaling $f(x)$ does not improve the asymptotic convergence speed. This is shown in Figure 5.8 where

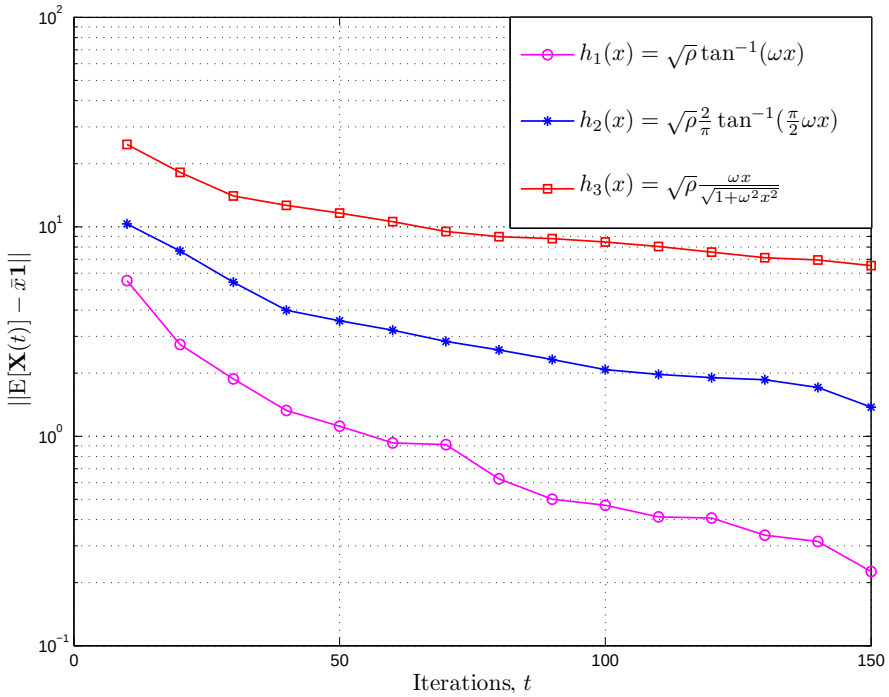

Figure 5.6: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: $f(x) = \frac{1.5x}{1+|1.5x|}$, $\omega = 0.01$, $N = 75$, $\bar{x} = 134.31$, $\rho = 10$ dB, $\gamma = 0.413$.

we see that when the iterations are large ($t > 120$), the speed of convergence of all the three functions are nearly the same. In Figure 5.9, we depict the robustness of the RNLC algorithm for various channel noise distributions. We observe that the performance is nearly the same for Gaussian and Laplacian distributions, whereas there is a significant gap between Cauchy and alpha-stable distributions considered in this simulation. The latter effect is due to the fact that $\mathrm{E}[f^2(n)]/(\mathrm{E}[f^{'}(n)])^2$ is significantly different for those two cases which justifies the performance gap. Finally, we illustrate the difference between the variance of θ^* and the asymptotic variance in Figure 5.10. Here we consider the evolution of the state value $x_1(t)$ of the first node for several consensus runs for the same initial conditions. Recall that in every consensus run the state value $x_1(t)$ converges to an instance of the limiting random variable θ^* and the variation among these several realizations is characterized by the

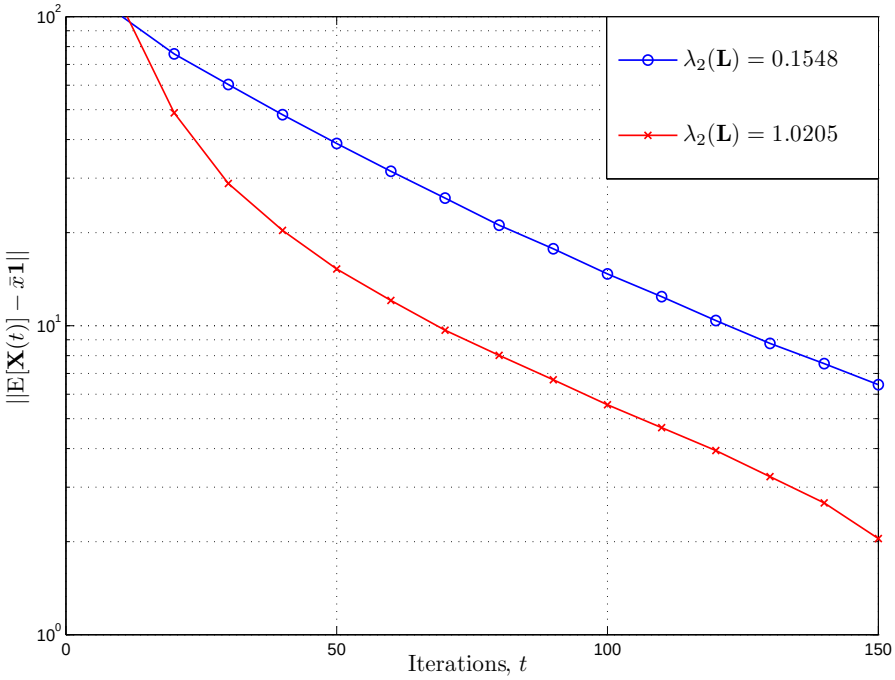

Figure 5.7: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: Cauchy noise, $h(x) = x$, $f(x) = \frac{1.5x}{1+|1.5x|}$, $N = 75$, $\bar{x} = 84.31$, $\gamma = 0.413$.

variance of θ^*. In contrast, how fast the state value $x_1(t)$ converges to the limiting value θ_0 is characterized by the asymptotic variance of $\sqrt{t}[x_1(t) - \theta_0]$ as $t \to \infty$.

Comments on the Sensing Model

We want to point out that the sensing model in Sections 4.3 and 5.2 are theoretical models, where the sensing measurement error is modeled as an additive noise with the true value of the parameter θ. However, the actual sensing characteristics is usually nonlinear in θ [130]. This type of nonlinear characteristic could be modeled as

$$x_i(0) = s(\theta) + n_i \ , \quad i = 1, \ldots, N \tag{5.49}$$

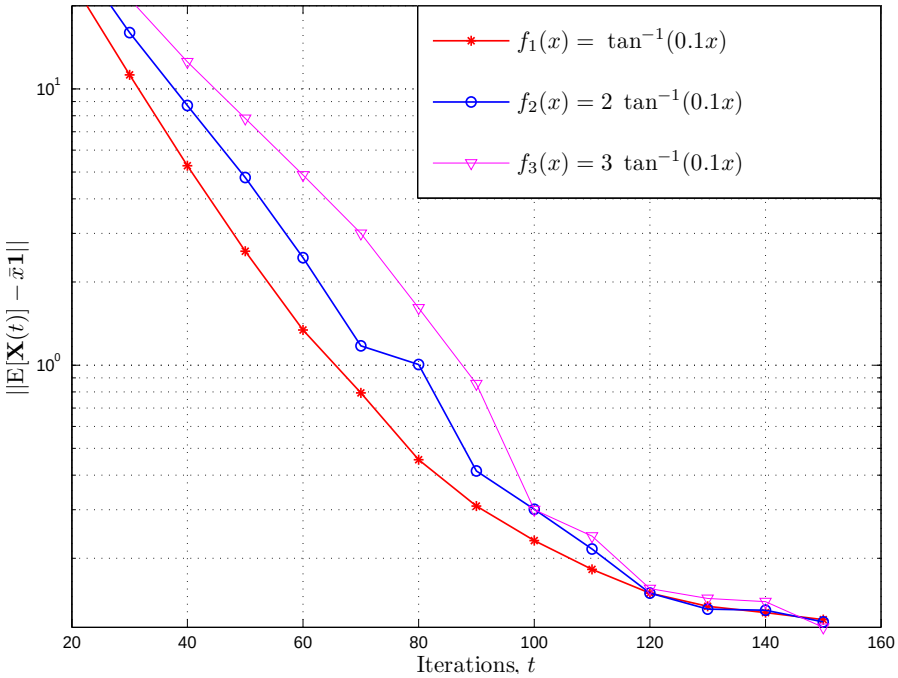

Figure 5.8: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: $h(x) = x$, $\omega = 1.5$, $N = 10$, $\bar{x} = 34.31$, $\gamma = 0.413$.

where $s(\theta)$ captures the inherent non-linearity of the sensor. The NLC and RNLC consensus algorithms discussed in Chapters 4 and 5 still would work without any change except that the consensus value will now depend on $s(\theta)$.

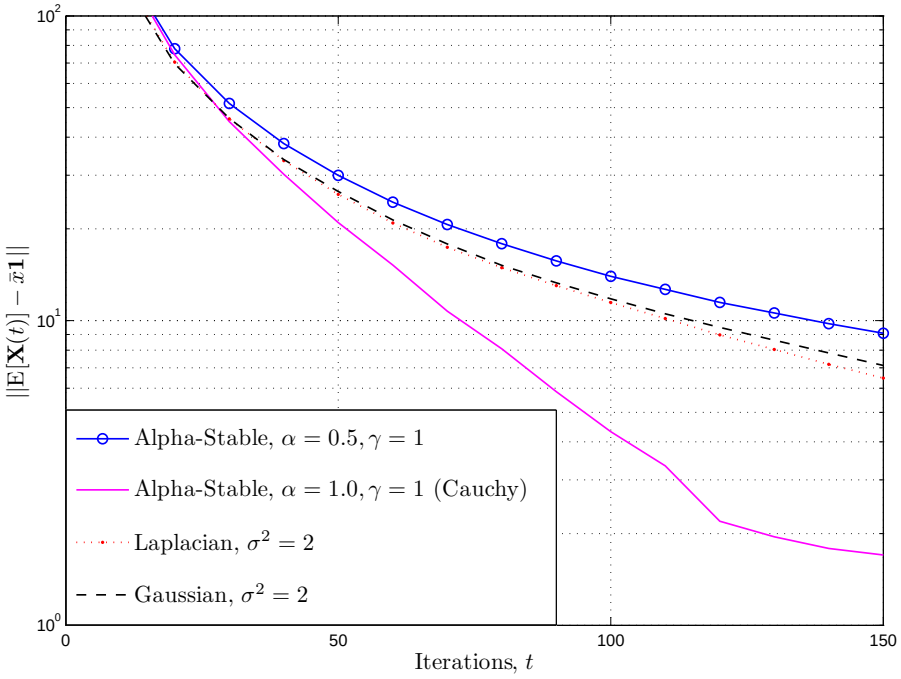

Figure 5.9: $||\mathrm{E}[\mathbf{X}(t)] - \bar{x}\mathbf{1}||$ versus Iterations t: Robustness to various noise distributions, $h(x) = x$, $f(x) = \tanh(2x)$, $N = 75$, $\bar{x} = 124.31$

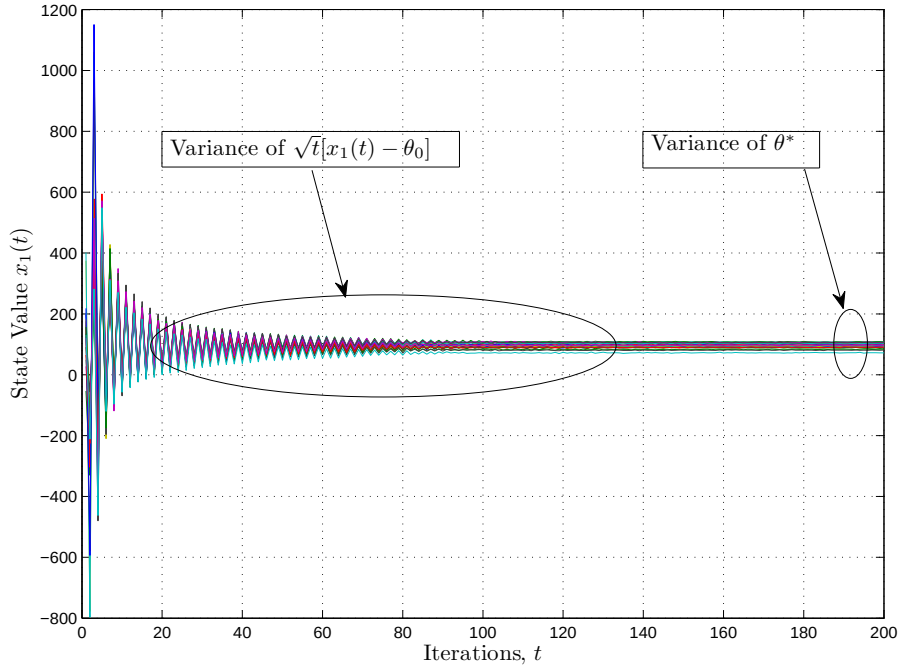

Figure 5.10: Entries of $\mathbf{X}(t)$ versus Iterations t: Difference between Variance of θ^* and Asymptotic Variance, $h(x) = x$, $f(x) = 3\tan^{-1}(0.05x)$, $N = 75$, $\bar{x} = 94.31$, $\gamma = 0.413$

Chapter 6

Conclusions

In Chapter 2 of this dissertation, a distributed detection scheme relying on constant modulus transmissions from the sensors is proposed over a Gaussian multiple access channel. The instantaneous transmit power does not depend on the random sensing noise, which is a desirable feature for low-power sensors with limited peak power capabilities. The DC of the proposed scheme is shown to depend on the characteristic function of the sensing noise and optimized with respect to ω for various sensing noise distributions. In addition to the desirable constant-power feature, the proposed detector is robust to impulsive noise, and performs well even when the moments of the sensing noise do not exist as in the case of the Cauchy distribution. Extensions to non-homogeneous sensors with non-identically distributed noise are also considered. It is shown that over Gaussian multiple access channels, the proposed detector outperforms AF, DF and MDF schemes consistently, and the MAF scheme when the sensing SNR is greater than 4 dB. The proposed detector is shown to work with the non-Gaussian channel noises as well. The error exponent is also derived for the proposed scheme and large deviation theory is used to approximate $P_{\mathrm{e}}(\omega)$ for large L. It is shown that while the DC has a simpler expression for the purpose of optimizing ω, the probability of error approximation based on (2.40) is shown to be an accurate indicator of detection performance for all distributions and moderate number of sensors. The effect of fading is also considered, and shown to be detrimental to the detection performance.

In Chapter 3, the more general problem of a distributed inference scheme using bounded transmissions from the sensors over Gaussian multi-

ple access channels is considered. A distributed inference scheme relying on bounded transmissions from the sensors is considered over Gaussian multiple access channels. The instantaneous transmit power is always constrained to be bounded irrespective of the random sensing noise, which is a desirable feature for low-power sensors with limited peak power capabilities. For the distributed estimation problem, the estimation scheme using bounded transmissions is shown to be strongly consistent provided that the variance of the noise samples are bounded and that the transmission function is one-to-one. For sensing noise distributions for which the sample mean is highly suboptimal or inconsistent, the proposed estimator is shown to be consistent. For heavy-tailed distributions with infinite variance like Cauchy, it is shown that the AF scheme fails, and that the proposed approach is superior to AF. As long as the variance of the noise samples grow to infinity slower than linearly, AF scheme is consistent, whereas the proposed scheme fails when the variance of the noise samples go to infinity at any rate. For the distributed detection problem, the regimes under which reliable detection is possible or impossible are also established. Monte Carlo simulations are presented to illustrate the performance of several bounded transmission functions for a variety of sensing noise distributions.

In Chapter 4 of this dissertation, a distributed consensus algorithm in which every sensor maps its state value through a bounded function before transmission to constrain the transmit power is proposed. The transmitted signal power at every node in every iteration is always bounded irrespective of the state value or the communication noise, which is a desirable feature for low-power sensors with limited peak power capabilities. In the presence of communication noise, it is proved using the theory of Markov processes

that the sensors reach consensus asymptotically on a finite random variable whose expectation contains the desired sample average of the initial sensor measurements, and whose mean-squared error is bounded. The asymptotic convergence speed of the proposed algorithm is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. While the proposed NLC algorithm has the desirable feature of bounded transmit power, it is shown that using the best case NLC algorithm results in larger asymptotic covariance compared to the best case linear consensus algorithm. In the absence of communication noise, it is illustrated that the network achieves consensus on the global sample average exponentially fast provided the step size is chosen appropriately and that by adjusting the step size, it is possible to achieve the same speed of convergence as that of the best case linear consensus algorithm using Laplacian heuristic.

Finally, in Chapter 5, a distributed average consensus algorithm in which every sensor performs a nonlinear processing at the receiver is proposed. Every sensor also maps its state value through a bounded function before transmission to constrain the transmit power. It is shown that nonlinearity at the receiver nodes makes the algorithm robust to a wide range of channel noise distributions including the impulsive ones. The proposed algorithm does not need the requirement of finite moments on the communication noise and thus it is proved to be not only more general than the existing consensus algorithms but is practically viable for WSNs deployed in adverse conditions. It is proved using the theory of Markov processes that the sensors reach consensus asymptotically on a finite random variable whose expectation contains the desired sample average of the initial sensor measurements, and whose mean-squared error is bounded. The asymptotic convergence speed of

the proposed algorithm is characterized by deriving the asymptotic covariance matrix using results from stochastic approximation theory. It is shown that scaling the receiver nonlinear function does not affect the convergence speed of the proposed algorithm and its robustness to a variety of channel noise distributions is highlighted. An interesting relationship between the Fisher information and the asymptotic covariance matrix is also shown. In the absence of communication noise, it is illustrated that the network achieves consensus on the global sample average exponentially fast provided the step size is chosen appropriately.

REFERENCES

[1] B. Sadler, "Fundamentals of energy-constrained sensor network systems," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 20, no. 8, pp. 17–35, Aug. 2005.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communication Magazine*, pp. 102–114, August 2002.

[3] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, p. pp. 5158, May 2000.

[4] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister, "Smart dust: communicating with a cubic-millimeter computer," *Computer*, vol. 34, no. 1, pp. 44–51, January 2001.

[5] M. K. Banavar, "Distributed inference over multiple-access channels with wireless sensor networks," Ph.D. dissertation, School of Electrical, Computer and Energy Engineering, Arizona State University, 2010.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. New Jersey: Prentice Hall, 1998.

[7] ——, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1998.

[8] S. Y. Chueng, S. C. Ergen, and P. Varaiya, "Traffic surveillance with wireless magnetic sensors," *Proceedings of the 12th ITS World Congress*, November 2005.

[9] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," in *NTC '78; National Telecommunications Conference, Birmingham, Ala., December 3-6, 1978, Conference Record. Volume 1. (A79-40501 17-32) Piscataway, N.J., Institute of Electrical and Electronics Engineers, Inc., 1978, p. 4.3.1-4.3.5.*, vol. 1, 1978, pp. 4.3.1–4.3.5.

[10] P. Gupta and P. R. Kumar, *Critical Power for Asymptotic Connectivity in Wireless Networks*. Birkhauser, 1998.

[11] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," *IEEE Infocom*, pp. 1079–1088, 2002.

[12] F. Xue and P. R. Kumar, *The Number of Neighbors Needed for Connectivity of Wireless Networks*. The Netherlands: Kluwer Academic Publishers, 2004, ch. 10, pp. 169–181.

[13] O. Dousse, F. Baccelli, and P. Thiran, "Impact of interferences on connectivity in ad-hoc networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 425–436, April 2005.

[14] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[15] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," *IEEE Infocom*, pp. 1360–1369, 2001.

[16] H. E. Gamal, "On the scaling laws of dense wireless sensor networks," *Proceedings of the Annual Allerton Conference on Communication, Control and Coding*, vol. 41, no. 3, pp. 1393–1401, 2003.

[17] F. Xue, L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 834–847, March 2005.

[18] L.-L. Xie and P. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2313–2328, June 2006.

[19] F. Xue and P. Kumar, *Scaling Laws for Ad-Hoc Wireless Networks: An Information Theoretic Approach*. Now Publishers, 2006, vol. 1, no. 2, pp. 145–270.

[20] M. Bohge and W. Trappe, "An authentication framework for hierarchical ad-hoc sensor networks. new york: Acm, 2003," *Proceedings of the 2nd ACM workshop on Wireless security*, vol. 22, no. 2, p. 7987, 2003.

[21] L. Sankaranarayanan, G. Kramer, and N. Mandayam, "Hierarchical sensor networks: capacity bounds and cooperative strategies using the

multiple-access relay channel model," in *Sensor and Ad-Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, oct. 2004, pp. 191 – 199.

[22] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors Part I: Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan 1997.

[23] R. Blum, S. Kassam, and H. Poor, "Distributed detection with multiple sensors Part II: Advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64 –79, jan 1997.

[24] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN'03)*, pp. 162–177, April 2003.

[25] M. Banavar, C. Tepedelenlioglu, and A. Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *Signal Processing, IEEE Transactions on*, vol. 58, no. 1, pp. 414 –425, jan. 2010.

[26] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 613–626, February 2006.

[27] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2617 – 2624, August 2005.

[28] Z. Hu and B. Li, "On the fundamental capacity and lifetime limits of energy-constrained wireless sensor networks," in *Proc. 10th IEEE Real-Time and Embedded Technology and Applications Symposium RTAS 2004*, May 25–28, 2004, pp. 2–9.

[29] M. Bhardwaj and A. P. Chandrakasan, "Bounding the lifetime of sensor networks via optimal role assignments," in *Proc. IEEE Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM 2002*, vol. 3, Jun. 23–27, 2002, pp. 1587–1596.

[30] M. Bhardwaj, T. Garnett, and A. P. Chandrakasan, "Upper bounds on the lifetime of sensor networks," in *Proc. IEEE International Conference on Communications ICC 2001*, vol. 3, Jun. 11–14, 2001, pp. 785–790.

[31] K. Kalpakis and S. Tang, "A combinatorial algorithm for the maximum lifetime data gathering and aggregation problem in sensor networks," in *Proc. International Symposium on a World of Wireless, Mobile and Multimedia Networks WoWMoM 2008*, Jun. 23–26, 2008, pp. 1–8.

[32] J. L. Gao, "Analysis of energy consumption for ad-hoc wireless sensor networks using a bit-meter-per-joule metric." in *IPN Progress Report 42-150, Jet Propulsion Laboratory*, vol. 3, Aug 2002, pp. 785–790.

[33] S. E, I. N. Cho S-H and, M. R, S. A, W. A, and C. A, "Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks," in *In Proceedings of 7th ACM Conference on Mobile Computing and Networkin*, July 2001.

[34] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1997.

[35] J. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, March 1993.

[36] F. Li and J. S. Evans, "Optimal strategies for distributed detection over multiaccess channels," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2008*, Mar. 2008, pp. 2417–2420.

[37] ——, "Design of distributed detection schemes for multiaccess channels," in *Proc. Australian Communications Theory Workshop AusCTW 2008*, Jan. 2008, pp. 51–57.

[38] C. Tepedelenlioglu and A. Narasimhamurthy, "Distributed estimation with constant modulus signaling over multiple access channels," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2010*, Mar. 2010.

[39] G. S. Lauer and N. R. S. Jr, "Distributed detection of known signal in correlated noise," *Rep. ALPHATECH, Burlington, MA*, vol. 2, pp. 297–344, March 1982.

[40] A. R. Reibman, "Performance and fault-tolerance of distributed detection networks,," *Ph.D. dissertation, Dept. Electrical Engineering, Duke Univ, Durham*, 1987.

[41] M. Xiang and C. Han, "Optimization of distributed detection networks with tree structures," in *Information Fusion, 2002. Proceedings of the Fifth International Conference on*, vol. 1, 2002, pp. 164–169 vol.1.

[42] C. Tepedelenlioglu and A. Narasimhamurthy, "Universal distributed estimation over multiple access channels with constant modulus signaling," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4783 –4794, Sept. 2010.

[43] M. Banavar, C. Tepedelenlioglu, and A. Spanias, "Distributed SNR estimation using constant modulus signaling over Gaussian multiple-access channels," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, jan. 2011, pp. 24 –29.

[44] ——, "Distributed SNR estimation with power constrained signaling over Gaussian multiple-access channels," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.

[45] J. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 167–182, April 1988.

[46] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access in sensor networks," in *Signal Processing Advances in Wireless Communications, 2005 IEEE 6th Workshop on*, June 2005, pp. 1018–1022.

[47] F. D. Hollander, *Large Deviations (Fields Institute Monographs,14)*. New York: Amer. Math. Soc., 2000.

[48] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, July 2006.

[49] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Joint estimation in sensor networks under energy constraints," *Poceedings of the First*

Annual IEEE Communications Society Conference on Sensor and Ad-Hoc Communications and Networks, pp. 264–271, October 2004.

[50] S. Cui, J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks - analog vs. digital," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, March 2005, pp. 745–748.

[51] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[52] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, February 2006.

[53] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[54] P. Gao and C. Tepedelenliglu, "Practical issues in estimation over multiaccess fading channels with TBMA wireless sensor networks," vol. 56, no. 3, pp. 1217–1229, Mar. 2008.

[55] J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "CTH15-1: Linear Coherent Decentralized Estimation," *GLOBECOM '06. IEEE*, pp. 1–5, November 2006.

[56] J. N. Tsitsiklis, "Problems in decision making and computation," Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, 1984. [Online]. Available: http://www.mit.edu/ jnt/Papers/PhD-84-jnt.pdf

[57] J. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *Automatic Control, IEEE Transactions on*, vol. 29, no. 1, pp. 42 – 50, Jan. 1984.

[58] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 5, Dec. 2003, pp. 4997 – 5002.

[59] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Analysis and optimization of randomized gossip algorithms," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 5, Dec. 2004, pp. 5310 – 5315.

[60] R. Saber and R. Murray, "Consensus protocols for networks of dynamic agents," in *American Control Conference, 2003. Proceedings of the 2003*, vol. 2, Apr. 2003, pp. 951 – 956.

[61] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215 –233, Jan. 2007.

[62] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1520 – 1533, Sept. 2004.

[63] M. Huang and J. Manton, "Stochastic consensus seeking with measurement noise: Convergence and asymptotic normality," in *American Control Conference, 2008*, June 2008, pp. 1337 –1342.

[64] B. Oreshkin, T. Aysal, and M. Coates, "Distributed average consensus with increased convergence rate," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, Mar. 2008, pp. 2285 –2288.

[65] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 355 –369, Jan. 2009.

[66] S. Kar, S. Aldosari, and J. Moura, "Topology for distributed inference on graphs," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2609 –2613, June 2008.

[67] S. Kar and J. Moura, "Distributed average consensus in sensor networks with random link failures and communication channel noise," in *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, Nov. 2007, pp. 676 –680.

[68] T. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 34, no. 2, pp. 469 –476, apr 1998.

[69] W. Li and H. Dai, "Distributed detection in wireless sensor networks using a multiple access channel," *Signal Processing, IEEE Transactions on*, vol. 55, no. 3, pp. 822 –833, march 2007.

[70] K. Liu and A. Sayeed, "Type-based decentralized detection in wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1899–1910, May 2007.

[71] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *Signal Processing, IEEE Transactions on*, vol. 55, no. 3, pp. 1081–1092, March 2007.

[72] A. Anandkumar and L. Tong, "Type-based random access for distributed detection over multiaccess fading channels," *Signal Processing, IEEE Transactions on*, vol. 55, no. 10, pp. 5032–5043, Oct. 2007.

[73] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[74] Q. Tian and E. J. Coyle, "Optimal distributed detection in clustered wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3892–3904, 2007.

[75] F. Li and J. S. Evans, "Decision fusion over noncoherent fading multiaccess channels," in *Proc. IEEE Global Telecommunications Conference IEEE GLOBECOM 2008*, Nov. 2008, pp. 1–5.

[76] M. Banavar, A. Smith, C. Tepedelenlioglu, and A. Spanias, "Distributed detection over fading MACs with multiple antennas at the fusion center," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 2894 –2897.

[77] T. Aysal and K. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *Signal Processing, IEEE Transactions on*, vol. 56, no. 4, pp. 1398 –1410, 2008.

[78] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Linear coherent decentralized estimation," *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 757 –770, 2008.

[79] B. Picinbono, "On deflection as a performance criterion in detection," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, no. 3, pp. 1072 –1081, jul 1995.

[80] R. Niu and P. Varshney, "Performance analysis of distributed detection in a random sensor field," *Signal Processing, IEEE Transactions on*, vol. 56, no. 1, pp. 339 –349, jan. 2008.

[81] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[82] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer; 1 edition, 1987.

[83] G. Samorodnitsky and M. S. Taqqu, "Stable non-Gaussian random processes," 1994, pp. 2–4, 10–13, 77–79, 84–86.

[84] Z. Zhang, "Inequalities for characteristic functions involving Fisher information," *Comptes Rendus Mathematique*, vol. 344, no. 5, pp. 327–330, March 2007.

[85] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 1246–1250, May 1998.

[86] B. Porat, *Digital processing of random signals: theory and methods*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[87] N. G. Ushakov, *Selected topics in characteristic functions*. VSP International Science Publishers, 1999.

[88] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley-Interscience, 1990.

[89] I. A. S. E. Milton Abramowitz (Editor), *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables (Paperback)*. Dover Publications, June 1, 1965.

[90] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[91] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, July 2006.

[92] M. Banavar, C. Tepedelenlioglu, and A. Spanias, "Distributed SNR estimation with power constrained signaling over Gaussian multiple-access channels," *Signal Processing, IEEE Transactions on*, vol. 60, no. 6, pp. 3289 –3294, June 2012.

[93] R. W. Santucci, M. K. Banavar, C. Tepedelenlioglu, and A. Spanias, "Energy-efficient distributed estimation by utilizing a nonlinear amplifier," *Signal Processing, Constantinides International Workshop on*, Jan. 2013.

[94] S. C. Cripps, *Advanced techniques in RF power amplifier design*. Norwood, MA: Artech House, 2002.

[95] ——, *RF Power Amplifiers for Wireless Communications, Second Edition (Artech House Microwave Library (Hardcover))*. Norwood, MA, USA: Artech House, Inc., 2006.

[96] M. Goldenbaum, S. Stanczak, and M. Kaliszan, "On function computation via wireless sensor multiple-access channels," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, april 2009, pp. 1 –6.

[97] M. Goldenbaum and S. Stanczak, "Computing the geometric mean over multiple-access channels: Error analysis and comparisons," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, nov. 2010, pp. 2172 –2178.

[98] S. Dasarathan, C. Tepedelenlioglu, M. Banavar, and A. Spanias, "Nonlinear distributed average consensus using bounded transmissions," 2013. [Online]. Available: http://arxiv.org/abs/1302.5371

[99] C. Tepedelenlioglu and S. Dasarathan, "Distributed detection over Gaussian multiple access channels with constant modulus signaling," *Signal*

Processing, IEEE Transactions on, vol. 59, no. 6, pp. 2875 –2886, june 2011.

[100] ——, "Distributed detection over Gaussian multiple access channels with constant modulus signaling," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, nov. 2010, pp. 2008 –2012.

[101] M. S. Chrysostomos L. Nikias, *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley-Interscience, 1 edition, 1995.

[102] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 2(Paperback)*. Wiley; 2nd edition, 1991.

[103] R. Bartle, *The Elements of Integration and Lebesgue Measure*, ser. Wiley Classics Library. Wiley, 2011.

[104] R. G. Bartle, *The Elements of Real Analysis*. Jhon Wiley and Sons, 1967.

[105] M. Gastpar and M. Vetterli, "Source-Channel communication in sensor networks." International Workshop on Information Processing in Sensor Networks (IPSN'03), March 2003, pp. 162–177.

[106] G. Pottie and W. Kaiser, *Principles of Embedded Networked Systems Design*. New York: Cambridge University Press, 2005.

[107] J. Fang and H. Li, "An adaptive quantization scheme for distributed consensus," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, Apr. 2009, pp. 2777 – 2780.

[108] ——, "Distributed consensus with quantized data via sequence averaging," *Signal Processing, IEEE Transactions on*, vol. 58, no. 2, pp. 944 – 948, Feb. 2010.

[109] T. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus with dithered quantization," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4905 – 4918, Oct. 2008.

[110] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1383 –1400, Mar. 2010.

[111] ——, "Distributed average consensus in sensor networks with quantized inter-sensor communication," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, Mar. 2008, pp. 2281 –2284.

[112] F. Chung, *Spectral Graph Theory*, ser. Regional Conference Series in Mathematics. Conference Board of the Mathematical Sciences, 1997, no. no. 92.

[113] U. Khan, S. Kar, and J. Moura, "Distributed average consensus: Beyond the realm of linearity," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, nov. 2009, pp. 1337 – 1342.

[114] U. Munz, A. Papachristodoulou, and F. Allgower, "Nonlinear multi-agent system consensus with time-varying delays," *Proceedings of the 17th World Congress The International Federation of Automatic Control*, pp. 6 – 11, July 2008.

[115] Q. Hui and W. M. Haddad, "Distributed nonlinear control algorithms for network consensus," *Automatica*, vol. 44, no. 9, pp. 2375 – 2381, 2008.

[116] W. Yu, G. Chen, and M. Cao, "Consensus in directed networks of agents with nonlinear dynamics," *Automatic Control, IEEE Transactions on*, vol. 56, no. 6, pp. 1436–1441, 2011.

[117] A. Ajorlou, A. Momeni, and A. G. Aghdam, "Sufficient conditions for the convergence of a class of nonlinear distributed consensus algorithms," *Automatica*, vol. 47, no. 3, pp. 625 – 629, 2011.

[118] M. Nevelson and R. Khasminskiĭ, *Stochastic Approximation and Recursive Estimation*. American Mathematical Society, 1973.

[119] D. Williams, *Probability with Martingales*, ser. Cambridge Mathematical Textbooks. Cambridge University Press, 1991. [Online]. Available: http://books.google.com/books?id=e9saZ0YSi-AC

[120] B. Polyak and Y. Tsypkin, "Optimal pseudogradient adaptation procedures," *Automat, Remote Control*, vol. 41, pp. 1101–1110, 1981.

[121] J. Cortes, "Analysis and design of distributed algorithms for x-consensus," in *Decision and Control, 2006 45th IEEE Conference on*, Dec. 2006, pp. 3363 – 3368.

[122] S. Sundaram and C. Hadjicostis, "Distributed function calculation via linear iterations in the presence of malicious agents x2014; part ii: Overcoming malicious behavior," in *American Control Conference, 2008*, June 2008, pp. 1356 –1361.

[123] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, pp. 33–46, Jan. 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1222667.1222952

[124] B. Touri and A. Nedic, "Distributed consensus over network with noisy links," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, July 2009, pp. 146 –154.

[125] M. Huang and J. Manton, "Stochastic approximation for consensus seeking: Mean square and almost sure convergence," in *Decision and Control, 2007 46th IEEE Conference on*, Dec. 2007, pp. 306 –311.

[126] L. Pescosolido, S. Barbarossa, and G. Scutari, "Average consensus algorithms robust against channel noise," in *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, July 2008, pp. 261 –265.

[127] S. Barbarossa, T. Battisti, L. Pescosolido, S. Sardellitti, and G. Scutari, "Distributed processing algorithms for wireless sensor networks having fast convergence and robustness against coupling noise," in *Spread Spectrum Techniques and Applications, 2008. ISSSTA '08. IEEE 10th International Symposium on*, Aug. 2008, pp. 1 – 6.

[128] T. Aysal and K. Barner, "Convergence of consensus models with stochastic disturbances," *Information Theory, IEEE Transactions on*, vol. 56, no. 8, pp. 4101 –4113, Aug. 2010.

[129] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 4, pp. 772–790, 2011.

[130] J. Fraden, *Handbook of Modern Sensors: Physics, Designs and Applications*, ser. Handbook of Modern Sensors. AIP Press/Springer Verlag, 2004.