

OPTIMIZED RESOURCE ALLOCATION IN WIRELESS SYSTEMS

By

Umar Rashid

B.Sc. in Electrical Engineering,

University of Engineering and Technology, Lahore, Pakistan

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY



Faculty of Engineering and Information Technology,

University of Technology, Sydney

Australia

January 2013

Abstract

Modern wireless systems rely to a great extent on the judicious distribution of available resources (e.g. power, bandwidth) to meet an ever increasing demand of better quality-of-service (QoS). Scarcity of these resources with time, coupled with the tremendous growth in numbers of users, network throughput, and applications, have resulted in making the problem of optimal resource allocation extremely important especially in wireless networks.

Generally, optimization problems posed in the resource allocation framework are nonconvex and thus render it difficult to find an optimal solution. Previous studies on this subject have reported only numerically cumbersome and non-tractable solutions. This dissertation attempts to exploit the hidden convexity of the resource allocation problems under some given performance criteria such as minimum mean square error (MMSE) or signal-to-interference-plus-noise ratio (SINR) and then successfully finds tractable optimization formulations.

The first research problem deals with the optimal power allocation and sensor assignment in linear and nonlinear networks for static and dynamic target tracking. The proposed method casts power allocation as a semi-definite program (SDP) while sensor selection is solved via d.c. (difference of convex functions/sets) programming. The second problem considers optimal beamforming and source power allocation in relay-assisted multiuser communication. This problem is further extended to include multiple-antenna systems to exploit spatial diversity in modern cellular communication by jointly optimizing source precoding and relay processing matrices. Supremacy of the proposed d.c. programming based iterative algorithm over existing methods is demonstrated via extensive simulations.

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UTS or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UTS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed: Umar Rashid

Date: May 23, 2013

Acknowledgements

First and foremost, I would like to thank Almighty Allah Subhanahu Wataala for His countless bounties and blessings upon me. Then I am grateful and proud for being Ummati of my Lord Prophet Muhammad (Sallal Lahu Alaih-e-Wasallam).

I also express my sincere gratitude and appreciation to my supervisor, Professor Hoang D. Tuan, for all the guidance and encouragement he has offered me throughout the period of this research. In addition to learning tricks of the trade, I have tremendously benefited from him in terms of acquiring essential skills for my professional life.

I am also grateful to my co-supervisor Dr H. H. Kha for his valuable contribution and technical assistance throughout my whole PhD tenure. He has actively participated in almost all of my research publication for which I am greatly indebted to him. I am thankful to Professor Ha H. Nguyen from University of Saskatchewan, Canada for his contribution in my projects on relay beamforming for multiuser communication. He has also been a mentor and a constant contributor in my research work. I am also obliged to Professor Pierre Apkarian from Institut de Mathematiques, Universite Paul Sabatier, France for his guidance and assistance in the field of nonlinear filtering in data fusion problems.

I would also like to take this opportunity to extend my thanks to Dean of Faculty of Engineering and Information Technology at University of Technology, Sydney, Professor Hung Nguyen who granted me tuition waiving scholarship which enabled me to fully focus on my research. My gratitude also goes to the generous financial support provided by Australian Research Council (ARC) Discovery Project for my graduate studies. I am thankful to them for putting their faith in my capabilities and I hope I have delivered.

Finally, I would like to thank my parents for their constant love and support, and for encour-

aging me to pursue higher studies.

Contents

Abstract	i
Originality Statement	ii
Acknowledgements	iii
Contents	v
List of Figures	viii
List of Tables	xi
List of Acronyms	xii
1 Introduction	1
1.1 Motivation and Scope	1
1.2 Problems in Resource Allocation in Wireless Networks	3
1.2.1 Distributed Estimation	3
1.2.2 Resource Allocation for Cooperative Networks	4
1.2.3 Throughput Allocation with Antenna Array Beamforming	5
1.3 Dissertation Outline	5
1.4 List of Publications	7
1.5 Notations	10

2	Fundamentals of Statistical Estimation, Wireless Communication and Optimization	
	Theory	11
2.1	Statistical Estimation	11
2.1.1	Estimation Foundation	12
2.1.2	Basics of MMSE	14
2.2	Wireless Channel	20
2.2.1	Propagation and Fading	21
2.2.2	Diversity in Wireless Communication	24
2.3	Optimization Theory	25
2.3.1	Convex Optimization	26
2.3.2	Duality Theory	28
2.3.3	d.c. Programming	29
2.3.4	Penalty Function Method	30
3	Globally Optimized Power Allocation and Sensor Assignment for Linear and Nonlinear Networks	32
3.1	Introduction	33
3.2	Global optimized decentralized Bayes filtering	35
3.2.1	Decentralized Bayes filtering for LSN	37
3.2.2	Decentralized Bayes Filtering for NSN	38
3.3	Global optimal decentralized Bayes filtering for tracking dynamic objects	41
3.4	Joint Optimization of Active Sensor Assignment and Power Allocation in Sensor Networks	46
3.4.1	Problem Formulation	47
3.4.2	Computational Methodology	51
3.5	Simulation Results	53
3.5.1	Decentralized Bayes filtering for locating static targets and the optimized sensor selection	54
3.5.2	Decentralized Bayes filtering for locating dynamic targets	61
3.5.3	Joint sensor selection and power allocation	67

3.6	Summary	70
4	Relay Beamforming Designs in Multi-User Wireless Relay Networks	72
4.1	Introduction	73
4.2	Maximin Throughput Optimization: Problem Formulations and Challenges . . .	76
4.3	Maximin Throughput Optimization by d.c. Programming	81
4.4	Beamforming Design with Orthogonal Source Transmissions	85
4.5	Joint Optimization of Source Power Allocation and Relay Beamforming	88
4.5.1	System Model and Formulation of the Joint Design Problem	89
4.5.2	Proposed D.C. Formulation and Iterative Algorithm	92
4.6	Numerical Results	95
4.6.1	Non-Orthogonal (Concurrent) Source Transmissions	95
4.6.2	Orthogonal Source Transmissions	99
4.6.3	Joint Optimization	101
4.7	Summary	106
5	Joint Source and Relay Precoding Design in Wireless MIMO Relay Networks	109
5.1	Introduction	110
5.2	Joint Optimization Formulation for Source and Relay Precoding	112
5.3	Two-Way Relays with MMSE Receiver	118
5.4	Simulation Results	122
5.4.1	Multiuser MIMO Downlink Communication	124
5.4.2	Two-way Relays with MMSE Receiver	128
5.5	Summary	130
6	Conclusion	133
	Bibliography	135

List of Figures

2.1	Block diagram depicting system, measurement and estimator.	16
2.2	Rayleigh and Rician distribution.	22
2.3	Lognormal distribution.	23
2.4	A schematic diagram of a generic MIMO system.	25
2.5	Convexity of a set and a function.	27
3.1	Mean square error versus sum transmit power in minimizing MSE for scalar parameter.	53
3.2	Mean square error versus logarithm of sum transmit power in minimizing sum power for scalar parameter	54
3.3	(a) MSE estimation performance for proposed multisensor and single sensor. (b) MSE estimation performance for proposed SDP based power allocation and equal power allocation schemes.	55
3.4	Mean square error versus total transmit power while minimizing MSE for vector-valued parameter	56
3.5	Mean square error versus total transmit power while minimizing power for vector-valued parameter.	57
3.6	Mean square error versus total transmit power in nonlinear model of locating a static target.	60
3.7	(a) True trajectory and estimate of the state θ_k under single sensor and multisensor case for a nonlinear dynamic model. (b) Mean square error versus total transmit power for a nonlinear dynamic model in an LSN.	61

3.8	Trajectory of a maneuvering target and distribution of nodes over surveillance region.	62
3.9	Target tracking performance of multisensor and single sensor in terms of estimation error of x-coordinate.	63
3.10	Target tracking performance of multisensor and single sensor in terms of estimation error of y-coordinate.	64
3.11	True trajectory and estimates by single and multisensor network.	65
3.12	Distribution of power among different sensor nodes for the case $N = 10$	66
3.13	MSE estimation performance comparison in an LSN for $N = 30$	68
3.14	MSE estimation performance comparison in an NSN for $N = 12$	69
3.15	MSE estimation performance comparison in an NSN for $N = 20$	70
4.1	A multi-user amplify-and-forward wireless relay network.	76
4.2	A multi-user multi-relay network.	89
4.3	Information throughput versus total relaying power for $M = 3$ and $N = 10$ under non-orthogonal source transmissions.	94
4.4	Information throughput versus total relaying power for $M = 4$ and $N = 12$ under non-orthogonal source transmissions.	95
4.5	Information throughput versus total relaying power for $M = 4$ and $N = 16$ under non-orthogonal source transmissions.	96
4.6	Effect of channel uncertainty on information throughput for $M = 5, N = 20$ under non-orthogonal source transmissions.	99
4.7	Information throughput versus total relaying power for orthogonal source transmissions.	100
4.8	Effect of uplink and downlink channel uncertainties on the information throughput for $M = 5, N = 10$ under orthogonal source transmissions.	101
4.9	Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 3$ and $N = 10$	103
4.10	Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 4$ and $N = 12$	104

4.11	Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 4$ and $N = 16$	105
4.12	Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 5$ and $N = 16$	106
4.13	Distribution of source powers for a specific channel realization in a network with $M = 4$ and $N = 12$	107
5.1	A multi-user amplify-and-forward MIMO wireless relay network for downlink cellular communication.	112
5.2	A two-way MIMO Relay system.	118
5.3	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (4, 4, 4)$ for downlink cellular communication.	122
5.4	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (2, 2, 4)$ for downlink cellular communication.	123
5.5	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (3, 5, 5)$ for downlink cellular communication.	124
5.6	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (3, 6, 6)$ for downlink cellular communication.	125
5.7	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (4, 8, 8)$ for downlink cellular communication.	126
5.8	Signal-to-interference-plus-noise ration (SINR) plotted versus total relaying power for $(M, N, K) = (5, 8, 8)$ for downlink cellular communication.	127
5.9	Sum throughput plotted versus total relaying power for two-way and one-way MIMO relay-assisted communication.	129
5.10	Sum throughput plotted versus total relaying power for two-way and one-way MIMO relay-assisted communication.	130
5.11	Sum of MSE of both source \mathcal{S}_1 and \mathcal{S}_2 plotted versus total relaying power with $(M, N) = (2, 2)$ for two-way MIMO relay-assisted communication.	131
5.12	Bit error rate plotted versus total relaying power with $(M, N) = (4, 4)$ for two-way MIMO relay-assisted communication.	132

List of Tables

3.1	Power allocation α for locating random scalar	58
3.2	Power allocation α to sensors for locating random vector	59
3.3	Average No. of iterations for LSN with $N = 30$	68
3.4	Average No. of iterations for NSN with $N = 20$	69
4.1	Average number of iterations for obtaining solutions under non-orthogonal source transmissions.	97
4.2	Average rank i_{opt} of \mathbf{X}_{opt} by SDP relaxation.	98
4.3	Average number of iterations for obtaining solutions under orthogonal source transmissions.	102
4.4	Average number of iterations for convergence.	108
5.1	Average number of iterations for convergence for $(M, N, K) = (3, 5, 5)$	128

List of Acronyms

Acronyms

SN	Sensor Network
FC	Fusion Center
MMSE	Minimum Mean Square Error
SDP	Semi-definite Programming
DCP	Difference of Convex Programming
UKF	Unscented Kalman Filter
LMI	Linear Matrix Inequality
LFT	Linear Fractional Transformation
SINR	Signal to Interference Plus Noise Ratio
SDR	Semidefinite Relaxation
MIMO	Multiple Input Multiple Output
s.t.	Subject to

Chapter 1

Introduction

This chapter introduces the motivation and scope of this dissertation for exploiting the hidden convexity of resource allocation research problems in wireless networks under various scenarios. The brief survey of existing popular research issues in the context of wireless systems is also presented. Then, the outline of the dissertation is provided and the key contributions are listed. Finally, notations which are used extensively throughout the dissertation are introduced.

1.1 Motivation and Scope

Over the past decade, there has been a significant progress in the design of wireless networks to suit the demands of various civil and military applications [1, 20]. In the presence of fading channels, user mobility, energy/power constraints, and many other factors, the task of optimizing performance of wireless communication systems becomes quite challenging. Optimal allocation of resources to such wireless networks is, indeed, one of the most important issues, especially, when physical constraints of the system are taken into account.

One of the important characteristics of a wireless network is its diversity—different users/nodes at different locations and times that send same copy of a given signal may suffer from different channel conditions, and therefore may have different requirements and capabilities [110]. Fixing and allocating resources without considering such user diversity can result in wastage of system resources, and thus system performance also degrades. Furthermore, taking advantage of the

diversities in wireless networks can significantly improve communication performance. All these factors contribute to the need of careful consideration of resource allocation [38].

The advance of resource allocation has been witnessed with tremendous progress in recent years. As one can imagine, because of the number of degree of freedom on many different parameters, resource allocation in wireless networks is a broad issue that covers a wide range of problems. Therefore, the mathematical optimization tools used to handle these problems also vary a lot. Besides the commonly used convex optimization in communication system design, majority of the resource allocation problems are nonlinear and nonconvex in nature. For instance, problem of maximizing the worst throughput in multi-user communication is highly nonconvex and indefinite, or when it comes to optimal power allocation to a nonlinear sensor network with an LMMSE (linear minimum mean square error) estimator [29], the problem becomes intractable. Moreover, if one takes into account time-varying conditions, then the problem evolves into one of dynamic optimizations. Nevertheless, it is fair to say that there is no single optimization method to solve all resource-allocation problems at once.

The fact that makes resource allocation more challenging is that for various applications, different wireless networks aim at different service goals, and therefore have different design specifications. A given network can be severely energy sensitive and power constrained, whereas the other can be throughput hungry or bandwidth limited. In some situations, a network may have a high degree of dynamic measurements with limited nonlinearity, while in other cases a network has highly nonlinear measurements under a strictly static case. As such, different networks face different resource-allocation problems, different aspects of problems employ different optimization techniques, and joint consideration of different objectives encounter different constrained optimization issues.

The scope of this dissertation is to exploit and utilize the inherent convexity of various power and throughput allocation optimization problem in wireless network framework. We investigate that optimal power allocation to wireless sensor network, in particular, can be cast as semi-definite programming problems, and hence efficiently solved using recently developed interior-point methods. Furthermore, wireless network problems which aim at achieving optimum throughput for multiple users can also be handled by using d.c. programming and penalty function based opti-

mization methods. In particular, the goals of research in this dissertation are

- To study the power allocation and scheduling problems in nonlinear sensor networks to guarantee an improvement in system performance while increasing network lifetime. An analysis based on designing Bayesian filter in conjunction with power allocation is also presented to accommodate system nonlinearity in both static and dynamic cases.
- To investigate binary/integer programming problem related to the optimal design of joint power allocation and sensor assignment which is an NP hard problem. An efficient d.c. (difference of convex functions/sets) programming based iterative methods is presented to account for the hard discrete binary constraint for the sensor selection vector.
- To analyze the nonconvex properties of the maximin SINR (signal-to-interference-plus-noise ratio) optimization for multiuser communication and propose efficient equivalent d.c. form for which sequential convex approximation is developed.
- To investigate that the design problem of joint optimization of source and relay precoding matrices for multi-antenna sources and MIMO relays can be expressed as d.c. functions with the help of exact penalty function method and hence can be solved under various design goals such as sum-rate maximization, minimax mean squared error (MSE).

1.2 Problems in Resource Allocation in Wireless Networks

This section reviews popular conventional design issues and active research problems for wireless networks. In other words, a strong motivation is provided on why such research problems are important enough to be considered by research community.

1.2.1 Distributed Estimation

Distributed estimation is based on measurements from multiple wireless sensors. It is assumed that a group of sensors observe the same quantity in independent additive observation noises with possibly different variances. The observations are transmitted using amplify-and-forward (analog) transmissions over nonideal fading wireless channels from the sensors to a fusion center,

where they are combined to generate an estimate of the observed quantity. Various problems and researchable issues arise in this case out of which a few important ones are outlined below:

Power Allocation

Increasing transmission power is one way to counteract channel impairments and improve the quality of the received signal. However, as the energy resources provided by the sensor networks are extremely limited, a power allocation problem naturally arises [10, 25, 29]. The objective of power allocation is to find an optimum strategy to assign power among different sensors, aiming at minimizing the estimation error under certain transmission power constraints or its converse: satisfying a target distortion performance with a minimum energy consumption.

Bandwidth Allocation

A related problem for both types of WSN (ad hoc, or with a fusion center) is that bandwidth is limited, necessitating the estimator to be formed using quantized versions of the original observations [55]. In this setup, quantization becomes an integral part of the estimation process, since one may think of quantization as a means of constructing binary observations.

Joint Sensor Selection and Power Allocation

Active sensor assignment to enhance the performance of FC (fusion center) is also equally important [49, 115]. Owing to the varying surrounding sensing environment, the networked sensors experience quite different channel conditions, which certainly affect their observability. It is desirable to deactivate sensors of poor observation and assign more power to those with better observation to preserve limited on-board battery power.

1.2.2 Resource Allocation for Cooperative Networks

Cooperative communication have gained attention as an emerging technology for future wireless networks. Cooperative communications efficiently take advantage of the broadcasting nature of wireless networks. The basic idea is that users or nodes in a network share their information

and transmit cooperatively to provide diversity that can significantly improve system performance [26,27].

1.2.3 Throughput Allocation with Antenna Array Beamforming

Advantage of having spatial diversity is best utilized when array or network of single or multiple antenna nodes adjust their beam patterns such that the received desired signal has maximum strength toward the desired direction, while the aggregate interference power is minimized at their output. Antenna-array processing techniques such as beamforming can inadvertently be applied to a multi-user communication scenario where it is desired to allocate balanced, or equal throughput to all the users in network [111, 113]. Thus, multiple co-channel users can be supported in each cell to increase the capacity by exploring the spatial diversity. Two important sub-problems of this category are: beamforming in relay network for throughput maximization in multi-user communication, specifically when exact channel state information is not available; second, joint power control and beamforming design which potentially grants better throughput than separate designs of power allocation and beamforming.

1.3 Dissertation Outline

In this dissertation, we focus on power allocation and beamforming vector assignment in wireless networks to enhance their performances. Particularly, we consider the optimized power allocation and sensor scheduling, beamforming design for multiple relay-assisted communication, and then MIMO precoding designs for multiple-antenna systems. The outline of the dissertation is as follows.

- Chapter 1, the introduction chapter, presents the motivation and scope, the outline and the contributions of the dissertation.
- In Chapter 2, a brief review of important background materials for estimation theory and wireless communication is presented. An overview of convex optimization and its special class, semi-definite programming as well as nonconvex optimization is provided.

- Chapter 3 is concerned with a sensor network, where each sensor is modeled by either a linear or nonlinear sensing system. These sensors team up in observing either static or dynamic random targets and transmit their observations through noisy communication channels to a fusion center (FC) for locating/tracking the targets. Regardless of whether the sensor measurements are linear or nonlinear, the targets are scalar or vectors, static or dynamic, the corresponding optimization problems are shown to be semi-definite programs (SDPs) of tractable optimization and thus are globally and efficiently solved by any existing SDP solver.

Next part of this chapter presents an effective solution for joint optimization of active sensor selection and power allocation for linear and nonlinear sensor networks. In particular, binary (discrete) constraints of the sensor selection problem are represented by a continuous d.c. constraint in tandem with an exact penalty function approach. Thus, the problem is expressed as minimization of a d.c. function subject to convex constraints, for which the proposed d.c procedure is able to locate its global optimal solution within a few iterations. Subsequent simulation results validate the effectiveness of the proposed method and prove that the network capacity can be fully achieved by invoking as few as about half of the available sensor nodes.

- Chapter 4 focuses on beamforming design for multi-user wireless relay networks under the criterion of *maximin* information throughput which is an important but also very hard optimization problem due to its nonconvexity nature. The existing approach to reformulate the design to a matrix rank-one constrained optimization only makes the problem harder. This paper exploits the d.c. structure of the objective function and the convex structure of the constraints in such a global optimization problem to develop efficient iterative algorithms of very low complexity to find the solutions. Both cases of concurrent and orthogonal transmissions from sources to relays are considered. Numerical results demonstrate that the developed algorithms are able to locate the global optimal solutions by a few iterations and they are superior to other methods in both performance and computation complexity.

Next part considers the joint design of source power allocation and relay beamforming for multi-user multi-relay wireless networks. This computationally intractable nonconvex prob-

lem is recast as an equivalent d.c. programming. By exploiting specific structures of the d.c. program, an iterative algorithm with low computational complexity is then developed to obtain the optimal solution. Furthermore, a simplified sub-optimal solution with equally constrained source powers is also suggested, which is efficient in both computation and required communication overhead but still has very good performance.

- Chapter 5 presents a novel joint design of optimal source and relay processing matrices for multi-antenna one/two-way MIMO relay-assisted multi-user communication. In the first part, an optimal relaying strategy that uses multiple input multiple output (MIMO) fixed relays with linear processing to support multiuser transmission in cellular networks is proposed. The fixed relay processes the received signal with linear operations and forwards the processed signal to multiple users creating a multiuser MIMO relay. Objective function of maximizing the minimum SINR is cast as a d.c. function with the assistance of an exact penalty function method.
- Finally, Chapter 6 summarizes the main contributions of this dissertation, and points out future research directions.

1.4 List of Publications

The main contribution of this dissertation is to investigate the convex properties in the resource allocation problems for wireless sensors and relays networks under various system-level constraints. As a result, the algorithms with lower complexity and higher robustness are applied. The allocation strategies for power and/or beamforming with improved performance can be efficiently obtained. Details of the research contributions in each chapter are as follows.

Chapter 3

The main results in this chapter involve the development of power allocation and active sensor assignment problem in nonlinear sensor network. Allocation of power is performed under both static and dynamic frameworks for which effective methods are established to track parameter of interest. Moreover, joint optimization algorithm for active sensor selection and power allocation is considered. Combinatoric nature of the sensor selection problem is shown to have been effectively

handled by the proposed d.c. programming based method in tandem with penalty function. The results have been published in one journal paper and five conference papers while one journal paper has been submitted:

- U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 903-915, 2012.
- U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Joint optimization of active sensor assignment and power allocation in sensor networks," accepted in *IEEE Transactions on Vehicular Technology*, 2013.
- U. Rashid, H. D. Tuan, and H. H. Kha, "Optimized power allocation in nonlinear sensor networks via semidefinite programming," in *IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, 2010, Ottawa.
- U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Optimized power allocation by semidefinite programming and unscented transformation for nonlinear sensor network," in *4th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2010, Gold Coast.
- U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Semi-definite programming for distributed tracking of dynamic objects by nonlinear sensor network," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, Prague.
- U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Multisensor data fusion in nonlinear Bayesian filtering," in *IEEE International Conference on Communications and Electronics (ICCE)*, 2012, Vietnam.
- U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Jointly Optimizing Sensor Selection and Power Control for Nonlinear Sensor Networks," in *12th International Symposium on Communications and Information Technologies (ISCIT)*, 2012, Gold Coast.

The main results in this chapter are related to optimal source power allocation and beamforming vector assignment to a relay network to maximize the minimum SINR in a multi-user communication scheme. An efficient algorithm called DCI is proposed which deals with the nonconvex and indefinite objective function of the said problem to generate optimized complex beamforming vectors which guarantee optimal increase in the system capacity. Later on, a joint design of source power allocation and relay beamforming is presented that has been shown to outperform separate optimization of source power and/or relay beamforming. The results have been published in two conference papers, and one journal paper while one journal paper is under review:

- U. Rashid, H. D. Tuan, and H. H. Nguyen, "Relay beamforming designs in multi-user wireless relay networks based on throughput maximin optimization," to appear in *IEEE Transactions on Communications*, 2013.
- U. Rashid, H. Kha, H.D. Tuan, and H. Nguyen, "Joint Optimization of Source Power Allocation and Relay Beamforming in Multi-User Multi-Relay Wireless Networks," accepted in *IEEE Transactions on Vehicular Technology*, 2013.
- U. Rashid, H. Tuan, and H. Nguyen, "Maximin relay beamforming in multi-user amplify-forward wireless relay networks," in *4th International Conference on Wireless Communications and Networking Conference (WCNC)*, 2012, Paris.
- U. Rashid, H. H. Kha, H.D. Tuan, and H. H. Nguyen, "Joint Design of Source Power Allocation and Relay Beamforming in Multi-User Multi-Relay Wireless Networks," in *6th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2012, Gold Coast.

Chapter 5

The key contribution of this chapter is to propose joint design of multi-antenna source and relay precoding design for MIMO relay-assisted multiuser communication for both one-way and two-way relays. The results are under one journal submission and one conference paper submission:

- U. Rashid, H. D. Kha, H. D. Tuan, and H. H. Nguyen, "Joint Design of Source and Relay

Precoding Matrices in One/Two-Way MIMO Relay Networks,” submitted to *IEEE Transactions on Communications*, 2013.

- U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, ”Joint Source and Relay Precoding Design in Wireless MIMO Relay Networks,” submitted to *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, Vancouver.

1.5 Notations

The notations of the dissertation are rather standard. In particular, our notations and conventions are as follows.

- *Matrices and column vectors are denoted by uppercase and lowercase characters, respectively.*
- *Notation $\mathbf{A} \geq 0$ means \mathbf{A} is a (Hermitian) positive semi-definite matrix.*
- *$\mathbf{0}_N$ and \mathbf{I}_N are zero and identity matrices of dimension $N \times N$, respectively.*
- *We denote $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}\mathbf{B})$ for matrices \mathbf{A} and \mathbf{B} of appropriate size, but $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}$ (their dot product), so $\|\mathbf{a}\|^2 = \langle \mathbf{a}, \mathbf{a} \rangle$ and $\|\mathbf{a}\|^2 = \langle \bar{\mathbf{a}}, \mathbf{a} \rangle$, where $\bar{\mathbf{a}}$ is the conjugate of \mathbf{a} .*
- *Notation $\lambda_{\max}(\mathbf{X})$ means the maximum eigenvalue of \mathbf{X} while $\rho(\mathbf{X}) := \max_{i=1,2,\dots} |\lambda_i(\mathbf{X})|$ with its eigenvalues $\lambda_i(\mathbf{X})$ is its spectral radius.*
- *$\text{diag}[a_1, a_2, \dots, a_N]$ is a diagonal matrix with ordered diagonal entries a_1, a_2, \dots, a_N , which may be scalars or matrices while $\mathbf{a} \odot \mathbf{b}$ is the element-wise Hadamard product of two vectors \mathbf{a} and \mathbf{b} .*
- *$E[\cdot]$ is the expectation operator. For a random variable (RV) \mathbf{x} , the notation $\bar{\mathbf{x}}$ is referred to its expectation $E[\mathbf{x}]$, while $R_{\mathbf{x}}$ is its auto-covariance $R_{\mathbf{x}} := E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$ and $R_{\mathbf{xy}}$ is its cross-covariance $R_{\mathbf{xy}} := E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^T]$ with another RV \mathbf{y} . $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, R_{\mathbf{x}})$ means \mathbf{x} is Gaussian RV with the moments $\bar{\mathbf{x}}$ and $R_{\mathbf{x}}$.*

Chapter 2

Fundamentals of Statistical Estimation, Wireless Communication and Optimization Theory

In this chapter we shall present a review of the concepts and techniques used throughout the dissertation. Beginning with a brief overview of estimation theory, we will explore the statistical characterization of various parameters under linear as well as nonlinear system models. After this brief introduction to the statistical estimation, we will shift our focus to signal propagation through a fading wireless channel, and then diversity is explored to explain how to mitigate fading and interference in multiuser communication regimes. Finally, we will provide the basic definitions and important concepts of optimization theory that are used extensively in our work. Without going into great detail, we emphasize the significance of these techniques from an algorithmic point of view.

2.1 Statistical Estimation

Estimation is the process of extracting information from data which can be used to infer the desired information and may contain errors. Modern estimation methods use known relationships to compute the desired information from the measurements, taking into account measurement errors,

the effects of disturbances and control actions on the system as well as prior knowledge of the information. Diverse measurements can be blended to produce 'best' estimates, and information which is not available for measurement can be approximated in an optimal fashion.

An optimal filter is basically a computational algorithm that processes measurements to deduce a minimum error estimate of a system by utilizing: knowledge of system and measurement dynamics, assumed statistics of system noises and measurement errors, and initial condition information. Advantage of such an estimator is that it minimizes the estimation error in a well defined statistical sense and that it utilizes all measurement data plus prior knowledge about the system. The potential disadvantages are its sensitivity to erroneous a priori models and statistics, and the inherent computational burden.

When the time at which an estimate is desired coincides with the last measurement point, the problem is referred to as *filtering*; when the time of interest falls within the span of available measurement data, the problem is termed *smoothing*; and when the time of interest occurs after the last available measurement, the problem is called *prediction*. However, the current dissertation considers only estimation under various conditions and scenarios. Furthermore, it should be noted that a linear MMSE (LMMSE) estimator has been employed in all of the estimation problems covered in the thesis. The reason for using the LMMSE estimator in distributed estimation of random parameters is that the input source has always been assumed to be following Gaussian distribution throughout the thesis. The LMMSE estimator is known to be an optimal estimator when the input source has Gaussian distribution. Moreover, with the known statistics of the input source of upto second order, the estimator offers a closed form expression for the estimate which makes it easier and fairly efficient to implement. This is, however, not the case with MAP, ML, etc., estimators.

2.1.1 Estimation Foundation

In what follows [97], for two random variables X, Y with the expectation/means $E\{X\} = \hat{x}, E\{Y\} = \hat{y}$, its covariance is

$$Cov(X, Y) = E\{(X - \hat{x})(Y - \hat{y})^T\}.$$

For two RVs X and $y = f(X)$, the most important question is how to estimate X by an affine

function $A_K Y + b_K$. There may be several estimate criteria (mutual estimation, MMSE) but let's focus on linear MMSE (LMMSE), for which A and b are found from

$$\min_{A_K, b_K} E\{\|X - (A_K Y + b_K)\|^2\} \quad (2.1)$$

Obviously, whenever $\hat{x} = 0$ and $\hat{y} = 0$ it follows that $b_K = 0$ and then whenever X and Y are uncorrelated it must be $A_K = 0$ as $E\{\|X - A_K Y\|^2\} = R_x + A_K R_y A_K^T$, i.e., 0 is the optimal LMMSE of X conditional on Y . Following theorem [4, 42] describes an important result related to the key results of this dissertation

Theorem 2.1.1. *It is true that*

$$(R_{yx}^T R_y^\dagger \hat{x} - R_{yx}^T R_y^\dagger \hat{y}) = \arg \min_{A_K, b_K} E\{\|X - (A_K Y + b_K)\|^2\} \quad (2.2)$$

Here R_x, R_y and R_{yx} are auto-covariance of x, y and their correlation, respectively, and R_y^\dagger is the pseudo-inverse of R_y .

Consequently, the LMMSE estimation of x based on the observation $Y = y$ for any random variables X and Y is

$$\hat{x} + R_{yx}^T R_y^\dagger (y - \hat{y}).$$

Moreover, the exact first-moment model for the relation of x and y is

$$x = R_{yx}^T R_y^\dagger (y - \hat{y}) + \hat{x} + e, \quad (2.3)$$

where the RV error $e = x - \hat{x} - R_{yx}^T R_y^\dagger (y - \hat{y})$ is always uncorrelated with y ($\because R_{ey} = R_{yx}^T - R_{yx}^T R_y^{-1} R_y = 0$) and is zero means with the auto-covariance

$$P_e = R_x - R_{yx}^T R_y^\dagger R_{yx}. \quad (2.4)$$

Thus, in the linear case $y = Ax + w$ with uncorrelated x and noise w ,

$$P_e = (R_x^{-1} + A^T R_w^{-1} A)^{-1}$$

This result bears an immense importance in terms of providing an optimal linear estimator for all types of random variables.

Proof. Note that for the covariance matrix of the augmented vector $[X^T, Y^T]^T$

$$R_{x,y} = \begin{bmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{bmatrix} \geq 0$$

so that any vector orthogonal to columns of R_y (R_x) must be orthogonal to columns of R_{yx} (R_{yx}^T). Let $\mathcal{M}(X)$ denote the space spanned by the columns of X , then we have $\mathcal{M}(R_{yx}) \subseteq \mathcal{M}(R_y)$ ($\mathcal{M}(R_{yx}^T) \subseteq \mathcal{M}(R_x)$) and there is a matrix B such that $R_{yx}^T = BR_y$. Using the property $R_y R_y^\dagger R_y = R_y$, it follows that

$$R_{yx}^T R_y^\dagger R_y = BR_y = R_{yx}^T \quad (2.5)$$

Then

$$\begin{aligned} \text{Cov}(X - \hat{x} - R_{yx}^T R_y^\dagger (Y - \bar{y}), Y - \hat{y}) &= \\ \langle X - \hat{x}, Y - \hat{y} \rangle - R_{yx}^T R_y^\dagger \langle Y - \hat{y}, Y - \hat{y} \rangle &= \\ R_{yx}^T - R_{yx}^T R_y^\dagger R_y &= 0 \end{aligned}$$

In other words, $e = X - \hat{x} - R_{yx}^T R_y^\dagger (Y - \bar{y})$ and $Y - \hat{y}$ are zero-means and uncorrelated. That implies that 0 is the optimal LMMSE of $X - \hat{x} - R_{yx}^T R_y^\dagger (Y - \bar{y})$ conditional on $Y - \hat{y}$. \square

A consequence of the above theorem is the following most fundamental result on Gaussian random variables.

Theorem 2.1.2. *Suppose that Y and X are two Gaussian variables of dimensions r and p with means \hat{y} and \hat{x} , written as $Y \sim \mathcal{N}_r(\hat{y}, R_y)$ and $X \sim \mathcal{N}_p(\hat{x}, R_x)$, respectively. Then the conditional distribution $x|y$ of X given Y is Gaussian, i.e.*

$$X|Y \sim \mathcal{N}_p(\hat{x} + R_{yx}^T R_y^\dagger (y - \hat{y}), R_x - R_{yx}^T R_y^\dagger R_{yx})$$

It is clear that for $Y = f(X)$ for a deterministic and generally nonlinear function f , the central question is the computations of the following statistical parameters:

$$\begin{aligned} \bar{Y} &= \int f(x) p_X(x) dx \\ R_y &= \int (f(X) - E[Y])(f(X) - E[Y])^T p_X(x) dx \\ R_{yx} &= \int (f(X) - E[Y])(X - \bar{X})^T p_X(x) dx \end{aligned}$$

Of course, for $y = Ax$, all of the above parameters can be expressed in terms of $\hat{x} = \int x p_X(x) dx$ and $R_x = \int x x^T p_X(x) dx$:

$$\bar{y} = A\hat{x}, R_y = AR_x A^T, R_{yx} = AR_x$$

which hold true for all kind of distributions for x .

2.1.2 Basics of MMSE

Throughout in this dissertation, we will be interested only in *minimum mean square error* (MMSE) estimates. According to this criteria, given a random variable Y that depends on another random variable X , an estimate \hat{x} should be obtained such that the mean square error given by $E(\|X - \hat{x}\|^2)$ is minimized. Thus, in a more compact form, the MMSE estimate of a random variable X conditioned on a random variable Y is defined as:

$$\hat{x} = \arg \min_z E(\|X - z\|^2 | Y = y)$$

MMSE estimates are important because for Gaussian variables, the estimate is linear in the state variable which allows us to restrict our attention to the linear estimators only [4, 97].

Theorem 2.1.3. $\hat{x} = \int x p_{X|Y}(x|y) dx$, so $E[\|X - \hat{x}\|^2] = E[\|X\|^2 | Y = y] - \|\hat{x}\|^2$. This means that the random variable $X|Y$ is the estimator of X in terms of Y .

Proof. It is based on the least squares

$$\begin{aligned} E[\|X - z\|^2 | Y = y] &= \int \|x - z\|^2 p_{X|Y}(x|y) dx \\ &= \|z - \int x p_{X|Y}(x|y) dx\|^2 + \int \|x\|^2 p_{X|Y}(x|y) dx - \|\int x p_{X|Y}(x|y) dx\|^2 \end{aligned}$$

which attains the minimum at $z = \int x p_{X|Y}(x|y) dx$.

Wiener Filter

The basic MMSE estimator is to estimate x from the observation equation

$$z = Hx + v, z \in \mathbb{R}^k, x \in \mathbb{R}^n, v \in \mathbb{R}^k \quad (2.6)$$

where x and v are independent random processes with $E(x) = \bar{x}$, $\text{Cov}(x) = P_0 > 0$ and $E(v) = 0$, $\text{Cov}(v) = R > 0$. Calculating the covariance of the observation z :

$$\text{Cov}\left(\begin{bmatrix} z \\ x \end{bmatrix}, \begin{bmatrix} z \\ x \end{bmatrix}\right) = \begin{bmatrix} Hp_0H^T + R & HP_0 \\ P_0H^T & P_0 \end{bmatrix}$$

and then according to Theorem 2.1.2 and Matrix Inverse Lemma [78]

$$x|z = \mathcal{N}(\bar{x} + P_0 H^T (H P_0 H^T + R)^{-1} (z - H \bar{x}), (H^T R^{-1} H + P_0^{-1})^{-1}) \quad (2.7a)$$

$$= \mathcal{N}(\bar{x} + (H^T R^{-1} H + P_0^{-1})^{-1} H^T R^{-1} (z - H \bar{x}), (H^T R^{-1} H + P_0^{-1})^{-1}) \quad (2.7b)$$

Hence, the Bayes estimate of x under the observation z (i.e. expectation of $x|z$) is

$$\hat{x}|z = \bar{x} + (P_0^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} (z - H \bar{x}). \quad (2.8)$$

which is the same as its estimator by the conventional Wiener filter $x = \bar{x} + Gz$ with

$$\begin{aligned} G &= \arg \min_G E(|\bar{x} + G(z - H \bar{x}) - x|^2) \\ &= P_0 H^T (H P_0 H^T + R)^{-1} \\ &= (P_0^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \end{aligned} \quad (2.9)$$

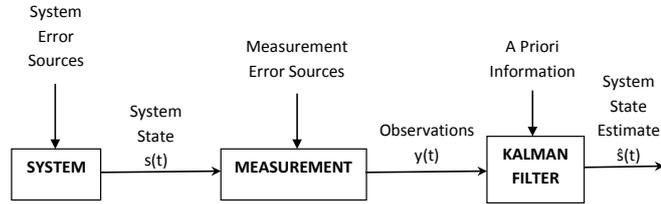


Figure 2.1: Block diagram depicting system, measurement and estimator.

Kalman Filter

Probably the most common optimal filtering technique is that developed by Kalman for estimating the state of a linear time-varying system shown in Figure 2.1. Furthermore, for a problem of estimating a state x by using the state and measurement equation at time k [42]

$$x_{k+1} = A_k x_k + B_k u_k, \quad (2.10)$$

$$z_k = C_k x_k + D_k n_k \quad (2.11)$$

where x_k is the system state variable, z_k is measurement variable, u_k is process noise and n_k is the measurement noise. C_k and A_k denote state and measurement gain matrices, respectively. When

x_k, u_k and n_k are Gaussian, z_k and x_{k+1} must be Gaussian too. Suppose at the initial time $k = 0$, the random variable x_0 has mean $\bar{x}_0 = m_{0| -1}$ and covariance $R_{x,0} = P_{0| -1}$. By (2.10), the random variable z_0 has mean $\eta_0 = C_0 m_{0| -1}$ and covariance $R_{z,0} = C_0 P_{0| -1} C_0^T + D_0 R_0 D_0^T$ with the cross-covariance of x_0 and z_0 given by $R_{zx,0} = C_0 P_{0| -1}$. On arrival of data z_0 , the conditional mean $\mathbb{E}\{x_0|z_0\}$ of the state $x_0|z_0$ is

$$m_0 = m_{0| -1} + K_0(z_0 - \eta_0)$$

and the covariance $R_{zx,0}$ of $x_0|z_0$ is $P_0 = P_{0| -1} - K_0 C_0 P_{0| -1}$, where $K_0 = R_{zx,0}^T R_{z,0}^{-1} = P_{0| -1} C_0^T (C_0 P_{0| -1} C_0^T + D_0 R_0 D_0^T)^{-1}$. Predicted state $x_1|z_0$ at the next time step conditional on data z_0 is the expectation $m_{1|0} = \mathbb{E}\{x_1|z_0\} = A_0 \mathbb{E}\{x_0|z_0\} = A_0 m_0$ and the covariance of $x_1|z_0$ is $P_{1|0} = A_0 P_0 A_0^T + B_0 Q_0 B_0^T$. A similar implementation of Kalman filter for $k \geq 1$ is as follows:

- Suppose the estimate of the state x_{k-1} at time $k-1$ given the history z_{k-1} has mean m_{k-1} and covariance P_{k-1} . Then, the conditional expectation of the predicted state at time k is $m_{k|k-1}$ with covariance $P_{k|k-1}$ where

$$m_{k|k-1} = A_{k-1} m_{k-1} \quad (2.12)$$

$$P_{k|k-1} = A_{k-1} P_{k-1} A_{k-1}^T + B_{k-1} Q_{k-1} B_{k-1}^T \quad (2.13)$$

- Taking $m_{k|k-1}$ as the estimate of the predicted state x_k given z_{k-1} and $P_{k|k-1}$ as its covariance, the conditional expectation of $x|z_{k-1}$ also conditional on the data z_k at time k is m_k with covariance P_k where

$$m_k = m_{k|k-1} + K_k(z_k - \eta_k) \quad (2.14)$$

$$P_k = P_{k|k-1} - K_k C_k P_{k|k-1} \quad (2.15)$$

with

$$\eta_k = C_k m_{k|k-1} \quad (2.16)$$

$$K_k = P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + D_k R_k D_k^T)^{-1}. \quad (2.17)$$

When either X or Y is non-Gaussian, Theorem 1 still provides the optimal linear estimator for X conditional on Y .

MMSE for nonlinear systems

As stated earlier, when observations are related to the state variable via nonlinear functions, the statistical parameters, required to produce estimate, cannot be generated via direct analytical expressions. Below are three methods which can be used under nonlinear functions:

1. *Extended Kalman filter.* The extended Kalman filter (EKF) applies a local linearization to the nonlinear mapping around the state estimate. For any smooth nonlinear function $y = f(x)$, we can This method is predicated on the weak premise that the estimate lies in the neighborhood of the global true value. As a result, stability of the filter and convergence of the estimate are not guaranteed. The EKF employs an approximation which is to linearize the nonlinear function f around \bar{x} where the higher order terms of the Taylor series are truncated under the assumption that the expected value \bar{x} lies in the proximity of the distributed values x . The approximation of f leads to the following:

$$\bar{y} \approx A\bar{x} + f(\bar{x}), \quad R_y \approx AR_xA^T, \quad R_{xy} = R_xA^T. \quad (2.18)$$

2. *Unscented Kalman filter.* The unscented Kalman filter (UKF) [40, 105] applies the unscented transformation, which uses linear regression technique, to approximate the moments of random variables. Thus, UKF aims at the direct approximation of R_{xy}, R_y and \bar{y} so its linearized model is more accurate than that of the EKF. Regression points $x_i, i = 1, \dots, p$ where $p = 2n$ are selected for n -dimensional x around \bar{x} in a manner such that the sample mean and covariance of the points are identical to the mean and covariance of x

$$\begin{aligned} \bar{x} &= \frac{1}{p+1} \sum_{i=0}^p x_i \\ R_x &= \frac{1}{p+1} \sum_{i=0}^p (x_i - \bar{x})(x_i - \bar{x})^T \end{aligned} \quad (2.19)$$

As $R_x > 0$ and thus admits Cholesky decomposition $R_x = \sum_{i=1}^n q_i q_i^T$, an obvious choice of these regression points is

$$\begin{aligned} x_0 &= \bar{x}, & x_i &= x_0 + \sqrt{\frac{p+1}{2}} q_i \\ x_{n+i} &= & x_0 &- \sqrt{\frac{p+1}{2}} q_i \end{aligned}$$

Let $y_i = f(x_i), i = 1, \dots, p$, then the mean and covariance of the random variable y and the cross-covariance of y and x are approximated by the distribution of the regression points x_i and y_i as

$$\bar{y} = \frac{1}{p+1} \sum_{i=0}^p y_i \quad (2.20a)$$

$$R_y = \frac{1}{p+1} \sum_{i=0}^p (y_i - \bar{y})(y_i - \bar{y})^T \quad (2.20b)$$

$$R_{yx} = \frac{1}{p+1} \sum_{i=0}^p (y_i - \bar{y})(x_i - \bar{x})^T \quad (2.20c)$$

One can see that (2.20) are indeed approximations of the continuous distribution $p_x(\cdot)$ to the discrete uniform distribution

$$P(x = x_i) = \frac{1}{p+1}, \quad i = 1, \dots, p$$

i.e., the distribution $p_x(\cdot)$ is statistically linearized around the regression points $x_i, i = 0, 1, \dots, p$ in the UKF.

3. *Linear Fractional Transformation.* The linear fractional transformation (LFT) representation for a nonlinear system comprises of a linear model and a simple nonlinear structure in the feedback loop with sparse representation [6,7,67]. This structure offers two advantages: first, any approximation involved is localized to the feedback loop only. Second, the highly uncorrelated nature of the nonlinear structure gives better approximation of the second-order moments. According to LFT model, any nonlinear mapping f differentiable at any order admits an equivalent representation:

$$\begin{bmatrix} y \\ y_\Delta \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ \omega_\Delta \end{bmatrix} \quad (2.21a)$$

$$\omega_\Delta = \Delta(x)y_\Delta, \quad (2.21b)$$

where $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times n_\Delta}, C \in \mathbb{R}^{n_\Delta \times n}$ and $D \in \mathbb{R}^{n_\Delta \times n_\Delta}$. The auxiliary variables $\omega_\Delta \in \mathbb{R}^{n_\Delta}$ and $y_\Delta \in \mathbb{R}^{n_\Delta}$ are related via the feedback path $\Delta(x)$, which admits a simple structure of the form $\Delta(x) = \sum_{i=1}^n \Delta_i x(i)$. In a compact form, the above model can be written as:

$$y = (A + B\Delta(x)(I - D\Delta(x))^{-1}C)x \quad (2.22)$$

where $\Delta(x)$ enters the relation in a highly nonlinear fashion. Under this representation, an approximation is localized to the feedback path for estimation of the auxiliary random variables ω_Δ . Define the regression points $\omega_{\Delta i} = \Delta(x_i)y_{\Delta i}$, where

$$y_{\Delta i} = Cx_i + D\bar{\omega}_\Delta \quad (2.23)$$

and $\bar{\omega}_\Delta \approx E[\omega_\Delta]$ is

$$\bar{\omega}_\Delta = (I - \bar{\Delta}D)^{-1} \left(\frac{1}{p+1} \sum_{i=0}^p \Delta(x_i)Cx_i \right) \quad (2.24)$$

with

$$\bar{\Delta} = \frac{1}{p+1} \sum_{i=0}^p \Delta(x_i) = \Delta(x_0). \quad (2.25)$$

The covariance of ω_Δ and the cross-covariance with x are computed as

$$R_\Delta = \frac{1}{p+1} \sum_{i=0}^p (\omega_{\Delta i} - \bar{\omega}_\Delta)(\omega_{\Delta i} - \bar{\omega}_\Delta)^T \quad (2.26a)$$

$$R_{\Delta x} = \frac{1}{p+1} \sum_{i=0}^p (\omega_{\Delta i} - \bar{\omega}_\Delta)(x_i - \bar{x})^T \quad (2.26b)$$

Hence, the moments of y are:

$$\bar{y} = A\bar{x} + B\bar{\omega}_\Delta \quad (2.27a)$$

$$R_y = AR_x A^T + BR_\Delta B^T + AR_{\Delta x}^T B^T + BR_{\Delta x} A^T \quad (2.27b)$$

$$R_{yx} = AR_x + BR_{\Delta x} \quad (2.27c)$$

2.2 Wireless Channel

Freedom from wires is an attractive, and often indispensable, feature for many communication applications. Examples of wireless communication include radio and television broadcast, point-to-point microwave links, cellular communications, and wireless local area networks (WLANs). Increasing integration of transceiver functionality using DSP-centric design has driven down implementation costs, and has led to explosive growth in consumer and enterprise applications of wireless, especially cellular telephony and WLANs.

Performance of a wireless communication system is largely dependent on the wireless channel environment. As opposed to the typically static and predictable characteristics of a wired channel,

a wireless channel is rather dynamic which makes an exact analysis of the wireless communication system more difficult. It is, therefore, important to develop understanding of wireless channels to lay the foundation of high performance and bandwidth-efficient wireless transmission technology.

2.2.1 Propagation and Fading

In wireless communication, radio propagation refers to the behaviour of radio waves when they are propagated from transmitter to receiver. In the course of their propagation, these waves are affected by three types of physical phenomena: reflection, diffraction and scattering [80, 90]. *Reflection* occurs when a wave impinges on an object with very large dimension compared to the wavelength, for example, surface of the earth and building. It forces the transmit signal to be reflected back to its origin. *Diffraction* refers to a phenomenon that occurs when the radio path between the transmitter and the receiver is obstructed by a surface with small and sharp irregularities. It appears as bending of waves around the small obstacles and spreading out of waves past small openings. *Scattering* is a phenomenon that forces radio waves to deviate from a straight path by one or more local obstacles.

A unique characteristic in a wireless channel is a phenomenon called 'fading', which is the variation of the signal amplitude over time and frequency. In addition to the additive noise as the most common source of signal degradation, fading is another source of signal degradation that is characterized as a non-additive signal disturbance in the wireless channel. Fading may either result from multipath propagation, referred to as multi-path fading, or from shadowing from obstacles, called shadow fading. The fading phenomenon can be broadly classified into two different types: *large-scale fading* and *small-scale fading*.

Large-scale fading occurs as a receiver moves through a large distance, for example, a distance of the order of cell size [90]. It is thus caused by path loss of signal as a function of distance as well as by shadowing by large objects. Small-scale fading refers to rapid variation of signal levels due to constructive and destructive interference of multiple signal paths when the receiver moves through short distances. Depending on the relative extent of a multipath, frequency selectivity of a channel is characterized (e.g., by frequency-selective or frequency flat) for small-scale fading. On the other hand, depending on the time variation in a channel due to mobile station speed

(characterized by the Doppler spread), short term fading can be classified as either fast fading or slow fading. In summary, small-scale fading is attributed to multi-path propagation, mobile speed, speed of surrounding objects and transmission bandwidth of signal.

Channel Fading Statistical Characterization

As discussed earlier, the rapid variations (fast fading) in signal power caused by local multipaths are represented by Rayleigh distribution. The long-term variations in the mean level are denoted by lognormal distribution. With a LOS propagation path, the Rician distribution is often used for fast fading. Thus, the fading characteristics of a mobile radio signal are described by the following statistical distributions:

1. *Rician Distribution.* When there is a dominant stationary (nonfading) signal component present, such as a LOS propagation path, the small-scale fading envelope distribution is Rician. The Rician distribution has a probability density function (PDF) given by:

$$p(r) = \frac{r}{\sigma^2} e^{-\left(\frac{r^2+A^2}{2\sigma^2}\right)} I_0\left(\frac{Ar}{\sigma^2}\right) \quad \text{for } A \geq 0, \quad r \geq 0 \quad (2.28)$$

where A = peak amplitude of the dominant signal, I_0 = Bessel function of the first kind and order zero, $r^2/2$ = instantaneous power, and σ = standard deviation of the local power. As the dominant path decreases in amplitude, the Rician distribution degenerates to a Rayleigh distribution (see Figure 2.2).

2. *Rayleigh Distribution.* The Rayleigh distribution is used to describe the statistical time-varying nature of the received envelope of a flat fading signal, or the envelope of an individual multipath component. The Rayleigh distribution is given as:

$$p(r) = \frac{r}{\sigma^2} e^{-\left(\frac{r^2}{2\sigma^2}\right)} \quad 0 \leq r \leq \infty \quad (2.29)$$

One can rightfully say that a flat fading signal is exponentially fading in power. This model, which is called Rayleigh fading, is quite reasonable for scattering mechanisms where there are many small reflectors, but is adopted primarily for its simplicity in typical cellular situations with a relatively small number of reflectors.

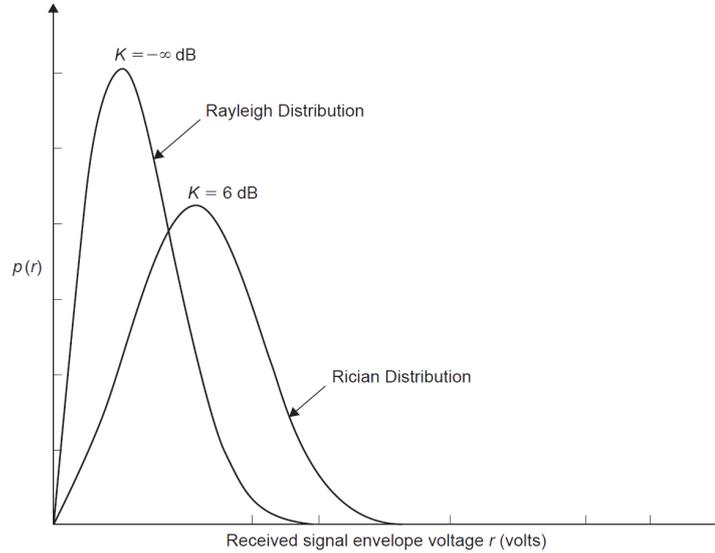


Figure 2.2: Rayleigh and Rician distribution.

3. *Lognormal Distribution.* Lognormal distribution describes the random shadowing effects which occur over a large number of measurement locations which have the same transmitter and receiver separation, but have different levels of clutter on the propagation path. The signal, $s(t)$, typically follows the Rayleigh distribution but its mean square value or its local mean power is lognormal in dBm with variance equal to σ_s^2 . The lognormal distribution is given by (see Figure 2.3):

$$p(S) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\left[\frac{(S-S_m)^2}{2\sigma_s^2}\right]} \quad (2.30)$$

where S_m = mean value of S in dBm, σ_s = standard deviation of S in dBm, $S = 10 \log s$ in dBm, and s = signal power in mW .

2.2.2 Diversity in Wireless Communication

Appropriate resource sharing mechanisms must be put in place if multiple users are to co-exist in a particular frequency band. The wireless channel can be shared among multiple users using several different approaches. One possibility is to eliminate potential interference by assigning different frequency channels to different users; this is termed frequency division multiple access (FDMA). Similarly, we can assign different time slots to different users; this is termed time division multiple

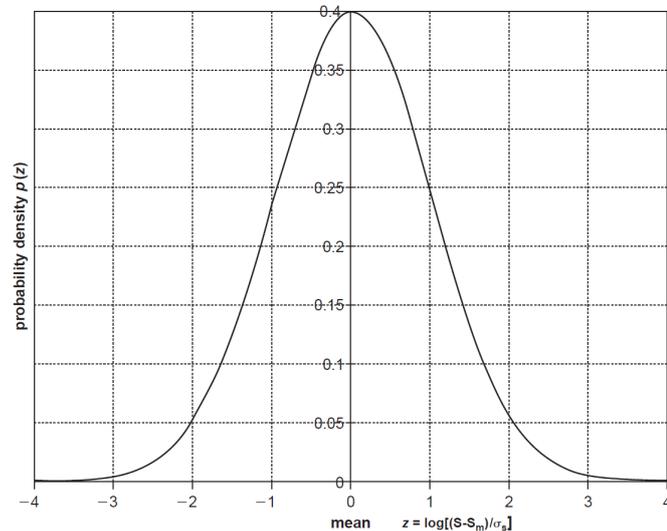


Figure 2.3: Lognormal distribution.

access (TDMA). If we use orthogonal multiple access such as FDMA or TDMA, then we can focus on single-user wireless link design. However, there are also nonorthogonal forms of multiple access, in which different users can signal at the same time over the same frequency band. In this case, the users would be assigned different waveforms, or "codes," which leads to the name code division multiple access (CDMA) for these techniques.

In addition to time and bandwidth, another resource available in wireless systems is space. For example, if one transmitter receiver pair is far enough away from another, then the mutual interference between them is attenuated enough so as to be negligible. Thus, wireless resources can be utilized more efficiently by employing spatial reuse, which forms the basis for cellular communication systems.

MIMO Communications

In recent years, Multiple-Input Multiple-Output (MIMO) systems have emerged as a most promising technology in these measures. MIMO communication systems can be defined intuitively by considering that multiple antennas are used at the transmitting end as well as at the receiving end. The core idea behind MIMO is that signals sampled in the spatial domain at both ends are combined in such a way that they either create effective multiple parallel spatial data pipes (therefore

increasing the data rate), and/or add diversity to improve the quality of the communication.

It is largely due to the seminal works of [31,93], that a proper utilization of the extra dimension of space culminated in the so called MIMO techniques. In our discussion, we will mainly limit ourselves to the fundamentals of these techniques. For an excellent elaboration, the reader is referred to [96]. Let us focus on the narrowband, frequency flat, point-to-point MIMO channel.

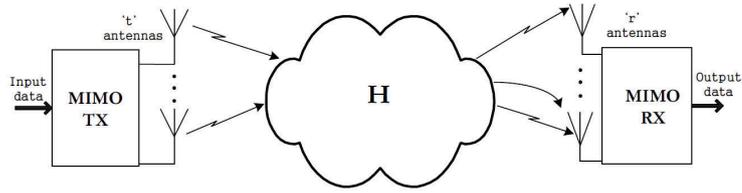


Figure 2.4: A schematic diagram of a generic MIMO system.

Assume that the transmitter is equipped with t and the receiver with r antennas as shown in Figure 2.4. The output at the receiver is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (2.31)$$

where $\mathbf{y} \in \mathbb{C}^r$, $\mathbf{x} \in \mathbb{C}^t$, \mathbf{H} lies in the space of $r \times t$ matrices having complex and possibly random entries i.e., $\mathbf{H} \in \mathbb{C}^{r \times t}$ with h_{ij} denoting the channel between j^{th} transmitter and the i^{th} receiver and \mathbf{n} is a zero mean circularly symmetric complex Gaussian (ZMCSCG) noise vector with independent and identically distributed (i.i.d.) entries. If we assume a noise variance of σ_n^2 at each receive antenna, we have $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_r$. In addition to this, we also suppose that $\mathbb{E}\{|h_{ij}|^2\} = 1 \forall i, j$. With these assumptions we impose a constraint that the total transmit power is $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = P$. The SNR at the i^{th} receiver branch is

$$\frac{\mathbb{E}\{\mathbf{x}^H \mathbf{H}^H(i, :) \mathbf{H}(i, :) \mathbf{x}\}}{\sigma_n^2} = \frac{\mathbb{E}\{\mathbf{x}^H \mathbf{I}_t \mathbf{x}\}}{\sigma_n^2} = \frac{P}{\sigma_n^2}. \quad (2.32)$$

where $\mathbf{H}(i, :)$ denotes the i^{th} row of \mathbf{H} .

2.3 Optimization Theory

In wireless networks, the available radio resources such as power, bandwidth are very limited. Resource allocation and its optimization are general method to improve network performance. Many wireless resource-allocation problems can be formulated as constrained optimization problems, which can be generally written as [15]:

$$\begin{aligned} & \min_{x \in \Omega} f(x), \\ \text{s.t. } & \begin{cases} g_i(x) \leq 0, & i = 1, \dots, m \\ h_j(x) = 0, & j = 1, \dots, l \end{cases} \end{aligned} \quad (2.33)$$

where x is the parameter for optimizing the resource allocation. $\Omega := \{x \in \mathbb{R}^N : g_i(x) \leq 0, h_j(x) = 0\}$ is the feasible range for the parameter, and $f(x)$ is the optimization objective function that represents the performance or cost. Here $g_i(x)$ and $h_j(x)$ are the inequality and equality constraints, respectively. The optimization process finds the solution x^* that satisfies all inequality and equality constraints. For an optimal solution, $f(x^*) \leq f(x), \forall x \in \Omega$. This is also called global optimal solution. On the other hand, a point $x^* \in \Omega$ is a local minimum of $f(x)$ if $f(x^*) \leq f(x)$ when $|x - x^*| < \varepsilon$.

The solvability of the problem (2.33) and the complexity of the algorithm to obtain the solution mainly depend on the properties of the function $f(x)$ and the feasibility set Ω . It is well known that the challenges of solving the optimization problem are related to its convexity, rather than its nonlinearity [83]. The nonconvex optimization problems are in general very difficult to solve, especially when the number of decision variables in x is large. The primary reason is that the problem may have many local optima. Furthermore, it might be very hard to find a feasible point. Other reasons are that the stopping criteria used in general optimization algorithms are often arbitrary, and that optimization algorithms may have very poor convergence rates. However, optimization problems can be efficiently solved if they are convex.

2.3.1 Convex Optimization

If the optimization function, the set of equality and inequality constraints are all linear functions of x , then the problem in (2.33) is called a linear program. While it is fairly easy to obtain a global

optimal point by linear programming, most of the practical problems in resource allocation are nonlinear. In general, there are multiple local optima in a nonlinear program and to find global optima is not an easy task. One special kind of nonlinear programs is the convex optimization problem in which the feasible set Ω is a convex set, and the objective function and constraints are convex/linear functions [83, 102]. A convex set is defined as: A set Ω is convex if for any $x, y \in \Omega$ and any θ with $0 \leq \theta \leq 1$, we have $\theta x + (1 - \theta)y \in \Omega$. Similarly, a convex function is defined as: A function f is convex over x if the feasible set Ω of x is a convex set, and if for all $x, y \in \Omega$ and $0 \leq \theta \leq 1$, we have $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$. Moreover if the function is

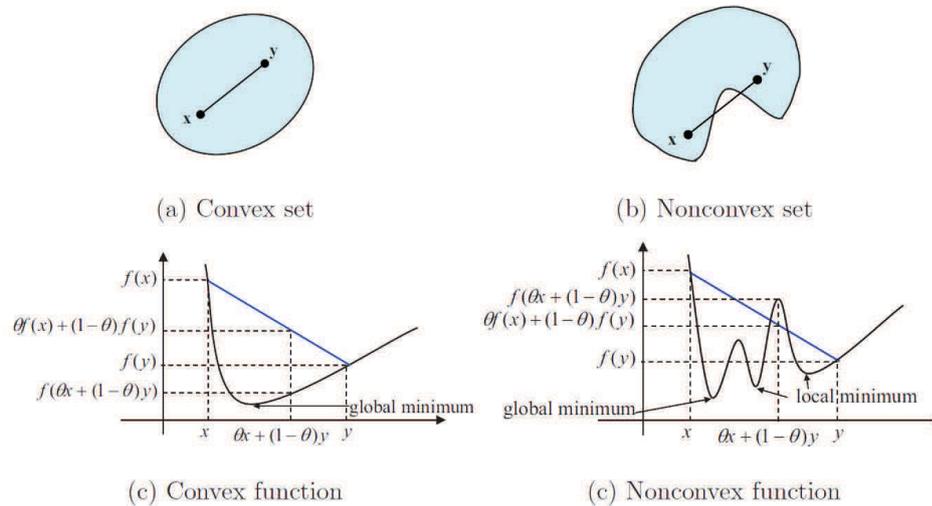


Figure 2.5: Convexity of a set and a function.

differentiable, and if either of the following two conditions hold, the function is a convex function:

$$\text{First-order condition: } f(y) \geq f(x) + \nabla f(x)^T (y - x); \quad (2.34)$$

$$\text{Second-order condition: } \nabla^2 f(x) \succeq 0 \quad (2.35)$$

As compared to general nonconvex optimization problems, the convex optimization problems offer a number of advantages:

- Any local optimum is also a global optimum.
- There are very efficient path-following algorithms with non-heuristic stopping criteria for computing the global minimizer.

- Convex optimization algorithms are also very tractable from theoretical point of view since the worst-case complexity is polynomial .

Although most convex optimization problems do not have analytical solutions, they can be solved numerically very efficiently. Unfortunately, the natural formulation of many engineering problems is nonconvex. Therefore, it is significantly important to cast the nonconvex optimization problems into the convex ones. There are two key techniques to reformulate the problems into the convex forms. The first method is to apply changes of variables. After finding the optimal solution from the convex problem, the original variables of the nonconvex problem can be recovered, and two problems are said to be equivalent. However, in certain problems it is hard to devise a convex problem which is exactly equivalent to the nonconvex one. In this case, the nonconvex problem can be convexified by relaxing constraints. Then, the optimal value of the relaxed problem is a lower bound of that of the original one.

An important class of convex programming is semidefinite programming (SDP) which can be efficiently solved using recently developed interior-point methods [59]. An SDP is an optimization problem of the minimization of a linear objective function subject to the constraint that an affine combination of symmetric matrices is positive semidefinite.

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & F(x) = F_0 + x_1 F_1 + \dots + x_n F_n \succeq 0 \end{aligned} \quad (2.36)$$

where $F_i \in \mathbb{R}^{m \times m}$ are known symmetric matrices and $f(x) \succeq 0$ denotes that $F(x)$ is positive semidefinite at x .

2.3.2 Duality Theory

The concept of duality as applied to optimization is essentially a problem transformation that leads to an indirect but sometimes more efficient solution method. In a duality-based method the original problem, which is referred to as the primal problem, is transformed into a problem in which the parameters are the Lagrange multipliers of the primal. The transformed problem is called the dual problem. In the case where the number of inequality constraints is much greater than the

dimension of x , solving the dual problem to find the Lagrange multipliers and then finding x^* for the primal problem becomes an attractive alternative.

Consider (2.33) as the primal problem. The Lagrange function associated with (2.33) can be defined as

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^l \nu_j h_j(x) \quad (2.37)$$

where λ_i is the Lagrange multiplier associated with the i^{th} inequality constraint $g_i(x) \leq 0$, and ν_j is the Lagrange multiplier associated with the j^{th} equality constraint $h_j(x) = 0$. It is obvious that

$$\sup_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu) = \begin{cases} f(x) & \text{if } x \in \Omega \\ +\infty & \text{otherwise} \end{cases}$$

Hence, the dual of (2.33) is stated as:

$$\min_x \max_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu) \quad (2.38)$$

Then, the dual problem associated with the problem (2.38) is defined as a maximization problem

$$\max_{\lambda \geq 0, \nu} \min_x \mathcal{L}(x, \lambda, \nu) \quad (2.39)$$

Here, the objective function of the dual problem is called the dual function $g(\lambda, \nu)$ defined as $g(\lambda, \nu) = \min_x \mathcal{L}(x, \lambda, \nu)$. It should be noted that the dual problem is a convex optimization problem because it involves the maximization of the concave function over a convex constraint set. It can be proved that

$$\max_{\lambda \geq 0, \nu} \min_x \mathcal{L}(x, \lambda, \nu) \leq \min_x \max_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu)$$

Here, the duality gap, which is defined as $f(x^*) - g(\lambda^*, \nu^*)$, is zero if the primal problem is convex. Therefore, most of robust convex optimization algorithms are based on primal-dual methods which generate a sequence of primal and dual feasible points in iteration to reduce the duality gap to zero for optimality [59].

2.3.3 d.c. Programming

A function is called d.c. (difference of convex functions/sets) if it can be expressed as the difference of two convex functions. Mathematical programs which deal with d.c. functions are called

d.c. programming problems. d.c. programming plays an important role in the field of nonconvex optimization because of its wide range of applications. Theoretically, every continuous function can be represented by a d.c function with any desired precision [45, 102]. Formally, it can be defined as: A real-valued function $f : \Omega \rightarrow \mathbb{R}$ is called d.c. on a convex subset Ω , if there exist two convex functions $f_1, f_2 : \Omega \rightarrow \mathbb{R}$ such that f can be expressed in the form

$$f(x) = f_1(x) - f_2(x) \quad (2.40)$$

We follow an iterative procedure based on expressing f_2 as its convex minorization to obtain an optimal solution for the d.c. programs over set of convex constraints. Suppose that $x^{(\kappa)}$ is a feasible point of (2.40) and $\nabla f_2(x^{(\kappa)})$ is gradient of f_2 at $x^{(\kappa)}$. Then one has

$$f_1(x) - f_2(x) \leq f_1(x) - f_2(x^{(\kappa)}) - \langle \nabla f_2(x^{(\kappa)}), x - x^{(\kappa)} \rangle \quad \forall x.$$

It follows that for any feasible $x^{(\kappa)}$ to (2.40), the following convex program provides a global upper bound minimization for d.c. program (2.40):

$$[f_1(x) - f_2(x) \leq f_1(x) - f_2(x^{(\kappa)}) - \langle \nabla f_2(x^{(\kappa)}), x - x^{(\kappa)} \rangle] : \quad x \in \Omega \quad (2.41)$$

Moreover, for the optimal solution $x^{(\kappa+1)}$ of (2.41), one has

$$\begin{aligned} f_1(x^{(\kappa+1)}) - f_2(x^{(\kappa+1)}) &\leq f_1(x^{(\kappa+1)}) - g(x^{(\kappa)}) - \langle \nabla f_2(x^{(\kappa)}), x^{(\kappa+1)} - x^{(\kappa)} \rangle \\ &\leq f_1(x^{(\kappa)}) - g(x^{(\kappa)}) - \langle \nabla f_2(x^{(\kappa)}), x^{(\kappa)} - x^{(\kappa)} \rangle \\ &\leq f_1(x^{(\kappa)}) - g(x^{(\kappa)}) \end{aligned}$$

which means that $x^{(\kappa+1)}$ is better than $x^{(\kappa)}$, i.e., convex program (2.41) generates a proper solution $x^{(\kappa)}$. Thus, initialized from a feasible $x^{(0)}$, recursively generating $x^{(\kappa)}$ for $\kappa = 0, 1, \dots$, by the optimal solution of convex program (2.41) is a path-following algorithm, which converges to an optimized solution.

2.3.4 Penalty Function Method

An important analytical and algorithmic technique in nonlinear programming involves the use of penalty functions, whereby the equality and inequality constraints are discarded, at the same time some terms are added in the objective function that penalize the violation of the constraints [77]. A

typical example for penalty function for mixed equality and inequality constraint problem (2.33) is given by

$$F_{\mu}(x) = f(x) + \mu \left\| \sum_{i=1}^m g_i(x) + \sum_{j=1}^l h_j(x) \right\|, \quad (2.42)$$

where μ is a positive penalty parameter. If x^* denotes the optimal solution to (2.33) and λ^* is the corresponding optimal Lagrange multiplier vector, then the solutions to (2.33) and (2.42) are equal if $\mu > \|\lambda^*\|$, and (2.42) is called exact penalty function.

Chapter 3

Globally Optimized Power Allocation and Sensor Assignment for Linear and Nonlinear Networks

Sensor networks (SNs) hold key to a wide range of future applications and have the potential to play a significant role in the realm of modern technology. Deployment of these networks ensures controlling instrumentation in industrial automation, sensing data remotely in a data collecting environment and providing surveillance in defense related applications [23]. These networks were initially used in military applications but their potential has recently been exposed in other areas of science and engineering such as process monitoring in industrial plants, navigational and guidance systems, radar tracking, sonar ranging [62], [9], [56], [41].

This chapter is organized as follows. Introduction is presented in Section 3.1. Sections 3.4.1 and 3.3 address distributed Bayes filtering for static and dynamic objects, respectively. Section 3.4 presents joint power allocation and sensor selection problem for nonlinear sensor networks. Section 3.4.1 describes system model and then formulates the joint problem. The proposed method/algorithm is described in detail in Section 3.4.2. Section 3.5 presents simulation results to validate superior performance of the proposed power allocation strategy. This is followed by the concluding remarks.

3.1 Introduction

The present chapter is concerned with a sensor network, where each sensor is modeled by either a linear or nonlinear sensing system. These sensors team up in observing either static or dynamic random targets and transmit their observations through noisy communication channels to a fusion center (FC) for locating/tracking the targets. Physically, the network is limited by energy resource. According to the available sum power budget, we develop a novel technique for power allocation to the sensor nodes that enables the FC produce the best linear estimate in terms of the mean square error (MSE). Regardless of whether the sensor measurements are linear or nonlinear, the targets are scalar or vectors, static or dynamic, the corresponding optimization problems are shown to be semi-definite programs (SDPs) of tractable optimization and thus are globally and efficiently solved by any existing SDP solver. In other words, new tractably computational algorithms of distributed Bayes filtering are derived with full multi-sensor diversity achieved. Intensive simulation shows that these algorithms clearly outperform previously known algorithms.

An SN is said to be linear (LSN) when each of its sensors is modeled by a linear input-output system. On the other hand, the most popular sensors are range and/or bearing sensing (see e.g. [12, 13, 35]), which are nonlinear input-output systems and accordingly an SN of such sensors is called nonlinear (NSN). Typically, sensors are geographically distributed and operating in an amplify-and-forward (AF) mode [32,33]. Through orthogonal noisy wireless communication channels, they send their own local measurements of a target to a central system, called the fusion center (FC). The FC fuses these local measurements to produce a global estimate of the target. Obviously, the sensors consume power during transmission of their observations to the FC, which must be economical because of low battery power of the sensors. Power efficiency is highly crucial for the network lifetime. An optimized power allocation to minimize estimate distortion of scalar parameters has been considered in [25] and [94] for LSNs. Instead of the optimal linear minimum square error estimator (LMMSE), the FC filtering in [25, 94] is the BLUE (Best Linear Unbiased Estimator) for accommodation of tractable convex optimization. Again for LSNs with FC filtering by LMMSE estimator, convex optimization for power allocation has been given in [10] but it is still not attractive enough for computational implementation. An NSN has been touched in [29] but unfortunately its nonlinear sensing modeling does not seem to correspond to any practically

used scenario. Its assumption that sensor output statistics are known at the FC does not look quite conventional as well. It is fair enough to say that decentralized estimation by NSNs for a static target has practically not been considered and tracking a dynamic object by NSNs has been completely open for research.

This chapter is divided in two parts. The first part develops an efficient strategy for allocating power to the sensor nodes in a globally optimized manner. The novelty of our approach is accentuated by a combination of computationally tractable semi-definite programming [14, 60] (for globally optimal power allocation) with unscented transformation [40] and linear fractional transformation (LFT) [36, 67, 99, 101] (for nonlinear statistic function approximation). The major contributions of the paper are:

- Unlike all previous works [10, 25, 29, 94] mainly of LSNs for locating a static target, the globally optimal decentralized Bayes filtering for both LSNs and NSNs are shown computationally tractable in our approach. Further extensive computation show its full diversity as well.
- Unlike previous works [65, 89, 92] which mainly focus on LSNs for tracking a dynamic target, and which, at the end, do not admit computationally tractable and optimal solutions of decentralized estimation, the globally optimal power allocation at each time instant is solved by computationally tractable SDP in both LSNs and NSNs. In other words, the globally optimal distributed Bayes filtering is solved by a sequence of tractably computational SDPs in our approach with multi-sensor diversity shown computationally;
- Our globally optimal power allocation for sensor nodes also leads to an efficient and practical sensor node selection. In fact, our simulation shows that the power allocation is rather concentrated at a few active sensor nodes, which means that only these sensor nodes contribute to the FC filtering performance. Other sensor nodes with zero or almost zero power allocation obviously have no impact in the FC filtering performance and thus should be put to sleep to prolong the network lifetime. In short, the optimal selection of linear or nonlinear sensor nodes in an SN can be effectively solved via our globally optimal power allocation.

The second part of this chapter deals with somewhat related yet equally important problem

of joint power allocation and sensor assignment in a nonlinear sensor network. Active sensor selection for the best performance of FC filtering is another important issue [115]. In general, the problem of optimally selecting a fixed number of sensor nodes among a set is a very difficult combinatoric problem and thus is not tractable at all. Note that the global optimal solution presented in the first half can also provide solution for the sensor selection problem. On the other hand, it will be shown that the proposed algorithm which is developed specifically for the sensor selection and power allocation, can outperform the previous method.

3.2 Global optimized decentralized Bayes filtering

In SN context, localization of a static target is based on the knowledge of its statistics along with the sensor noisy observations. The goal is to produce an estimator that has MMSE under constraints on transmission power at the sensors. Target localization is immensely important to the modern research arena such as target localization for active sonar systems [44], video sensor nodes selection for localization in wireless camera sensor networks [52] and source localization based on range and bearing information [35].

Consider a target $\boldsymbol{\theta} \sim \mathcal{N}(\bar{\boldsymbol{\theta}}, \mathbf{R}_{\boldsymbol{\theta}})$ in N -dimensional space (i.e. rough initial information on $\boldsymbol{\theta}$ expressed by $\bar{\boldsymbol{\theta}}$ and $\mathbf{R}_{\boldsymbol{\theta}}$ is given), which is observed by M spatially distributed sensors. The sensors send their noise corrupted observations to a FC over wireless flat-fading time-orthogonal communication channels [32, 33]. Thus, all these interactions can be compactly modeled by the following behavioral equations

$$\mathbf{y} = g(\boldsymbol{\theta}) + \mathbf{n}, \quad (3.1)$$

$$\mathbf{z} = \mathbf{H}(\boldsymbol{\alpha})\mathbf{y} + \mathbf{w}, \quad (3.2)$$

where $g(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_M(\boldsymbol{\theta}))^T$ with each component $g_i(\boldsymbol{\theta})$ a (linear or nonlinear) deterministic function for expression of i -sensor measuring quantity such as range and/or bearing. Accordingly, $\mathbf{y} = (y_1, y_2, \dots, y_M)^T$ is the sensor observations and $\mathbf{n} \sim \mathcal{N}(0, \mathbf{R}_n)$ with diagonal \mathbf{R}_n is a corrupt noise, which is uncorrelated with the source $\boldsymbol{\theta}$. These observations are relayed to the FC and so $\mathbf{H}(\boldsymbol{\alpha}) \in \mathbb{R}^{M \times M}$ is called the relay matrix defined by

$$\mathbf{H}(\boldsymbol{\alpha}) = \text{diag}[\sqrt{\alpha_i} \sqrt{h_i}]_{i=1,2,\dots,M} \quad (3.3)$$

which includes the channel gains $\sqrt{h_i}$ between i -th sensor node and FC and amplifier coefficients $\sqrt{\alpha_i}$ to control the transmit power of i -th sensor node. $\mathbf{w} \sim \mathcal{N}(0, \mathbf{R}_w)$ with diagonal \mathbf{R}_w is the communication noise. It follows that the power consumed by i -th sensor node is

$$P_i = \mathbf{R}_y(i, i)\alpha_i.$$

Hence, for $\Sigma_\alpha := \text{diag}[\alpha_i]_{i=1,2,\dots,M}$, the sum power consumed by the entire SN is $\sum_{i=1}^M P_i = \text{Trace}(\Sigma_\alpha \mathbf{R}_y)$ is normally constrained by a fixed budget $P_T > 0$

$$\text{Trace}(\Sigma_\alpha \mathbf{R}_y) \leq P_T. \quad (3.4)$$

By [42, Theorem 12.1], the LMMSE estimate for $\boldsymbol{\theta}$ based on FC output \mathbf{z} is

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \bar{\boldsymbol{\theta}} + \mathbf{R}_{z\boldsymbol{\theta}}^T \mathbf{R}_z^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \\ &= \bar{\boldsymbol{\theta}} + \mathbf{R}_{y\boldsymbol{\theta}}^T \mathbf{H}^T(\boldsymbol{\alpha}) (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} (\mathbf{z} - \mathbf{H}(\boldsymbol{\alpha}) \bar{\mathbf{y}}), \end{aligned} \quad (3.5)$$

while the covariance of LMMSE estimator of $\boldsymbol{\theta}$ given \mathbf{z} is

$$\mathbf{R}_\theta - \mathbf{R}_{z\boldsymbol{\theta}}^T \mathbf{R}_z^{-1} \mathbf{R}_{z\boldsymbol{\theta}} = \mathbf{R}_\theta - \mathbf{R}_{y\boldsymbol{\theta}}^T \mathbf{H}^T(\boldsymbol{\alpha}) (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{R}_{y\boldsymbol{\theta}} \mathbf{H}(\boldsymbol{\alpha}). \quad (3.6)$$

This covariance is also the covariance of the estimator error

$$\mathbf{e} := \boldsymbol{\theta} - \hat{\boldsymbol{\theta}},$$

so

$$\begin{aligned} \mathbf{R}_e &= \mathbf{R}_\theta - \mathbf{R}_{z\boldsymbol{\theta}}^T (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{R}_{z\boldsymbol{\theta}} \\ &= \mathbf{R}_\theta - \mathbf{R}_{y\boldsymbol{\theta}}^T \mathbf{H}^T(\boldsymbol{\alpha}) (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_{y\boldsymbol{\theta}}. \end{aligned}$$

Using the Inverse Matrix Lemma [37]

$$\begin{aligned} \mathbf{R}_y - \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y &= (\mathbf{R}_y^{-1} + \mathbf{H}^T(\boldsymbol{\alpha}) \mathbf{R}_w^{-1} \mathbf{H}(\boldsymbol{\alpha}))^{-1} \\ \Leftrightarrow \mathbf{H}^T(\boldsymbol{\alpha}) (\mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\boldsymbol{\alpha}) &= \mathbf{R}_y^{-1} - (\mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) \mathbf{R}_w^{-1} \mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y)^{-1}. \end{aligned}$$

Therefore,

$$\mathbf{R}_e = \mathbf{R}_\theta - \mathbf{R}_{y\boldsymbol{\theta}}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\boldsymbol{\theta}} + \mathbf{R}_{y\boldsymbol{\theta}}^T (\mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T(\boldsymbol{\alpha}) \mathbf{R}_w^{-1} \mathbf{H}(\boldsymbol{\alpha}) \mathbf{R}_y)^{-1} \mathbf{R}_{y\boldsymbol{\theta}} \quad (3.7)$$

We are now in a position to formulate the problem of minimization of MSE, subject to the power budget constraint (3.4) as

$$\min_{\alpha_i \geq 0, i=1,2,\dots,M} \text{Trace}(\mathbf{R}_e) \quad \text{subject to} \quad (3.4), \quad (3.8)$$

which by (3.70) is equivalent to

$$\min_{\alpha_i \geq 0, i=1,2,\dots,M} \text{Trace}(\mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T (\alpha) \mathbf{R}_w^{-1} \mathbf{H} (\alpha) \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta}) \quad \text{subject to} \quad (3.4). \quad (3.9)$$

Note that by Schur's complement [37]

$$\begin{aligned} & \text{Trace}(\mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T (\alpha) \mathbf{R}_w^{-1} \mathbf{H} (\alpha) \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta}) \leq t \\ \Leftrightarrow & \mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T (\alpha) \mathbf{R}_w^{-1} \mathbf{H} (\alpha) \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta} \preceq \mathbf{Z}, \text{Trace}(\mathbf{Z}) \leq t \\ \Leftrightarrow & \begin{bmatrix} \mathbf{Z} & \mathbf{R}_{y\theta}^T \\ \mathbf{R}_{y\theta} & \mathbf{R}_y + \mathbf{R}_y \mathbf{H}^T (\alpha) \mathbf{R}_w^{-1} \mathbf{H} (\alpha) \mathbf{R}_y \end{bmatrix} \succeq 0, \text{Trace}(\mathbf{Z}) \leq t \end{aligned}$$

while $\mathbf{H}^T (\alpha) \mathbf{R}_w^{-1} \mathbf{H} (\alpha) = \Sigma_\alpha \mathbf{R}_w^{-1} \Sigma_h$ (because \mathbf{R}_w^{-1} is diagonal) with $\Sigma_h = \text{diag} \{[h_i]_{i=1,2,\dots,M}\}$.

This leads to the following SDP formulation for (3.9),

$$\min_{t, \mathbf{Z}, \alpha_i \geq 0, i=1,2,\dots,M} t \quad \text{subject to} \quad (3.10)$$

$$(3.4), \text{Trace}(\mathbf{Z}) \leq t, \begin{bmatrix} \mathbf{Z} & \mathbf{R}_{y\theta}^T \\ \mathbf{R}_{y\theta} & \mathbf{R}_y + \mathbf{R}_y \Sigma_\alpha \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y \end{bmatrix} \succeq 0. \quad (3.11)$$

Alternatively, the problem of minimization of the total power consumption under MSE threshold ε is also formulated by the following SDP

$$\min_{\mathbf{Z}, t, \alpha_i \geq 0, i=1,2,\dots,M} \text{Trace}(\Sigma_\alpha \mathbf{R}_y) \quad \text{subject to} \quad (3.12)$$

$$\begin{bmatrix} \mathbf{Q} & \mathbf{R}_{y\theta}^T \\ \mathbf{R}_{y\theta} & \mathbf{R}_y + \mathbf{R}_y \Sigma_\alpha \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y \end{bmatrix} \succeq 0, \text{Trace}(\mathbf{Q}) \leq \varepsilon - \text{Trace}(\mathbf{R}_\theta) + \text{Trace}(\mathbf{R}_{y\theta}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\theta}). \quad (3.13)$$

Both SDPs (3.71)-(3.11) and (3.12)-(3.13) are computationally tractable and can be globally solved by any existing SDP solver such as YALMIP [53], provided that the sensor output covariance \mathbf{R}_y and its cross-covariance $\mathbf{R}_{y\theta}$ with the source θ can be calculated. The remainder of this section is devoted to the computational issue for these covariance matrices to make SDPs (3.71)-(3.11) and (3.12)-(3.13) completely realizable.

3.2.1 Decentralized Bayes filtering for LSN

In LSNs, model (3.1)-(3.2) is completely linear, i.e. the input-output system (3.1) of the sensor measurements is represented by

$$\mathbf{y} = \mathbf{G}\theta + \mathbf{n}, \quad (3.14)$$

where $\mathbf{G} \in \mathbb{R}^{M \times N}$ is a matrix representing effects of path loss, fading and shadowing, which is known to the FC. Therefore their analytical forms are available

$$\begin{aligned}\mathbf{R}_{\theta\mathbf{y}} &= \mathbb{E}[(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\mathbf{y} - \bar{\mathbf{y}})^T] = \mathbf{R}_{\theta}\mathbf{G} \\ \mathbf{R}_{\mathbf{y}} &= \mathbb{E}[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T] = \mathbf{G}\mathbf{R}_{\theta}\mathbf{G}^T + \mathbf{R}_n\end{aligned}\quad (3.15)$$

Theorem 3.2.1. *The optimal decentralized Bayes for locating the target $\boldsymbol{\theta}$ by a LSN modeled by (3.52), (3.2) under the power constraint (3.4) is (3.57) where $\mathbf{R}_{\mathbf{y}\theta}$ and $\mathbf{R}_{\mathbf{y}}$ are defined by (3.15) while α is found from the SDP*

$$\min_{t, \mathbf{Z}, \alpha_i \geq 0, i=1,2,\dots,M} t \quad \text{subject to} \quad (3.15), (3.11). \quad (3.16)$$

3.2.2 Decentralized Bayes Filtering for NSN

Due to nonlinearity of sensing map g in (3.1), analytical expressions of $\mathbf{R}_{\theta\mathbf{y}}$ and $\mathbf{R}_{\mathbf{y}}$ for computational implementation of (3.9) and (3.12) are not expected. However, we now present their efficient and attractive approximations. The first one is the unscented transformation based approximation [40], which works reasonably well for moderately nonlinear maps whereas, the second one is the linear fractional transformation (LFT) based approximation [67], which works reasonably well for higher order nonlinear or fractional maps g . Without loss of generality, assume that $g(\boldsymbol{\theta})$ admits form

$$g(\boldsymbol{\theta}) = \begin{pmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.17)$$

where $g_1(\boldsymbol{\theta})$ is moderately nonlinear in $\boldsymbol{\theta}$ (e.g. a range function) while $g_2(\boldsymbol{\theta})$ is highly nonlinear (e.g. a bearing function) and admits a tractable LFT [36, 67, 99, 101]

$$\begin{bmatrix} g_2(\boldsymbol{\theta}) \\ \mathbf{y}_{\Delta} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{D} \\ \mathbf{G}_{\Delta} & \mathbf{D}_{\Delta} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega}_{\Delta} \end{bmatrix}, \quad \boldsymbol{\omega}_{\Delta} = \Delta(\boldsymbol{\theta})\mathbf{y}_{\Delta}, \quad (3.18)$$

with deterministic matrices $\mathbf{G}, \mathbf{D}, \mathbf{G}_{\Delta}, \mathbf{D}_{\Delta}$ and simple nonlinear feedback $\boldsymbol{\omega}_{\Delta} = \Delta(\boldsymbol{\theta})\mathbf{y}_{\Delta}$ capturing all nonlinearity of g in (3.1) but nevertheless $\Delta(\boldsymbol{\theta})$ a linear map in $\boldsymbol{\theta}$.

As shown in [67], for either higher order nonlinearity or forms involving fractional terms like the above $g_2(\boldsymbol{\theta})$, the above equivalent LFT model (3.18), consisting of a linear model and a simple

nonlinear structure in the virtual feedback, performs better than the conventional global unscented transformations. In the LFT setting, better approximations for moments are obtained by applying unscented transformations in the feedback loop $\Delta(\boldsymbol{\theta})$ only. It is well known from robust control (see e.g. [116]) that any smooth nonlinear map g_2 in (3.17) admits an LFT (3.18). The name of LFT (3.18) comes from the following actual linear fractional form of g_2

$$g_2(\boldsymbol{\theta}) = [\mathbf{G} + \mathbf{D}\Delta(\boldsymbol{\theta})(\mathbf{I} - \mathbf{D}\Delta(\boldsymbol{\theta}))^{-1}\mathbf{G}_\Delta]\boldsymbol{\theta}.$$

Unscented Transformations for Moderately Nonlinear Maps

Unlike linearizing the deterministic map g_1 in (3.17) as done in the Extended Kalman Filter (EKF), the unscented transformation [40] provides a first order approximation for its distribution moments as follows. Since $\mathbf{R}_\theta \succeq 0$, it admits Cholesky decomposition

$$\mathbf{R}_\theta = \sum_{r=1}^N \boldsymbol{\psi}_r \boldsymbol{\psi}_r^T, \boldsymbol{\psi}_r \in \mathbb{R}^N. \quad (3.19)$$

Accordingly, $2N + 1$ regression points $\boldsymbol{\theta}^{(r)}, r = 0, 1, \dots, 2N$ are defined

$$\boldsymbol{\theta}^{(0)} = \bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^{(r)} = \bar{\boldsymbol{\theta}} + \sqrt{\frac{2N+1}{2}} \boldsymbol{\psi}_r, \boldsymbol{\theta}^{(M+r)} = \bar{\boldsymbol{\theta}} - \sqrt{\frac{2N+1}{2}} \boldsymbol{\psi}_r, r = 1, 2, \dots, 2N. \quad (3.20)$$

Clearly,

$$\bar{\boldsymbol{\theta}} = \frac{1}{2N+1} \sum_{r=0}^{2N} \boldsymbol{\theta}^{(r)}, \mathbf{R}_\theta = \frac{1}{2N+1} \sum_{r=0}^{2N} (\boldsymbol{\theta}^{(r)} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(r)} - \bar{\boldsymbol{\theta}})^T,$$

and thereby transform $\mathbf{y}^{(r)} := g_1(\boldsymbol{\theta}^{(r)})$, $r = 0, 1, \dots, 2N$ for approximations

$$\begin{aligned} \bar{g}_1(\boldsymbol{\theta}) &= \frac{1}{2N+1} \sum_{r=0}^{2N} \mathbf{y}^{(r)} \\ \mathbf{R}_{g_1(\boldsymbol{\theta})} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{y}^{(r)} - \bar{g}_1(\boldsymbol{\theta}))(\mathbf{y}^{(r)} - \bar{g}_1(\boldsymbol{\theta}))^T \\ \mathbf{R}_{g_1(\boldsymbol{\theta})\boldsymbol{\theta}} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{y}^{(r)} - \bar{g}_1(\boldsymbol{\theta}))(\boldsymbol{\theta}^{(r)} - \bar{\boldsymbol{\theta}})^T. \end{aligned} \quad (3.21)$$

Linear Fractional Transformations for Highly Nonlinear Maps

For the regression points $\theta^{(r)}, r = 0, 1, \dots, 2N$ defined by (3.20) set

$$\begin{aligned}
\bar{\Delta} &= \frac{1}{2N+1} \sum_{r=0}^{2N} \Delta(\theta^{(r)}) = \Delta(\bar{\theta}), \\
\bar{\omega}_{\Delta} &= (I - \bar{\Delta}D_{\Delta})^{-1} \left(\frac{1}{2N+1} \sum_{r=0}^{2N} \Delta(\theta^{(r)}) G_{\Delta} \theta^{(r)} \right), \\
\mathbf{y}_{\Delta}^{(r)} &= G_{\Delta} \theta^{(r)} + D_{\Delta} \bar{\omega}_{\Delta}, \quad r = 0, 1, \dots, 2N, \\
\omega_{\Delta}^{(r)} &= \Delta(\theta^{(r)}) \mathbf{y}_{\Delta}^{(r)}, \quad r = 0, 1, \dots, 2N, \\
\mathbf{R}_{\Delta} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\omega_{\Delta}^{(r)} - \bar{\omega}_{\Delta})(\omega_{\Delta}^{(r)} - \bar{\omega}_{\Delta})^T, \\
\mathbf{R}_{\Delta\theta} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\omega_{\Delta}^{(r)} - \bar{\omega}_{\Delta})(\theta^{(r)} - \bar{\theta})^T.
\end{aligned} \tag{3.22}$$

Utilize the following alternative approximations to (3.21)

$$\begin{aligned}
\bar{g}_2(\theta) &= G\bar{\theta} + D\bar{\omega}_{\Delta} \\
\mathbf{R}_{g_2(\theta)} &= GR_{\theta}G^T + DR_{\Delta}D^T, \\
\mathbf{R}_{g_2(\theta)\theta} &= GR_{\theta} + DR_{\Delta\theta}.
\end{aligned} \tag{3.23}$$

SDP based decentralized Bayes filtering for NSN

Having the moment approximations (3.21) and (3.22) we can easily approximate the moments of \mathbf{y} in (3.1) as follows

$$\begin{aligned}
\bar{\mathbf{y}} &= \begin{pmatrix} \bar{g}_1(\theta) \\ \bar{g}_2(\theta) \end{pmatrix}, \quad \mathbf{R}_{\mathbf{y}\theta} = \begin{pmatrix} \mathbf{R}_{g_1(\theta)\theta} \\ \mathbf{R}_{g_2(\theta)\theta} \end{pmatrix}, \\
\mathbf{R}_{g_1(\theta)g_2(\theta)} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (g_1(\theta^{(r)}) - \bar{g}_1(\theta))(G\theta^{(r)} + D\omega_{\Delta}^{(r)} - \bar{g}_2(\theta)), \\
\mathbf{R}_{\mathbf{y}} &= \begin{pmatrix} \mathbf{R}_{g_1(\theta)} & \mathbf{R}_{g_1(\theta)g_2(\theta)} \\ \mathbf{R}_{g_1(\theta)g_2(\theta)}^T & \mathbf{R}_{g_2(\theta)} \end{pmatrix} + \mathbf{R}_n.
\end{aligned} \tag{3.24}$$

Theorem 3.2.2. *The optimal decentralized Bayes for locating the target θ by NSN modeled by (3.1), (3.17) under the power constraint (3.4) is (3.57) with $\mathbf{R}_{\mathbf{y}\theta}$ and $\mathbf{R}_{\mathbf{y}}$ defined by (3.24) and α found from the SDP*

$$\min_{t, \mathbf{Z}, \alpha_i \geq 0, i=1,2,\dots,M} t \quad \text{subject to} \quad (3.19), (3.20), (3.21), (3.22), (3.23), (3.24), (3.11). \tag{3.25}$$

□**Remark.** It is obvious that $\mathbf{G}\boldsymbol{\theta}$ in (3.52) is a particular (linear) case of (3.17) with $g(\boldsymbol{\theta}) = g_2(\boldsymbol{\theta})$ and $\mathbf{D} = 0$ in (3.18). Then (3.15) is (3.23) corresponding to $\mathbf{D} = 0$. Thus, Theorem 3.2.1 is a particular case of Theorem 3.2.2 with $\mathbf{D} = 0$.

A more general representation of g than (3.17) is

$$g(\boldsymbol{\theta}) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}), \quad (3.26)$$

where like g_1 and g_2 are at the same structure as in (3.17), i.e. g_1 is moderately nonlinear while g_2 is represented by LFT (3.18). Then, still using approximations (3.21) and (3.23) leads to the following approximation instead of (3.24)

$$\begin{aligned} \bar{\mathbf{y}} &= \bar{g}_1(\boldsymbol{\theta}) + \bar{g}_2(\boldsymbol{\theta}), \quad \mathbf{R}_{\mathbf{y}\boldsymbol{\theta}} = \mathbf{R}_{g_1(\boldsymbol{\theta})\boldsymbol{\theta}} + \mathbf{R}_{g_2(\boldsymbol{\theta})\boldsymbol{\theta}}, \\ \mathbf{R}_{\mathbf{y}} &= \frac{1}{2N+1} \sum_{r=0}^{2N} [(g_1(\boldsymbol{\theta}^{(r)}) - \bar{g}_1(\boldsymbol{\theta})) + (\mathbf{G}\boldsymbol{\theta}^{(r)} + \mathbf{D}\boldsymbol{\omega}_{\Delta}^{(r)} - \bar{g}_2(\boldsymbol{\theta}))] \\ &\quad \times [(g_1(\boldsymbol{\theta}^{(r)}) - \bar{g}_1(\boldsymbol{\theta})) + (\mathbf{G}\boldsymbol{\theta}^{(r)} + \mathbf{D}\boldsymbol{\omega}_{\Delta}^{(r)} - \bar{g}_2(\boldsymbol{\theta}))]^T + \mathbf{R}_n. \end{aligned} \quad (3.27)$$

3.3 Global optimal decentralized Bayes filtering for tracking dynamic objects

This section discusses tracking of a dynamic target by SN under a power constraint. We consider a scenario consisting of a target moving within a surveillance area. The sensor nodes of a SN are assigned to carry out measurements necessary for tracking the object and send their measurements to the FC to output the final estimate of the target's trajectory. Such a process can be modeled by the following input-output system

$$\boldsymbol{\theta}_{k+1} = f_k(\boldsymbol{\theta}_k) + \mathbf{v}_k, \quad (3.28)$$

$$\mathbf{y}_k = g_k(\boldsymbol{\theta}_k) + \mathbf{n}_k, \quad (3.29)$$

$$\mathbf{z}_k = \mathbf{H}(\boldsymbol{\alpha}(k))\mathbf{y}_k + \mathbf{w}_k. \quad (3.30)$$

Here at time instants k , (3.28) is the evolution equation of the target state $\boldsymbol{\theta}_k \in \mathbf{R}^N$ transition, (3.29) is the measurement equation of all sensors and (3.30) is the FC equation. $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{v}_k})$, $\mathbf{n}_k \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{n}_k})$ and $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{w}_k})$ are additive noises. Note the similarity of (3.29)-(3.30) and

(3.1)-(3.2), i.e. the physical meaning of variables $\mathbf{y}_k, \boldsymbol{\theta}_k, \mathbf{z}_k$ in (3.29)-(3.30) is exactly the same as $\mathbf{y}, \boldsymbol{\theta}, \mathbf{z}$ in (3.1)-(3.2), while like $\mathbf{H}(\boldsymbol{\alpha})$ in (3.2) defined by (3.3), the matrix $\mathbf{H}(\boldsymbol{\alpha}(k))$ is defined by

$$\mathbf{H}(\boldsymbol{\alpha}(k)) = \text{diag} \{ \sqrt{\alpha_i(k)} \sqrt{h_i} \}_{i=1,2,\dots,M} \quad (3.31)$$

where $\sqrt{h_i}$ is the channel gain between i -th sensor node and FC, while $\sqrt{\alpha_i(k)}$ is the amplifier coefficient to control the transmit power of i -th sensor node at time instant k , i.e. at each time instant k , the sum power of the SN is constrained by a budget P_T :

$$\text{Trace}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}(k)} \mathbf{R}_{\mathbf{y}_k}) \leq P_T, \quad (3.32)$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}(k)} := \text{diag} \{ \alpha_i(k) \}_{i=1,2,\dots,M}$. On the other hand, (3.29)-(3.30) is also a particular case of (3.1)-(3.2) with $f_k(\boldsymbol{\theta}_k) = \boldsymbol{\theta}_k, \mathbf{v}_k \equiv 0$ and $g_k(\boldsymbol{\theta}_k) \equiv g(\boldsymbol{\theta}_k)$.

From now on, we adopt the following form of $f_k(\boldsymbol{\theta}_k)$ and $g_k(\boldsymbol{\theta}_k)$,

$$f_k(\boldsymbol{\theta}_k) = \begin{pmatrix} f_{1k}(\boldsymbol{\theta}_k) \\ f_{2k}(\boldsymbol{\theta}_k) \end{pmatrix}, \quad g_k(\boldsymbol{\theta}_k) = \begin{pmatrix} g_{1k}(\boldsymbol{\theta}_k) \\ g_{2k}(\boldsymbol{\theta}_k) \end{pmatrix} \quad (3.33)$$

where f_{1k}, g_{1k} are moderately nonlinear maps while f_{2k}, g_{2k} are highly nonlinear map, which are then transformed to an LFT

$$\begin{bmatrix} f_{2k}(\boldsymbol{\theta}_k) \\ g_{2k}(\boldsymbol{\theta}_k) \\ \boldsymbol{\theta}_{\Delta k} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_k & \mathbf{B}_k \\ \mathbf{G}_k & \mathbf{D}_k \\ \mathbf{F}_{\Delta k} & \mathbf{B}_{\Delta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{\omega}_{\Delta k} \end{bmatrix}, \quad \boldsymbol{\omega}_{\Delta k} = \Delta_k(\boldsymbol{\theta}_k) \boldsymbol{\theta}_{\Delta k}, \quad (3.34)$$

with deterministic matrices $\mathbf{F}_k, \mathbf{B}_k, \mathbf{G}_k, \mathbf{D}_k, \mathbf{F}_{\Delta k}, \mathbf{B}_{\Delta k}$ and linear map $\Delta_k(\boldsymbol{\theta}_k)$. We have seen at the end of the previous section that the linear case of f_k or g_k corresponding to (3.33) with $f_k = f_{2k}, \mathbf{B}_k = 0$ or $g_k = g_{2k}, \mathbf{D}_k = 0$ in (3.34), so (3.33)-(3.34) is an universal representation for whatever possible (linear or nonlinear) modeling.

Given the initial information

$$\mathbb{E}[\boldsymbol{\theta}_0] = \bar{\boldsymbol{\theta}}_{0|-1} \quad \text{and} \quad \mathbf{R}_{\boldsymbol{\theta}_0} = \mathbf{R}_{0|-1}, \quad (3.35)$$

the problem at the FC level is to track the state $\boldsymbol{\theta}_k$ based on the instant information \mathbf{z}_k , which is power constrained by (3.32). If one sets $\boldsymbol{\theta}_{0|-1} \sim \mathcal{N}(\bar{\boldsymbol{\theta}}_{0|-1}, \mathbf{R}_{0|-1})$, as an initial estimator of $\boldsymbol{\theta}_0$, then the following recursive Bayes filtering to track the target $\boldsymbol{\theta}_k$ is most natural. At time instant

k , suppose $\theta_{k|k-1}$ is the estimator of θ_k by using $k-1$ past measurements $\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{k-1}$ with the moments $\bar{\theta}_{k|k-1}$ and $\mathbf{R}_{k|k-1} := \mathbf{R}_{\theta_{k|k-1}}$. The FC process two iterations at each time instant k .

- Treat θ_k in (3.29)-(3.30) as $\theta_{k|k-1}$:

$$\mathbf{y}_k = g_k(\theta_{k|k-1}) + \mathbf{n}_k, \quad \mathbf{z}_k = \mathbf{H}(\alpha(k))\mathbf{y}_k + \mathbf{w}_k \quad (3.36)$$

to produce the LMMSE estimator $\theta_{k|k}$ for $\theta_{k|k-1}$ under the power constraint (3.32).

Like (3.71), for $\Sigma_\alpha := \text{diag}\{[\] \alpha_i\}_{i=1,2,\dots,M}$, solve the following SDP

$$\begin{aligned} & \min_{t, \mathbf{Z}, \alpha_i \geq 0, i=1,2,\dots,M} t \quad \text{subject to} \quad (3.37) \\ & \text{Trace}(\Sigma_\alpha \mathbf{R}_{\mathbf{y}_k}) \leq P_T, \quad \text{Trace}(\mathbf{Z}) \leq t, \quad \begin{bmatrix} \mathbf{Z} & \mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}}^T \\ \mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}} & \mathbf{R}_{\mathbf{y}_k} + \mathbf{R}_{\mathbf{y}_k} \Sigma_\alpha \mathbf{R}_{\mathbf{w}_k}^{-1} \Sigma_h \mathbf{R}_{\mathbf{y}_k} \end{bmatrix} \succeq 0 \end{aligned} \quad (3.38)$$

for the optimal solution $\alpha(k)$. Then, like (3.57) and (3.58), the moments of $\theta_{k|k}$ are approximated by

$$\begin{aligned} \bar{\theta}_{k|k} &= \bar{\theta}_{k|k-1} + \mathbf{H}^T(\alpha(k)) \mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}}^T (\mathbf{H}(\alpha(k)) \mathbf{R}_{\mathbf{y}_k} \mathbf{H}^T(\alpha(k)) + \mathbf{R}_{\mathbf{w}_k})^{-1} \\ &\quad \times (\mathbf{z}_k - \mathbf{H}(\alpha(k)) \bar{\mathbf{y}}_k), \end{aligned} \quad (3.39)$$

$$\begin{aligned} \mathbf{R}_{k|k} &:= \mathbf{R}_{\theta_{k|k}} \\ &= \mathbf{R}_{k|k-1} - \mathbf{H}^T(\alpha(k)) \mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}}^T (\mathbf{H}(\alpha(k)) \mathbf{R}_{\mathbf{y}_k} \mathbf{H}^T(\alpha(k)) + \mathbf{R}_{\mathbf{w}_k})^{-1} \\ &\quad \times \mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}} \mathbf{H}(\alpha(k)) \end{aligned} \quad (3.40)$$

which makes use of the following approximations of $\mathbf{R}_{\mathbf{y}_k}$ and $\mathbf{R}_{\mathbf{y}_k \theta_{k|k-1}}$.

- Unscented Transformations. Like (3.21), make Cholesky decomposition $\mathbf{R}_{k|k-1} = \sum_{r=1}^N \psi_r \psi_r^T$ for the definition of $2N+1$ regression points $\theta^{(r)}$, $r=0, 1, \dots, 2N$ by (3.20) (with $\bar{\theta} \rightarrow \bar{\theta}_{k|k-1}$) and then transform $\mathbf{y}_k^{(r)} := g_{1k}(\theta^{(r)})$, $r=0, 1, \dots, 2N$ for approximations,

$$\begin{aligned} \bar{g}_{1k}(\theta_{k|k-1}) &= \frac{1}{2N+1} \sum_{r=0}^{2N} \mathbf{y}_k^{(r)} \\ \mathbf{R}_{g_{1k}(\theta_{k|k-1})} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{y}_k^{(r)} - \bar{\mathbf{y}}_k)(\mathbf{y}_k^{(r)} - \bar{\mathbf{y}}_k)^T \\ \mathbf{R}_{g_{1k}(\theta_{k|k-1}) \theta_{k|k-1}} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\mathbf{y}_k^{(r)} - \bar{\mathbf{y}}_k)(\theta_k^{(r)} - \bar{\theta}_{k|k-1})^T. \end{aligned} \quad (3.41)$$

- Linear fraction transformation. Determine $2N + 1$ corresponding regression points $\boldsymbol{\theta}^{(r)}$, $r = 0, 1, \dots, 2N$ by (3.20) (with $\bar{\boldsymbol{\theta}} \rightarrow \bar{\boldsymbol{\theta}}_{k|k-1}$). Set

$$\begin{aligned}
\bar{\Delta}_k &= \frac{1}{2N+1} \sum_{r=0}^{2N} \Delta_k(\boldsymbol{\theta}^{(r)}), \\
\bar{\boldsymbol{\omega}}_{\Delta k} &= (\mathbf{I} - \bar{\Delta}_k \mathbf{D}_{\Delta k})^{-1} \left(\frac{1}{2N+1} \sum_{r=0}^{2N} \Delta_k(\boldsymbol{\theta}^{(r)}) \mathbf{F}_{\Delta k} \boldsymbol{\theta}^{(r)} \right), \\
\mathbf{y}_{\Delta k}^{(r)} &= \mathbf{F}_{\Delta k} \boldsymbol{\theta}^{(r)} + \mathbf{B}_{\Delta k} \bar{\boldsymbol{\omega}}_{\Delta k}, \quad r = 0, 1, \dots, 2N \\
\boldsymbol{\omega}_{\Delta k}^{(r)} &= \Delta_k(\boldsymbol{\theta}^{(r)}) \mathbf{y}_{\Delta k}^{(r)}, \quad r = 0, 1, \dots, 2N, \\
\mathbf{R}_{\Delta k} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{\boldsymbol{\omega}}_{\Delta k})(\boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{\boldsymbol{\omega}}_{\Delta k})^T, \\
\mathbf{R}_{\Delta k \boldsymbol{\theta}_{k|k-1}} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{\boldsymbol{\omega}}_{\Delta k})(\boldsymbol{\theta}^{(r)} - \bar{\boldsymbol{\theta}}_{k|k-1})^T.
\end{aligned} \tag{3.42}$$

Accordingly,

$$\begin{aligned}
\bar{g}_{2k}(\boldsymbol{\theta}_{k|k-1}) &= \mathbf{G}_k \bar{\boldsymbol{\theta}}_{k|k-1} + \mathbf{D}_k \bar{\boldsymbol{\omega}}_{\Delta k} \\
\mathbf{R}_{g_{2k}(\boldsymbol{\theta}_{k|k-1})} &= \mathbf{G}_k \mathbf{R}_{k|k-1} \mathbf{G}_k^T + \mathbf{D}_k \mathbf{R}_{\Delta k} \mathbf{D}_k^T + \mathbf{R}_{n_k} \\
\mathbf{R}_{g_{2k}(\boldsymbol{\theta}_{k|k-1}) \boldsymbol{\theta}_{k|k-1}} &= \mathbf{G}_k \mathbf{R}_{\boldsymbol{\theta}_{k|k-1}} + \mathbf{D}_k \mathbf{R}_{\Delta k \boldsymbol{\theta}_{k|k-1}}.
\end{aligned} \tag{3.43}$$

- Finalize

$$\begin{aligned}
\mathbf{R}_{g_{1k}(\boldsymbol{\theta}_{k|k-1}) g_{2k}(\boldsymbol{\theta}_{k|k-1})} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (g_{1k}(\boldsymbol{\theta}^{(r)}) - \bar{g}_{1k}(\boldsymbol{\theta}_{k|k-1})) (\mathbf{G}_k \boldsymbol{\theta}^{(r)} + \mathbf{D}_k \boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{g}_{2k}(\boldsymbol{\theta}_{k|k-1}))^T, \\
\bar{\mathbf{y}}_k &= \begin{pmatrix} \bar{g}_{1k}(\boldsymbol{\theta}_{k|k-1}) \\ \bar{g}_{2k}(\boldsymbol{\theta}_{k|k-1}) \end{pmatrix}, \quad \mathbf{R}_{\mathbf{y}_k \boldsymbol{\theta}_{k|k-1}} = \begin{pmatrix} \mathbf{R}_{g_{1k}(\boldsymbol{\theta}_{k|k-1}) \boldsymbol{\theta}_{k|k-1}} \\ \mathbf{R}_{g_{2k}(\boldsymbol{\theta}_{k|k-1}) \boldsymbol{\theta}_{k|k-1}} \end{pmatrix}, \\
\mathbf{R}_{\mathbf{y}_k} &= \begin{pmatrix} \mathbf{R}_{g_{1k}(\boldsymbol{\theta}_{k|k-1})} & \mathbf{R}_{g_{1k}(\boldsymbol{\theta}_{k|k-1}) g_{2k}(\boldsymbol{\theta}_{k|k-1})} \\ \mathbf{R}_{g_{1k}(\boldsymbol{\theta}_{k|k-1}) g_{2k}(\boldsymbol{\theta}_{k|k-1})}^T & \mathbf{R}_{g_{2k}(\boldsymbol{\theta}_{k|k-1})} \end{pmatrix} + \mathbf{R}_{\mathbf{v}_k},
\end{aligned} \tag{3.44}$$

- Treat $\boldsymbol{\theta}_{k+1}$ and $\boldsymbol{\theta}_k$ in (3.28) as $\boldsymbol{\theta}_{k+1|k}$ and $\boldsymbol{\theta}_{k|k}$, respectively

$$\boldsymbol{\theta}_{k+1|k} = f_k(\boldsymbol{\theta}_{k|k}) + \mathbf{v}_k \tag{3.45}$$

and apply [42, Theorem 12.1] to produce the moments $\bar{\boldsymbol{\theta}}_{k+1|k}$ and $\mathbf{R}_{k+1|k} := \mathbf{R}_{\boldsymbol{\theta}_{k+1|k}}$ for the next recursive time $k + 1$,

$$\bar{\boldsymbol{\theta}}_{k+1|k} = \bar{f}_k(\boldsymbol{\theta}_{k|k}), \quad \mathbf{R}_{k+1|k} = \mathbf{R}_{f_k(\boldsymbol{\theta}_{k|k})} + \mathbf{R}_{\mathbf{v}_k}. \tag{3.46}$$

Like (3.19), (3.20), (3.21), (3.22), (3.23), (3.24) making Cholesky factorization $\mathbf{R}_{k|k} = \sum_{r=1}^N \psi_r \psi_r^T$ to define $2N + 1$ regression points $\boldsymbol{\theta}^{(r)}, r = 0, 2, \dots, 2N$ by (3.20) with $\bar{\boldsymbol{\theta}}$ obviously replaced by $\bar{\boldsymbol{\theta}}_{k|k}$, the resultant approximation equation for (3.46) is

$$\begin{aligned} \bar{\boldsymbol{\theta}}_{k+1|k} &= \begin{pmatrix} \bar{f}_{1k}(\boldsymbol{\theta}_{k|k}) \\ \bar{f}_{2k}(\boldsymbol{\theta}_{k|k}) \end{pmatrix}, \\ \mathbf{R}_{k+1|k} &= \begin{pmatrix} \mathbf{R}_{f_{1k}(\boldsymbol{\theta}_{k|k})} & \mathbf{R}_{f_{1k}(\boldsymbol{\theta}_{k|k})f_{2k}(\boldsymbol{\theta}_{k|k})} \\ \mathbf{R}_{f_{1k}(\boldsymbol{\theta}_{k|k})f_{2k}(\boldsymbol{\theta}_{k|k})}^T & \mathbf{R}_{f_{2k}(\boldsymbol{\theta}_{k|k})} \end{pmatrix} + \mathbf{R}_{\mathbf{v}_k}, \end{aligned} \quad (3.47)$$

where

$$\begin{aligned} \bar{f}_{1k}(\boldsymbol{\theta}_{k|k}) &= \frac{1}{2N+1} \sum_{r=0}^{2N} f_{1k}(\boldsymbol{\theta}^{(r)}), \\ \mathbf{R}_{f_{1k}(\boldsymbol{\theta}_{k|k})} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (f_{1k}(\boldsymbol{\theta}^{(r)}) - \bar{f}_{1k}(\boldsymbol{\theta}_{k|k}))(f_{1k}(\boldsymbol{\theta}^{(r)}) - \bar{f}_{1k}(\boldsymbol{\theta}_{k|k}))^T, \end{aligned} \quad (3.48)$$

and

$$\begin{aligned} \bar{f}_{2k}(\boldsymbol{\theta}_{k|k}) &= \mathbf{F}_k \bar{\boldsymbol{\theta}}_{k|k} + \mathbf{B}_k \bar{\boldsymbol{\omega}}_{\Delta k}, \quad \mathbf{R}_{f_{2k}(\boldsymbol{\theta}_{k|k})} = \mathbf{F}_k \mathbf{R}_{k|k} \mathbf{F}_k^T + \mathbf{B}_k \mathbf{R}_{\Delta k} \mathbf{B}_k^T, \\ \mathbf{R}_{f_{1k}(\boldsymbol{\theta}_{k|k})f_{2k}(\boldsymbol{\theta}_{k|k})} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (f_{1k}(\boldsymbol{\theta}^{(r)}) - \bar{f}_{1k}(\boldsymbol{\theta}_{k|k}))(\mathbf{F}_k \boldsymbol{\theta}^{(r)} + \mathbf{B}_k \boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{f}_{2k}(\boldsymbol{\theta}_{k|k}))^T. \end{aligned} \quad (3.49)$$

with

$$\begin{aligned} \bar{\Delta}_k &= \frac{1}{2N+1} \sum_{r=0}^{2N} \Delta_k(\boldsymbol{\theta}^{(r)}), \\ \bar{\boldsymbol{\omega}}_{\Delta k} &= (\mathbf{I} - \bar{\Delta}_k \mathbf{B}_{\Delta k})^{-1} \left(\frac{1}{2N+1} \sum_{r=0}^{2N} \Delta_k(\boldsymbol{\theta}^{(r)}) \mathbf{F}_{\Delta k} \boldsymbol{\theta}^{(r)} \right), \\ \boldsymbol{\theta}_{\Delta k}^{(r)} &= \mathbf{F}_{\Delta k} \boldsymbol{\theta}^{(r)} + \mathbf{B}_{\Delta k} \bar{\boldsymbol{\omega}}_{\Delta k}, \quad r = 0, 1, \dots, 2N, \\ \boldsymbol{\omega}_{\Delta k}^{(r)} &= \Delta_k(\boldsymbol{\theta}^{(r)}) \boldsymbol{\theta}_{\Delta k}^{(r)}, \quad r = 0, 1, \dots, 2N, \\ \mathbf{R}_{\Delta k} &= \frac{1}{2N+1} \sum_{r=0}^{2N} (\boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{\boldsymbol{\omega}}_{\Delta k})(\boldsymbol{\omega}_{\Delta k}^{(r)} - \bar{\boldsymbol{\omega}}_{\Delta k})^T \end{aligned} \quad (3.50)$$

Theorem 3.3.1. *The decentralized Bayes filtering for tracking the dynamic target $\boldsymbol{\theta}_k$ by a NSN modeled by (3.28)-(3.30) with f_k and g_k in form (3.33), (3.34), under the sum power constraint (3.32) and initialized condition (3.35) is the following recursive procedure for $k = 0, 1, 2, \dots$,*

- Solve the SDP (3.37)-(3.38) with $\bar{\mathbf{y}}_k$, $\mathbf{R}_{\mathbf{y}_k}$ and $\mathbf{R}_{\mathbf{y}_k \boldsymbol{\theta}_{k|k-1}}$ approximated by (3.44)
- Execute (3.39)-(3.40);

- Execute (3.47).

□**Remark.** It should be noted that Theorem 3.3.1 in the particular case of $\mathbf{B}_k = 0$ and $\mathbf{D}_k = 0$ while $f_k = f_{1k}$ and $g_k = g_{2k}$, i.e. (3.28)-(3.30) is the following linear system

$$\boldsymbol{\theta}_{k+1} = \mathbf{F}_k \boldsymbol{\theta}_k + \mathbf{v}_k, \mathbf{y}_k = \mathbf{G}_k \boldsymbol{\theta}_k + \mathbf{n}_k, \mathbf{z}_k = \mathbf{H}(\boldsymbol{\alpha}(k)) \mathbf{y}_k + \mathbf{w}_k \quad (3.51)$$

is nothing but the decentralized Kalman filter, which is also a new result.

3.4 Joint Optimization of Active Sensor Assignment and Power Allocation in Sensor Networks

Owing to the varying surrounding sensing environment, the networked sensors experience quite different channel conditions, which certainly affect their observability. It is desirable to deactivate sensors of poor observation and assign more power to those with better observation to preserve limited on-board battery power. However, in contrast to sensor power allocation, which has been shown convex and thus computationally tractable [82], active sensor assignment is a combinatoric (binary) program and thus is \mathcal{NP} -complete. The existing assignment is either through computationally intensive enumeration or relaxation of binary constraints by box constraints, which is far from optimality.

For linear sensor networks, power allocation is considered jointly with sensor assignment in [57] but their applicability is limited to the estimation of the scalar parameter only. In the present paper, we consider joint optimization of active sensor assignment and power allocation to both linear and nonlinear sensor nodes. This problem is recognized as mixed binary program, which is very challenging because of original coupling constraints in continuous power variables and binary assignment variables. Nevertheless, through elegant variable changes, it is firstly transformed to a much more tractable convex program with additional binary constraints. Since these binary constraints are represented by a continuous d.c. (difference of two convex functions/sets) constraint, this program is actually a convex program with an additional reverse constraint [102]. In order to accommodate iterative d.c. programming [98] we follow the exact penalty function approach [70] to equivalently re-express it as minimization of a d.c. function subject to convex

constraints only. Our simulation shows that the global optimal solution of the latter can be located by quite a few convex programs within the proposed d.c. procedure.

The purpose of this section is two-fold: to develop a solution procedure of practical computational complexity for the mixed binary program of joint active sensor assignment and power optimization and to show a striking fact that the capacity of a WSN can be fully achievable by activating only about half of its sensor nodes. Although we consider joint optimization of assignment and power allocation for estimation of a static target only, its extension to tracking a dynamic target is obvious [82].

3.4.1 Problem Formulation

Let $\theta \sim \mathcal{N}(\bar{\theta}, \mathbf{R}_\theta)$ be an M -dimensional parameter that is to be estimated by a network of N sensors. We refer to θ as a parameter, although it is a vector of M parameters. Each i -th sensor can be called for making the following i.i.d. noise corrupted observation of the targeted parameter θ :

$$y_i = \mathbf{f}_i(\theta) + v_i, \quad (3.52)$$

where $\mathbf{f}_i(\theta)$ generally is a nonlinear function of θ and y_i is the corresponding sensor observation corrupted by noise $v_i \sim \mathcal{N}(0, \mathbf{R}_v(i, i))$ for diagonal \mathbf{R}_v . This observation is scaled by an amplification factor $\sqrt{\alpha_i}$ if the i -th sensor is activated for transmission to FC. Therefore, the received signal from the i th sensor at FC takes the form

$$z_i = x_i(\sqrt{\alpha_i h_i} y_i + w_i), \quad (3.53)$$

where $w_i \sim \mathcal{N}(0, \mathbf{R}_w(i, i))$ is the channel noise and $\sqrt{h_i}$ denotes the i th channel power gain. Here x_i is the assignment variable, i.e. $x_i = 1$ means the i -th sensor is active and will transmit its sensing data to FC through the AF protocol. Otherwise, $x_i = 0$ means that the i -th sensor is deactivated in data sensing operation. Define an assignment vector as

$$\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \{0, 1\}^N, \quad (3.54)$$

which is constrained by a given number N_0 of activated sensors

$$\sum_{i=1}^N x_i = N_0. \quad (3.55)$$

Rewrite (3.53) in vector form

$$\mathbf{z} = \mathbf{X}\mathbf{H}(\alpha)\mathbf{y} + \mathbf{X}\mathbf{w}, \quad (3.56)$$

where $\mathbf{X} = \text{diag}[x_1, \dots, x_N]$, $\mathbf{H}(\alpha) := \text{diag}[\sqrt{\alpha_1 h_1}, \dots, \sqrt{\alpha_N h_N}]$.

On the background that auto-covariance matrix \mathbf{R}_y and cross-covariance matrix $\mathbf{R}_{y\theta}$ admit either the analytical form whenever \mathbf{f}_i in (3.52) are linear in θ , or an effective approximation based on linear fraction transformation (LFT) technique when they are nonlinear in θ [82], the Linear Minimum Mean Square Error (LMMSE) estimate of θ based on FC output \mathbf{z} is

$$\begin{aligned} \hat{\theta} &= \bar{\theta} + \mathbf{R}_{z\theta}^T \mathbf{R}_z^\dagger (\mathbf{z} - \bar{\mathbf{z}}) \\ &= \bar{\theta} + \mathbf{R}_{y\theta}^T \mathbf{H}(\alpha) \mathbf{X} (\mathbf{X} \mathbf{H}(\alpha) \mathbf{R}_y \mathbf{H}(\alpha) \mathbf{X} + \mathbf{X} \mathbf{R}_w)^\dagger (\mathbf{z} - \mathbf{X} \mathbf{H}(\alpha) \bar{\mathbf{y}}), \end{aligned} \quad (3.57)$$

while the covariance of LMMSE estimator of θ given \mathbf{z} is

$$\mathbf{R}_e := \mathbf{R}_\theta - \mathbf{R}_{z\theta}^T \mathbf{R}_z^\dagger \mathbf{R}_{z\theta} = \mathbf{R}_\theta - \mathbf{R}_{y\theta}^T \mathbf{H}(\alpha) \mathbf{X} (\mathbf{X} \mathbf{H}(\alpha) \mathbf{R}_y \mathbf{H}(\alpha) \mathbf{X} + \mathbf{X} \mathbf{R}_w)^\dagger \mathbf{X} \mathbf{H}(\alpha) \mathbf{R}_{y\theta}. \quad (3.58)$$

Here and after, by \mathbf{A}^\dagger we denote the pseudo-inverse matrix of a matrix \mathbf{A} .

Thus the joint optimization of active sensor assignment and power allocation under total sensor power constraint P_T can be formulated as

$$\min_{\mathbf{x} \in \{0,1\}^N} \langle \mathbf{R}_e \rangle : (3.55), \sum_{i=1}^N x_i \alpha_i \mathbf{R}_y(i, i) \leq P_T. \quad (3.59)$$

It can be seen from the definition (3.58) that the objective $\langle \mathbf{R}_e \rangle$ is a highly nonlinear function in both binary assignment variable $\mathbf{x} = (x_1, \dots, x_N)^T$ and continuous power variable $\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_N)^T$. Moreover, these variables are coupled in this objective as well as in constraints in (3.59). Consequently, (3.59) looks to be too hard for a mixed nonconvex program. We now equivalently re-express it by more computationally tractable mixed convex program.

For every $\mathbf{x} \in \{0, 1\}^N$ satisfying (3.55) define

$$\mathbf{I}(\mathbf{x}) := \{i_j \in \{1, 2, \dots, N\} : x_{i_j} = 1\}, \quad (3.60)$$

and accordingly

$$\begin{aligned} \mathbf{y}_x &:= (y_{i_1}, \dots, y_{i_{N_0}})^T, \quad \mathbf{z}_x := (z_{i_1}, \dots, z_{i_{N_0}})^T, \quad \mathbf{w}_x := (w_{i_1}, \dots, w_{i_{N_0}})^T, \\ \bar{\mathbf{z}}_x &:= (\bar{z}_{i_1}, \dots, \bar{z}_{i_{N_0}})^T, \quad \bar{\mathbf{H}}(\boldsymbol{\alpha}(\mathbf{x})) = \text{diag}[\sqrt{\alpha_{i_1} h_{i_1}}, \dots, \sqrt{\alpha_{i_{N_0}} h_{i_{N_0}}}], \quad \Sigma_{h(\mathbf{x})} = \text{diag}[h_{i_1}, \dots, h_{i_{N_0}}]. \end{aligned} \quad (3.61)$$

Then (3.56) is rewritten by

$$\mathbf{z}_x = \bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{y}_x + \mathbf{w}_x. \quad (3.62)$$

Accordingly, (3.57) is rewritten by

$$\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}} + \mathbf{R}_{\mathbf{y}_x\boldsymbol{\theta}_x}^T \bar{\mathbf{H}}(\alpha(\mathbf{x})) (\bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\bar{\mathbf{H}}(\alpha(\mathbf{x})) + \mathbf{R}_{\mathbf{w}_x})^{-1} (\mathbf{z}_x - \bar{\mathbf{H}}(\alpha(\mathbf{x}))\bar{\mathbf{y}}_x), \quad (3.63)$$

while (3.58) becomes

$$\mathbf{R}_e = \mathbf{R}_\boldsymbol{\theta} - \mathbf{R}_{\mathbf{y}_x\boldsymbol{\theta}_x}^T \bar{\mathbf{H}}(\alpha(\mathbf{x})) (\bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\bar{\mathbf{H}}(\alpha(\mathbf{x})) + \mathbf{R}_{\mathbf{w}_x})^{-1} \bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\boldsymbol{\theta}. \quad (3.64)$$

The key observation to ravel complexity of (3.63) is the following equality

$$\begin{aligned} \langle \mathbf{R}_{\mathbf{y}_x\boldsymbol{\theta}_x}^T \bar{\mathbf{H}}(\alpha(\mathbf{x})) (\bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\bar{\mathbf{H}}(\alpha(\mathbf{x})) + \mathbf{R}_{\mathbf{w}_x})^{-1} \bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\boldsymbol{\theta} \rangle &= \\ \langle \mathbf{R}_{\mathbf{y}_\boldsymbol{\theta}}^T \mathbf{H}(\alpha)\mathbf{X}(\mathbf{X}\mathbf{H}(\alpha)\mathbf{R}_{\mathbf{y}}\mathbf{H}(\alpha)\mathbf{X} + \mathbf{R}_{\mathbf{w}})^{-1} \mathbf{X}\mathbf{H}(\alpha)\mathbf{R}_{\mathbf{y}}\boldsymbol{\theta} \rangle. \end{aligned} \quad (3.65)$$

Indeed, for each such \mathbf{x} we re-label sensor nodes such that $I(\mathbf{x}) = \{1, 2, \dots, N_0\}$ so we can write

$$\mathbf{y} = (\mathbf{y}_x^T, \mathbf{y}_2^T)^T, \quad \mathbf{w} = (\mathbf{w}_x^T, \mathbf{w}_2^T)^T, \quad \mathbf{X} = \text{diag}[\mathbf{I}_{N_0}, \mathbf{0}_{N-N_0}], \quad \bar{\mathbf{H}}(\alpha(\mathbf{x})) = \text{diag}[\sqrt{\alpha_1 h_1}, \dots, \sqrt{\alpha_{N_0} h_{N_0}}]$$

Hence

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_\boldsymbol{\theta}} &= \begin{bmatrix} \mathbf{R}_{\mathbf{y}_x\boldsymbol{\theta}} \\ \mathbf{R}_{\mathbf{y}_2\boldsymbol{\theta}} \end{bmatrix}, \quad \mathbf{X}\mathbf{H}(\alpha)\mathbf{R}_{\mathbf{y}_\boldsymbol{\theta}} = \begin{bmatrix} \bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x\boldsymbol{\theta}} \\ \mathbf{0}_{(N-N_0)M} \end{bmatrix}, \\ (\mathbf{X}\mathbf{H}(\alpha)\mathbf{R}_{\mathbf{y}}\mathbf{H}(\alpha)\mathbf{X} + \mathbf{R}_{\mathbf{w}})^{-1} &= \text{diag}[(\bar{\mathbf{H}}(\alpha(\mathbf{x}))\mathbf{R}_{\mathbf{y}_x}\bar{\mathbf{H}}(\alpha(\mathbf{x})) + \mathbf{R}_{\mathbf{w}_x})^{-1}, \mathbf{R}_{\mathbf{w}_2}^{-1}] \end{aligned}$$

which verifies (3.65).

Now, set new variables $\bar{\alpha}_i = \alpha_i x_i$, $i = 1, 2, \dots, N$ which are constrained by

$$0 \leq \bar{\alpha}_i \leq x_i P_T, \quad i = 1, 2, \dots, N. \quad (3.66)$$

For $\mathbf{H}(\bar{\boldsymbol{\alpha}}) := \mathbf{X}\mathbf{H}(\alpha)$, $\Sigma_{\bar{\boldsymbol{\alpha}}} := \text{diag}[\bar{\alpha}_1, \dots, \bar{\alpha}_N]$ and $\Sigma_h := \text{diag}[h_1, \dots, h_N]$, note that

$$\mathbf{H}^2(\bar{\boldsymbol{\alpha}}) = \text{diag}[\bar{\alpha}_1 h_1, \dots, \bar{\alpha}_N h_N] = \Sigma_{\bar{\boldsymbol{\alpha}}}\Sigma_h$$

Accordingly, the sum power constraint in (3.59) is

$$\langle \Sigma_{\bar{\boldsymbol{\alpha}}}\mathbf{R}_{\mathbf{y}} \rangle \leq P_T. \quad (3.67)$$

Thus

$$\langle \mathbf{R}_e \rangle = \langle \mathbf{R}_\theta \rangle - \langle \mathbf{R}_{y\theta}^T \mathbf{H}(\bar{\alpha})(\mathbf{H}(\bar{\alpha})\mathbf{R}_y\mathbf{H}(\bar{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\bar{\alpha})\mathbf{R}_{y\theta} \rangle. \quad (3.68)$$

Using the Inverse Matrix Lemma

$$\begin{aligned} \mathbf{R}_y - \mathbf{R}_y\mathbf{H}(\bar{\alpha})(\mathbf{H}(\bar{\alpha})\mathbf{R}_y\mathbf{H}(\bar{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\bar{\alpha})\mathbf{R}_y &= (\mathbf{R}_y^{-1} + \mathbf{H}(\bar{\alpha})\mathbf{R}_w^{-1}\mathbf{H}(\bar{\alpha}))^{-1} \\ \Leftrightarrow \mathbf{H}(\bar{\alpha})(\mathbf{H}(\bar{\alpha})\mathbf{R}_y\mathbf{H}(\bar{\alpha}) + \mathbf{R}_w)^{-1} \mathbf{H}(\bar{\alpha}) &= \mathbf{R}_y^{-1} - (\mathbf{R}_y + \mathbf{R}_y\mathbf{H}(\bar{\alpha})\mathbf{R}_w^{-1}\mathbf{H}(\bar{\alpha})\mathbf{R}_y)^{-1}. \end{aligned}$$

Therefore, (3.68) becomes

$$\langle \mathbf{R}_e \rangle = \langle \mathbf{R}_\theta \rangle - \langle \mathbf{R}_{y\theta}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\theta} + \mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y\mathbf{H}(\bar{\alpha})\mathbf{R}_w^{-1}\mathbf{H}(\bar{\alpha})\mathbf{R}_y)^{-1} \mathbf{R}_{y\theta} \rangle \quad (3.69)$$

$$\begin{aligned} &= \langle \mathbf{R}_\theta \rangle - \langle \mathbf{R}_{y\theta}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\theta} + \mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \Sigma_{\bar{\alpha}} \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta} \rangle \\ &= \langle \mathbf{R}_\theta \rangle - \langle \mathbf{R}_{y\theta}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\theta} \rangle - \langle \mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \Sigma_{\bar{\alpha}} \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta} \rangle. \end{aligned} \quad (3.70)$$

Since $\langle \mathbf{R}_\theta \rangle$ and $\langle \mathbf{R}_{y\theta}^T \mathbf{R}_y^{-1} \mathbf{R}_{y\theta} \rangle$ are constants, minimizing $\langle \mathbf{R}_e \rangle$ is the same as maximizing $\langle \mathbf{Z} \rangle := \langle \mathbf{R}_{y\theta}^T (\mathbf{R}_y + \mathbf{R}_y \Sigma_{\bar{\alpha}} \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y)^{-1} \mathbf{R}_{y\theta} \rangle$. Using Shur's complement, (3.59) is now equivalently rewritten by

$$\min_{t, \mathbf{x}, \bar{\alpha}, \mathbf{Z}} t : \quad (3.54), (3.55), (3.66), (3.67), \quad (3.71a)$$

$$\langle \mathbf{Z} \rangle \leq t, \quad \begin{bmatrix} \mathbf{Z} & \mathbf{R}_{y\theta}^T \\ \mathbf{R}_{y\theta} & \mathbf{R}_y + \mathbf{R}_y \Sigma_{\bar{\alpha}} \mathbf{R}_w^{-1} \Sigma_h \mathbf{R}_y \end{bmatrix} \geq 0. \quad (3.71b)$$

The objective function in (3.71) is linear while all but the binary constraint (3.54) are convex, so (3.71) is a convex program with binary constraints, which is still nonconvex as a mixed convex program but is already much more computationally tractable than the original mixed nonconvex program (3.59). Interestingly, relaxing binary constraint (3.54) by $\mathbf{x} \in [0, 1]^N$ leads to the following convex program of optimized power allocations (for all sensors active) [82]

$$\min_{t, \bar{\alpha}, \mathbf{Z}} t : (3.67), (3.71b), 0 \leq \bar{\alpha}_i \leq P_T. \quad (3.72)$$

Indeed, after obtaining the optimal solution $\bar{\alpha}^{(0)}$ of convex program (3.72), it is obvious that

$$\mathbf{x}^{(0)} := \bar{\alpha}^{(0)} / P_T \in [0, 1]^N \quad (3.73)$$

and constraint (3.55) in (3.71) is actually always verified.

Since the convex program (3.72) for all sensor power allocation is a relaxation of the mixed convex

program (3.71) for joint sensor assignment and power allocation, the optimal value of the former represents the estimation capacity of the overall WSN and the estimation capacity of the later can be judged by how its optimal value approaches the capacity of the former.

3.4.2 Computational Methodology

This section is devoted to computation for the optimal solution of the mixed convex program (3.71). One can see that binary constraint (3.54) is equivalent to

$$\mathbf{x} \in [0, 1]^N, \quad (3.74a)$$

$$\sum_{i=1}^N x_n - \sum_{i=1}^N x_n^2 \leq 0. \quad (3.74b)$$

Indeed, (3.74a) implies $x_n^2 \leq x_n$, $n = 1, 2, \dots, N$, which together with (3.74b) yield $\sum_{i=1}^N x_n = \sum_{i=1}^N x_n^2$ and then $x_n^2 = x_n$, $n = 1, 2, \dots, N$, which is (3.54).

Since (3.74a) is linear and (3.74b) is reverse convex, (3.71) is a convex program with additional reverse convex constraint [102], which belongs to the realm of d.c. optimization [102]. To avoid the well developed global optimization algorithms [102], which are computationally expensive due to their global search nature, we now amend (3.71) to a program of minimization of a d.c. function subject to convex constraints only. This is another important class in d.c. optimization but can be efficiently solved by iterative d.c. algorithms of local search nature with much less computational complexity but nevertheless being capable of locating the global optimal solutions in many cases of interest [43, 100].

Following our recently developed exact penalty function approach [70], there is $0 < \mu_0 < +\infty$ such that (3.71) is equivalent to the following optimization problem whenever $\mu > \mu_0$

$$\min_{\mathbf{x}, \bar{\alpha}} [t + \mu (\sum_{i=1}^N x_n - \sum_{i=1}^N x_n^2)] : (3.55), (3.66), (3.67), (3.71b), (3.74a). \quad (3.75)$$

Thus, the only nonconvex constraint (3.74) in (3.71) is relegated to the objective function in its equivalent program (3.75). Then (3.75) can be compactly written as

$$\min_{t, \mathbf{x}, \bar{\alpha}} [f_\mu(t, \mathbf{x}) - g_\mu(\mathbf{x})] : (3.55), (3.66), (3.67), (3.71b), (3.74a), \quad (3.76)$$

where both f_μ and g_μ are convex functions defined by

$$f_\mu(\mathbf{x}) := t + \sum_{i=1}^N x_n + (\sum_{i=1}^N x_n)^2, \quad g_\mu(\mathbf{x}) := \sum_{i=1}^N x_n^2 + (\sum_{i=1}^N x_n)^2.$$

The following iterative d.c. procedure [43, 98] is readily applied for solution of d.c. program (3.76):

- Initialize by $(t^{(0)}, \bar{\alpha}^{(0)}, x^{(0)})$, whereas $(t^{(0)}, \bar{\alpha}^{(0)})$ is the optimal solution of relaxed convex program (3.72) (for all sensor power allocation) and $x^{(0)}$ is defined by (3.73).
- For $\kappa = 0, 1, 2, \dots$, iteratively generate feasible solution $(t^{(\kappa+1)}, \bar{\alpha}^{(\kappa+1)}, x^{(\kappa+1)})$ by the optimal solution of the following convex program

$$\min_{t, \mathbf{x}, \bar{\alpha}} [f_{\mu}(t, \mathbf{x}) - (g_{\mu}(x^{(\kappa)}) + \langle \nabla g_{\mu}(x^{(\kappa)}), \mathbf{x} - x^{(\kappa)} \rangle)] : (3.55), (3.66), (3.67), (3.71b), (3.74a), \quad (3.77)$$

$$\text{where } \langle \nabla g_{\mu}(x^{(\kappa)}), \mathbf{x} - x^{(\kappa)} \rangle = 2\mu (\sum_{n=1}^N (x_n^{(\kappa)} + \sum_{n=1}^N x_n^{(\kappa)}) (x_n - x_n^{(\kappa)})).$$

- Given a tolerance $\varepsilon > 0$, stop to output the solution $(\bar{\alpha}^*, x^*) = (\bar{\alpha}^{(\kappa)}, x^{(\kappa)})$ whenever

$$\frac{f_{\mu}(t^{(\kappa+1)}, x^{(\kappa+1)}) - f_{\mu}(t^{(\kappa)}, x^{(\kappa)}) - g_{\mu}(x^{(\kappa+1)}) + g_{\mu}(x^{(\kappa)})}{|f_{\mu}(t^{(\kappa)}, x^{(\kappa)}) - g_{\mu}(x^{(\kappa)})|} < \varepsilon$$

Due to convexity of functions g_{μ} ,

$$g_{\mu}(\mathbf{x}) \geq g_{\mu}(x^{(\kappa)}) + \langle \nabla g_{\mu}(x^{(\kappa)}), \mathbf{x} - x^{(\kappa)} \rangle \quad \forall \mathbf{x},$$

so the objective function of (3.77) is a global upper bound for that of (3.76). Therefore, as shown by [43, 98], the above iterative d.c. procedure generates a sequence $\{(t^{(\kappa)}, \bar{\alpha}^{(\kappa)}, x^{(\kappa)})\}$ of its improved feasible solution:

$$f_{\mu}(t^{(\kappa)}, x^{(\kappa)}) - g_{\mu}(x^{(\kappa)}) < f_{\mu}(t^{(\kappa-1)}, x^{(\kappa-1)}) - g_{\mu}(x^{(\kappa-1)}),$$

which converges to its local optimal solution. With an appropriate initial solution $(t^{(0)}, \bar{\alpha}^{(0)}, x^{(0)})$ as that defined above, this local optimal solution is often the global optimal one. Indeed, for the above specific program (3.76), the following power re-optimization often leads to its global optimal solution.

Recalling that the assignment vector x^* is found through the above iterative d.c. procedure, $I(x^*)$ and then \mathbf{y}_{x^*} , \mathbf{w}_{x^*} and $\Sigma_{h(x^*)}$ are defined according to (3.60) and (3.61). The optimal power allocation for active sensors defined by x^* is provided by the following convex program

$$\min_{t, \bar{\alpha}, \mathbf{Z}} t : \langle \mathbf{Z} \rangle \leq t, \quad \begin{bmatrix} \mathbf{Z} & \mathbf{R}_{\mathbf{y}_{x^*}}^T \theta \\ \mathbf{R}_{\mathbf{y}_{x^*}} \theta & \mathbf{R}_{\mathbf{y}_{x^*}} + \mathbf{R}_{\mathbf{y}_{x^*}} \text{diag}[\bar{\alpha}_1, \dots, \bar{\alpha}_{N_0}] \mathbf{R}_{\mathbf{w}_{x^*}}^{-1} \Sigma_{h(x^*)} \mathbf{R}_{\mathbf{y}_{x^*}} \end{bmatrix} \geq 0 \quad (3.78)$$

The power re-optimization (3.78) has been previously used in [82] but with x^* such that its N_0 nonzero (one) components correspond to the N_0 largest components of the optimal solution $\bar{\alpha}^{(0)}$ of all sensor power program (3.72).

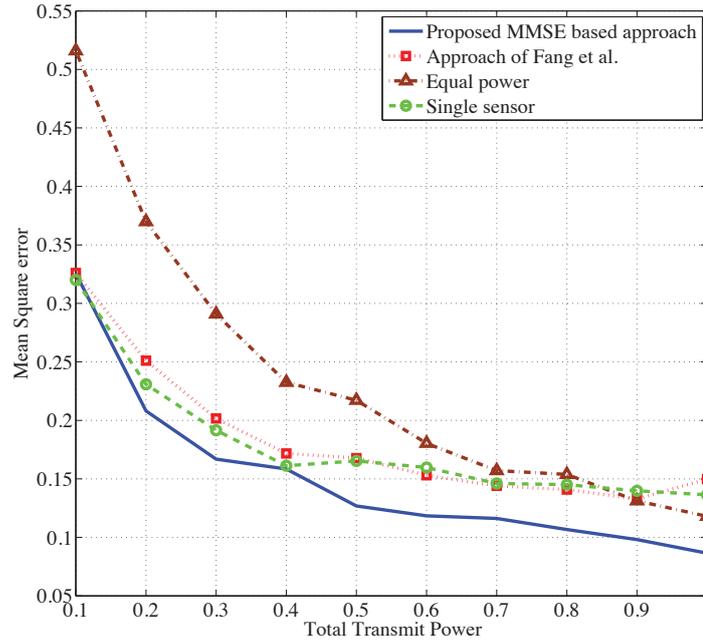


Figure 3.1: Mean square error versus sum transmit power in minimizing MSE for scalar parameter.

3.5 Simulation Results

Effectiveness of the proposed strategies is validated via 10000 Monte Carlo channel realizations through simulation results presented for both static and dynamic targets. There are ten sensors ($M = 10$) and channel gains h_i between these sensors and the FC are generated by normal distribution. It is also assumed that

$$\mathbf{R}_{n_k} \equiv \mathbf{R}_n = \mathbf{R}_{w_k} \equiv \mathbf{R}_w = \text{diag} \{ [] 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 \}. \quad (3.79)$$

Following is the graphical demonstration of how the proposed filtering algorithms discussed in previous sections perform under various scenarios.

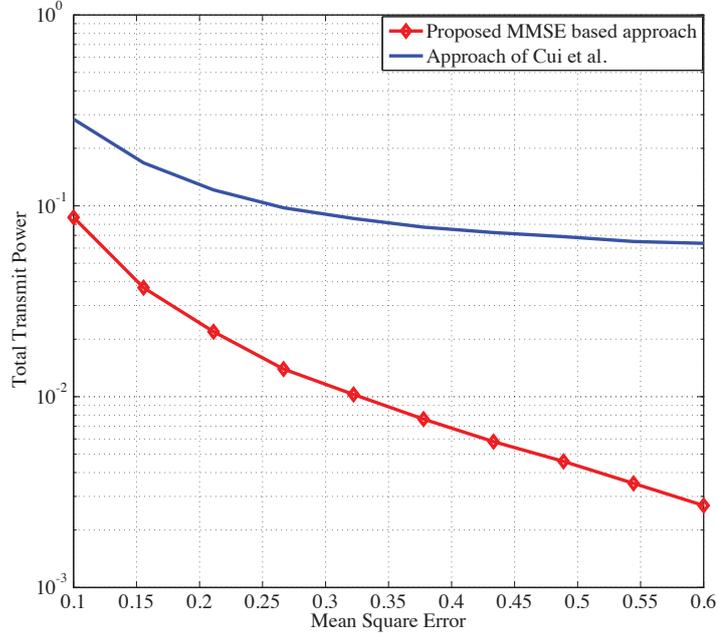


Figure 3.2: Mean square error versus logarithm of sum transmit power in minimizing sum power for scalar parameter

3.5.1 Decentralized Bayes filtering for locating static targets and the optimized sensor selection

Consider (3.52), (3.2) with

$$\mathbf{G} = \text{diag} \{ [] 1.00, 1.11, 1.22, 1.33, 1.44, 1.55, 1.66, 1.77, 1.88, 2.0 \}, \quad (3.80)$$

i.e. the sensors are in different channel conditions. The decentralized Bayes filter for static target based on SDP (3.16), is simulated by applying Theorem 3.2.1. Depending on the target's state dimension, there are two possible cases: Scalar target case consists of sensor nodes transmitting their observations of one dimensional parameter of interest to the FC. Figure 3.1 shows the MSE performances at the FC for four different power allocation schemes. From the analytical view point, the resultant solutions offered by [29] lose their tractability for a relatively general channel environment in which each sensor experiences a different noise variance. It can be observed from Figure 3.1 that the proposed LMMSE based technique (by using Theorem 3.2.1) yields better

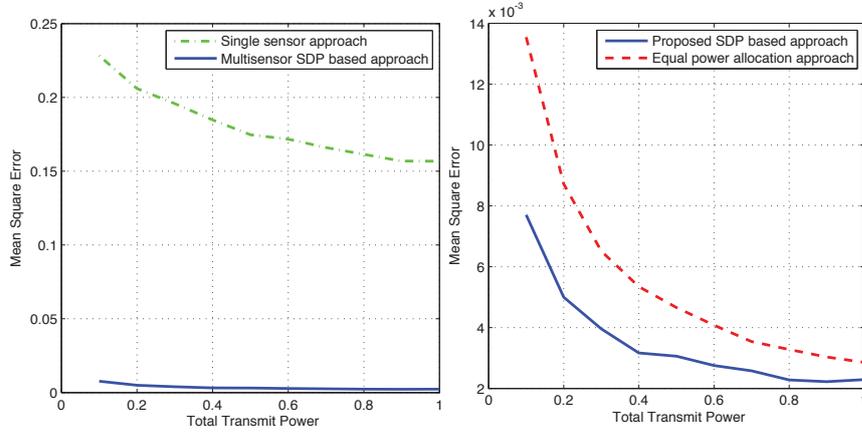


Figure 3.3: (a) MSE estimation performance for proposed multisensor and single sensor. (b) MSE estimation performance for proposed SDP based power allocation and equal power allocation schemes.

results than the single best sensor approach, which assigns all the power to a single sensor of the highest SNR, the sub-optimal solution of [29] and the equal power strategy. Our solution takes full advantage of using a sensor network over a single sensor, while all other solution could not; the suboptimal solution of [29] is actually worse than that by a single sensor and the equal power based solution is much worse than that by a single sensor. Furthermore, Table 3.1 with the optimal power allocations under different sum powers reveals that our globally optimal decentralized Bayes filters assigns the power to only a selective number $k < M$ of sensors (sensors # 4, 2 and 3 for $P_T = 0.8$, $P_T = 0.6$ and $P_T = 0.4$, respectively). These Bayes filters, thus, effectively solves the optimal sensor selection by SDP instead of a hard conventional combinatoric problem of selection. For comparison with the result of [25], consider the problem of minimizing the sum transmit power subject to a threshold of MSE distortion, whose global optimal solution is computed based on (3.12)-(3.13). The corresponding simulation is sketched in Figure 3.2, from which it is clear that for a given distortion, the proposed solution requires a much smaller amount of the sum power than does the BLUE based solution of [25]. This demonstrates that BLUE is far from the optimal LMMSE, despite that its theoretical diversity has been shown.

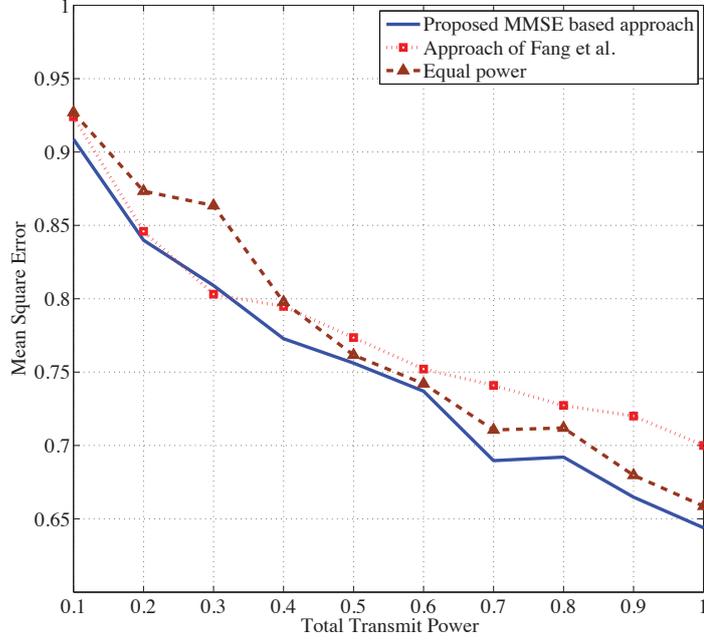


Figure 3.4: Mean square error versus total transmit power while minimizing MSE for vector-valued parameter

Next, we study estimation performance of various power allocation schemes for a random vector θ . We consider localizing a static object in R^3 in which the vector parameter to be estimated consists of three Cartesian coordinates. Observations performed by sensor nodes include measurements of range, elevation angle and azimuth angle, all of which are then fused at the FC to estimate the target's position. A formal representation of such a model is $\mathbf{y} = (g_1(\theta), g_2(\theta), \dots, g_M(\theta))^T + \mathbf{n}$, where

$$g_i(\theta) = \left(\sqrt{(\theta(1) - s_{i,x})^2 + (\theta(2) - s_{i,y})^2 + (\theta(3) - s_{i,z})^2}, \frac{\theta(2) - s_{i,y}}{\theta(1) - s_{i,x}}, \frac{\theta(3) - s_{i,z}}{\sqrt{(\theta(1) - s_{i,x})^2 + (\theta(2) - s_{i,y})^2}} \right)^T$$

Under LSN framework, nonlinear maps $g_i(\theta)$ are linearized at $\bar{\theta}$ to have the linear sensor model (3.52) with $\mathbf{G} = [\mathbf{G}_1^T \ \mathbf{G}_2^T \ \dots \ \mathbf{G}_M^T]^T$ with $\mathbf{G}_i = \partial g_i(\bar{\theta}) / \partial \theta$. From Fig 3.3(a), it can be seen that, by exploiting spatial diversity, LSN provides a far better estimate of the position vector as compared to a single sensor. In Fig 3.3(b), it is demonstrated that our proposed strategy of optimal power allocation gives lower MSE for the same transmit power when compared to the scheme of

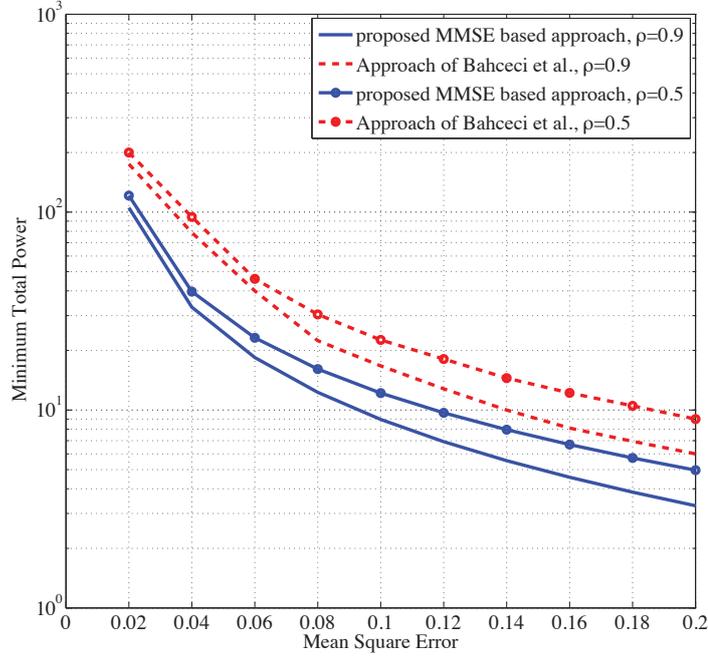


Figure 3.5: Mean square error versus total transmit power while minimizing power for vector-valued parameter.

assigning equal power among nodes.

In another scenario of vector estimation, we consider a model [10, 29] for comparison purpose. According to this model, each sensor in an LSN measures one scalar component of the N -dimensional parameter. The resulting simulation result is given in Figure 3.4. Once again, our proposed approach performs by far the best in terms of the sum power spent for different MSE thresholds. Table 3.2 describes distribution of power to the sensors. As expected, the roles of the sensors are more or less equal so the decentralized Bayes filters pick almost every sensor for activity. Alternatively, for comparison with the result of [11] we consider the sum power minimization subject to different MSE thresholds. Figure 3.5, providing the different performances, clearly shows how our approach using SDP (3.12)-(3.13) outperforms that by [11], under the same simulation conditions as given in [11]: observation noise variance $\sigma_n^2 = 0.01$, channel noise variance $\sigma_w^2 = 1$, with 2 sensor nodes and channel gains $h_1 = 0.8, h_2 = 1$, while

$$[\mathbf{R}_\theta]_{i,j} = \rho^{|j-i|}, \quad |\rho| < 1$$

Table 3.1: Power allocation α for locating random scalar

Sensor	$P_T = 0.8$	$P_T = 0.6$	$P_T = 0.4$
1	0.580	0.583	0.561
2	0.388	0.396	0.000
3	0.000	0.000	0.000
4	0.171	0.000	0.109
5	0.241	0.000	0.000
6	0.000	0.000	0.000
7	0.000	0.000	0.000
8	0.000	0.000	0.083
9	0.000	0.000	0.000
10	0.000	0.000	0.000

with ρ used to express the correlation between the sensor observations.

Simulation environment for NSN is similar to its counterpart in the previous subsection, except that the sensors carry out nonlinear range and bearing measurements here [12, 13, 35]. It is assumed that sensors are randomly distributed over a region where they perform range and bearing measurement of a target lying on the x -axis. Thus the target is located by its x -axis coordinate θ based on range and bearing sensors

$$y_i = \begin{pmatrix} \sqrt{(s_{i,x} - \theta)^2 + s_{i,y}^2} \\ \frac{s_{i,y}}{s_{i,x} - \theta} \end{pmatrix} + n_i,$$

Table 3.2: Power allocation α to sensors for locating random vector

Sensor	$P_T = 8$	$P_T = 6$	$P_T = 4$
1	0.262	0.284	0.242
2	0.240	0.233	0.194
3	0.235	0.217	0.176
4	0.213	0.153	0.160
5	0.185	0.112	0.120
6	0.179	0.152	0.128
7	0.166	0.128	0.121
8	0.093	0.141	0.089
9	0.127	0.112	0.000
10	0.105	0.000	0.000

Here $(s_{i,x}, s_{i,y})$ are (x, y) -coordinates of i -sensor. Thus, the nonlinear sensing map g is in the form of (3.17) with $g_1(\theta) = \sqrt{(s_{i,x} - \theta)^2 + s_{i,y}^2}$ and $g_2(\theta) = \frac{s_{i,y}}{s_{i,x} - \theta}$, which admits an LFT (3.18) with

$$\mathbf{G} = 0_{M \times 4}, \mathbf{D} = \text{diag} \left\{ \left[\frac{1}{s_{i,y}^2} [s_{i,y} \quad -s_{i,x}] \right] \right\}, \mathbf{G}_\Delta = 0_{M \times 4}, \mathbf{D}_\Delta = \text{diag} \left\{ \left[\frac{1}{s_{i,x}} \quad 0 \right] \right\}, \Delta(\theta) = \theta.$$

Accordingly, the approximation (3.24) is used in implementation of SDP (3.25) of Theorem 3.2.2. Figure 3.6 shows the performance comparison between the proposed power allocation by Theorem 3.2.2, the equal power allocation and the strategy of assigning all power to a single sensor. It is notable that the single sensor performance is not even comparable to the proposed and equal power schemes due to lack of spatial diversity, i.e a full diversity is achievable for multi nonlinear sensors. The proposed scheme performs better than the equal power allocation throughout the entire range. One can also note that the difference between these two strategies become larger for smaller values of P_T , which demonstrates that our scheme ensures best use of the resources under severe power budget constraints.

To determine MSE performance for vector target in NSN, we consider an object in a 2D plane, which is located by its (x, y) -axis coordinates $\theta = (\theta(1), \theta(2))^T$.

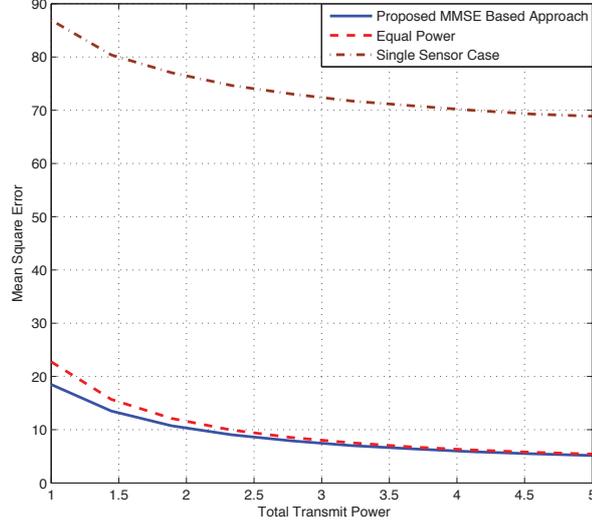


Figure 3.6: Mean square error versus total transmit power in nonlinear model of locating a static target.

A typical sensor's measurements consist of the following ranging and bearing

$$y_i = \begin{pmatrix} \sqrt{(s_{i,x} - \theta(1))^2 + (s_{i,y} - \theta(2))^2} \\ \frac{s_{i,y} - \theta(2)}{s_{i,x} - \theta(1)} \end{pmatrix} + n_i \quad (3.81)$$

where each map $g_i(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}))^T$ in form (3.17) with $g_2(\boldsymbol{\theta}) := (s_{i,y} - \theta(2))/(s_{i,x} - \theta(1))$ admits an LFT (3.18) with

$$\mathbf{G} = \mathbf{0}_{M \times 4}, \mathbf{D} = \text{diag} \left\{ \left[\frac{1}{s_{i,x}^2} [s_{i,y} \quad -s_{i,x}] \right] \right\}, \mathbf{G}_\Delta = \mathbf{0}_{M \times 4}, \mathbf{D}_\Delta = \text{diag} \left\{ \left[\frac{1}{s_{i,x}} \quad 0 \right] \right\}, \Delta(\boldsymbol{\theta}) = \boldsymbol{\theta}.$$

Accordingly, the approximation (3.24) is used in implementation of SDP (3.25) of Theorem 3.2.2. Simulation results in Fig. 3.6 indicate that the SDP based approach outperforms the single sensor approach, and shows better results than the equal power based scheme.

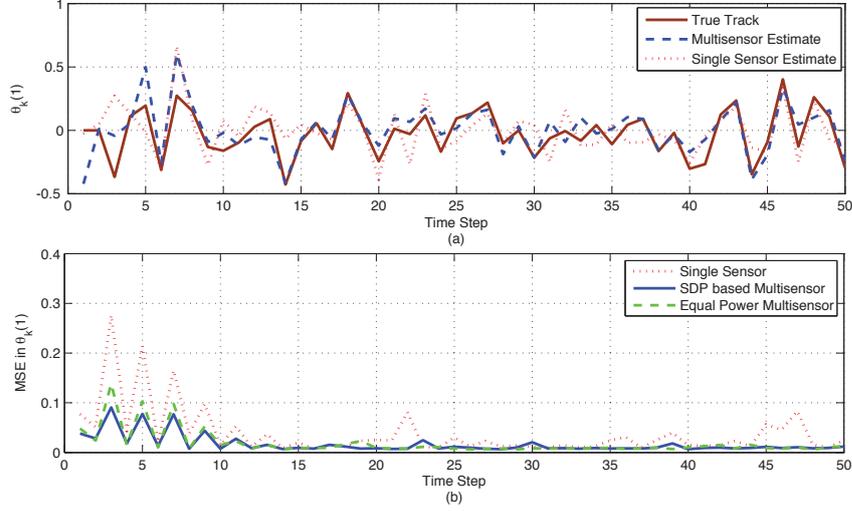


Figure 3.7: (a) True trajectory and estimate of the state θ_k under single sensor and multisensor case for a nonlinear dynamic model. (b) Mean square error versus total transmit power for a nonlinear dynamic model in an LSN.

3.5.2 Decentralized Bayes filtering for locating dynamic targets

We begin our discussion of simulation results for dynamic target's state estimation with a nonlinear state transition model of the third order. In this example, we consider a typical third order nonlinear autoregressive process described mathematically as $\mathbf{q}_{k+2} = -0.1\mathbf{q}_{k+1} - \mathbf{q}_k^3 + \mathbf{w}_k$ with the noise corrupted observations $\mathbf{y}_k = \mathbf{q}_k + \mathbf{n}_k$. Addressed in previous work [67], this system admits the following state-space formulations with the state $\boldsymbol{\theta}_k = (\boldsymbol{\theta}_k(1), \boldsymbol{\theta}_k(2))^T := (\mathbf{q}_k, \mathbf{q}_{k+1})^T$

$$\begin{aligned}\boldsymbol{\theta}_{k+1} &= \begin{pmatrix} 0 & 1 \\ -\boldsymbol{\theta}_k(1)^2 & -0.1 \end{pmatrix} \boldsymbol{\theta}_k + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathbf{v}_k \\ \mathbf{y}_k &= [\mathbf{G} \ 0] \boldsymbol{\theta}_k(1) + \mathbf{n}_k\end{aligned}$$

with $\mathbf{R}_{v_k} = 0.04$. An equivalent LFT model (3.34) and (3.49) for this third-order nonlinearity is achieved with the following deterministic parameters

$$\mathbf{F}_k = \begin{bmatrix} 0 & 1 \\ 0 & -0.1 \end{bmatrix}, \quad \mathbf{B}_k = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\mathbf{F}_{\Delta k} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{B}_{\Delta k} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \Delta(\theta_k) = \theta_k(1)\mathbf{I}_2$$

where \mathbf{I}_2 is the 2×2 identity matrix.

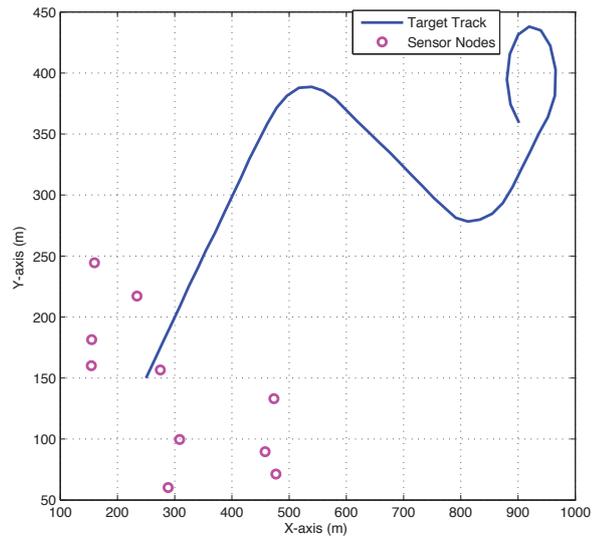


Figure 3.8: Trajectory of a maneuvering target and distribution of nodes over surveillance region.

Using $\bar{\theta}_{0|-1} = (0, 0)^T$ as the initial conditional estimate of the state with covariance $[\mathbf{R}_{\theta_{0|-1}}]_{i,j} = \rho^{|j-1|}$, $\rho = 0.75$, trajectory of the state θ_k for 50 time steps along with state estimates are shown in Fig. 3.7 where mean square error for the first component of the state variable obtained from 10^2 Monte Carlo runs is plotted using Theorem 3.3.1. These results suggest that the proposed multi-sensors approach outperforms the single sensor scheme for the measured state estimation. Also, the proposed optimal power allocation technique offers less MSE than equal power allocation.

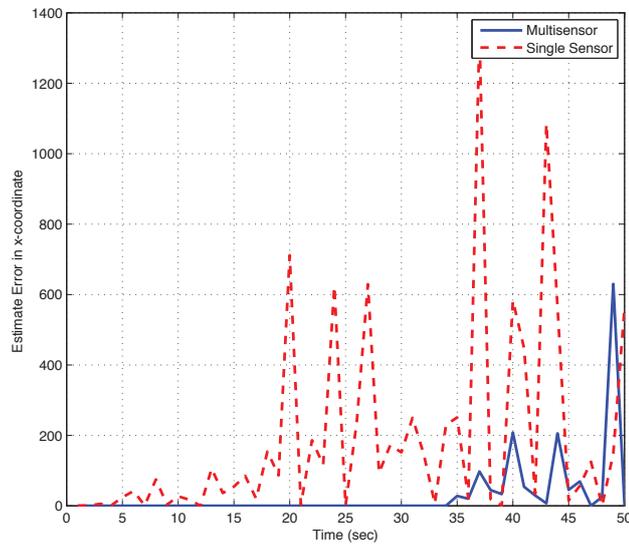


Figure 3.9: Target tracking performance of multisensor and single sensor in terms of estimation error of x-coordinate.

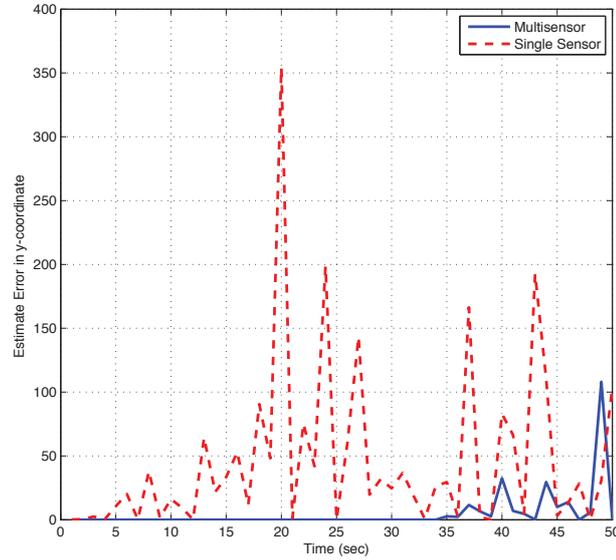


Figure 3.10: Target tracking performance of multisensor and single sensor in terms of estimation error of y-coordinate.

The problem of dynamic target tracking has been extensively addressed in previous literature (see, e.g. [12] and [13] and the references therein). In this example, we consider a vehicle moving along a trajectory as specified in Fig 3.8, with a dynamic model based on constant velocity and a coordinated turn model [12] to account for nonmaneuver and maneuver motion of the target. Corresponding state and the i th measurement equations are

$$\begin{aligned} \boldsymbol{\theta}_{k+1} &= \mathbf{A}\boldsymbol{\theta}_k + \mathbf{w}_k \\ y_{ik} &= \begin{pmatrix} \sqrt{(s_{i,x} - \boldsymbol{\theta}_k(1))^2 + (s_{i,y} - \boldsymbol{\theta}_k(3))^2} \\ \frac{s_{i,y} - \boldsymbol{\theta}_k(3)}{s_{i,x} - \boldsymbol{\theta}_k(1)} \end{pmatrix} + n_{ik}, \end{aligned}$$

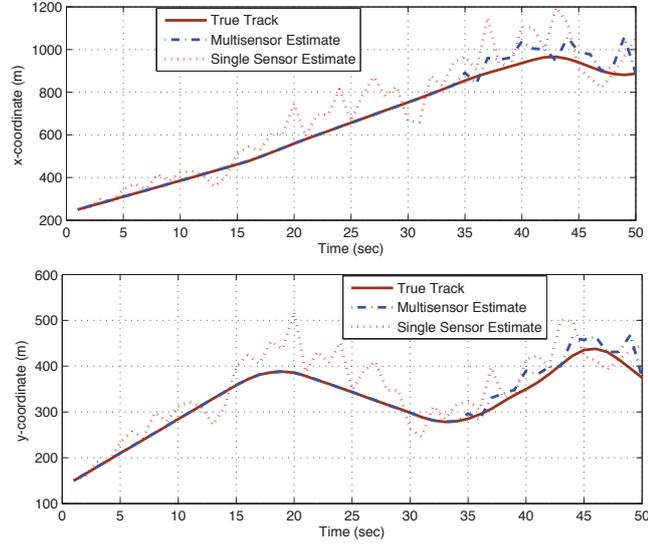


Figure 3.11: True trajectory and estimates by single and multisensor network.

where $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{w}_k})$ is the process noise, $\boldsymbol{\theta}_k = (p_{x_k}, \dot{p}_{x_k}, p_{y_k}, \dot{p}_{y_k})^T$ denotes the kinetic state of the target at time k consisting of the target coordinate $(p_{i,x}, p_{i,y})$ and its velocity $(\dot{p}_{i,x}, \dot{p}_{i,y})$, and with the system matrices

$$A = \begin{pmatrix} 1 & \frac{\sin \omega T}{\omega} & 0 & -\frac{1 - \cos \omega T}{\omega} \\ 0 & \cos \omega T & 0 & -\sin \omega T \\ 0 & \frac{1 - \cos \omega T}{\omega} & 1 & \frac{\sin \omega T}{\omega} \\ 0 & \sin \omega T & 0 & \cos \omega T \end{pmatrix}, \quad \mathbf{R}_{\mathbf{w}_k} = \begin{pmatrix} \frac{T^4}{4} & \frac{T^3}{2} & 0 & 0 \\ \frac{T^3}{2} & T^2 & 0 & 0 \\ 0 & 0 & \frac{T^4}{4} & \frac{T^3}{2} \\ 0 & 0 & \frac{T^3}{2} & T^2 \end{pmatrix}.$$

Here T is the sampling period and ω is the turn rate of the maneuvering target. $\mathbf{R}_{\boldsymbol{\theta}_k}$ is initialized as $\mathbf{R}_{\boldsymbol{\theta}_{0-1}} = \text{diag} [10, 2.5, 10, 2.5]$ with initial position and velocity as $(250, 150)$ and $(15, 15)$, respectively. The sensor nonlinear measurements include range and bearing information of the vehicle, corrupted by noise vector $\mathbf{n}_k \sim (0, \mathbf{R}_{\mathbf{n}_k})$. It is notable that range measurement function bears a moderate nonlinear map, hence (3.41) is used to approximate mean and other moments, whereas for highly nonlinear map of the bearing information, we use (3.43) to evaluate the required statistics with the same LFT system matrices as used in the static nonlinear scenario. In other words, implementation of Theorem 3.3.1 is sought to track the ever evolving state vector under linear state dynamics for this target tracking problem.

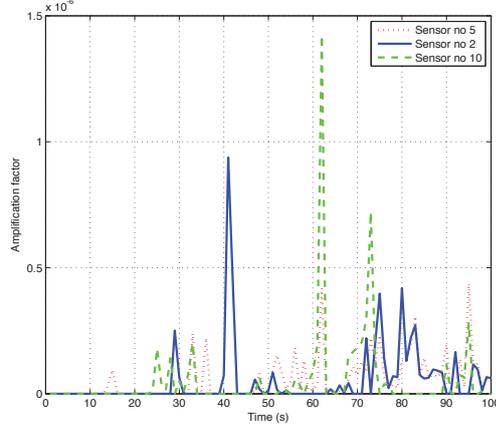


Figure 3.12: Distribution of power among different sensor nodes for the case $N = 10$.

Tracking of the moving vehicle is executed over a 2D surveillance region of $[100, 50] \times [1000, 450]m^2$. A multitude of sensor nodes $M = 10$, randomly deployed over the region of interest, takes measurements at a sampling period of 1 sec. For process noise, we choose $\sigma_w^2 = 0.5$. Similarly, variance of local as well as global channel noise while measuring range and bearing are $\sigma_{n,range}^2 = 0.5$, $\sigma_{n,bearing}^2 = \frac{\pi}{180}$ and $\sigma_{v,range}^2 = 1.0$, $\sigma_{v,bearing}^2 = \frac{\pi}{180}$, respectively. The target's trajectory, along with sensor nodes distribution have been shown in Fig 3.8. A one-dimensional view of the true and estimated trajectory for single and multiple sensors converted to Cartesian coordinates is shown in Fig. 3.11. At $k = 1$ sec target sets off from $(250, 150)$ at a constant velocity of 15 m/sec and after 14 sec performs a clockwise turn for 6 sec at a turn rate of $\omega = -0.2$ rad/sec. It takes two counterclockwise turns; one after 16 sec and then after 4 sec. Simulation results shown in Fig 3.9 and 3.10 suggest that multisensors perform better than a single sensor even when all available power is allocated entirely to the single sensor.

Of particular interest is the observation that for optimized estimation performance at all time instants, nodes with the best channel conditions, determined by the proposed strategy, are selected to sense and send their data. This is further elaborated in Fig 3.12 where distribution of power among 10 sensor nodes is plotted.

3.5.3 Joint sensor selection and power allocation

This section provides simulation results obtained from 5000 Monte Carlo realizations of the random channel gains h_i from sensors to the FC which follow normal distribution. Perfect channel state information is acknowledged at the FC. For computational implementation, the penalty parameter is set as $\mu = 10$ and the tolerance is $\varepsilon = 0.01$. Zero-mean additive white Gaussian noise (AWGN) with variance $\mathbf{R}_w(i, i) = 0.1$ is considered for all sensors which are assumed to be randomly deployed in a given geographic area to localize an object through linear/nonlinear noisy observations.

Firstly, consider localizing a static object in R^3 , i.e. the vector parameter to be estimated consists of three Cartesian coordinates. Observations performed by sensor nodes include measurements of range with $f_i(\boldsymbol{\theta}) = \sqrt{(\theta_1 - s_{i,x})^2 + (\theta_2 - s_{i,y})^2 + (\theta_3 - s_{i,z})^2}$, elevation angle by sensors with $f_i(\boldsymbol{\theta}) = \frac{\theta_2 - s_{i,y}}{\theta_1 - s_{i,x}}$ and azimuth angle by sensors with $f_i(\boldsymbol{\theta}) = \frac{\theta_3 - s_{i,z}}{\sqrt{(\theta_1 - s_{i,x})^2 + (\theta_2 - s_{i,y})^2}}$, where $(s_{i,x}, s_{i,y}, s_{i,z})$ are the Cartesian coordinates of the i th sensor [82]. Under the linear sensor network (LSN) framework, nonlinear maps $f_i(\boldsymbol{\theta})$ are linearized at $\bar{\boldsymbol{\theta}}$ to have the linear sensor model so $f_i(\boldsymbol{\theta}) = \mathbf{F}_i \boldsymbol{\theta}$ with $F_i = \partial f_i(\bar{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta}$ for definition of (3.52). The noise v_i is assumed to be zero-mean AWGN with its variance for the range measurement $\mathbf{R}_{v,range}(i, i) = 0.1$ whereas for bearing measurement (e.g. for both elevation and azimuth angles) $\mathbf{R}_{v,bearing}(i, i) = \pi/180$. Accordingly, $\mathbf{R}_{y\theta} = \mathbf{F}\mathbf{R}_\theta$ and $\mathbf{R}_y = \mathbf{F}\mathbf{R}_\theta\mathbf{F}^T + \mathbf{R}_v$, where $\mathbf{F} := (\mathbf{F}_1^T, \dots, \mathbf{F}_N^T)^T$. Figure 3.13 shows MSE plot versus total transmit power budget under for $N = 30$ and $N_0 = 6$ and $N_0 = 15$. A comparison is drawn among the proposed technique, the power re-optimization method of [82] for solution of mixed convex program (3.71), and the lower bound provided by the relaxed program (3.72) (for power allocation to all sensors). Note that the power re-optimization method jointly performs sensor selection and power allocation in two steps: first, it solves SDP under MSE minimization to determine the active set of sensors, and second, it solves another SDP to optimally allocate power over the reduced set of active sensors. As illustrated in the figure, the MSE curves of the proposed method with $N_0 = 15$ is closest to the lower bound curve (for optimal value of (3.72)) which implies that the proposed method is capable of locating the optimal solution of (3.71) by just using half of available sensor nodes. In particular, the curve of the proposed method with $N_0 = 15$ gradually approaches its lower bound as total power increases. The joint optimized

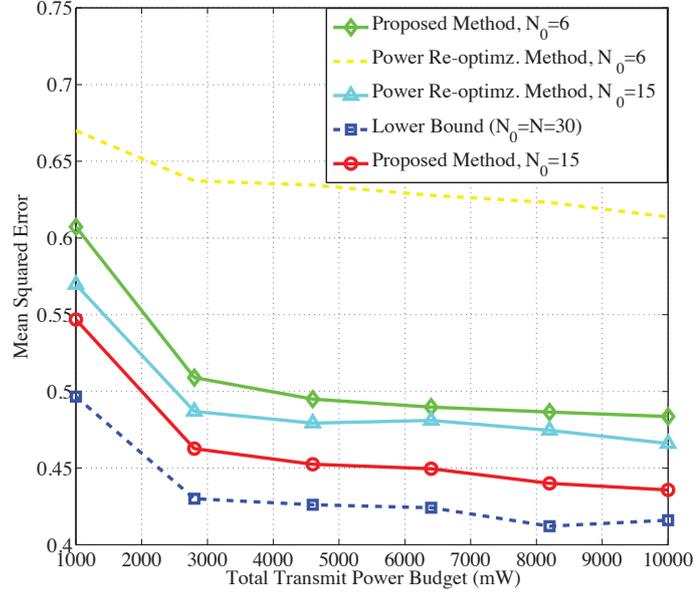


Figure 3.13: MSE estimation performance comparison in an LSN for $N = 30$.

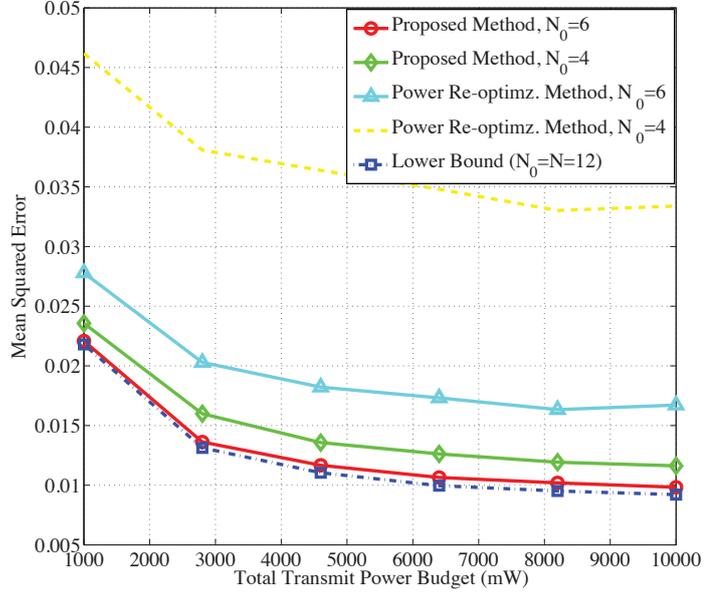
sensor assignment and power allocation are well motivated. It can be seen that the performance of the power re-optimization method of [82] does not perform better than the proposed method, and its performance is even far worse when the active sensor number N_0 is decreased to 6.

Table 3.3: Average No. of iterations for LSN with $N = 30$

P_T (mW)	500	1000	1500	2000	2500	3000	3500	4000
Iterations ($N_0 = 3$)	3.0280	2.0380	2.1140	2.1720	2.3160	2.3520	2.3880	2.4560
Iterations ($N_0 = 6$)	2.0640	2.1460	2.2360	2.3540	2.4220	2.4820	2.4780	2.6860

Next, consider the simulation example for nonlinear sensing function used in [82], where an object in a 2D plane, which is located by its (x, y) -axis coordinates $\theta = (\theta_1, \theta_2)^T$. The sensor's measurements consist of the range with $f_i(\theta) = \sqrt{(s_{i,x} - \theta_1)^2 + (s_{i,y} - \theta_2)^2}$ and bearing $f_i(\theta) = \frac{s_{i,y} - \theta_2}{s_{i,x} - \theta_1}$. Local sensor noise is zero-mean AWGN with $\mathbf{R}_{v,range}(i, i) = 0.1$ and $\mathbf{R}_{v,bearing}(i, i) = \pi/180$.

Of course, there are no analytical forms of the auto-covariance \mathbf{R}_y and cross-covariance $\mathbf{R}_{y\theta}$ but they can be approximately computed well based on the linear fraction transformation technique

Figure 3.14: MSE estimation performance comparison in an NSN for $N = 12$.Table 3.4: Average No. of iterations for NSN with $N = 20$

P_T (mW)	500	1000	1500	2000	2500	3000	3500	4000
Iterations ($N_0 = 6$)	2.1148	2.0312	2.0672	2.1200	2.2012	2.2692	2.3384	2.3848
Iterations ($N_0 = 14$)	4.1704	4.0172	3.9592	3.8868	3.6976	3.5236	3.2752	3.0064

[82], which is also used in our simulation. Like Figure 3.13, Figures 3.14 and 3.15 present a performance comparison between the proposed method and that of [82]. Again, the lower bound provided by program (3.72) is found to be very close to the proposed curve of $N_0 = 6$ (for $N = 12$) and $N_0 = 10$ (for $N = 20$) in Figures 3.14 and 3.15, respectively. On the other hand, the above mentioned power re-optimization method of [82] performs poorly compared to the corresponding proposed technique for both cases.

Tables I and II provide the averaged number of iterations, i.e. the averaged number of the convex program (3.77) used in the proposed d.c. procedure for solution of (3.71). Just up to three iterations are needed for its solutions. The computational efficiency of our d.c. procedure is obvious.

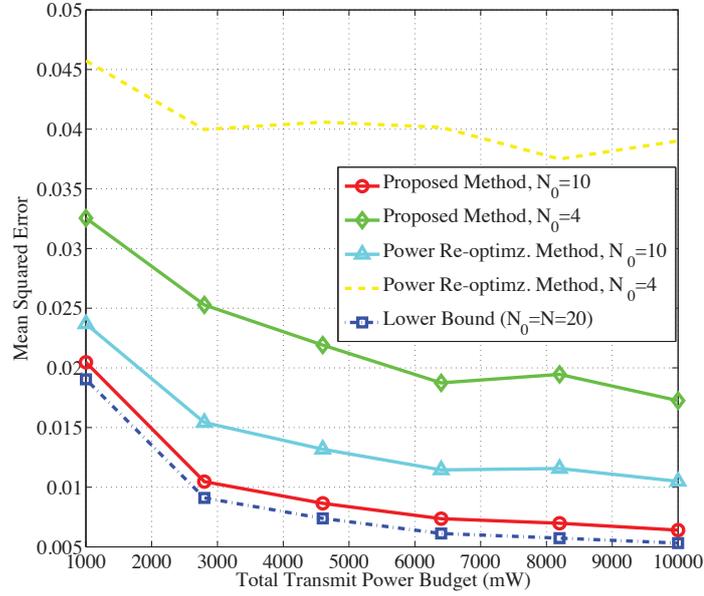


Figure 3.15: MSE estimation performance comparison in an NSN for $N = 20$.

3.6 Summary

The problem of power allocation among sensor nodes for locating a static target or for tracking a dynamic target in either linear or nonlinear sensing systems has been addressed. These sensors observe the targets and then transmit their noisy observations through noisy wireless channels to the FC where the final estimate is carried out. Due to limited energy resources, it is desired to develop an optimized power allocation technique which is able to minimize mean square error of the estimate under a given power budget. A novel technique based on tractable optimization (SDP) and approximation (unscented and linear fractional transformations) has been proposed. The multi-sensor diversity has been fully exploited to arrive at an accurate estimate of the target's state. Accompanying simulation results clearly showed the viability of the theoretical results.

In the next part, a joint program of active sensor assignment and power allocation in linear and nonlinear sensor network is considered. Its optimized solutions are quickly located by the developed d.c. procedure of local search. Nevertheless, simulation results indicate that the global optimal solutions are obtained in a few iterations, which result in the full WSN capacity within

only about half of sensor nodes invoked.

Chapter 4

Relay Beamforming Designs in Multi-User Wireless Relay Networks

Relay-assisted wireless communication is currently one of the most active research topics (see e.g. [34, 48, 86]). The key advantage of relay-assisted communication is that spatial diversity can be exploited via cooperation of relays/nodes to improve the link reliability, and to extend the communication coverage area [16, 39, 108]. In a multi-user relaying framework, distributed relay nodes are employed to assist the communication among a number of multiple sources and destinations [50]. Relay-assisted communication schemes are generally classified into three main categories: decode-and-forward (DF), compress-and-forward (CF) and amplify-and-forward (AF) [46, 48]. Due to its simplicity in mathematical modeling and low cost in implementation, AF relaying scheme has been extensively studied. In AF relaying, the relays simply amplify the signals received from the sources and then forward the amplified versions to all the destinations. In a multi-user relay network, beamforming is implemented at relay nodes such that the desired signal of each user at the destination can be constructively combined, while the interferences and noise are efficiently mitigated [28, 30, 63, 70, 71].

The outline of this chapter is organized as follows. A brief introduction and motivation of the problem is given in Section 4.1. Section 4.2 describes the system model when the sources simultaneously transmit to the relays (i.e., under non-orthogonal source transmissions). It also formulates the optimization problem and discusses challenges in obtaining solutions. Section 4.3

presents a formulation of the optimization problem as a d.c. program and develops an iterative algorithm to obtain the solutions. Section 4.4 considers a beamforming design problem when the sources operate in orthogonal channels in communicating with the relays in the first phase (so as to improve the quality of service for networks with a larger number of source-destination pairs). Section 4.5.1 formulates the joint source power allocation and relay beamforming design problem for which proposed solution is presented in Section 4.5.2. A simplified joint optimization based on equally constrained source powers is also developed in Section 4.5.2. Simulation results that support the algorithm developments are presented in Section 4.6. Conclusions are drawn in Section 4.7.

4.1 Introduction

A large body of works considered beamforming design in the context of relay power minimization subject to the signal-to-interference-plus-noise ratio (SINR) constraints. While the objective function of beamforming power is quadratic convex in the complex beamforming vector, the SINR constraints are indefinite (nonconvex), which makes the overall program nonconvex quadratic. Such a nonconvex quadratic program can be trivially rewritten as a rank-one constrained matrix program with the variable dimension substantially increased. Without individual relay power constraints, this nonconvex program can be transformed into a relaxed semi-definite (convex) program (SDP) by dropping the rank-one constraint because the relaxed SDP often outputs its rank-one optimal solution [30, 63]. However, as shown both theoretically and numerically in [70–72], the presence of the individual relay power constraints (which are necessary to reflect practical limitations of relay hardware [16]) would make the optimized solution of the corresponding relaxed SDP just low-rank, but no longer rank-one. Not having rank-one with the optimal solution of the SDP relaxation means that it is not possible to locate even feasible beamforming solutions. On the other hand, the low-rank of its optimal solution means that a further randomization [30], which in fact tries to randomly generate feasible solutions in the low-dimensional subspace of eigenvectors corresponding to nonzero eigenvalues of its optimal solutions, is highly inefficient. It can be easily shown that a nontrivial randomization would simply perform the same. Our previous works [70, 71] developed efficient algorithms to address this beamforming power minimization,

where the original matrix rank-one (nonconvexity) constrained optimization is exactly reformulated as minimization of nonsmooth matrix spectral objective functions under convex constraints. Additionally, the objective functions of the reformulated optimizations have been shown to be d.c. (difference of two convex) functions in the matrix variables [102]. The d.c. path-following procedures [8] have been known to effectively locate the global optimal solutions.

One of the main objectives of emerging wireless technologies is to provide reliable services at high throughput under limited power resources. In the context of multi-user relay-assisted wireless networks, the goal is to maximize the information throughput among multiple sources and destinations in a fair manner. Accordingly, the beamforming design problem is to maximize the minimum information throughput among source-destination pairs. In other words, the problem is *maximin* throughput optimization, which is related to the maximization problem of the minimum SINR under limited beamforming power constraints. However, in contrast to the above mentioned power beamforming program, the objective function of maximin information throughput optimization is neither smooth nor concave. This means that the program is maximization of nonconcave and nonsmooth objective functions subject to (nonconvex) rank-one constraints. Such a program belongs to the most challenging optimization class [102].

Specifically, even by dropping the rank-one constraints, the relaxed maximization problem is still not computationally tractable as the objective function is still not concave. The conventional rank-one dropping relaxation simply does not work. This problem has been considered by [17] for a particular case of a total relay power constraint only. A standard bisection procedure to iteratively update feasible SINR thresholds has been employed. As mentioned above, without individual relay power constraints, the feasibility/infeasibility of a SINR threshold can be accurately solved by the SDP relaxation. Under the individual relay power constraints, this nonconvex program cannot be addressed by SDP relaxation and so the bisection method is no longer a good solution procedure. Again, a trivial randomization would perform the same. In the end, this maximin throughput optimization is convex constrained optimization in beamforming vector. Such a matrix rank-one constrained re-formulation not only increases the problem dimension substantially but also unnecessarily invites the nonconvex hard rank-one constraint that destroys the tractable convexity of the original constraints, while the new objective function is still nonconvex.

In other words, the existing matrix optimization approach reformulates the original hard convex constrained minimization of a nonconvex objective function to a much harder and much larger dimension nonconvex constrained minimization of a nonconvex objective function.

In this chapter, we follow [8, 98] and adopt d.c. programming [45, 102] to directly address this nonconvex maximin throughput optimization problem, bypassing the above mentioned matrix rank-one constrained optimization. The main issue is how to recognize and then explore hidden partial convex structures of the problem at hand in order to develop an effective algorithm to find the solutions. As both dimension and nonconvexity rank of this optimization problem are surely high for applicability of global optimization algorithms [102], we pursue alternative iterations of the local search, which however are able to locate approximately global optimal solutions. The advantages of the developed algorithm are summarized as follows:

- Unlike other sequential quadratic/SDP iterations (see e.g. [6, 7]) for locating solutions of the Karush-Kuhn-Tucker (KKT) necessary optimality condition, the proposed iterations can locate optimal solutions.
- Unlike other nonsmooth optimization iterations (see e.g. [5]), which may suffer the so-called zero-progression steps that slows down the convergence, the proposed iterations surely improve the solution at every step and thus converge quickly.
- The implementation of the proposed iterations is simple and does not involve the control of step size (which can be both sensitive and difficult to determine) [5–7].

Furthermore, this chapter also considers joint design of source power allocation and relay beamforming for multi-user multi-relay wireless networks. The design problem is formulated as maximizing the worst signal-to-interference-plus-noise ratio (SINR) among users (SINR maximin optimization) subject to meaningful power constraints. This computationally intractable nonconvex problem is recast as an equivalent d.c. (difference of two convex functions) programming. By exploiting specific structures of the d.c. program, an iterative algorithm with low computational complexity is then developed to obtain the optimal solution. Furthermore, a simplified sub-optimal solution with equally constrained source powers is also suggested, which is efficient in both computation and required communication overhead but still has very good performance. Extensive

simulation results show that our optimal and sub-optimal solutions of the joint optimization problem significantly outperform the solution obtained by separate designs of source power allocation and/or relay beamforming.

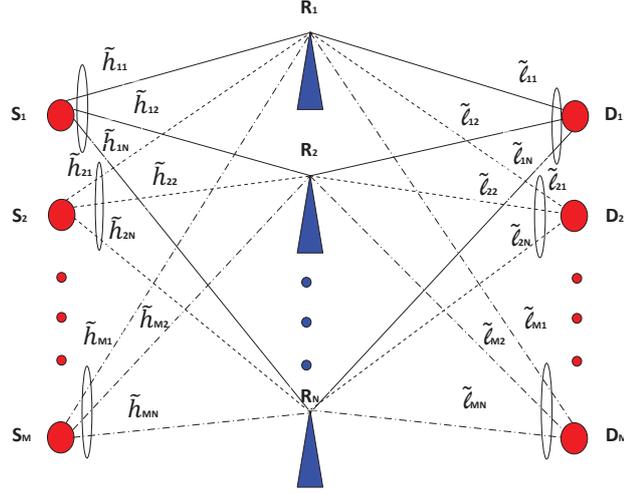


Figure 4.1: A multi-user amplify-and-forward wireless relay network.

4.2 Maximin Throughput Optimization: Problem Formulations and Challenges

Figure 5.1 illustrates a wireless relay network, in which M pairs of source-destination communicates with the help of N relays. All relay and user nodes are equipped with a single antenna and operate in half-duplex mode. In the first time-slot, the sources simultaneously send their signals to the relays. The relays “amplify” their received signals by multiplying with certain weights and then simply forward these processed signals to all destinations. The case that sources communicate with the relays over orthogonal channels shall be treated in Section 4.4.

Let $\mathbf{s} = (s_1, s_2, \dots, s_M)^T \in \mathcal{C}^M$ be the vector of signals sent by M sources, which is assumed to be zero mean and component-wise independent with variance $\sigma_s^2 = \mathbb{E}[|s_i|^2]$. Let $\tilde{\mathbf{h}}_m = (\tilde{h}_{m1}, \tilde{h}_{m2}, \dots, \tilde{h}_{mN})^T \in \mathcal{C}^N$, $m = 1, 2, \dots, M$, be the vector of uplink channel coefficients between the m th source and all the relays. Likewise, let $\tilde{\mathbf{l}}_i = (\tilde{l}_{i1}, \tilde{l}_{i2}, \dots, \tilde{l}_{iN})^T \in \mathcal{C}^N$, $i = 1, 2, \dots, M$, be the vector of downlink channel coefficients between all the relays and the i th destination. The received signals at all the

relays can be collectively written as

$$\mathbf{y}_{\text{up}} = \sum_{m=1}^M \tilde{\mathbf{h}}_m s_m + \mathbf{n}_{\text{R}}, \quad (4.1)$$

where $\mathbf{n}_{\text{R}} = (n_{\text{R},1}, \dots, n_{\text{R},N})^T \in \mathcal{C}^N$ represents the additive noises at the relay receivers, which is modeled as zero-mean white Gaussian random variables of variance $\sigma_{\text{R}}^2 = \mathbb{E}[|n_{\text{R},n}|^2]$, $n = 1, 2, \dots, N$.

Let $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ be the vector of beamforming weights. Then the relays send the following signals to the destinations:

$$\mathbf{y}_{\text{amp}} = \mathbf{x} \odot \mathbf{y}_{\text{up}} = \sum_{m=1}^M \mathbf{x} \odot \tilde{\mathbf{h}}_m s_m + \mathbf{x} \odot \mathbf{n}_{\text{R}}. \quad (4.2)$$

Accordingly, the received signal at the i th destination is

$$\mathbf{y}_{\text{D},i} = \langle \tilde{\mathbf{l}}_i, \mathbf{y}_{\text{amp}} \rangle + n_{\text{D},i} = \sum_{m=1}^M \langle \tilde{\mathbf{c}}_{mi}, \mathbf{x} \rangle s_m + \langle \tilde{\mathbf{l}}_i \odot \mathbf{n}_{\text{R}}, \mathbf{x} \rangle + n_{\text{D},i}, \quad (4.3)$$

where $n_{\text{D},i}$ is the additive Gaussian noise at the i th destination with variance σ_{D}^2 . The vector $\tilde{\mathbf{c}}_{mi} = \tilde{\mathbf{l}}_i \odot \tilde{\mathbf{h}}_m$ represents the *compound* channel coefficient from source m to destination i .

To incorporate the channel uncertainties in the beamforming design, assume that

$$\begin{aligned} \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^H &= \mathbf{h}_m \mathbf{h}_m^H + \Delta \mathbf{H}_m, \quad m = 1, 2, \dots, M, \\ \tilde{\mathbf{l}}_i \tilde{\mathbf{l}}_i^H &= \mathbf{l}_i \mathbf{l}_i^H + \Delta \mathbf{L}_i, \quad i = 1, 2, \dots, M, \\ \tilde{\mathbf{c}}_{mi} \tilde{\mathbf{c}}_{mi}^H &= \mathbf{c}_{mi} \mathbf{c}_{mi}^H + \Delta \mathbf{C}_{mi}, \quad m, i = 1, 2, \dots, M, \end{aligned} \quad (4.4)$$

where

- \mathbf{h}_m , \mathbf{l}_i and \mathbf{c}_{mi} are nominal values that can be obtained through channel estimation (see e.g. [95], where the compound channel coefficients \mathbf{c}_{mi} can be directly obtained at the destination side);
- $\Delta \mathbf{H}_m \in \mathcal{C}^{N \times N}$, $\Delta \mathbf{L}_i \in \mathcal{C}^{N \times N}$ and $\Delta \mathbf{C}_{mi} \in \mathcal{C}^{N \times N}$ are Hermitian symmetric full block uncertainty matrices [116], which satisfy the following spectral constraints:

$$\rho(\Delta \mathbf{H}_m) \leq \zeta^2, \quad \rho(\Delta \mathbf{L}_i) \leq \zeta^2, \quad \rho(\Delta \mathbf{C}_{mi}) \leq \zeta^2, \quad m, i = 1, 2, \dots, M. \quad (4.5)$$

The uncertainties in the form of (4.4) are called *unstructured* (see e.g. [116]). On the other hand, the *structured* uncertainties have the following forms:

$$\tilde{\mathbf{h}}_m = \mathbf{h}_m + \delta \mathbf{h}_m, \tilde{\boldsymbol{\ell}}_i = \boldsymbol{\ell}_i + \delta \boldsymbol{\ell}_i, \tilde{\mathbf{c}}_{mi} = \mathbf{c}_{mi} + \delta \mathbf{c}_{mi}, m, i = 1, 2, \dots, M \quad (4.6)$$

with

$$\|\delta \mathbf{h}_m\| \leq \bar{\zeta}, \|\delta \boldsymbol{\ell}_i\| \leq \bar{\zeta}, \|\delta \mathbf{c}_{mi}\| \leq \bar{\zeta}, m, i = 1, 2, \dots, M. \quad (4.7)$$

In fact, (4.6) can be rewritten in the form of (4.4) with

$$\begin{aligned} \Delta \mathbf{H}_m &= \mathbf{h}_m (\delta \mathbf{h}_m)^H + (\delta \mathbf{h}_m) \mathbf{h}_m^H + \delta \mathbf{h}_m (\delta \mathbf{h}_m)^H, \\ \Delta \mathbf{L}_i &= \boldsymbol{\ell}_i (\delta \boldsymbol{\ell}_i)^H + (\delta \boldsymbol{\ell}_i) \boldsymbol{\ell}_i^H + \delta \boldsymbol{\ell}_i (\delta \boldsymbol{\ell}_i)^H, \\ \Delta \mathbf{C}_{mi} &= \mathbf{c}_{mi} (\delta \mathbf{c}_{mi})^H + (\delta \mathbf{c}_{mi}) \mathbf{c}_{mi}^H + \delta \mathbf{c}_{mi} (\delta \mathbf{c}_{mi})^H, m, i = 1, 2, \dots, M \end{aligned} \quad (4.8)$$

and

$$\zeta^2 := \bar{\zeta} \left(\bar{\zeta} + 2 \max \left\{ \max_{m=1,2,\dots,M} \|\mathbf{h}_m\|, \max_{i=1,2,\dots,M} \|\boldsymbol{\ell}_i\|, \max_{m,i=1,2,\dots,M} \|\mathbf{c}_{mi}\| \right\} \right). \quad (4.9)$$

For the received signal given in (4.3), since only the signal component $\langle \mathbf{c}_{ii}, \mathbf{x} \rangle s_i$ is of interest, its “robust” power is defined by

$$\begin{aligned} S_i(\mathbf{x}) &= \sigma_s^2 \inf_{\rho(\Delta \mathbf{C}_{ii}) \leq \zeta^2} |\langle \mathbf{x}, \tilde{\mathbf{c}}_{ii} \rangle|^2 \\ &= \sigma_s^2 \inf_{\rho(\Delta \mathbf{C}_{ii}) \leq \zeta^2} [|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 + \mathbf{x}^H \Delta \mathbf{C}_{ii} \mathbf{x}] \\ &= \sigma_s^2 [|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 \|\mathbf{x}\|^2]. \end{aligned} \quad (4.10)$$

Analogously, the “robust” interference power in (4.3) is defined by

$$\begin{aligned} \text{INT}_i(\mathbf{x}) &= \sup_{\rho(\Delta \mathbf{L}_i) \leq \zeta^2, \Delta \mathbf{C}_{mi} \leq \zeta^2} \left[\sigma_s^2 \sum_{m \neq i} |\langle \mathbf{x}, \tilde{\mathbf{c}}_{mi} \rangle|^2 + \sigma_{\mathbf{R}}^2 |\langle \tilde{\boldsymbol{\ell}}_i, \mathbf{x} \rangle|^2 \right] \\ &= \sigma_s^2 \sum_{m \neq i} (|\langle \mathbf{x}, \mathbf{c}_{mi} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2) + \sigma_{\mathbf{R}}^2 (|\langle \boldsymbol{\ell}_i, \mathbf{x} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2). \end{aligned} \quad (4.11)$$

It follows that the “robust” signal-to-interference-plus-noise ratio (SINR) at destination i can be expressed as

$$\text{SINR}_i(\mathbf{x}) = \frac{S_i(\mathbf{x})}{\text{INT}_i(\mathbf{x}) + \sigma_{\mathbf{D}}^2}. \quad (4.12)$$

Note that by (4.2) the total beamforming power across all the relays is

$$P_T(\mathbf{x}) = \mathbb{E}\{\|\mathbf{y}_{\text{amp}}\|^2\} = \sigma_s^2 \sup_{\rho(\Delta \mathbf{H}_m) \leq \zeta^2} \sum_{m=1}^M \|\mathbf{x} \odot \tilde{\mathbf{h}}_m\|^2 + \sigma_{\mathbf{R}}^2 \|\mathbf{x}\|^2 = \langle \mathbf{x} \mathbf{x}^H, \mathbf{R} \rangle,$$

with

$$\mathbf{R} := \text{diag}[\mathbf{r}], \mathbf{r} := (r_1, \dots, r_N)^T, r_n = \sigma_s^2 \sum_{m=1}^M (|h_{mn}|^2 + \zeta^2) + \sigma_R^2, n = 1, 2, \dots, N.$$

On the other hand, the individual beamforming power at relay n is $P_n(x_n) = r_n|x_n|^2$.

In order to quantify the quality of service (QoS) for the multi-user relay communication framework under consideration, we use the metric of information throughput computed for each source-destination pair. This quantity is expressed for the i th source-destination pair as

$$I_i(\mathbf{x}) = \log_2(1 + \text{SINR}_i(\mathbf{x})). \quad (4.13)$$

In particular, the program of maximin information throughput under the individual relay power constraints, $P_n(x_n) \leq \gamma_n$, $n = 1, 2, \dots, N$, is formulated as

$$\max_{\mathbf{x} \in \mathcal{C}^N} \min_{i=1,2,\dots,M} \log_2(1 + \text{SINR}_i(\mathbf{x})) : r_n|x_n|^2 \leq \gamma_n, n = 1, 2, \dots, N. \quad (4.14)$$

The equivalent program in terms of SINR threshold maximin optimization is

$$\max_{\mathbf{x} \in \mathcal{C}^N} \min_{i=1,2,\dots,M} \text{SINR}_i(\mathbf{x}) := \frac{\sigma_s^2 (|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 |\mathbf{x}|^2)}{\sigma_s^2 \sum_{m \neq i} (|\langle \mathbf{x}, \mathbf{c}_{mi} \rangle|^2 + \zeta^2 |\mathbf{x}|^2) + \sigma_R^2 (|\langle \mathbf{l}_i, \mathbf{x} \rangle|^2 + \zeta^2 |\mathbf{x}|^2) + \sigma_D^2} : r_n|x_n|^2 \leq \gamma_n, n = 1, 2, \dots, N. \quad (4.15)$$

While the constraints in maximin programs (4.14) and (4.15) are convex, their objective functions are not concave nor convex. To the best of our knowledge, effective methods to solve this program were not known.

For the particular case of $\zeta = 0$, i.e., there is no channel uncertainty, the original minimax problem, which is essentially minimization of relay transmit power subject to constraints on SINR, becomes:

$$\min_{\mathbf{x}} \max_{n=1,2,\dots,N} [r_n|x_n|^2/\gamma_n] : \text{SINR}_i(\mathbf{x}) \geq \alpha, i = 1, 2, \dots, M \quad (4.16)$$

where α is the lower bound of individual SINR. The above problem is reformulated in reference [30] by using the equivalent matrix rank-one constrain:

$$\begin{aligned} \min_{\mathbf{X} \in \mathcal{C}^{N \times N}} \max_{n=1,2,\dots,N} [r_n \mathbf{X}(n,n)/\gamma_n] : \mathbf{X} \geq 0, \text{rank}(\mathbf{X}) = 1, \\ \sigma_s^2 \langle \mathbf{X}, \mathbf{C}_{ii} \rangle \geq \alpha [\sigma_s^2 \sum_{m \neq i} \langle \mathbf{X}, \mathbf{C}_{mi} \rangle + \sigma_R^2 \langle \mathbf{X}, \mathbf{L}_i \rangle + \sigma_D^2], \end{aligned} \quad (4.17)$$

which is then relaxed to the following SDP by dropping the only nonconvex rank-one constraint, also known as semidefinite relaxation (SDR):

$$\min_{\mathbf{X} \in \mathcal{C}^{N \times N}} \max_{n=1,2,\dots,N} [r_n \mathbf{X}(n,n) / \gamma_n] : \sigma_s^2 \langle \mathbf{X}, \mathbf{C}_{ii} \rangle \geq \alpha [\sigma_s^2 \sum_{m \neq i} \langle \mathbf{X}, \mathbf{C}_{mi} \rangle + \sigma_R^2 \langle \mathbf{X}, \mathbf{L}_i \rangle + \sigma_D^2], \mathbf{X} \geq 0. \quad (4.18)$$

It then follows that one can use the bisection method to find the maximum α_{opt} of those α such that the optimal value of SDP (4.18) is less than one. Thus α_{opt} provides an upper bound for the maximin program (4.15). If SDP (4.18) at $\alpha = \alpha_{\text{opt}}$ has rank-one optimization solution $\mathbf{X}_{\text{opt}} = \mathbf{x}_{\text{opt}} \mathbf{x}_{\text{opt}}^H$ then such \mathbf{x}_{opt} is surely optimal solution of the nonconvex program (4.16) and then of maximin program (4.15). However, \mathbf{X}_{opt} is often not rank-one [70, 71], so it admits a singular value decomposition (SVD) $\mathbf{X}_{\text{opt}} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H$ with unitary matrix $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ and diagonal matrix $\mathbf{\Sigma}$, whose diagonal entries are arranged in decreasing order and so $\Sigma(i,i) = 0$ for $i \geq i_{\text{opt}} := \text{rank}(\mathbf{X}_{\text{opt}}) \geq 2$. The SDR-based randomization generates

$$\mathbf{x}^{(v)} = \mathbf{U} \mathbf{\Sigma}^{1/2} \mathbf{v} = \sum_{i=1}^{i_{\text{opt}}} \Sigma^{1/2}(i,i) v_i \mathbf{u}_i, \quad (4.19)$$

where $\mathbf{v} = (v_1, \dots, v_N)$ is generated as unit-variance complex Gaussian vector with uncorrelated components. Hence, $\mathbf{x}^{(v)}$ always belongs to $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{i_{\text{opt}}}\}$ of low dimension i_{opt} , i.e., randomization is performed in a low-dimensional sub-space. Each randomly generated vector $\mathbf{x}^{(v)}$ is further scaled by $t_v := \min_{n=1,2,\dots,N} \sqrt{\gamma_n / r_n |\mathbf{x}_n^{(v)}|^2}$ to satisfy the power constraint in maximin program (4.15) and then $\min_{n=1,2,\dots,N} \text{SINR}_i(t_v \mathbf{x}^{(v)})$ is calculated to update its current best value. Obviously, the optimized solution of maximin program (4.15) hardly resides in subspace $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{i_{\text{opt}}}\} \subset \mathcal{C}^N$. As such, no matter how many $\mathbf{x}^{(v)}$ are randomly generated, the value of maximin program (4.15) is not much improved. As a matter of fact, the simulation results in Section 6 will show that a more trivial randomization will simply work quicker because it calls only one SDP solver.

From the view point of finding solutions, the above standard bisection method is backward because the original maximin program (4.15) as maximization of a nonconcave objective function over convex constraints is in fact easier than the minimax program (4.16) used for the bisection routine, which is minimization of a convex objective function over nonconvex constraints [102]. Moreover, variable \mathbf{X} in matrix rank-one constrained reformulation (4.17) for program (4.16) is of dimension $N(N+1)/2$, which is $(N+1)/2$ times larger than that of variable \mathbf{x} in the maximin program (4.15).

Although both the minimax program (4.16) and maximin program (4.15) in nonrobust scenarios ($\zeta = 0$) can be solved by our nonsmooth matrix spectral optimization algorithm with the matrix variable setting $\mathbf{X} = \mathbf{x}\mathbf{x}^H$ of dimension $N(N+1)/2$ [71], the next section develops and presents direct approaches for obtaining solutions of maximin programs (4.14) and (4.15), hence bypassing any bisection approach and computationally demanding matrix optimization.

4.3 Maximin Throughout Optimization by d.c. Programming

Our strategy is to express the maximin program (4.15) in the following canonical form of d.c. optimization [102]:

$$\min_{\mathbf{z}} [f(\mathbf{z}) - g(\mathbf{z})] : \mathbf{z} \in \mathcal{H}, \quad (4.20)$$

where \mathbf{z} is a matrix and/or vector variable, \mathcal{H} is a compact and convex set, $f(\cdot)$ is a quasi-convex function (i.e., for each t the level set $\{\mathbf{z} : f(\mathbf{z}) \leq t\}$ is either empty or convex) and $g(\cdot)$ is a convex and smooth function. Suppose that $\mathbf{z}^{(\kappa)} \in \mathcal{H}$ and $\nabla g(\mathbf{z}^{(\kappa)})$ is the gradient of $g(\cdot)$ at $\mathbf{z}^{(\kappa)}$. Then [102]

$$f(\mathbf{z}) - g(\mathbf{z}) \leq f(\mathbf{z}) - g(\mathbf{z}^{(\kappa)}) - \langle \nabla g(\mathbf{z}^{(\kappa)}), \mathbf{z} - \mathbf{z}^{(\kappa)} \rangle \quad \forall \mathbf{z} \in \mathcal{H}$$

It then follows that the following convex program provides a global upper bound minimization for d.c. program (5.17):

$$\min_{\mathbf{z}} [f(\mathbf{z}) - g(\mathbf{z}^{(\kappa)}) - \langle \nabla g(\mathbf{z}^{(\kappa)}), \mathbf{z} - \mathbf{z}^{(\kappa)} \rangle] : \mathbf{z} \in \mathcal{H}, \quad (4.21)$$

where $\mathbf{z}^{(\kappa)}$ is also its feasible solution. Moreover, for the optimal solution $\mathbf{z}^{(\kappa+1)}$ of (5.17),

$$\begin{aligned} f(\mathbf{z}^{(\kappa+1)}) - g(\mathbf{z}^{(\kappa+1)}) &\leq f(\mathbf{z}^{(\kappa+1)}) - g(\mathbf{z}^{(\kappa)}) - \langle \nabla g(\mathbf{z}^{(\kappa)}), \mathbf{z}^{(\kappa+1)} - \mathbf{z}^{(\kappa)} \rangle \\ &\leq f(\mathbf{z}^{(\kappa)}) - g(\mathbf{z}^{(\kappa)}) - \langle \nabla g(\mathbf{z}^{(\kappa)}), \mathbf{z}^{(\kappa)} - \mathbf{z}^{(\kappa)} \rangle \\ &= f(\mathbf{z}^{(\kappa)}) - g(\mathbf{z}^{(\kappa)}), \end{aligned}$$

which means that $\mathbf{z}^{(\kappa+1)}$ is better than $\mathbf{z}^{(\kappa)}$ toward (5.17). Thus, initialized from a feasible $\mathbf{z}^{(0)} \in \mathcal{H}$, for $\kappa = 0, 1, \dots$, generating $\mathbf{z}^{(\kappa)}$ by the optimal solution of convex program (5.18) is a path-following algorithm, which converges to an optimal solution.

It can be seen that there are infinite number of d.c. representations for the same nonconvex problems and it is clear from (5.18) that the efficiency of the d.c. path-following procedure critically depends on the choice of d.c. representation. The next section shall examine this issue in

more details. Unlike the approach of [69], which attempts to locate the global optimal solution of d.c. program (5.17) by combining iterations (5.18) with customized branch-and-bound of high computational complexity, here we develop effective equivalent d.c. decompositions that make iterations (5.18) converge to the global optimal solution.

Proposition 1. *The maximin program (4.15) is equivalent to*

$$\max_{\mathbf{x} \in \mathcal{C}^N, \mathbf{y} \in \mathbb{R}_+^M} \min_{i=1,2,\dots,M} \varphi_i(\mathbf{x}, y_i) := \frac{|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} : \quad (4.22a)$$

$$\sum_{m \neq i} (|\langle \mathbf{x}, \mathbf{c}_{mi} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2) + \frac{\sigma_R^2}{\sigma_s^2} (|\langle \boldsymbol{\ell}_i, \mathbf{x} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2) \leq y_i, \quad i = 1, 2, \dots, M, \quad (4.22b)$$

$$r_n |x_n|^2 \leq \gamma_n, \quad n = 1, 2, \dots, N. \quad (4.22c)$$

Proof. For any feasible solution (\mathbf{x}, \mathbf{y}) of (4.22), it is obvious that

$$\frac{|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} \leq \frac{|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 \|\mathbf{x}\|^2}{\sum_{m \neq i} (|\langle \mathbf{x}, \mathbf{c}_{mi} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2) + \frac{\sigma_R^2}{\sigma_s^2} (|\langle \boldsymbol{\ell}_i, \mathbf{x} \rangle|^2 + \zeta^2 \|\mathbf{x}\|^2)}.$$

Therefore,

$$\max (4.22) \leq \max (4.15). \quad (4.23)$$

On the other hand, for the optimal solution \mathbf{x}_{opt} of the maximin program (4.15), it is also obvious that $(\mathbf{x}_{\text{opt}}, \mathbf{y}_{\text{opt}})$ with $y_{\text{opt},i} = \sum_{m \neq i} (|\langle \mathbf{x}_{\text{opt}}, \mathbf{c}_{mi} \rangle|^2 + \zeta^2 \|\mathbf{x}_{\text{opt}}\|^2) + \frac{\sigma_R^2}{\sigma_s^2} (|\langle \boldsymbol{\ell}_i, \mathbf{x}_{\text{opt}} \rangle|^2 + \zeta^2 \|\mathbf{x}_{\text{opt}}\|^2)$ is feasible to program (4.22). Then

$$\max (4.15) = \min_{i=1,2,\dots,M} \varphi_i(\mathbf{x}_{\text{opt}}, y_{\text{opt},i}) \leq \max (4.22). \quad (4.24)$$

It is concluded from (4.23) and (4.24) that $(\mathbf{x}_{\text{opt}}, \mathbf{y}_{\text{opt}})$ must be the optimal solution of (4.22). \square

Obviously, (4.22b)-(4.22c) are convex quadratic constraints, which can also be represented by the linear matrix inequality (LMI):

$$\begin{bmatrix} \mathbf{Q}_i & \mathbf{Q}_i \mathbf{x} \\ \mathbf{x}^H \mathbf{Q}_i & y_i \end{bmatrix} \succeq 0; \quad \begin{bmatrix} \frac{\gamma_n}{r_n} & x_n \\ x_n^H & 1 \end{bmatrix} \succeq 0, \quad n = 1, 2, \dots, N, \quad (4.25)$$

$$\mathbf{Q}_i := \sum_{m \neq i} (\mathbf{c}_{mi} \mathbf{c}_{mi}^H + \zeta^2 \mathbf{I}_N) + \frac{\sigma_R^2}{\sigma_s^2} (\boldsymbol{\ell}_i \boldsymbol{\ell}_i^H + \zeta^2 \mathbf{I}_N), \quad i = 1, 2, \dots, M.$$

Proposition 2. *Each fractional function $|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 / (y_i + \sigma_D^2 / \sigma_s^2)$ is convex*

Proof. First, each function $\phi_i(s, y_i) := |s|^2 / (y_i + \sigma_D^2 / \sigma_s^2)$ is convex in $(s, y_i) \in \mathcal{C} \times \mathbb{R}_+$ because its Hessian, which is defined by

$$H_{s, y_i} := \begin{bmatrix} \frac{\partial^2 \phi_i(s, y_i)}{\partial^2 s} & \frac{\partial^2 \phi_i(s, y_i)}{\partial s \partial y_i} \\ \frac{\partial^2 \phi_i(s, y_i)}{\partial s \partial y_i} & \frac{\partial^2 \phi_i(s, y_i)}{\partial^2 y_i} \end{bmatrix} = \frac{1}{y_i + \sigma_D^2 / \sigma_s^2} \begin{bmatrix} 2 & -\frac{s + \bar{s}}{y_i + \sigma_D^2 / \sigma_s^2} \\ -\frac{s + \bar{s}}{y_i + \sigma_D^2 / \sigma_s^2} & \frac{2|s|^2}{(y_i + \sigma_D^2 / \sigma_s^2)^2} \end{bmatrix}$$

is positive definite [102]:

$$\begin{aligned} H_{s, y_i}(1, 1) &= 2 > 0, \quad H_{s, y_i}(2, 2) = 2|s|^2 / (y_i + \sigma_D^2 / \sigma_s^2)^3 \geq 0, \\ \det(H_{s, y_i}) &= 4(|s|^2 - \operatorname{Re}(s)^2) / (y_i + \sigma_D^2 / \sigma_s^2)^3 \geq 0. \end{aligned}$$

This means that, for any $0 \leq \theta \leq 1$ and $(s, y_i), (s', y'_i) \in \mathcal{C} \times \mathbb{R}_+$, one has

$$\frac{|\theta s + (1 - \theta)s'|^2}{\theta y_i + (1 - \theta)y'_i + \sigma_D^2 / \sigma_s^2} \leq \frac{\theta |s|^2}{y_i + \sigma_D^2 / \sigma_s^2} + \frac{(1 - \theta)|s'|^2}{y'_i + \sigma_D^2 / \sigma_s^2}.$$

Therefore,

$$\begin{aligned} \frac{|\langle \theta \mathbf{x} + (1 - \theta)\mathbf{x}', \mathbf{c}_{ii} \rangle|^2}{\theta y_i + (1 - \theta)y'_i + \sigma_D^2 / \sigma_s^2} &= \frac{|\theta \langle \mathbf{x}, \mathbf{c}_{ii} \rangle + (1 - \theta)\langle \mathbf{x}', \mathbf{c}_{ii} \rangle|^2}{\theta y_i + (1 - \theta)y'_i + \sigma_D^2 / \sigma_s^2} \\ &\leq \theta \frac{|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2}{y_i + \sigma_D^2 / \sigma_s^2} + (1 - \theta) \frac{|\langle \mathbf{x}', \mathbf{c}_{ii} \rangle|^2}{y'_i + \sigma_D^2 / \sigma_s^2}, \end{aligned}$$

which shows that $|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 / (y_i + \sigma_D^2 / \sigma_s^2)$ is convex. \square

Also from [102] one has

$$\begin{aligned} \min_{i=1,2,\dots,M} \varphi_i(\mathbf{x}, y_i) &= \min_{i=1,2,\dots,M} \left[\frac{|\langle \mathbf{x}, \mathbf{c}_{ii} \rangle|^2 - \zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} \right] \\ &= \sum_{j=1}^M \frac{|\langle \mathbf{x}, \mathbf{c}_{jj} \rangle|^2}{y_j + \sigma_D^2 / \sigma_s^2} - \max_{i=1,2,\dots,M} \left(\frac{\zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} + \sum_{m \neq i} \frac{|\langle \mathbf{x}, \mathbf{c}_{mm} \rangle|^2}{y_m + \sigma_D^2 / \sigma_s^2} \right) \\ &= f_{02}(\mathbf{x}, \mathbf{y}) - f_{01}(\mathbf{x}, \mathbf{y}), \end{aligned} \quad (4.26)$$

where

$$f_{02}(\mathbf{x}, \mathbf{y}) := \sum_{j=1}^M \frac{|\langle \mathbf{x}, \mathbf{c}_{jj} \rangle|^2}{y_j + \sigma_D^2 / \sigma_s^2}, \quad f_{01}(\mathbf{x}, \mathbf{y}) := \max_{i=1,2,\dots,M} \left(\frac{\zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} + \sum_{m \neq i} \frac{|\langle \mathbf{x}, \mathbf{c}_{mm} \rangle|^2}{y_m + \sigma_D^2 / \sigma_s^2} \right). \quad (4.27)$$

Note that $f_{02}(\mathbf{x}, \mathbf{y})$ is convex since it is a sum of convex functions [102].

Proposition 3. *The function $f_{01}(\mathbf{x}, \mathbf{y})$ is quasi-convex.*

Proof. For every $t \geq 0$, the level set

$$\{(\mathbf{x}, \mathbf{y}) : f_{01}(\mathbf{x}, \mathbf{y}) \leq t\} \quad (4.28)$$

is fully described by the following constraints:

$$\frac{\zeta^2 \|\mathbf{x}\|^2}{y_i + \sigma_D^2 / \sigma_s^2} \leq t_{0i}, \quad i = 1, 2, \dots, M; \quad \frac{|\langle \mathbf{x}, \mathbf{c}_{mm} \rangle|^2}{y_m + \sigma_D^2 / \sigma_s^2} \leq t_{1m}, \quad m = 1, 2, \dots, M, \quad (4.29a)$$

$$t_{0i} + \sum_{m \neq i} t_{1m} \leq t, \quad i = 1, 2, \dots, M. \quad (4.29b)$$

On the other hand, the LMIs representations for convex constraints (4.29a) are

$$\begin{bmatrix} t_{0i} & \mathbf{x}^H \\ \mathbf{x} & \frac{y_i + \sigma_D^2 / \sigma_s^2}{\zeta^2} I_N \end{bmatrix} \succeq 0, \quad i = 1, 2, \dots, M; \quad \begin{bmatrix} t_{1m} & \langle \mathbf{x}, \mathbf{c}_{mm} \rangle \\ \langle \mathbf{x}, \mathbf{c}_{mm} \rangle & y_m + \sigma_D^2 / \sigma_s^2 \end{bmatrix} \succeq 0, \quad m = 1, \dots, M. \quad (4.30)$$

Thus, the level set (4.28) is fully described by LMIs (4.29b) and (4.30), so it is convex, proving that f_{01} is quasi-convex. \square

The preceding analysis implies that the maximin program (4.22) can be recast as the following d.c. program:

$$- \min_{\mathbf{x} \in \mathcal{C}^N, \mathbf{y} \in \mathcal{R}_+^M} [f_{01}(\mathbf{x}, \mathbf{y}) - f_{02}(\mathbf{x}, \mathbf{y})] : (4.22b) - (4.22c) \quad (4.31)$$

which belongs to the d.c. class (5.17) with $\mathbf{z} = (\mathbf{x}, \mathbf{y})$, $f(\cdot) = f_{01}(\cdot)$, $g(\cdot) = f_{02}(\cdot)$ and \mathcal{K} described by convex constraints (4.22b)-(4.22c).

Initialized from a feasible solution $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ of (4.31), by (5.18), the path-following $(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)})$ is the optimal solution of the following convex program:

$$\min_{\mathbf{x} \in \mathcal{C}^N, \mathbf{y} \in \mathcal{R}_+^M} \left[f_{01}(\mathbf{x}, \mathbf{y}) - f_{02}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}) - \sum_{i=1}^M \langle \nabla \varphi_i(\mathbf{x}^{(k)}, y_i^{(k)}), (\mathbf{x}, y_i) - (\mathbf{x}^{(k)}, y_i^{(k)}) \rangle \right] : (4.22b) - (4.22c) \quad (4.32)$$

where

$$\langle \nabla \varphi_i(\mathbf{x}^{(k)}, y_i^{(k)}), (\mathbf{x}, y_i) - (\mathbf{x}^{(k)}, y_i^{(k)}) \rangle = \frac{2\operatorname{Re}(\overline{\langle \mathbf{x}^{(k)}, \mathbf{c}_{ii} \rangle} \cdot \langle \mathbf{c}_{ii}, \mathbf{x} - \mathbf{x}^{(k)} \rangle)}{y_i^{(k)} + \sigma_D^2 / \sigma_s^2} - \frac{|\langle \mathbf{x}^{(k)}, \mathbf{c}_{ii} \rangle|^2 (y_i - y_i^{(k)})}{(y_i^{(k)} + \sigma_D^2 / \sigma_s^2)^2}.$$

The computational complexity of convex program (4.32) solver is $O((N+M)^4)$.

The main result in this section is summarized in the following algorithm.

Path-following d.c. iterations (DCI)

Initialization: Set $\kappa = 0$ and choose an initial feasible solution $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ of (4.31).

κ -th iteration: Solve the convex program (4.32) to obtain the optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$

and set $\kappa \rightarrow \kappa + 1$, $(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}) \rightarrow (\mathbf{x}^*, \mathbf{y}^*)$. Given the tolerance level ε , stop if $|f_{01}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}) - f_{01}(\mathbf{x}^{(\kappa-1)}, \mathbf{y}^{(\kappa-1)}) - f_{02}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}) + f_{02}(\mathbf{x}^{(\kappa-1)}, \mathbf{y}^{(\kappa-1)})| / (f_{01}(\mathbf{x}^{(\kappa-1)}, \mathbf{y}^{(\kappa-1)}) - f_{01}(\mathbf{x}^{(\kappa-1)}, \mathbf{y}^{(\kappa-1)})) \leq \varepsilon$.

In order to use the existing convex program software such as SeDuMi [91] for solution of (4.32), we express it by the following SDP:

$$\min_{\mathbf{x} \in \mathcal{C}^N, \mathbf{y} \in \mathcal{R}_+^M, t, t_{0i}, t_{1i}, i=1,2,\dots,M} \left[t - f_{02}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}) - \sum_{i=1}^M \left(\frac{2\operatorname{Re}(\langle \mathbf{x}^{(\kappa)}, \mathbf{c}_{ii} \rangle \cdot \langle \mathbf{c}_{ii}, \mathbf{x} - \mathbf{x}^{(\kappa)} \rangle)}{y_i^{(\kappa)} + \sigma_D^2 / \sigma_s^2} - \frac{|\langle \mathbf{x}^{(\kappa)}, \mathbf{c}_{ii} \rangle|^2 (y_i - y_i^{(\kappa)})}{(y_i^{(\kappa)} + \sigma_D^2 / \sigma_s^2)^2} \right) \right] : (4.25), (4.29b), (4.30), (4.33)$$

4.4 Beamforming Design with Orthogonal Source Transmissions

Our previous studies (see e.g. [70, 71]) show that it is not practical to aim for a very high SINR threshold under nonorthogonal (concurrent) transmissions of the source-destination pairs if there are more than five pairs in the network. This motivates us to consider a system model where the uplink channels $\tilde{\mathbf{h}}_m, m = 1, 2, \dots, M$ are orthogonal [32, 72]. This allows the beamforming to be applied individually on the received signal from each source before combining them for forwarding to the destinations.

Let $\mathbf{x}_m = (x_{m1}, x_{m2}, \dots, x_{mN})^T \in \mathcal{C}^N$ be the beamforming weight applied by the relays to the signals received from source m , namely $\tilde{\mathbf{h}}_m s_m + \mathbf{n}_R$. After beamforming is applied, the signal received from source m becomes

$$\mathbf{y}_{\text{amp}}^{(m)} = \mathbf{x}_m \odot (\tilde{\mathbf{h}}_m s_m + \mathbf{n}_R) \quad (4.34)$$

Thus, the signals forwarded by the relays to the destination are given in the following vector:

$$\mathbf{y}_{\text{amp}} = \sum_{m=1}^M \mathbf{x}_m \odot (\tilde{\mathbf{h}}_m s_m + \mathbf{n}_R). \quad (4.35)$$

The received signal at destination i is thus

$$\mathbf{y}_{D,i} = \langle \tilde{\boldsymbol{\ell}}_i, \mathbf{y}_{\text{amp}} \rangle + n_{D,i} = \sum_{m=1}^M [\langle \tilde{\mathbf{c}}_{mi}, \mathbf{x}_m \rangle s_m + \langle \tilde{\boldsymbol{\ell}}_i \odot \mathbf{n}_R, \mathbf{x}_m \rangle] + n_{D,i}. \quad (4.36)$$

As in Section II, all the concerned channels gains \mathbf{h}_m , \mathbf{c}_{mi} and $\boldsymbol{\ell}_i$ are subject to uncertainties as described in (4.4)-(4.5). Similarly to (4.10) and (4.11), the robust power of the desired signal component at destination i is

$$S_i(\mathbf{x}_i) = \sigma_s^2 (|\langle \mathbf{c}_{ii}, \mathbf{x}_i \rangle|^2 - \zeta^2 \|\mathbf{x}_i\|^2),$$

while the robust interference power is

$$\text{INT}_i(\mathbf{x}_1, \dots, \mathbf{x}_M) = \sigma_s^2 \sum_{m \neq i}^M (|\langle \mathbf{c}_{mi}, \mathbf{x}_m \rangle|^2 + \zeta^2 \|\mathbf{x}_m\|^2) + \sigma_R^2 \sum_{m=1}^M (|\langle \boldsymbol{\ell}_i, \mathbf{x}_m \rangle|^2 + \zeta \|\mathbf{x}_m\|^2) + \sigma_D^2.$$

The total beamforming power can be computed as

$$P_T(\mathbf{x}_1, \dots, \mathbf{x}_M) = \sum_{\rho(\Delta \mathbf{H}_i) \leq \zeta^2} \mathbb{E}\{\|\mathbf{y}_{\text{amp}}\|^2\} = \sum_{i=1}^M \langle \sigma_s^2 \text{diag}[\mathbf{h}_i \odot \mathbf{h}_i^H] + (\sigma_s^2 \zeta^2 + \sigma_R^2) \mathbf{I}_N, \mathbf{x}_i \mathbf{x}_i^H \rangle, \quad (4.37)$$

while the individual beamforming power at relay n is

$$P_n(\mathbf{x}_1, \dots, \mathbf{x}_M) = \sum_{m=1}^M (\sigma_s^2 |h_{mn}|^2 + \sigma_s^2 \zeta^2 + \sigma_R^2) |x_{mn}|^2. \quad (4.38)$$

The maximin optimization of the information throughput can now be expressed as

$$\max_{\mathbf{x}_m \in \mathcal{C}^N, m=1,2,\dots,M} \min_{i=1,2,\dots,M} \log_2 \left(1 + \frac{S_i(\mathbf{x}_i)}{\text{INT}_i(\mathbf{x}_1, \dots, \mathbf{x}_M)} \right) \quad (4.39a)$$

$$\text{s.t.} \quad \sum_{m=1}^M (\sigma_s^2 |h_{mn}|^2 + \sigma_s^2 \zeta^2 + \sigma_R^2) |x_{mn}|^2 \leq \gamma_n, \quad n = 1, 2, \dots, N \quad (4.39b)$$

Like the maximin program (4.15), (4.39) is equivalent to

$$\max_{\mathbf{x}_i, \mathbf{y}_i} \min_{i=1,2,\dots,M} \varphi_i(\mathbf{x}_i, \mathbf{y}_i) := \frac{|\langle \mathbf{c}_{ii}, \mathbf{x}_i \rangle|^2 - \zeta^2 \|\mathbf{x}_i\|^2}{\mathbf{y}_i + \frac{\sigma_D^2}{\sigma_s^2}} \quad : \quad (4.39b), \quad (4.40a)$$

$$\sum_{m \neq i}^M (|\langle \mathbf{c}_{mi}, \mathbf{x}_m \rangle|^2 + \zeta^2 \|\mathbf{x}_m\|^2) + \frac{\sigma_R^2}{\sigma_s^2} \sum_{m=1}^M (|\langle \boldsymbol{\ell}_i, \mathbf{x}_m \rangle|^2 + \zeta^2 \|\mathbf{x}_m\|^2) \leq \mathbf{y}_i, \quad i = 1, 2, \dots, M. \quad (4.40b)$$

Define

$$\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_M^T)^T \in \mathcal{C}^{NM}, \quad \tilde{\mathbf{C}}_i = \text{diag}(\mathbf{c}_{1i}^H, \dots, \mathbf{c}_{(i-1)i}^H, \mathbf{0}_{1 \times N}, \mathbf{c}_{(i+1)i}^H, \dots, \mathbf{c}_{Mi}^H) \in \mathcal{C}^{M \times (NM)},$$

$$\mathbf{J}_i = \text{diag}(\mathbf{I}_N, \dots, \mathbf{0}_N, \dots, \mathbf{I}_N), \quad \tilde{\boldsymbol{\ell}}_i = \text{diag}(\boldsymbol{\ell}_i^H, \dots, \boldsymbol{\ell}_i^H) \in \mathcal{C}^{M \times (NM)}.$$

It then follows that

$$\sum_{m \neq i}^M (|\langle \mathbf{c}_{mi}, \mathbf{x}_m \rangle|^2 + \zeta^2 \|\mathbf{x}_m\|^2) + \frac{\sigma_{\mathbf{R}}^2}{\sigma_s^2} \left(\sum_{m=1}^M |\langle \ell_i, \mathbf{x}_m \rangle|^2 + \zeta^2 \|\mathbf{x}_m\|^2 \right) = \mathbf{x}^H (\tilde{\mathbf{C}}_i^H \tilde{\mathbf{C}}_i + \zeta^2 \mathbf{J}_i + \frac{\sigma_{\mathbf{R}}^2}{\sigma_s^2} (\tilde{\mathbf{L}}_i^H \tilde{\mathbf{L}}_i + \zeta^2 \mathbf{I}_{MN})) \mathbf{x}$$

and the convex constraints (4.40b) are expressed by the following LMIs:

$$\begin{bmatrix} \tilde{\mathbf{Q}}_i & \tilde{\mathbf{Q}}_i \mathbf{x} \\ \mathbf{x}^H \tilde{\mathbf{Q}}_i & y_i \end{bmatrix} \succeq 0, \quad \tilde{\mathbf{Q}}_i := \tilde{\mathbf{C}}_i^H \tilde{\mathbf{C}}_i + \zeta^2 \mathbf{J}_i + \frac{\sigma_{\mathbf{R}}^2}{\sigma_s^2} (\tilde{\mathbf{L}}_i^H \tilde{\mathbf{L}}_i + \zeta^2 \mathbf{I}_{MN}), \quad i = 1, 2, \dots, M. \quad (4.41)$$

Similarly, the LMI representation for robust power constraint (4.39b) is

$$\begin{bmatrix} \mathbf{I}_M & A_n(x_{1n}, \dots, x_{Mn}) \\ A_n^H(x_{1n}, \dots, x_{Mn}) & \gamma_n \end{bmatrix} \succeq 0, \quad n = 1, 2, \dots, N, \quad (4.42)$$

where $A_n(x_{1n}, \dots, x_{Mn}) = \left(\sqrt{\sigma_s^2 |h_{1n}|^2 + \sigma_{\mathbf{R}}^2 + \zeta^2 x_{1n}}, \dots, \sqrt{\sigma_s^2 |h_{Mn}|^2 + \sigma_{\mathbf{R}}^2 + \zeta^2 x_{Mn}} \right)^T$. Therefore, like (4.31), the maximin program (4.40) can be recast by the following d.c. program:

$$- \min_{\mathbf{x} \in \mathcal{C}^{NM}, \mathbf{y} \in \mathbf{R}_+^M} [f_{01}(\mathbf{x}, \mathbf{y}) - f_{02}(\mathbf{x}, \mathbf{y})] : (4.41), (4.42) \quad (4.43)$$

where the function $f_{01}(\mathbf{x}, \mathbf{y}) := \max_{i=1,2,\dots,M} [\zeta^2 \|\mathbf{x}_i\|^2 / (y_i + \sigma_{\mathbf{D}}^2 / \sigma_s^2) + \sum_{m \neq i} |\langle \mathbf{c}_{mm}, \mathbf{x}_m \rangle|^2 / (y_m + \sigma_{\mathbf{D}}^2 / \sigma_s^2)]$ is quasi-convex while the function $f_{02}(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^M |\langle \mathbf{c}_{ii}, \mathbf{x}_i \rangle|^2 / (y_i + \sigma_{\mathbf{D}}^2 / \sigma_s^2)$ is convex and smooth.

Obviously, DCI is also applicable to find the solutions of (4.43) by initializing from a feasible solution $(\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_M^{(0)}, \mathbf{y}^{(0)})$ of (4.43). In particular, the following convex program is required at κ -th iteration for generating $(\mathbf{x}_1^{(\kappa+1)}, \dots, \mathbf{x}_M^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)})$ instead of (4.32):

$$\min_{\mathbf{x} \in \mathcal{C}^{NM}, \mathbf{y} \in \mathbf{R}_+^M} \left[f_{01}(\mathbf{x}, \mathbf{y}) - f_{02}(\mathbf{x}^{(\kappa)}, \mathbf{y}^{(\kappa)}) - \sum_{i=1}^M \left[\frac{2 \operatorname{Re}(\langle \mathbf{x}_i^{(\kappa)}, \mathbf{c}_{ii} \rangle \cdot \langle \mathbf{c}_{ii}, \mathbf{x}_i - \mathbf{x}_i^{(\kappa)} \rangle)}{y_i^{(\kappa)} + \sigma_{\mathbf{D}}^2 / \sigma_s^2} - \frac{|\langle \mathbf{x}_i^{(\kappa)}, \mathbf{c}_{ii} \rangle|^2 (y_i - y_i^{(\kappa)})}{(y_i^{(\kappa)} + \sigma_{\mathbf{D}}^2 / \sigma_s^2)^2} \right] \right] : (4.39b), (4.41) \quad (4.44)$$

Similarly to (4.33), the above convex program can be easily converted to a SDP format so that the existing SDP solvers can be readily used.

4.5 Joint Optimization of Source Power Allocation and Relay Beamforming

Recently, reference [21] has applied a path-following procedure of d.c. programming [98, 102] to jointly design the source powers and relay beamforming for minimizing the total transmitted power while meeting the SINR requirements of all users. The limitation in [21] is that the individual transmission capacities of both sources and relays expressed via the individual power constraints could not be incorporated into the optimization. Although these additional meaningful individual power constraints are convex, their feasibility when enforced together with the practical SINR requirements is not quite computationally tractable. In other words, a feasible solution needed for initialization of any iterative optimization is hardly found by random generation [21]. Moreover, the resultant feasible set may be disconnected that makes the iterative solutions converge to a wrong solution if it is initialized from a feasible solution, which is disconnected from the optimal solution. It is no wonder only impractically low SINR target 3dB was set in [21]. To deal with the infeasibility issue, the admission control should be taken into consideration, which leads to other highly intractable optimization problems [76, 88].

This section examines an optimization problem that maximizes the minimum SINR subject to the mentioned individual power constraints. Different to the min-power problem in [21], which can be infeasible, the max-min SINR problem considered in this paper is always feasible. The admission control is not only absolutely unnecessary but also can be resolved consequently in our setting. The fairness among users is also taken into account by considering the quality of the worst user. Inspired by the efficiency of a special class of d.c. algorithm [43, 81, 98], we cast the optimization problem into a d.c. programming through elegant variable changes. In particular, the design problem is shown as minimization of a d.c. objective function over the set of convex constraints. A low complexity iterative algorithm (DCI), which exhibits a fast convergence, is then devised. Unlike other sequential quadratic iterations (see e.g. [6, 7]) for targeting solutions of the Karush-Kuhn-Tucker (KKT) necessary optimality condition, DCI is capable of locating the optimal solutions. Also, different from nonsmooth optimization iterations (see e.g. [5]) which may suffer from the so-called zero-progression steps that slow down the convergence, DCI surely

improves the solution at every step and thus converges quickly. The implementation of DCI is simple and does not involve the control of step size, which can be both sensitive and difficult to determine [5–7].

It should also be pointed out that more complicated programs of joint precoder design for MIMO and joint beamforming for multiantenna relay broadcast channels have been considered in [111, 113], but their solutions appear to be far from being efficient in both performance and computation. In addition to the joint optimization strategy, we also suggest a simplified sub-optimal strategy in which the source powers are constrained equal. Thus, the corresponding d.c. program has only one more power variable as compared to that for the beamforming-only program [81]. Consequently, the simplified joint optimization becomes much less computationally demanding. It will be shown by simulation results that the proposed simplified sub-optimal solution also gives better results than the results obtained by separate optimizations of source powers and relay beamforming. This means that the number of design parameters can be substantially reduced in the joint optimization. As these parameters are found at the destination and fed back to the corresponding sources and relays, a reduction in the overhead communications is provided.

4.5.1 System Model and Formulation of the Joint Design Problem

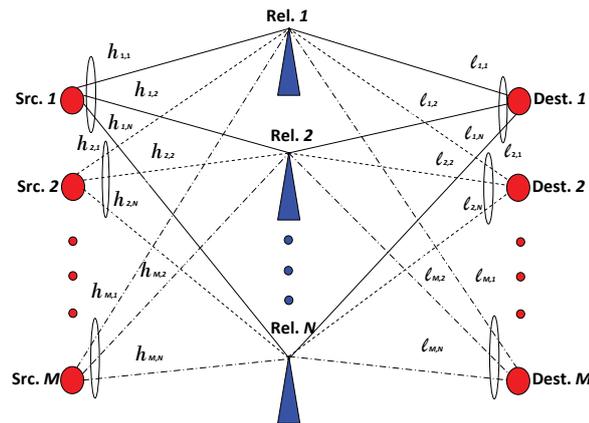


Figure 4.2: A multi-user multi-relay network.

Figure 5.2 shows a wireless communication network under consideration in which M sources communicate in pair to M destinations with the assistance of N relays. This scenario applies when

the direct links from sources to destinations cannot be used due to the severe path loss and shadowing. All user terminals (sources and destinations) and relays are equipped with a single antenna and operate in half-duplex mode. As mentioned, we consider the amplify-and-forward signal processing strategy for relays. The transmission from sources to destinations is carried out in two time slots. Let $\mathbf{s} = (s_1, s_2, \dots, s_M)^T \in \mathcal{C}^M$ with $\mathbf{s} \sim \mathcal{CN}(0, \mathbf{I}_M)$ represent the transmitted symbols from all the sources. Also denote the transmitted powers at the sources by $\mathbf{p} = (p_1, p_2, \dots, p_M)^T \in \mathcal{R}_+^M$. In the first time slot, all the sources send their signals to all the relays. All the channels in the network are assumed to be independent Rayleigh flat-fading channels and let

$$\mathbf{h}_m = (h_{m,1}, h_{m,2}, \dots, h_{m,N})^T \in \mathcal{C}^N, \quad m = 1, 2, \dots, M \quad (4.45)$$

be the vector of the ‘‘uplink’’ channels between the m th source and all the relays. The assumption of statistical independence of channels over time (4.45) is based on the consideration of a block fading channel model where the channel state remains constant within a fading block, but becomes independent across a different fading block.

The received signals at the relays can be collectively given by vector \mathbf{r}_r as

$$\mathbf{r}_r = \sum_m^M \mathbf{h}_m \sqrt{p_m} s_m + \mathbf{n}_r \in \mathcal{C}^N \quad (4.46)$$

where $\mathbf{n}_r = (n_{r,1}, \dots, n_{r,N})^T \in \mathcal{C}^N$ represents the additive white Gaussian noise components at the relays and $\mathbf{n}_r \sim \mathcal{CN}(0, \sigma_r^2 \mathbf{I}_N)$. In practice, the source terminals have their own power limits. As such, the following constraints are imposed on the source transmitted powers:

$$0 \leq p_m \leq P_m^{S,\max}, \quad m = 1, \dots, M. \quad (4.47)$$

where $P_m^{S,\max}$ is a maximum allowable transmit power of the m th source terminal.

As the channel state information is available at the relays, the relays perform beamforming to adjust the phases and amplitudes of their received signals so that the signals are constructively combined at destinations. In general, let $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \mathcal{C}^N$ be the beamforming weights to be used by the relays. Then, in the second time slot, the relays send the following signals to the destinations:

$$\mathbf{r}_b = \mathbf{x} \odot \mathbf{r}_r = \sum_{m=1}^M \mathbf{x} \odot \mathbf{h}_m \sqrt{p_m} s_m + \mathbf{x} \odot \mathbf{n}_r \in \mathcal{C}^N.$$

Typically the relays are geographically distributed and one needs to impose the following individual power constraints on the relays:

$$P_n^R(x_n, \mathbf{p}) := |x_n|^2 \left(\sum_{m=1}^M |h_{m,n}|^2 p_m + \sigma_r^2 \right) \leq P_n^{R,\max}, \quad n = 1, \dots, N, \quad (4.48)$$

where $P_n^{R,\max}$ is the maximum transmitted power of the n th relay. Furthermore, it is also practical and desirable to constraint the total transmitted power of all relays and sources as follows:

$$P_{\text{sum}}(\mathbf{x}, \mathbf{p}) := \sum_{m=1}^M p_m + \sum_{n=1}^N P_n^R(x_n, \mathbf{p}) \leq P_{\text{sum}}^{\max}. \quad (4.49)$$

Let $\boldsymbol{\ell}_i = (\ell_{i,1}, \ell_{i,2}, \dots, \ell_{i,N})^T \in \mathcal{C}^N$, $i = 1, 2, \dots, M$ be the vector of the ‘‘downlink’’ channels between all the relays and the i th destination. Then, the received signal at the i th destination is

$$y_i = \langle \boldsymbol{\ell}_i, \mathbf{r}_b \rangle + n_{d,i} = \sum_m^M \langle \mathbf{c}_{m,i}, \mathbf{x} \rangle \sqrt{p_m} s_m + \langle \boldsymbol{\ell}_i \odot \mathbf{n}_r, \mathbf{x} \rangle + n_{d,i} \in \mathcal{C}, \quad i = 1, \dots, M, \quad (4.50)$$

where $\mathbf{c}_{m,i} = \boldsymbol{\ell}_i \odot \mathbf{h}_m$ presents the effective channel gains from sources to destinations via the relays.

Note that the noise components at all M destinations can be described as $\mathbf{n}_d = (n_{d,1}, \dots, n_{d,M}) \sim \mathcal{CN}(0, \sigma_d^2 \mathbf{I}_M)$. For the multi-user wireless network under consideration, the performance metric of interest is the signal-to-interference-plus-noise ratio (SINR). It follows from (4.50) that the SINR at destination i is given as

$$\mathbf{SINR}_i(\mathbf{x}) := \frac{|\langle \mathbf{x}, \mathbf{c}_{i,i} \rangle|^2 p_i}{\sum_{m=1, m \neq i}^M |\langle \mathbf{x}, \mathbf{c}_{m,i} \rangle|^2 p_m + \mathbf{x}^H \mathbf{L}_i \mathbf{x} + \sigma_d^2}, \quad (4.51)$$

where $\mathbf{L}_i = \sigma_r^2 \text{diag}(|\ell_{i,1}|^2, \dots, |\ell_{i,N}|^2)$.

The aim of this paper is to jointly find the power allocation vectors \mathbf{p} and the beamforming vector \mathbf{x} to maximize the minimum SINR subject to the individual power constraints, which is mathematically formulated as

$$\max_{\mathbf{x} \in \mathcal{C}^N} \min_{\mathbf{p} \in \mathcal{P}_+^M} \min_{i=1,2,\dots,M} \mathbf{SINR}_i(\mathbf{x}) : (4.47), (4.48), (4.49). \quad (4.52)$$

It can be seen that the optimization problem in (4.52) is highly nonlinear and nonconvex and hence it is challenging to find the optimal solution.

4.5.2 Proposed D.C. Formulation and Iterative Algorithm

In this section, we introduce an efficient iterative algorithm to jointly optimize (\mathbf{p}, \mathbf{x}) as stated in problem (4.52). To this end, we first recast the optimization problem (4.52) into a specific class of d.c. programming which minimizes a d.c. objective function over a set of convex constraints.

Define

$$q_m = \frac{1}{p_m^2} \geq 0, \quad m = 1, \dots, M. \quad (4.53)$$

Then, problem (4.52) is rewritten as

$$\max_{\mathbf{x} \in \mathcal{C}^N, \mathbf{q} \in \mathbb{R}_+^M, \boldsymbol{\alpha} \in \mathbb{R}_+^M} \min_{i=1, \dots, M} \varphi_i(\mathbf{x}, q_i, \alpha_i) := \frac{|\langle \mathbf{x}, \mathbf{c}_{i,i} \rangle|^2}{\sqrt{q_i \alpha_i}} \quad \text{s.t.} \quad (4.54a)$$

$$q_m \geq 1/(P_m^{S, \max})^2, \quad m = 1, \dots, M, \quad (4.54b)$$

$$\Pi_n^R(x_n, \mathbf{q}) := \sum_{m=1}^M |h_{m,n}|^2 \frac{|x_n|^2}{\sqrt{q_m}} + \sigma_r^2 |x_n|^2 \leq P_n^{R, \max}, \quad n = 1, \dots, N, \quad (4.54c)$$

$$\Pi_{\text{sum}}(\mathbf{x}, \mathbf{q}) := \sum_{m=1}^M \frac{1}{\sqrt{q_m}} + \sum_{n=1}^N \left(\sum_{m=1}^M |h_{m,n}|^2 \frac{|x_n|^2}{\sqrt{q_m}} + \sigma_r^2 |x_n|^2 \right) \leq P_{\text{sum}}^{\max}, \quad (4.54d)$$

$$\Pi_i(\mathbf{x}, \boldsymbol{\alpha}) := \sum_{m \neq i}^M \frac{|\langle \mathbf{x}, \mathbf{c}_{m,i} \rangle|^2}{\sqrt{q_m}} + \mathbf{x}^H \mathbf{L}_i \mathbf{x} + \sigma_d^2 - \sqrt{\alpha_i} \leq 0, \quad i = 1, \dots, M. \quad (4.54e)$$

To analyze the convexity of the above problem, it is relevant to notice that the function $\varphi(x, y) = x^2/\sqrt{y}$ is convex in its variable $(x, y) \in \mathbb{R}_+^2$. Accordingly, the function $\phi(x, y, z) = x^2/\sqrt{yz}$ is also convex in variables $(x, y, z) \in \mathbb{R}_+^3$ [18]. Therefore, all the functions $\Pi_n^R(\cdot)$, $\Pi_{\text{sum}}^R(\cdot)$, $\Pi_i(\cdot)$ and $\varphi_i(\mathbf{x}, q_i, \alpha_i)$ are convex. Consequently, constraints (4.54b)-(4.54e) are convex. However, the objective function (4.54a) is not convex as it is point-wise minimum of convex functions $\varphi_i(\cdot)$. Nevertheless, its d.c. decomposition is available [102]:

$$\min_{i=1, \dots, M} \varphi_i(\mathbf{x}, q_i, \alpha_i) = f_1(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha}) - f_2(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha})$$

where

$$f_1(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha}) := \sum_{i=1}^M \varphi_i(\mathbf{x}, q_i, \alpha_i), \quad f_2(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha}) := \max_{j=1, \dots, M} \sum_{i=1, i \neq j}^M \varphi_i(\mathbf{x}, q_i, \alpha_i)$$

are both convex as they are a sum and maximum of convex functions, respectively. Thus, problem (4.54) is equivalent to the following d.c. programming [102]:

$$\min_{\mathbf{x} \in \mathcal{C}^N, \mathbf{q} \in \mathbb{R}_+^M, \boldsymbol{\alpha} \in \mathbb{R}_+^M} [f_2(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha}) - f_1(\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha})] \quad \text{s.t.} \quad (4.54b), (4.54c), (4.54e). \quad (4.55)$$

It should be emphasized that, although problem (4.55) is equivalent to problem (4.54), the latter exposes the separate nonconvex term which can be exploited to develop an efficient iterative algorithm.

We now develop an iterative algorithm, called DCI, for the specific problem in (4.55). Initialized from a feasible solution $\mathbf{z}^{(0)} := (\mathbf{x}^{(0)}, \mathbf{q}^{(0)}, \boldsymbol{\alpha}^{(0)})$, we obtain $\mathbf{z}^{(\kappa+1)} := (\mathbf{x}^{(\kappa+1)}, \mathbf{q}^{(\kappa+1)}, \boldsymbol{\alpha}^{(\kappa+1)})$ as the optimal solution of the convex program

$$\begin{aligned} \min_{\mathbf{z} := (\mathbf{x}, \mathbf{q}, \boldsymbol{\alpha}) \in \mathcal{C}^N \times \mathcal{R}_+^M \times \mathcal{R}_+^M} & [f_2(\mathbf{z}) - f_1(\mathbf{z}^{(\kappa)}) - \langle f_1(\mathbf{z}^{(\kappa)}), \mathbf{z} - \mathbf{z}^{(\kappa)} \rangle] \\ \text{s.t.} & (4.54\text{b}), (4.54\text{c}), (4.54\text{d}), (4.54\text{e}), \end{aligned} \quad (4.56)$$

where

$$\begin{aligned} \langle f_1(\mathbf{z}^{(\kappa)}), \mathbf{z} - \mathbf{z}^{(\kappa)} \rangle &= \sum_{i=1}^M \langle \nabla \varphi_i(\mathbf{x}^{(\kappa)}, q_i^{(\kappa)}, \alpha_i^{(\kappa)}), (\mathbf{x}, q_i, \alpha_i) - (\mathbf{x}^{(\kappa)}, q_i^{(\kappa)}, \alpha_i^{(\kappa)}) \rangle \\ &= \sum_{i=1}^M \left[2 \frac{\text{Re}(\langle \mathbf{x}^{(\kappa)}, \mathbf{c}_{i,i} \rangle \cdot \langle \mathbf{c}_{i,i}, \mathbf{x} - \mathbf{x}^{(\kappa)} \rangle)}{\sqrt{q_i^{(\kappa)}, \alpha_i^{(\kappa)}}} - \frac{1}{2} \frac{|\langle \mathbf{x}^{(\kappa)}, \mathbf{c}_{i,i} \rangle|^2 (q_i - q_i^{(\kappa)})}{\sqrt{(q_i^{(\kappa)})^3 \alpha_i^{(\kappa)}}} \right. \\ &\quad \left. - \frac{1}{2} \frac{|\langle \mathbf{x}^{(\kappa)}, \mathbf{c}_{i,i} \rangle|^2 (\alpha_i - \alpha_i^{(\kappa)})}{\sqrt{q_i^{(\kappa)} (\alpha_i^{(\kappa)})^3}} \right]. \end{aligned}$$

The computational complexity of program (4.56) is $\mathcal{O}((N+2M)^3)$. Furthermore, in view of convexity of $f_1(\mathbf{z})$,

$$\begin{aligned} f_2(\mathbf{z}^{(\kappa+1)}) - f_1(\mathbf{z}^{(\kappa+1)}) &\leq f_2(\mathbf{z}^{(\kappa+1)}) - f_1(\mathbf{z}^{(\kappa)}) - \langle \nabla f_1(\mathbf{z}^{(\kappa)}), \mathbf{z}^{(\kappa+1)} - \mathbf{z}^{(\kappa)} \rangle \\ &\leq f_2(\mathbf{z}^{(\kappa)}) - f_1(\mathbf{z}^{(\kappa)}) - \langle \nabla f_1(\mathbf{z}^{(\kappa)}), \mathbf{z}^{(\kappa)} - \mathbf{z}^{(\kappa)} \rangle \\ &= f_2(\mathbf{z}^{(\kappa)}) - f_1(\mathbf{z}^{(\kappa)}), \end{aligned}$$

showing that $\mathbf{z}^{(\kappa+1)}$ is a better solution of (4.56) than $\mathbf{z}^{(\kappa)}$. Therefore, initialized by feasible $\mathbf{z}^{(0)}$, the sequential convex program generates a sequence of $\mathbf{z}^{(\kappa)}$ of improved solutions of (4.56). Since the feasible set of (4.55) is bounded and closed, the sequence $\{\mathbf{z}^{(\kappa)}\}$ is compact and converges to an optimized solution. The proposed iterative algorithm repeatedly solves the convex optimization (4.56) until the objective value converges, i.e., $|\frac{F^{(\kappa+1)} - F^{(\kappa)}}{F^{(\kappa)}}| \leq \varepsilon$, where ε is an error tolerant number and $F^{(\kappa)} = f_2(\mathbf{z}^{(\kappa)}) - f_1(\mathbf{z}^{(\kappa)})$.

It is pointed out that that the convex program (4.56) solves for a total of $M+N$ variables, M for source power variables and N for relay beamforming variables. A simpler joint optimization

problem can be considered when all source power variables are constrained equal to a single scalar variable, namely $p_m = p$, $m = 1, 2, \dots, M$. As a consequence, the number of variables is $N + 1$, regardless of the number of user pairs. Letting $q = 1/p^2$, the optimization problem (4.54) simplifies to the following:

$$\begin{aligned} & \max_{\mathbf{x} \in \mathcal{C}^N, q \geq 0, \boldsymbol{\alpha} \in \mathcal{R}_+^M} \min_{i=1, \dots, M} \varphi_i(\mathbf{x}, q, \boldsymbol{\alpha}_i) \\ \text{s.t. } & (4.54\text{e}), \Pi_n^R(x_n, q) \leq P_n^{R, \max}, n = 1, \dots, N; \Pi_{\text{sum}}(\mathbf{x}, q) \leq P_{\text{sum}}^{\max} \end{aligned} \quad (4.57)$$

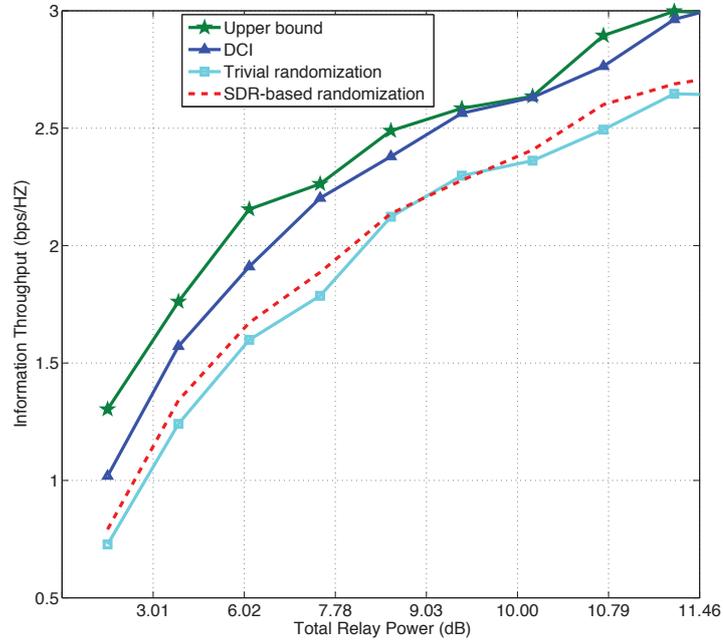


Figure 4.3: Information throughput versus total relaying power for $M = 3$ and $N = 10$ under non-orthogonal source transmissions.

For convenience, we call the above formulation with scalar power variable $p \in \mathcal{R}_+$ the structured formulation, whereas the original problem in (4.54) with vector power variable $p \in \mathcal{R}_+^M$ is called structure-free formulation. Finally, it can be observed that, the computational complexity of the joint optimization is reduced to $\mathcal{O}((N + M)^3)$ because of the inherent flexibility of the joint optimal design.

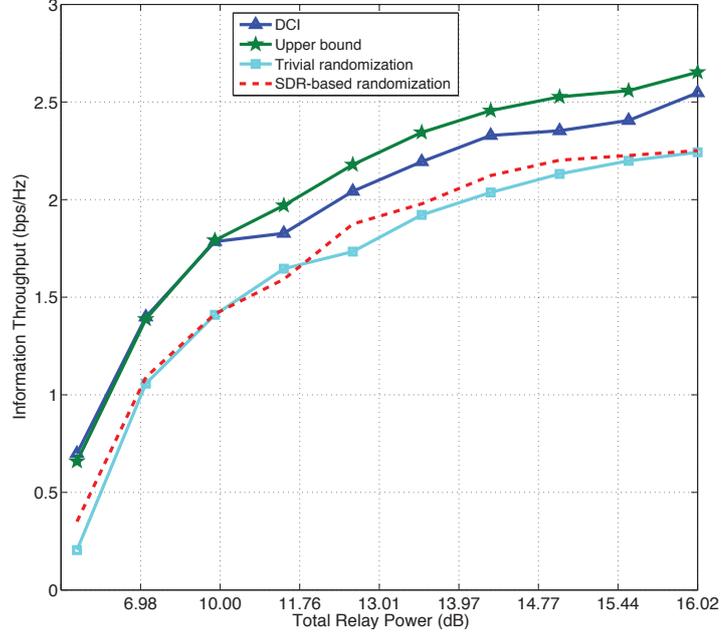


Figure 4.4: Information throughput versus total relaying power for $M = 4$ and $N = 12$ under non-orthogonal source transmissions.

4.6 Numerical Results

In all the simulations carried out in this section, the power of AWGN at relays and destination nodes is normalized to $\sigma_R^2 = \sigma_D^2 = 1$, while the signal power of all the sources is set at $\sigma_s^2 = 100$. Both uplink and downlink channel gains are randomly generated according to a circularly symmetric complex Gaussian distribution (i.e., Rayleigh distribution of their magnitudes). All results are averaged over 100 Monte-Carlo simulation runs to capture the variation of the channel coefficients.

4.6.1 Non-Orthogonal (Concurrent) Source Transmissions

First, consider the case of no channel uncertainty, i.e., $\zeta = 0$ in (4.5). This means that

$$\tilde{\mathbf{h}}_m \equiv \mathbf{h}_m, \tilde{\mathbf{l}}_i \equiv \mathbf{l}_i, \tilde{\mathbf{c}}_{mi} \equiv \mathbf{c}_{mi} \quad (4.58)$$

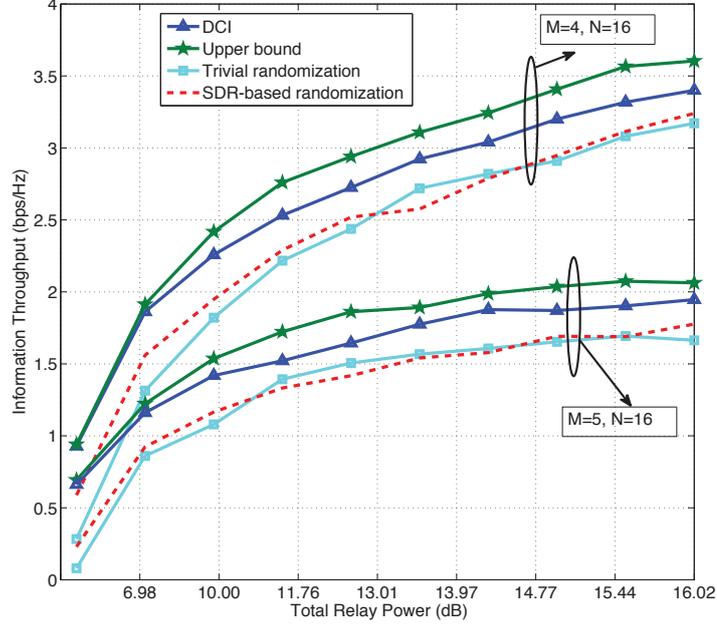


Figure 4.5: Information throughput versus total relaying power for $M = 4$ and $N = 16$ under non-orthogonal source transmissions.

in (4.4). We use the following convex program for any $10\log_{10}(\alpha_i) \geq 1\text{dB}$,

$$\min_{\mathbf{X} \in \mathcal{C}^{N \times N}} \max_{n=1,2,\dots,N} \mathbf{X}(n,n) \quad \text{s.t.} \quad \mathbf{X} \geq 0, \quad \sigma_s^2 \langle \mathbf{C}_{ii}, \mathbf{X} \rangle \geq \alpha_i \left(\sigma_s^2 \sum_{m \neq i}^M \langle \mathbf{C}_{mi}, \mathbf{X} \rangle + \sigma_R^2 \langle \mathbf{L}_i, \mathbf{X} \rangle + \sigma_D^2 \right), \quad i = 1, 2, \dots, M. \quad (4.59)$$

We take as an initial $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ for the DCI the eigenvector $\mathbf{x}^{(0)}$ corresponding to the eigenvalue $\lambda_{\max}(\mathbf{X}^{(0)})$ of the optimal solution $\mathbf{X}^{(0)}$ of (4.61) such that $\|\mathbf{x}^{(0)}\|^2 = \lambda_{\max}(\mathbf{X}^{(0)})$, and accordingly $y_i^{(0)} = \sum_{m \neq i} |\langle \mathbf{x}^{(0)}, \mathbf{c}_{mi} \rangle|^2 + \frac{\sigma_R^2}{\sigma_s^2} |\langle \mathbf{l}_i, \mathbf{x}^{(0)} \rangle|^2$.

Table 5.1 lists the numbers of iterations used, while Figures 4.3, 4.4 and 4.5 present simulation results for different numbers of users and relay nodes. Specifically, plotted in these figures are the minimum information throughput among all users achieved by our proposed DCI approach given by (4.31). Since the DCI method achieves the balanced throughput for all users, this minimum throughput is the actual information throughput for all users as well. The plots of the upper bound for the throughput performance obtained by solving a sequence of relaxed programs (4.18) are also included in these figures (i.e., the SDR curves $\log_2(1 + \alpha_{\text{opt}})$, where α_{opt} is the maximum of

Table 4.1: Average number of iterations for obtaining solutions under non-orthogonal source transmissions.

$M = 3, N = 10$		$M = 4, N = 16$		$M = 5, N = 20$	
P_T (dB)	Iterations	P_T (dB)	Iterations	P_T (dB)	Iterations
0.04	13.57	0.04	21.03	-10	23.71
4.08	13.25	7.27	20.39	7.51	21.88
6.14	12.62	9.85	20.16	10.48	21.20
7.53	12.11	11.46	18.77	12.23	20.25
8.58	12.23	12.63	18.67	13.48	19.08
9.43	11.33	13.55	17.60	14.44	17.99
10.14	10.15	14.31	16.59	15.27	17.07
10.75	9.16	14.96	15.73	15.90	14.83
11.28	7.95	15.52	12.69	16.48	12.72
11.76	6.11	16.02	7.48	16.99	8.84

those α such that the optimal value of SDP (4.18) is not more than one). In each of these figures, the DCI curve is observed to be very close to the SDR curve which establishes the superiority of our proposed method.

It is naturally expected that an increase in throughput would happen with an increased value of the total relay power. In this paper the total relay power budget is equally divided among the relay nodes. We start with a scenario consisting of $M = 3$ users and $N = 10$ relays. It can be observed in Figure 4.3 that with our proposed d.c. programming based approach, a throughput of 2.5 bps/Hz is achieved for each user when the total power budget is 9.03 dB, while the SDR-based randomization approach only gives 1.7 bps/Hz for the same amount of total relay power. The randomization techniques fail to distribute the information throughput fairly among all users.

The effect of the number of relay nodes on throughput performance can be illustrated by comparing results of Figure 4.4 and 4.5 where, as expected, an increase in the number of nodes boosts the corresponding throughput for the same amount of total power.

Table 4.2 provides the averaged rank of the optimal solution \mathbf{X}_{opt} of SDP (4.18) at α_{opt} for

Table 4.2: Average rank i_{opt} of \mathbf{X}_{opt} by SDP relaxation.

$M = 3, N = 10$		$M = 4, N = 16$		$M = 5, N = 20$	
P_T (dB)	Avg. i_{opt}	P_T (dB)	Avg. i_{opt}	P_T (dB)	Avg. i_{opt}
0.04	2.18	0.04	2.55	0.04	2.55
4.08	2.09	7.27	2.52	7.27	2.58
6.14	2.11	9.85	2.47	9.85	2.62
7.53	2.11	11.46	2.50	11.46	2.64
8.58	2.11	12.63	2.52	12.63	2.62
9.43	2.14	13.55	2.65	13.55	2.62
10.14	2.11	14.31	2.46	14.31	2.57
10.75	2.14	14.96	2.49	14.96	2.63
11.28	2.15	15.52	2.54	15.52	2.59
11.76	2.14	16.02	2.52	16.02	2.55

different choices of users and relay nodes. It reveals that vectors $\mathbf{x}^{(v)}$ in (4.19) are generated in 2 or 3 dimensional subspace of 10, 16 and 20-dimensional space. This explains why such SDR based randomization performs poorly as Figures 4.3-4.5 show, even 5,000 such \mathbf{x}^v have been generated for each case. This randomization is only as good as a more trivial randomization, which needs only one SDP solver for a feasible solution \mathbf{X} of SDP (4.18) at some α to generate $\mathbf{x}^{(v)}$ according to (4.19) with i_{opt} replaced by the rank of \mathbf{X} and unitary \mathbf{U} and diagonal $\mathbf{\Sigma}$ in SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$. Figures 4.3-4.5 show that the performance of this simple randomization is still poor although it is comparable to that of the more computationally-intensive SDR based randomization (which requires tens of SDP solvers for solution of SDP (4.18) in the bisection procedure to locate the optimal α_{opt}).

Next, Figure 4.6 presents throughput for the scenario that $\zeta \neq 0$ in (4.5), whose value can be set to represent varying degrees of channel uncertainties. For each user, throughput curves are plotted versus the total relay power for $\zeta = 0.1, 0.3, 0.6, 0.8$. Except for the fifth user in $\zeta = 0.3$ case, which is assigned a slightly higher throughput, all users achieve approximately the same throughput when beamforming is performed according to our proposed DCI method.

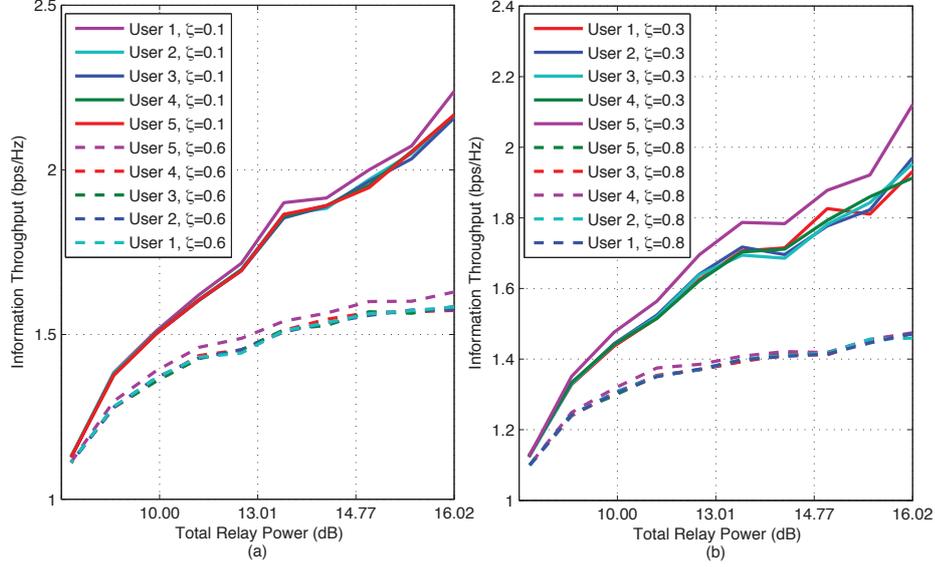


Figure 4.6: Effect of channel uncertainty on information throughput for $M = 5, N = 20$ under non-orthogonal source transmissions.

4.6.2 Orthogonal Source Transmissions

Again, consider first the case of no channel uncertainty case, i.e., $\zeta = 0$ in (4.5) and (4.58) is realized as in (4.4). With the bisection procedure in α , one can find the maximum α_{opt} such that the optimal value of the following SDP is less than one:

$$\begin{aligned}
 & \min_{\mathbf{X}_m \in \mathcal{C}^{N \times N}, m=1,2,\dots,M} \max_{n=1,2,\dots,N} \left[\frac{1}{\gamma_n} \sum_{m=1}^M (\sigma_s^2 |h_{mn}|^2 + \zeta^2 + \sigma_R^2) \mathbf{X}_m(n, n) \right] : \\
 & \sigma_s^2 \langle \mathbf{C}_{ii} - \zeta^2 \mathbf{I}_N, \mathbf{X}_i \rangle \geq \alpha \left[\sigma_s^2 \sum_{m \neq i}^M \langle \mathbf{C}_{mi} + \zeta^2 \mathbf{I}_N, \mathbf{X}_m \rangle + \sigma_R^2 \sum_{m=1}^M \langle \mathbf{L}_i + \zeta^2 \mathbf{I}_N, \mathbf{X}_m \rangle + \sigma_D^2 \right], \quad (4.60) \\
 & \mathbf{X}_i \geq 0, \quad i = 1, 2, \dots, M
 \end{aligned}$$

where $\mathbf{C}_{mi} = \mathbf{c}_{mi} \mathbf{c}_{mi}^H$ and $\mathbf{L}_i = \mathbf{l}_i \mathbf{l}_i^H$. However, in contrast to the SDP (4.18), the SDP (4.60) always admits the optimal rank-one solution $\mathbf{X}_{\text{opt}}^{(i)} = \mathbf{x}_{\text{opt}}^{(i)} \mathbf{x}_{\text{opt}}^{(i)H}$, $i = 1, 2, \dots, M$. This means that $\log_2(1 + \alpha_{\text{opt}})$ is actually the global solution of the maximin program (4.39) and $\mathbf{x}_{\text{opt}}^{(i)}$, $i = 1, 2, \dots, M$ form the optimal solution. This also means that SDR is able to provide the optimal solution for the maximin program (4.39). Table 4.3 provides the iteration numbers of DCI and also of the above mentioned SDR bisection for finding the solution of the nonconvex program (4.39). These num-

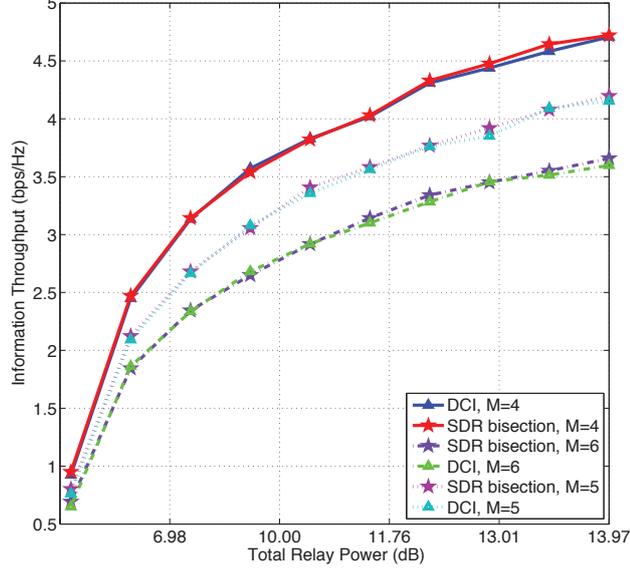


Figure 4.7: Information throughput versus total relaying power for orthogonal source transmissions.

bers correspond to the numbers of required SDP (4.44) to implement the DCI and the SDR-based bisection procedure, respectively. One can see that the required number of SDP (4.44) is always less than that of the required number of SDP (4.60). Moreover, the variable dimension of (4.44) is $(N+1)M$ while that of (4.60) is $MN(N+1)/2$. This clearly indicates that the DCI approach is much more computationally efficient than the SDR-based bisection approach. We solve (4.60) for some arbitrary value of α (e.g., $\alpha \geq 0.01$) to obtain $(\mathbf{X}_1^{(0)}, \dots, \mathbf{X}_M^{(0)})$, whose eigenvectors $\mathbf{x}_m^{(0)}$ corresponding to eigenvalues $\lambda_{\max}(\mathbf{X}_m^{(0)})$ constitute the initial $(\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_M^{(0)}, \mathbf{y}^{(0)})$ of DCI. The simulation results presented in Figure 4.7 illustrate throughput performance of the DCI algorithm. We consider typical scenarios for orthogonal transmission with $M = 4, 5, 6$ while the number of relay nodes $N = 10$ is fixed. The rank-one solution obtained by the aforementioned SDR (4.60) is also plotted to establish an upper bound. It is noted that the DCI method achieves approximately the same throughput curve as that of the SDR bisection method despite the lower dimensions of the DCI variables.

Finally, Figure 4.8 illustrates the impact of channel state uncertainty under the scenario of orthogonal source transmissions. Plotted in the figure is the minimum information throughput

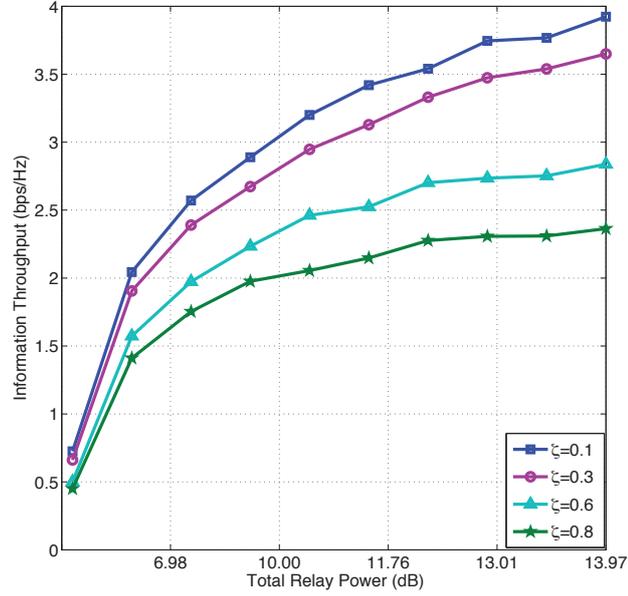


Figure 4.8: Effect of uplink and downlink channel uncertainties on the information throughput for $M = 5, N = 10$ under orthogonal source transmissions.

for different values of ζ , accounting for different amounts of channel uncertainty. As can be observed from the figure, an increase in channel uncertainty causes a reduction in the information throughput, and that the performance gaps among the curves increase for higher relay powers. It is important to note that the DCI method is able to maintain a balanced throughput (or equivalently SINR) for all users, even when the exact knowledge of channel state information is not available.

4.6.3 Joint Optimization

The iterative algorithm developed for joint optimization of source powers and beamforming vector is numerically evaluated in this section. In order to quantify the quality of service, we use metric of information throughput (bps/Hz) achievable for every source-destination pair, which is directly related to the SINR metric. Performance comparison is made among the following designs: (i) Optimal beamforming vector with fixed equal source powers [81]; (ii) Optimal source power allocation with known optimal beamforming vector; and (iii) Our proposed joint optimization of source power allocation and beamforming vector. Note that (i) employs d.c. iteration method to converge to the local optimal solution [81], while (ii) makes use of the bisection method to solve

Table 4.3: Average number of iterations for obtaining solutions under orthogonal source transmissions.

$M = 4, N = 10$			$M = 5, N = 10$			$M = 6, N = 10$		
P_T (dB)	DCI	SDR bisection	P_T (dB)	DCI	SDR bisection	P_T (dB)	DCI	SDR bisection
-3.01	11.43	13.19	-3.01	12.64	13.30	-3.01	8.72	13.21
5.05	11.41	13.19	5.05	12.82	13.29	5.05	7.60	13.22
7.74	11.27	13.24	7.74	12.59	13.23	7.74	5.37	13.29
9.38	11.39	13.28	9.38	12.70	13.23	9.38	4.31	13.25
10.56	11.11	13.25	10.56	12.63	13.18	10.56	3.23	13.22
11.50	10.89	13.22	11.50	12.06	13.22	11.50	2.13	13.22
12.26	10.34	13.18	12.26	11.38	13.29	12.26	1.25	13.19
12.91	9.53	13.22	12.91	10.32	13.31	12.91	1.26	13.18
13.48	7.63	13.29	13.48	7.85	13.33	13.48	2.28	13.29
13.98	2.24	13.21	13.98	2.89	13.25	13.98	8.84	13.33

the linear program in source power allocation.

The individual source power is set as $p_m \leq \nu p_0$ where $\nu = 1.5$ is a scaling factor, and $p_0 = 100$ is the equal source power used in [81]. On the other hand, the individual relay power is limited to $P_n^{R,\max} = \gamma_n = P_{\text{rel}}^{\text{tot}}/N$. Thus, the total sum power budget is upper bounded by $P_{\text{sum}}^{\max} = (Mp_0 + \sum_{n=1}^N \gamma_n)$.

In Fig. 4.9, the information throughput achieved for an arbitrary user pair (since other user pairs also have the same throughput because of the balanced optimal results) is plotted against the total relay power for a 3-user network with 10 relays. The results clearly shows that the performance of the proposed joint optimization scheme is much better than the other two alternative optimization methods in which the beamforming vector and source power allocation are optimized separately. Specifically, it can be seen that there is a capacity gain of around 0.5 bps/Hz achieved by our proposed method over the optimized source power allocation scheme at $P_{\text{rel}}^{\text{tot}} = 11.76$ dB. Of particular interest is the observation that for the structured optimization in (4.57), the sub-

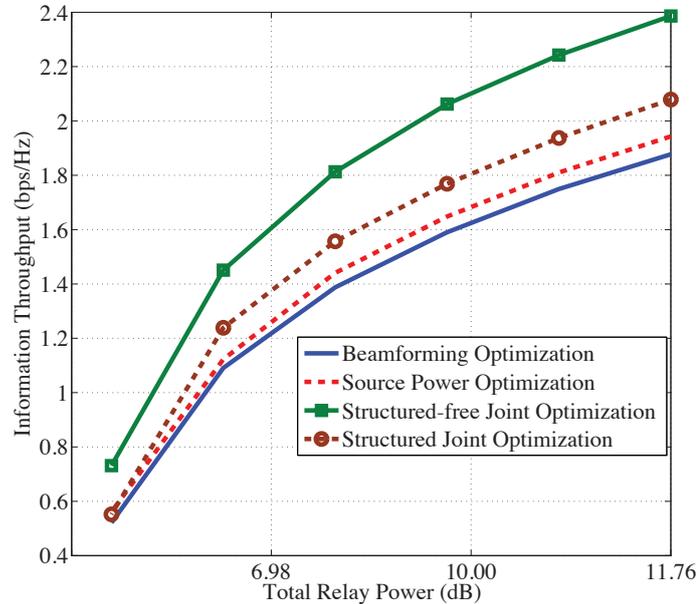


Figure 4.9: Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 3$ and $N = 10$.

optimal solution, even with a reduced number of variables, still offers a much better throughput than that obtained with the optimal source power allocation and fixed optimized beamforming. In other words, the joint optimization over only one scalar power variable and beamforming vector improves the spectral efficiency of multi-user relay network. On the other hand, as shown in Figures 4.9-4.12, when the source power is optimized separately while using a fixed optimized beamforming, the throughput performance improves very modestly.

Fig. 4.10 illustrates the throughput performance when $N = 12$ relays are used to assist $M = 4$ users. While there is a decline in the throughput value as compared to the network with 3 users, the proposed joint optimization method still offers a better throughput over the entire range of the total relay power. Upon increasing the number of relays to $N = 16$ for the same number of users, we can observe an increase in the information throughput as shown in Fig. 4.11. To evaluate the impact of adding an additional user into this network, the throughput plot is shown in Fig. 4.12 for $M = 5$ users. Again, the proposed joint optimization method also outperforms the alternative optimization

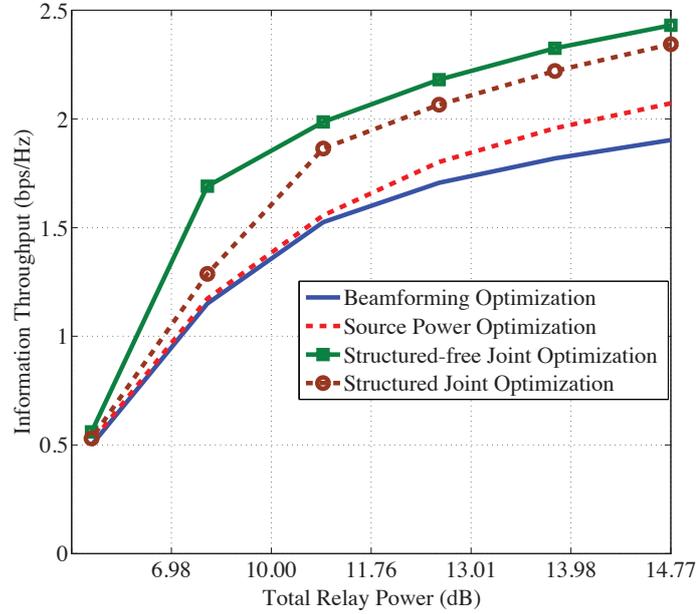


Figure 4.10: Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 4$ and $N = 12$.

techniques for such a network scenario. Moreover, as the total relay power increases, the capacity gain of the proposed algorithm over the alternative optimizations also increases. The distribution of the source powers determined by the proposed method for a given channel realization is plotted in Fig. 4.13 for $M = 4$ and $N = 12$. The fact that the source powers are vastly different for different sources signifies the importance of our approach of jointly optimizing source power and beamforming vector compared to the equal power setting used in [81]. It also shows that power control among the sources is adaptive to the corresponding channels.

One can easily find initial feasible solution of (4.56) by solving convex feasibility problem defined by the set of convex constraints (4.54b)-(4.54e). It should be emphasized however that successful implementation of DCI demands a good initial feasible point to guarantee convergence towards right optimized solution. Such an initial point is obtained by solving the following convex

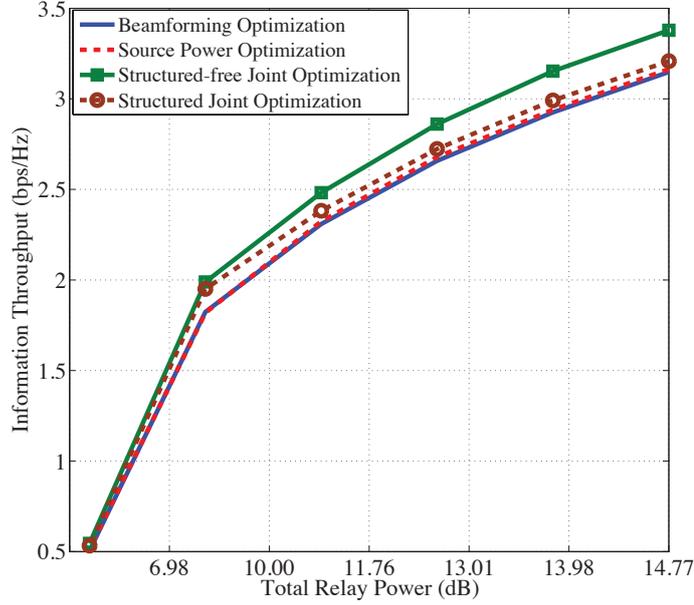


Figure 4.11: Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 4$ and $N = 16$.

program

$$\begin{aligned}
& \min_{\mathbf{X} \in \mathcal{C}^{N \times N}} \max_{n=1,2,\dots,N} \mathbf{X}(n,n) \\
& \text{s.t. } \mathbf{X} \geq 0, \\
& p_i^{(0)} \langle \mathbf{C}_{i,i}, \mathbf{X} \rangle \geq \alpha_i \left(\sum_{m \neq i}^M p_m^{(0)} \langle \mathbf{C}_{m,i}, \mathbf{X} \rangle + \sigma_R^2 \langle \tilde{\mathbf{L}}_i, \mathbf{X} \rangle + \sigma_D^2 \right), \quad i = 1, 2, \dots, M, \\
& \left(\sum_{m=1}^M p_m^{(0)} |h_{m,n}|^2 + \sigma_r^2 \right) \mathbf{X}(n,n) \leq \gamma_n, \quad n = 1, 2, \dots, N, \\
& \sum_{m=1}^M p_m^{(0)} + \sum_{n=1}^N \left(\sum_{m=1}^M p_m^{(0)} |h_{m,n}|^2 + \sigma_r^2 \right) \mathbf{X}(n,n) \leq P_{\text{sum}}^{\max}
\end{aligned} \tag{4.61}$$

for a small value of $\alpha \geq 0$, with $\mathbf{C}_{m,i} = \mathbf{c}_{m,i} \mathbf{c}_{m,i}^H$, $\tilde{\mathbf{L}}_i = \boldsymbol{\ell}_i \boldsymbol{\ell}_i^H$ and $p_i^{(0)} \equiv p_0 = 100$. Suppose $\mathbf{x}^{(0)}$ is the eigenvector corresponding to the maximal eigenvalue $\lambda_{\max}(\mathbf{X}^{(0)})$ of the optimal solution $\mathbf{X}^{(0)}$ of (4.61) such that $\|\mathbf{x}^{(0)}\|^2 = \lambda_{\max}(\mathbf{X}^{(0)})$ and $p^{(0)} = (p_0, \dots, p_0)^T \in \mathcal{R}^M$, so that the individual source power constraint (4.47) is also satisfied. Defining $\mathbf{q}^{(0)} = (1/p_0^2, \dots, 1/p_0^2)^T \in \mathcal{R}^M$ and

$$\alpha_i^{(0)} = \left(\sum_{m=1, m \neq i}^M |\langle \mathbf{x}^{(0)}, \mathbf{c}_{m,i} \rangle|^2 p_0 + \mathbf{x}^{(0)H} \mathbf{L}_i \mathbf{x}^{(0)} + \sigma_d^2 \right)^2, \quad i = 1, 2, \dots, M,$$

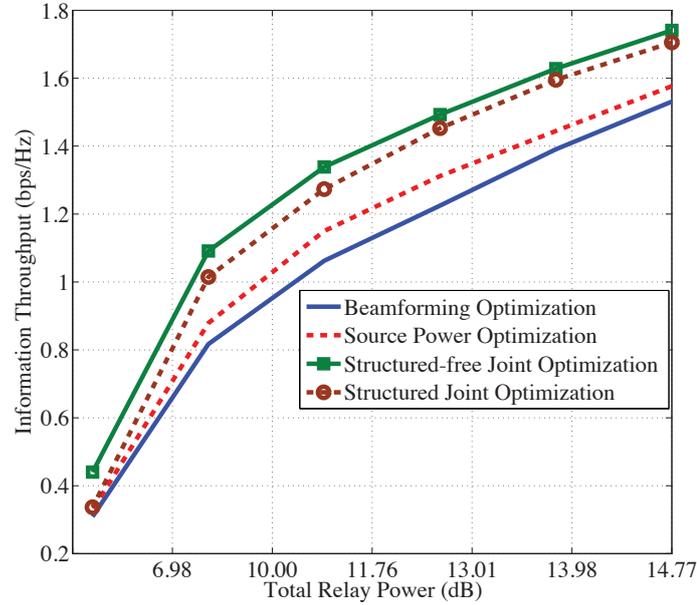


Figure 4.12: Performance comparison of proposed joint optimization methods and separate optimizations of source powers and beamforming vector, $M = 5$ and $N = 16$.

the initial feasible solution of (4.55) is taken as $\mathbf{z}^{(0)} = (\mathbf{x}^{(0)}, \mathbf{q}^{(0)}, \boldsymbol{\alpha}^{(0)})$. Also the tolerance $\varepsilon = 0.01$ is set. The convergence performance of the proposed DCI is illustrated in Table 5.1, which confirms that optimized solutions are obtained within a few iterations.

4.7 Summary

This chapter has solved the beamforming design problems in multi-user wireless relay networks to maximize the minimum information throughput among all users. Both cases of concurrent and orthogonal transmissions from sources to relays are considered. Different from the existing approach which reformulates the design problems to matrix rank-one constrained optimizations, our approach exploits the d.c. structure of the objective function and the convex structure of the constraints to develop efficient iterative algorithms of very low complexity to find the solutions. Numerical results demonstrate that the developed algorithms are able to locate the global optimal solutions by a few iterations and they illustrate the superiority of our method over the existing

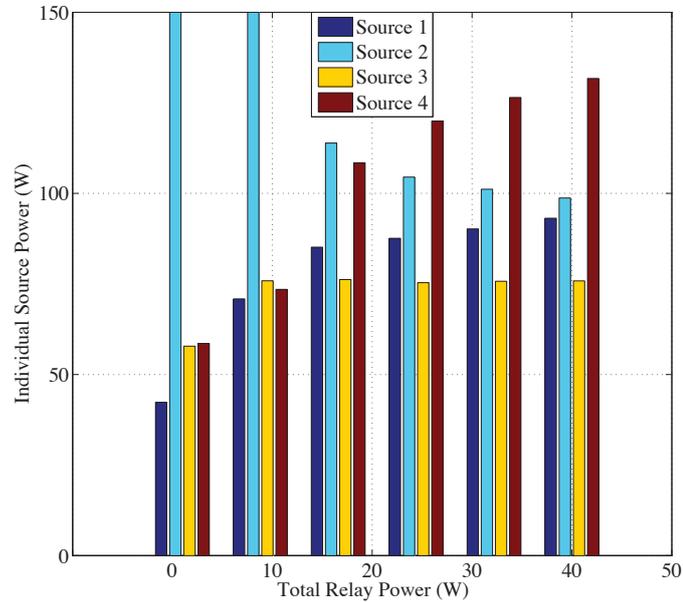


Figure 4.13: Distribution of source powers for a specific channel realization in a network with $M = 4$ and $N = 12$.

methods.

In the next part, joint design problem of source power allocation and relay beamforming in multi-user multi-relay wireless networks has been addressed. Different from existing approaches which use a fixed and equal power at all the sources, the proposed method simultaneously optimizes the source powers and relay beamforming weights. To handle the difficulty of the non-convexity in the design problem, we first cast the design problem into a specific class of d.c. programming and develop an iterative algorithm to solve it. Furthermore, a simplified joint optimization problem is also suggested whose solution exhibits good performance and low computational complexity. The simulation results indicate that our joint optimization methods outperform the separate optimizations of either the source power allocation or the relay beamforming vector.

Table 4.4: Average number of iterations for convergence.

$M = 3, N = 10$		$M = 4, N = 12$		$M = 5, N = 20$	
$P_{\text{rel}}^{\text{tot}}$ (dB)	Iterations	$P_{\text{rel}}^{\text{tot}}$ (dB)	Iterations	$P_{\text{rel}}^{\text{tot}}$ (dB)	Iterations
0	5.20	0	6.60	0	4.30
5.98	3.00	9.44	4.35	9.44	9.20
8.20	5.00	12.20	3.40	12.20	9.80
9.73	4.80	13.87	5.10	13.87	13.75
10.86	2.75	15.07	5.40	15.07	20.25
11.76	2.85	16.02	5.60	16.02	10.00

Chapter 5

Joint Source and Relay Precoding Design in Wireless MIMO Relay Networks

Next generation wireless systems seek to extend the information throughput capacities of the existing cellular systems without consuming additional spectral resources. The thirst for greater data rates exhibited by users of mobile wireless services has been on an exponential trajectory. Additionally, the current multiple-input multiple-output (MIMO) antenna techniques in the uplink and downlink is a well-known technology to enhance the spectral efficiency and link reliability [31]. Deployment of multiple antennas at wireless stations offer increase in channel capacity through spatial multiplexing as well as spatial modulation and coding [93].

The outline of this chapter is organized as follows. A brief introduction and motivation of the problem is given in Section 5.1. Section 4.2 describes the system model for MIMO broadcast downlink cellular communication through MIMO relay and formulates joint optimization problem for source and relay precoding design. Section 5.3 presents a formulation of the optimization problem for two-way MIMO relay-assisted communication between multi-antenna sources as minimization of the worst mean-squared error and sum-throughput maximizations. Simulation results that support the effectiveness of the proposed algorithm are presented in Section 5.4.

Conclusions are drawn in Section 5.5.

5.1 Introduction

Recent release of long-term evolution-advanced (LTE-A) has adopted the idea of relaying techniques for wireless transmissions to further increase the data throughput and coverage range of wireless networks [68]. It is known that the direct wireless link between a source and destination fails to provide the desired data throughput in the presence of channel impairments such as path-loss, shadowing and small-scale fading. Under such circumstances, wireless relay nodes can be deployed to assist the communication between source and destination, especially when there are multiple users at the destination side. Fixed relays are low cost and low transmit power elements that receive and forward data from the base station to the users via wireless channels, and vice versa [66]. Conventionally, the use of single-input single-output (SISO) relays has been focused in most of the multi-user communication systems [30, 47, 61, 74, 81]. However, recently much attention has been paid on investigating MIMO relay-assisted multi-user systems [3, 22, 64, 73, 107]. This is because using MIMO relay node can enhance transmission quality of a multi-user system more efficiently than using multiple SISO relay nodes. In general, relay can be either regenerative (e.g., decode-and-forward) or non-regenerative (e.g., amplify-and-forward). The amplify-and-forward (AF) protocol is of practical interest due to its ease of implementation. For MIMO relays, the AF protocol can essentially be extended to filter-and-forward (FF) protocol in which the MIMO relay transforms the input vector through its precoding matrix and then forwards it to the destination [85].

The MIMO broadcast channel, where a BS communicates with multiple users with multiple antennas, has been studied in terms of information theory [19, 87]. With MIMO relay station interplay between the BS and the single antenna users, increasing attention has been given to the joint design of source and relay precoding matrices to attain high throughput performances [111, 113]. Particularly, [113] was dealt with the total transmit power minimization under certain quality-of-service (QoS) requirements expressed by signal-to-interference-ratio (SINR) constraints. Its solutions are based on alternating optimization between source and relay precoding matrices, which is computationally cumbersome. Beside a well-known premature termination inherent in any al-

ternating optimization, even optimization routine in the relay precoding matrix with source precoding matrix held fixed remains highly nonconvex and thus computationally intractable. Furthermore, [111] considered the sum-rate maximization, which is maximization of SINR function under transmit power constraints. The alternating optimization is hardly applied as any optimization routine in one matrix variable with another one held fixed is still intractable. As a matter of fact, the iterations proposed in [111] are just classical smooth function local approximation that does not perform better than the advanced sequential quadratic programming code `fmincon` in standard Matlab [58]. Moreover, the sum-rate is not quite appropriate target in multiuser communication scenario as it leads to unbalanced rate distributions for the users. For fair QoS, it is more natural to maximize the lowest rate among the users, especially the users which are located at the edge of a cell. The resultant nonconvex objective function even becomes nonsmooth so all the aforementioned tools of smooth optimization are not applicable.

Meantime, two-way relay networks [79] have attracted considerable attention in research community [2, 54, 112, 114] due their higher spectral efficiency than the conventional one-way relay networks. By two-way relaying, the overall network throughput rate can be increased by conducting information exchange in two-time slots as opposed to the four time slots used in case of one-way relay system [24, 54]. Obviously, the two-way relaying scheme suffers even more severely from interference arising from concurrent transmission from both communication end compared with one-way relaying. However, with "source" precoding matrix now including that at the both ends of communication, all the design problems in two-way relaying are still mathematically similar to that for one-way relaying. Similar challenges and unsolved problems in one-way relaying are easily observed in joint design of users and relay precoding matrices in two-way MIMO relaying [51, 84, 106]. Particularly, [84, 106] considered sum-MMSE (minimum mean-square error) minimization under the transmission power constraint by alternating optimization. As mentioned above, as the optimization routine in relay precoding matrix with source precoding matrices held fixed remains highly nonconvex, even one more artificial vector variable is introduced to introduced for alternating optimization of this optimization routine. As the problems of either sum-rate maximization or sum-MMSE under transmission power constraints can be trivially by unconstrained optimization of highly nonlinear function, [51] simply applied a standard

gradient algorithm for its solution.

In summary, the joint designs of source and relay precoding in FF protocol are very complex nonconvex programs of matrix variables with no known efficient computational solutions. Actually, they are polynomial fractional programs, which are among the hardest classes of nonconvex optimization [103]. Nevertheless, following our approach [72, 75, 98] we will show in this paper that these programs can be very efficiently treated by d.c. (difference of two convex functions/sets) programming [102], where the partial convexity structures are recognized for locating the optimized solutions. Firstly, these programs are efficiently transferred to minimization of d.c. function under d.c. constraints. They are still not tractable enough for computation purpose and so the nonconvex duality with zero gap [104] is applied to equivalently transfer them to minimization of d.c. functions under convex constraints, for which d.c. iteration (DCI) is ready to apply.

5.2 Joint Optimization Formulation for Source and Relay Precoding

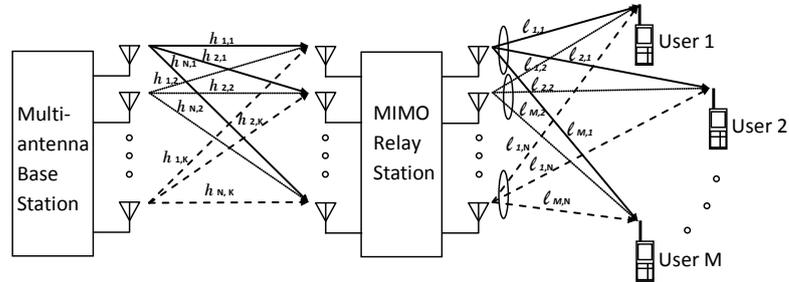


Figure 5.1: A multi-user amplify-and-forward MIMO wireless relay network for downlink cellular communication.

Consider a scenario of multi-user downlink cellular communication with a base station (BS), equipped with K antennas, serving M single-antenna users through a N -antenna relay station (RS) [111, 113]. The transmission is conducted in two independent time-slots. In the first time slot, the BS transmits M data streams to the relay station which processes the received signal by mixing them with a linear transformation and then simply forward these processed signals to all destinations in the second time slot.

Let $\mathbf{s} = (s_1, s_2, \dots, s_M)^T \in \mathcal{C}^M$ present transmit symbols from the sources, with $\mathbf{s} \sim \mathcal{CN}(0, \mathbf{I}_M)$.

Before transmitting \mathbf{s} to the RS, the BS applies precoding transformation $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_M] \in \mathcal{C}^{K \times M}$ to \mathbf{s} . Since \mathbf{s} is assumed to have unit power, the transmit power of the signal at BS is given by :

$$P_B(\mathbf{W}) := \mathbb{E}[\mathbf{W}\mathbf{s}\mathbf{s}^H\mathbf{W}^H] = \langle \mathbf{W}\mathbf{W}^H \rangle \quad (5.1)$$

The transmission from the BS to the RS can be treated as a typical MIMO communication for which the received signal at RS is expressed as :

$$\mathbf{y}_r = \mathbf{H}\mathbf{W}\mathbf{s} + \mathbf{n}_r \quad (5.2)$$

where $\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_K) \in \mathcal{C}^{N \times K}$, with $\mathbf{h}_i = (h_{i1}, \dots, h_{iN})^T \in \mathcal{C}^N, i = 1, \dots, K$, represents MIMO complex channel gain from the i^{th} BS antenna to the RS, and $\mathbf{n}_r = (n_{r,1}, \dots, n_{r,N})^T \in \mathcal{C}^N$ is complex noise vector at RS which is assumed $\mathbf{n}_r \sim \mathcal{C}\mathcal{N}(0, \sigma_r^2 \mathbf{I}_N)$. Suppose $\mathbf{F} \in \mathcal{C}^{N \times N}$ is the beamforming precoding matrix used to pre-process the relay received signal. The RS hence sends the following signals to the destinations:

$$\mathbf{y}_{amp} = \mathbf{F}\mathbf{y}_r = \mathbf{X}\mathbf{s} + \mathbf{F}\mathbf{n}_r$$

for

$$\mathbf{X} := [\mathbf{x}_1 \ \dots \ \mathbf{x}_M] = \mathbf{F}\mathbf{H}\mathbf{W} = [\mathbf{F}\mathbf{H}\mathbf{w}_1 \ \dots \ \mathbf{F}\mathbf{H}\mathbf{w}_M] \quad (5.3)$$

Power of the transmitted signal at the RS is expressed as:

$$P_R(\mathbf{W}, \mathbf{F}) := \mathbb{E}[\mathbf{y}_{amp}\mathbf{y}_{amp}^H] = \langle \mathbf{X}\mathbf{X}^H \rangle + \sigma_r^2 \langle \mathbf{I}_F \mathbf{F}^H \rangle.$$

The transmission from the RS to all users can be modelled as typical MISO broadcasting. Namely, the received signal at destination i is

$$y_{d,i} = \mathbf{l}_i \mathbf{x}_i s_i + \sum_{j \neq i}^M \mathbf{l}_j \mathbf{x}_j s_j + \mathbf{l}_i \mathbf{F} \mathbf{n}_r + n_{d,i}, \quad i = 1, \dots, M \quad (5.4)$$

where the complex noise at destination i is $n_{d,i} \sim \mathcal{C}\mathcal{N}(0, \sigma_d^2)$ and the channel vector between the RS and destination i is $\mathbf{l}_i = (l_{i1}, \dots, l_{iN}) \in \mathcal{C}^{1 \times N}, i = 1, \dots, M$. As only the signal component $\mathbf{l}_i \mathbf{F} \mathbf{H} \mathbf{w}_i s_i$ in (5.4) is of interest, the other components are treated as noise and interference. Hence, the signal-to-interference-plus-noise ratio (SINR) at destination i is given as:

$$\text{SINR}_i(\mathbf{W}, \mathbf{F}) = \frac{|\mathbf{l}_i \mathbf{x}_i|^2}{\sum_{j \neq i}^M |\mathbf{l}_i \mathbf{x}_j|^2 + \sigma_r^2 \|\mathbf{l}_i \mathbf{F}\|^2 + \sigma_d^2} \quad (5.5)$$

Since the information rate at destination i is defined by $R_i(\mathbf{W}, \mathbf{F}) = \log_2(1 + \text{SINR}_i(\mathbf{W}, \mathbf{F}))$, the rate maximin optimization $\max_{\mathbf{W}, \mathbf{F}} \min_{i=1,2,\dots,M} R_i(\mathbf{W}, \mathbf{F})$ is equivalent to the following SINR maximin optimization

$$\max_{\mathbf{W} \in \mathcal{C}^{K \times M}, \mathbf{F} \in \mathcal{C}^{N \times N}, \mathbf{X} \in \mathcal{C}^{N \times M}} \min_{i=1,2,\dots,M} \frac{|\mathbf{l}_i \mathbf{x}_i|^2}{\sum_{j \neq i}^M |\mathbf{l}_i \mathbf{x}_j|^2 + \sigma_r^2 \|\mathbf{l}_i \mathbf{F}\|^2 + \sigma_d^2} \quad \text{s.t. (5.3), (5.6a)}$$

$$[\mathbf{W}\mathbf{W}^H]_{k,k} \leq \nu p_0, \quad k = 1, \dots, K \quad (5.6b)$$

$$[\mathbf{X}\mathbf{X}^H + \sigma_r^2 \mathbf{F}\mathbf{F}^H]_{n,n} \leq \gamma_n, \quad n = 1, \dots, N \quad (5.6c)$$

$$\langle \mathbf{W}\mathbf{W}^H \rangle + \langle \mathbf{X}\mathbf{X}^H + \sigma_r^2 \mathbf{F}\mathbf{F}^H \rangle \leq P_{sum}^{max}. \quad (5.6d)$$

It can be seen that the objective function in (5.6) is highly nonlinear and nonconvex accompanied by the nonlinear equality constraint (5.3). This renders the difficulty to obtain the optimal solution. Furthermore, the problem is still nonconvex even if one fixes one matrix e.g. \mathbf{W} and solve for the relay matrix \mathbf{F} , and vice versa, i.e. alternating optimization is not applicable at all.

[113] considered the total transmission power minimization under SINR constraints

$$\min_{\mathbf{W}, \mathbf{F}} P_B(\mathbf{W}, \mathbf{F}) + \beta \cdot P_R(\mathbf{W}, \mathbf{F}) \quad \text{s.t.} \quad \text{SINR}_i(\mathbf{F}, \mathbf{W}) \geq \gamma_i, \quad i = 1, 2, \dots, M \quad (5.7)$$

for given $\beta > 0$ and $\gamma_i > 0$, $i = 1, 2, \dots, M$. Alternating optimization has been used in [113] to address (5.7), which optimize alternatively on only each of variables \mathbf{W} and \mathbf{F} with the another variable held fixed. While the objective function in (5.7) is obviously separably convex in \mathbf{W} and \mathbf{F} , by a result of [109] the SINR constraints in (5.7) can be equivalently to (convex) second-order cone constraint only when \mathbf{F} is held fixed. This means only the optimization routine in \mathbf{W} is convex and thus tractable. The optimization routine in \mathbf{F} remains difficult and could be efficiently addressed only very recently [73]. On the other hand [111] attempted to address the sum-rate maximization

$$\max_{\mathbf{W}, \mathbf{F}} \sum_{i=1}^M \ln(1 + \text{SINR}_i(\mathbf{W}, \mathbf{F})) : (5.6b), (5.6c) \quad (5.8)$$

by a rough local approximation of each function $\ln(1 + \text{SINR}_i(\mathbf{W}, \mathbf{F}))$ and $P_s(\mathbf{W}, \mathbf{F})$ ($s \in \{B, R\}$) by quadratic concave and affine functions, respectively. Particularly, $\text{SINR}_i(\mathbf{W}, \mathbf{F})$ is firstly linearised by an affine function $\mathcal{A}_i(\mathbf{W}, \mathbf{F})$ of the first order Taylor expansion and then $\ln(1 + \text{SINR}_i(\mathbf{W}, \mathbf{F}))$ is approximated by the second-order Taylor expansion for function $\ln(1 + \mathcal{A}_i(\mathbf{W}, \mathbf{F}))$. The authors

of [111] do not seem to know that function $\ln(1 + \mathcal{A}_i(\mathbf{W}, \mathbf{F}))$ is already concave so its maximization is already tractable and efficiently solved by a standard convex program software. In other words, approximation for $\ln(1 + \mathcal{A}_i(\mathbf{W}, \mathbf{F}))$ is not necessary and only results in deterioration of the iterative performance. In any case, such approach cannot perform better than advanced sequential quadratic programming approach coded by Matlab `fmincon` [58].

We now develop an efficient approach for solution of nonsmooth and nonconvex program (5.6) which is also easily workable for solution of nonconvex but smooth program (5.8) as well. Following [73], to rewrite (5.6) equivalently as

$$\max_{\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y} \in \mathcal{R}_+^M} \min_{i=1,2,\dots,M} \frac{|\mathbf{l}_i \mathbf{x}_i|^2}{y_i + \sigma_d^2} \quad \text{s.t.} \quad (5.6b), (5.3), (5.6c), (5.6d), \quad (5.9a)$$

$$\sum_{j \neq i}^M |\mathbf{l}_j \mathbf{x}_j|^2 + \sigma_r^2 \|\mathbf{l}_i \mathbf{F}\|^2 \leq y_i, \quad i = 1, 2, \dots, M, \quad (5.9b)$$

or equivalently,

$$\min_{\mathbf{W} \in \mathcal{C}^{K \times M}, \mathbf{F} \in \mathcal{C}^{N \times N}, \mathbf{X}, \mathbf{y}} [f_{01}(\mathbf{X}, \mathbf{y}) - f_{02}(\mathbf{X}, \mathbf{y})] : (5.6b), (5.3), (5.9b), (5.6c) \quad (5.10)$$

with

$$f_{01}(\mathbf{X}, \mathbf{y}) := \max_{i=1,2,\dots,M} \sum_{j \neq i} \frac{|\mathbf{l}_j \mathbf{x}_j|^2}{y_j + \sigma_d^2}, \quad f_{02}(\mathbf{X}, \mathbf{y}) := \sum_{i=1}^M \frac{|\mathbf{l}_i \mathbf{x}_i|^2}{y_i + \sigma_d^2} \quad (5.11)$$

Since each function $|\mathbf{l}_j \mathbf{x}_j|^2 / (y_j + \sigma_d^2)$ is convex in its variables [73], both functions f_{01} and f_{02} are convex as maximum and sum of convex function [102]. The objective function in (5.10) is thus a d.c. function [102]. On the other hand, all constraints (5.6b), (5.9b) and (5.6c) are convex. The main issue is how to recognize partial convexity structures of the quadratic equality constraint (5.3). We have the following result with its proof given below.

Lemma 1. *The equality constraint (5.3) is equivalent to the following equality constraints in additional variables \mathbf{W}_{11} and \mathbf{W}_{22} ,*

$$\begin{bmatrix} \mathbf{W}_{11} & \mathbf{X} & \mathbf{F}\mathbf{H} \\ \mathbf{X}^H & \mathbf{W}_{22} & \mathbf{W}^H \\ \mathbf{H}^H \mathbf{F}^H & \mathbf{W} & \mathbf{I}_K \end{bmatrix} \succeq 0, \quad (5.12a)$$

$$\langle \mathbf{W}_{11} - \mathbf{F}\mathbf{H}\mathbf{H}^H \mathbf{F}^H \rangle \leq 0. \quad (5.12b)$$

Proof. Let's begin with an auxiliary result.

Lemma 2. For given matrix W_{12}, W_{22} of sizes $n \times m$ and $m \times m$ with W_{22} positive definite, one has

$$\begin{bmatrix} 0 & W_{12} \\ W_{12}^T & W_{22} \end{bmatrix} \succeq 0 \quad (5.13)$$

if and only if $W_{12} = 0$.

Proof. Since the “if” part is obvious, let's prove the “only if” part. First, (5.13) means

$$2x^T W_{12}y + y^T W_{22}y \geq 0, \quad \forall (x, y) \in \mathbb{R}^{n+m} \quad (5.14)$$

Now, suppose the contrary, that $W_{12} \neq 0$, i.e. there is (i, j) such that $W_{12}(i, j) \neq 0$. Take (x, y) such that $x_k = 0$ for $k \neq i$ and $y_\ell = 0$ for $\ell \neq j$. Then (5.14) gives

$$2W_{12}(i, j)x_i y_j + W_{22}(j, j)y_j^2 \geq 0, \quad \forall (x_i, y_j) \in \mathbb{R}^2 \quad (5.15)$$

However, for every $y_j \neq 0$, taking $x_i = -\frac{W_{22}(j, j)+1}{2W_{12}(i, j)}y_j$ will make LHS of (5.15) negative, a contradiction. □

By Shur's complement, (5.12a) is equivalent to

$$\begin{bmatrix} \mathbf{W}_{11} & \mathbf{X} \\ \mathbf{X}^H & \mathbf{W}_{22} \end{bmatrix} - \begin{bmatrix} \mathbf{F}\mathbf{H} \\ \mathbf{W}^H \end{bmatrix} \begin{bmatrix} \mathbf{H}^H \mathbf{F}^H & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{11} - \mathbf{F}\mathbf{H}\mathbf{H}^H \mathbf{F}^H & \mathbf{X} - \mathbf{F}\mathbf{H}\mathbf{W} \\ \mathbf{X}^H - \mathbf{W}^H \mathbf{H}^H \mathbf{F}^H & \mathbf{W}_{22} - \mathbf{W}\mathbf{W}^H \end{bmatrix} \succeq 0 \quad (5.16)$$

which implies $\mathbf{W}_{11} - \mathbf{F}\mathbf{H}\mathbf{H}^H \mathbf{F}^H \succeq 0$. This together with (5.12b) yields $\mathbf{W}_{11} = \mathbf{F}\mathbf{H}\mathbf{H}^H \mathbf{F}^H$. Then applying Lemma 2 to (5.16) gives $\mathbf{X} = \mathbf{F}\mathbf{H}\mathbf{W}$ as desired. □

Define a convex set

$$\mathcal{D} := \{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \quad \text{s.t.} \quad (5.6b), (5.9b), (5.6c), (5.12a)\}$$

and use Lemma 1 to write (5.10) as a minimization of a d.c. function subject to d.c. constraint

$$\min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} [f_{01}(\mathbf{X}, \mathbf{y}) - f_{02}(\mathbf{X}, \mathbf{y})] : (5.12b), \quad (5.17)$$

which is written as

$$\min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} \max_{\mu \geq 0} \mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu) \quad (5.18)$$

for

$$\mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu) := f_{01}(\mathbf{X}, \mathbf{y}) - f_{02}(\mathbf{X}, \mathbf{y}) + \mu \langle \mathbf{W}_{11} - \mathbf{F} \mathbf{H} \mathbf{H}^H \mathbf{F}^H \rangle.$$

A nonconvex duality for (5.18) is

$$\max_{\mu \geq 0} \min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} \mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu) \quad (5.19)$$

It is known that there is a gap between a primal nonconvex program and its duality [102], even for indefinite quadratic programs with just one nonconvex quadratic constraint [104]. However, we have the following result.

Lemma 3. *The following minimax equality holds true*

$$\min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} \max_{\mu \geq 0} \mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu) = \max_{\mu \geq 0} \min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} \mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu). \quad (5.20)$$

Moreover, the function

$$g(\mu) := \min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} \mathcal{L}((\mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{W}_{11}), \mu)$$

is increasing in μ .

Proof. Note that (5.12a) implies that

$$\langle \mathbf{W}_{11} - \mathbf{F} \mathbf{H} \mathbf{H}^H \mathbf{F}^H \rangle \geq 0 \quad (5.21)$$

□

By the above Lemma 3, we solve (5.10) by the exact penalty function:

$$\min_{\mathbf{W} \in \mathcal{C}^{K \times M}, \mathbf{F} \in \mathcal{C}^{N \times N}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}} [f_{01}(\mathbf{X}, \mathbf{y}) - f_{02}(\mathbf{X}, \mathbf{y}) + \mu \langle \mathbf{W}_{11} - \mathbf{F} \mathbf{H} \mathbf{H}^H \mathbf{F}^H \rangle] : \quad (5.22)$$

(5.6b), (5.9b), (5.6c), (5.12a),

where the difficult constraint (5.12b) is losslessly incorporated into its objective function.

Rewrite (5.22) in the clearly canonical d.c. form [102]

$$\min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} [F_{01}(\mathbf{X}, \mathbf{y}, \mathbf{W}_{11}) - F_{02}(\mathbf{X}, \mathbf{y}, \mathbf{F})], \quad (5.23)$$

with the convex functions $F_{01}(\mathbf{X}, \mathbf{y}, \mathbf{W}_{11}) := f_{01}(\mathbf{X}, \mathbf{y}) + \mu \langle \mathbf{W}_{11} \rangle$ and $F_{02}(\mathbf{X}, \mathbf{y}, \mathbf{F}) = f_{02}(\mathbf{X}, \mathbf{y}) + \mu \langle \mathbf{F} \mathbf{H} \mathbf{H}^H \mathbf{F}^H \rangle$. Initialized from $(\mathbf{W}^{(0)}, \mathbf{F}^{(0)}, \mathbf{X}^{(0)}, \mathbf{y}^{(0)}, \mathbf{W}_{11}^{(0)}, \mathbf{W}_{22}^{(0)}) \in \mathcal{D}$, our proposed d.c. iteration (DCI) at κ -th iteration solves the following convex program to generate $(\mathbf{W}^{(\kappa+1)}, \mathbf{X}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{F}^{(\kappa+1)})$,

$$\begin{aligned} \min_{(\mathbf{W}, \mathbf{F}, \mathbf{X}, \mathbf{y}, \mathbf{W}_{11}, \mathbf{W}_{22}) \in \mathcal{D}} & [F_{01}(\mathbf{X}, \mathbf{y}) - F_{02}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) - \\ & \langle \nabla F_{02}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}), (\mathbf{X}, \mathbf{y}, \mathbf{F}) - (\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) \rangle] \end{aligned} \quad (5.24)$$

where due to convexity of F_{02} ,

$$\begin{aligned} \langle \nabla F_{02}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}), (\mathbf{X}, \mathbf{y}, \mathbf{F}) - (\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) \rangle & := \\ \sum_{i=1}^M \frac{2\text{Re}(\overline{\mathbf{l}_i \mathbf{x}_i^{(\kappa)}}) \cdot \mathbf{l}_i (\mathbf{x}_i - \mathbf{x}_i^{(\kappa)})}{y_i^{(\kappa)} + \sigma_d^2} - \frac{|\mathbf{l}_i \mathbf{x}_i^{(\kappa)}|^2 (y_i - y_i^{(\kappa)})}{(y_i^{(\kappa)} + \sigma_d^2)^2} + 2\text{Re} \langle (\mathbf{F} - \bar{\mathbf{F}}) \mathbf{H} \mathbf{H}^H \bar{\mathbf{F}}^H \rangle & \leq \\ F_{02}(\mathbf{X}, \mathbf{y}, \mathbf{F}) - F_{02}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) \quad \forall (\mathbf{X}, \mathbf{y}, \mathbf{F}). \end{aligned} \quad (5.25)$$

By (5.25), it is easily seen that

$$F_{01}(\mathbf{X}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{F}^{(\kappa+1)}) - F_{02}(\mathbf{X}^{(\kappa+1)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa+1)}) \leq F_{01}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) - F_{02}(\mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)})$$

so $(\mathbf{W}^{(\kappa+1)}, \mathbf{X}^{(\kappa+1)}, \mathbf{y}^{(\kappa+1)}, \mathbf{F}^{(\kappa+1)}) \in \mathcal{D}$ is better than $(\mathbf{W}^{(\kappa)}, \mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)}) \in \mathcal{D}$ and the proposed DCIs indeed generates a sequence $\{(\mathbf{W}^{(\kappa)}, \mathbf{X}^{(\kappa)}, \mathbf{y}^{(\kappa)}, \mathbf{F}^{(\kappa)})\}$ of improved solutions that converges to a local optimal solution $(\bar{\mathbf{W}}, \bar{\mathbf{X}}, \bar{\mathbf{y}}, \bar{\mathbf{F}})$. It is important to have a good initial $(\mathbf{W}^{(0)}, \mathbf{F}^{(0)}, \mathbf{X}^{(0)}, \mathbf{y}^{(0)}, \mathbf{W}_{11}^{(0)}, \mathbf{W}_{22}^{(0)}) \in \mathcal{D}$, so that convergence towards optimized solution is obtained efficiently. This will be further addressed in the Simulation section.

5.3 Two-Way Relays with MMSE Receiver

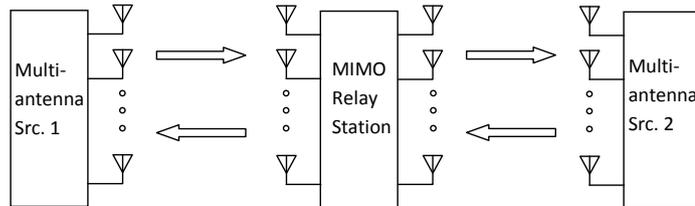


Figure 5.2: A two-way MIMO Relay system.

This section explores throughput capacity achievable by using two-way MIMO relay systems. Consider a two-way MIMO relay communication system where nodes \mathcal{S}_1 and \mathcal{S}_2 exchange information with the aid of a MIMO relay \mathcal{R} [106]. We assume that both nodes \mathcal{S}_1 and \mathcal{S}_2 are equipped with M antennas, whereas the relay \mathcal{R} node has N antennas. The generalization to a system with different number of antennas at each node is straightforward. The information exchange between nodes \mathcal{S}_1 and \mathcal{S}_2 is completed in two time slots. In the first time slot, nodes \mathcal{S}_1 and \mathcal{S}_2 concurrently transmit, and the signal vector from node i is $\mathbf{x}_i = \mathbf{W}_i \mathbf{s}_i$, $i = 1, 2$, where \mathbf{s}_i is the $N \times 1$ modulated source signal vector covariance matrix $\mathbf{R}_{s_i} = \mathbb{E}[\mathbf{s}_i \mathbf{s}_i^H]$ and \mathbf{W}_i is the $M \times M$ source precoding matrix at node i . Assume that the source signal vector satisfies $\mathbb{E}[\mathbf{s}_i \mathbf{s}_i^H] = \mathbf{I}_M$, $i = 1, 2$, where \mathbf{I}_M stands for an $M \times M$ identity matrix. The $N \times 1$ received signal vector \mathbf{r} at the relay node is

$$\mathbf{r} = \mathbf{H}_1 \mathbf{x}_1 + \mathbf{H}_2 \mathbf{x}_2 + \mathbf{n}_r \quad (5.26)$$

where \mathbf{H}_i is the $N \times M$ MIMO channel from node \mathcal{S}_i to the relay node and \mathbf{n}_r is the additive Gaussian noise vector with covariance matrix $\sigma_r^2 \mathbf{I}_N$.

In the second time slot, the MIMO relay node filters \mathbf{r} by $\mathbf{F} \in \mathcal{N}^{N \times N}$ and broadcasts the signal vector $\mathbf{F}\mathbf{r}$ to nodes \mathcal{S}_1 and \mathcal{S}_2 . After cancelling the self-interference term, the received signal vector \mathbf{y}_i at node i^* can be written as

$$\mathbf{y}_i = \mathbf{L}_{2/i} \mathbf{F} \mathbf{H}_{2/i} \mathbf{W}_{2/i} \mathbf{s}_{2/i} + \mathbf{L}_{2/i} \mathbf{F} \mathbf{n}_r + \mathbf{n}_{d,i}, i = 1, 2 \quad (5.27)$$

where $\mathbf{L}_{2/i}$ is the $M \times N$ forward channel from the relay \mathcal{R} to terminal \mathcal{S}_i , and $\mathbf{n}_{d,i}$ is the additive Gaussian noise vector with zero means and covariance $\sigma_d^2 \mathbf{I}_M$.

Using linear MMSE receivers at the both nodes \mathcal{S}_i , $i = 1, 2$ the estimated signal waveforms are given by $\hat{\mathbf{s}}_{2/i} = \hat{\mathbf{B}}_i \mathbf{y}_i$, where [42]

$$\hat{\mathbf{B}}_i = \arg \min_{\mathbf{B}_i} \mathbb{E}(\|\mathbf{s}_{2/i} - \mathbf{B}_i \mathbf{y}_i\|^2) = \mathbf{R}_{y_i s_{2/i}}^H \mathbf{R}_{y_i}^{-1}, \quad (5.28)$$

with $\mathbf{R}_{y_i s_{2/i}} = \mathbb{E}[\mathbf{y}_i \mathbf{s}_{2/i}^H] = \mathbf{L}_{2/i} \mathbf{X}_{2/i}$ and $\mathbf{R}_{y_i} = \mathbb{E}[\mathbf{y}_i \mathbf{y}_i^H] = \mathbf{L}_{2/i} \mathbf{X}_{2/i} \mathbf{X}_{2/i}^H \mathbf{L}_{2/i}^H + \sigma_r^2 \mathbf{L}_{2/i} \mathbf{F} \mathbf{F}^H \mathbf{L}_{2/i}^H + \sigma_d^2 \mathbf{I}_M$ for

$$\mathbf{X}_j = \mathbf{F} \mathbf{H}_j \mathbf{W}_j, j = 1, 2. \quad (5.29)$$

* $2/i := \frac{2}{i}$ in common sense

The MMSE at terminal i is thus

$$\mathbb{E}(\|\mathbf{s}_{2/i} - \hat{\mathbf{s}}_{2/i}\|^2) = e_{2/i}(\mathbf{W}_1, \mathbf{W}_2, \mathbf{F}) := \langle [\mathbf{I}_M + (\mathbf{L}_{2/i} \mathbf{X}_{2/i})^H (\sigma_r^2 \mathbf{L}_{2/i} \mathbf{F} \mathbf{F}^H \mathbf{L}_{2/i}^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_{2/i} \mathbf{X}_{2/i}]^{-1} \rangle, \quad (5.30)$$

which is a very complicated nonlinear functions in $(\mathbf{W}_1, \mathbf{W}_2, \mathbf{F})$. Obviously its minimization remains intractable even when $(\mathbf{W}_1, \mathbf{W}_2)$ is held fixed. To apply alternating optimization, with this MMSE is expressed as optimization in $(\mathbf{W}_1, \mathbf{W}_2), \mathbf{F}$ and $(\mathbf{B}_1, \mathbf{B}_2)$ (from (5.28)) [106], which is convex when two of them are held fixed.

Now, with the introduction of nonlinear variables

$$\mathbf{W}_{11} = \mathbf{F} \mathbf{F}^H, \quad (5.31)$$

$$\mathbf{Y}_j = (\mathbf{L}_j \mathbf{X}_j)^H (\sigma_r^2 \mathbf{L}_j \mathbf{W}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j \mathbf{X}_j, j = 1, 2, \quad (5.32)$$

the MMSE minimax optimization subject to transmission power constraints is formulated as

$$\min_{\mathbf{W}_1, \mathbf{F}, \mathbf{Y}_j, \mathbf{X}_j} \max_{j=1,2} \langle (\mathbf{I}_M + \mathbf{Y}_j)^{-1} \rangle \quad \text{s.t.} \quad (5.29), (5.31), (5.32), \quad (5.33a)$$

$$[\mathbf{W}_j \mathbf{W}_j^H]_{m,m} \leq \nu p_0, \quad m = 1, \dots, M, j = 1, 2, \quad (5.33b)$$

$$\left[\sum_{j=1}^2 \mathbf{X}_j \mathbf{X}_j^H + \sigma_r^2 \mathbf{F} \mathbf{F}^H \right]_{n,n} \leq \gamma_n, \quad n = 1, \dots, N, \quad (5.33c)$$

$$\left\langle \sum_{j=1}^2 (\mathbf{W}_j \mathbf{W}_j^H + \mathbf{X}_j \mathbf{X}_j^H) + \sigma_r^2 \mathbf{F} \mathbf{F}^H \right\rangle \leq P_{sum}^{max} \quad (\text{Total Power}) \quad (5.33d)$$

The key observation is that the nonlinear constraints (5.29), (5.31), and (5.32) are equivalently expressed by

$$\begin{bmatrix} \mathbf{W}_{11} & \mathbf{X}_j & \mathbf{F} \\ \mathbf{X}_j^H & \mathbf{W}_{22j} & \mathbf{W}_j^H \mathbf{H}_j^H \\ \mathbf{F}^H & \mathbf{H}_j \mathbf{W}_j & \mathbf{I}_N \end{bmatrix} \succeq 0, j = 1, 2, \quad (5.34a)$$

$$\begin{bmatrix} \mathbf{Y}_j & \mathbf{X}_j^H \mathbf{L}_j^H \\ \mathbf{L}_j \mathbf{X}_j & \sigma_r^2 \mathbf{L}_j \mathbf{W}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M \end{bmatrix} \succeq 0, j = 1, 2, \quad (5.34b)$$

$$\langle \mathbf{W}_{11} - \mathbf{F} \mathbf{F}^H \rangle + \sum_{j=1}^2 \langle \mathbf{Y}_j - (\mathbf{L}_j \mathbf{X}_j)^H (\sigma_r^2 \mathbf{L}_j \mathbf{W}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j \mathbf{X}_j \rangle \leq 0. \quad (5.34c)$$

In summary, the exact penalty formulation for minimizing the maximum mean-squared error of the estimates of the two messages by two different sources can be formulated as under the power

constraints is

$$\begin{aligned} & \min_{\mathbf{F}, \mathbf{W}_{11}, \mathbf{X}_j, \mathbf{Y}_j, \mathbf{W}_{22j}, j=1,2} [\max_{j=1,2} \langle (\mathbf{I}_M + \mathbf{Y}_j)^{-1} \rangle + \mu \langle \langle \mathbf{W}_{11} - \mathbf{F}\mathbf{F}^H \rangle \rangle] \\ & + \sum_{j=1}^2 \langle \mathbf{Y}_j - (\mathbf{L}_j \mathbf{X}_j)^H (\sigma_r^2 \mathbf{L}_j \mathbf{W}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j \mathbf{X}_j \rangle \quad : (5.33b) - (5.33d), (5.34a), (5.34b). \end{aligned} \quad (5.35)$$

In Appendix C, it is shown that function

$$g_j(\mathbf{X}_j, \mathbf{W}_{11}) := \langle (\mathbf{L}_j \mathbf{X}_j)^H (\sigma_r^2 \mathbf{L}_j \mathbf{W}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j \mathbf{X}_j \rangle \quad (5.36)$$

is convex with

$$\begin{aligned} & \langle \nabla g_j(\bar{\mathbf{X}}_j, \bar{\mathbf{W}}_{11}), (\mathbf{X}_j, \mathbf{W}_{11}) - (\bar{\mathbf{X}}_j, \bar{\mathbf{W}}_{11}) \rangle = \\ & 2 \cdot \text{Re} \{ \langle \bar{\mathbf{X}}_j^H \mathbf{L}_j^H (\sigma_r^2 \mathbf{L}_j \bar{\mathbf{W}}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j (\mathbf{X}_j - \bar{\mathbf{X}}_j) \rangle \} - \\ & \sigma_r^2 \cdot \langle \bar{\mathbf{X}}_j^H \mathbf{L}_j^H (\sigma_r^2 \mathbf{L}_j \bar{\mathbf{W}}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j (\mathbf{W}_{11} - \bar{\mathbf{W}}_{11}) \mathbf{L}_j^H (\sigma_r^2 \mathbf{L}_j \bar{\mathbf{W}}_{11} \mathbf{L}_j^H + \sigma_d^2 \mathbf{I}_M)^{-1} \mathbf{L}_j \bar{\mathbf{X}}_j \rangle. \end{aligned} \quad (5.37)$$

On the other hand, by (5.27), the information throughput at terminal \mathcal{S}_i is

$$\ln I(\mathbf{s}_{i/2}; \mathbf{y}_i) = \ln \det(\mathbf{I}_M + \mathbf{L}_{i/2} \mathbf{X}_{i/2} (\sigma_r^2 \mathbf{L}_{i/2} \mathbf{F}\mathbf{F}^H \mathbf{L}_{i/2}^H + \sigma_d^2 \mathbf{I}_M)^{-1} (\mathbf{L}_{i/2} \mathbf{X}_{i/2})^H). \quad (5.38)$$

Under the variable introductions (5.29), (5.31), (5.32), the sum throughput maximization is written by

$$\min_{\mathbf{F}, \mathbf{W}_{11}, \mathbf{Z}_j, \mathbf{Y}_j, \mathbf{W}_{22j}, j=1,2} \left[- \sum_{j=1}^2 \log_2 \det(\mathbf{I}_M + \mathbf{Y}_j) \right] : (5.33b) - (5.33d), (5.34a) - (5.34c). \quad (5.39)$$

Its exact penalty formulation through nonconvex duality is thus

$$\begin{aligned} & \min_{\mathbf{F}, \mathbf{W}_{11}, \mathbf{Z}_j, \mathbf{Y}_j, \mathbf{W}_{22j}, j=1,2} \left[- \sum_{j=1}^2 \log_2 \det(\mathbf{I}_M + \mathbf{Y}_j) + \mu \langle \langle \mathbf{W}_{11} - \mathbf{F}\mathbf{F}^H \rangle \rangle \right. \\ & \left. + \sum_{j=1}^2 (\langle \mathbf{Y}_j \rangle - g_j(\mathbf{X}_j, \mathbf{W}_{11})) \right] : (5.33b) - (5.33d), (5.34a) - (5.34b), \end{aligned} \quad (5.40)$$

with convex functions $g_j(\mathbf{X}_j, \mathbf{W}_{11})$ defined from (5.36).

Obviously, the information throughput maximization at terminal \mathcal{S}_i in one way relaying is described by (5.38) by setting $\mathbf{W}_{i/2} = 0$ and can be solved similarly by DCI. Suppose $p(P_{sum})$ and $p_i(P_{sum}/2)$ are the optimal values of (5.39) and the mentioned information throughput maximization at terminal \mathcal{S}_i (in one way relaying) with power sum $P_{sum}/2$. Since two-way relaying uses 2

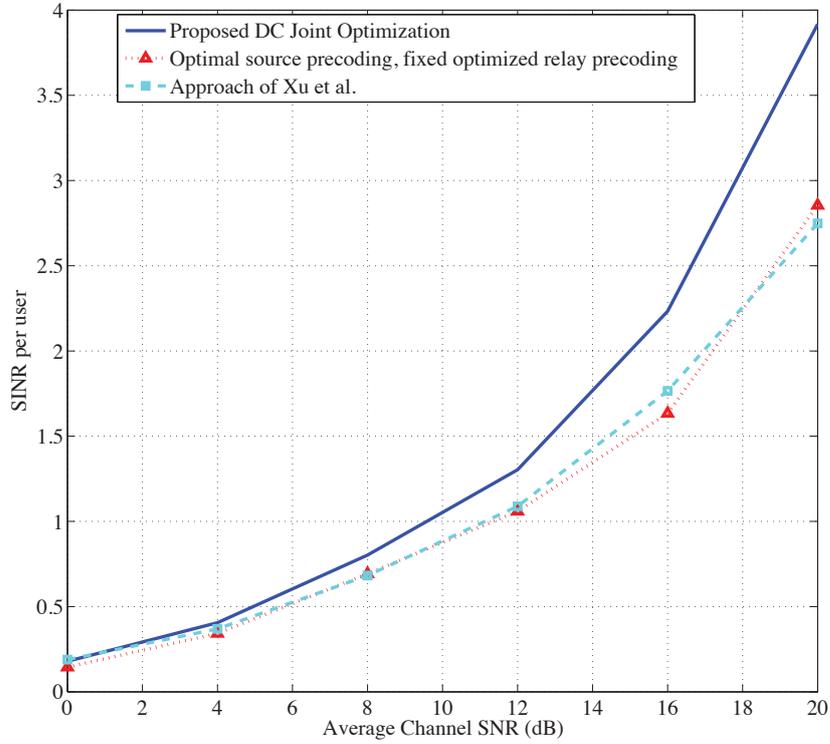


Figure 5.3: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (4, 4, 4)$ for downlink cellular communication.

channels for communication between two terminals while one-way relaying uses 4 channels in total for their communication, the overall information throughput efficiency is defined by $p(P_{sum})/2$ and $(p_1(P_{sum}/2) + p_2(P_{sum}/2))/4$, respectively. Therefore the advantage of two-way relaying over one-way one can be verified by compare these values.

5.4 Simulation Results

In this section, simulation results and discussions are presented to illustrate the performance of the proposed DCI method. It is assumed that the knowledge of both forward channel (i.e. between BS and RS) as well as backward channel (i.e. between RS and users) is perfectly known. All channel elements are generated as zero-mean, complex Gaussian distributed with unit variance

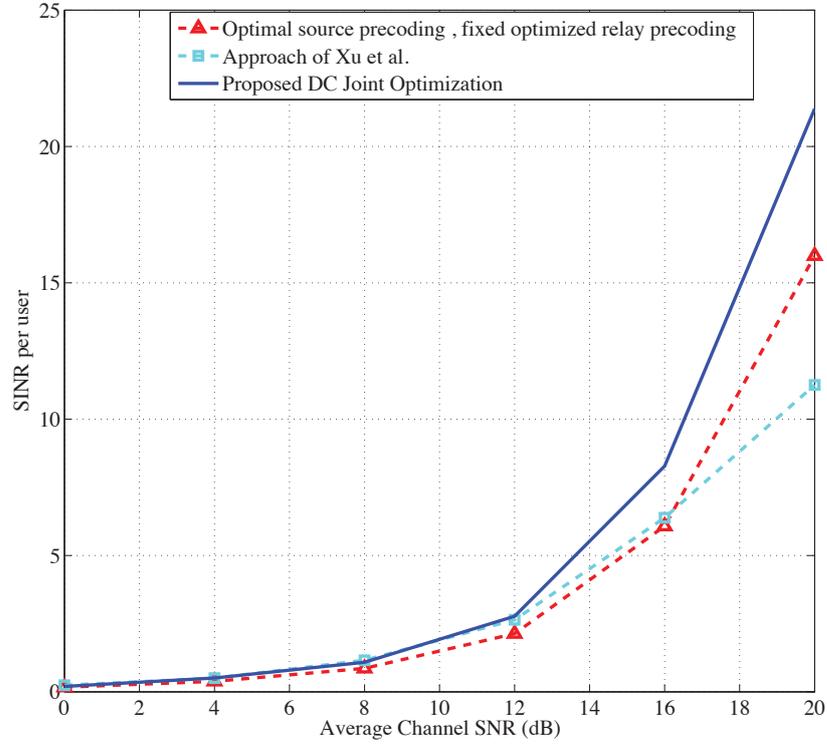


Figure 5.4: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (2, 2, 4)$ for downlink cellular communication.

and the results are averaged out over 100 Monte Carlo runs. Power of the input symbols \mathbf{s} as well as AWGN at relay station and destination nodes are normalized to unity. Recall that M is the number of user pairs, N is the number of antennas at the relay station (RS), and K is the number of antenna at base station (BS). Simulation is carried out for different combinations of parameters (M, N, K) . The individual antenna power is set as $p_m \leq \nu p_0$ where $\nu = 1.5$ is a scaling factor, and $p_0 = 10$. On the other hand, the individual relay antenna power is limited to $P_n^{R, \max} = \gamma_n = P_{\text{rel}}^{\text{tot}}/N$. Thus, the total sum power budget is upper bounded by $P_{\text{sum}}^{\max} = (Kp_0 + \sum_{n=1}^N \gamma_n)$. In the following subsections, we divide simulation results into two parts based upon two MIMO communication system models described in previous sections.

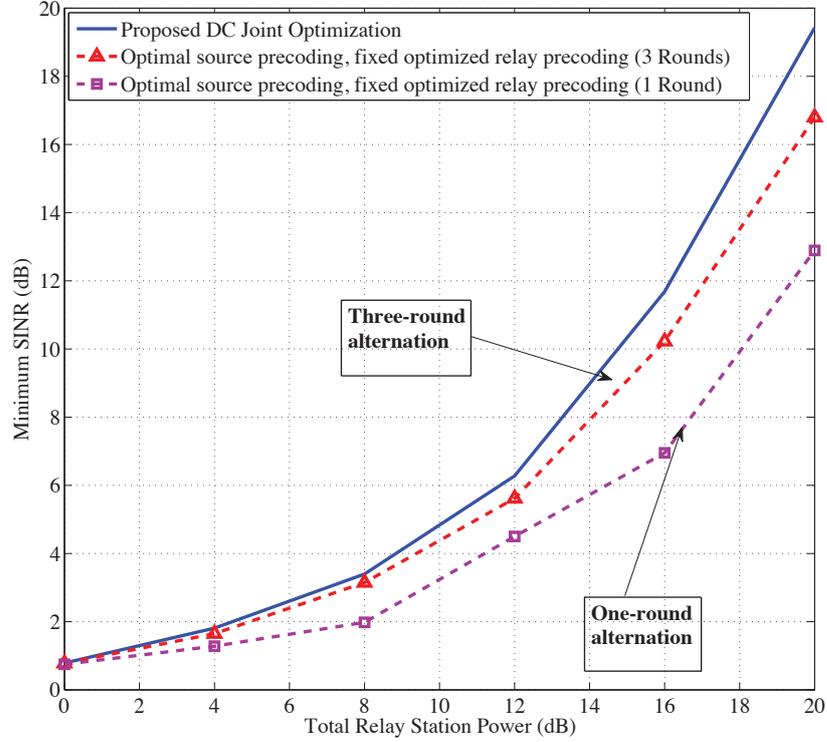


Figure 5.5: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (3, 5, 5)$ for downlink cellular communication.

5.4.1 Multiuser MIMO Downlink Communication

First, we consider MIMO broadcast channel downlink communication model where a BS broadcasts message signals towards multiple users with the assistance of a MIMO relay station. For $(M, N, K) = (2, 2, 4)$ and $(M, N, K) = (2, 2, 4)$, we plot SINR achieved for each user against total relay station power in Figure 5.3 and Figure 5.4. It is assumed equal transmit power at base station and relay station, i.e., $P_{tot}^B = P_{tot}^R$ is utilized. Simulation results compare performance of the minimum SINR achieved by proposed joint d.c. algorithm with the average SINR achieved by method of [111] and alternating optimization of \mathbf{F} and \mathbf{W} . It is observed that the numerical results demonstrate better performance of the proposed optimization method than the rest of aforementioned methods. Although, curve of [111] denotes average SINR per user based on the sum-rate

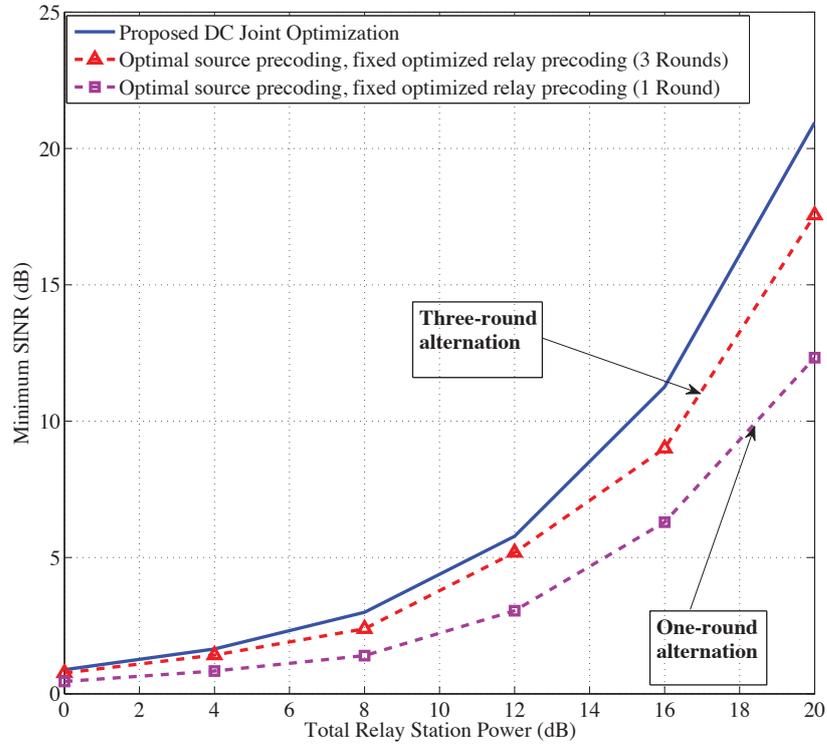


Figure 5.6: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (3, 6, 6)$ for downlink cellular communication.

maximization considered in [111], it is still not better than the minimum SINR achieved by the proposed method. Note that the SINR performance of both the [111] algorithm and alternating method are almost similar to each other. It is worth mentioning that throughput sum achieved by the proposed method for, say, 4 users at power level of 15 dB is 6.34 bps/Hz compared to 5.51 bps/Hz achieved by [111] despite the fact that the proposed method does not even aim at sum rate maximization.

Next examples for this broadcast model attempts to present an extensive comparison between the proposed method and several rounds of alternating optimization method from Figure 5.5 to Figure 5.8. Essentially, we implement the following two steps to yield results from alternating optimization:

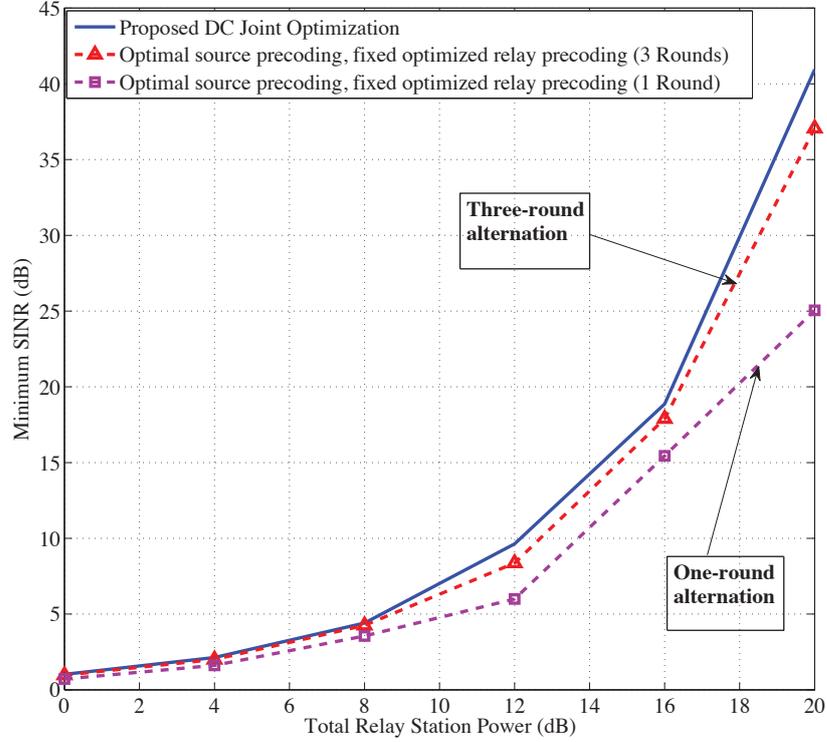


Figure 5.7: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (4, 8, 8)$ for downlink cellular communication.

1. With fixed BS precoding matrix \mathbf{W} , we maximize minimum SINR as a function of \mathbf{F} in the form of a d.c. program which can be solved by [81].
2. With fixed \mathbf{F} obtained from the previous step, we solve d.c program in \mathbf{W} .

This procedure is repeated recursively, until there is no further improvement in the resulting SINR. Figure 5.5 illustrates that the proposed joint source and relay optimization algorithms obtain an improvement of approximately 2 dB of SINR at power level of 16 dB compared with the other approaches. Although the three-round algorithm performs better than the one-round of alternation, the former algorithm requires a larger number of iterations than the latter one to converge to final solution as listed in Table I. Furthermore, it is observed that the joint optimization algorithm converges to the same final result whether it is initiated from the solution of the three-round or

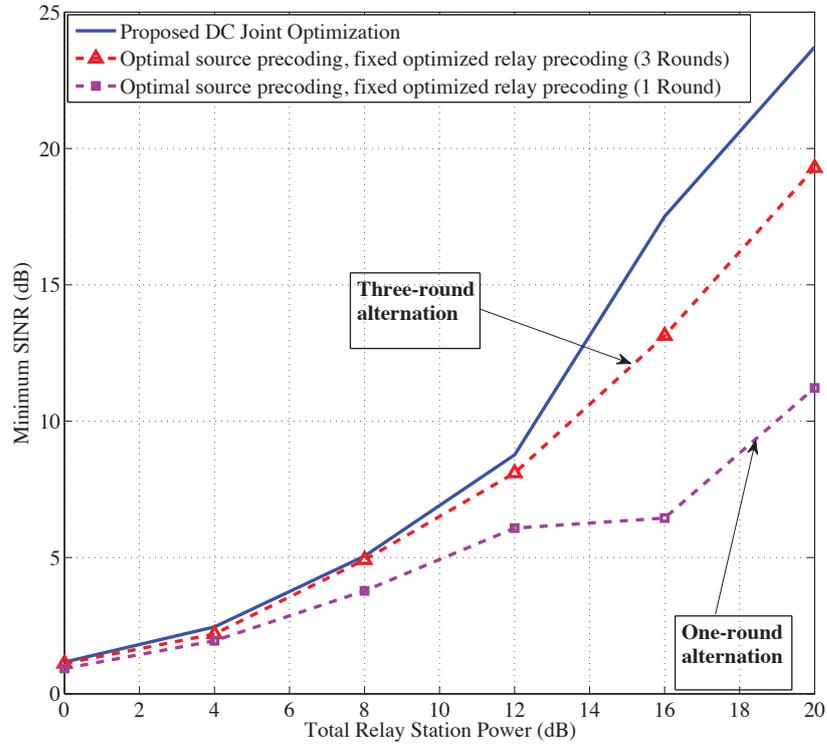


Figure 5.8: Signal-to-interference-plus-noise ratio (SINR) plotted versus total relaying power for $(M, N, K) = (5, 8, 8)$ for downlink cellular communication.

one-round alternating algorithm. Based on the number of iterations they need to converge, the overall computational complexity of the one-round alternating algorithm is smaller than that of the three-round algorithm. Such performance-complexity trade-off is very important for practical multiuser MIMO relay communication systems.

By increasing number of antenna, as shown in Figure 5.7, significant improvement in SINR is achievable by virtue of spatial diversity. In the later examples, we compare the SINR performance of the proposed algorithms for different number of antennas at the RS and BS for various combinations of user mobiles. Figure 5.7 compares minimum SINR performance of the proposed algorithms versus one- and three-round alternating optimizations for the same number of source and relay antennas as used in Figure 5.8 with different number of users. It can be clearly seen

Table 5.1: Average number of iterations for convergence for $(M, N, K) = (3, 5, 5)$.

Power (dB)	Joint Algorithm	One-Round	Three-Round
0.00	12.00	22.20	34.10
4.00	13.15	26.35	36.45
8.00	13.20	19.40	29.20
12.00	13.75	27.80	39.55
16.00	14.85	16.75	29.10
20.00	14.00	18.75	25.00

from Figure 5.7 that as we increase the number of antennas at the relay and/or source(s), the performance of the proposed algorithms improve significantly.

5.4.2 Two-way Relays with MMSE Receiver

Next we consider two-way MIMO relay model with MMSE receiver under two performance criteria: 1) Sum-rate maximization, and 2) Minimization of maximum MSE. Maximization of sum of information throughput is solved by both DCI and exact penalty function approach of (5.40) by choosing an appropriate value of μ (which is $\mu = 10$ in our case). We use a randomly chosen initial point to start the proposed d.c. iterations similar to the previous scenario. Four system configurations of $(M, N) = \{(2, 2), (3, 4), (4, 4), (5, 5)\}$ are studied in two-way MIMO relay case. The individual power thresholds are similarly set as in previous simulation results with source power for each antenna fixed at $p_0 = 10$ dB while the relay station power is varied from 1 dB to 10 dB. The maximized sum throughput for two-way and one-way relay communication are obtained for comparison.

A comparison is drawn in terms of sum throughput between two-way and on-way MIMO relays in Figure 5.9 for $(M, N) = (3, 4)$ and $(M, N) = (3, 5)$. As can be observed, there is a gap of 5 bps/Hz at relay power of 10 dB between the two modes of communication. In Figure 5.10, we further investigate the sum throughput performance when the antenna number N at \mathcal{S}_j changes from 3 to 4 with $N = 4, 5$ that allows the system performance to significantly improve when N is

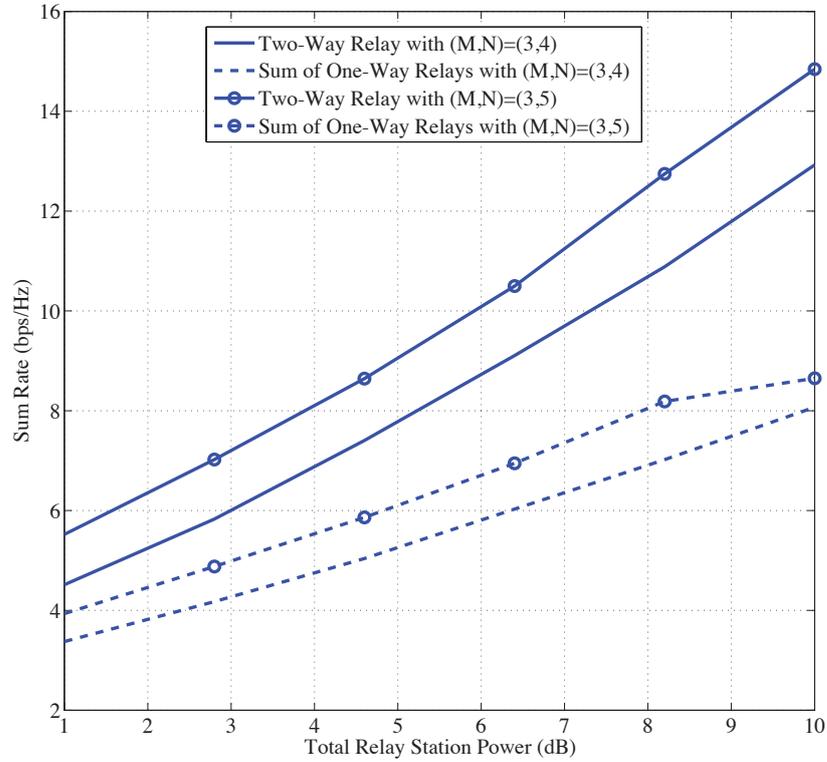


Figure 5.9: Sum throughput plotted versus total relaying power for two-way and one-way MIMO relay-assisted communication.

greater. This owes to extra diversity gain that can be exploited to enhance the reliability of data transmission.

Furthermore, in Figure 5.11, we plot sum of MSEs for both sources \mathcal{S}_1 and \mathcal{S}_2 when they communicate with each other simultaneously through the MIMO relay for a configuration of $(M, N) = (2, 2)$. By solving (5.35), we obtain solution for minimizing the maximum MSE among the given sources/users for which a comparison is drawn between [106] and the proposed d.c. based method. Note that it is rather the sum of MSE of both sources which is used along the vertical axis of the plot, and despite the fact that the objective function in the proposed method does not aim at minimizing the sum of mean square error, the proposed method shows improvement over method of [106] by just aiming at minimizing the sum of MSEs. This is further evident when

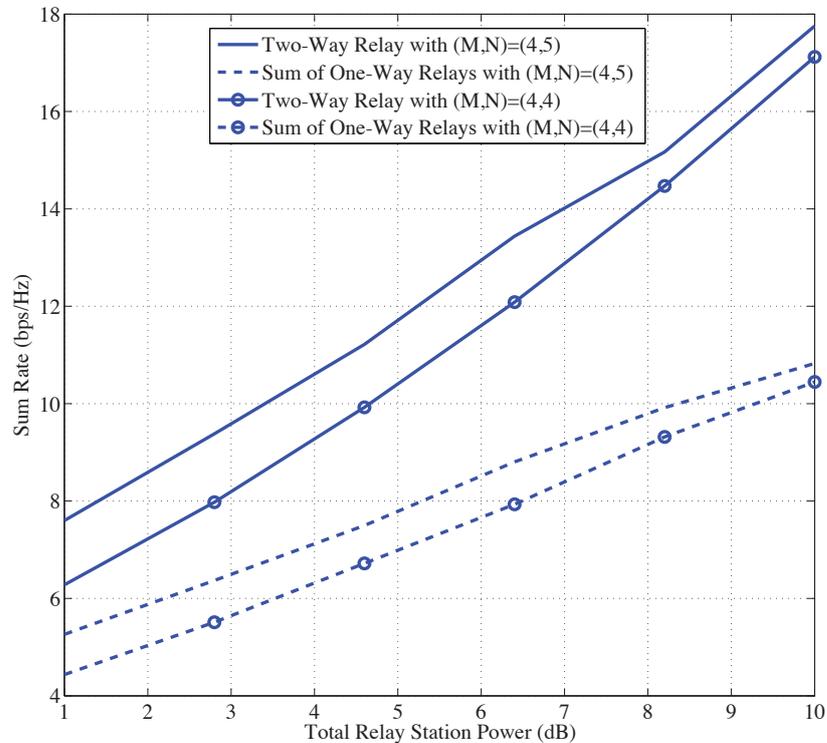


Figure 5.10: Sum throughput plotted versus total relaying power for two-way and one-way MIMO relay-assisted communication.

bit error rate is plotted with QPSK modulation scheme for $(M, N) = (4, 4)$ to show in Figure 5.12 that the proposed method achieves lower bit error rate than the method of [106].

5.5 Summary

Joint optimization of source precoding and MIMO relay processing matrices is considered in this chapter in two parts. In the first part, a maximin SINR optimization problem is jointly solved for one-way MIMO relays in multiuser broadcast communication in cellular network via proposed DCI method. In the second part, two-way MIMO relays with MMSE receiver are considered for both minimax MSE and sum-throughput maximization. Both of these problems have been shown to be solved via DCI joint iterative algorithm. Simulation results demonstrate supremacy of the

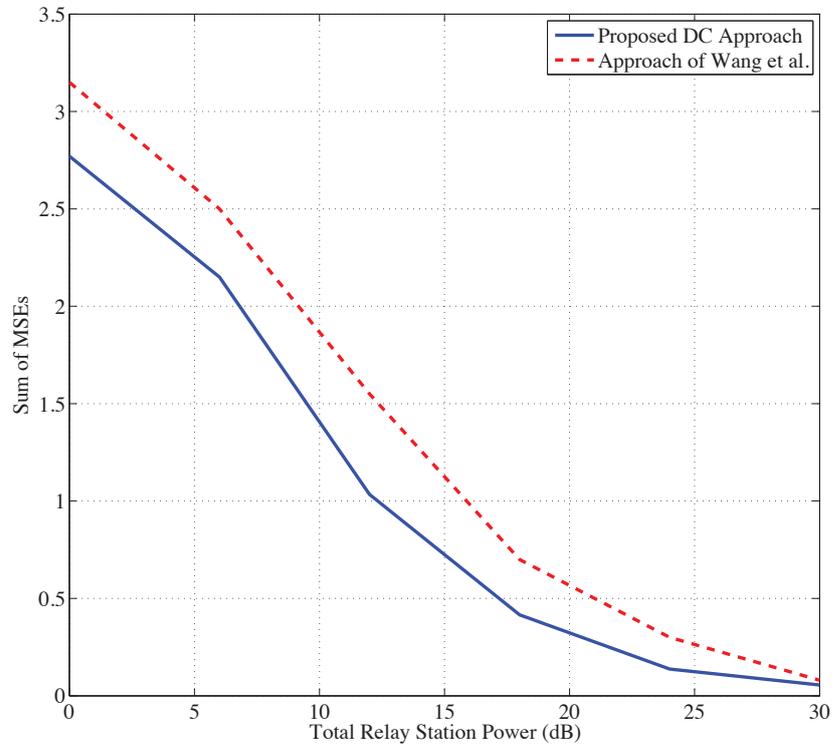


Figure 5.11: Sum of MSE of both source \mathcal{S}_1 and \mathcal{S}_2 plotted versus total relaying power with $(M, N) = (2, 2)$ for two-way MIMO relay-assisted communication.

proposed methods over previously known algorithms.

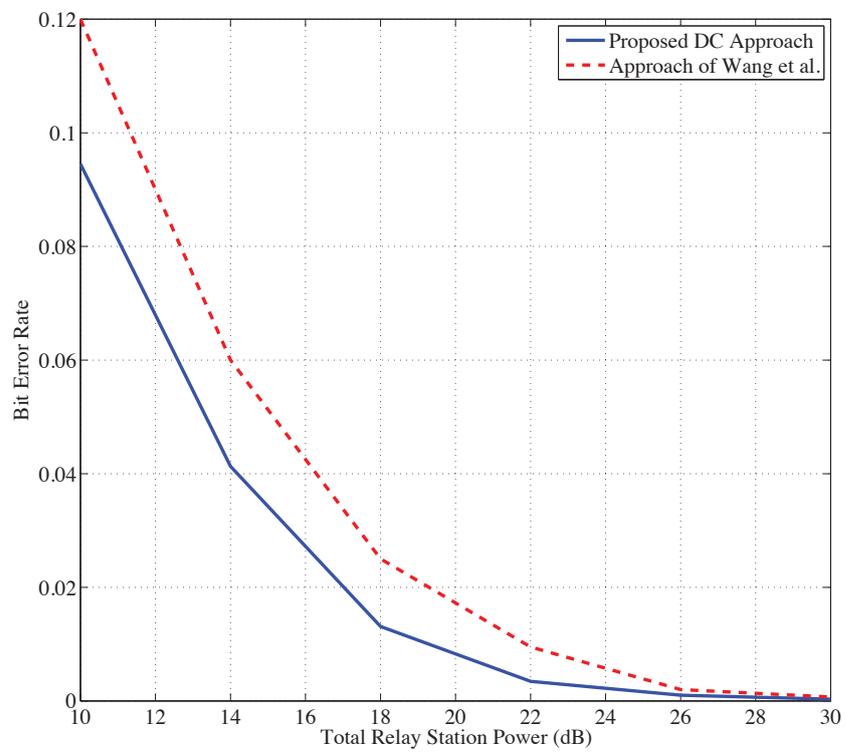


Figure 5.12: Bit error rate plotted versus total relaying power with $(M, N) = (4, 4)$ for two-way MIMO relay-assisted communication.

Chapter 6

Conclusion

In this chapter, we recap the main contributions of the dissertation and point out some possible future research areas. This dissertation attempts to exploit the hidden convexity of the resource allocation problems under some given performance criteria such as minimum mean square error (MMSE), information throughput or signal-to-interference-plus-noise ratio (SINR) and then successfully finds tractable optimization formulations. We have considered the optimal design problems of power allocation, active sensor selection and antenna beamforming vector assignment problems to meet various performance goals in an optimized manner. In short, the contributions of this dissertation are summarized as follows.

After giving the motivation of the dissertation in Chapter 1, an overview of statistical estimation, wireless communication and convex optimization theory on which many results of this dissertation are based on have been briefly introduced in Chapter 2.

Chapter 3 has addressed the problem of power allocation among sensor nodes for locating a static target or for tracking a dynamic target in either linear or nonlinear sensing systems. These sensors observe the targets and then transmit their noisy observations through noisy wireless channels to the FC where the final estimate is carried out. Due to limited energy resources, it is desired to develop an optimized power allocation technique which is able to minimize mean square error of the estimate under a given power budget. A novel technique based on tractable optimization (SDP) and approximation (unscented and linear fractional transformations) has been proposed. The multi-sensor diversity has been fully exploited to arrive at an accurate estimate of the target's

state. Moreover, Chapter 3 has also formulated joint program of active sensor assignment and power allocation in linear and nonlinear sensor network is considered. Its optimized solutions are quickly located by the developed d.c. procedure of local search. Accompanying simulation results clearly showed the viability of the theoretical results.

Chapter 4 has analyzed the beamforming design problems in multi-user wireless relay networks to maximize the minimum information throughput among all users. Both cases of concurrent and orthogonal transmissions from sources to relays are considered. It has also considered joint design problem of source power allocation and relay beamforming in multi-user multi-relay wireless networks. Different from the existing approach which reformulates the design problems to matrix rank-one constrained optimizations, the proposed approach exploits the d.c. structure of the objective function and the convex structure of the constraints to develop efficient iterative algorithms of very low complexity to find the solutions. Numerical results demonstrate that the developed algorithms are able to locate the global optimal solutions by a few iterations and they illustrate the superiority of our method over the existing methods.

In Chapter 5 joint optimization of source precoding and MIMO relay processing matrices is considered in this paper. The paper is divided into two parts. In the first part, a maximin SINR optimization problem is jointly solved for one-way MIMO relays in multiuser broadcast communication in cellular network via proposed DCI method. In the second part, two-way MIMO relays with MMSE receiver are considered for both minimax MSE and sum-throughput maximization. Both of these problems have been shown to be solved via DCI joint iterative algorithm. Simulation results demonstrate supremacy of the proposed methods over previously known algorithms.

Bibliography

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: A survey,” *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [2] G. Amarasuriya, C. Tellambura, and M. Ardakani, “Two-way amplify-and-forward multiple-input multiple-output relay networks with antenna selection,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, pp. 1513–1529, 2012.
- [3] K. Amiri, M. Wu, J. Cavallaro, and J. Lilleberg, “Cooperative partial detection using mimo relays,” *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5039–5049, 2011.
- [4] B. Anderson and J. Moore, *Optimal filtering*. Prentice-hall Englewood Cliffs, NJ, 1979, vol. 11.
- [5] P. Apkarian, D. Noll, and A. Rondepierre, “Mixed H_2H_∞ control via nonsmooth optimization,” *SIAM J. Control Optimz.*, vol. 47, no. 3, pp. 1516–1546, 2008.
- [6] P. Apkarian, D. Noll, J. B. Thevenet, and H. D. Tuan, “A spectral quadratic-SDP method with applications to fixed-order H_2 and H_∞ synthesis,” *European Journal of Control*, vol. 10, no. 6, pp. 527–538, 2004.
- [7] P. Apkarian, D. Noll, and H. D. Tuan, “Fixed-order H_∞ control design via a partially augmented Lagrangian method,” *International Journal of Robust and Nonlinear Control*, vol. 12, no. 6, pp. 1137–1148, 2003.
- [8] P. Apkarian and H. D. Tuan, “Concave programming in control theory,” *J. of Global Optimization*, vol. 15, pp. 243–270, 1999.

- [9] M. Arik and O. B. Akan, "Collaborative mobile target imaging in UWB wireless radar sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 6, pp. 950–961, 2010.
- [10] I. Bahceci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Transactions on Communications*, vol. 56, no. 7, pp. 1146–1156, 2008.
- [11] I. Bahceci and A. K. Khandani, "Energy-efficient estimation of correlated data in wireless sensor networks," in *Proc. 40th Annual Conf. Information Sciences and Systems*, 2006, pp. 973–978.
- [12] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS, 1995.
- [13] Y. Bar-Shalom, X.-R. Li, and T. Kirubajan, *Estimation with applications to tracking and navigation*. Wiley, 2001.
- [14] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization, 2001.
- [15] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
- [16] A. Bletsas and A. Lippman, "Implementing cooperative diversity antenna arrays with commodity hardware," *IEEE Communication Magazine*, vol. 44, no. 12, pp. 33–40, 2006.
- [17] G. Bournaka, K. Cumanan, S. Lambbotharan, J. Chambers, and F. Lazarakis, "An SINR balancing based multiuser relaying scheme," in *Proc. of IEEE International Conference on Communication Systems (ICCS)*, Singapore, Dec. 2010.
- [18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.
- [19] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

- [20] D. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 418–425, 1985.
- [21] Y. Cheng and M. Pesavento, "Joint optimization of source power allocation and distributed relay beamforming in multiuser peer-to-peer relay networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2962–2973, 2012.
- [22] G. Choi, W. Zhang, and X. Ma, "Achieving joint diversity in decode-and-forward mimo relay networks with zero-forcing equalizers," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1545–1554, 2012.
- [23] C. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.
- [24] H. Chung, N. Lee, B. Shim, and T. Oh, "On the beamforming design for mimo multipair two-way relay channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 7, pp. 3301–3306, 2012.
- [25] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, 2007.
- [26] X. Deng and A. Haimovich, "Power allocation for cooperative relaying in wireless networks," *IEEE Communications Letters*, vol. 9, no. 11, pp. 994–996, 2005.
- [27] Z. Ding, W. Chin, and K. Leung, "Distributed beamforming and power allocation for cooperative networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1817–1822, 2008.
- [28] L. Dong, A. P. Petropulu, and H. V. Poor, "Weighted cross-layer cooperative beamforming for wireless networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3240–3252, 2009.
- [29] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, 2009.

- [30] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Fazor, "Multiple peer-to-peer communications using a network of relays," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3053–3062, 2009.
- [31] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [32] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Lecture Notes in Computer Science*, vol. 2634, pp. 162–177, 2003.
- [33] —, "Power, spatio-temporal bandwidth, and distortion in large sensor network," *IEEE Journal of Selected Areas in Communications*, vol. 23, no. 4, pp. 745–754, 2005.
- [34] M. O. Hasna and M. S. Alouni, "End-to-end performance of transmission systems with relays over Rayleigh fading channel," *IEEE Transactions on Wireless Communication*, vol. 2, no. 6, pp. 1126–1131, 2003.
- [35] K. C. Ho and L. M. Vicente, "Sensor allocation for source localization with decoupled range and bearing estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5773–5789, 2008.
- [36] N. T. Hoang, H. D. Tuan, P. Apkarian, and S. Hosoe, "Gain-scheduled filtering for time-varying discrete systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2464–2476, 2004.
- [37] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ Press, 1990.
- [38] L. Huang, S. Kumar, and C. Kuo, "Adaptive resource allocation for multimedia qos management in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 2, pp. 547–558, 2004.
- [39] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1414–1423, 2009.

- [40] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.
- [41] S. Kar and J. Moura, "A mixed time-scale algorithm for distributed parameter estimation : Nonlinear observation models and imperfect communication," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2009*, 2009, pp. 3669–3672.
- [42] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
- [43] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 510–515, 2012.
- [44] S. Kim, B. Ku, W. Hong, and H. Ko, "Performance comparison of target localization for active sonar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1371–1380, 2008.
- [45] H. Konno, P. T. Thach, and H. Tuy, *Optimization on low rank nonconvex structure*. Kluwer Academic, 1997.
- [46] G. Kramer, M. Gastpar, and G. Gupta, "Cooperative strategies and capacity theorems," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.
- [47] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorem for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.
- [48] J. Laneman, D. Tse, and G. Wornel, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

- [49] M. Lazaro, M. Sanchez-Fernandez, and A. Artes-Rodriguez, "Optimal sensor selection in binary heterogeneous sensor networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1577–1587, 2009.
- [50] L. B. Le and E. Hossain, "Multihop cellular networks: potential gains, research challenges, and a resource allocation framework," *IEEE Communication Magazine*, vol. 45, no. 9, pp. 66–73, 2007.
- [51] K. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple relay systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3671–3681, 2010.
- [52] L. Liu, X. Zhang, and H. Ma, "Optimal node selection for target localization in wireless camera sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 7, pp. 3562–3576, 2010.
- [53] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int Computer Aided Control Systems Design Symp*, 2004, pp. 284–289.
- [54] R. Louie, Y. Li, and B. Vucetic, "Practical physical layer network coding for two-way relay channels: performance analysis and comparison," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 764–777, 2010.
- [55] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, 2005.
- [56] J. Majchrzak, M. Michalski, and G. Wiczynski, "Distance estimation with a long-range ultrasonic sensor system," *IEEE Sensors Journal*, vol. 9, no. 7, pp. 767–773, 2009.
- [57] J. Matamoros and C. Anton-Haro, "Opportunistic power allocation and sensor selection schemes for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 534–539, 2010.
- [58] MATLAB, *Optimization Toolbox*. The MathWorks Inc.: Natick, MA, U.S.A., 2012.

- [59] Y. Nesterov and A. Nemirovskii, "Interior-point polynomial algorithms in convex programming," *Society for Industrial and Applied Mathematics*, 1994.
- [60] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [61] D. Nguyen, H. Nguyen, and H. Tuan, "Distributed beamforming in relay-assisted multiuser communications," in *Proceedings of IEEE International Conference on Communications, ICC 2009, Dresden, Germany*, 2009, pp. 1–5.
- [62] D. Nguyen and M. Bagajewicz, "Design of nonlinear sensor networks for process plants," *Industrial & Engineering Chemistry Research*, vol. 47, no. 15, pp. 5529–5542, 2008.
- [63] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Distributed beamforming in relay-assisted multiuser communications," in *Proc. of IEEE International Conference on Communications (ICC), Dresden, Germany*, June 2009.
- [64] K. Nishimori, N. Honma, T. Murakami, and T. Hiraguri, "Effectiveness of relay MIMO transmission by measured outdoor channel state information," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 615–623, 2012.
- [65] R. Otafi-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. of 46 Conf. on Decision and Control (CDC)*, 2007, pp. 5492–5498.
- [66] R. Pabst, B. H. Walke, and D. C. Schultz, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, 2004.
- [67] S. A. Pasha, H. D. Tuan, and B.-N. Vo, "Nonlinear Bayesian filtering using the unscented linear fractional transformation model," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 477–489, 2010.
- [68] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. H. Jr., "Relay architectures for 3gpp lte-advanced," *EURASIP J. Wireless Communication and Networking*, vol. 2009, 2009.

- [69] T. Pham-Dinh, N. Nguyen-Canh, and H. A. Le-Thi, "An efficient combined DCA and B&B using DC/SDP relaxation for globally solving binary quadratic programs," *J. of Global Optimization*, vol. 48, pp. 595–632, 2010.
- [70] A. Phan, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Beamforming optimization in multi-user amplify-and-forward wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1510–1520, 2012.
- [71] A. H. Phan, H. D. Tuan, and H. H. Kha, "Optimized solutions for beamforming problems in amplify-and-forward wireless relay networks," in *Proc. of Global Communication Conference (Globecom), Houston, TX, USA, Dec. 2011*.
- [72] —, "Space-time beamforming for multiuser wireless relay networks," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech, May 2011*.
- [73] A. H. Phan, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Iterative D.C. optimization of precoding in wireless MIMO relaying," *to appear in IEEE Transactions on Wireless Communications*, 2013.
- [74] A. Phan, H. Tuan, H. H. Kha, and H. Nguyen, "Beamforming optimization in multi-user amplify-and-forward wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1510–1520, 2012.
- [75] A. Phan, H. Tuan, H. Kha, and D. Ngo, "Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2941–2951, 2012.
- [76] K. T. Phan, L. B. Le, S. A. Vorobyov, and T. Le-Ngoc, "Power allocation and admission control in multiuser relay networks via convex programming: Centralized and distributed schemes," *EURASIP J. Wireless Comm. and Networking*, vol. 2009, 2009.
- [77] G. D. Pillo, *Exact penalty methods in: Algorithms for continuous optimization: the state-of-the-art*, E. Spedicato (ed.). Kluwer Academic Publishers, Boston, 1994.

- [78] C. J. R. Horn, *Matrix analysis*. Cambridge University Press, 1985.
- [79] B. Rankov and A. Wittneben, "Spectral efficient protocols for half- duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, p. 379389, 2007.
- [80] T. Rappaport, *Wireless communications: principles and practice*. Pearson Education, Inc., Upper Saddle River, New Jersey., 2002.
- [81] U. Rashid, H. Tuan, and H. H. Nguyen, "Relay beamforming designs in multi-user wireless relay networks based on throughput maximin optimization," *to appear in IEEE Transactions on Communications*, 2013.
- [82] U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 903–915, 2012.
- [83] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 1970.
- [84] Y. Rong, "Joint source and relay optimization for two-way MIMO relay multi-relay networks," *IEEE Communication Letters*, vol. 15, no. 12, pp. 1329–1351, 2011.
- [85] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier mimo relay communication systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4837–4851, 2009.
- [86] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: system description-Part II: implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1948, 2003.
- [87] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 506–522, 2005.
- [88] Y. Shen, G. Feng, B. Yang, and X. Guan, "Fair resource allocation and admission control in wireless multiuser amplify-and-forward relay networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 1383–1397, 2012.

- [89] V. Shin, G. Shevlyakov, and K. Kim, "A new fusion formula and its application to continuous-time linear systems with multisensor environment," *Computational Statistics and Data Analysis*, vol. 52, pp. 840–854, 2007.
- [90] B. Sklar, *Digital communications*. Prentice Hall, 2001, vol. 2.
- [91] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization Methods Software*, vol. 11-12, pp. 625–653, 1999.
- [92] S.-L. Sun and Z.-L. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, pp. 1017–1023, 2004.
- [93] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [94] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2649–2661, 2008.
- [95] N. N. Tran, H. H. Nguyen, H. D. Tuan, and D. E. Dodds, "Training designs for amplify-and-forward relaying with spatially correlated antennas," *IEEE Transactions on Vehicular Technology*, vol. 61, pp. 2659–2671, 2012.
- [96] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [97] H. D. Tuan, *Gaussian tracking and filtering*. Unpublished Notes.
- [98] H. D. Tuan, P. Apkarian, S. Hosoe, and H. Tuy, "D.C. optimization approach to robust controls: the feasibility problems," *International Journal of Control*, vol. 73, pp. 89–104, Feb. 2000.
- [99] H. D. Tuan, P. Apkarian, T. Narikiyo, and M. Kanota, "New fuzzy control model and dynamic output feedback parallel distributed compensation," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 1, pp. 13–21, 2004.

- [100] H. Tuan, S. Hosoe, and H. Tuy, "Dc optimization approach to robust controls: the optimal scaling value problem," *IEEE Transactions on Automatic Control*, vol. 45, no. 10, pp. 1903–1909, 2000.
- [101] H. D. Tuan, P. Apkarian, and T. Q. Nguyen, "Robust filtering for uncertain nonlinearly parameterized plants," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1806–1815, 2003.
- [102] H. Tuy, *Convex Analysis and Global Optimization*. Kluwer Academic, 1998.
- [103] H. Tuy, P. T. Thach, and H. Konno, "Optimization of polynomial fractional functions," *J. of Global Optimization*, vol. 29, no. 1, pp. 19–44, May 2004.
- [104] H. Tuy and H. D. Tuan, "Generalized S-lemma and strong duality in nonconvex quadratic programming," *J. of Global Optimization*, 2012.
- [105] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. and Control Symp Adaptive Systems for Signal Processing, Communications 2000. AS-SPCC. The IEEE 2000*, 2000, pp. 153–158.
- [106] R. Wang and M. Tao, "Joint source and relay precoding designs for mimo two-way relaying based on mse criterion," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1352–1365, 2012.
- [107] C. Wen, K. Wong, and C. T. K. Ng, "On the asymptotic properties of amplify-and-forward MIMO relay channels," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 590–602, 2011.
- [108] S. Wie, D. L. Goeckel, and M. C. Valenti, "Asynchronous cooperative diversity," *IEEE Transactions on Wireless Communication*, vol. 5, no. 6, pp. 1547–1557, 2006.
- [109] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 161–176.

- [110] A. Willsky, M. Bello, D. Castanon, B. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 799–813, 1982.
- [111] W. Xu, X. Dong, and W. Lu, "Joint precoding optimization for multiuser multi-antenna relaying downlinks using quadratic programming," *IEEE Transactions on Communications*, vol. 59, no. 5, pp. 1228–1235, 2011.
- [112] M. Zeng, R. Zhang, and S. Cui, "On design of collaborative beamforming for two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2284–2295, 2011.
- [113] R. Zhang, C. Chai, and Y. Liang, "Joint beamforming and power control for multiantenna relay broadcast channel with qos constraints," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 726–737, 2009.
- [114] R. Zhang, Y. C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. May, pp. 699–712, 2009.
- [115] F. Zhao and L. Guibas, *Wireless sensor network: an information processing approach*. Morgan Kaufman, San Francisco, CA, 2004.
- [116] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice-Hall, 1996.