

國立交通大學

電控工程研究所

博士論文

最佳化功率分配於無線感測網路之線性一致離

散估測

Linear Coherent Distributed Estimation in Wireless
Sensor Networks with Optimal Power Allocation

研究生：吳建賢

指導教授：林清安教授

中華民國一百年七月

最佳化功率分配於無線感測網路之線性一致離散估測

Linear Coherent Distributed Estimation in Wireless Sensor
Networks with Optimal Power Allocation

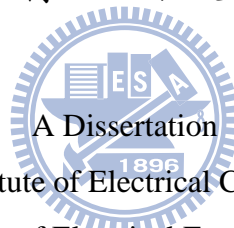
研究生：吳建賢

Student: Chien-Hsien Wu

指導教授：林清安 教授

Advisor: Ching-An Lin

國立交通大學
電控工程研究所
博士論文



Submitted to Institute of Electrical Control Engineering

College of Electrical Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Department of Electrical Engineering

July 2011

Hsinchu, Taiwan, Republic of China

中華民國一百年七月

最佳化功率分配於無線感測網路之線性一致離散估測

研究生：吳建賢

指導教授：林清安

國立交通大學電控工程研究所

中文摘要

本論文以一致性多路存取通道與線性最小均方差融合規則為基礎，探討無線感測網路的離散估測，我們討論三種不同感測網路的離散估測：(i) 以群聚單輸入單輸出感測網路為基礎的估測、(ii) 單輸入單輸出感測網路在未知通道下的估測、以及 (iii) 多輸入多輸出感測網路下的估測；於每一種感測網路下，我們皆討論系統總功率有限的情況下，利用最佳化功率分配來得到最小估測失真。首先，我們討論一純量訊號於群聚感測網路中的離散估測，我們證明最佳的放大矩陣為一個利用兩已知向量做外積所得到的秩一矩陣，由此最佳的放大矩陣，我們證明利用感測器合作能夠改善系統的效能。第二，我們探討一純量訊號在未知通道的無線網路中的離散估測，我們利用二階段法來估測訊號：先分配訓練功率於各感測器來估測通道，再用最佳功率分配策略或平均功率分配策略結合前一步驟所得到的估測通道來估測信號，我們證明不論是最佳功率分配策略或是平均功率分配策略，最佳的訓練功率是相同的；當感測器的數量增加時，在通道未知的情況下所造成的的效能損失也會增加。最後，我們討論一向量訊號於多輸入多輸出感測網路的離散估測，我們利用奇異值分解法將求編碼矩陣的問題表示為凸面最佳化問題，藉此，我們推導出最佳編碼矩陣的封閉解，我們利用數值模擬來證明於三種感測網路下的分析結果。

Linear Coherent Distributed Estimation in Wireless Sensor Networks with Optimal Power Allocation

Student: Chien-Hsien Wu

Advisor: Ching-An Lin

Institute of Electrical Control Engineering
National Chiao-Tung University

Abstract

We study distributed estimation in wireless sensor networks with coherent multiple access channel model and LMMSE fusion rule. In this thesis, three different sensor network systems for distributed estimation are discussed: (i) estimation using single-input single-output (SISO) cluster-based sensor network, (ii) estimation using SISO sensor network over unknown channels, and (iii) estimation using multiple-input multiple-output (MIMO) sensor network. In each network system, we study the problem of minimizing estimation distortion by optimally allocating power under a total power constraint. We first discuss distributed estimation of a scalar signal with cluster-based sensor network. We show that the optimal amplification matrix for each cluster is a rank one matrix, which is a scaled outer product of two known vectors. With the optimal amplification matrices, we also show that collaboration can improve performance. Secondly, we consider distributed estimation of a scalar signal using sensor network with unknown channels. We adopt a two-phase approach, which first optimally allocates the training power for channel estimation, and then uses the equal power scheme or optimal power scheme for source signal estimation. We reveal that the optimal training powers for the optimal and equal power schemes are the same. Moreover, as the number of sensors increases, the penalty caused by channel estimation becomes worse. Finally, we discuss distributed estimation of a vector signal using MIMO sensor network. Based on singular value decomposition technique, the problem of choosing coding matrices can be formulated as a convex optimization problem, based on which we derive closed form expression of optimal coding matrices. We use simulations to verify the analytical results in the three network systems.

致 謝

首先我要向我的指導教授林清安老師表達最誠摯的感謝，由於老師耐心的指導及幫助，使我得以順利完成論文；老師專業領域的博學、研究上嚴謹的態度、簡潔清楚的觀念闡述、發現與解決問題的方法，以及對學術真理的堅持，為我對學術研究的態度上樹立最佳典範，跟老師學習的這幾年讓我獲益良多。

我亦要感謝論文口試委員交大電控所林源倍教授、蔡尚澤教授，交大電信所吳卓諭教授，中山大學通訊所王藏億教授，暨南大學通訊所李彥文教授，以及逢甲大學通訊所陳益生教授對本論文提供建議並給予指導，使本論文更加完善；我要特別感謝吳卓諭教授在我研究所這幾年對我研究上的幫忙和課業上的指導，還有陳益生教授在學術上給予的建議。

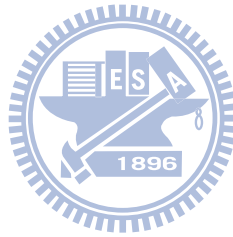
此外，感謝實驗室同學們的陪伴，你們讓我研究所的生活更加充實，感謝文達、敬淵、炅嶽、建宇以及傑智陪我打球，使我在生活上有所調劑；感謝正哲、威德以及哲嘉對生活上的關照；泰偉、惟庭以及彥伯講述旅遊見聞，為我生活添增樂趣；瑞華、江龍、育霖、政興、世修、國展、逸棋，有你們讓我的研究所生活更多彩。

最後，我要感謝我的家人，特別是我的父母及兩位弟弟，沒有你們的支持與鼓勵，我可能無法拿到學位，亦謝謝妻子如忻及岳父母給我的關心與幫助，你們的支持讓我更有動力堅持下去並完成學業。我要將這份論文獻給我最愛的家人以及我們家庭即將到來的新成員。

Contents

Chinese Abstract	i
Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Overview of Wireless Sensor Networks	1
1.2 Literature Review on Distributed Estimation	2
1.3 Problem Studied in This Thesis	3
1.4 Organization of the Thesis	5
2 Distributed Estimation Using Cluster-Based Sensor Network	6
2.1 System Model and Problem Formulation	6
2.2 Optimal Amplification Matrices	9
2.3 Numerical Results and Discussion	12
3 Distributed Estimation Using Sensor Network with Unknown Channels	20
3.1 System Model	21
3.2 LMMSE Estimation	22
3.2.1 <i>Channel Estimation</i>	22
3.2.2 <i>Source Estimation</i>	23
3.3 Optimal Power Allocation	24
3.3.1 <i>When channels are known</i>	24
3.3.2 <i>When channels are estimated</i>	26

3.3.3	<i>Comparison of two cases</i>	28
3.4	Equal Power Allocation	29
3.4.1	<i>When channels are known</i>	29
3.4.2	<i>When channels are estimated</i>	30
3.4.3	<i>Comparison of two cases</i>	31
3.5	Numerical Results and Discussion	32
4	Distributed Estimation Using MIMO Sensor Network	39
4.1	System Model and Problem Formulation	39
4.2	Proposed Approach	42
4.3	Numerical Results and Discussion	46
5	Conclusions	53
A	Proof of Proposition 2.1	55
B	Proof of Lemma 2.3	58
C	Derivation of (3.9)	60
D	Proof of Proposition 3.1	61
E	Derivation of (3.25)	63
F	Convexity of Objective Function in (3.33)	64
G	Proof of Proposition 3.2	65
	Bibliography	67



List of Figures

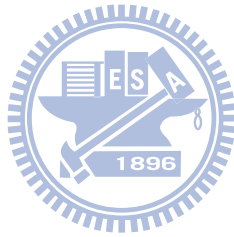
2.1	Cluster-based sensor network with coherent MAC	7
2.2	MSE of full collaboration case with different numbers of transmitters	13
2.3	MSE of full collaboration and non-collaboration cases with different noise power levels	14
2.4	MSE of full collaboration and non-collaboration cases with different power levels	15
2.5	MSEs for $K_l = 1, 4, 8$, and K with different number of sensors	15
2.6	MSEs with different number of clusters	16
2.7	Cluster-based sensor network with orthogonal MAC	17
2.8	Comparison of the coherent MAC model to that of the orthogonal MAC model	17
3.1	Coherent MAC wireless sensor network	21
3.2	Mean square error (MSE) with optimal power allocation	32
3.3	Mean square error (MSE) with equal power allocation	33
3.4	Comparison between equal and optimal power allocation schemes	34
3.5	Orthogonal MAC wireless sensor network	35
3.6	Performance of optimal power allocation scheme: coherent model and orthogonal model	35
3.7	MSE ratio versus total power ratio: optimal power scheme with different P^e	36
3.8	MSE ratio versus total power ratio: equal power scheme with different K	37
4.1	Linear distributed estimation with coherent MAC	40
4.2	MSE versus N with different power levels	47

4.3	MSE versus k_l with different power levels	48
4.4	MSE comparison between the proposed method and the equal power method	49
4.5	Network structure in [28]	50
4.6	MSE comparison between the proposed scheme and that in [28]	50
4.7	MSE versus P with different number of sensors	51



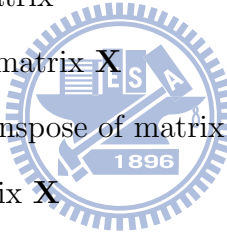
List of Tables

2.1 Different number of sensors in clusters 18



List of Notation

\mathbb{R}^n	Real n -vectors
$\mathbb{R}^{l \times m}$	Real $l \times m$ matrices
\mathbb{C}^n	Complex n -vectors
$\mathbb{C}^{l \times m}$	Complex $l \times m$ matrices
\mathbf{I}_n	$n \times n$ identity matrix
$\mathbf{0}$	Zero vector
$\mathbf{0}_{l \times m}$	$l \times m$ zero matrix
\mathbf{X}^T	Transpose of matrix \mathbf{X}
\mathbf{X}^H	Hermitian transpose of matrix \mathbf{X}
$\text{tr}(\mathbf{X})$	Trace of matrix \mathbf{X}
$\text{rank}(\mathbf{X})$	Rank of matrix \mathbf{X}
$\text{diag}(x_1, \dots, x_n)$	Diagonal matrix with diagonal entries x_1, \dots, x_n
x^*	Complex conjugate of x
$ x $	Magnitude of x
$\ \mathbf{x}\ $	Norm of vector \mathbf{x}



Chapter 1

Introduction

1.1 Overview of Wireless Sensor Networks

Wireless sensor networks (WSNs), made up of a large number of low cost devices, have attracted much attention due to the advancements in micro-electro-mechanical systems (MEMS) technology. The technological advancements enable multi-function sensors to be small in size and communicate in short distances. In WSNs, the sensors, which consist of sensing, data processing, and communication components, are densely deployed in different regions of a field to aggregate data. For example, sensor networks are used to monitor temperature in environmental applications, to track enemy movement in military applications, and to report life-signs in health applications [1, 2].

Due to simple structure of sensors and dense deployment across a spatial domain, two key resources – communication bandwidth and energy – are an important challenge in the design of WSNs. Many previous research works, e.g. [3–7], addressed network-structure-based protocols which save communication resources by self-organization of large number of sensors so that the transmitted redundancy is reduced. Recently, WSNs based on distributed signal processing perspectives are used for parameters detection and estimation [8, 9], and object tracking [10] by developing energy and bandwidth efficient algorithms. The WSN topology used to design distributed signal processing algorithms can be classified into (i) the presence of a fusion center (FC), (ii) the absence of a FC named

ad hoc WSNs, and (iii) hybrids of (i) and (ii) named cluster-based WSNs.

1.2 Literature Review on Distributed Estimation

In WSNs, signal processing sensors are capable of simple local computation, short range low data-rate communication, and the fusion center (FC) has more powerful communication and processing capability. The fusion center receives data transmitted from the sensors over wireless channels and combines it for a specific processing purpose. One example of such distributed signal processing scheme is distributed estimation. A certain parameter or variable is measured by the sensors and the measurements are sent to the FC, and the objective is to estimate the parameter based on distributed sensor measurements. Due to bandwidth and power limitations, many algorithms have been proposed to make efficient use of these two resources.

In the distributed estimation scenario, existing algorithms for modelling the bandwidth constraint can be classified into the quantized strategy and the unquantized strategy. The quantized strategy is to limit the number of bits that each sensor can send to the fusion center per observation period, i.e., the measurements sent from the sensors are quantized, encoded, and transmitted via digital modulation. Some papers discuss minimization of mean squared error (MSE) via bit length assignment [11–14], while others focus on the search of optimal quantization threshold for one bit quantization schemes [15–17]. The unquantized strategy is to limit the number of messages that each sensor can send to the fusion center per observation period. In this strategy, the sensors send raw measurements directly through channels without quantization, and thus analog transmission schemes, such as amplify-and-forward approach, are used. It is asserted in [18] that the amplified-and-forward approach is optimal over additive white Gaussian noise channels. Along this line of approach, many papers study the minimization of MSE under a total network power constraint by optimally allocating the transmitted power for each sensor [19–28] and others analyze asymptotic behavior as the network power or the number of sensors increases [29–31].

In the amplify-and-forward approach, two types of channel models are used. The first one is the orthogonal multiple access channel (MAC), in which each sensor has an independent noninterfering channel to the FC. This MAC model was widely used in previous studies on optimal power allocation problems [19, 20, 24, 25]. In [19, 20], the source signal is assumed deterministic and the best linear unbiased estimator (BLUE) is used at the FC. For random source signal, distributed linear minimum mean squared error (LMMSE) estimation based on sensor clustering is considered in [24] and in [25] the results are extended to the case where channels are assumed unknown. The second type of channel model is the coherent MAC, in which all sensors transmit simultaneously to the FC via the same frequency. In this MAC model, previous works study optimal power allocation problems under LMMSE fusion rule both scalar signal [27] as well as vector signal [28] are considered.

All proposed schemes for solving the power allocation problems in [19, 20, 24, 25, 27, 28] used the single-input single-output (SISO) systems in which the sensor with a scalar measurement/transmitter is regarded as a unit. Recently, multiple-input multiple-output (MIMO) systems have attracted much attention due to advantages of increase in data rate and improvement of performance [33]. In MIMO wireless sensor networks, each sensor has a vector measurement and multiple transmitters. Recent works on MIMO sensor networks addressed dimensionality reduction problems based on ideal channel assumption [21–23] and the optimal design of coding matrices [26].

1.3 Problem Studied in This Thesis

In this thesis, we study linear distributed estimation in WSNs based on the LMMSE fusion rule. The coherent MAC model, in which the transmitted signals from all sensors are received by the FC as a coherent sum, is taken into account. The problem we investigate is to design optimal power allocation schemes so that the estimation distortion is minimized under a total power constraint. We consider distributed estimation with three different

sensor network systems:

- (i) Estimation of a scalar signal with SISO cluster-based sensor network. In cluster-based sensor network, the sensors are divided into a number of small groups called clusters. The sensors in the same cluster are allowed to collaborate, while collaboration is prohibited for sensors in different clusters. For each cluster, the collaboration is through an amplification matrix that forms a message from the measurements for transmission to the FC, where a final estimate is formed. We study the problem of choosing the amplification matrices so that the MSE of the estimated source signal is minimized under a total power constraint. We show that the optimal amplification matrix is a rank one matrix and that, with optimal amplification matrices, collaboration improves performance in terms of MSE.
- (ii) Estimation of a scalar signal using SISO sensor network with unknown flat fading channels. Since channels are unknown at the FC and need to be estimated, we derive training-based LMMSE channel estimator. The channel estimates are then used to obtain LMMSE estimation of the source signal. We study the allocation of power for training and for data transmission to each sensor, under a total power constraint, so as to minimize the MSE of the estimated source signal. We consider the optimal power allocation scheme in which training power and data power for each sensor are optimized, and the equal power allocation scheme in which training power is optimized while data power for each sensor is set equal. We show that the optimal training powers are the same for both two schemes. Also, compared with the case when channels are known, the penalty caused by the channel estimation error is roughly proportional to the number of sensor when the total power is large.
- (iii) Estimation of a vector signal using MIMO sensor network. In MIMO sensor networks, each sensor has multiple measurements which are multiplied by a linear coding matrix to form messages for transmission to the FC through a channel matrix; at the FC, messages from all sensors are aggregated by multiple received antennas to form an estimated vector signal. In the sensor network, we study the problem of design of linear coding matrices so that the MSE of the estimated vector signal is

minimized under a total power constraint. We show that the problem can be written as a convex optimization problem which follows a water-filling type solution. Hence, closed form expressions for the optimal coding matrices are obtained.

1.4 Organization of the Thesis

The thesis is organized as follows. In Chapter 2, we consider distributed estimation using cluster-based sensor network. We derive the optimal design of amplification matrix and give the performance comparison between the collaboration case and the non-collaboration case. In Chapter 3, we consider distributed estimation using sensor network with unknown channels. We use a two-phase approach, which first estimates channels and then estimates the source signal. We study optimal allocation of power for training and for data transmission based on the optimal and equal power allocation schemes. The performance comparison between the case when channels are known and when channels are unknown are also discussed. In Chapter 4, we consider distributed estimation of a vector signal using MIMO sensor network. Based on singular value decomposition technique, we propose a method that can solve the estimation problem efficiently, which then leads to closed form expressions of the optimal coding matrices. Chapter 5 concludes this thesis and discusses the future research.

Chapter 2

Distributed Estimation Using Cluster-Based Sensor Network

It is shown in [5,7] that sensor clustering can make efficient use of energy and thus prolong network lifetime. In this chapter, we study distributed estimation of a scalar parameter based on cluster-based sensor networks. The problem is formulated as the choice of the amplification matrices so that the MSE of the estimated source signal is minimized under a total power constraint. We also discuss two special cases: the full collaboration case, in which all sensors are in the same cluster, and the non-collaboration case, in which each cluster has only one sensor. Numerical examples are used to illustrate performance improvement based on sensor collaboration.

2.1 System Model and Problem Formulation

We consider a wireless sensor network consisting of K spatially deployed sensors for estimating a random source signal θ . The sensors in the network are divided into L clusters, as shown in Figure 2.1. The l th cluster has K_l sensors and the measurement at the k th sensor is given by

$$x_{l,k} = f_{l,k}\theta + n_{l,k}, \quad 1 \leq l \leq L, 1 \leq k \leq K_l, \quad (2.1)$$

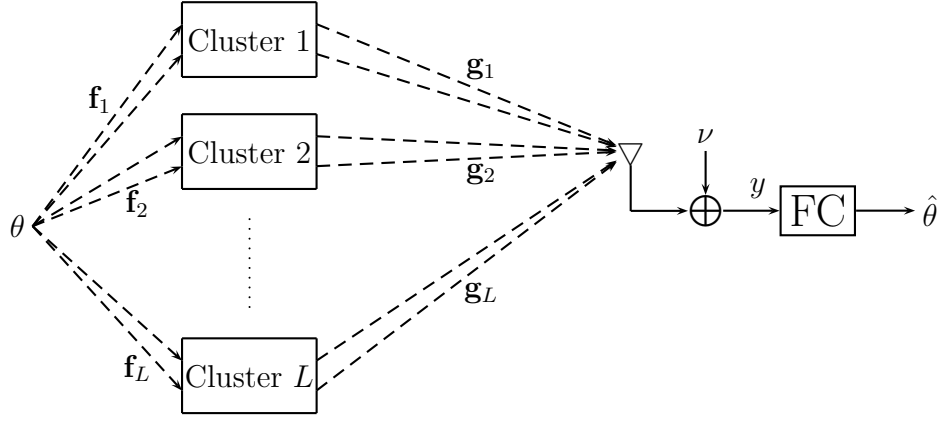


Figure 2.1: Cluster-based sensor network with coherent MAC

where $f_{l,k}$ is the observation gain and $n_{l,k}$ is the measurement noise. In vector form, (2.1) becomes

$$\mathbf{x}_l = \mathbf{f}_l \theta + \mathbf{n}_l, \quad 1 \leq l \leq L, \quad (2.2)$$

where $\mathbf{x}_l = [x_{l,1} \cdots x_{l,K_l}]^T$, $\mathbf{f}_l = [f_{l,1} \cdots f_{l,K_l}]^T$, and $\mathbf{n}_l = [n_{l,1} \cdots n_{l,K_l}]^T$. The collaboration between sensors in the l th cluster is through an amplification matrix $\mathbf{A}_l \in \mathbb{R}^{N_l \times K_l}$, which takes $\mathbf{x}_l \in \mathbb{R}^{K_l}$ to form the message vector $\mathbf{A}_l \mathbf{x}_l \in \mathbb{R}^{N_l}$. The messages are then sent to the FC and the signal y received at the FC can be expressed as

$$y = \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l (\mathbf{f}_l \theta + \mathbf{n}_l) + \nu \quad (2.3)$$

where $\mathbf{g}_l = [g_{l,1} \cdots g_{l,N_l}]^T$ is the channel gain vector and ν is the additive noise at the receiver. In practice, the sensors which are geographically closely located can compose a cluster. The collaboration between sensors in the same cluster can be implemented by choosing one sensor as the cluster head whose task is to collect and process information sent from other sensors to form a message vector and transmit it to the FC.

The following assumptions are made in this chapter:

- i) $E[\theta] = 0$ and $E[\theta^2] = \sigma_\theta^2$.
- ii) The measurement noises are zero-mean and mutually uncorrelated, specifically $E[\mathbf{n}_l] = \mathbf{0}$, $E[\mathbf{n}_l \mathbf{n}_l^T] = \sigma_n^2 \mathbf{I}_{K_l}$ and $E[\mathbf{n}_l \mathbf{n}_m^T] = \mathbf{0}_{K_l \times K_m}$ for $l \neq m$.

iii) $E[\nu] = 0$ and $E[\nu^2] = \sigma_\nu^2$.

iv) The source signal, the measurement noises, and the receiver noise are uncorrelated, that is, $E[\theta \mathbf{n}_l] = \mathbf{0}$, $E[\theta \nu] = 0$, and $E[\nu \mathbf{n}_l] = \mathbf{0}$.

v) The observation gain vectors \mathbf{f}_l and the channel gain vectors \mathbf{g}_l are known to the FC.

Remarks: Assumption v) is reasonable in cases where the network condition and the signal observation model change slowly in a quasi-static manner. Hence, as \mathbf{f}_l and \mathbf{g}_l are obtained by the FC, they can be used for a long period of time.

For a given set of amplification matrices \mathbf{A}_l , the LMMSE estimate of θ using the received signal y in (2.3) is [35, p.382]

$$\begin{aligned} \hat{\theta} &= \frac{E[\theta y]}{E[y^2]} y \\ &= \frac{\sigma_\theta^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l}{\sigma_\theta^2 \left(\sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l \right)^2 + \sigma_n^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{A}_l^T \mathbf{g}_l + \sigma_\nu^2} y \end{aligned} \quad (2.4)$$

and the corresponding MSE is

$$\begin{aligned} J &= E[(\theta - \hat{\theta})^2] = \sigma_\theta^2 - \frac{(E[\theta y])^2}{E[y^2]} \\ &= \left(\frac{1}{\sigma_\theta^2} + \frac{\left(\sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l \right)^2}{\sigma_n^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{A}_l^T \mathbf{g}_l + \sigma_\nu^2} \right)^{-1} \end{aligned} \quad (2.5)$$

The problem is to minimize the MSE in (2.5) by choosing optimal amplification matrices \mathbf{A}_l under a total power constraint. The total transmitted power of the L clusters is $\sum_{l=1}^L E[\mathbf{x}_l^T \mathbf{A}_l^T \mathbf{A}_l \mathbf{x}_l]$. Hence if P is the amount of power that the clusters together can use, then we have the following constraint

$$\sum_{l=1}^L \text{tr} (E[\mathbf{A}_l \mathbf{x}_l \mathbf{x}_l^T \mathbf{A}_l^T]) = \sum_{l=1}^L \text{tr} (\sigma_\theta^2 \mathbf{A}_l \mathbf{f}_l \mathbf{f}_l^T \mathbf{A}_l^T + \sigma_n^2 \mathbf{A}_l \mathbf{A}_l^T) \leq P \quad (2.6)$$

where we use $E[\mathbf{x}_l \mathbf{x}_l^T] = \sigma_\theta^2 \mathbf{f} \mathbf{f}^T + \sigma_n^2 \mathbf{I}_{K_l}$. From (2.5) and (2.6), the optimization problem under consideration can be written as

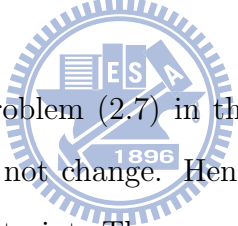
$$\begin{aligned} &\min_{\mathbf{A}_l, 1 \leq l \leq L} J \\ &\text{subject to } \sum_{l=1}^L \text{tr} (\sigma_\theta^2 \mathbf{A}_l \mathbf{f}_l \mathbf{f}_l^T \mathbf{A}_l^T + \sigma_n^2 \mathbf{A}_l \mathbf{A}_l^T) \leq P, \end{aligned} \quad (2.7)$$

where J is given in (2.5).

Remarks: We had assumed that the measurement noises are mutually uncorrelated across all sensors. If the measurement noises are correlated within the same cluster but uncorrelated across different clusters, the problem can still be formulated in the same form as (2.7). To see this, suppose $E[\mathbf{n}_l \mathbf{n}_l^T] = \mathbf{R}_{n_l}$, where $\mathbf{R}_{n_l} = \mathbf{R}_{n_l}^T \in \mathbb{R}^{K_l \times K_l}$ is positive definite and $E[\mathbf{n}_l \mathbf{n}_m^T] = \mathbf{0}_{K_l \times K_m}$ for $l \neq m$. Let $\mathbf{R}_{n_l} = \mathbf{U}_{n_l} \mathbf{\Lambda}_{n_l} \mathbf{U}_{n_l}^T$ be the eigenvalue decomposition with $\mathbf{\Lambda}_{n_l} = \text{diag}(\sigma_{n_{l,1}}^2, \dots, \sigma_{n_{l,K_l}}^2) > 0$. By setting $\tilde{\mathbf{A}}_l = \mathbf{A}_l \mathbf{U}_{n_l} \mathbf{\Lambda}_{n_l}^{1/2}$ and $\tilde{\mathbf{f}}_l = \mathbf{\Lambda}_{n_l}^{-1/2} \mathbf{U}_{n_l}^T \mathbf{f}_l$, the corresponding optimization problem has the same form as (2.7) with \mathbf{A}_l , \mathbf{f}_l , and σ_n^2 replaced by $\tilde{\mathbf{A}}_l$, $\tilde{\mathbf{f}}_l$, and 1, respectively.

2.2 Optimal Amplification Matrices

In this section, we consider the solution of the optimization problem (2.7) with the goal of obtaining a closed form expression for the optimal amplification matrices \mathbf{A}_l . We first make the following observations:

- 
- (i) If the inequality sign of problem (2.7) in the constraint is replaced by the equality sign, the solution does not change. Hence we could consider the optimization problem with equality constraint. The argument is as follows. Since the constraint function is quadratic in the elements of \mathbf{A}_l , if a set of \mathbf{A}_l is such that strict inequality holds, we can equally scale up each \mathbf{A}_l so that equality holds. And if we equally scale up each \mathbf{A}_l , we get a lower function value of J because in (2.5) the second term inside the parentheses becomes larger. Consequently, with optimal \mathbf{A}_l , the inequality constraint must be active.
 - (ii) Consider the optimal MSE in (2.7), say, J^* as a function of the power P , then J^* is a strictly decreasing function of P , that is, if $P_2 > P_1$, then $J^*(P_2) < J^*(P_1)$. The argument is similar: if the power level increases, we can equally scale up \mathbf{A}_l to obtain a lower value of J and thus a lower value of optimal MSE J^* can be obtained.
 - (iii) Since the function $J^*(P)$ is 1-1 and decreasing, the inverse function $P(J^*)$ is also 1-1 and decreasing. Hence instead of finding the matrices \mathbf{A}_l that minimize J in

(2.5) under an equality constraint on power level, we can find the matrices \mathbf{A}_l that minimize the power level subject to an equality constraint on MSE. And if the constraint value on MSE is such that the resulting minimum power level matches the given value P in (2.7), the corresponding matrices \mathbf{A}_l are the optimal ones we set out to find. We thus consider the following optimization problem

$$\begin{cases} \min_{\mathbf{A}_l, 1 \leq l \leq L} & \sum_{l=1}^L \text{tr}(\sigma_\theta^2 \mathbf{A}_l \mathbf{f}_l \mathbf{f}_l^T \mathbf{A}_l^T + \sigma_n^2 \mathbf{A}_l \mathbf{A}_l^T) \\ \text{subject to} & \left(\frac{1}{\sigma_\theta^2} + \frac{\left(\sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l \right)^2}{\sigma_n^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{A}_l^T \mathbf{g}_l + \sigma_\nu^2} \right)^{-1} = J^* \end{cases} \quad (2.8)$$

where $0 < J^* \leq \sigma_\theta^2$. We note that both the objective function and the constraint function in (2.8) are quadratic in the elements of \mathbf{A}_l . This problem is considerably easier to solve than the original one (2.7). The main result based on solving (2.8) is in the following proposition whose proof is given in Appendix A.

Proposition 2.1. *Consider the sensor network model described by (2.2) and (2.3). Suppose the total transmitted power from all sensors is no greater than P , then using the LMMSE estimator, the optimal amplification matrix of the l th cluster is given by*

$$\mathbf{A}_l^{\text{opt}} = \sqrt{\left(\sum_{i=1}^L \frac{\|\mathbf{f}_i\|^2 \|\mathbf{g}_i\|^2 (\sigma_\theta^2 \|\mathbf{f}_i\|^2 + \sigma_n^2)}{\phi_i^2} \right)^{-1} \frac{P}{\phi_l^2} \mathbf{g}_l \mathbf{f}_l^T}, \quad l = 1, \dots, L \quad (2.9)$$

where $\phi_i = \sigma_\nu^2 (\sigma_\theta^2 \|\mathbf{f}_i\|^2 + \sigma_n^2) + \sigma_n^2 \|\mathbf{g}_i\|^2 P$, and the corresponding minimum MSE is

$$J_M = \left(\frac{1}{\sigma_\theta^2} + \sum_{l=1}^L \frac{\|\mathbf{f}_l\|^2 \|\mathbf{g}_l\|^2 P}{\sigma_\nu^2 (\sigma_\theta^2 \|\mathbf{f}_l\|^2 + \sigma_n^2) + \sigma_n^2 \|\mathbf{g}_l\|^2 P} \right)^{-1} \quad (2.10)$$

The optimal amplification matrix $\mathbf{A}_l^{\text{opt}}$ is a rank one matrix, which is a scaled outer product of \mathbf{g}_l and \mathbf{f}_l . As expected the optimal MSE J_M decreases as P increases. Moreover, as $P \rightarrow \infty$, we have

$$\lim_{P \rightarrow \infty} J_M = \frac{\sigma_\theta^2}{1 + (\sigma_\theta^2 / \sigma_n^2) \sum_{l=1}^L \|\mathbf{f}_l\|^2} \quad (2.11)$$

The limit does not go to zero but approaches a finite value which depends on the signal to noise ratio $\sigma_\theta^2 \sum_{l=1}^L \|\mathbf{f}_l\|^2 / \sigma_n^2$, since the measured signal $\mathbf{f}_l \theta + \mathbf{n}_l$ is amplified by \mathbf{A}_l , $1 \leq l \leq L$.

For comparison, we consider two special cases: $L = 1$ and $L = K$. When $L = 1$, there is full collaboration among the K sensors. The observation gain is $\mathbf{f} \in \mathbb{R}^K$ and the channel gain is $\mathbf{g} \in \mathbb{R}^N$, $N \leq K$. It is easy to see from (2.9) that the optimal amplification matrix is

$$\mathbf{A}^{opt} = \sqrt{\left(\frac{\|\mathbf{f}\|^2 \|\mathbf{g}\|^2 (\sigma_\theta^2 \|\mathbf{f}\|^2 + \sigma_n^2)}{\bar{\phi}^2}\right)^{-1} \frac{P}{\bar{\phi}^2} \mathbf{g} \mathbf{f}^T} \quad (2.12)$$

where $\bar{\phi} = \sigma_\nu^2 (\sigma_\theta^2 \|\mathbf{f}\|^2 + \sigma_n^2) + \sigma_n^2 \|\mathbf{g}\|^2 P$, and the corresponding MSE is

$$J_C = \left(\frac{1}{\sigma_\theta^2} + \frac{\|\mathbf{f}\|^2 \|\mathbf{g}\|^2 P}{\sigma_\nu^2 (\sigma_\theta^2 \|\mathbf{f}\|^2 + \sigma_n^2) + \sigma_n^2 \|\mathbf{g}\|^2 P} \right)^{-1} \quad (2.13)$$

When $L = K$, each sensor is a cluster and no collaboration between sensor exists. The scalar observation gains and channel gains are respectively f_k and g_k , $1 \leq k \leq K$. Again from (2.9), the optimal amplification gain is

$$a_k^{opt} = \sqrt{\left(\sum_{i=1}^K \frac{f_i^2 g_i^2 (\sigma_\theta^2 f_i^2 + \sigma_n^2)}{\tilde{\phi}_i^2}\right)^{-1} \frac{P}{\tilde{\phi}_k^2} g_k f_k^T}, \quad k = 1, \dots, K \quad (2.14)$$

where $\tilde{\phi}_i = \sigma_\nu^2 (\sigma_\theta^2 f_i^2 + \sigma_n^2) + \sigma_n^2 g_i^2 P$, and the corresponding MSE is

$$J_N = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^K \frac{f_k^2 g_k^2 P}{\sigma_\nu^2 (\sigma_\theta^2 f_k^2 + \sigma_n^2) + \sigma_n^2 g_k^2 P} \right)^{-1} \quad (2.15)$$

To compare the performance of the general case and the two special cases, we assume $N_l = K_l$, $1 \leq l \leq L$, in (2.10) and $N = K$ in (2.13), that is, the number of measurements is equal to the number of transmitters in each cluster. Hence the observation gain and the channel gain vectors can be written as $\mathbf{f} = [\mathbf{f}_1^T \mathbf{f}_2^T \cdots \mathbf{f}_L^T]^T = [f_1 f_2 \cdots f_K]^T$ and $\mathbf{g} = [\mathbf{g}_1^T \mathbf{g}_2^T \cdots \mathbf{g}_L^T]^T = [g_1 g_2 \cdots g_K]^T$, respectively, where $\mathbf{f}_l, \mathbf{g}_l \in \mathbb{R}^{K_l}$ and $K_1 + \cdots + K_L = K$. In terms of the MSE, it is not unexpected that collaboration improves performance. Indeed, we have the following proposition.

Proposition 2.2. *The minimum MSEs J_M in (2.10), J_C in (2.13), and J_N in (2.15) satisfy*

$$J_C \leq J_M \leq J_N. \quad (2.16)$$

The proof of Proposition 2.2 is based on the following lemma.

Lemma 2.3. For $\mathbf{x} = [\mathbf{x}_1^T \mathbf{x}_2^T \cdots \mathbf{x}_L^T]^T \in \mathbb{R}^n$ and $\mathbf{y} = [\mathbf{y}_1^T \mathbf{y}_2^T \cdots \mathbf{y}_L^T]^T \in \mathbb{R}^n$, where \mathbf{x}_i and \mathbf{y}_i are nonzero vectors of dimension ≥ 1 , we have

$$\frac{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \geq \sum_{i=1}^L \frac{\|\mathbf{x}_i\|^2 \|\mathbf{y}_i\|^2}{\|\mathbf{x}_i\|^2 + \|\mathbf{y}_i\|^2}. \quad (2.17)$$

Proof. Please see Appendix B. □

We now establish Proposition 2.2. Let $\mathbf{x} = \sqrt{\sigma_\theta^2} \mathbf{f}$, $\mathbf{y} = \sqrt{(\sigma_n^2/\sigma_\nu^2)P} \mathbf{g}$, $\mathbf{x}_i = \sqrt{\sigma_\theta^2} \mathbf{f}_i$, and $\mathbf{y}_i = \sqrt{(\sigma_n^2/\sigma_\nu^2)P} \mathbf{g}_i$. Then by Lemma 2.3, we get

$$\frac{\|\mathbf{f}\|^2 \|\mathbf{g}\|^2}{\sigma_\theta^2 \|\mathbf{f}\|^2 + (\sigma_n^2/\sigma_\nu^2)P \|\mathbf{g}\|^2 + \sigma_n^2} \geq \sum_{i=1}^L \frac{\|\mathbf{f}_i\|^2 \|\mathbf{g}_i\|^2}{\sigma_\theta^2 \|\mathbf{f}_i\|^2 + (\sigma_n^2/\sigma_\nu^2)P \|\mathbf{g}_i\|^2 + \sigma_n^2}$$

and thus $J_C^{-1} \geq J_M^{-1}$, or equivalently, $J_C \leq J_M$. The second inequality in (2.16) follows similarly: apply Lemma 2.3 to each \mathbf{x}_i , \mathbf{y}_i , and their respective scalar components, and their sums give the desired inequality.

2.3 Numerical Results and Discussion

In this section, we use numerical simulations to verify the analytical result established in Section 2.2. In all simulations, the random parameters, θ , $n_{l,k}$, and ν , are zero-mean Gaussian. And we assume that $\sigma_\theta^2 = \sigma_\nu^2 = 1$. The observation gains $f_{l,k}$ are assumed to be uniformly distributed in the interval $[0.5, 1]$. The channel gains are taken as $c_g d^{-3.5}$, where d is uniformly drawn from the interval $[1, 10]$ and $c_g = 22.6$ is a normalization constant to make $E[g_{l,n}] = 1$ as in [26].

Simulation 2.1 - Effects of N : In this simulation, we demonstrate the effect of the number of transmitters. In particular, we consider the effect of different numbers of transmitters N , where $N \leq K$, in the full collaboration case. We set $K = 10$ and $\sigma_n^2 = 0.4$. In Figure 2.2, we plot the average MSE versus N with power levels $P = 0$ dB, 5 dB, and 10 dB. We note that as N increases, the MSEs decrease; also, we see that large power levels result in smaller MSEs.

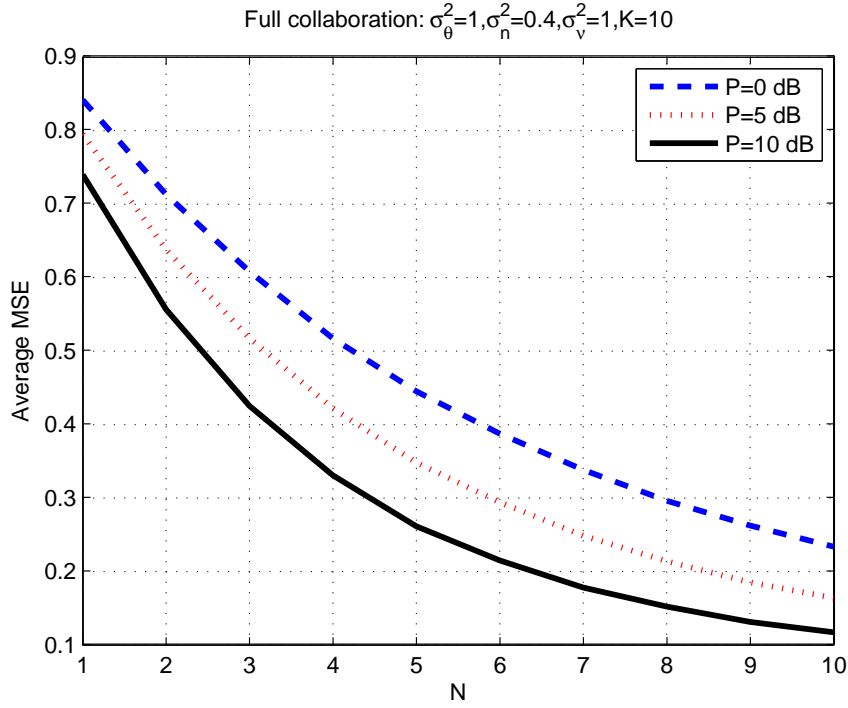


Figure 2.2: MSE of full collaboration case with different numbers of transmitters

Since the MSE decreases as N increases, in all the simulations to follow, the number of sensors and the number of transmitters are set equal.

Simulation 2.2 - Effects of Observation Noise Powers: In this simulation, we compare the MSE of the full collaboration case to that of the non-collaboration case for different observation noise powers. Figure 2.3 shows the average MSE versus P for the full collaboration and non-collaboration cases with different observation noises, $\sigma_n^2 = 0.4$ and $\sigma_n^2 = 0.8$. We set $K = 20$. For $\sigma_n^2 = 0.4$, the case with full collaboration performs better than the non-collaboration case. Moreover, as the transmitted power increases, the MSEs for both two cases decrease. In fact, from (2.11), these two cases approach identical MSE as $P \rightarrow \infty$. We also see that the MSE of the case with $\sigma_n^2 = 0.4$ is smaller than that of the case with $\sigma_n^2 = 0.8$, that is, a large signal to noise ratio results in a good performance.

Simulation 2.3 - Comparison between Full Collaboration and Non-collaboration: In this simulation, we compare the full collaboration case with two non-collaboration cases,

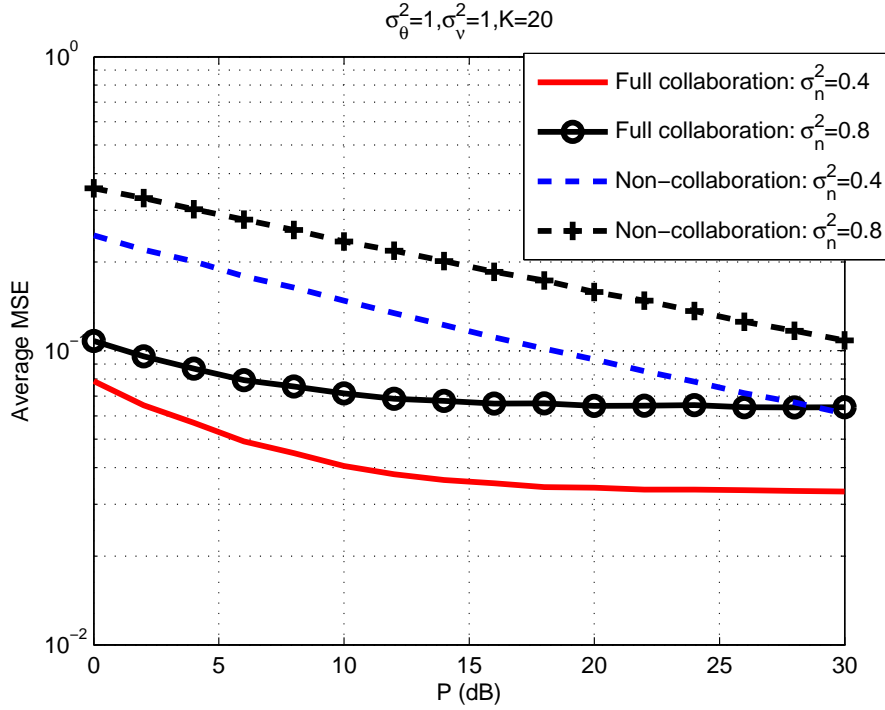


Figure 2.3: MSE of full collaboration and non-collaboration cases with different noise power levels

which, respectively, use the optimal power allocation scheme and the equal power allocation scheme. Note that the equal power allocation scheme is to set the amplification gains as $a_k = \sqrt{P/K}$, $1 \leq k \leq K$. We set $K = 50$ and $\sigma_n^2 = 0.4$. Figure 2.4 shows that optimal power allocation improves performance over equal power allocation for the non-collaboration case. In addition, the reduction in MSE by full collaboration with optimal power allocation is about 10 dB compared with the equal power allocation scheme.

Simulation 2.4 - Effects of Number of Sensors in Each Cluster: In this simulation, we demonstrate the effect of the number of sensors in each cluster. Specifically, we fix the number of sensors in the network and consider two multiple cluster cases: in case 1, each cluster consists of 4 sensors, and in case 2, each cluster consists of 8 sensors. Hence for a fixed number of sensors K , case 1 has $K/4$ clusters and case 2 has $K/8$ clusters. We compare their performance with the full collaboration and non-collaboration cases. We set $P = 0$ dB and $\sigma_n^2 = 0.4$. Figure 2.5 shows that MSE of case 2 is less than that of case 1

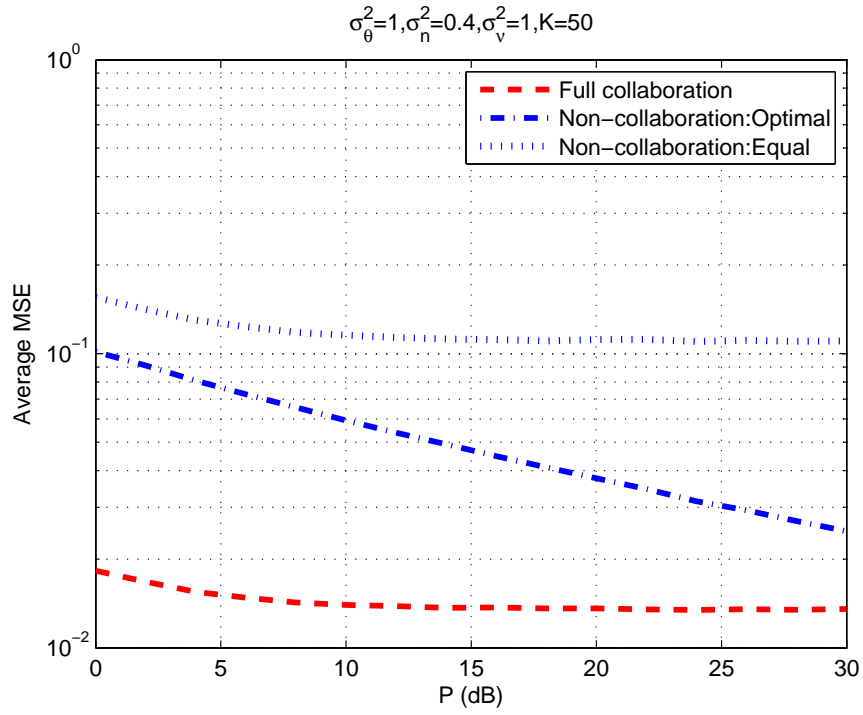


Figure 2.4: MSE of full collaboration and non-collaboration cases with different power levels

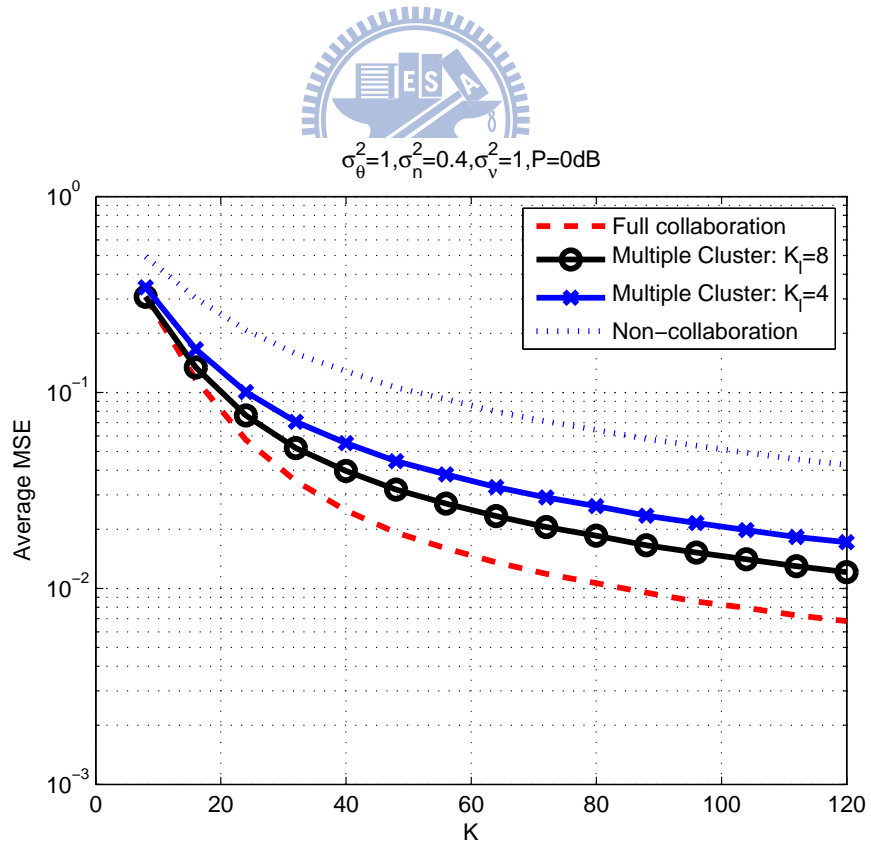


Figure 2.5: MSEs for $K_l = 1, 4, 8$, and K with different number of sensors

since for a fixed K , case 2 has a smaller number of clusters and thus more collaboration among sensors. That the full collaboration has the lowest MSE and the non-collaboration case has the highest MSE is as predicted by (2.16).

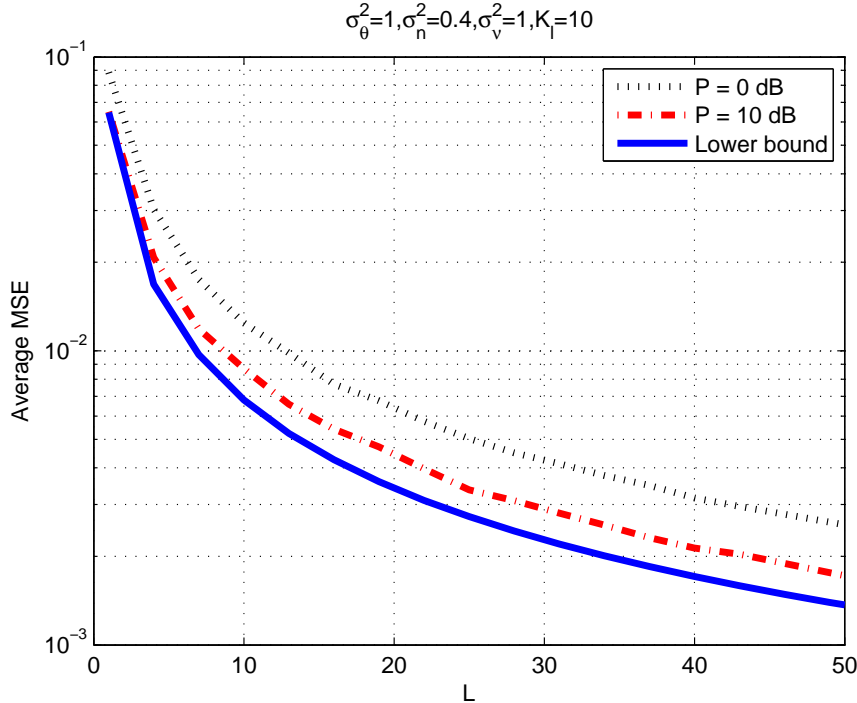


Figure 2.6: MSEs with different number of clusters

Simulation 2.5 - Effects of Cluster Number: In this simulation, we fix the number of sensors $K_l = 10$ in each cluster and investigate the relationship between the MSE and the number of cluster. Hence, as the number of clusters is L , the sensors in the network are $10 \times L$. Figure 2.6 shows the MSE as a function of L with, respectively, $P = 0$ dB, $P = 10$ dB, and $P = \infty$, which is equivalent to the performance lower bound as shown in (2.11). From this figure, we see that when the number of sensors in each cluster is fixed, the MSE decreases as L increases. In addition, compared the MSEs of $P = 0$ dB to that of $P = 10$ dB, we see that the performance gain for the increase of power approaches a constant as $L > 10$.

Simulation 2.6 - Comparison with Orthogonal MAC Model [24]: In this simulation, we

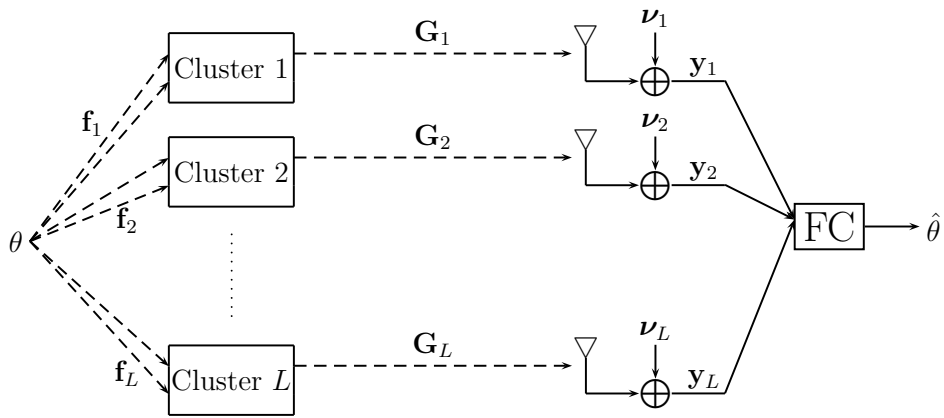


Figure 2.7: Cluster-based sensor network with orthogonal MAC

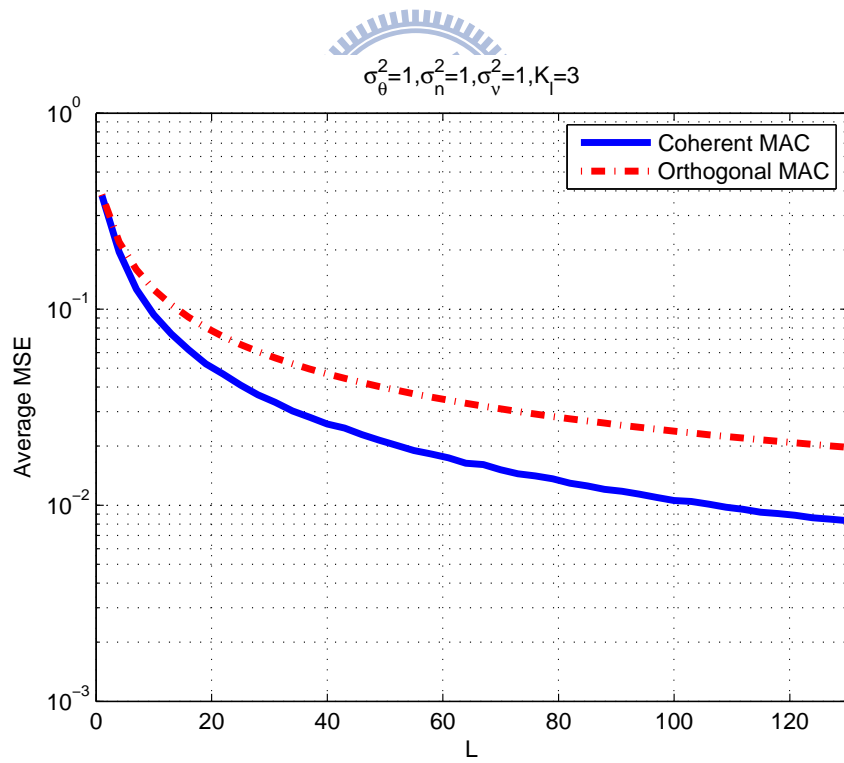


Figure 2.8: Comparison of the coherent MAC model to that of the orthogonal MAC model

compare the proposed scheme to that in [24], in which the orthogonal MAC model is taken into account. Figure 2.7 shows the model in [24]: the measurement vector for the l th cluster is $\mathbf{x}_l = \mathbf{f}_l\theta + \mathbf{n}_l$ which then transmits to the l th receiver through a diagonal channel gain matrix \mathbf{G}_l after multiplying by an amplification matrix \mathbf{A}_l ; at the FC, the received signal vector from the l th cluster is $\mathbf{y}_l = \mathbf{G}_l\mathbf{A}_l\mathbf{x}_l + \boldsymbol{\nu}_l$, $1 \leq l \leq L$, where the additive noise $\boldsymbol{\nu}_l$ is assumed to be $E[\boldsymbol{\nu}_l] = \mathbf{0}$, $E[\boldsymbol{\nu}_l\boldsymbol{\nu}_l^T] = \sigma_{\nu}^2\mathbf{I}_{K_l}$, and $E[\boldsymbol{\nu}_l\boldsymbol{\nu}_j^T] = \mathbf{0}_{K_l \times K_j}$ for $j \neq l$. After collecting L signal vectors at the FC, the LMMSE fusion rule is used for estimating the source signal. The performance comparison for the orthogonal and coherent MAC models is plotted in Figure 2.8, in which we take $P = 10$ dB, $K_l = 3$ for all clusters, and $\sigma_n^2 = 1$. We see that the MSE of the coherent MAC model performs better than that of the orthogonal MAC model. This is because by using the orthogonal MAC model, the number of receiver noises increases as the number of clusters increases. However, by using the coherent MAC model, there is only one receiver noise regardless of the number of clusters.

Table 2.1: Different number of sensors in clusters

	$K = 30$ $P = 0$ dB									
Clusters (L)	4	5	5	6	6	6	7	7	8	9
Sensors in each cluster (K_l)	9, 9, 9, 3	10, 8, 7, 3, 2	9, 9, 6, 4, 2	10, 8, 6, 4, 1, 1	10, 7, 5, 4, 3, 1	8, 7, 6, 5, 3, 1	9, 7, 6, 3, 2, 2, 1	9, 6, 4, 4, 3, 2, 2	8, 7, 5, 3, 3, 2, 1, 1	6, 6, 4, 3, 3, 3, 2, 2, 1
Number of entries	252	226	218	218	200	184	184	166	162	124
MSE (J_M)	0.0566	0.0603	0.0617	0.0635	0.0642	0.0659	0.0687	0.0690	0.0704	0.0790

Simulation 2.7 - Relation between Collaboration: In this simulation, we see quantitatively the relation between collaboration and MSE for the coherent MAC model, we consider a network with 30 sensors and $P = 0$ dB. We perform 10 simulations with the number of cluster ranging from 4 to 9. The number of sensors in each cluster is randomly chosen from 1 to 10. In each case, we count the total number of entries in the amplification matrices \mathbf{A}_l . For example, in the first case there are 4 clusters, the numbers of sensors in the clusters are respectively 9, 9, 9, and 3, and the number of entries is $9^2 + 9^2 + 9^2 + 3^2 = 252$. Table 1 shows the number of cluster, the number of entries, and the corresponding MSE

for each case. From the table, we see that the MSE decreases as the number of entries increases. For comparison, the MSE for the two special cases are respectively $J_N = 0.1689$ and $J_C = 0.0392$.

A Brief Summary and Discussion: We study optimal collaboration for distributed estimation in cluster-based wireless sensor network. We show that the optimal amplification matrix of each cluster is a rank one matrix obtained as a scaled outer product of the observation gain and the channel gain vectors. We also show that with optimal collaboration matrices, the performance of the collaboration case is better than that of the non-collaboration case. For a fixed number of sensors in the network, we demonstrate, through simulation results, that the amount of improvement is closely related to the amount of collaboration.



Chapter 3

Distributed Estimation Using Sensor Network with Unknown Channels

In wireless communication systems, channels are often unknown to the receiver and have to be estimated in practice [32, 34]. In this chapter, we extend the non-collaboration case when channels are known to the case when channels are unknown. In the sensor network with unknown channels, we use training-based LMMSE channel estimator. The channel estimates are then used to obtain LMMSE estimation of the source signal. The problem is formulated as the allocation of power to each sensor for training and for data transmission under a total network power constraint. We first consider the optimal power allocation scheme, in which the sensors get full channel or estimated channel information and consider optimizing the training power and data power for each sensor. Then we discuss the equal power allocation scheme, in which the sensors do not have channel information, except for the phases. We set equal data power to each sensor and consider the optimization of training power to minimize the MSE. In each scheme, we compare the performance of the distributed estimation with estimated channels to that with actual known channels.

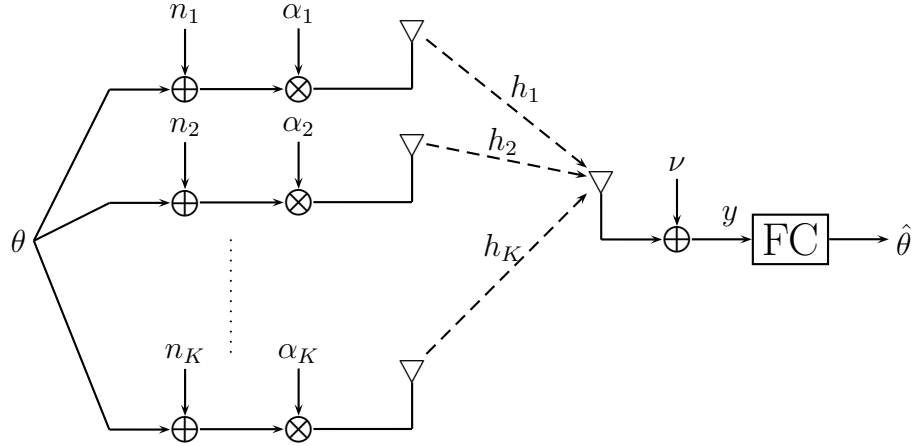


Figure 3.1: Coherent MAC wireless sensor network

3.1 System Model

We consider a wireless sensor network with K sensors for estimating a random source signal θ , as depicted in Figure 3.1. The measurement at the k th sensor is corrupted by an additive noise n_k and amplified by a factor α_k before it is transmitted to the FC through a flat fading channel, h_k . The signal y received at the FC can be expressed as

$$y = \sum_{k=1}^K h_k \alpha_k (\theta + n_k) + \nu \quad (3.1)$$

where ν is the additive noise at the receiver.

The following assumptions are made throughout this chapter.

- i) $E[\theta] = 0$ and $E[|\theta|^2] = \sigma_\theta^2$.
- ii) The measurement noises are independent and $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ for $k = 1, 2, \dots, K$, that is, n_k 's are independent and circular Gaussian with zero mean and variance σ_n^2 .
- iii) The channels are independent and $h_k \sim \mathcal{CN}(0, \sigma_h^2)$.
- iv) $\nu \sim \mathcal{CN}(0, \sigma_\nu^2)$.
- v) The source signal, the channels, the measurement noises, and the receiver noise are uncorrelated. Specifically, for $1 \leq k, j \leq K$, $E[\theta^* n_k] = 0$, $E[n_k^* h_l] = 0$, $E[\theta^* \nu] = 0$, $E[\theta^* h_k] = 0$, and $E[h_k^* \nu] = 0$.

The problem is to estimate the parameter θ based on the received signal y at the FC. The fading channels are assumed unknown. We consider a two-phase approach similar to that proposed in [25]: to estimate the channels first using training symbols sent from the sensors and then to estimate θ based on the estimated channels and y . In both phases, we seek the LMMSE estimator.

3.2 LMMSE Estimation

3.2.1 Channel Estimation

During the training phase, the sensors send training symbols in sequence: the training period is divided into K time intervals and only the k th sensor sends a training symbol t_k over the k th time interval. Thus, the received signal at the k th time interval can be expressed as $y_k = h_k t_k + \nu_k$, $k = 1, 2, \dots, K$, where $\nu_k \sim \mathcal{CN}(0, \sigma_\nu^2)$ and $E[\nu_i^* \nu_j] = 0$ for $i \neq j$. For a given training sequence t_k , the LMMSE estimator of h_k is given by

$$\hat{h}_k = \frac{\sigma_h^2}{|t_k|^2 \sigma_h^2 + \sigma_\nu^2} t_k^* y_k \quad (3.2)$$

and the corresponding MSE is

$$\delta_k^2 = E \left[|h_k - \hat{h}_k|^2 \right] = \frac{\sigma_h^2 \sigma_\nu^2}{|t_k|^2 \sigma_h^2 + \sigma_\nu^2}, \quad k = 1, \dots, K. \quad (3.3)$$

The MSE of \hat{h}_k decreases as the power of the training symbol $|t_k|^2$ increases. The LMMSE problem under the training power constraint $\sum_{k=1}^K |t_k|^2 \leq P_t$ can be formulated as

$$\min_{p_k: 1 \leq k \leq K} \frac{1}{K} \sum_{k=1}^K \delta_k^2, \quad \text{subject to} \quad \sum_{k=1}^K p_k \leq P_t \quad \text{and} \quad p_k \geq 0, \quad k = 1, \dots, K$$

where $p_k = |t_k|^2$ is the training power of sensor k . The problem can be solved using standard Karush-Kuhn-Tucker (KKT) condition [25] and the solution is $|t_k|^2 = P_t/K$, $\forall k$, as expected since the channels are independent and identically distributed. In particular, we choose the training symbol to be real and positive, that is, $t_k = \sqrt{P_t/K}$, and the resulting channel estimate is

$$\hat{h}_k = \frac{\sigma_h^2 \sqrt{K P_t}}{\sigma_h^2 P_t + K \sigma_\nu^2} y_k, \quad k = 1, 2, \dots, K \quad (3.4)$$

with the corresponding MSE

$$\delta_k^2 = \frac{K\sigma_h^2\sigma_\nu^2}{\sigma_h^2 P_t + K\sigma_\nu^2}, \quad k = 1, \dots, K. \quad (3.5)$$

We note that with such choice of training symbols, both the received signal y_k and the channel estimate \hat{h}_k are circular Gaussian.

3.2.2 Source Estimation

During the second phase, channel estimates \hat{h}_k are available at the FC, although the actual channels are unknown. We express the received signal y in (3.1) in terms of \hat{h}_k as

$$y = \sum_{k=1}^K \hat{h}_k \alpha_k \theta + \sum_{k=1}^K \hat{h}_k \alpha_k n_k + \varepsilon + \nu \quad (3.6)$$

where $\varepsilon = \sum_{k=1}^K (h_k - \hat{h}_k) \alpha_k (\theta + n_k)$ is contributed by channel estimation error. Let $\hat{\mathbf{h}} = [\hat{h}_1 \hat{h}_2 \dots \hat{h}_K]^T$ be the vector of channel estimates. The LMMSE estimate of θ given $\hat{\mathbf{h}}$ is

$$\hat{\theta} = ay, \quad \text{where } a = \frac{E[\theta y^* | \hat{\mathbf{h}}]}{E[|y|^2 | \hat{\mathbf{h}}]}. \quad (3.7)$$

From (3.6) it follows that

$$\begin{aligned} E[\theta y^* | \hat{\mathbf{h}}] &= E\left[\theta \left(\sum_{k=1}^K \hat{h}_k^* \alpha_k^* \theta^* + \sum_{k=1}^K \hat{h}_k^* \alpha_k^* n_k^* + \varepsilon^* + \nu^* \right) \middle| \hat{\mathbf{h}}\right] \\ &= \sum_{k=1}^K \hat{h}_k^* \alpha_k^* \sigma_\theta^2 \end{aligned} \quad (3.8)$$

where the last equality is from the assumptions that the source signal is uncorrelated with the measurement noise and the receiver noise, and that $\hat{h}_k = E[h_k | y_k] = E[h_k | \hat{h}_k]$ since \hat{h}_k is a linear function of y_k . It is derived in Appendix C that

$$E[|y|^2 | \hat{\mathbf{h}}] = \left| \sum_{k=1}^K \hat{h}_k \alpha_k \right|^2 \sigma_\theta^2 + \sum_{k=1}^K |\hat{h}_k|^2 |\alpha_k|^2 \sigma_n^2 + \sigma_\nu^2 + (\sigma_\theta^2 + \sigma_n^2) \delta_1^2 \sum_{k=1}^K |\alpha_k|^2. \quad (3.9)$$

The MSE incurred by (3.7) is

$$\begin{aligned} J &= E[|\theta - \hat{\theta}|^2 | \hat{\mathbf{h}}] \\ &= \sigma_\theta^2 - a E[\theta^* y | \hat{\mathbf{h}}] - a^* E[y^* \theta | \hat{\mathbf{h}}] + |a|^2 E[|y|^2 | \hat{\mathbf{h}}] \\ &= \left(\frac{1}{\sigma_\theta^2} + \frac{\left| \sum_{k=1}^K \hat{h}_k \alpha_k \right|^2}{\sum_{k=1}^K |\hat{h}_k|^2 |\alpha_k|^2 \sigma_n^2 + \sigma_\nu^2 + (\sigma_\theta^2 + \sigma_n^2) \delta_1^2 \sum_{k=1}^K |\alpha_k|^2} \right)^{-1} \end{aligned} \quad (3.10)$$

When the channel h_k is available at the FC, we can set $\hat{h}_k = h_k$ and $\delta_k^2 = 0$ in (3.10), and the corresponding MSE becomes

$$J_o = \left(\frac{1}{\sigma_{\hat{\theta}}^2} + \frac{\left| \sum_{k=1}^K h_k \alpha_k \right|^2}{\sum_{k=1}^K |h_k|^2 |\alpha_k|^2 \sigma_n^2 + \sigma_v^2} \right)^{-1} \quad (3.11)$$

The MSE J_o is a lower bound of J in (3.10) and can serve as a benchmark against which the performance of the estimator (3.7) can be compared.

3.3 Optimal Power Allocation

During the training phase, each sensor uses the same training symbol and thus consumes the same amount of training power P_t/K , where P_t is the total allocated training power. From (3.5), it is clear that as P_t increases, the MSE in channel estimation decreases. In a sensor network, there is likely a total power constraint, that is, there is an upper bound imposed on the sum of training power and the power used to transmit data. Hence, when more power is allocated for training, less power is available for data transmission and vice versa. Under the total power constraint, the minimum MSE of $\hat{\theta}$, that is, J in (3.10), depends on the training power P_t and how the remaining network power is allocated to each sensor for data transmission. In the following, we consider the optimal power allocation problem, that is, to choose P_t and data power for each sensor to minimize J under a total power constraint. For comparison, we will also consider the case when channel information is available, no training, no channel error, and all power is used for data transmission. The comparison of the two cases will show the penalty incurred due to the fact that the channel is unknown.

3.3.1 When channels are known

If the channel h_k is available at the sensor k , the phase of α_k can be chosen to match that of the channel, that is, $\angle \alpha_k = -\angle h_k$, so that $h_k \alpha_k = |h_k| |\alpha_k|$. The MSE J_o in (3.11) can

then be rewritten as

$$J_o = \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta \left(\sum_{k=1}^K g_k |\alpha_k| \right)^2}{\zeta \sigma_n^2 \left(\sum_{k=1}^K g_k^2 |\alpha_k|^2 \right) + 1} \right)^{-1} \quad (3.12)$$

where $\zeta = \sigma_h^2/\sigma_v^2$ is the channel SNR, and $g_k = |h_k|/\sigma_h$ is the normalized channel gain for the k th sensor. Such choices of phases make J_o smallest among α_k 's of the same magnitude. Note that g_k has a Rayleigh distribution with density function $f_g(x) = 2x \exp(-x^2)$, $x \geq 0$, and $E[g_k] = \sqrt{\pi/4}$; g_k^2 has an exponential distribution with density function $f_{g^2}(x) = \exp(-x)$, $x \geq 0$, and $E[g_k^2] = 1$ [37, p.51]. The signal transmitted from the k th sensor is $\alpha_k(\theta + n_k)$ with power $P_k = E[|\alpha_k(\theta + n_k)|^2] = |\alpha_k|^2(\sigma_\theta^2 + \sigma_n^2)$. From (3.12), the optimal power allocation problem with the total network power constrained to $P > 0$ can be formulated as the following optimization problem

$$\begin{cases} \min_{|\alpha_k|: 1 \leq k \leq K} & \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta \left(\sum_{k=1}^K g_k |\alpha_k| \right)^2}{\zeta \sigma_n^2 \left(\sum_{k=1}^K g_k^2 |\alpha_k|^2 \right) + 1} \right)^{-1} \\ \text{subject to} & \sum_{k=1}^K |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2) \leq P. \end{cases} \quad (3.13)$$

By the same reasons shown in the observations (i) to (iii) in Section 2.2 in Chapter 2, instead of solving (3.13), we consider an alternative problem in which the power is minimized subject to an MSE constraint:

$$\begin{cases} \min_{|\alpha_k|: 1 \leq k \leq K} & \sum_{k=1}^K |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2) \\ \text{subject to} & \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta \left(\sum_{k=1}^K g_k |\alpha_k| \right)^2}{\zeta \sigma_n^2 \left(\sum_{k=1}^K g_k^2 |\alpha_k|^2 \right) + 1} \right)^{-1} = J_o \end{cases} \quad (3.14)$$

where $0 < J_o \leq \sigma_\theta^2$. It is noted that the problem (3.14) can be regarded as the non-collaboration case of the problem (2.8) with \mathbf{A}_l , \mathbf{g}_l , and \mathbf{f}_l replaced by $|\alpha_k|$, g_k , and 1. The solution, from Proposition 2.1, can then be written as

$$|\alpha_k|^2 = \left(\sum_{k=1}^K \frac{g_k^2 (\sigma_\theta^2 + \sigma_n^2)}{[(\sigma_\theta^2 + \sigma_n^2) + \sigma_n^2 \zeta g_k^2 P]^2} \right)^{-1} \frac{g_k^2 P}{[(\sigma_\theta^2 + \sigma_n^2) + \sigma_n^2 \zeta g_k^2 P]^2} \quad (3.15)$$

and thus the optimal power is $P_k = |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2)$, $1 \leq k \leq K$; the corresponding MSE is

$$J_o = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^K \frac{g_k^2}{\sigma_n^2 g_k^2 + (\sigma_\theta^2 + \sigma_n^2) \frac{1}{\zeta P}} \right)^{-1} \quad (3.16)$$

which is also the minimum MSE of (3.13). Since the minimum MSE depends on the total network power P and the number of sensors K , we hereafter write the MSE J_o in (3.16) as $J_o(P, K)$.

As the power P increases, we expect J_o to decrease, which is easy to see from (3.16). For a fixed K , as $P \rightarrow \infty$, we have

$$\lim_{P \rightarrow \infty} J_o(P, K) = \frac{\sigma_\theta^2}{1 + K\beta} \quad (3.17)$$

where $\beta = \sigma_\theta^2/\sigma_n^2$ is the observation SNR. The limit does not go to zero but is roughly proportional to $1/K$ as we would expect. On the other hand, for a fixed $P > 0$, as K increases, we have

$$\begin{aligned} \lim_{K \rightarrow \infty} J_o(P, K) &= \left(\frac{1}{\sigma_\theta^2} + \lim_{K \rightarrow \infty} KE \left[\frac{g_k^2}{\sigma_n^2 g_k^2 + (\sigma_\theta^2 + \sigma_n^2) \frac{1}{\zeta P}} \right] \right)^{-1} \\ &= \lim_{K \rightarrow \infty} \frac{1}{K} \left\{ E \left[\frac{\zeta g_k^2 P}{(\sigma_\theta^2 + \sigma_n^2) + \sigma_n^2 \zeta g_k^2 P} \right] \right\}^{-1} = 0 \end{aligned} \quad (3.18)$$

where in the first equality we used the law of large numbers [39]. From (3.18), we conclude that in the coherent MAC model, the MSE decreases in the order of $1/K$ as K goes to infinity even though the total network power P is finite. Similar conclusion for the unit variance case, $\sigma_\theta^2 = \sigma_n^2 = \sigma_\nu^2 = 1$, appeared in [26].

3.3.2 When channels are estimated

Suppose training for channel estimation consumes power P_t , then the remaining power for data transmission is $P - P_t$. The power allocation problem now is to optimally choose training power P_t and data power for each sensor. The phase of α_k is chosen to match that of \hat{h}_k , i.e., $\angle \alpha_k = -\angle \hat{h}_k$. Write $h_k = \hat{h}_k + (h_k - \hat{h}_k)$, since \hat{h}_k and $h_k - \hat{h}_k$ are uncorrelated we have $\sigma_h^2 = \sigma_{\hat{h}}^2 + \delta_k^2$, where $\delta_k^2 = E[|h_k - \hat{h}_k|^2]$. Use (3.5) and $\sigma_h^2 = \sigma_{\hat{h}}^2 + \delta_k^2$, we can express the MSE in (3.10) as

$$J = \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta^2 \left(\sum_{k=1}^K \hat{g}_k |\alpha_k| \right)^2 P_t}{\zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2 \right) P_t + \zeta P_t + K \zeta (\sigma_\theta^2 + \sigma_n^2) \left(\sum_{k=1}^K |\alpha_k|^2 \right) + K} \right)^{-1} \quad (3.19)$$

where $\hat{g}_k = |\hat{h}_k|/\sigma_{\hat{h}}$ is the normalized estimated channel gain for the k th sensor. Since \hat{h}_k is circular Gaussian, \hat{g}_k and g_k have identical distribution. From (3.19), the MMSE

optimization problem under a total network power constraint can be formulated as

$$\begin{cases} \min_{P_t; |\alpha_k|: 1 \leq k \leq K} & \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta^2 (\sum_{k=1}^K \hat{g}_k |\alpha_k|)^2 P_t}{\zeta^2 \sigma_n^2 (\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2) P_t + \zeta P_t + K \zeta (\sigma_\theta^2 + \sigma_n^2) (\sum_{k=1}^K |\alpha_k|^2) + K} \right)^{-1} \\ \text{subject to} & \sum_{k=1}^K |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2) + P_t \leq P. \end{cases} \quad (3.20)$$

Again instead of solving problem (3.20) directly, we consider a problem in which the roles of objective function and constraint are interchanged. The solution to problem (3.20) is given in the following proposition, the proof of which is given in Appendix D.

Proposition 3.1. *For $K > 1$, the solution to (3.20) gives the optimal training power*

$$P_t^{opt} = \frac{K(\zeta P + 1) - \sqrt{K(\zeta P + 1)(\zeta P + K)}}{\zeta(K - 1)} \quad (3.21)$$

where $\zeta = \sigma_h^2 / \sigma_v^2$ is the channel SNR, and the associated optimal data power for the k th sensor is

$$P_k^{opt} = \left(\sum_{k=1}^K \hat{g}_k^2 / \hat{\phi}_k^2 \right)^{-1} \frac{\hat{g}_k^2}{\hat{\phi}_k^2} (P - P_t^{opt}) \quad (3.22)$$

where $\hat{\phi}_k = [(\sigma_\theta^2 + \sigma_n^2) + \lambda_0 \zeta^2 \sigma_n^2 \hat{g}_k^2 P_t^{opt} + \lambda_0 K \zeta (\sigma_\theta^2 + \sigma_n^2)]$ and $\lambda_0 = P_t^{opt} / (K + K \zeta (P - P_t^{opt}))$. The incurred MSE is

$$J(P, K) = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^K \frac{\hat{g}_k^2}{\sigma_n^2 \hat{g}_k^2 + (\sigma_\theta^2 + \sigma_n^2) \left(\frac{K-1}{\sqrt{K(\zeta P + 1)} - \sqrt{\zeta P + K}} \right)^2} \right)^{-1} \quad (3.23)$$

Note that the optimal training power in (3.21) depends on the number of sensors K , the channel SNR ζ , and the total network power P . Moreover, we have

$$P_t^{opt} - \frac{P}{2} = \frac{\left(\sqrt{K(\zeta P + 1)} - \sqrt{\zeta P + K} \right)^2}{2\zeta(K - 1)} > 0,$$

that is, for a total power constraint P , the power used for training is greater than that used for data transmission.

From (3.23), we see that the MSE decreases as the power P increases. For a fixed K , as $P \rightarrow \infty$, we obtain

$$\lim_{P \rightarrow \infty} J(P, K) = \frac{\sigma_\theta^2}{1 + K\beta} \quad (3.24)$$

which is the same as (3.17). This makes sense since $P \rightarrow \infty$ implies $P_t^{opt} \rightarrow \infty$ and thus the MSE of channel estimation in (3.5) approaches to zero, that is, $\hat{h}_k \rightarrow h_k$ as $P \rightarrow \infty$ in the mean square sense. It is shown in Appendix E that, for a fixed P ,

$$\lim_{K \rightarrow \infty} J(P, K) = \sigma_\theta^2 \left(1 + \frac{\beta}{1 + \beta} \left(\sqrt{\zeta P + 1} - 1 \right)^2 \right)^{-1}. \quad (3.25)$$

The MSE does not approach to zero. The reason is that the order of $1/K$ decrease in MSE in (3.18) is offset by the order of K increase in the power of the error term $E[|\varepsilon|^2 | \hat{\mathbf{h}}]$ in (C.3) in Appendix C. Therefore, in the presence of the channel estimation error, the MSE reaches a finite nonzero value as K goes to infinity.

3.3.3 Comparison of two cases

If the total network power and number of sensors are fixed, with estimated channel, the estimation performance is worse than when channel information is available due to the presence of channel estimation error. To quantitative compare the two cases, we set the same MSE objective, use optimal power allocation for both cases, and determine the respective total network power that would be required. Suppose to achieve the selected MSE, total network power P^a is required when channel information is available and the required total network power is P^e when channels are estimated. The ratio P^a/P^e gives an indication of the penalty incurred by the consumption of training power and the presence of channel estimation error. A small ratio would imply a heavy penalty. But the MSE expressions in (3.16) and (3.23) are random variables, we instead derive the condition on P^a and P^e under which the distributions of MSEs are identical. This is possible due to the fact that the random variables g_k and \hat{g}_k have identical Rayleigh distribution. From (3.16) and (3.23), the distributions of MSE expressions are identical if the deterministic terms in the denominator are equal, that is,

$$\left(\frac{K - 1}{\sqrt{K(\zeta P^e + 1)} - \sqrt{\zeta P^e + K}} \right)^2 = \frac{1}{\zeta P^a} \quad (3.26)$$

Rearranging (3.26), we get

$$\frac{P^a}{P^e} = \left(\frac{\sqrt{K(1 + 1/(\zeta P^e))} - \sqrt{1 + K/(\zeta P^e)}}{K - 1} \right)^2 \quad (3.27)$$

Note that the ratio in (3.27) is less than one, and for P^e large

$$\frac{P^a}{P^e} \approx \frac{1}{(\sqrt{K} + 1)^2}. \quad (3.28)$$

The ratio decreases as the number of sensors K increases. This means that the penalty caused by channel estimation becomes worse as the number of sensor increases.

3.4 Equal Power Allocation

The optimal power allocation scheme discussed in the previous section requires that the actual channel or estimated channel information is available at the sensors. If the sensors do not get full channel information feedback from the FC, a reasonable strategy is to allocate equal power for each sensor for data transmission. In the following, we study the performance of the equal power allocation scheme. We again consider two cases: i) channels are known at the FC and ii) channels are estimated. In the latter case, we consider the optimal choice of training power P_t to achieve the smallest MSE. We compare performance of the two cases in terms of the power ratio P^a/P^e as in the previous section.

3.4.1 When channels are known

We will assume that the channel phase $\angle h_k$ is available at the sensor k and choose the phase of α_k as $\angle \alpha_k = -\angle h_k$. With equal power allocation, we have $|\alpha_k|^2 = P/(K(\sigma_\theta^2 + \sigma_n^2))$, for $k = 1, \dots, K$ and the MSE in (3.12) can be rewritten as

$$J_o(P, K) = \sigma_\theta^2 \left(1 + \frac{\frac{\beta}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K g_k \right)^2}{\frac{1}{K} \frac{1}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K g_k^2 \right) + \frac{1}{K\zeta P}} \right)^{-1} \quad (3.29)$$

It is easy to see from (3.29) that J_o decreases as P increases. For a fixed K , as $P \rightarrow \infty$, we have

$$\lim_{P \rightarrow \infty} \frac{1}{J_o(P, K)} = \sigma_\theta^{-2} \left(1 + \beta \frac{(\sum_{k=1}^K g_k)^2}{\sum_{k=1}^K g_k^2} \right) \leq \sigma_\theta^{-2} (1 + K\beta)$$

where the last inequality uses the Cauchy-Schwartz inequality and the equal sign holds if and only if $g_1 = \dots = g_K$. Therefore, as $P \rightarrow \infty$, we have a MSE lower bound as follows

$$\lim_{P \rightarrow \infty} J_o(P, K) \geq \frac{\sigma_\theta^2}{1 + K\beta}. \quad (3.30)$$

Since equality holds in (3.17), we see that the performance of the equal power scheme is usually worse than that of the optimal power scheme as $P \rightarrow \infty$. On the other hand, for a fixed P , as $K \rightarrow \infty$, we have $(1/K) \sum_{k=1}^K g_k \rightarrow E[g_k] = \sqrt{\pi/4}$ and $(1/K) \sum_{k=1}^K g_k^2 \rightarrow E[g_k^2] = 1$, thus (3.29) becomes

$$\lim_{K \rightarrow \infty} J_o(P, K) = \lim_{K \rightarrow \infty} \frac{\sigma_\theta^2}{K} \left(\frac{\frac{\beta}{1+\beta} \frac{\pi}{4}}{\frac{1}{1+\beta} + \frac{1}{\zeta P}} \right)^{-1} = 0. \quad (3.31)$$

Hence, the MSE decreases in the order of $1/K$ and approaches to zero as $K \rightarrow \infty$ even though the total power P is finite. Similar conclusion appeared in [26] for the unit variance case.

3.4.2 When channels are estimated

If the power P_t is used for channel estimation, the transmitted data power for the k th sensor is $P_k = (P - P_t)/K$, or equivalently, $|\alpha_k|^2 = (P - P_t)/(K(\sigma_\theta^2 + \sigma_n^2))$. Again the phase of α_k is chosen as $\angle \alpha_k = -\angle \hat{h}_k$ and the MSE derived from (3.19) is

$$J(P, K) = \sigma_\theta^2 \left(1 + \frac{\zeta^2 K \frac{\beta}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2 (P - P_t) P_t}{\zeta^2 \frac{1}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k^2 \right) (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K} \right)^{-1} \quad (3.32)$$

From (3.32), the optimization problem becomes to choose P_t so that the MSE J is minimum under the total network power constraint. From (3.32) the MMSE optimization problem can be formulated equivalently as

$$\begin{cases} \min_{P_t} & - \frac{\zeta^2 K \frac{\beta}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2 (P - P_t) P_t}{\zeta^2 \frac{1}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k^2 \right) (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K} \\ \text{subject to} & 0 \leq P_t \leq P. \end{cases} \quad (3.33)$$

It is shown in Appendix F that the objective function in (3.33) with respect to P_t is positive. Hence, the optimization problem (3.33) is convex since the objective function is convex and the constraint is linear. The following proposition gives the optimal training power and the corresponding MSE.

Proposition 3.2. *For $K > 1$, the solution to (3.33) gives the optimal training power*

$$P_t^{opt} = \frac{K(\zeta P + 1) - \sqrt{K(\zeta P + 1)(\zeta P + K)}}{\zeta(K - 1)} \quad (3.34)$$

where $\zeta = \sigma_h^2/\sigma_v^2$, and the incurred MSE

$$J(P, K) = \sigma_\theta^2 \left(1 + \frac{\frac{\beta}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2}{\frac{1}{K} \frac{1}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k^2 \right) + \frac{(K-1)^2}{K^2 (\sqrt{\zeta P+1} - \sqrt{1+(\zeta P/K)})^2}} \right)^{-1} \quad (3.35)$$

Proof. Please see Appendix G. □

Note that the optimal training powers for both the equal and optimal power allocation schemes are the same. This is because for the optimal power allocation scheme, if we set $P_k = (P - P_t^{opt})/K$ for all k , the MSE and the optimal training power must be the same as the equal power allocation scheme.

From (3.35) with fixed K , as $P \rightarrow \infty$, we obtain $\lim_{P \rightarrow \infty} (1/J(P, K)) \leq \sigma_\theta^{-2} (1 + K\beta)$ and thus

$$\lim_{P \rightarrow \infty} J(P, K) \geq \frac{\sigma_\theta^2}{1 + K\beta} \quad (3.36)$$

which is the same as (3.30). On the other hand, for a fixed P , when $K \rightarrow \infty$, we obtain

$$\lim_{K \rightarrow \infty} J(P, K) = \sigma_\theta^2 \left(1 + \frac{\pi}{4} \frac{\beta}{1+\beta} \left(\sqrt{\zeta P+1} - 1 \right)^2 \right)^{-1} \quad (3.37)$$

which is worse than (3.25). Note that the MSE in (3.37) also approaches a finite nonzero value as the number of sensors goes to infinity due to the same reason as stated in Section 3.3.2.

3.4.3 Comparison of two cases

To compare performance of the two cases, we set P^a and P^e respectively so that the MSE expressions in (3.29) and (3.35) have the same distribution as in Section 3.3.3. From (3.29) and (3.35), the distributions of the MSE are identical if the deterministic terms in the denominator are equal, that is,

$$\left(\frac{K-1}{\sqrt{K(\zeta P^e+1)} - \sqrt{\zeta P^e+K}} \right)^2 = \frac{1}{\zeta P^a}. \quad (3.38)$$

This equation is the same as (3.26) and thus we have the same ratio of penalty incurred by the training power consumption and the channel estimation error as shown in (3.27) and (3.28).

3.5 Numerical Results and Discussion

In this section, we use a number of numerical simulations to verify the analytical results obtained in previous sections. All random parameters, θ , n_k , h_k , and ν , are set as zero-mean circular Gaussian. The parameter θ and the channel h_k are assumed to have unit variance, that is, we set $\sigma_\theta^2 = \sigma_h^2 = 1$ (0 dB). The observation noise variance $\sigma_n^2 = -10$ dB and the receiver noise variance $\sigma_\nu^2 = -1$ dB so that the observation SNR $\beta = \sigma_\theta^2/\sigma_n^2$ and the channel SNR $\zeta = \sigma_h^2/\sigma_\nu^2$ are 10 dB and 1 dB, respectively.

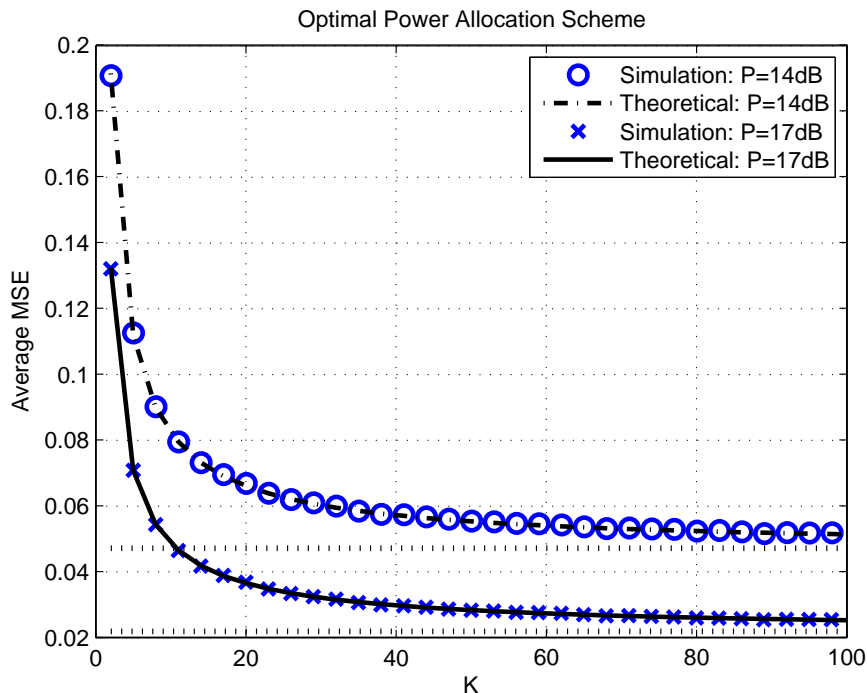


Figure 3.2: Mean square error (MSE) with optimal power allocation

Simulation 3.1 - Comparison of Theoretical and Simulation MSEs Based on Optimal Power Scheme: In this simulation, we compute the average MSEs of theoretical and simulation results based on the optimal power allocation scheme. The average MSE is the average of 10^5 independent runs. The theoretical MSE is given in (3.23), where only the normalized channel gains \hat{g}_k are random. To obtain the simulation MSE, we use the LMMSE estimators in (3.2) and (3.7) with all random variables independently generated, and take the average MSE of $\hat{\theta}$. It is clear from Figure 3.2 that the theoretical and simu-

lation values of MSE are very close. For a fixed P , we see that the MSE decreases as the number of sensors K increases and approaches to the lower bound (3.25). The results for $P = 14$ dB and $P = 17$ dB show that the 3 dB difference in total power leads to about 3 dB difference in MSE for $K \geq 20$.

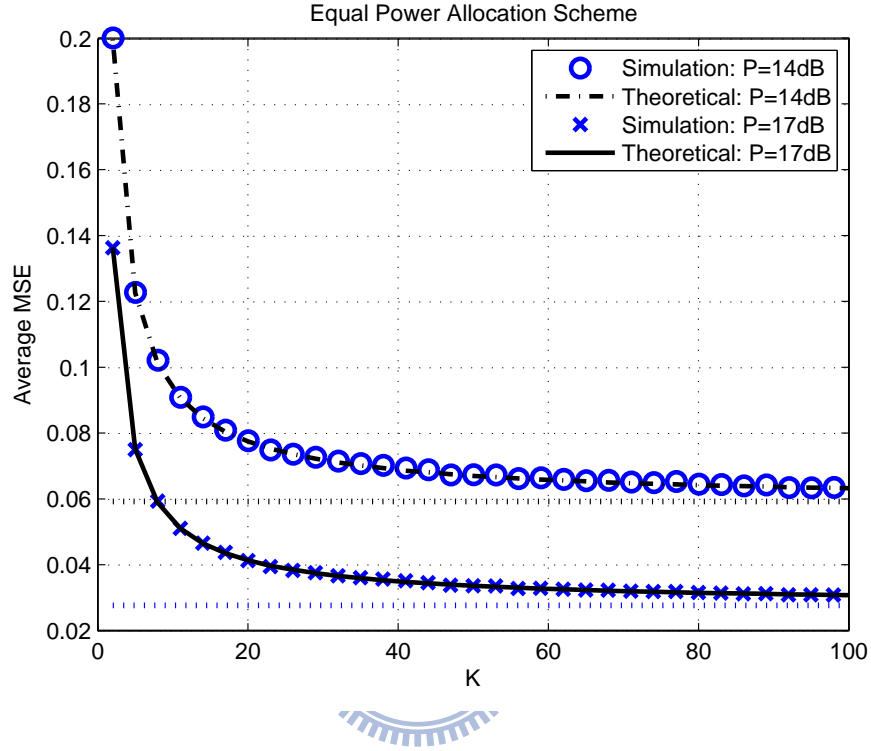


Figure 3.3: Mean square error (MSE) with equal power allocation

Simulation 3.2 - Comparison of Theoretical and Simulation MSEs based on Equal Power Scheme: In this simulation, we compute the average MSEs of theoretical and simulation results based on the equal power allocation scheme. The comparison is shown in Figure 3.3, where the theoretical result averages the MSE in (3.35). Again the figure shows that the theoretical and simulation values are very close. As K increases, the average MSE decreases and approaches to the lower bound in (3.37). The results for $P = 14$ dB and $P = 17$ dB also show roughly 3 dB difference in MSE for $K \geq 20$.

Simulation 3.3 - Comparison of Optimal and Equal Power Schemes: In this simulation, we compare the MSE of the optimal power scheme to that of the equal power scheme.

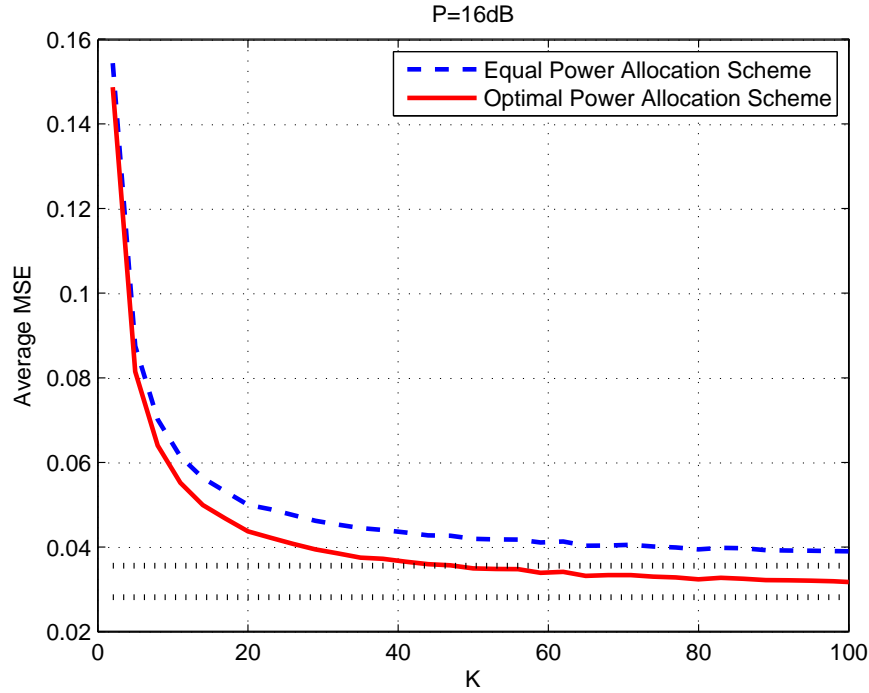


Figure 3.4: Comparison between equal and optimal power allocation schemes

Figure 3.4 shows the comparison result for a fixed $P = 16$ dB. It shows that the optimal power scheme performs better than the equal power scheme. For $K \geq 20$, the difference in MSE between the two schemes approaches to a constant value 0.007 (approximately 20% difference), which is about the difference between the respective low bounds.

Simulation 3.4 - Comparison with Orthogonal MAC model [25]: In this simulation, we compare the proposed scheme to that in [25]. The two-phase approach proposed in [25] is based on the orthogonal model, where the k th sensor transmits the measured signal $\alpha_k(\theta + n_k)$ to the k th receiver through an unknown fading channel h_k , $k = 1, \dots, K$, as shown in Figure 3.5. The k th receiver is corrupted by an additive noise $\nu_k \sim \mathcal{CN}(0, \sigma_\nu^2)$, where $\sigma_\nu^2 = -1$ dB, and $E[\nu_i \nu_j] = 0$ for $i \neq j$; then K received data is collected by the FC with the LMMSE fusion rule for estimating the signal. Figure 3.6 shows that with a fixed $P = 17$ dB, the MSE of the orthogonal model exhibits a conspicuous degradation as $K > 40$, while the MSE of the coherent model approaches to a constant value. Also compared with the orthogonal model, the coherent model has a lower average MSE re-

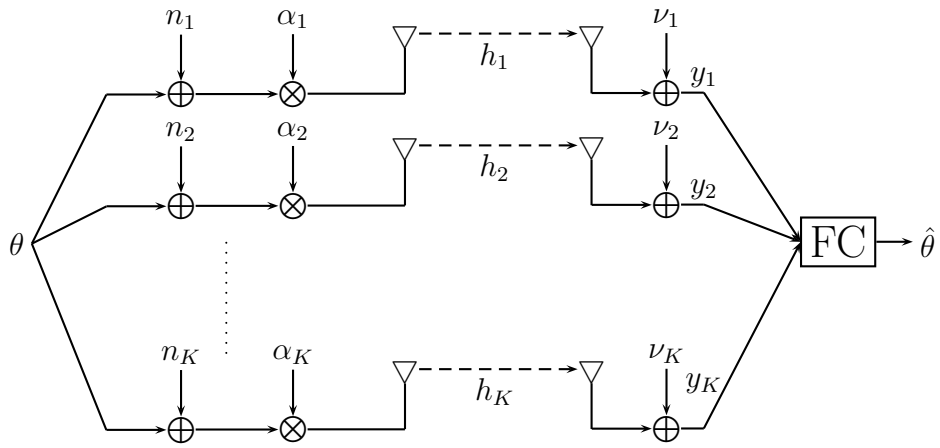


Figure 3.5: Orthogonal MAC wireless sensor network

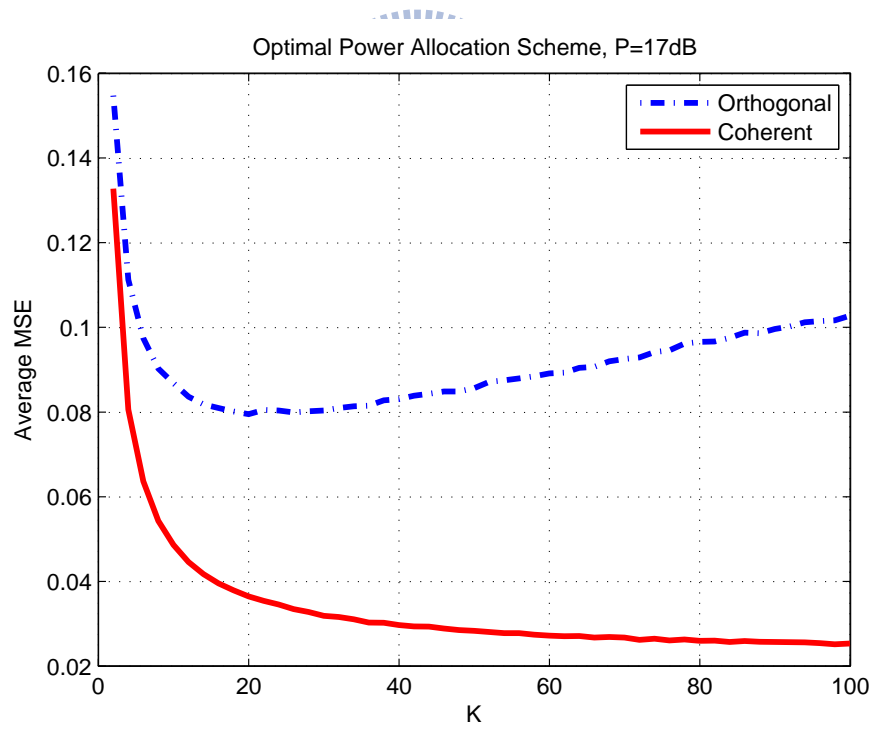


Figure 3.6: Performance of optimal power allocation scheme: coherent model and orthogonal model

regardless of the number of sensors used. This is a consequence of using orthogonal model, which results in K different receiver noise ν_k at the FC so that the increase of K does not reduce the effect of receiver noise; while in the coherent model, only one receiver noise is generated at the FC, which leads to increased signal to noise ratio as K increases. In the figure, we see that as K increases, the MSE of the coherent model is less sensitive to the channel estimation error than that of the orthogonal model.

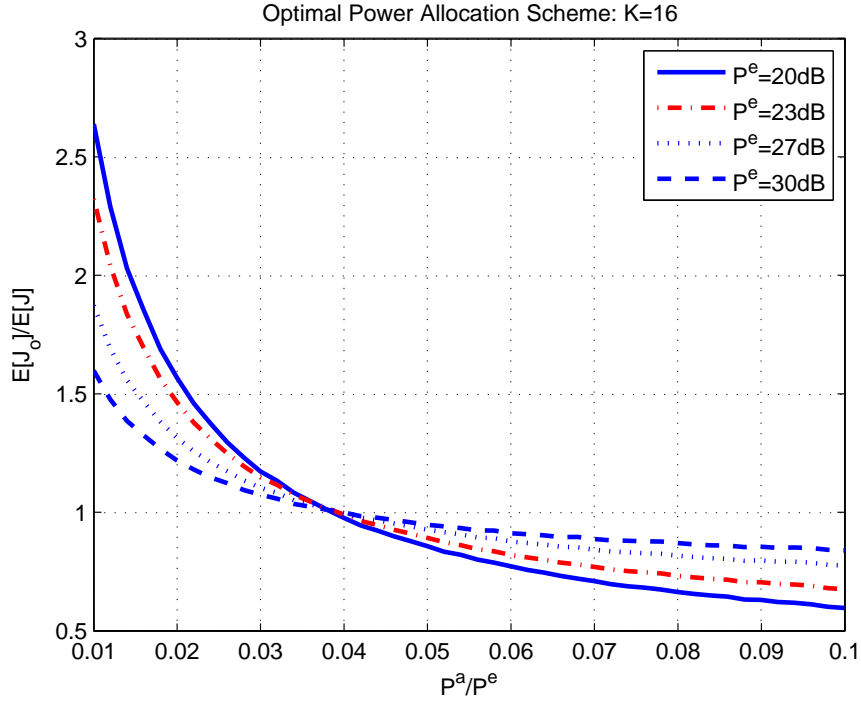


Figure 3.7: MSE ratio versus total power ratio: optimal power scheme with different P^e

Simulation 3.5 - Comparison with the Case When Channels are Known: In this simulation, we use simulations to show the analytical results given in Section 3.3.3 and Section 3.4.3. Figure 3.7 shows the ratio of average MSE $E[J_o]/E[J]$ versus the ratio P^a/P^e for the optimal power allocation scheme with a fixed $K = 16$ sensors. In the figure, the curves corresponding to total network power $P^e = 20$ dB, 23 dB, 27 dB, and 30 dB, respectively. The curves all cross the horizontal lines $E[J_o]/E[J] = 1$ at about 0.04 very close to the predicted $1/(\sqrt{K} + 1)^2$ in (3.28). Figure 3.8 shows the ratio of average MSE $E[J_o]/E[J]$ versus the ratio P^a/P^e for the equal power scheme. The total

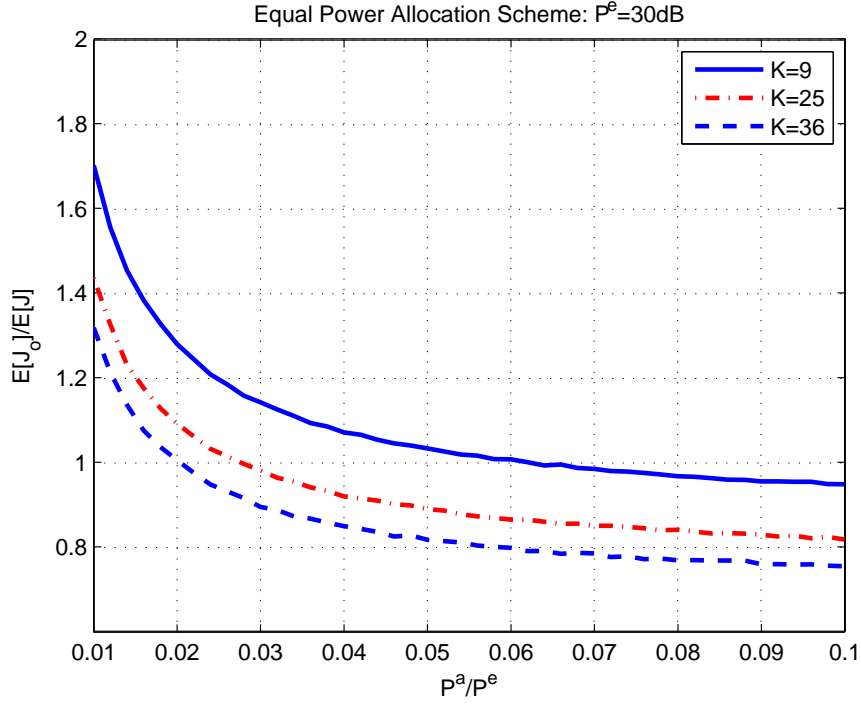
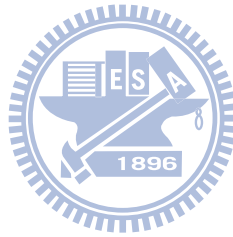


Figure 3.8: MSE ratio versus total power ratio: equal power scheme with different K

network power is fixed at $P^e = 30$ dB and the number of sensors $K = 9, 25,$ and 36 . We see that the curve for $K = 9$ crosses $E[J_o]/E[J] = 1$ at about 0.06, the curve for $K = 25$ at about 0.03, and the curve for $K = 36$ at about 0.02. The curves show that the penalty caused by channel estimation becomes worse as the number of sensors K increases.

A Brief Summary and Discussion: We use a two-phase approach for channel and source signal estimations; in both phases, the LMMSE criterion is used. We consider two power allocation schemes. For the optimal power allocation scheme, we obtain expressions of optimal training power and optimal data power for each sensor and the resulting MSE as a function of total network power P and the number of sensor K . For the equal power allocation scheme, we obtain an expression for the optimal training power and the resulting MSE. In both schemes, the optimal training powers are equal. Our results show that with estimated channels, the MSEs approach to finite nonzero values as the number of sensors increases. We note that this is in contrast with the result obtained for orthogonal MAC model [25] which shows the MSE performance eventually deteriorates

as the number of sensor increases. The MSE performance compared with the case when channels are known shows the penalty caused by channel estimation becomes worse as the number of sensors increases.



Chapter 4

Distributed Estimation Using MIMO Sensor Network

Multiple-input multiple-output (MIMO) techniques can potentially improve data rate and link reliability and thus increase the capacity of wireless systems [33, 34]. In this chapter, we consider distributed estimation of a vector signal using MIMO wireless sensor network, in which each sensor is composed of multiple measurements and multiple transmitters, and the FC consists of multiple received antennas. We study the problem of the design of linear coding matrices so that the estimated error is minimized under a total power constraint. We show that the problem can be formulated as a convex optimization problem, based on which we derive closed form expressions of the optimal coding matrices.

4.1 System Model and Problem Formulation

We consider a wireless sensor network consisting of L sensors for estimating p random source signals, written in vector form $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T \in \mathbb{C}^p$, as shown in Figure 4.1. The l th sensor has k_l measurements, which can be expressed in vector form as

$$\mathbf{x}_l = \mathbf{F}_l \boldsymbol{\theta} + \mathbf{n}_l, \quad 1 \leq l \leq L \quad (4.1)$$

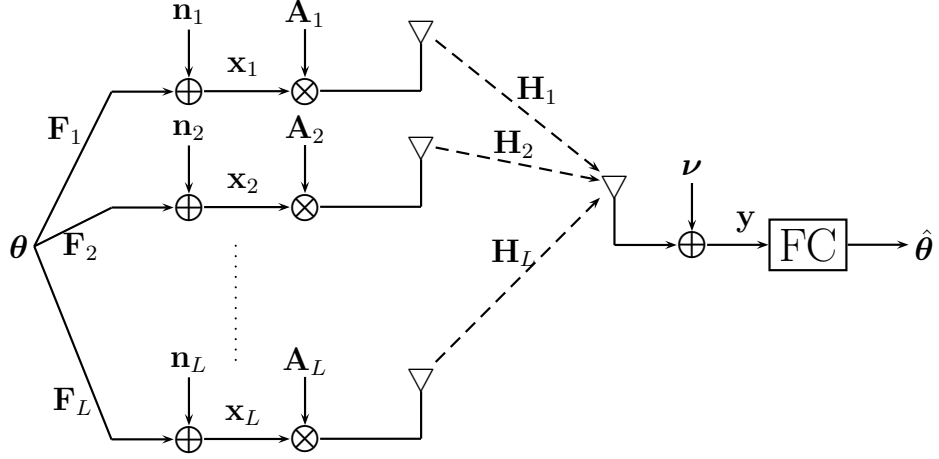


Figure 4.1: Linear distributed estimation with coherent MAC

where $\mathbf{F}_l \in \mathbb{C}^{k_l \times p}$ is the observation matrix and $\mathbf{n}_l \in \mathbb{C}^{k_l}$ is the additive noise. Let $K = \sum_{l=1}^L k_l$. In vector form, (4.1) can be written as

$$\mathbf{x} = \mathbf{F}\boldsymbol{\theta} + \mathbf{n}, \quad (4.2)$$

where $\mathbf{x} = [\mathbf{x}_1^T \mathbf{x}_2^T \cdots \mathbf{x}_L^T]^T \in \mathbb{C}^K$, $\mathbf{F} = [\mathbf{F}_1^T \mathbf{F}_2^T \cdots \mathbf{F}_L^T]^T \in \mathbb{C}^{K \times p}$, and $\mathbf{n} = [\mathbf{n}_1^T \mathbf{n}_2^T \cdots \mathbf{n}_L^T]^T \in \mathbb{C}^K$. The measurement of the l th sensor is encoded by a linear coding matrix $\mathbf{A}_l \in \mathbb{C}^{N \times k_l}$ to form the message vector $\mathbf{A}_l \mathbf{x}_l \in \mathbb{C}^N$, which is then sent to the FC through a channel matrix $\mathbf{H}_l \in \mathbb{C}^{N \times N}$. From (4.1), the received signal \mathbf{y} at the FC can be expressed as

$$\mathbf{y} = \sum_{l=1}^L \mathbf{H}_l \mathbf{A}_l (\mathbf{F}_l \boldsymbol{\theta} + \mathbf{n}_l) + \boldsymbol{\nu}, \quad (4.3)$$

where $\boldsymbol{\nu} \in \mathbb{C}^N$ is additive noise at the receiver. Let $\mathbf{B}_l = \mathbf{H}_l \mathbf{A}_l \in \mathbb{C}^{N \times k_l}$ and $\mathbf{B} = [\mathbf{B}_1 \mathbf{B}_2 \cdots \mathbf{B}_L] \in \mathbb{C}^{N \times K}$, and the received signal in (4.3) can be rewritten as

$$\begin{aligned} \mathbf{y} &= [\mathbf{B}_1 \mathbf{B}_2 \cdots \mathbf{B}_L] \left(\begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_L \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_L \end{bmatrix} \right) + \boldsymbol{\nu} \\ &= \mathbf{B}(\mathbf{F}\boldsymbol{\theta} + \mathbf{n}) + \boldsymbol{\nu}. \end{aligned} \quad (4.4)$$

The following assumption are made in the sequel.

- i) The source signal vector is $E[\boldsymbol{\theta}] = \mathbf{0}$ and $E[\boldsymbol{\theta}\boldsymbol{\theta}^H] = \sigma_\theta^2 \mathbf{I}_p$.
- ii) The measurement noise vectors are zero-mean and mutually uncorrelated, that is, $E[\mathbf{n}_l] = \mathbf{0}$, $E[\mathbf{n}_l \mathbf{n}_l^H] = \sigma_n^2 \mathbf{I}_{k_l}$, and $E[\mathbf{n}_l \mathbf{n}_j^H] = \mathbf{0}$ for $j \neq l$.
- iii) The received noise vector is $E[\boldsymbol{\nu}] = \mathbf{0}$ and $E[\boldsymbol{\nu}\boldsymbol{\nu}^H] = \sigma_\nu^2 \mathbf{I}_N$.
- iv) The signal vector, the measurement noise vectors, and the received noise vector are uncorrelated, that is, $E[\boldsymbol{\theta}\mathbf{n}^H] = \mathbf{0}$, $E[\boldsymbol{\theta}\boldsymbol{\nu}^H] = \mathbf{0}$, and $E[\mathbf{n}\boldsymbol{\nu}^H] = \mathbf{0}$.
- v) The observation matrices \mathbf{F}_l and the channel matrices \mathbf{H}_l are known at the FC.
- vi) The number of source signals p , measurements K , and transmitters N should satisfy $K \geq N$, $K \geq p$, and $\text{rank}(\mathbf{F}) = p$.
- vii) The channel matrices are diagonal, that is, $\mathbf{H}_l = \text{diag}(h_1^l, \dots, h_N^l)$ with $h_i^l \neq 0, \forall i, l$.

Remarks: Assumption vii) simplifies the derivations to follow. In practice, if the N signals are transmitted using N different frequencies, the assumption is reasonable.

According to the received signal \mathbf{y} in (4.4), the LMMSE estimate of $\boldsymbol{\theta}$ is

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= E[\boldsymbol{\theta}\mathbf{y}^H] (E[\mathbf{y}\mathbf{y}^H])^{-1} \mathbf{y} \\ &= \sigma_\theta^2 \mathbf{F}^H \mathbf{B}^H [\mathbf{B}\mathbf{R}_x \mathbf{B}^H + \sigma_\nu^2 \mathbf{I}_N]^{-1} \mathbf{y}, \end{aligned}$$

where $\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H] = \sigma_\theta^2 \mathbf{F}\mathbf{F}^H + \sigma_n^2 \mathbf{I}_K$, and the corresponding MSE is

$$\begin{aligned} J &= \text{tr}(E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^H]) \\ &= \text{tr} \left(\sigma_\theta^2 \mathbf{I}_p - \sigma_\theta^4 \mathbf{F}^H \mathbf{B}^H [\mathbf{B}\mathbf{R}_x \mathbf{B}^H + \sigma_\nu^2 \mathbf{I}_N]^{-1} \mathbf{B}\mathbf{F} \right). \end{aligned} \quad (4.5)$$

The problem is to minimize J in (4.5) by the design of the coding matrices \mathbf{A}_l under a total power constraint. The transmitted power for the l th sensor is defined as $E[\mathbf{x}_l^H \mathbf{A}_l^H \mathbf{A}_l \mathbf{x}_l] = \text{tr}(\mathbf{A}_l \mathbf{R}_{x_l} \mathbf{A}_l^H)$, where $\mathbf{R}_{x_l} = E[\mathbf{x}_l \mathbf{x}_l^H] = \sigma_\theta^2 \mathbf{F}_l \mathbf{F}_l^H + \sigma_n^2 \mathbf{I}_{k_l}$. If P is the total power that the L sensors can use, then the constraint can be expressed as

$$\sum_{l=1}^L \text{tr}(\mathbf{H}_l^{-1} \mathbf{B}_l \mathbf{R}_{x_l} \mathbf{B}_l^H \mathbf{H}_l^{-H}) \leq P, \quad (4.6)$$

where we use $\mathbf{A}_l = \mathbf{H}_l^{-1}\mathbf{B}_l$. Let $D = \text{tr} \left(\mathbf{F}^H \mathbf{B}^H [\mathbf{B}\mathbf{R}_x \mathbf{B}^H + \sigma_\nu^2 \mathbf{I}_N]^{-1} \mathbf{B}\mathbf{F} \right)$. Thus $J = p\sigma_\theta^2 - \sigma_\theta^4 D$. Since σ_θ and p are fixed, minimization of J subject to (4.6) can be expressed equivalently as

$$\begin{aligned} & \max_{\mathbf{B}_l, 1 \leq l \leq L} D \\ & \text{subject to } \sum_{l=1}^L \text{tr} \left(\mathbf{H}_l^{-1} \mathbf{B}_l \mathbf{R}_{x_l} \mathbf{B}_l^H \mathbf{H}_l^{-H} \right) \leq P. \end{aligned} \quad (4.7)$$

Remarks: In general, if $E[\boldsymbol{\theta}\boldsymbol{\theta}^H] = \mathbf{R}_\theta$ and $E[\mathbf{n}_l \mathbf{n}_l^H] = \mathbf{R}_{n_l}$ are Hermitian and positive definite, we can write $\mathbf{R}_\theta = \mathbf{R}_\theta^{1/2} \mathbf{R}_\theta^{1/2}$ and $\mathbf{R}_{n_l} = \mathbf{R}_{n_l}^{1/2} \mathbf{R}_{n_l}^{1/2}$, and introduce $\tilde{\boldsymbol{\theta}} = \mathbf{R}_\theta^{-1/2} \boldsymbol{\theta}$, $\tilde{\mathbf{n}}_l = \mathbf{R}_{n_l}^{-1/2} \mathbf{n}_l$, $\tilde{\mathbf{A}}_l = \mathbf{A}_l \mathbf{R}_{n_l}^{1/2}$, and $\tilde{\mathbf{F}}_l = \mathbf{R}_{n_l}^{-1/2} \mathbf{F}_l \mathbf{R}_\theta^{1/2}$. Then \mathbf{A}_l , \mathbf{F}_l , $\boldsymbol{\theta}$, and \mathbf{n}_l in (4.4) can be replaced by $\tilde{\mathbf{A}}_l$, $\tilde{\mathbf{F}}_l$, $\tilde{\boldsymbol{\theta}}$, and $\tilde{\mathbf{n}}_l$, respectively, and the equivalent model satisfies assumption i) and ii). Hence, the optimization problem can still be formulated in the same form as (4.7).

4.2 Proposed Approach

In this section, we first consider the objective function of (4.7) and determine its maximum through singular value decomposition technique. After obtaining the maximum objective function, we consider the power constraint and formulate the problem (4.7) as a convex optimization problem, which then yields a solution in closed form.

Since $\mathbf{R}_x = \sigma_\theta^2 \mathbf{F}\mathbf{F}^H + \sigma_n^2 \mathbf{I}_K$ is positive definite, it can be expressed as $\mathbf{R}_x = \mathbf{R}_x^{1/2} \mathbf{R}_x^{1/2}$, where $\mathbf{R}_x^{1/2}$ is Hermitian and positive definite. Since $\text{rank}(\mathbf{F}) = p$ by assumption vi), $\text{rank}(\mathbf{R}_x^{-1/2} \mathbf{F}\mathbf{F}^H \mathbf{R}_x^{-1/2}) = p$ and we have

$$\mathbf{R}_x^{-1/2} \mathbf{F}\mathbf{F}^H \mathbf{R}_x^{-1/2} = \mathbf{U}_C \boldsymbol{\Lambda}_C \mathbf{U}_C^H, \quad (4.8)$$

where $\boldsymbol{\Lambda}_C = \text{diag}(c_1, \dots, c_p, 0, \dots, 0)$ with $c_1 \geq \dots \geq c_p > 0$ and $\mathbf{U}_C \in \mathbb{C}^{K \times K}$ is unitary. Multiplying (4.8) by $\mathbf{R}_x^{-1/2}$ on the right and $\mathbf{R}_x^{1/2}$ on the left shows that the nonzero eigenvalues c_i , $1 \leq i \leq p$, are also the nonzero eigenvalues of $\mathbf{F}\mathbf{F}^H (\sigma_\theta^2 \mathbf{F}\mathbf{F}^H + \sigma_n^2 \mathbf{I}_K)^{-1}$ and it follows that $c_i \leq 1/\sigma_\theta^2$, $1 \leq i \leq p$. Let $\bar{\mathbf{B}} = \mathbf{B}\mathbf{R}_x^{1/2}$ and use (4.8), the objective function D in (4.7) can be expressed as

$$\begin{aligned} D &= \text{tr} \left(\bar{\mathbf{B}}^H [\bar{\mathbf{B}}\bar{\mathbf{B}}^H + \sigma_\nu^2 \mathbf{I}_N]^{-1} \bar{\mathbf{B}}\mathbf{R}_x^{-1/2} \mathbf{F}\mathbf{F}^H \mathbf{R}_x^{-1/2} \right) \\ &= \text{tr} \left(\bar{\mathbf{B}}^H [\bar{\mathbf{B}}\bar{\mathbf{B}}^H + \sigma_\nu^2 \mathbf{I}_N]^{-1} \bar{\mathbf{B}}\mathbf{U}_C \boldsymbol{\Lambda}_C \mathbf{U}_C^H \right). \end{aligned} \quad (4.9)$$

Express $\bar{\mathbf{B}}$ as a singular value decomposition

$$\bar{\mathbf{B}} = \mathbf{U}_{\bar{\mathbf{B}}} \mathbf{\Lambda}_{\bar{\mathbf{B}}} \mathbf{V}_{\bar{\mathbf{B}}}^H, \quad (4.10)$$

where $\mathbf{U}_{\bar{\mathbf{B}}} \in \mathbb{C}^{N \times N}$ is unitary, $\mathbf{\Lambda}_{\bar{\mathbf{B}}} = \text{diag}(\sqrt{b_1}, \dots, \sqrt{b_N})$, $b_1 \geq \dots \geq b_N \geq 0$, and $\mathbf{V}_{\bar{\mathbf{B}}} \in \mathbb{C}^{K \times N}$ has orthonormal columns. With (4.10), D in (4.9) can be further simplified as

$$D = \text{tr} \left(\mathbf{\Lambda}_{\bar{\mathbf{B}}} [\mathbf{\Lambda}_{\bar{\mathbf{B}}}^2 + \sigma_\nu^2 \mathbf{I}_N]^{-1} \mathbf{\Lambda}_{\bar{\mathbf{B}}} \mathbf{V}_{\bar{\mathbf{B}}}^H \mathbf{U}_C \mathbf{\Lambda}_C \mathbf{U}_C^H \mathbf{V}_{\bar{\mathbf{B}}} \right). \quad (4.11)$$

To find an upper bound on (4.11), we need the follow fact.

Fact 4.1. [36, p.326] Let $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{n \times n}$ be positive semidefinite matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ and $\delta_1 \geq \delta_2 \geq \dots \geq \delta_n \geq 0$ respectively. Then

$$\text{tr}(\mathbf{X}\mathbf{Y}) \leq \sum_{i=1}^n \lambda_i \delta_i.$$

Applying the fact to (4.11) with $\mathbf{X} = \mathbf{\Lambda}_{\bar{\mathbf{B}}} [\mathbf{\Lambda}_{\bar{\mathbf{B}}} \mathbf{\Lambda}_{\bar{\mathbf{B}}} + \sigma_\nu^2 \mathbf{I}_N]^{-1} \mathbf{\Lambda}_{\bar{\mathbf{B}}}$ and $\mathbf{Y} = \mathbf{V}_{\bar{\mathbf{B}}}^H \mathbf{U}_C \mathbf{\Lambda}_C \mathbf{U}_C^H \mathbf{V}_{\bar{\mathbf{B}}}$, we have

$$D \leq \sum_{i=1}^p \frac{c_i b_i}{\sigma_\nu^2 + b_i}. \quad (4.12)$$

We note that since \mathbf{X} is diagonal, if we choose $\mathbf{V}_{\bar{\mathbf{B}}}$ so that $\mathbf{V}_{\bar{\mathbf{B}}}^H \mathbf{U}_C = [\mathbf{I}_N \mathbf{0}] \in \mathbb{C}^{N \times K}$, where $N \geq p$, then equality in (4.12) holds. Hence the upper bound in (4.12) is achieved if we choose

$$\mathbf{V}_{\bar{\mathbf{B}}} = \mathbf{U}_C(:, 1 : N), \quad (4.13)$$

where $\mathbf{U}_C(:, 1 : N)$ denotes the first N columns of \mathbf{U}_C . Since the upper bound is the same for all $N > p$, we choose $N = p$. This keeps the number of transmitters for each sensor at a minimum.

With the choice $\mathbf{V}_{\bar{\mathbf{B}}} = \mathbf{U}_C(:, 1 : p)$, we have

$$D = \sum_{i=1}^p \frac{c_i b_i}{\sigma_\nu^2 + b_i} \quad (4.14)$$

where $c_i > 0$ and σ_ν^2 are fixed. Hence the problem of choosing \mathbf{B}_l in \mathbf{B} to maximize D amounts to choosing the singular values $\sqrt{b_i}$ of $\bar{\mathbf{B}}$ in (4.10) since the choice of $\mathbf{U}_{\bar{\mathbf{B}}}$ is

irrelevant and $\mathbf{B} = \bar{\mathbf{B}}\mathbf{R}_x^{-1/2}$. It is clear from (4.14) that to maximize D , we should choose each $b_i > 0$ as large as possible; however, it can not be chosen arbitrarily large due to the total power constraint.

By taking into account the power constraint in (4.7), we set $\mathbf{U}_{\bar{B}} = \mathbf{I}_N$ to simplify our analysis and thus we have $\mathbf{B} = \Lambda_{\bar{B}}\mathbf{V}_{\bar{B}}^H\mathbf{R}_x^{-1/2}$, or equivalently,

$$\mathbf{B}_l = \Lambda_{\bar{B}}\hat{\mathbf{U}}_{C,l}, \quad 1 \leq l \leq L, \quad (4.15)$$

where $\hat{\mathbf{U}}_{C,l} \in \mathbb{C}^{p \times k_l}$ is the l th block matrix in $\mathbf{V}_{\bar{B}}^H\mathbf{R}_x^{-1/2} = [\hat{\mathbf{U}}_{C,1}, \dots, \hat{\mathbf{U}}_{C,L}]$ with $\mathbf{V}_{\bar{B}} = \mathbf{U}_C(:, 1:p)$. Substituting (4.15) into (4.7), since \mathbf{H}_l are diagonal matrices from assumption vii), the power constraint can be further simplified as

$$\text{tr}(\bar{\mathbf{R}}\Lambda_{\bar{B}}^2) = \sum_{i=1}^p r_i b_i \leq P, \quad (4.16)$$

where $\bar{\mathbf{R}} = \sum_{l=1}^L \mathbf{H}_l^{-1}\hat{\mathbf{U}}_{C,l}\mathbf{R}_{x_l}\hat{\mathbf{U}}_{C,l}^H\mathbf{H}_l^{-H}$ with diagonal entries r_i , $1 \leq i \leq p$. Note that since $\mathbf{R}_{x_l} = \sigma_\theta^2\mathbf{F}_l\mathbf{F}_l^H + \sigma_n^2\mathbf{I}_{k_l}$, we see that the diagonal entries of $\mathbf{H}_l^{-1}\hat{\mathbf{U}}_{C,l}\mathbf{R}_{x_l}\hat{\mathbf{U}}_{C,l}^H\mathbf{H}_l^{-H}$ are positive and thus $r_i > 0$. From (4.14) and (4.16), the problem in (4.7) with $\mathbf{V}_{\bar{B}} = \mathbf{U}_C(:, 1:p)$ and $\mathbf{U}_{\bar{B}} = \mathbf{I}_N$ can be written as

$$\begin{aligned} & \min_{b_i, 1 \leq i \leq p} - \sum_{i=1}^p \frac{c_i b_i}{\sigma_\nu^2 + b_i} \\ & \text{subject to } \sum_{i=1}^p r_i b_i \leq P, \\ & \quad b_i \geq 0, \quad i = 1, \dots, p. \end{aligned} \quad (4.17)$$

This is a convex optimization problem since the cost function is convex and the constraints are linear.

To solve the problem (4.17), we form the Lagrangian as

$$L(b_i, \mu_0, \mu_i) = - \sum_{i=1}^p \frac{c_i b_i}{\sigma_\nu^2 + b_i} + \mu_0 \left(\sum_{i=1}^p r_i b_i - P \right) - \sum_{i=1}^p \mu_i b_i,$$

where $\mu_0 \geq 0$ and $\mu_i \geq 0$, and the associated KKT conditions [38] are

$$- \frac{c_i \sigma_\nu^2}{(\sigma_\nu^2 + b_i)^2} + \mu_0 r_i - \mu_i = 0 \quad (4.18)$$

$$\mu_0 \left(\sum_{i=1}^p r_i b_i - P \right) = 0 \quad (4.19)$$

$$\mu_i b_i = 0 \quad (4.20)$$

From (4.18), we obtain

$$b_i = \sigma_\nu^2 \left(\sqrt{\frac{c_i}{\sigma_\nu^2(\mu_0 r_i + \mu_i)}} - 1 \right). \quad (4.21)$$

From (4.20), if $b_i > 0$, we have $\mu_i = 0$ and thus (4.21) can be written as

$$b_i = \sigma_\nu^2 \left(\sqrt{\frac{c_i}{\sigma_\nu^2 \mu_0 r_i}} - 1 \right)^+ \quad (4.22)$$

where $(x)^+ = \max(0, x)$. From (4.18) with $b_i > 0$ and $\mu_i = 0$, we have $\mu_0 > 0$ otherwise we have a contradiction $\mu_0 < 0$. Hence, from (4.19), we get

$$\sum_{i=1}^p r_i b_i = P. \quad (4.23)$$

Let $w_{m_i} = \sqrt{c_{m_i}/r_{m_i}}$ and assumed $w_{m_1} \geq \dots \geq w_{m_p}$, where $m_i \in \{1, \dots, p\}$. We define a function

$$f(m_n) = w_{m_n} \times \frac{\frac{P}{\sigma_\nu^2} + \sum_{i=m_1}^{m_n} r_i}{\sum_{i=m_1}^{m_n} \sqrt{r_i c_i / \sigma_\nu^2}}.$$

Let $1 \leq p_1 \leq p$ be such that $f(m_{p_1}) > 1$ and $f(m_{p_1+1}) \leq 1$. Then we have

$$b_{m_i} = \begin{cases} \sqrt{\frac{\sigma_\nu^2}{\mu_0}} w_{m_i} - \sigma_\nu^2, & i \leq p_1 \\ 0, & i > p_1 \end{cases} \quad (4.24)$$

where

$$\mu_0 = \left(\frac{\sum_{i=m_1}^{m_{p_1}} \sqrt{r_i c_i / \sigma_\nu^2}}{\frac{P}{\sigma_\nu^2} + \sum_{i=m_1}^{m_{p_1}} r_i} \right)^2$$

is obtained by substituting (4.24) into (4.23).

With the choice $\mathbf{\Lambda}_{\bar{B}} = \text{diag}(\sqrt{b_1}, \dots, \sqrt{b_p})$ according to (4.24), the coding matrix can be written from (4.15) and $\mathbf{A}_l = \mathbf{H}_l^{-1} \mathbf{B}_l$, and is given by

$$\mathbf{A}_l = \mathbf{H}_l^{-1} \mathbf{\Lambda}_{\bar{B}} \hat{\mathbf{U}}_{C,l}, \quad 1 \leq l \leq L; \quad (4.25)$$

moreover, from (4.5) and (4.14), the corresponding MSE can be written as

$$J = p\sigma_\theta^2 - \sigma_\theta^4 \sum_{i=1}^p \frac{c_i b_i}{\sigma_\nu^2 + b_i}. \quad (4.26)$$

As $P \rightarrow \infty$, from (4.16), we have $b_i \rightarrow \infty$ and thus the lower bound of the MSE is

$$J_{low} = p\sigma_\theta^2 - \sigma_\theta^4 \sum_{i=1}^p c_i, \quad (4.27)$$

Note that since $c_i \leq 1/\sigma_\theta^2$, we have $J_{low} \geq 0$.

So far we have developed an analytical method for the design of optimal coding matrices. For clarity, we summarize the steps of our proposed method as follows:

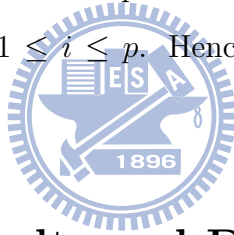
Step 1: Give the prior knowledge of $\sigma_\theta^2 \mathbf{I}_p$, $\sigma_n^2 \mathbf{I}_{k_l}$, $\sigma_\nu^2 \mathbf{I}_p$, the observation matrices \mathbf{F}_l , and the channel matrices \mathbf{H}_l .

Step 2: Stack \mathbf{F}_l into matrix \mathbf{F} , and compute $\mathbf{R}_{x_l} = \sigma_\theta^2 \mathbf{F}_l \mathbf{F}_l^H + \sigma_n^2 \mathbf{I}$ and $\mathbf{R}_x = \sigma_\theta^2 \mathbf{F} \mathbf{F}^H + \sigma_n^2 \mathbf{I}$.

Step 3: Decompose $\mathbf{R}_x^{-1/2} \mathbf{F} \mathbf{F}^H \mathbf{R}_x^{-1/2}$ into $\mathbf{U}_C \mathbf{\Lambda}_C \mathbf{U}_C^H$, where $\mathbf{\Lambda}_C = \text{diag}(c_1, \dots, c_p, \dots, 0)$, and set $\mathbf{V}_{\bar{B}} = \mathbf{U}_C(:, 1:p)$.

Step 4: Write $\mathbf{V}_{\bar{B}}^H \mathbf{R}_x^{-1/2} = [\hat{\mathbf{U}}_{C,1}, \dots, \hat{\mathbf{U}}_{C,L}]$ with $\hat{\mathbf{U}}_{C,l} \in \mathbb{R}^{p \times k_l}$ and compute $\bar{\mathbf{R}} = \sum_{l=1}^L \hat{\mathbf{U}}_{C,l} \mathbf{R}_{x_l} \hat{\mathbf{U}}_{C,l}^H \times (\mathbf{H}_l \mathbf{H}_l^H)^{-1}$ to get its diagonal entries r_i , $1 \leq i \leq p$.

Step 5: With $\mathbf{\Lambda}_C$ and $\bar{\mathbf{R}}$ obtained in Step 3 and Step 4, solve the optimization problem (4.17) to get b_i in (4.24), $1 \leq i \leq p$. Hence, the optimal coding matrices \mathbf{A}_l are obtained from (4.25).



4.3 Numerical Results and Discussion

In this section, we use numerical simulations to illustrate the analytical results established in Section 4.2. In all simulations, the random vectors $\boldsymbol{\theta}$, \mathbf{n}_l , and $\boldsymbol{\nu}$, are complex Gaussian. Specifically, each entry in the vectors is set as a complex Gaussian random variable with zero mean and unit variance. We also take the entries of \mathbf{F}_l and the channels h_i^l as complex Gaussian random variables with zero mean and unit variance. The number of source signals is set as $p = 5$.

Simulation 4.1 - Effects of N : In this simulation, we demonstrate the effect of the number of transmitters. The average MSE versus N with different power levels $P = 5$ dB, 10 dB, and 20 dB is shown in Figure 4.2, in which we take $L = 10$ and $k_l = 8$, $\forall l$. The dash-line denotes the MSE lower bound in (4.27). We can see that the MSE decreases as N

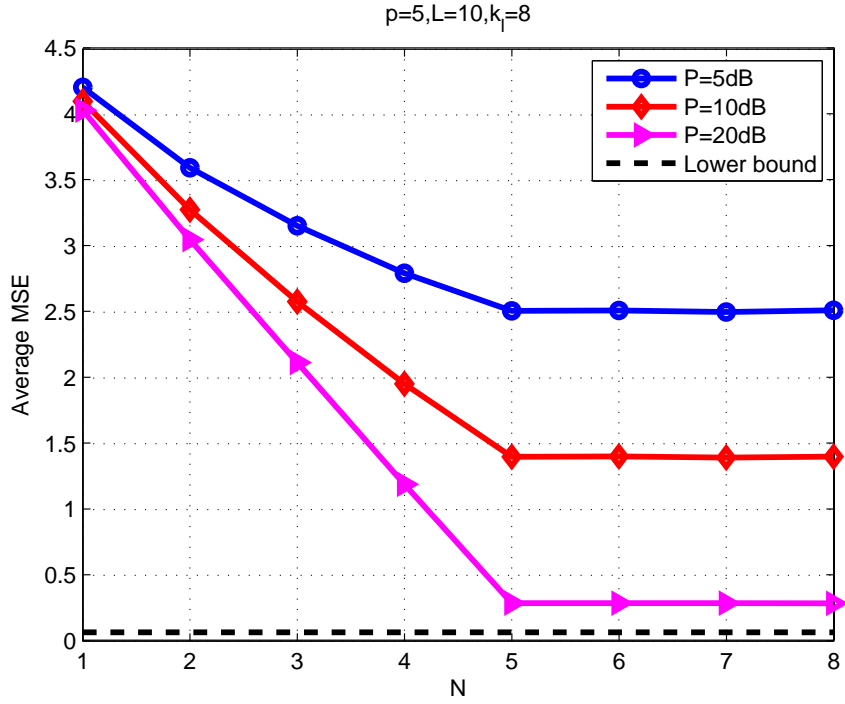


Figure 4.2: MSE versus N with different power levels

increases and remains a constant as $N \geq p$. That is, there is no improvement in network performance as the number of transmitters for each sensor is greater than the number of source signals. The result is consistent with our analysis in Section 4.2 that under a total power constraint, the minimum achieved MSE is (4.26) for all $N > p$. From the figure we can also see that the MSE decreases as the power level increases as is expected.

Since the performance is the same for all $N > p$, we set $N = p$ in all the simulations to follow.

Simulation 4.2 - Effects of k_l : In this simulation, we demonstrate the effect of the number of measurements. We consider $L = 10$ and assume that k_l are equal for all sensors. Figure 4.3 shows the average MSE versus k_l with different power levels $P = 5$ dB, 10 dB, and 20 dB. We note that as k_l increases, the MSEs decreases; moreover, the gap between the lower bound and the power constraint case becomes large as k_l increases. This is because although the increase of k_l leads to the increase of measurement power, the transmitted

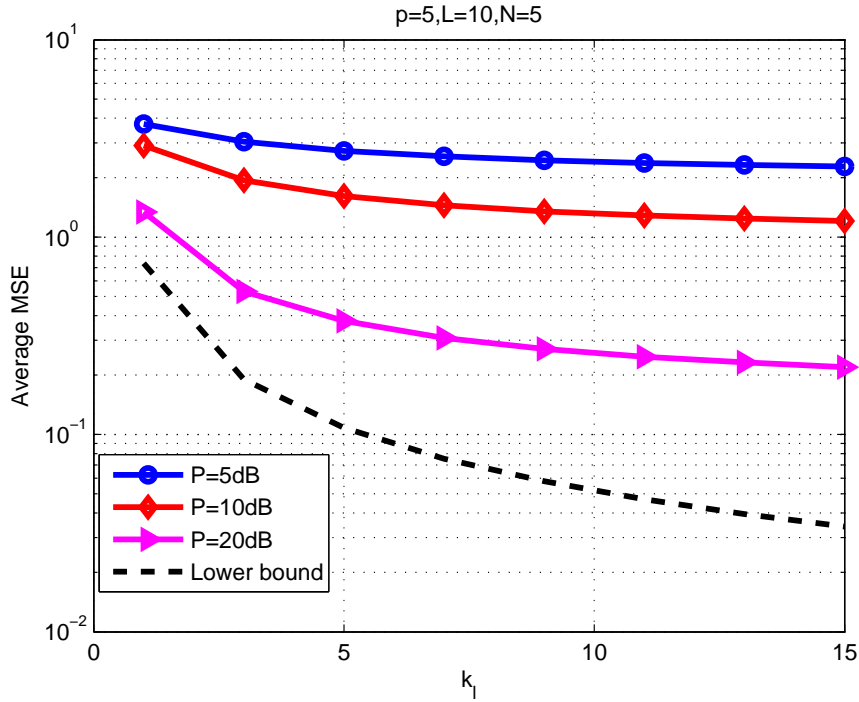


Figure 4.3: MSE versus k_l with different power levels

power for each sensor is restricted if there is a power constraint. Hence, the performance of the power constraint case has no significant improvement as k_l increases compared with that of the unconstrained power case (the lower bound).

Simulation 4.3 - Comparison with Equal Allocation Scheme: In this simulation, we compare the MSE of the optimal allocation scheme to that of the equal allocation scheme. We take $L = 10$ and $k_l = 8, \forall l$. The equal allocation scheme is to set the transmitted powers for all sensors to be equal, that is, we choose the coding matrix $\mathbf{A}_l = \alpha_l \cdot [\mathbf{I}_5 \mathbf{0}]$, where $\alpha_l = \sqrt{P/[L\text{tr}(\mathbf{R}_{x_l}(1:5, 1:5))]}$, so that $\text{tr}(\mathbf{A}_l \mathbf{R}_{x_l} \mathbf{A}_l^H) = P/L, 1 \leq l \leq L$, here $\mathbf{R}_{x_l}(1:5, 1:5)$ denotes the first 5 rows and columns of \mathbf{R}_{x_l} and $[\mathbf{I}_5 \mathbf{0}]$ is a 5×8 matrix with its diagonal entries equal to 1 and other entries equal to 0. Figure 4.4 shows that the proposed scheme performs better than the equal allocation scheme. Moreover, the MSE of the proposed scheme goes to the lower bound as P increases while the MSE of the equal allocation scheme approaches a constant MSE for $P > 25$ dB. This is because the proposed scheme takes into account the effects of the observation and channel ma-

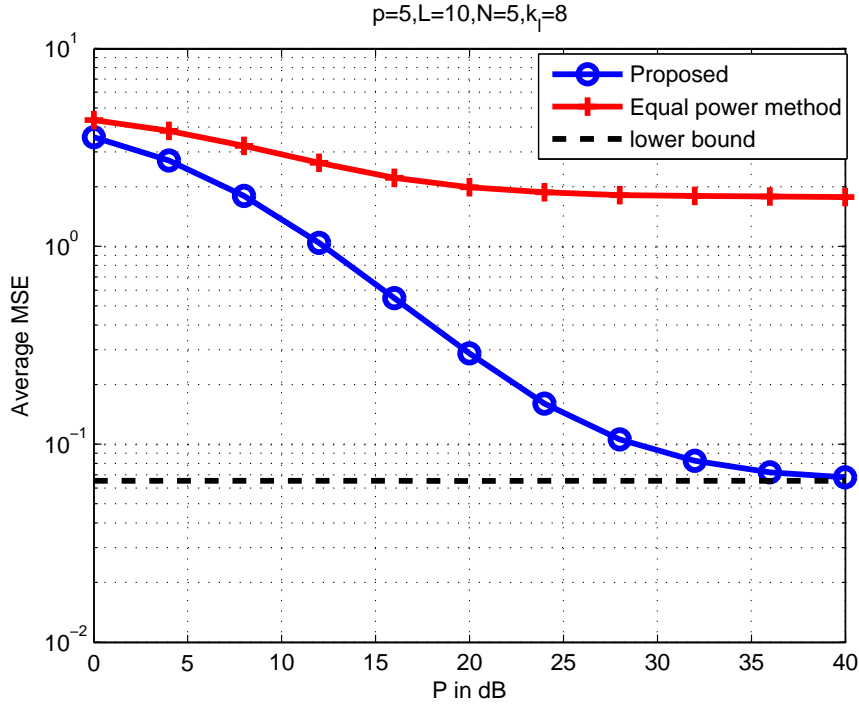


Figure 4.4: MSE comparison between the proposed method and the equal power method

trices in the design of coding matrices, while the equal allocation scheme does not use the information of observation matrices and channel matrices and thus the performance improvement is limited as P increases.

Simulation 4.4 - Comparison with Previous Work [28]: In this simulation, we compare the scheme of our MIMO model to the scheme of the SISO model in [28]. The model used in [28] is shown in Figure 4.5, where a scalar measurement $x_l = f_l \theta + n_l$ at the l th sensor is multiplied by an amplified factor $a_l(n)$ at the n th time instant before it is transmitted to the receiver through a fading channel h_l , $1 \leq l \leq Q$. At time n , the received signal is $y(n) = \sum_{l=1}^Q h_l a_l(n) x_l + \nu(n)$, where $\nu(n)$ is an additive noise. After collecting p received signals $\{y(1), \dots, y(p)\}$ at the FC, the source vector θ is then estimated based on the LMMSE fusion rule. The proposed MIMO scheme is modelled in (4.4) with $L = 1$, a single vector sensor. Figure. 4.6 shows that the average MSE of the proposed scheme with $k_l = 10$ is slightly lower than that of the scheme in [28] with $Q = 10$ when power P is small; as the power increases, the performance of two schemes are very close. Moreover,

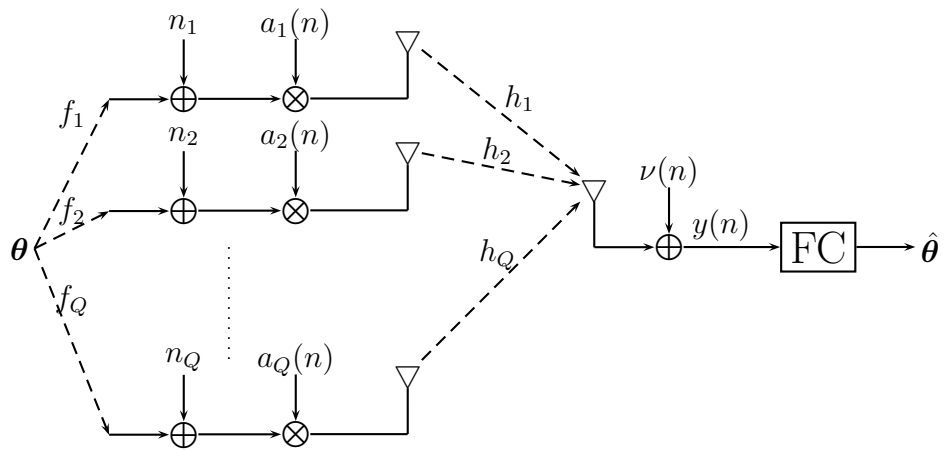


Figure 4.5: Network structure in [28]

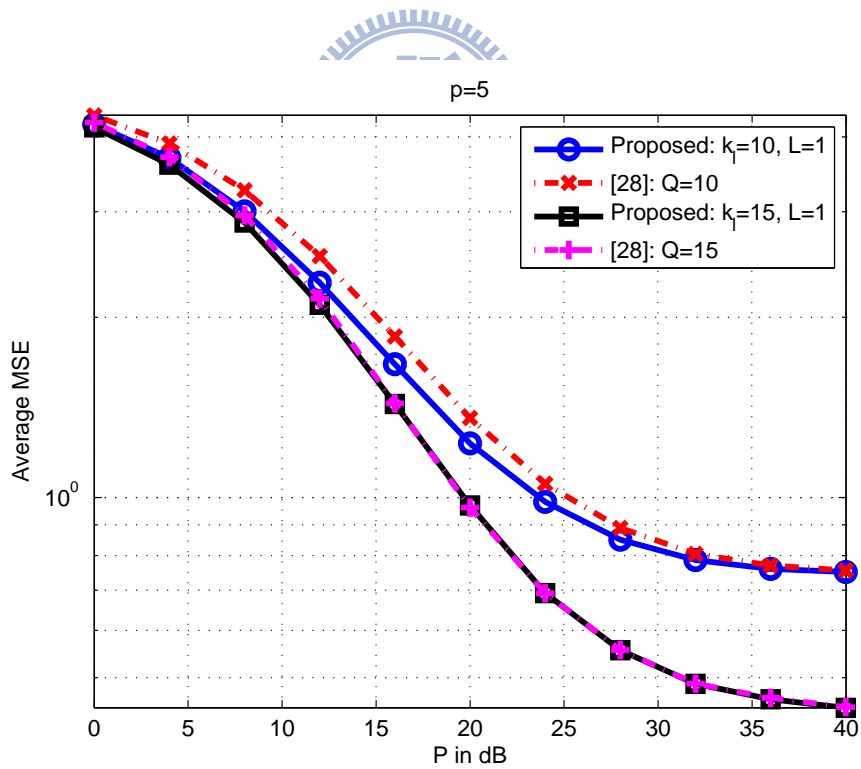


Figure 4.6: MSE comparison between the proposed scheme and that in [28]

for the proposed scheme with $k_l = 15$ and the scheme in [28] with $Q = 15$, we see that the MSEs are roughly the same for over all P .

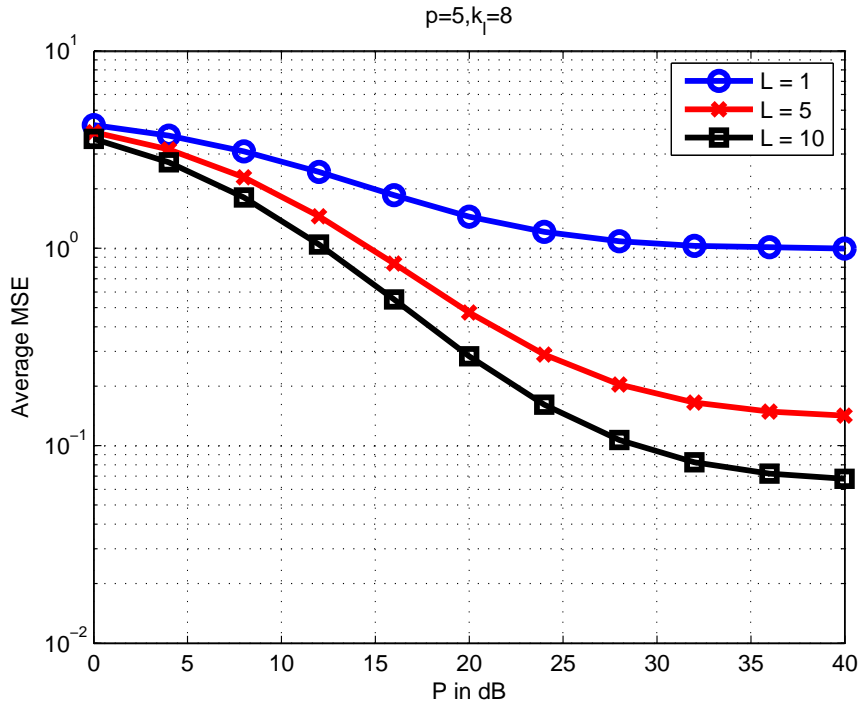
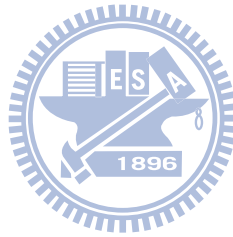


Figure 4.7: MSE versus P with different number of sensors

Simulation 4.5 - Different Number of Sensors: In this simulation, we illustrate the performance of the proposed method for different number of sensors L . The optimal coding matrices in (4.25) are used with $k_l = 8$ for $L = 1, 5$, and 10 . Figure 4.7 shows that for a fixed total power, the MSE performs better as L increases. In addition, the MSE of $L = 1$, whose performance is roughly the same as that in [28], reaches the lower bound (4.27) for $P \geq 32$ dB.

A Brief Summary and Discussion: We study distributed estimation of a random vector signal in MIMO wireless sensor networks. Based on the singular value decomposition, we first derive an upper bound of the objective function, which is an inverse proportion of the MSE. We then show that the minimum number of transmitters to reach the upper bound is equal to the number of the source signals. The extra number of transmitters beyond

the number of the sources does not improve the performance. As taking into account the total power constraint, we propose a method based on specific choices of singular vectors to formulate the problem as a convex optimization problem. As a result, we obtain closed form expressions of the coding matrices for the MIMO sensor network.



Chapter 5

Conclusions

In this thesis, we give investigations on the distributed estimation problem in wireless sensor networks with the coherent multiple access channel model and LMMSE fusion rule. With a total power constraint, we propose optimal power allocation schemes so that the estimated distortion is minimized.

We consider distributed estimation with three different sensor networks. In Chapter 2 and Chapter 3, the SISO sensor networks are discussed. To enhance power efficiency, we study the estimation problem based on cluster-based sensor networks in Chapter 2. The result reveals that for the achievement of the same performance, the collaboration case uses less power than the non-collaboration case. In practice, since channels are often unknown to the receiver in wireless communication systems, we study the estimation problem of non-collaboration sensor network with unknown channels in Chapter 3. In Chapter 4, we extend SISO sensor networks to MIMO sensor networks. Our main contributions are as follows:

- (i) In Chapter 2, we study distributed estimation of a scalar signal using cluster-based sensor network. We show that the optimal amplification matrices are scaled outer products of the observation gain and the channel gain vectors. Based on the result, we give the optimal power schemes for two special cases: the full collaboration case and the non-collaboration case. We show that the collaboration can improve performance; moreover, for a fixed number of sensors, the amount of improvement

is closely related to the amount of collaboration.

- (ii) In Chapter 3, we study distributed estimation of a scalar signal using sensor network with unknown channels. Based on a two-phase approach, we study the optimal and equal power allocation schemes. In each schemes, we compare the performance of the distributed estimation with estimated channels to that with actual known channels. The results reveal that: i) the MSE with estimated channels at the FC approaches a finite nonzero value as the number of sensors increases; ii) the optimal training powers are the same for both schemes; iii) compared with the case when channels are known, the penalty caused by the channel estimation error becomes worse as the number of sensor increases.
- (iii) In Chapter 4, we consider distributed estimation of a vector signal using MIMO sensor network. Based on the singular value decomposition, the problem can be formulated as a convex optimization problem, which follows a water-filling type solution. By the solution and the singular vectors of the observation matrices, closed form expressions for the coding matrices are easy to obtained.

Future investigation can focus on taking into account the communication noise between sensors and the cluster head and the energy consumption caused by sensor collaboration in cluster-based sensor networks. Another interesting direction is vector signal estimation in MIMO wireless sensor networks for i) non-diagonal channel matrices and ii) unknown channel matrices at the fusion center.

Appendix A

Proof of Proposition 2.1

Write $\mathbf{A}_l^T = [\mathbf{a}_{l,1} \ \mathbf{a}_{l,2} \ \cdots \ \mathbf{a}_{l,N_l}]$, where $\mathbf{a}_{l,j} \in \mathbb{R}^{K_l}$. Define slack variables $t_l = \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l = \sum_{n=1}^{N_l} g_{l,n} \mathbf{a}_{l,n}^T \mathbf{f}_l$, $1 \leq l \leq L$. The problem (2.8) is rewritten as

$$\left\{ \begin{array}{l} \min_{\mathbf{a}_{l,n}, t_l} \quad \sum_{l=1}^L \left(\sigma_\theta^2 \sum_{n=1}^{N_l} (\mathbf{a}_{l,n}^T \mathbf{f}_l)^2 + \sigma_n^2 \sum_{n=1}^{N_l} (\mathbf{a}_{l,n}^T \mathbf{a}_{l,n}) \right) \\ \text{subject to} \quad \sum_{n=1}^{N_l} g_{l,n} \mathbf{a}_{l,n}^T \mathbf{f}_l = t_l, \quad 1 \leq l \leq L \\ \sigma_n^2 \sum_{l=1}^L \left(\sum_{n=1}^{N_l} g_{l,n} \mathbf{a}_{l,n}^T \right) \left(\sum_{m=1}^{N_l} g_{l,m} \mathbf{a}_{l,m} \right) + \sigma_\nu^2 = \left(\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} \right)^{-1} \left(\sum_{l=1}^L t_l \right)^2 \end{array} \right. \quad (\text{A.1})$$

The Lagrangian function for (A.1) is

$$\begin{aligned} L(\mathbf{a}_{l,n}, t_l, \lambda_l, \lambda_0) = & \sum_{l=1}^L \left(\sigma_\theta^2 \sum_{n=1}^{N_l} (\mathbf{a}_{l,n}^T \mathbf{f}_l)^2 + \sigma_n^2 \sum_{n=1}^{N_l} (\mathbf{a}_{l,n}^T \mathbf{a}_{l,n}) \right) + \sum_{l=1}^L \lambda_l \left(t_l - \sum_{n=1}^{N_l} g_{l,n} \mathbf{a}_{l,n}^T \mathbf{f}_l \right) + \\ & \lambda_0 \left[\sigma_n^2 \sum_{l=1}^L \left(\sum_{n=1}^{N_l} g_{l,n} \mathbf{a}_{l,n}^T \right) \left(\sum_{m=1}^{N_l} g_{l,m} \mathbf{a}_{l,m} \right) + \sigma_\nu^2 - \left(\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} \right)^{-1} \left(\sum_{l=1}^L t_l \right)^2 \right] \end{aligned}$$

where $\lambda_l, \lambda_0 \in \mathbb{R}$, and the associated necessary conditions for optimality are

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_{l,n}} &= 2\mathbf{a}_{l,n}^T (\sigma_\theta^2 \mathbf{f}_l \mathbf{f}_l^T + \sigma_n^2 \mathbf{I}_{K_l}) - \lambda_l g_{l,n} \mathbf{f}_l^T + 2\lambda_0 \sigma_n^2 g_{l,n} \mathbf{g}_l^T \mathbf{A}_l = \mathbf{0}_{1 \times K_l}, \quad 1 \leq n \leq N_l, 1 \leq l \leq L \\ &\Rightarrow 2\mathbf{A}_l (\sigma_\theta^2 \mathbf{f}_l \mathbf{f}_l^T + \sigma_n^2 \mathbf{I}_{K_l}) - \lambda_l \mathbf{g}_l \mathbf{f}_l^T + 2\lambda_0 \sigma_n^2 \mathbf{g}_l \mathbf{g}_l^T \mathbf{A}_l = \mathbf{0}_{N_l \times K_l}, \quad 1 \leq l \leq L \end{aligned} \quad (\text{A.2})$$

$$\frac{\partial L}{\partial t_l} = \lambda_l - 2\lambda_0 \left(\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} \right)^{-1} \left(\sum_{i=1}^L t_i \right) = 0, \quad 1 \leq l \leq L \quad (\text{A.3})$$

$$\frac{\partial L}{\partial \lambda_l} = t_l - \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l = 0, \quad 1 \leq l \leq L \quad (\text{A.4})$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda_0} &= \sigma_n^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{A}_l^T \mathbf{g}_l + \sigma_\nu^2 - \left(\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} \right)^{-1} \left(\sum_{l=1}^L t_l \right)^2 = 0 \\ &\Rightarrow \sigma_n^2 \sum_{l=1}^L \mathbf{g}_l^T \mathbf{A}_l \mathbf{A}_l^T \mathbf{g}_l + \sigma_\nu^2 = \left(\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} \right)^{-1} \left(\sum_{l=1}^L t_l \right)^2 \end{aligned} \quad (\text{A.5})$$

It follows from (A.3) that $\lambda_1 = \dots = \lambda_L$. Let $\lambda_l = \lambda, \forall l$. It follows from (A.2) that

$$\begin{aligned} \mathbf{A}_l + \lambda_0 \sigma_n^2 \mathbf{g}_l \mathbf{g}_l^T \mathbf{A}_l (\sigma_\theta^2 \mathbf{f}_l \mathbf{f}_l^T + \sigma_n^2 \mathbf{I}_{K_l})^{-1} &= \frac{\lambda}{2} \mathbf{g}_l \mathbf{f}_l^T (\sigma_\theta^2 \mathbf{f}_l \mathbf{f}_l^T + \sigma_n^2 \mathbf{I}_{K_l})^{-1} \\ \Rightarrow \mathbf{A}_l + \lambda_0 \mathbf{g}_l \mathbf{g}_l^T \mathbf{A}_l - \frac{\sigma_\theta^2 \lambda_0 t_l}{\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2} \mathbf{g}_l \mathbf{f}_l^T &= \frac{\lambda}{2(\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2)} \mathbf{g}_l \mathbf{f}_l^T \\ \Rightarrow \mathbf{A}_l &= \frac{\sigma_\theta^2 \lambda_0 t_l}{\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2} (\mathbf{I}_{N_l} + \lambda_0 \mathbf{g}_l \mathbf{g}_l^T)^{-1} \mathbf{g}_l \mathbf{f}_l^T + \frac{\lambda}{2(\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2)} (\mathbf{I}_{N_l} + \lambda_0 \mathbf{g}_l \mathbf{g}_l^T)^{-1} \mathbf{g}_l \mathbf{f}_l^T \\ \Rightarrow \mathbf{A}_l &= \frac{\sigma_\theta^2 \lambda_0 t_l}{(\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2)(1 + \lambda_0 \|\mathbf{g}_l\|^2)} \mathbf{g}_l \mathbf{f}_l^T + \frac{\lambda}{2(\sigma_n^2 + \sigma_\theta^2 \|\mathbf{f}_l\|^2)(1 + \lambda_0 \|\mathbf{g}_l\|^2)} \mathbf{g}_l \mathbf{f}_l^T \end{aligned} \quad (\text{A.6})$$

where in the second equation we use the matrix inversion lemma [36, p.45] and $t_l = \mathbf{g}_l^T \mathbf{A}_l \mathbf{f}_l$. Substituting (A.6) into (A.4), we obtain $t_l = \lambda \|\mathbf{g}_l\|^2 \|\mathbf{f}_l\|^2 / (2\varphi_l)$, where $\varphi_l = \sigma_\theta^2 \|\mathbf{f}_l\|^2 + \sigma_n^2 + \lambda_0 \sigma_n^2 \|\mathbf{g}_l\|^2$, and thus from (A.6), we get $\mathbf{A}_l = \lambda / (2\varphi_l) \mathbf{g}_l \mathbf{f}_l^T$. From (A.3) and $t_l = \lambda \|\mathbf{g}_l\|^2 \|\mathbf{f}_l\|^2 / 2\varphi_l$, we have

$$\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} = \sum_{i=1}^L \frac{\lambda_0 \|\mathbf{g}_i\|^2 \|\mathbf{f}_i\|^2}{\varphi_i} \quad (\text{A.7})$$

Substituting $t_l = \lambda \|\mathbf{g}_l\|^2 \|\mathbf{f}_l\|^2 / (2\varphi_l)$, $\mathbf{A}_l = \lambda / (2\varphi_l) \mathbf{g}_l \mathbf{f}_l^T$, and (A.7) into (A.5), we have

$$\frac{\lambda}{2} = \sqrt{\left(\sum_{i=1}^L \frac{\|\mathbf{g}_i\|^2 \|\mathbf{f}_i\|^2 (\sigma_\theta^2 \|\mathbf{f}_i\|^2 + \sigma_n^2)}{\varphi_i} \right)^{-1} \sigma_\nu \lambda_0} \quad (\text{A.8})$$

From (A.8) and $\mathbf{A}_l = \lambda / (2\varphi_l) \mathbf{g}_l \mathbf{f}_l^T$, the minimum total power can be written as

$$P_{\min} = \sum_{l=1}^L \text{tr} (\sigma_\theta^2 \mathbf{A}_l \mathbf{f}_l \mathbf{f}_l^T \mathbf{A}_l^T + \sigma_n^2 \mathbf{A}_l \mathbf{A}_l^T) = \lambda_0 \sigma_\nu^2.$$

It follows that

$$\lambda_0 = P_{\min}/\sigma_\nu^2 \quad (\text{A.9})$$

and thus from (A.7), we get

$$\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} = \sum_{i=1}^L \frac{P_{\min} \|\mathbf{g}_i\|^2 \|\mathbf{f}_i\|^2}{\sigma_\nu^2 (\sigma_\theta^2 \|\mathbf{f}_i\|^2 + \sigma_n^2) + P_{\min} \sigma_n^2 \|\mathbf{g}_i\|^2} \quad (\text{A.10})$$

Equation (A.10) gives the relation between the achieved minimum power and the constraint J^* on MSE. The optimal amplification matrices are $\mathbf{A}_l = \lambda/(2\varphi_l) \mathbf{g}_l \mathbf{f}_l^T$, where λ and φ_l depends on P_{\min} through (A.9). In view of observation (iii) in Section III, if we set $P_{\min} = P$, the corresponding MSE is given in (2.10) and the corresponding amplification matrices is in (2.9).



Appendix B

Proof of Lemma 2.3

We first show that for $\mathbf{x} = [x_1 \cdots x_n]^T \in \mathbb{R}^n$ and $\mathbf{y} = [y_1 \cdots y_n]^T \in \mathbb{R}^n$, the following inequality holds

$$\frac{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \geq \sum_{i=1}^n \frac{x_i^2 y_i^2}{x_i^2 + y_i^2}, \quad (\text{B.1})$$

or equivalently,

$$\begin{aligned} & \frac{\sum_{i=1}^n x_i^2 \sum_{j=1}^n y_j^2}{\sum_{j=1}^n (x_j^2 + y_j^2)} \geq \frac{\sum_{i=1}^n \left(x_i^2 y_i^2 \prod_{k \neq i}^n (x_k^2 + y_k^2) \right)}{\prod_{k=1}^n (x_k^2 + y_k^2)} \\ \Leftrightarrow & \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{j=1}^n y_j^2 \right) \prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{j=1}^n (x_j^2 + y_j^2) \right] \left[\sum_{i=1}^n \left(x_i^2 y_i^2 \prod_{k \neq i}^n (x_k^2 + y_k^2) \right) \right] \geq 0 \end{aligned}$$

The left hand side of the above inequality can be written as

$$\begin{aligned} & \left[\sum_{i=1}^n \sum_{j=1}^n x_i^2 y_j^2 \right] \prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n \sum_{j=1}^n x_i^2 y_i^2 (x_j^2 + y_j^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \\ = & \left[\sum_{i=1}^n \sum_{j=i}^n x_i^2 y_j^2 \right] \prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n x_i^2 y_i^2 \right] \prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n \sum_{j \neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \\ = & \left[\sum_{i=1}^n \sum_{j \neq i}^n x_i^2 y_j^2 \right] \prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n \sum_{j \neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \\ = & \left[\sum_{i=1}^n \sum_{j \neq i}^n x_i^2 y_j^2 (x_i^2 + y_i^2) - \sum_{i=1}^n \sum_{j \neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \\ = & \left[\sum_{i=1}^n \sum_{j \neq i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \\ = & \left[\sum_{i=1}^n x_i^2 \sum_{j > i}^n (x_i^2 y_j^2 - x_j^2 y_i^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) + \left[\sum_{i=1}^n x_i^2 \sum_{j < i}^n (x_i^2 y_j^2 - x_j^2 y_i^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) \end{aligned}$$

$$\begin{aligned}
&= \left[\sum_{i=1}^n \sum_{j>i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2) \right] \prod_{k \neq i}^n (x_k^2 + y_k^2) + \left[\sum_{i=1}^n \sum_{j>i}^n x_j^2 (x_j^2 y_i^2 - x_i^2 y_j^2) \right] \prod_{k \neq j}^n (x_k^2 + y_k^2) \\
&= \left[\sum_{i=1}^n \sum_{j>i}^n x_i^2 (x_j^2 + y_j^2) (x_i^2 y_j^2 - x_j^2 y_i^2) - x_j^2 (x_i^2 + y_i^2) (x_i^2 y_j^2 - x_j^2 y_i^2) \right] \cdot \prod_{k \neq i, j}^n (x_k^2 + y_k^2) \\
&= \sum_{i=1}^n \sum_{j>i}^n (x_i^2 y_j^2 - x_j^2 y_i^2)^2 \prod_{k \neq i, j}^n (x_k^2 + y_k^2) \geq 0
\end{aligned}$$

Thus, we obtain (B.1). Now, let $\mathbf{x} = [\mathbf{x}_1^T \cdots \mathbf{x}_L^T]^T = [x_1 \cdots x_n]^T$ and $\|\mathbf{x}_l\|^2 = \tilde{x}_l^2$, then we have

$$\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2 = \sum_{l=1}^L \|\mathbf{x}_l\|^2 = \sum_{l=1}^L \tilde{x}_l^2 = \|\tilde{\mathbf{x}}\|^2 \quad (\text{B.2})$$

where $\tilde{\mathbf{x}} = [\tilde{x}_1 \cdots \tilde{x}_L]^T$. By the same way, we have

$$\|\mathbf{y}\|^2 = \sum_{i=1}^n y_i^2 = \sum_{l=1}^L \|\mathbf{y}_l\|^2 = \sum_{l=1}^L \tilde{y}_l^2 = \|\tilde{\mathbf{y}}\|^2 \quad (\text{B.3})$$

From (B.1), (B.2), and (B.3), we have

$$\frac{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} = \frac{\|\tilde{\mathbf{x}}\|^2 \|\tilde{\mathbf{y}}\|^2}{\|\tilde{\mathbf{x}}\|^2 + \|\tilde{\mathbf{y}}\|^2} \geq \sum_{l=1}^L \frac{\tilde{x}_l^2 \tilde{y}_l^2}{\tilde{x}_l^2 + \tilde{y}_l^2} = \sum_{l=1}^L \frac{\|\mathbf{x}_l\|^2 \|\mathbf{y}_l\|^2}{\|\mathbf{x}_l\|^2 + \|\mathbf{y}_l\|^2}$$

and the result follows.

Appendix C

Derivation of (3.9)

We first show that given $\hat{\mathbf{h}}$, ε is uncorrelated with ν , $\sum_{k=1}^K \hat{h}_k \alpha_k n_k$, and $\sum_{k=1}^K \hat{h}_k \alpha_k \theta$. Since $E[\nu^* \theta] = 0$, $E[h_k^* \nu] = 0$, and $E[\nu^* n_l] = 0$, $E[\nu^* \varepsilon | \hat{\mathbf{h}}] = 0$. We show that ε and $\sum_{k=1}^K \hat{h}_k \alpha_k n_k$ are uncorrelated as follows:

$$\begin{aligned} E \left[\left(\sum_{k=1}^K \hat{h}_k \alpha_k n_k \right)^* \varepsilon \mid \hat{\mathbf{h}} \right] &= E \left[\left(\sum_{k=1}^K \hat{h}_k^* (h_k - \hat{h}_k) |\alpha_k|^2 |n_k|^2 \right) \mid \hat{\mathbf{h}} \right] \\ &= \sum_{k=1}^K \hat{h}_k^* \left(E[h_k \mid \hat{\mathbf{h}}] - \hat{h}_k \right) |\alpha_k|^2 \sigma_n^2 = 0 \end{aligned} \quad (\text{C.1})$$

where the first equality uses that $E[n_k^* \theta] = 0$ and $E[n_k^* n_l] = 0$ for $k \neq l$. The last equality follows because $\hat{h}_k = E[h_k | y_k] = E[h_k | \hat{h}_k]$. Similarly,

$$\begin{aligned} E \left[\left(\sum_{k=1}^K \hat{h}_k \alpha_k \theta \right)^* \varepsilon \mid \hat{\mathbf{h}} \right] &= E \left[\left(\sum_{k=1}^K \sum_{l=1}^K \hat{h}_k^* (h_l - \hat{h}_l) \alpha_k^* \alpha_l |\theta|^2 \right) \mid \hat{\mathbf{h}} \right] \\ &= \sum_{k=1}^K \sum_{l=1}^K \hat{h}_k^* \left(E[h_l \mid \hat{\mathbf{h}}] - \hat{h}_l \right) \alpha_k^* \alpha_l \sigma_\theta^2 = 0 \end{aligned} \quad (\text{C.2})$$

Finally the conditional variance

$$\begin{aligned} E \left[|\varepsilon|^2 \mid \hat{\mathbf{h}} \right] &= E \left[\sum_{k=1}^K (h_k - \hat{h}_k)^* (h_k - \hat{h}_k) |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2) \mid \hat{\mathbf{h}} \right] \\ &= (\sigma_\theta^2 + \sigma_n^2) \sum_{k=1}^K E \left[(h_k - \hat{h}_k)^* (h_k - \hat{h}_k) \mid \hat{\mathbf{h}} \right] |\alpha_k|^2 \\ &= (\sigma_\theta^2 + \sigma_n^2) \sum_{k=1}^K \delta_k^2 |\alpha_k|^2 = (\sigma_\theta^2 + \sigma_n^2) \delta_1^2 \sum_{k=1}^K |\alpha_k|^2 \end{aligned} \quad (\text{C.3})$$

where the last equality uses $\delta_1 = \dots = \delta_k$ in (3.5). By (C.1) to (C.3) and (3.6), equality (3.9) follows.

Appendix D

Proof of Proposition 3.1

According to the observations (i) to (iii) in Section 2.2, instead of solving (3.20) directly, we consider the following problem

$$\begin{cases} \min_{P_t, |\alpha_k|} & (\sigma_\theta^2 + \sigma_n^2) \sum_{k=1}^K |\alpha_k|^2 + P_t \\ \text{subject to} & \left(\frac{1}{\sigma_\theta^2} + \frac{\zeta^2 (\sum_{k=1}^K \hat{g}_k |\alpha_k|)^2 P_t}{\zeta^2 \sigma_n^2 (\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2) P_t + \zeta P_t + K \zeta (\sigma_\theta^2 + \sigma_n^2) (\sum_{k=1}^K |\alpha_k|^2) + K} \right)^{-1} = J \end{cases}$$

where $0 < J \leq \sigma_\theta^2$. Let $t = \sum_{k=1}^K \hat{g}_k |\alpha_k|$, the optimization problem becomes

$$\begin{cases} \min_{P_t, |\alpha_k|, t} & (\sigma_\theta^2 + \sigma_n^2) \sum_{k=1}^K |\alpha_k|^2 + P_t \\ \text{subject to} & \sum_{k=1}^K \hat{g}_k |\alpha_k| - t = 0 \\ & \zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2 \right) P_t + \zeta P_t + K \zeta (\sigma_\theta^2 + \sigma_n^2) (\sum_{k=1}^K |\alpha_k|^2) + K = \left(\frac{1}{J} - \frac{1}{\sigma_\theta^2} \right)^{-1} \zeta^2 t^2 P_t \end{cases}$$

The Lagrangian is

$$\begin{aligned} L(|\alpha_k|, P_t, t, \lambda, \lambda_0) = & (\sigma_\theta^2 + \sigma_n^2) \sum_{k=1}^K |\alpha_k|^2 + P_t + \lambda \left(\sum_{k=1}^K \hat{g}_k |\alpha_k| - t \right) + \lambda_0 \left[\zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2 \right) P_t + \right. \\ & \left. \zeta P_t + K \zeta (\sigma_\theta^2 + \sigma_n^2) (\sum_{k=1}^K |\alpha_k|^2) + K - \left(\frac{1}{J} - \frac{1}{\sigma_\theta^2} \right)^{-1} \zeta^2 t^2 P_t \right] \end{aligned}$$

where $\lambda, \lambda_0 \in \mathbb{R}$, and the associated necessary conditions for optimality are

$$\frac{\partial L}{\partial |\alpha_k|} = 2(\sigma_\theta^2 + \sigma_n^2)|\alpha_k| + \lambda \hat{g}_k + \lambda_0 [2\zeta^2 \sigma_n^2 \hat{g}_k^2 P_t |\alpha_k| + 2K\zeta(\sigma_\theta^2 + \sigma_n^2)|\alpha_k|] = 0 \quad (\text{D.1})$$

$$\frac{\partial L}{\partial P_t} = 1 + \lambda_0 \left[\zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2 \right) + \zeta - \left(\frac{1}{J} - \frac{1}{\sigma_\theta^2} \right)^{-1} \zeta^2 t^2 \right] = 0 \quad (\text{D.2})$$

$$\frac{\partial L}{\partial t} = -\lambda - 2\lambda_0 \left(\frac{1}{J} - \frac{1}{\sigma_\theta^2} \right)^{-1} \zeta^2 P_t t = 0 \quad (\text{D.3})$$

$$\frac{\partial L}{\partial \lambda} = \sum_{k=1}^K \hat{g}_k |\alpha_k| - t = 0 \quad (\text{D.4})$$

$$\frac{\partial L}{\partial \lambda_0} = \zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^2 |\alpha_k|^2 \right) P_t + \zeta P_t + K\zeta(\sigma_\theta^2 + \sigma_n^2) \left(\sum_{k=1}^K |\alpha_k|^2 \right) + K - \left(\frac{1}{J} - \frac{1}{\sigma_\theta^2} \right)^{-1} \zeta^2 t^2 P_t = 0 \quad (\text{D.5})$$

From (D.1), $|\alpha_k| = -\lambda \hat{g}_k / (2\hat{\phi}_k)$, where $\hat{\phi}_k = (\sigma_\theta^2 + \sigma_n^2) + \lambda_0 \zeta^2 \sigma_n^2 \hat{g}_k^2 P_t + \lambda_0 K \zeta (\sigma_\theta^2 + \sigma_n^2)$, thus it follows from (D.4) that $t = -(\lambda/2) \sum_{k=1}^K (\hat{g}_k^2 / \hat{\phi}_k)$ and then from (D.3), we have

$$\frac{1}{J} - \frac{1}{\sigma_\theta^2} = \lambda_0 \zeta^2 P_t \sum_{k=1}^K \frac{\hat{g}_k^2}{\hat{\phi}_k} = \zeta^2 P_t \sum_{k=1}^K \frac{\hat{g}_k^2}{(\sigma_\theta^2 + \sigma_n^2)/\lambda_0 + \zeta^2 \sigma_n^2 \hat{g}_k^2 P_t + K\zeta(\sigma_\theta^2 + \sigma_n^2)} \quad (\text{D.6})$$

Use $t = -\lambda(1/J - 1/\sigma_\theta^2)/(2\mu\zeta^2 P_t)$ from (D.3) and $|\alpha_k| = -\lambda \hat{g}_k / (2\hat{\phi}_k)$ in (D.2) to get

$$\frac{\lambda^2}{4} = \frac{\frac{1}{\lambda_0} + \zeta}{\frac{\frac{1}{J} - \frac{1}{\sigma_\theta^2}}{\lambda_0^2 \zeta^2 P_t^2} - \zeta^2 \sigma_n^2 \left(\sum_{k=1}^K \hat{g}_k^4 / \hat{\phi}_k^2 \right)} = \frac{(1 + \lambda_0 \zeta)/(1 + K\lambda_0 \zeta)}{\sum_{k=1}^K \hat{g}_k^2 (\sigma_\theta^2 + \sigma_n^2) / \hat{\phi}_k^2} P_t. \quad (\text{D.7})$$

where the last equality follows from (D.6). Since the data power for the k th sensor is $P_k = |\alpha_k|^2 (\sigma_\theta^2 + \sigma_n^2) = (\lambda^2/4) (\hat{g}_k^2 (\sigma_\theta^2 + \sigma_n^2) / \hat{\phi}_k^2)$, the total power for data transmission is $\sum_{k=1}^K P_k = (\lambda^2/4) \sum_{k=1}^K (\hat{g}_k^2 (\sigma_\theta^2 + \sigma_n^2) / \hat{\phi}_k^2) = (1 + \lambda_0 \zeta) P_t / (1 + K\lambda_0 \zeta)$. With the total network power constraint P , it follows from (D.2) and (D.5) that

$$\lambda_0 = \frac{P_t}{K + K\zeta(P - P_t)}. \quad (\text{D.8})$$

where we use $\sum_{k=1}^K P_k = P - P_t$. Moreover, since $\sum_{k=1}^K P_k + P_t = P$, we have

$$\frac{2 + (K + 1)\zeta\lambda_0}{1 + K\zeta\lambda_0} P_t = P \quad (\text{D.9})$$

Substituting (D.8) into (D.9), we get the optimal training power in (3.21). With P_t^{opt} and λ_0 , we get P_k^{opt} in (3.22) and the MSE in (3.23) follows from (D.6).

Appendix E

Derivation of (3.25)

Rewrite (3.23) as

$$J(P, K) = \left(\frac{1}{\sigma_\theta^2} + \frac{1}{\sigma_\theta^2 + \sigma_n^2} b(K) \sum_{k=1}^K \frac{\hat{g}_k^2}{\gamma b(K) \hat{g}_k^2 + K} \right)^{-1}$$

where $b(K) = [K (\sqrt{\zeta P + 1} - \sqrt{1 + \zeta P/K}) / (K - 1)]^2$ and $\gamma = \sigma_n^2 / (\sigma_\theta^2 + \sigma_n^2)$. Note that $\lim_{K \rightarrow \infty} b(K) = (\sqrt{\zeta P + 1} - 1)^2$. We will show that the sum inside the parentheses converges to $E[\hat{g}_k^2] = 1$ as $K \rightarrow \infty$. Since

$$\frac{\hat{g}_k^2}{K} - \frac{\hat{g}_k^2}{\gamma b(K) \hat{g}_k^2 + K} = \frac{\gamma b(K) \hat{g}_k^4}{K [\gamma b(K) \hat{g}_k^2 + K]} \leq \frac{\gamma b(K) \hat{g}_k^4}{K^2},$$

we have $\hat{g}_k^2/K - \gamma b(K) \hat{g}_k^4/K^2 \leq \hat{g}_k^2/[\gamma b(K) \hat{g}_k^2 + K] \leq \hat{g}_k^2/K$ and thus

$$\sum_{k=1}^K \frac{\hat{g}_k^2}{K} - \sum_{k=1}^K \frac{\gamma b(K) \hat{g}_k^4}{K^2} \leq \sum_{k=1}^K \frac{\hat{g}_k^2}{\gamma b(K) \hat{g}_k^2 + K} \leq \sum_{k=1}^K \frac{\hat{g}_k^2}{K}$$

It follows from the law of large numbers that as $K \rightarrow \infty$, we have $\sum_{k=1}^K \hat{g}_k^2/K = E[\hat{g}_k^2] = 1$,

$$\sum_{k=1}^K \frac{\gamma b(K) \hat{g}_k^4}{K} = \gamma (\sqrt{\zeta P + 1} - 1)^2 E[\hat{g}_k^4], \text{ and } \sum_{k=1}^K \frac{\gamma b(K) \hat{g}_k^4}{K^2} = 0$$

because $E[\hat{g}_k^4]$ is finite. Therefore,

$$\sum_{k=1}^K \frac{\hat{g}_k^2}{\gamma b(K) \hat{g}_k^2 + K} = 1, \text{ as } K \rightarrow \infty$$

and (3.25) follows.

Appendix F

Convexity of Objective Function in (3.33)

To show that the second derivative of the objective function in (3.33) is positive, we let $c_1 = K \frac{\beta}{1+\beta} (\frac{1}{K} \sum_{k=1}^K \hat{g}_k)^2$ and $c_2 = \frac{1}{1+\beta} (\frac{1}{K} \sum_{k=1}^K \hat{g}_k^2)$. Then the objective function in (35) can be written as

$$\begin{aligned} f(P_t) &= -\frac{\zeta^2 c_1 (P - P_t) P_t}{\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K} \\ \frac{df(P_t)}{dP_t} &= -\frac{\zeta^2 c_1 [\zeta (K - 1) P_t^2 - 2K (\zeta P + 1) P_t + KP (\zeta P + 1)]}{[\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K]^2} \\ \frac{d^2 f(P_t)}{dP_t^2} &= -\frac{2\zeta^2 c_1 X}{[\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K]^3} \end{aligned}$$

where

$$\begin{aligned} X &= [\zeta (K - 1) P_t - K (\zeta P + 1)] [\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K] \\ &\quad - [\zeta (K - 1) P_t^2 - 2K (\zeta P + 1) P_t + KP (\zeta P + 1)] [\zeta^2 c_2 (P - 2P_t) + \zeta - \zeta K] \\ &= \zeta^2 c_2 \{ (P - P_t) P_t [\zeta (K - 1) P_t - K (\zeta P + 1)] - (P - 2P_t) [\zeta (K - 1) P_t^2 + K (\zeta P + 1) (P - 2P_t)] \} \\ &\quad + [\zeta P_t + \zeta K (P - P_t) + K] [\zeta (K - 1) P_t - K (\zeta P + 1)] + \zeta (K - 1) [\zeta (K - 1) P_t^2 + K (\zeta P + 1) (P - 2P_t)] \\ &= -\zeta^2 c_2 c_3 - K (\zeta P + 1) (\zeta P + K) \end{aligned}$$

with $c_3 = K \zeta (P - P_t)^3 + K (P - 2P_t)^2 + KP_t (P - P_t) + \zeta P_t^3$. Since $P - P_t > 0$, we have $c_3 > 0$ and $X < 0$, and thus $(d^2 f(P_t)/dP_t^2) > 0$. Hence the result follows.

Appendix G

Proof of Proposition 3.2

Let $c_1 = K \frac{\beta}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2$ and $c_2 = \frac{1}{1+\beta} \left(\frac{1}{K} \sum_{k=1}^K \hat{g}_k^2 \right)$, then the Lagrangian of the problem (3.33) is

$$L(P_t, \mu_1, \mu_2) = -\frac{\zeta^2 c_1 (P - P_t) P_t}{\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K} + \mu_1 (P_t - P) - \mu_2 P_t$$

and the associated KKT conditions are

$$-\frac{\zeta^2 c_1 [\zeta(K-1)P_t^2 - 2K(\zeta P + 1)P_t + KP(\zeta P + 1)]}{(\zeta^2 c_2 (P - P_t) P_t + \zeta P_t + \zeta K (P - P_t) + K)^2} + \mu_1 - \mu_2 = 0 \quad (\text{G.1})$$

$$\mu_1 (P_t - P) = 0, \quad \mu_1 \geq 0 \quad (\text{G.2})$$

$$\mu_2 P_t = 0, \quad \mu_2 \geq 0 \quad (\text{G.3})$$

Since the training power have to be greater than 0, we have $\mu_2 = 0$. If $\mu_1 > 0$, then $P_t = P$, but then (G.1) leads to $\mu_1 < 0$ a contradiction. Therefore, we have $\mu_1 = \mu_2 = 0$ and $P > P_t > 0$. From (G.1), we have

$$\begin{aligned} & \zeta(K-1)P_t^2 - 2K(\zeta P + 1)P_t + KP(\zeta P + 1) = 0 \\ \Rightarrow P_t &= \frac{K(\zeta P + 1) \pm \sqrt{K(\zeta P + 1)(\zeta P + K)}}{\zeta(K-1)} \end{aligned}$$

where we take negative term since positive term can not satisfy the constraint $P \geq P_t$. Let $a = K(\zeta P + 1)$ and $b = \zeta P + K$, then we have $P_t^{opt} = (a - \sqrt{ab})/[\zeta(K-1)]$ and

$P - P_t^{opt} = (b - \sqrt{ab})/[\zeta(K - 1)]$, and from (3.32), (3.35) follows:

$$\begin{aligned} J(P, K) &= \sigma_\theta^2 \left(1 + \frac{c_1[(a + b)\sqrt{ab} - 2ab]}{c_2[(a + b)\sqrt{ab} - 2ab] + (K - 1)^2\sqrt{ab}} \right)^{-1} \\ &= \sigma_\theta^2 \left(1 + \frac{c_1}{c_2 + \frac{(K-1)^2}{(\sqrt{a}-\sqrt{b})^2}} \right)^{-1} \end{aligned}$$

where the first equality uses that

$$\begin{aligned} &(K - 1)(a - \sqrt{ab}) + K(K - 1)(\sqrt{ab} - b) + K(K - 1)^2 \\ &= (K - 1)[(K - 1)\sqrt{ab} + \underbrace{a - Kb + K^2 - K}_{=0}] = (K - 1)^2\sqrt{ab}. \end{aligned}$$



Bibliography

- [1] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM Mobile Computing and Communications Review*, vol. 5, no. 2, pp. 28-36, Apr. 2002.
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communication Magazine*, vol. 40, no. 8, pp. 102-114, Aug. 2002.
- [3] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6-28, Dec. 2004.
- [4] J. Kulik, W. R. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, no. 2/3, pp. 169-185, Mar.-May 2002.
- [5] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Communications*, vol. 1, no. 4, pp. 660-669, Oct. 2002.
- [6] P. Kumarawadu, D. J. Dechene, M. Luccini, and A. Sauer, "Algorithms for node clustering in Wireless Sensor Networks: a survey," *Proc. of IEEE ICIAFS 2008*, pp. 295-300, Dec. 2008.
- [7] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *IEEE Network*, vol. 20, no. 3, pp. 20-25, May/June. 2006.

- [8] B. Chen, L. Tong, and P. K. Varshney, "Channel-Aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16-26, Jul. 2006.
- [9] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27-41, Jul. 2006.
- [10] M. Çetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42-55, Jul. 2006.
- [11] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in sensor network with correlated sensor noise," *EURASIP J. Wireless Communications and Networking*, vol. 4, pp. 473-482, 2005.
- [12] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 413-422, Feb. 2006.
- [13] J. Y. Wu, Q. Z. Huang, and T. S. Lee, "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 2171-2176, May 2008.
- [14] J. Li and G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 57, no. 10, pp. 3746-3758, Oct. 2009.
- [15] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, part I: Gaussian case," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1131-1143, Mar. 2006.
- [16] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Trans. Signal Processing*, vol. 56, no. 4, pp. 1398-1410, Apr. 2008.

- [17] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," *IEEE Trans. Information Theory*, vol. 55, no. 12, pp. 5682-5699, Dec. 2009.
- [18] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source channel communication revisited," *IEEE Trans. Information Theory*, vol. 49, no. 5, pp. 1147-1158, May 2003.
- [19] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Processing*, vol. 55, no. 9, pp. 4683-4695, Sep. 2007.
- [20] I. Bahceci and A. J. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Trans. Communications*, vol. 56, no. 7, pp. 1146-1156, Jul. 2008.
- [21] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1631-1639, May. 2005.
- [22] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Processing*, vol. 55, no. 8, pp. 4284-4299, Aug. 2007.
- [23] J. Fang and H. Li, "Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4339-4353, Aug. 2010.
- [24] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. Wireless Communications*, vol. 8, no. 7, pp. 3822-3832, Jul. 2009.
- [25] H. Şenol and C. Tepedelenlioğlu, "Performance of distributed estimation over unknown parallel fading channels," *IEEE Trans. Signal Processing*, vol. 56, no. 12, pp. 6057-6068, Dec. 2008.

- [26] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 757-770, Feb. 2008.
- [27] N. Khajehnouri and A. H. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3336-3348, Jul. 2007.
- [28] W. Guo, J-J Xiao, and S. Cui, "An efficient water-filling solution for linear coherent joint estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5301-5305, Oct. 2008.
- [29] H. Behroozi, F. Alajaji, and T. Linder, "On the optimal performance in asymmetric Gaussian wireless sensor networks with fading," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2436-2441, Apr. 2010.
- [30] M. K. Banavar, C. Tepedelenlioglu, and A. Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *IEEE Trans. Signal Processing*, vol. 58, no. 1, pp. 414-425, Jan. 2010.
- [31] K. Liu, H. El-Gamal, and A. Sayeed, "On optimal parametric field estimation in sensor networks," *Proc. IEEE/SP 13th Workshop on Statist. Signal Processing*, pp. 1170-1175, Jul. 2005.
- [32] B. Su and P. P. Vaidyanathan, "A generalized algorithm for blind channel identification with linear redundant precoders," *EURASIP J. Advances in Signal Processing*, vol. 2007, no. 1, pp. 1-13, Jan. 2007.
- [33] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communication*, Cambridge, U.K: Cambridge Univ. Press, 2003.
- [34] D. Tse and P. P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge, U.K: Cambridge Univ. Press, 2005.
- [35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall PTR, 1993.

- [36] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, 2005.
- [37] G. L. Stüber, *Principles of Mobile Communication*, Norwell, MA: Kluwer, 2001.
- [38] S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge, U.K: Cambridge Univ. Press, 2003.
- [39] H. Stark and J. W. Woods, *Probability and Random Processes with Application to Signal Processing*, Prentice-Hall, Inc., 2002.

