

**University of Alberta**

Generalized Quadratically Constrained Quadratic Programming and its  
Applications in Array Processing and Cooperative Communications

by

Arash Khabbazibasmenj

A thesis submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Communications

Department of Electrical and Computer Engineering

©Arash Khabbazibasmenj  
Fall 2013  
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

*To My Beloved Parents and Sister...*

# Abstract

In this thesis, we introduce and solve a particular generalization of the quadratically constrained quadratic programming (QCQP) problem which is frequently encountered in the fields of communications and signal processing. Specifically, we consider such generalization of the QCQP problem which can be precisely or approximately recast as the difference-of-convex functions (DC) programming problem. Although the DC programming problem can be solved through the branch-and-bound methods, these methods do not have any worst-case polynomial time complexity guarantees. Therefore, we develop a new approach with worst-case polynomial time complexity that can solve the corresponding DC problem of a generalized QCQP problem. It is analytically guaranteed that the point obtained by this method satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions. Furthermore, there is a great evidence of global optimality in polynomial time for the proposed method. In some cases the global optimality is proved analytically as well. In terms of applications, we focus on four different problems from array processing and cooperative communications. These problems boil down to QCQP or its generalization. Specifically, we address the problem of transmit beamspace design for multiple-input multiple-output (MIMO) radar in the application to the direction-of-arrival estimation when certain considerations such as enforcement of the rotational invariance property or energy focusing are taken into account. We also study the robust adaptive beamforming (RAB) problem from a new perspective that allows to develop a new RAB method for the rank-one signal model which uses as little as possible and easy to obtain prior information. We also develop a new general-rank RAB method which outperforms other existing state-of-the-art methods. Finally, we concentrate on the mathematical issues of the relay amplification matrix design problem in a two-way amplify-and-forward (AF) MIMO relaying system when the sum-rate, the max-min rate, and the proportional fairness are used as the design criteria.

# Acknowledgements

First of all, I would like to express my sincerest appreciation to my supervisor Dr. Sergiy A. Vorobyov. Without Dr. Vorobyov's precious advice and encouragements, this work would have never been possible. I have been very lucky to have had the opportunity to be Dr. Vorobyov's student. He has always been actively involved in my PhD research and we had discussions in all aspects of my work. I can not describe how much I have learnt from him during my studies.

I would like to thank Dr. Tim Davidson, Dr. Hai Jiang, Dr. Yindi Jing and Dr. Ioanis Nikolaidis for their efforts as my examining committee and for their valuable comments and suggestions.

My great appreciation goes to my colleagues in our research group and in our laboratory. In particular, I would like to thank Dr. Aboulnasr Hassanien for his valuable discussions and ideas which significantly helped me throughout my research, Dr. Omid Taheri, Mahdi Shaghaghi and Jie Gao for their helpful conversations.

My special thanks also goes to my good friends that helped me to experience a very nice and pleasant life during my studies. In particular, I would like to thank Siamak Abdollahi, Arash Talebi, Xi Li, Farzam Nejabatkhah, Omid Taheri, Zohreh Abdyazdan, and Amir Behrouzi whom I spent most of my time with during my PhD studies. Without their friendship, it would have been very difficult to overcome the difficulties of the life.

Last, but not the least, I would like to express my sincerest thanks and appreciation to my parents and my sister. Their never ending love and support has always been the source of any motivation in my life. I can not imagine how my life would have been without their never ending support and care.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Proposed research problems . . . . .	3
1.2	Organization of the thesis . . . . .	8
<b>2</b>	<b>Preliminaries</b>	<b>9</b>
2.1	Convex optimization . . . . .	9
2.1.1	Composition of convex/concave functions . . . . .	12
2.1.2	Optimization problem . . . . .	12
2.1.3	Lagrangian dual functions . . . . .	14
2.1.4	Local optimality conditions . . . . .	15
2.1.5	Optimization problems with generalized inequalities . . . . .	16
2.1.6	Semi-definite programming . . . . .	17
2.1.7	Semi-definite programming relaxation . . . . .	18
2.2	MIMO radar . . . . .	19
2.2.1	Direction of arrival estimation . . . . .	23
2.3	Array processing and beamforming . . . . .	25
2.3.1	Adaptive beamforming . . . . .	26
2.3.2	Robust adaptive beamforming . . . . .	27
2.3.3	General-rank signal model and beamforming . . . . .	31
2.4	Two-way cooperative communications . . . . .	33
<b>3</b>	<b>Generalized Quadratically Constrained Quadratic Programming</b>	<b>36</b>
3.1	Number of quadratic functions in the constraints less than or equal to three . . . . .	38
3.1.1	Polynomial time DC algorithm . . . . .	42
3.2	Number of quadratic functions in the constraints greater than three . . . . .	46
<b>4</b>	<b>Transmit Beamspace Design for DOA Estimation in MIMO Radar</b>	<b>48</b>
4.1	System model . . . . .	51
4.2	Problem formulation . . . . .	54
4.3	Transmit beamspace design . . . . .	55
4.3.1	Two transmit waveforms . . . . .	55
4.3.2	Even number of transmit waveforms . . . . .	58
4.3.3	Optimal rotation of the transmit beamspace matrix . . . . .	61
4.3.4	Spatial-division based design . . . . .	63
4.4	Simulation results . . . . .	64
4.4.1	Example 1 : Effect of the uniform power distribution across the waveforms . . . . .	65
4.4.2	Example 2 : Two separated sectors of width $20^\circ$ degrees each . . . . .	66
4.4.3	Example 3 : Single and centrally located sector of width $10^\circ$ degrees . . . . .	67
4.4.4	Example 4 : Single and centrally located sector of width $30^\circ$ degrees . . . . .	70
4.5	Chapter summary . . . . .	72

<b>5</b>	<b>A New Robust Adaptive Beamforming</b>	<b>74</b>
5.1	The unified principle to MVDR RAB design . . . . .	76
5.2	New beamforming problem formulation . . . . .	77
5.3	Steering vector estimation via SDP relaxation . . . . .	86
5.3.1	SDP relaxation . . . . .	86
5.3.2	Rank of the optimal solution . . . . .	87
5.4	Simulation results . . . . .	88
5.4.1	Example 1 : Exactly known signal steering vector . . . . .	89
5.4.2	Example 2 : Desired signal steering vector mismatch due to wavefront distortion . . . . .	90
5.4.3	Example 3 : Effect of the error in the knowledge of the antenna array geometry . . . . .	91
5.4.4	Example 4 : Desired signal steering vector mismatch due to coherent local scattering [120] . . . . .	93
5.4.5	Example 5 : Comparison with eigenvalue beamforming-based methods of [97] . . . . .	95
5.5	Chapter summary . . . . .	97
<b>6</b>	<b>Robust Adaptive Beamforming for General-Rank Signal Model with Positive Semi-Definite Constraint</b>	<b>100</b>
6.1	Problem formulation . . . . .	102
6.2	New proposed method . . . . .	104
6.2.1	Iterative POTDC algorithm . . . . .	108
6.2.2	Lower-bounds for the optimal value . . . . .	112
6.3	Simulation results . . . . .	113
6.3.1	Example 1 : Parameter mismatch . . . . .	113
6.3.2	Example 2 : Effect of the rank of desired signal covariance matrix . . . . .	115
6.3.3	Example 3 : Distribution mismatch . . . . .	116
6.3.4	Example 4 : Complexity Comparison . . . . .	117
6.4	Chapter summary . . . . .	120
<b>7</b>	<b>Two-Way AF MIMO Relay Amplification Matrix Design</b>	<b>122</b>
7.1	System model . . . . .	124
7.2	Problem statement for sum-rate maximization . . . . .	128
7.3	POTDC algorithm in the sum-rate maximization . . . . .	130
7.4	An upper-bound for the optimal value . . . . .	137
7.5	Proportional fair and max-min rate fair approaches . . . . .	139
7.5.1	Proportional fairness . . . . .	140
7.5.2	Max-min rate fairness . . . . .	142
7.6	Simulation results for sum-rate maximization problem . . . . .	143
7.6.1	Example 1 : Symmetric channel conditions . . . . .	145
7.6.2	Example 2 : Asymmetric channel conditions . . . . .	146
7.6.3	Example 3 : Effect of the number of relay antenna elements . . . . .	147
7.6.4	Example 4 : Performance comparison for the scenario of two-way relaying via multiple single antenna relays . . . . .	148
7.7	Simulation results for PF and MMRF relay amplification matrix design . . . . .	149
7.7.1	Example 1 : Symmetric channel conditions . . . . .	150
7.8	Chapter summery . . . . .	150
<b>8</b>	<b>Conclusion</b>	<b>153</b>
8.1	Summary of contributions . . . . .	153
8.1.1	Transmit beamspace design for DOA estimation in MIMO radar . . . . .	154
8.1.2	A new robust adaptive beamforming . . . . .	154
8.1.3	Robust adaptive beamforming for general-rank signal model with positive semi-definite constraint . . . . .	155
8.1.4	Two-way AF MIMO relay transmit strategy design . . . . .	155
8.1.5	Probable future research . . . . .	156

A Proof of Theorem 5.1	157
B Proof of Lemma 5.1	161
C Proof of Lemma 6.1	163
D Proof of Lemma 6.2	165
E Proof of Lemma 6.3	169
F Proof of Theorem 6.1	173
G Proof of Lemma 7.1	176
H Proof of Lemma 7.2	178
I Rank deficient received signal covariance matrix	179
Bibliography	182

# List of Tables

5.1	Different robust adaptive beamforming methods. . . . .	78
5.2	Width of the feasible set $\Theta_a$ versus the level of perturbations $\alpha$ . . . . .	93
6.1	Average number of the iterations . . . . .	119
6.2	Average CPU time . . . . .	119

# List of Figures

2.1	Simple convex and non-convex sets [70]. . . . .	10
2.2	Convex hull of the non-convex set shown in Fig. 2.1(b) [70]. . . . .	10
2.3	Phased-array radar and MIMO radar. . . . .	20
2.4	A mono-static radar system. . . . .	21
4.1	Example 1: Performance of the new proposed method with and without uniform power distribution across transmit waveforms. . . . .	66
4.2	Example 2: Transmit beampatterns of the traditional MIMO and the proposed transmit beamspace design-based methods. . . . .	68
4.3	Example 2: Individual beampatterns associated with individual waveforms and the overall beampattern. . . . .	68
4.4	Example 2: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods. . . . .	69
4.5	Example 2: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods. . . . .	69
4.6	Example 3: Transmit beampatterns of the traditional MIMO and the proposed transmit beamspace design-based method. . . . .	70
4.7	Example 3: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based method. . . . .	71
4.8	Example 3: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods. . . . .	71
4.9	Example 4: Transmit beampatterns of the traditional MIMO and the proposed methods. . . . .	72
4.10	Example 4: Performance comparison between the traditional MIMO and the proposed methods. . . . .	73
5.1	Squared norm of the projection of vector $\mathbf{d}(\theta)$ on the linear subspaces spanned by the columns of $\mathbf{U}_1$ and $\mathbf{U}_2$ versus $\theta$ . . . . .	80
5.2	Values of the term $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$ in the constraint (5.6) for different angles. . . . .	83
5.3	Comparison between the quadratic term $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$ with and without array perturbations. . . . .	84
5.4	Example 1: Output SINR versus training sample size $K$ for fixed SNR = 20 dB and INR = 30 dB. . . . .	90
5.5	Example 1: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB. . . . .	91
5.6	Example 2: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB. . . . .	92
5.7	Example 3: Output SINR versus training sample size $K$ for fixed SNR = 20 dB and INR = 30 dB for the case of perturbations in antenna array geometry. . . . .	93
5.8	Example 3: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB for the case of perturbations in antenna array geometry. . . . .	94
5.9	Example 4: Output SINR versus training sample size $K$ for fixed SNR = 20 dB and INR = 30 dB. . . . .	95

5.10	Example 4: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB. . . . .	96
5.11	Example 5: Bearing beampattern corresponding to the eigenvalue beamforming method of [97]. . . . .	97
5.12	Example 5: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB. . . . .	98
5.13	Example 5: Output SINR versus SNR for training data size of $K = 30$ and INR = 30 dB. . . . .	99
6.1	Iterative method for minimizing the optimal value function $k(\alpha)$ . The convex optimal value function $l(\alpha, \alpha_0)$ is an upper-bound to $k(\alpha)$ which is tangent to it at $\alpha = \alpha_0$ , and its minimum is denoted as $\alpha_1$ . The point $\alpha_1$ is used to establish another convex upper-bound function denoted as $l(\alpha, \alpha_1)$ and this process continues. . . . .	109
6.2	Example 1: Output SINR versus SNR, INR=30 dB and $K = 50$ . . . . .	114
6.3	Example 1: Objective function value of the problem (6.4) versus SNR, INR=30 dB and $K = 50$ . . . . .	115
6.4	Example 1: Objective function value of the problem (6.4) versus the number of iterations, SNR=15 dB, INR=30 dB and $K = 50$ . . . . .	116
6.5	Example 2: Output SINR versus the actual rank of $\mathbf{R}_s$ , SNR=10 dB, INR=30 dB and $K = 50$ . . . . .	117
6.6	Example 2: Objective function value of the problem (6.4) versus the actual rank of $\mathbf{R}_s$ , SNR=10 dB, INR=30 dB and $K = 50$ . . . . .	118
6.7	Example 3: Actual and presumed angular power densities of general-rank source. . . . .	119
6.8	Example 3: Output SINR versus SNR, INR=30 dB and $K = 50$ . . . . .	120
7.1	Two-way relaying system model. . . . .	125
7.2	Linear approximation of $\ln(\beta)$ around $\beta_c$ . The region above the dashed curve is non-convex. . . . .	136
7.3	Feasible region of the constraint $\ln(\beta) \leq t$ and the convex hull in each sub-division. . . . .	139
7.4	Example 1: The case of symmetric channel conditions. Sum-rate $r_1 + r_2$ versus $\sigma^{-2}$ for $M_R = 3$ antennas. . . . .	145
7.5	Example 2: The case of asymmetric channel conditions. Sum-rate $r_1 + r_2$ versus the distance between the relay and the second terminal $d_2$ for $M_R = 3$ antennas. . . . .	146
7.6	Example 3: The case of asymmetric channel conditions. Sum-rate $r_1 + r_2$ versus the number of the relay antenna elements $M_R$ . . . . .	148
7.7	Example 4: The case of symmetric channel conditions and distributed single antenna relays. Sum-rate $r_1 + r_2$ versus $\sigma^{-2}$ for $M_R = 3$ antennas. . . . .	149
7.8	Example 1: Fairness index versus $\sigma^{-2}$ . . . . .	151
7.9	Example 1: Minimum data rate versus $\sigma^{-2}$ . . . . .	152

# List of Symbols

Symbol	Description	First use
$\mathbb{R}^{N \times M}$	$N \times M$ -dimensional Euclidean space . . . . .	9
$\mathcal{S}_+^N$	Set of $N \times N$ symmetric positive semi-definite matrices . .	11
$\mathcal{H}_+^N$	Set of $N \times N$ Hermitian positive semi-definite matrices . .	11
$(\cdot)^T$	Transpose operator . . . . .	11
$(\cdot)^H$	Hermitian operator . . . . .	11
$\succeq$	Generalized inequality over the cone of positive semi-definite matrices . . . . .	11
$\mathbb{C}^{N \times M}$	$N \times M$ -dimensional complex space . . . . .	11
$\nabla$	Gradient operator . . . . .	11
$\nabla^2$	Hessian matrix operator . . . . .	11
$\ \cdot\ $	Euclidean norm of vector or Frobenius norm of matrix . . .	13
$\text{int}\{\cdot\}$	Interior of a set . . . . .	15
$\text{tr}\{\cdot\}$	Trace operator . . . . .	18
$\text{rank}\{\cdot\}$	Rank operator . . . . .	18
$\mathbf{a}(\theta)$	Transmit array steering vector . . . . .	22
$\otimes$	Kronecker product . . . . .	23
$\mathbb{E}\{\cdot\}$	Statistical expectation . . . . .	27
$\text{Pr}\{\cdot\}$	Probability operator . . . . .	29
$\mathbf{I}$	Identity matrix . . . . .	30

$\mathcal{P}\{\cdot\}$	Eigenvector associated with the largest eigenvalue operator	32
$\mathcal{H}^N$	Set of $N \times N$ Hermitian matrices . . . . .	36
$\text{card}\{\cdot\}$	Cardinality operator . . . . .	38
$(\cdot)'$	Derivative operator . . . . .	46
$\delta\{\cdot\}$	Kronecker delta operator . . . . .	51
$(\cdot)^*$	Conjugate operator . . . . .	51
$\tilde{(\cdot)}$	Flipping operator . . . . .	55
$\text{Re}\{\cdot\}$	Real part operator . . . . .	62
$\lambda_{\max}\{\cdot\}$	Largest eigenvalue operator . . . . .	104
$\mathcal{O}(\cdot)$	Big-O notation . . . . .	112
$\langle \cdot, \cdot \rangle$	Inner product . . . . .	117
$\text{vec}\{\cdot\}$	Vectorization operator . . . . .	125
$\mathbb{R}^+$	Set of positive real number . . . . .	129
$\lambda_{\min}\{\cdot\}$	Smallest eigenvalue operator . . . . .	131
$\text{diag}\{\cdot\}$	Diagonal matrix . . . . .	148
$\text{Null}\{\cdot\}$	Null space . . . . .	179

# List of Abbreviations

<b>Abbreviation</b>	<b>Description</b>	<b>First use</b>
AF	Amplify-and-forward . . . . .	7
ANOMAX	Algebraic norm-maximizing . . . . .	144
CSI	Channel state information . . . . .	7
DC	Difference-of-convex functions . . . . .	3
DF	Decode-and-forward . . . . .	35
DFT	Discrete Fourier transform . . . . .	144
DOA	Direction-of-arrival . . . . .	4
ESPRIT	Estimation of signal parameters via rotational invariance	24
i.i.d.	Independently and identically distributed . . . . .	82
INR	Interference-to-noise ratio . . . . .	89
KKT	Karush-Kuhn-Tucker . . . . .	4
LS	Least squares . . . . .	25
LSMI	Diagonally loaded SMI . . . . .	89
MIMO	Multiple-input multiple-output . . . . .	4
MISO	Multiple-input single-output . . . . .	2
MMRF	Max-min rate fair . . . . .	8
MUSIC	Multiple signal classification . . . . .	24
MV	Minimum variance . . . . .	27

MVDR	Minimum variance distortionless response . . . . .	5
PF	Proportional fairness . . . . .	2
POTDC	Polynomial time DC . . . . .	3
PSD	Positive semi-definite . . . . .	6
QCQP	Quadratically constrained quadratic programming . . . . .	2
QoS	Quality of service . . . . .	7
RAB	Robust adaptive beamforming . . . . .	5
RAGES	Rate-maximization via generalized eigenvectors . . . . .	144
RCS	Radar cross section . . . . .	19
RIP	Rotational invariance property . . . . .	4
RMSE	Root mean square error . . . . .	65
SDD	Spatial-division based design . . . . .	50
SDP	Semi-definite programming . . . . .	17
SDR	Semi-definite relaxation . . . . .	2
SIMO	Single-input multiple output . . . . .	48
SINR	Signal-to-interference-plus-noise ratio . . . . .	3
SISO	Single-input single-output . . . . .	126
SMI	Sample matrix inversion . . . . .	27
SNR	Signal-to-noise ratio . . . . .	7
TWR	Two-way relaying . . . . .	7
ULA	Uniform linear array . . . . .	4

# Chapter 1

## Introduction

In the last decade, there has been a tremendous increase in the number of applications of optimization theory in signal processing and communications. Indeed, due to the ever increasing complexity of electronic systems and networks, the optimization aspects of such systems come at the frontier of signal processing and communications research [1] and [2]. Moreover, many other fundamental signal processing problems such as, for example, subspace tracking [3], parameter estimation [4], and robust design problems [5] can be converted into optimization problems. The complexity of handling an optimization problem generally depends on the properties of the problem. Convex optimization problems form the largest known class of optimization problems that can be efficiently addressed. Once a design problem is cast as a convex optimization problem, it can be considered solved as there are powerful polynomial time numerical methods for solving such problems globally optimally. Moreover, there exists rigorous optimality conditions and a duality theory which can specify the structure of the optimal solution and reveal design insights [6]. The worst-case-based robust adaptive beamforming [7], [8] and optimal power allocation in two hop decode-and-forward relay networks [9] are just two among many examples of signal processing problems that can be cast as convex optimization problems.

As opposed to the convex optimization problems, the non-convex problems are usually extremely hard to deal with. Particularly, unlike the convex problems, there exist no rigorous sufficient globally optimality conditions for such problems. Non-convex problems are very frequently encountered in signal processing and communications applications. For example, the robust adaptive beamforming design problem for the general-rank (distributed source) signal model with a positive semi-definite constraint [10], the power control for wireless cellular systems when the rate

is used as a utility function [11], the dynamic spectrum management for digital subscriber lines [12], the problems of finding the weighted sum-rate optimal point, the proportional fairness (PF) operating point, and the max-min optimal point (egalitarian solution) for the two-user multiple-input single-output (MISO) interference channel [13] are all non-convex programming problems.

Although the non-convex optimization problems are inherently very challenging, it is still possible to solve some of these problems by means of convex optimization techniques. Specifically, it is sometimes possible to relax a non-convex problem into a set of convex problems and then extract the optimal solution of the original problem through the solution of the convexly relaxed problems. Relaxation of a non-convex problem and also the extraction of the optimal solution is typically problem dependent and there is no general universal solution as in convex optimization. There is an extensive research on such relaxations of the non-convex problems with different structures in the optimization theory and thereof applications [14]–[16].

Quadratically constrained quadratic programming (QCQP) problem is one of the important classes of the non-convex optimization problems which is very frequently encountered in different fields of signal processing and communications. Despite being exceedingly difficult, QCQP problems can be approximately solved using the semi-definite programming relaxation (SDR) techniques [14], [17]–[22]. In particular, SDR is a very powerful and computationally efficient method which relaxes the non-convex problem into a convex problem and then extracts a suboptimal solution of the QCQP problem through the optimal solution of the convexly relaxed problem. SDR has been successfully adopted in many practical problems that involve some sort of non-convex QCQP problem [23]–[25]. The authors of [23] consider the application of SDR method for developing an energy-based localization method in wireless sensor networks. The coded waveform design for the radar performance optimization in the presence of colored Gaussian disturbance by means of the SDR for addressing the non-convex QCQP problem has been considered in [24]. Besides, the SDR has been applied for the computationally demanding problem of multiuser detection for decreasing the computational complexity [25].

Despite the profound importance of the QCQP optimization problem and its approximate solution in the related fields, the more general form of these problems have not been studied thoroughly. As an instance, the fractional quadratically constrained quadratic programming problem [26] is an example of such generalization.

Specifically, the QCQP problems can be generalized to include the composition of one-dimensional convex and quadratic functions in the objective and the constraint functions. The motivation behind this generalization is the observation of such composite functions in several interesting problems in signal processing and communications. For example, such important quantities as the transmit power of the multi antenna systems and/or the received power of a multi channel single antenna receiver have quadratic forms with respect to the transmit beamforming vector/precoding matrix [27]. As a result, the transmission rate, which is a common quality of service indicator, and the signal-to-interference-plus-noise ratio (SINR) are in the forms of such compositions. Moreover, the corresponding objective function of the rate allocation schemes based on different criteria such as, for example, the sum-rate maximization [28], proportionally fair, and the max-min fair [29] are one form or another form of the composition of one-dimensional and quadratic functions. The other example is the output power of a minimum variance distortionless response beamformer which is inversely proportional to a quadratic function with respect to beamforming vector. Besides, the uncertainty sets [7] and the robustness constraints [30] are also usually in the forms of the composition of norm and quadratic functions.

## 1.1 Proposed research problems

In this dissertation, we mostly concentrate on a generalization of the QCQP problems which can be precisely or approximately represented as the difference-of-convex functions (DC) programming problems. The existing most typical approaches developed for addressing DC programming problems are based on the so-called branch-and-bound methods [13] and [31]–[36]. However, these methods do not have any (worst-case) polynomial time complexity guarantees which considerably limits or most often prohibits their applicability in practical communication systems. Accordingly, methods with guaranteed polynomial time complexity that can solve such DC programming problems at least suboptimally are of great importance. Thus, we develop a new approach with (worst-case) polynomial time complexity that can solve the corresponding DC problem of a generalized QCQP problem. The proposed method is referred to as the polynomial time DC (POTDC) and is based on semi-definite relaxation, linearization, and an iterative Newton-type search over a small set of parameters of the problem. It is analytically guaranteed that the point

obtained by this method satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions. Furthermore, in the application to the problems addressed in this thesis, there is a great evidence of global optimality for the proposed method. Specifically, this evidence is shown to reduce to the conjecture that can be checked easily numerically. Such check can be viewed as a simple numerical global optimality test as it will be explained later in the thesis.

The problems addressed in this thesis are in the general area of multi-antenna and cooperative systems. Specifically, we have considered the application of SDR and the proposed POTDC method in multiple-input multiple-output (MIMO) radar transmit beamspace design with practical considerations, robust adaptive beamforming design methods for rank-one and general-rank signal models, and the relay amplification matrix design of a two-way relaying system. All these design problems correspond to certain QCQP or its generalizations. A brief description of the tackled problems is as follows.

*Transmit beamspace design in multiple-input multiple-output radar*

MIMO radar is a new emerging technique which offers significant performance improvements compared to its traditional counterpart, that is, single input multiple-output radar [37]. The performance improvements can be attributed to the fact that the transmit signals as well as transmit beamforming techniques in MIMO radar can be chosen/designed quite freely [37], [38]. The additional degrees of freedom provided by MIMO radar allow for improved direction of arrival (DOA) estimation performance using, for example, search-free estimation of signal parameters via rotational invariance techniques [39]. However, the known search-free DOA estimation techniques in application to MIMO radar use the conventional idea of the virtual/receive array partitioning [39]. Thus, for achieving the rotational invariance property (RIP) between different subarrays, either the transmit or receive arrays must be a uniform linear array (ULA). It has been shown in [40], [41], however, the RIP for MIMO radar can be also achieved for transmit and receive arrays of arbitrary geometry in unconventional way through an appropriate design of the transmit beamspace matrix. Specifically, in [41] the idea that the RIP could be enforced at the transmit antenna array was introduced, and a solution utilizing the method of spheroidal sequences was used to obtain  $K$  data sets with RIP. In addition to the RIP due to a proper transmit beamspace design, one can obtain a further DOA estimation performance improvement by focusing the transmit energy on the desired

sectors.

The problem of designing such transmit beamspace matrix which satisfies a number of practical constraints is an open problem of significant interest. In this thesis, we introduce a new approach based on the optimization theory to guarantee the satisfaction of the RIP at the transmit array while other practical requirements are satisfied. Specifically, these requirements are that the transmit beampattern must be as close as possible to the desired beampattern and the transmit power is uniform across the transmit antennas [42], [43]. For the case of even but otherwise arbitrary number of transmit waveforms, the corresponding problem is cast as a non-convex QCQP problem which is solved using SDR. Our numerical results confirm the superiority of the proposed transmit beamspace design method.

#### *Robust adaptive beamforming*

Robust adaptive beamforming (RAB) is one of the classic array processing problems with ubiquitous applicability in wireless communications, radar, sonar, microphone array speech processing, radio astronomy, medical imaging, etc. Various robust adaptive beamforming techniques have been developed in the literature [44] and [45]. Among first robust adaptive beamforming techniques are the diagonal loading [46], [47] and the eigenspace-based beamformers [48]. More recent and more rigorous techniques are the worst-case-based adaptive beamforming [7] and [49]–[51], the probabilistically constrained robust adaptive beamforming [52], doubly constrained robust Capon beamforming [53], [54] and the method of [55] based on steering vector estimation. In general, most of the known minimum variance distortionless response (MVDR) robust adaptive beamforming techniques can be unified under one framework which can be summarized as follows. Use minimum variance distortionless response principle for beamforming vector computation in tandem with sample covariance matrix estimation and steering vector estimation based on some prior information about steering vector. Based on such unified framework to robust adaptive beamforming, we develop a new beamforming technique in which the steering vector is estimated by the beamformer output power maximization under the constraint on the norm of the steering vector estimate and the requirement that the estimate does not converge to an interference steering vector. To satisfy the latter requirement, we develop a new constraint which is different from the one in [55] and is convex quadratic. In general, our new robust adaptive beamforming technique differs from other techniques by the prior information about steering

vector. The prior information used in our technique is only the imprecise knowledge of the antenna array geometry and angular sector in which the actual steering vector lies. Mathematically, the proposed MVDR RAB is expressed as a non-convex QCQP problem with two constraints, which can be efficiently and exactly solved. Some new results for the corresponding optimization problem such as a new algebraic way of finding the rank-one solution from the general-rank solution of the relaxed problem and the condition under which the solution of the relaxed problem is guaranteed to be rank-one are derived. Our simulation results demonstrate the superiority of the proposed method over other previously developed RAB techniques.

*General-rank robust adaptive beamforming*

Most of the robust adaptive beamforming methods have been developed for the case of point source for which the rank of the desired source covariance matrix is equal to one [44] and [45]. However, in some practical applications such as the incoherently scattered signal source or source with fluctuating (randomly distorted) wavefronts, the rank of the source covariance matrix is higher than one [8]. The robust adaptive beamformer for the general-rank signal model that is based on explicit modeling of the error mismatches has been developed in [8]. Despite its simplicity, the robust adaptive beamformer of [8] is known to be overly conservative [10], [56]. Thus, less conservative approaches have been developed in [10] and [56] by adding an additional positive semi-definite (PSD) constraint to the beamformer of [8] which eventually leads to a non-convex optimization problem. The existing approaches for this non-convex problem are either suboptimal iterative methods for which the convergence is not guaranteed or a closed-form solution that may be far from the optimal solution. These shortcomings have motivated us to look for a new efficient and exact solution for the aforementioned non-convex problem. Specifically, the robust adaptive beamforming problem for general-rank signal model with the additional positive semi-definite constraint is shown to belong to the class of the generalized QCQP problems which can be precisely expressed as a DC optimization problem. We solve the corresponding non-convex DC problem by using the proposed POTDC algorithm and give arguments suggesting that the solution is globally optimal. Particularly, we rewrite the problem as the minimization of a one-dimensional optimal value function whose corresponding optimization problem is non-convex. Then, the optimal value function is replaced with another equivalent one, for which the corresponding optimization problem is convex. The new one-dimensional optimal

value function is then minimized iteratively via POTDC algorithm. The solution is guaranteed to satisfy the KKT optimality conditions and there is a strong evidence that such solution is also globally optimal. The new RAB method shows superior performance compared to the other state-of-the-art general-rank RAB methods.

*Two-way relaying transmit strategy design*

Cooperative relay networks enjoy the advantages of the MIMO systems such as, for example, high data rate and low probability of outage without applying multiple antennas at the nodes and by exploiting the inherent spatial diversity [57]–[59]. Although the conventional cooperative relaying systems address the practical problem of packing multiple antennas in low-cost receivers, they are not spectrally efficient. Indeed, due to the orthogonal channel assignments of the cooperative relay networks, the bidirectional data transmission between two different nodes through the relays needs to be accomplished in four time slots [60] which is twice larger than that of the regular transmission. In order to resolve this problem and for a more efficient spectral usage, the two-way relaying (TWR) systems has been recently proposed [61]. The main idea behind the two-way relaying is to reduce the number of the required time slots by relaxing the requirement of orthogonal transmissions between the nodes and the relays [62]. In other words, the simultaneous transmissions by the nodes to the relay on the same frequencies are allowed in the first time slot, while a combined signal is broadcast by the relay in the second time slot.

One fundamental problem associated with TWR systems is the relay transmit strategy design based on the available channel state information (CSI) [61]–[69]. It is usually designed so that a specific performance criterion is optimized under constraints on the available resources and/or quality-of-service (QoS) requirements. The design of rate optimal strategies in two-way relaying systems is one of the most important problems in the area [61]. Although the rate optimal strategy for two-way relaying is in general unknown, the achievable rate for the case of amplify-and-forward (AF) MIMO relaying system with two single antenna users has been discussed in [62]. Moreover, the importance of the user fairness in asymmetric TWR systems has been recently demonstrated in [67], [68], and [69]. The authors of [67] study the optimal power allocation problem for single antenna users and single antenna relay where the sum-rate is maximized under the fairness constraint. Relay beamforming and optimal power allocation for a pair of single antenna users and several single antenna relays based on max-min signal-to-noise ratio (SNR) or

equivalently (data rate) has been also considered in [68] and [69].

In this thesis, we focus on the optimal relay transmit strategy design in a AF TWR system with two single antenna users and one multi-antenna relay based on the maximum sum-rate, PF and the max-min rate fairness (MMRF) criteria. This is a basic model which can be extended in several different ways. First, we consider the relaying transmit strategy or equivalently the relay amplification matrix design when the maximum sum-rate is used as the design criterion. It is shown that the objective function of the sum-rate maximization problem can be represented as a product of quadratic ratios which is a generalized QCQP optimization problem. Using the proposed POTDC algorithm, this problem is precisely expressed as a DC programming problem and then solved to find at least a KKT point. There is a strong evidence, however, that such a point is actually globally optimal. We derive an upper-bound for the optimal value of the corresponding optimization problem and show by simulations that this upper-bound is achieved by POTDC algorithm. It provides an evidence that the algorithm finds a global optimum.

Next, we consider the relay amplification matrix design based on the max-min rate fair and the PF criteria. Similar to the maximum sum-rate problem, we show that the corresponding optimization problems also belong to the class of the generalized QCQP optimization problems which can be precisely recast as DC problems. The corresponding DC problems are also efficiently addressed by using POTDC method. In addition, based on our numerical results, the POTDC method always results in the globally optimal solution of the corresponding optimization problem. The global optimality of the resulted solution is equivalent to the concavity of a certain optimal value function which can be easily checked numerically.

## 1.2 Organization of the thesis

This thesis is organized as follows. Preliminaries are given in Chapter 2 while a generalization of QCQP and its solution is discussed in Chapter 3. The proposed transmit beamspace design for DOA estimation in MIMO radar with considering the practical considerations is given in Chapter 4. In Chapters 5 and 6, respectively, the robust adaptive beamforming design problem for a rank-one and general-rank signal models are developed and investigated. Finally, Chapter 7 studies the relay amplification matrix design in two-way relaying systems. Our conclusions and future potential research directions are given in Chapter 8.

# Chapter 2

## Preliminaries

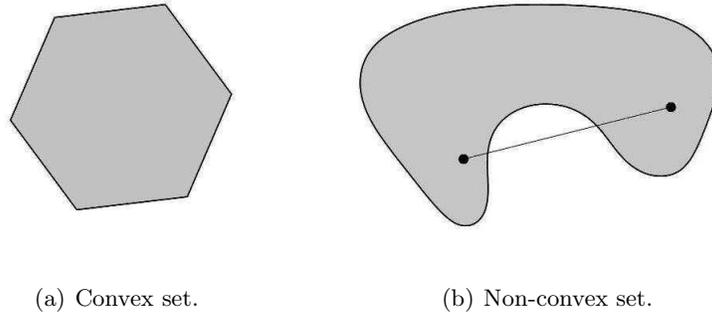
In this chapter, we first review the basics of the convex optimization theory and particularly the semi-definite programming. These principles will be required in Chapter 3 for introducing a general form of optimization problems that are frequently encountered in different fields of communications and signal processing. Next, we will briefly review the preliminaries of the tackled problems. Specifically, the principles of MIMO radar, robust adaptive beamforming, and two-way cooperative relay networks are shortly introduced.

### 2.1 Convex optimization

Let us consider any two arbitrary points  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$  where  $\mathbf{x}_1 \neq \mathbf{x}_2$  and  $\mathbb{R}^{N \times M}$  denotes the  $N \times M$ -dimensional Euclidean space. Note that for simplicity, the space of  $\mathbb{R}^{N \times 1}$  will be denoted as  $\mathbb{R}^N$  hereafter. The points of the form  $\mathbf{y}(\theta) = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$  where  $\theta \in \mathbb{R}$  correspond to the line which connects the points  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in  $\mathbb{R}^N$ . Moreover, the line segment between  $\mathbf{x}_1$  and  $\mathbf{x}_2$  corresponds to the case where  $\theta$  is restricted to lie between 0 and 1.

*Convex set:* The set  $C \subset \mathbb{R}^N$  is defined to be convex if the line segment between any two arbitrary points in  $C$  is a subset of  $C$ . More specifically, if for any arbitrary points  $\mathbf{x}_1, \mathbf{x}_2 \in C$  and  $0 \leq \theta \leq 1$ , the point  $\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$  belongs to  $C$ , then  $C$  is a convex set.

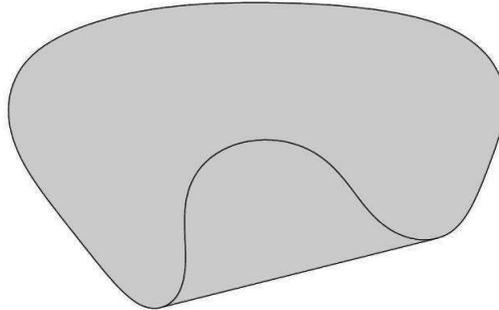
Fig. 2.1 shows two example sets in the Euclidean space of  $\mathbb{R}^2$ . It is easy to observe that the line segment between any two arbitrary points in the set of Fig. 2.1(a) lies inside this set and as a result it is convex. However, the set shown in Fig. 2.1(b) is non-convex because the connecting line segment between the shown two points does not lie inside this set.



**Figure 2.1:** Simple convex and non-convex sets [70].

*Convex hull:* The convex hull of a set  $C$  is defined as the set of all points that are in the form of  $\sum_{i=1}^K \theta_i \mathbf{x}_i$  referred to as the *convex combination* of the points  $\mathbf{x}_i \in C, i = 1, \dots, K$  where  $\sum_{i=1}^K \theta_i = 1, \theta_i \geq 0$  and  $K \geq 1$ . The convex hull of a set is basically the smallest convex set which contains that set [70].

From the definition, it can be immediately concluded that the convex hull of a convex set is equal to itself. Fig. 2.2 shows the convex hull of the non-convex set shown in Fig. 2.1(b).



**Figure 2.2:** Convex hull of the non-convex set shown in Fig. 2.1(b) [70].

*Cone:* The set  $C \subset \mathbb{R}^N$  is said to be a cone if for any arbitrary  $\mathbf{x} \in C$  and  $\theta \geq 0$ , the new point  $\theta \mathbf{x}$  also belongs to  $C$ .

A set is said to be a *convex cone* if it is a cone and at the same time it is convex. In other words,  $C$  is a convex cone if and only if for any arbitrary points  $\mathbf{x}_1, \mathbf{x}_2 \in C$  and  $\theta_1 \geq 0$  and  $\theta_2 \geq 0$  the point  $\theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2$  also belongs to  $C$  [70].

The set of symmetric and Hermitian positive semi-definite matrices which are

denoted as  $\mathcal{S}_+^N$  and  $\mathcal{H}_+^N$ , respectively, are among the important convex cones and are defined as

$$\mathcal{S}_+^N = \{\mathbf{X} \in \mathbb{R}^{N \times N} \mid \mathbf{X} = \mathbf{X}^T \text{ and } \mathbf{X} \succeq \mathbf{0}\} \quad (2.1)$$

and

$$\mathcal{H}_+^N = \{\mathbf{X} \in \mathbb{C}^{N \times N} \mid \mathbf{X} = \mathbf{X}^H \text{ and } \mathbf{X} \succeq \mathbf{0}\} \quad (2.2)$$

where  $(\cdot)^T$  and  $(\cdot)^H$  stands for the transpose and Hermitian operators, respectively,  $\mathbf{A} \succeq \mathbf{B}$  means that  $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$  is a PSD matrix and  $\mathbb{C}^{N \times M}$  denotes the  $N \times M$ -dimensional complex space. For simplicity, the space of  $\mathbb{C}^{N \times 1}$  will be denoted hereafter as  $\mathbb{C}^N$ .

*Convex function:* The function  $f : \mathcal{D} \subset \mathbb{R}^N \rightarrow \mathbb{R}$  is said to be convex if its domain,  $\mathcal{D}$ , is convex and for all  $\mathbf{x}, \mathbf{y} \in \mathcal{D}$  and  $0 \leq \theta \leq 1$  the following inequality holds

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}). \quad (2.3)$$

Similarly, the function  $f$  is concave, if  $-f$  is convex [70].

For the differentiable function  $f(\mathbf{x})$ , the convexity is equivalent to the following inequality

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{D} \quad (2.4)$$

where the vector  $\nabla f(\mathbf{x})$  denotes the gradient of  $f(\mathbf{x})$  at  $\mathbf{x}$ . Moreover, a twice differentiable function  $f(\mathbf{x})$  is convex if and only if its domain is convex and its Hessian denoted as  $\nabla^2 f(\mathbf{x})$  is positive semi-definite for all  $\mathbf{x} \in \mathcal{D}$  [70]. A linear function in the form of  $\mathbf{a}^T \mathbf{x} + b$  where  $\mathbf{a}, \mathbf{x} \in \mathbb{R}^N$  is an example of a convex function and a quadratic function of the form of  $\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{a}^T \mathbf{x} + b$  where  $\mathbf{A} \in \mathbb{R}^{N \times N}$  and  $\mathbf{a}, \mathbf{x} \in \mathbb{R}^N$  is convex if and only if the matrix  $\mathbf{A}$  is positive semi-definite [70].

Two important property of the convex functions is as follows

- Pointwise maximum of the convex functions  $f_i(\mathbf{x}), i = 1, \dots, k$  defined as  $g(\mathbf{x}) = \max_i f_i(\mathbf{x})$  is a convex function.
- The  $\alpha$ -sublevel set of a convex function  $f(\mathbf{x})$  defined as  $C_\alpha = \{\mathbf{x} \mid f(\mathbf{x}) \leq \alpha\}$  is a convex set.

### 2.1.1 Composition of convex/concave functions

In the next chapters of this thesis, we will encounter the functions which are the compositions of convex and concave functions. In this part, we will briefly review the different possible cases and the sufficient conditions which guarantee the composition function to be convex/concave. Let us consider the functions  $h : \mathcal{D}_h \subset \mathbb{R}^k \rightarrow \mathbb{R}$  and  $g : \mathcal{D}_g \subset \mathbb{R}^n \rightarrow \mathbb{R}^k$  and define their decomposition as  $f \triangleq h \circ g = h(g(\mathbf{x})) = h(g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_k(\mathbf{x}))$  where  $g_i : \mathcal{D}_g \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, \dots, k$  and  $\mathcal{D}_f \triangleq \{\mathbf{x} \mid g(\mathbf{x}) \in \mathcal{D}_h, \mathbf{x} \in \mathcal{D}_g\}$ . Depending on the monotonicity and convexity/concavity of the functions  $g(\mathbf{x})$  and  $h(\mathbf{x})$ , the following sufficient conditions hold true [70]

- $f$  is convex if  $h$  is convex,  $h$  is non-decreasing in each argument, and  $g_i, i = 1, \dots, k$  are convex.
- $f$  is convex if  $h$  is convex,  $h$  is non-increasing in each argument, and  $g_i, i = 1, \dots, k$  are concave.
- $f$  is concave if  $h$  is concave,  $h$  is non-decreasing in each argument, and  $g_i, i = 1, \dots, k$  are concave.
- $f$  is concave if  $h$  is concave,  $h$  is non-increasing in each argument, and  $g_i, i = 1, \dots, k$  are convex.

### 2.1.2 Optimization problem

An optimization problem has the following generic form

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p. \end{aligned} \tag{2.5}$$

Problem (2.5) describes the problem of minimizing the function  $f_0(\mathbf{x})$  among all  $\mathbf{x}$  which satisfy the constraints  $f_i(\mathbf{x}) \leq 0$ ,  $i = 1, \dots, m$  and  $h_i(\mathbf{x}) = 0$ ,  $i = 1, \dots, p$ .

The function  $f_0(\mathbf{x})$  is called the objective function while the functions  $f_i$ ,  $i = 1, \dots, m$  and  $h_i$ ,  $i = 1, \dots, p$  are called the inequality constraint and the equality constraint functions, respectively. Domain of the optimization problem (2.5), denoted as  $\mathcal{D}$ , is defined as the intersection of the domain of the objective function and all the constraint functions. The point  $\mathbf{x} \in \mathcal{D}$  is said to be a feasible point

of the optimization problem (2.5), if it satisfies all the equality and inequality constraints. The set of all the feasible points is referred to as the feasible set. The feasible point  $\mathbf{x}_{\text{opt}}$  is said to be the globally optimal solution of the problem (2.5), if  $p^* \triangleq f_0(\mathbf{x}_{\text{opt}}) \leq f_0(\mathbf{x})$  for any feasible point  $\mathbf{x}$ . Furthermore, the point  $\mathbf{x}^*$  is said to be locally optimal, if there exists  $\epsilon > 0$  for which the inequality  $f(\mathbf{x}^*) \leq f(\mathbf{x})$  holds for all  $\mathbf{x}$  such that  $\|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon$  where  $\|\cdot\|$  stands for the Euclidean norm of a vector or Frobenius norm of a matrix.

One of the most important classes of optimization problems is the QCQP. The general form of a QCQP problem is as follows

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^H \mathbf{A}_0 \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^H \mathbf{A}_i \mathbf{x} \leq \alpha_i, \quad i = 1, \dots, m \\ & \mathbf{x}^H \mathbf{B}_j \mathbf{x} = \beta_j, \quad j = 1, \dots, p \end{aligned} \tag{2.6}$$

where the matrices  $\mathbf{A}_i, i = 0, \dots, m$  and  $\mathbf{B}_j, j = 1, \dots, p$  are real symmetric or Hermitian.

The complexity of addressing an optimization problem generally depends on its properties. Convex optimization problems form the largest known class of optimization problems that can be efficiently addressed. Compared to the convex problems, non-convex problems are generally hard to solve and there is no general efficient algorithm to solve them [6]. In what follows, the definition of a convex problem is explained in details.

*Convex Optimization:*

The optimization problem (2.5) is said to be a convex problem if the objective function and all the inequality constraint functions,  $f_i, i = 0, \dots, m$ , are convex with respect to  $\mathbf{x}$  while all the equality constraint functions,  $h_i, i = 1, \dots, p$ , are affine with respect to  $\mathbf{x}$ . In a convex optimization problem, the feasible set is convex. The latter is due to the fact that the feasible set of the problem (2.5) is the intersection of the convex feasible set of the constraints and as a result it is convex.

The QCQP problem (2.6), which commonly appears in signal processing and communications applications, is generally a non-convex problem unless, there is only inequality constraints and all the matrices  $\mathbf{A}_i, i = 0, \dots, m$  are positive semi-definite. As a result, it is usually hard to deal with such problems.

### 2.1.3 Lagrangian dual functions

In optimization theory, the so-called dual problem provides a different perspective to the optimization problems. The essence of the definition of dual problem is to utilize the Lagrangian dual function as it is shortly explained.

*Lagrangian:* The corresponding Lagrangian of the optimization problem (2.5) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x}) \quad (2.7)$$

where  $\lambda_i$  is called the Lagrangian multiplier associated with the inequality constraint  $f_i(\mathbf{x}) \leq 0$ ,  $\nu_i$  is the Lagrangian multiplier that corresponds to the equality constraint  $h_i(\mathbf{x}) = 0$ , and  $\boldsymbol{\lambda} \triangleq (\lambda_1, \dots, \lambda_m)$  and  $\boldsymbol{\nu} \triangleq (\nu_1, \dots, \nu_p)$  are the set of Lagrangian multipliers.

*The Lagrangian dual function:* For a fixed value of the Lagrangian multipliers  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$ , the Lagrangian dual function associated with the optimization problem (2.5) is defined as

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &\triangleq \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &\triangleq \min_{\mathbf{x} \in \mathcal{D}} f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x}) \end{aligned} \quad (2.8)$$

where  $\mathcal{D}$  is the domain of the optimization problem (2.5). The Lagrangian dual function is only defined for the such values of  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$  for which  $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$  is finite. The Lagrangian multipliers  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$  are said to be dual feasible if  $\lambda_i \geq 0$ ,  $i = 1, \dots, m$  and  $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$  is finite. It should be emphasized that since the Lagrangian dual function is the pointwise minimum of a family of the functions that are linear with respect to  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$ , it is always concave with respect to  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$  regardless of the primal problem (2.5).

For any dual feasible variables  $\boldsymbol{\lambda}$  and  $\boldsymbol{\nu}$ , the Lagrangian dual function  $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$  gives a lower-bound for the optimal value of the optimization problem (2.5), i.e.,  $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$  [70]. By utilizing the latter fact, one can find the tightest lower-bound for the optimization problem (2.5) as it follows

*Lagrangian dual problem:* Lagrangian dual problem is the problem of finding the best lower-bound for the optimal value of the problem (2.5) using the Lagrangian dual function (2.8). Lagrangian dual problem can be expressed as the following optimization problem

$$\begin{aligned}
d^* \triangleq & \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\
\text{subject to} & \lambda_i \geq 0, \quad i = 1, \dots, m.
\end{aligned} \tag{2.9}$$

Note that since the Lagrangian dual function (2.8) is concave, the Lagrangian dual problem (2.9) is convex even if the primal problem is not convex. As it was mentioned before the Lagrangian dual function gives a lower-bound for the optimal value of the primal problem (2.5) and as a result, it can be concluded that  $d^* \leq p^*$ , where this property is referred to as the weak duality. The difference between the  $p^*$  and  $d^*$  is called the duality gap between the primal and dual problems.

Duality gap is generally nonzero. When duality gap is zero, i.e.,  $d^* = p^*$ , it is said that the strong duality holds. Convex optimization problems are a big category of the problems for which the strong duality holds under some mild conditions. More specifically, if the problem (2.8) is convex and it satisfies certain constraint qualifications, it is convex. One of the simple constraint qualifications is the Slater's condition. Slater's condition holds if there exists a feasible point  $\mathbf{x} \in \text{int}\{\mathcal{D}\}$  for which all the inequalities  $f_i(\mathbf{x}) < 0, i = 1, \dots, m$  hold true and  $\text{int}\{\cdot\}$  denotes the interior of a set [70].

#### 2.1.4 Local optimality conditions

In this part, we will discuss the necessary local optimality conditions for the problem (2.5) which is not restricted to be convex in general. For this goal, we need to introduce the notion of regularity. The feasible point  $\mathbf{x}$  is said to be a regular point if the gradients of the equality constraints, i.e., the gradients  $\nabla h_i(x), i = 1, \dots, p$ , and the active inequality constraints, i.e., the gradients  $\nabla f_i(x), i \in \{l \mid f_l(x) = 0, i = 1, \dots, m\}$ , at this point are all linearly independent.

Let the point  $\mathbf{x}^*$  be a feasible point of the problem (2.5). The necessary condition for the regular point  $\mathbf{x}^*$  to be locally optimal is that there exist Lagrangian

multipliers  $\lambda^*$  and  $\nu^*$  such that the following conditions hold true

$$\begin{aligned}
f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \dots, m \\
h_i(\mathbf{x}^*) &= 0, \quad i = 1, \dots, p \\
\lambda_i^* &\geq 0, \quad i = 1, \dots, m \\
\lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \dots, m \\
\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(\mathbf{x}^*) &= 0
\end{aligned} \tag{2.10}$$

which are called the KKT conditions [70]. It is noteworthy to mention that if the optimization problem (2.5) is convex and the strong duality holds then the KKT conditions are the necessary and sufficient conditions for the point  $\mathbf{x}^*$  to be globally optimal.

### 2.1.5 Optimization problems with generalized inequalities

A straightforward generalization of the optimization problem (2.5) is obtained by allowing generalized inequalities. A generalized inequality is a partial ordering in the space of  $\mathbb{R}^N$  rather than  $\mathbb{R}$ . Moreover, generalized inequalities are defined over specific cones that are referred to as the proper cones. The convex cone  $\mathcal{K} \subset \mathbb{R}^N$  is said to be proper, if it has the following properties

- it is closed, in other words, it includes its boundary.
- It does not have any non-empty interior.
- if  $\mathbf{x} \in \mathcal{K}$  and  $-\mathbf{x} \in \mathcal{K}$  then  $\mathbf{x} = \mathbf{0}$ .

The partial ordering over the proper cone  $\mathcal{K}$  is defined as

$$\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y} \quad \text{if and only if} \quad \mathbf{y} - \mathbf{x} \in \mathcal{K} \tag{2.11}$$

Allowing the generalized inequalities, the generic optimization problem (2.5) can be generalized as

$$\begin{aligned}
\min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & \mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}, \quad i = 1, \dots, m \\
& h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p
\end{aligned} \tag{2.12}$$

where  $\mathcal{K}_i \subset \mathbb{R}^{N_i}, i = 1, \dots, m$  are proper cones and  $\mathbf{f}_i : \mathcal{D}_i \subset \mathbb{R}^N \longrightarrow \mathbb{R}^{N_i}, i = 1, \dots, m$  are multi-valued functions.

In order to define the convexity of an optimization problem with generalized inequalities, the definition of convex functions is required to be extended to multi-valued functions. In what follows a convex multi-valued function is defined.

$\mathcal{K}_i$ -convexity: Let us consider the proper cone  $\mathcal{K}_i \subset \mathbb{R}^{N_i}$  with the corresponding generalized inequality  $\preceq_{\mathcal{K}_i}$ . The multi-valued function  $\mathbf{f}_i : \mathcal{D}_i \subset \mathbb{R}^N \longrightarrow \mathbb{R}^{N_i}$  is said to be  $\mathcal{K}_i$ -convex, if for all  $\mathbf{x}, \mathbf{y} \in \mathcal{D}_i$  and  $0 \leq \theta \leq 1$ , the following inequality holds

$$\mathbf{f}(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \preceq_{\mathcal{K}_i} \theta \mathbf{f}(\mathbf{x}) + (1 - \theta) \mathbf{f}(\mathbf{y}) \quad (2.13)$$

Based on the new definition, the optimization problem (2.12) is said to be convex if  $f_0(\mathbf{x})$  is convex,  $\mathbf{f}_i(\mathbf{x})$  is  $\mathcal{K}_i$ -convex and all the equality constraints are affine.

The Lagrangian for an optimization problem with generalized inequality is similarly defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_m, \boldsymbol{\nu}) \triangleq f_0(\mathbf{x}) + \sum_{i=1}^m \boldsymbol{\lambda}_i^T \mathbf{f}_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x}) \quad (2.14)$$

where  $\boldsymbol{\lambda}_i \in \mathbb{R}^{N_i}$  is the Lagrangian multiplier associated with the generalized inequality constraint  $\mathbf{f}_i(\mathbf{x}) \preceq_{\mathcal{K}_i} \mathbf{0}$ ,  $\nu_i$  is the Lagrangian multiplier associated with the equality constraint  $h_i(\mathbf{x}) = 0$  and  $\boldsymbol{\nu} \triangleq (\nu_1, \dots, \nu_p)$ .

The Lagrangian dual function (2.8), the Lagrangian dual problem (2.9) are defined very similarly and the weak duality and strong duality results are also valid for an optimization problem with generalized inequality [70].

### 2.1.6 Semi-definite programming

In this thesis, we will frequently encounter the optimization problems with generalized inequality constraints over the cone of symmetric positive semi-definite matrices  $\mathcal{S}_+^N$  or the cone of Hermitian positive semi-definite matrices  $\mathcal{H}_+^N$  whose objective function is linear and there are other linear constraints. The general form of these problems is as follows and they are referred to as the semi-definite programming problems (SDP)

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{tr}\{\mathbf{A}_0 \mathbf{X}\} \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}_i \mathbf{X}\} \leq \alpha_i, \quad i = 1, \dots, m \\ & \text{tr}\{\mathbf{B}_j \mathbf{X}\} = \beta_j, \quad j = 1, \dots, p \\ & \mathbf{X} \succeq \mathbf{0} \end{aligned} \quad (2.15)$$

where  $\text{tr}\{\cdot\}$  stands for the trace operator and the matrices  $\mathbf{A}_i, i = 0, \dots, m$  and  $\mathbf{B}_j, j = 1, \dots, p$  are real symmetric or Hermitian matrices. Semi-definite programming is very useful for solving the non-convex QCQP problems approximately through the semi-definite programming relaxation as it is explained shortly.

### 2.1.7 Semi-definite programming relaxation

As it was mentioned earlier, QCQP optimization problems are generally non-convex. One of the most efficient ways of dealing with such generally NP-hard non-convex problems is through their approximation with SDP problems [17]–[22]. For this goal, let us consider the following general QCQP problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^H \mathbf{A}_0 \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^H \mathbf{A}_i \mathbf{x} \leq \alpha_i, \quad i = 1, \dots, m \\ & \mathbf{x}^H \mathbf{B}_j \mathbf{x} = \beta_j, \quad j = 1, \dots, p. \end{aligned} \tag{2.16}$$

By defining the new additional variable  $\mathbf{X} \triangleq \mathbf{x}\mathbf{x}^H$  and considering the facts that  $\mathbf{X} = \mathbf{x}\mathbf{x}^H$  implies that  $\mathbf{X}$  is a rank-one positive semi-definite matrix and also the fact that for any arbitrary matrix  $\mathbf{D}$ , the following relationship holds  $\mathbf{x}^H \mathbf{D} \mathbf{x} = \text{tr}\{\mathbf{D}\mathbf{x}\mathbf{x}^H\}$ , the QCQP problem of (2.16) can be equivalently expressed as

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{tr}\{\mathbf{A}_0 \mathbf{X}\} \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}_i \mathbf{X}\} \leq \alpha_i, \quad i = 1, \dots, m \\ & \text{tr}\{\mathbf{B}_j \mathbf{X}\} = \beta_j, \quad j = 1, \dots, p \\ & \mathbf{X} \succeq \mathbf{0}, \quad \text{rank}\{\mathbf{X}\} = 1 \end{aligned} \tag{2.17}$$

where  $\mathbf{X}$  is a real symmetric or Hermitian matrix and  $\text{rank}\{\cdot\}$  denotes the rank of a matrix. The only non-convex constraint in (2.17) is the rank-one constraint while all other constraint functions and the objective function are linear with respect to  $\mathbf{X}$ . The non-convex rank constraint is the main difficulty of QCQP problems. The SDP relaxation is based on approximating the problem (2.17) by dropping the non-convex rank constraint and trying to somehow extract the solution of the original problem from the solution of the relaxed problem [14], [17]–[22]. By dropping the

non-convex rank-one constraint the following SDP problem is resulted

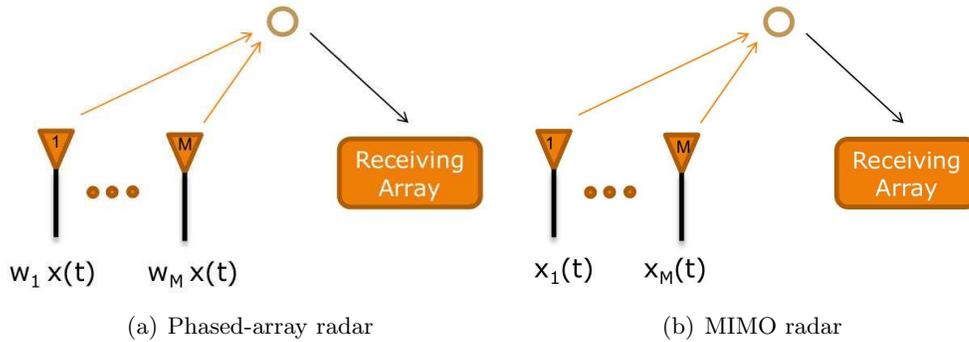
$$\begin{aligned}
& \min_{\mathbf{X}} && \text{tr}\{\mathbf{A}_0\mathbf{X}\} \\
& \text{subject to} && \text{tr}\{\mathbf{A}_i\mathbf{X}\} \leq \alpha_i, \quad i = 1, \dots, m \\
& && \text{tr}\{\mathbf{B}_j\mathbf{X}\} = \beta_j, \quad j = 1, \dots, p. \\
& && \mathbf{X} \succeq \mathbf{0}.
\end{aligned} \tag{2.18}$$

When the relaxed problem (2.18) is solved, extracting the approximate solution of the original problem (2.17) through the optimal solution of the relaxed problem (2.18) is done through randomization techniques. In particular, let  $\mathbf{X}_{\text{opt}}$  denote the optimal solution of the relaxed problem (2.18). If the rank of  $\mathbf{X}_{\text{opt}}$  is one, then the optimal solution of the original problem (2.16) can be obtained by simply finding the principal eigenvector of  $\mathbf{X}_{\text{opt}}$ . However, if the rank of the matrix  $\mathbf{X}_{\text{opt}}$  is higher than one, we need to resort to the randomization techniques to extract an approximate solution. A number of different randomization techniques has been developed in the literature [14]. Briefly, the essence of such techniques is to generate first a set of candidate vectors using  $\mathbf{X}_{\text{opt}}$  and then to choose the best vector among all candidate vectors.

In the next section, the essences of the MIMO radar will be introduced.

## 2.2 MIMO radar

Radar systems can be used for detecting and measuring the parameters such as velocity, acceleration, range/location, and radar cross section (RCS) of moving targets. The essence of any radar systems is to transmit energy to the specific directions of the space and then to process the received echo from the targets to detect or estimate the parameters of interest [71]–[73]. The radar systems are usually formed by a transmit and a receive array. The transmit array is utilized for transmitting specific waveforms towards the targets in the space while the receive array collects the reflected echo from the targets for further processing. Depending on the distance between the transmit and receive arrays, the radar systems are classified as mono-static and bi-static [74]. Particularly, in a mono-static radar, the distance between the transmit and receive arrays is comparatively small and the targets in the far-field are viewed from the same angle while in a bi-static radar, the array are widely separated and they view the targets from different angles.



**Figure 2.3:** Phased-array radar and MIMO radar.

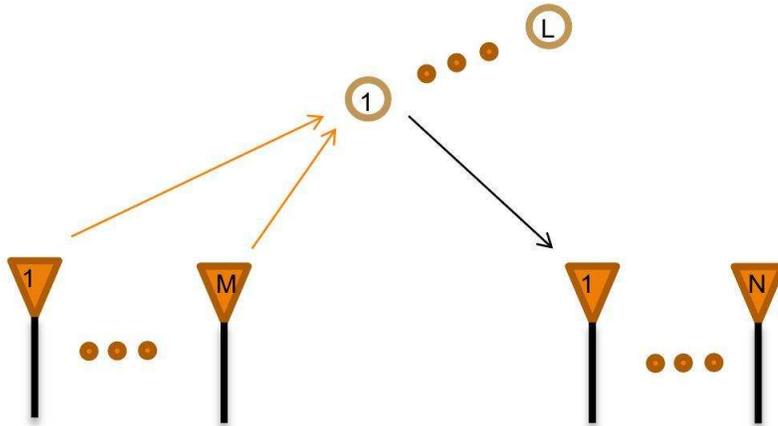
The essence of the conventional radar systems referred to as the phased-array radar is to exploit the transmit and receive array coherent processing gains [71]. More specifically, phased-array radar coheres beams towards the specific directions of the space by transmitting a single waveform from different antenna elements whose phase has been properly rotated. By such phase rotations, the received signal reflected by the desired targets can be added coherently which can give rise to the coherent processing gains at the receive array.

MIMO radar is a new emerging technology [41], [75]–[80]. As apposed to the phased-array radar, the fundamental and enabling concept of the MIMO radar is the waveform diversity. Specifically, compared to the phased-array radar that transmits scaled versions of a single waveform, in MIMO radar, the transmitted waveforms can be chosen quite freely and only are restricted by hardware issues (see Fig. 2.3). Due to the so provided additional degrees of freedom which is referred to as the waveform diversity, MIMO radar may show performance improvements compared to the phased-array radar systems [75]–[80].

The transmit/receive array in MIMO radar can be either widely separated or colocated. In a MIMO radar with widely separated antennas, the antenna elements view different aspects of the target [76]. In other words, the RCS factor of each antenna element is independent from that of the other antenna elements. The waveform diversity in this case is very similar to the concept of the multi-path diversity in wireless communication channels [75]. The multi-path diversity of a wireless communication system is usually realized by transmitting the same source data to the destination through several independently faded channels [81]. Since the probability of an unfavorable fading in all the channels is negligible, this form

of diversity allows more reliable decoding of the transmitted information. In a very similar way, the waveform diversity of a widely separated MIMO radar allows more reliable detection as the probability of simultaneous scintillation for all the transmit antennas is very small [75]. Many works have recently been reported in the literature that demonstrate the benefits of applying the MIMO radar concept using widely separated antennas [75]–[78].

MIMO radar with colocated antennas consists of closely spaced transmit antennas [79]. Compared to the widely separated MIMO radar, the antennas cohere the beam to a certain direction in the space. Using the colocated antennas, the virtual aperture of the received antenna array can be significantly increased [41]. The increased virtual array size results in improvement in the maximum number of the targets that can be uniquely identified, enhancement in the angular resolution and the parameter estimation performance [37]. In this dissertation, we mostly concentrate on the MIMO radar with colocated antennas. In order to explain why the corresponding aperture size of a colocated MIMO radar is increased, let us consider the mono-static radar in Fig. 2.4 with the transmit and receive arrays of size  $M$  and  $N$ , respectively.



**Figure 2.4:** A mono-static radar system.

It is assumed that the transmitted signals from the transmit array are all narrow-band and orthogonal. Moreover, the point source signal model is adopted for the targets. Let  $x_m(t), m = 1, \dots, M$  denote the transmitted signal from antenna element  $m$  at time instant  $t$ , which has unit energy, i.e.,  $\int_T x_m^2(t)dt = 1$ , where  $T$  is the signal duration. Under the condition that the channel is non-dispersive, the received sig-

nal at a target with location parameter  $\theta$  can be written as  $\sum_{m=1}^M e^{-j2\pi f_0 \tau_m(\theta)} x_m(t)$  where  $f_0$  is the carrier frequency and  $\tau_m(\theta)$  is the time needed for the emitted signal from the antenna  $m$  to be received at a target whose location parameter is  $\theta$  [37]. By defining following vectors

$$\mathbf{a}(\theta) = [e^{-j2\pi f_0 \tau_1(\theta)}, e^{-j2\pi f_0 \tau_2(\theta)}, \dots, e^{-j2\pi f_0 \tau_M(\theta)}]^T \quad (2.19)$$

and

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)] \quad (2.20)$$

the received signal at a target with location parameter  $\theta$  can be expressed in a closed-form as  $\mathbf{a}^T(\theta)\mathbf{x}(t)$ . The vector  $\mathbf{a}(\theta)$  is a known function of  $\theta$  which depends on the transmit antenna array geometry and it is referred to as the steering vector of the transmit array. Assuming that  $L$  targets are present, the echoes which are reflected from the targets and are received at the receive array can be written as

$$\mathbf{y}(t, \varrho) = \sum_{l=1}^L \beta_l(\varrho) \mathbf{b}(\theta_l) \mathbf{a}^T(\theta_l) \mathbf{x}(t) + \mathbf{z}(t, \varrho) \quad (2.21)$$

where  $t$  is the fast time index, i.e., time within a frame,  $\varrho$  is the slow time index, i.e., number of pulses,  $\beta_l(\varrho)$  is the reflection coefficient of the target located at the angle  $\theta_l$  with variance  $\sigma_\beta^2$ ,  $\mathbf{b}(\theta) = [e^{-j2\pi f_0 \tilde{\tau}_1(\theta)}, e^{-j2\pi f_0 \tilde{\tau}_2(\theta)}, \dots, e^{-j2\pi f_0 \tilde{\tau}_N(\theta)}]^T$  is the steering vector of receive array,  $\tilde{\tau}_k(\theta)$  is the time needed for the reflected signal to propagate from the target at location  $\theta$  to the  $k$ th receive antenna element, and  $\mathbf{z}(t, \varrho)$  is the  $N \times 1$  vector of zero-mean white Gaussian noise.

Since the transmit signals  $x_m(t), m = 1, \dots, M$  are orthogonal, multiple orthogonal waveforms will be received at the receive array. By extracting the orthogonal components from the received signal  $\mathbf{y}(t, \varrho)$ , a larger virtual aperture can be achieved [41]. More specifically, by matched filtering the received signal  $\mathbf{y}(t, \varrho)$  to each of the orthogonal waveforms  $x_m(t), m = 1, \dots, M$ , the  $N \times 1$  virtual data vector that corresponds to  $x_m(t)$  can be obtained as

$$\begin{aligned} \mathbf{y}_m(\varrho) &\triangleq \int_{\mathbb{T}} \mathbf{y}(t, \varrho) x_m(t) dt \\ &\triangleq \sum_{l=1}^L \beta_l(\varrho) e^{-j2\pi f_0 \tau_m(\theta_l)} \mathbf{b}(\theta_l) + \mathbf{z}_m(\varrho), \quad m = 1, \dots, M \end{aligned} \quad (2.22)$$

where  $\mathbf{z}_m(\varrho) \triangleq \int_{\mathbb{T}} \mathbf{z}(t, \varrho) x_m(t) dt$  is the  $N \times 1$  noise term whose covariance is  $\sigma_z^2 \mathbf{I}_N$ . Note that  $\mathbf{z}_m(\varrho)$  and  $\mathbf{z}_n(\varrho)$  ( $m \neq n$ ) are independent due to the orthogonality

between  $x_m(t)$  and  $x_n(t)$ . By stacking the data vectors (2.22) obtained by matched filtering the received signal (2.21), the increased virtual data vector can be expressed as

$$\begin{aligned}\mathbf{y}(\varrho) &= [\mathbf{y}_1^T(\varrho), \mathbf{y}_2^T(\varrho), \dots, \mathbf{y}_M^T(\varrho)]^T \\ &= \sum_{l=1}^L \beta_l(\varrho) \mathbf{a}(\theta_l) \otimes \mathbf{b}(\theta_l) + \mathbf{z}(\varrho)\end{aligned}\quad (2.23)$$

where  $\otimes$  denotes the Kronecker product,  $\mathbf{z}(\varrho) = [\mathbf{z}_1^T(\varrho), \mathbf{z}_2^T(\varrho), \dots, \mathbf{z}_M^T(\varrho)]^T$  is the stacked noise vector and

$$\mathbf{c}(\theta) = \mathbf{a}(\theta) \otimes \mathbf{b}(\theta) \quad (2.24)$$

is the steering vector of the virtual antenna array of size of  $M \times N$ .

By defining  $\mathbf{A} = [\mathbf{c}(\theta_1), \mathbf{c}(\theta_2), \dots, \mathbf{c}(\theta_L)]$  as the matrix of the steering vectors and  $\boldsymbol{\beta} = [\beta_1(\varrho) \cdots \beta_L(\varrho)]$ , the received signal of the virtual array (2.23) can be rewritten as

$$\mathbf{y}(\varrho) = \mathbf{A}\boldsymbol{\beta}(\varrho) + \mathbf{z}(\varrho). \quad (2.25)$$

The sample matrix estimate of the received signal correlation matrix which is of dimension  $MN \times MN$  is defined as

$$\tilde{\mathbf{R}} \triangleq \frac{1}{K} \sum_{\varrho=1}^K \mathbf{y}(\varrho) \mathbf{y}^H(\varrho). \quad (2.26)$$

Under the condition that the additive noise is white, the signal subspace matrix denoted as  $\mathbf{E}$  is constructed by considering the first  $L$  dominant eigenvectors of the received signal correlation matrix. For the case that the number of pulses  $K$  approaches infinity, the signal subspace matrix is related to the matrix of steering vectors  $\mathbf{A}$  in (2.25) through the following equation

$$\mathbf{A} = \mathbf{E}\mathbf{T} \quad (2.27)$$

where  $\mathbf{T}$  is an  $L \times L$  non-singular matrix.

### 2.2.1 Direction of arrival estimation

Estimating the DOA of the targets based on the received reflected signal from the targets, for example, (2.21) in the case of MIMO radar is one of the most important radar problems. DOA estimation is a classical parameter estimation problem and

many different DOA estimation methods, ranging from non-parametric to parametric high-resolution methods, have been developed in the literature [82]–[90]. Among the high resolution DOA estimation methods, the most popular ones are multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance (ESPRIT) [86]–[88]. These methods are the high resolution algorithms that utilize the underlying model of the received signal and as a result enjoy substantial performance improvements compared to the non-parametric methods [87]. Spectral MUSIC is based on the exhaustive search over the parameter space which makes it computationally demanding while ESPRIT exploits the RIP and therefore, it is very computationally efficient [87]. Both of the aforementioned methods require the exact knowledge of the antenna array geometry, however, in most practical applications only approximate knowledge of array geometry is available. The ESPRIT algorithm is more robust to such array geometry knowledge imperfections and as a result it is often more interesting than the MUSIC DOA estimation method in practice [87].

ESPRIT was initially developed for the phased-array radar systems, however, it has been recently extended to MIMO radar systems as well [39]. In order to understand how the rotational invariance property is utilized in a MIMO radar system, let us consider the mono-static radar system of Fig. 2.4 with uniform linear transmit and receive arrays. In this case, the steering vector of the transmit array can be expressed as

$$\mathbf{a}(\theta) = [1, e^{-j\frac{2\pi d}{\lambda} \sin(\theta)}, \dots, e^{-j(M-1)\frac{2\pi d}{\lambda} \sin(\theta)}]^T \quad (2.28)$$

where  $d$  denotes the inter-element spacing between adjacent antenna elements and  $\lambda$  is the wavelength. Let us define  $\mathbf{a}_0(\theta)$  and  $\mathbf{a}_1(\theta)$  as the first and last  $M-1$  elements of the transmit steering vector  $\mathbf{a}(\theta)$ . Since the adjacent antennas are spaced uniformly, it is obvious that  $\mathbf{a}_1(\theta) = e^{-j\frac{2\pi d}{\lambda} \sin(\theta)} \mathbf{a}_0(\theta)$ . Let us define the following submatrices of  $\mathbf{A}$  in (2.25) as

$$\mathbf{A}_0 = [\mathbf{a}_0(\theta_1) \otimes \mathbf{b}(\theta_1) \quad \mathbf{a}_0(\theta_2) \otimes \mathbf{b}(\theta_2) \cdots \mathbf{a}_0(\theta_L) \otimes \mathbf{b}(\theta_L)] \quad (2.29)$$

$$\mathbf{A}_1 = [\mathbf{a}_1(\theta_1) \otimes \mathbf{b}(\theta_1) \quad \mathbf{a}_1(\theta_2) \otimes \mathbf{b}(\theta_2) \cdots \mathbf{a}_1(\theta_L) \otimes \mathbf{b}(\theta_L)]. \quad (2.30)$$

Based on the definitions of  $\mathbf{a}_0(\theta)$  and  $\mathbf{a}_1(\theta)$ , it can be concluded that the submatrix  $\mathbf{A}_0$  is related to the submatrix  $\mathbf{A}_1$  through the following equation

$$\mathbf{A}_1 = \mathbf{A}_0 \mathbf{D} \quad (2.31)$$

where  $\mathbf{D}$  is an  $L \times L$  diagonal matrix whose  $i$ th diagonal element is equal to  $e^{-j\frac{2\pi d}{\lambda}\sin(\theta_i)}$ . By forming the submatrices  $\mathbf{E}_0$  and  $\mathbf{E}_1$  from the signal subspace matrix  $\mathbf{E}$  in the same way that the submatrices  $\mathbf{A}_0$  and  $\mathbf{A}_1$  are formed and by considering (2.27), it can be concluded that

$$\mathbf{A}_0 = \mathbf{E}_0\mathbf{T} \quad (2.32)$$

$$\mathbf{A}_1 = \mathbf{E}_1\mathbf{T}. \quad (2.33)$$

Based on (2.31), (2.32), and (2.33) and using the fact that the matrix  $\mathbf{T}$  is invertible, it can be concluded that

$$\mathbf{E}_1 = \mathbf{E}_0\mathbf{T}\mathbf{D}\mathbf{T}^{-1}. \quad (2.34)$$

The matrices  $\mathbf{E}_0$  and  $\mathbf{E}_1$  are known and therefore, the matrix product  $\mathbf{T}\mathbf{D}\mathbf{T}^{-1}$  can be estimated using, for example, the least squares (LS) estimation. The diagonal elements of  $\mathbf{D}$  are the eigenvalues of the matrix product  $\mathbf{T}\mathbf{D}\mathbf{T}^{-1}$  and therefore the location of the targets can be estimated [39].

## 2.3 Array processing and beamforming

In many practical applications such as wireless communications, radar and sonar, the desired signal is usually received through an array of antenna elements. The main difficulty of extracting the desired signal from the received signal is due to the presence of other interfering sources and the noise. If the desired signal and the interfering sources occupy different frequency bands, then the desired signal can be easily separated from the interferences by using the temporal filters. However, it is not usually the case and some of the interference sources are in the same frequency band as the desired signal [89]. In this particular case, the fact that the desired signal and the interference sources impinge to the antenna array from different directions can be utilized to spatially filter the received signal and separate the desired signal. The process of spatially filtering the received signal is generally referred to as beamforming [89]. In order to separate the desired signal from the interference sources, the beamformer linearly combines the received signals from different antenna elements. The weights of the linear combination are designed in such a way that the interference sources are suppressed [89]. Depending on whether the knowledge of the received data is utilized or not, the beamformer may be data-independent or adaptive. In the case of data-independent beamformer, the

weights are designed independent of the received data so that the beampattern has a particular shape and specific directions in the space are suppressed, while, in the case of adaptive beamformer, the statistics of the received data are used to optimally adopt the beamforming vector to the data.

Despite the fact that the data-independent beamformers suppress the interference sources to some extent, in most practical applications, the data-independent beamformers do not provide satisfactory performance and adaptive beamforming methods are required. The main idea behind adaptive beamforming algorithms is to adjust the weight vectors according to the statistics of the received data.

### 2.3.1 Adaptive beamforming

In this subsection, we briefly introduce the main concepts behind adaptive beamforming techniques. For this goal, let us consider a linear antenna array with  $M$  omnidirectional antenna elements. The narrowband signal received by this array at the time instant  $k$  can be written as

$$\mathbf{x}(k) = \mathbf{s}(k) + \mathbf{i}(k) + \mathbf{n}(k) \quad (2.35)$$

where  $\mathbf{s}(k)$ ,  $\mathbf{i}(k)$ , and  $\mathbf{n}(k)$  are the  $M \times 1$  vectors of the desired signal, interference, and noise, respectively. The desired signal, interference, and noise components of the received signal (2.35) are statistically independent to each other. The desired signal can be written as  $\mathbf{s}(k) = s(k)\mathbf{a}$  where  $s(k)$  is the signal waveform and  $\mathbf{a}$  is the associated steering vector. As it was mentioned earlier, the output of a beamformer is a linear combination of the received signal from different antenna elements and at the time instant  $k$ , it can be written as

$$y(k) = \mathbf{w}^H \mathbf{x}(k) \quad (2.36)$$

where  $\mathbf{w}$  is the  $M \times 1$  complex weight (beamforming) vector of the antenna array. Assuming that the steering vector  $\mathbf{a}$  is known, the adaptive beamformer aims at maximizing the beamformer output SINR [90]

$$\text{SINR} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{a}|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (2.37)$$

where  $\sigma_s^2$  is the desired signal power,  $\mathbf{R}_{i+n} \triangleq \text{E}\{(\mathbf{i}(k) + \mathbf{n}(k))(\mathbf{i}(k) + \mathbf{n}(k))^H\}$  is the  $M \times M$  interference-plus-noise covariance matrix, and  $\text{E}\{\cdot\}$  stands for the statistical

expectation. Since  $\mathbf{R}_{i+n}$  is unknown in practice, it is substituted in (2.37) by the data sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{i=1}^K \mathbf{x}(i)\mathbf{x}^H(i) \quad (2.38)$$

where  $K$  is the number of training data samples which also include the desired signal component.

The problem of maximizing (2.37), where the sample estimate (2.38) is used instead of  $\mathbf{R}_{i+n}$ , is known as the minimum variance (MV) sample matrix inversion (SMI) beamforming and it is mathematically equivalent to the following convex optimization problem

$$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{a} = 1. \quad (2.39)$$

The adaptive beamformer obtained by solving the problem (2.39) preserves the desired signal at the beamformer output and meanwhile minimizes the corresponding power due to the interference sources and the noise. As a result, such adaptive beamformer has usually null points at the location of the interference sources in its corresponding beampattern.

The solution of (2.39) can be easily found as  $\mathbf{w}_{\text{MVDR-SMI}} = \alpha \hat{\mathbf{R}}^{-1} \mathbf{a}$  where  $\alpha = 1/\mathbf{a}^H \hat{\mathbf{R}}^{-1} \mathbf{a}$  [90].

### 2.3.2 Robust adaptive beamforming

When the desired signal is present in the training data, the performance of adaptive beamforming methods degrades dramatically in the presence of even a very slight mismatch in the knowledge of the desired signal steering vector  $\mathbf{a}$ . The mismatch between the presumed and actual steering vectors of the desired signal occurs because of, for example, the displacement of antenna elements, time varying environment, imperfections of propagation medium, etc. Under the presence of such mismatches between the actual and presumed steering vectors, the robust adaptive beamformer preserves the received signal from a direction other than the actual desired one and can even put a null at the corresponding direction of the desired source. As a result, the desired signal can be suppressed from the beamformer output and it leads to severe performance degradation. The main goal of any RAB technique is to provide robustness against any such mismatches.

The traditional design approaches to adaptive beamforming [90]–[93] do not provide sufficient robustness and are not applicable in such situations. Thus, various RAB techniques have been developed [44]. Some examples of popular conventional RAB approaches are the diagonal loading technique [46], [47], the projection beamforming techniques [48], [93], and the eigenspace-based beamforming technique [94]. The disadvantages of these approaches such as, for example, the ad hoc nature of the former one and high probability of subspace swap at low SNRs for the latter one [95] are well known.

Among more recent RAB techniques based on MVDR principle are (i) the worst-case-based adaptive beamforming technique proposed in [7], [49] and further developed in [50]–[51]; (ii) the doubly constrained robust Capon beamforming method [53], [54] (it is based on the same idea of the worst-case performance optimization as (i)); (iii) the probabilistically constrained RAB technique [52]; (iv) the RAB technique based on steering vector estimation [55]; and others.

All the MVDR RAB methods need some sort of prior information about the desired signal. In what follows, we briefly review the recent RAB techniques and explain the type of prior information that they use.

*Eigenspace-based beamformer [48], [94]:* Taking an inaccurate knowledge of the actual steering vector  $\mathbf{a}$ , i.e., the presumed steering vector  $\mathbf{p}$ , as a prior, the eigenspace-based beamformer finds and uses the projection of  $\mathbf{p}$  onto the sample signal-plus-interference subspace as a corrected estimate of the steering vector. The eigendecomposition of (2.38) yields

$$\hat{\mathbf{R}} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H + \mathbf{G}\mathbf{\Gamma}\mathbf{G}^H \quad (2.40)$$

where the  $M \times (J+1)$  matrix  $\mathbf{E}$  and  $M \times (M-J-1)$  matrix  $\mathbf{G}$  contain the signal-plus-interference subspace eigenvectors of  $\hat{\mathbf{R}}$  and the noise subspace eigenvectors, respectively, while the  $(J+1) \times (J+1)$  matrix  $\mathbf{\Lambda}$  and  $(M-J-1) \times (M-J-1)$  matrix  $\mathbf{\Gamma}$  contain the eigenvalues corresponding to  $\mathbf{E}$  and  $\mathbf{G}$ , respectively. Here,  $J$  is the number of interfering signals. The eigenspace-based beamformer is obtained by substituting the so-obtained estimate of the steering vector to the MVDR-SMI beamformer, and is expressed as

$$\mathbf{w}_{\text{eig}} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{a}} = \hat{\mathbf{R}}^{-1}\mathbf{E}\mathbf{E}^H\mathbf{p} = \mathbf{E}\mathbf{\Lambda}^{-1}\mathbf{E}^H\mathbf{p} \quad (2.41)$$

where  $\hat{\mathbf{a}} = \mathbf{E}\mathbf{E}^H\mathbf{p}$  and  $\mathbf{E}\mathbf{E}^H$  is the projection matrix.

The prior information used in this method is the presumed steering vector  $\mathbf{p}$  and the number of interfering signals. It is known that the eigenspace-based beamformer may suffer from a high probability of subspace swap as well as incorrect estimation of the signal-plus-interference subspace dimension [95].

*Worst-case-based and doubly constrained RAB techniques [7], [50], [53]:* These techniques assume the knowledge of the presumed steering vector and model the actual steering vector  $\mathbf{a}$  as  $\mathbf{a} = \mathbf{p} + \boldsymbol{\delta}$ , where  $\boldsymbol{\delta}$  is a deterministic unknown mismatch vector with bounded norm, i.e.,  $\|\boldsymbol{\delta}\| \leq \varepsilon$  for some value  $\varepsilon$ . Assuming, for example, spherical uncertainty for  $\boldsymbol{\delta}$ , i.e.,  $\mathcal{A}(\varepsilon) \triangleq \{\mathbf{a} = \mathbf{p} + \boldsymbol{\delta} \mid \|\boldsymbol{\delta}\| \leq \varepsilon\}$ , the worst-case-based MVDR RAB can be interpreted as the standard MVDR-SMI used in tandem with a steering vector estimation obtained by solving the following covariance fitting problem [50]

$$\begin{aligned} \max_{\sigma^2, \hat{\mathbf{a}}} \sigma^2 \quad \text{subject to} \quad & \hat{\mathbf{R}} - \sigma^2 \hat{\mathbf{a}} \hat{\mathbf{a}}^H \geq 0 \\ & \text{for any } \hat{\mathbf{a}} \text{ satisfying } \|\boldsymbol{\delta}\| \leq \varepsilon. \end{aligned} \quad (2.42)$$

The doubly constrained MVDR RAB differs from the aforementioned one only by adding an additional natural constraint to the norm of the estimate of steering vector, i.e.,  $\|\hat{\mathbf{a}}\|^2 = M$ . The latter optimization problem can be, however, significantly more complex since this additional constraint is non-convex.

The prior information used in these MVDR RAB techniques is the presumed steering vector and the bound to the norm of the steering vector mismatch  $\varepsilon$ , which is difficult to obtain in practice. The uncertainty region for these methods can be generalized by considering an ellipsoidal region [51], [54]. However, in this case, a more sophisticated prior information that is needed to define the shape of the ellipsoid has to be available, which is even more difficult to reliably obtain in practice.

*Probabilistically constrained robust adaptive beamforming [52]:* This MVDR RAB technique assumes that the mismatch vector  $\boldsymbol{\delta}$  is random and is formulated as

$$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \Pr\{|\mathbf{w}^H \mathbf{a}| \geq 1\} \geq p_0 \quad (2.43)$$

where  $\Pr\{\cdot\}$  denotes probability and  $p_0$  is preselected probability value.

The prior information for this technique is again the presumed steering vector and a distribution of  $\boldsymbol{\delta}$  together with  $p_0$ . This prior information knowledge is more relaxed as compared to the worst-case-based approach since it is typically easier to estimate statistics of mismatch distribution reliably and  $p_0$  has a physical meaning

of a non-outage probability for the distortionless response constraint. Moreover, for the cases of Gaussian and the worst-case distributions of  $\boldsymbol{\delta}$ , the probabilistically constrained RAB can be approximated by the same optimization problem as for the worst-case-based MVDR RAB technique with  $\varepsilon$  being a function of covariance matrix of  $\boldsymbol{\delta}$  and  $p_0$ .

*The MVDR RAB of [55]:* In the mismatched case, the solution of (2.39) can be written as a function of unknown mismatch vector  $\boldsymbol{\delta}$ , that is,  $\mathbf{w}(\boldsymbol{\delta}) = \alpha \hat{\mathbf{R}}^{-1}(\mathbf{p} + \boldsymbol{\delta})$ . Thus, the beamformer output power can be also written as a function of  $\boldsymbol{\delta}$  as

$$P(\boldsymbol{\delta}) = \frac{1}{(\mathbf{p} + \boldsymbol{\delta})^H \hat{\mathbf{R}}^{-1}(\mathbf{p} + \boldsymbol{\delta})}. \quad (2.44)$$

Then the best estimate of  $\boldsymbol{\delta}$  or, equivalently,  $\mathbf{a}$  is the one which maximizes (2.44) under the requirement that this estimate does not converge to any of the interferences or their linear combinations. The aforementioned convergence is avoided in [55] by requiring that

$$\mathbf{P}^\perp(\mathbf{p} + \hat{\boldsymbol{\delta}}) = \mathbf{P}^\perp \hat{\mathbf{a}} = 0 \quad (2.45)$$

where  $\mathbf{P}^\perp \triangleq \mathbf{I} - \mathbf{L}\mathbf{L}^H$ ,  $\mathbf{L} \triangleq [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_L]$ ,  $\mathbf{l}_l$ ,  $l = 1, \dots, L$  are the  $L$  dominant eigenvectors of the matrix  $\mathbf{C} \triangleq \int_{\Theta} \mathbf{d}(\theta) \mathbf{d}^H(\theta) d\theta$ ,  $\mathbf{d}(\theta)$  is the steering vector associated with direction  $\theta$  and having the structure defined by the antenna geometry,  $\Theta$  is the angular sector in which the desired signal is assumed to be located,  $\hat{\boldsymbol{\delta}}$  and  $\hat{\mathbf{a}}$  stand for the estimate of the steering vector mismatch and the actual steering vector, respectively, and  $\mathbf{I}$  is the identity matrix. The steering vector estimation in [55] is based on splitting the mismatch vector  $\hat{\boldsymbol{\delta}}$  into the orthogonal component to the presumed steering vector  $\mathbf{p}$  and the parallel one, i.e.,  $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_\perp + \hat{\boldsymbol{\delta}}_\parallel$ , and then the estimation of  $\boldsymbol{\delta}$  and  $\mathbf{a}$  can be found iteratively by finding  $\hat{\boldsymbol{\delta}}_\perp$  as a solution of the following convex optimization problem

$$\begin{aligned} \min_{\hat{\boldsymbol{\delta}}_\perp} \quad & (\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp)^H \hat{\mathbf{R}}^{-1}(\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp) \\ \text{subject to} \quad & \mathbf{P}^\perp(\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp) = \mathbf{0}, \quad \mathbf{p}^H \hat{\boldsymbol{\delta}}_\perp = 0 \\ & \|\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp\|^2 \leq M \\ & (\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp)^H \tilde{\mathbf{C}}(\mathbf{p} + \hat{\boldsymbol{\delta}}_\perp) \leq \mathbf{p}^H \tilde{\mathbf{C}}\mathbf{p} \end{aligned} \quad (2.46)$$

where  $\tilde{\mathbf{C}} \triangleq \int_{\tilde{\Theta}} \mathbf{d}(\theta) \mathbf{d}^H(\theta) d\theta$ , the sector  $\tilde{\Theta}$  is the complement of the sector  $\Theta$ . Here, the last constraint limits the noise power collected outside  $\Theta$ , while the orthogonality between  $\hat{\boldsymbol{\delta}}_\perp$  and  $\mathbf{p}$  is imposed by adding the constraint  $\mathbf{p}^H \hat{\boldsymbol{\delta}}_\perp = 0$ .

It can be seen that the prior information used in this approach is the presumed steering vector, approximate knowledge of antenna array geometry and the angular sector  $\Theta$  in which the desired signal is located. The technique of [55] can be significantly simplified if it is known that the array is partly calibrated [96], but the amount of prior information about the type of uncertainty, i.e., the knowledge that the array is partially calibrated, then increases.

### 2.3.3 General-rank signal model and beamforming

Most of the beamforming methods have been developed for the case of point source signals when the rank of the desired signal covariance matrix is equal to one [44], [49]–[52] (see Subsection 2.3.2). However, in many practical applications such as, for example, the incoherently scattered signal source or source with fluctuating (randomly distorted) wavefronts, the rank of the source covariance matrix is higher than one. Although the RAB methods of [49]–[52] provide excellent robustness against any mismatch of the underlying point source assumption, they are not perfectly suited to the case when the rank of the desired signal covariance matrix is higher than one. In what follows, a brief description of the general-rank adaptive beamforming is given and then the robust adaptive beamforming for general-rank signal models is discussed.

Similar to the rank-one signal model, the beamforming problem in the case of a general-rank signal is formulated as finding the beamforming vector  $\mathbf{w}$  which maximizes the beamformer output SINR given as

$$\text{SINR} = \frac{\mathbf{w}^H \mathbf{R}_s \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (2.47)$$

where  $\mathbf{R}_s \triangleq \text{E}\{\mathbf{s}(k)\mathbf{s}^H(k)\}$  is the desired signal covariance matrix. Depending on the nature of the desired signal source, its corresponding covariance matrix can be of an arbitrary rank, i.e.,  $1 \leq \text{rank}\{\mathbf{R}_s\} \leq M$ . Indeed, in many practical applications, for example, in the scenarios with incoherently scattered signal sources or signals with randomly fluctuating wavefronts, the rank of the desired signal covariance matrix  $\mathbf{R}_s$  is greater than one [8]. The only particular case in which, the rank of  $\mathbf{R}_s$  is equal to one is the case of a point source.

As it was mentioned earlier, the interference-plus-noise covariance matrix  $\mathbf{R}_{i+n}$  is typically unavailable in practice and it is substituted by the data sample covariance matrix. By such a substitution, the corresponding MVDR beamforming problem

for general-rank source can be mathematically formulated as

$$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{R}_s \mathbf{w} = 1. \quad (2.48)$$

The solution to the MVDR beamforming problem (2.48) can be found as [44]

$$\mathbf{w}_{\text{MVDR-SMI}} = \mathcal{P}\{\hat{\mathbf{R}}^{-1} \mathbf{R}_s\} \quad (2.49)$$

which is known as the MVDR-SMI beamformer for general-rank signal model. Here  $\mathcal{P}\{\cdot\}$  stands for the principal eigenvector operator.

In practice, the actual desired signal covariance matrix  $\mathbf{R}_s$  is usually unknown and only its presumed value is available. The actual source correlation matrix can be modeled as  $\mathbf{R}_s = \tilde{\mathbf{R}}_s + \mathbf{\Delta}_1$ , where  $\mathbf{\Delta}_1$  and  $\tilde{\mathbf{R}}_s$  denote an unknown mismatch and the presumed correlation matrices, respectively. Similar to the point source signal model, the general-rank MVDR beamformer is very sensitive to such mismatches [8] and the ultimate goal of any general-rank RAB method is to provide robustness against such mismatches. Moreover, RABs address the situation when the sample estimate of the data covariance matrix (2.38) is inaccurate (for example, because of small sample size) and  $\mathbf{R} = \hat{\mathbf{R}} + \mathbf{\Delta}_2$ , where  $\mathbf{\Delta}_2$  is an unknown mismatch matrix to the data sample covariance matrix. The RAB for the general-rank signal model based on the explicit modeling of the error mismatches has been developed in [8] based on the worst-case performance optimization principle. Specifically, for providing robustness against the norm-bounded mismatches  $\|\mathbf{\Delta}_1\| \leq \epsilon$  and  $\|\mathbf{\Delta}_2\| \leq \gamma$ , the RAB of [8] uses the worst-case performance optimization principle of [7] and finds the solution as

$$\mathbf{w} = \mathcal{P}\{(\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1}(\tilde{\mathbf{R}}_s - \epsilon \mathbf{I})\}. \quad (2.50)$$

Although the RAB of [8] has a simple closed-form solution (2.50), it is overly conservative because the constraint that the matrix  $\tilde{\mathbf{R}}_s + \mathbf{\Delta}_1$  has to be PSD is not considered [10]. For example, the worst-case desired signal covariance matrix  $\tilde{\mathbf{R}}_s - \epsilon \mathbf{I}$  in (2.50) can be indefinite or even negative definite if  $\tilde{\mathbf{R}}_s$  is rank deficient. Indeed, in the case of incoherently scattered source,  $\tilde{\mathbf{R}}_s$  has the following form  $\tilde{\mathbf{R}}_s = \sigma_s^2 \int_{-\pi/2}^{\pi/2} \zeta(\theta) \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta$ , where  $\zeta(\theta)$  denotes the normalized angular power density,  $\sigma_s^2$  is the desired signal power, and  $\mathbf{a}(\theta)$  is the steering vector towards direction  $\theta$ . For a uniform angular power density on the angular bandwidth  $\Phi$ , the approximate numerical rank of  $\tilde{\mathbf{R}}_s$  is equal to  $(\Phi/\pi) \cdot M$  [97]. This leads to a rank deficient matrix  $\tilde{\mathbf{R}}_s$  if the angular power density does not cover all the directions.

Therefore, the worst-case covariance matrix  $\tilde{\mathbf{R}}_s - \epsilon \mathbf{I}$  is indefinite or negative definite. Note that the worst-case data sample covariance matrix  $\hat{\mathbf{R}} + \gamma \mathbf{I}$  is always positive definite. A less conservative robust adaptive beamforming problem formulation which enforces the matrix  $\tilde{\mathbf{R}}_s + \mathbf{\Delta}_1$  to be PSD has been considered in [10] and [56] will be explained in details in Chapter 6.

## 2.4 Two-way cooperative communications

Wireless channels have the serious challenge of the signal fading due to the multipath propagation and shadowing [81]. As a result of signal fading, the received signal has a random time-varying power that may give rise to severe loss in terms of the power usage. Diversity is one of the most efficient ways to alleviate such adverse effects of signal fading. Diversity means to transmit the same information over independently faded paths. The basic idea behind the diversity is the fact that the probability of having deep fading over several independently faded signals is vanishingly small [81]. Among the different forms of diversity, the spatial diversity is the most favorable and common one which can be easily combined with other forms of diversity such as, for example, time and frequency and can result in significant performance improvements even when other forms of diversity are not present in a system. The spatial diversity is usually established by having multiple antennas at transmitter or receiver. More specifically, if the antennas at the transmitter or receiver are far enough from each other, the received signals will experience independent fading. However, due to the practical restrictions, it is not always possible to locate multiple antennas in a small mobile terminal and ensure that the corresponding paths are independent. To overcome this drawback of the spatial diversity offered by means of applying multiple antennas at the transmitter or receiver, the idea of cooperative diversity has been introduced which offers most of the advantages of MIMO systems such as high data rate and low probability of outage [57]–[59], [98]. Cooperative diversity utilizes the inherent spatial diversity of wireless networks and, as a result, does not require to apply multiple antennas.

In order to explain how the inherent spatial diversity can be exploited, let us consider a simple wireless network in which the source nodes  $S_1$  and  $S_2$  communicate with their corresponding destination nodes  $D_1$  and  $D_2$ , respectively, through the orthogonal channels by means of time division multiplexing. The channel between the nodes  $\alpha$  and  $\beta$  is denoted as  $h_{\alpha,\beta}$  and is assumed to have Rayleigh flat fading

model [81]. At the first time slot,  $S_1$  transmits its information to  $D_1$  through the channel  $h_{S_1,D_1}$  and the received signal can be written as

$$y_{S_1,D_1} = \sqrt{p_1} h_{S_1,D_1} x_1 + n_{S_1,D_1} \quad (2.51)$$

where  $x_1$  is the transmit message of source node  $S_1$  of unit power,  $p_1$  is the transmit power of the source node  $S_1$ , and  $n_{S_1,D_1}$  is the additive white Gaussian noise in  $D_1$ . Due to the broadcast nature of the wireless networks, besides  $D_1$ , the source node  $S_2$  also receives the noisy and attenuated version of  $x_1$  as follows

$$y_{S_1,S_2} = \sqrt{p_1} h_{S_1,S_2} x_1 + n_{S_1,S_2}. \quad (2.52)$$

If the source node  $S_2$  processes the received signal  $y_{S_1,S_2}$  and then retransmits it to  $D_1$  in the second time slot, then  $D_1$  will receive two independently faded noisy versions of  $x_1$  [98]. The received signal in  $D_1$  from  $S_2$  in the second time slot can be written as

$$y_{S_2,D_1} = \sqrt{p_2} h_{S_2,D_1} f(y_{S_1,S_2}) + n_{S_2,D_1} \quad (2.53)$$

where  $p_2$  is the transmit power of the source node  $S_2$  and  $f(y_{S_1,S_2})$  denotes the received signal  $y_{S_1,S_2}$  which has been processed by the source node  $S_2$  to remove the channel imperfections and the noise. By collaboration of  $S_2$  and acting as a relay, a spatial diversity with two different paths can be achieved. The data rate between  $S_1$  and  $D_1$  is a function of the received SNR and can be expressed as

$$r_{S_1,D_1} = \frac{1}{2} \log(1 + SNR(S_1 \rightarrow D_1)) \quad (2.54)$$

where  $SNR(S_1 \rightarrow D_1)$  is the received SNR at  $D_1$ , which depends on type of the diversity combining at the receiver, and the factor of 1/2 appears in this scheme because two time slots are needed for every data transmission from  $S_1$  to  $D_1$ . This type of relaying is referred to as the conventional one-way relaying and it is not optimal in terms of the rate. In particular, by using the one-way relaying scheme, a bidirectional data exchange between a source and destination would require four time slots [60].

The idea of the TWR has been proposed to address this spectral inefficiency of the conventional one-way relaying systems [99], [100]. TWR can be viewed as a certain form of network coding [101] which allows to reduce the number of time slots used for transmission in one-way relaying by relaxing the requirement of ‘orthogonal/non-interfering’ transmissions between the terminals and the relay [62].

Specifically, simultaneous transmissions by the terminals to the relay on the same frequencies are allowed in the first time slot, while a combined signal is broadcast by the relay in the second time slot.

The rate-optimal strategy for two-way relaying is in general unknown [61]. However, some efficient strategies have been developed. Depending on the ability of the relay to regenerate/decode the signals from the terminals, several two-way transmission protocols have been introduced and studied. The regenerative relay adopts the decode-and-forward (DF) protocol and performs the decoding process at the relay [102], while the non-regenerative relay typically adopts a form of AF protocol and does not perform decoding at the relay, but amplifies and possibly beamforms or precodes the signals to retransmit them to the terminals [61], [103], [104]. The advantages of the latter are a smaller delay in the transmission and lower hardware complexity of the relay. Most of the research on TWR systems concentrates on studying the corresponding sum-rate, the achievable rate region, and also the bit error probability of different schemes [105]. The trade-off between the error probability and the achievable rate has been recently studied in [105] using Gallager's random coding error exponent.

## Chapter 3

# Generalized Quadratically Constrained Quadratic Programming

In this thesis, we are mostly interested in the following optimization problem

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}} \quad & f_0(\mathbf{x}^H \mathbf{A}_0 \mathbf{x}) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \alpha_{2i-1} f_{2i-1}(\mathbf{x}^H \mathbf{A}_{2i-1} \mathbf{x}) - \alpha_{2i} f_{2i}(\mathbf{x}^H \mathbf{A}_{2i} \mathbf{x}) + h_i(\mathbf{y}) \leq 0, \\ & i = 1, \dots, M \end{aligned} \quad (3.1)$$

where  $\mathbf{x} \in \mathbb{C}^m$ ,  $\mathbf{y} \in \mathbb{R}^n$ , and  $f_i : \mathcal{D}_{f_i} \subset \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 0, \dots, 2M$  are one-dimensional convex differentiable functions. Moreover, the function  $f_0$  is assumed to be a monotonic function which is bounded from the below over the feasible set of the problem. The matrices  $\mathbf{A}_i \in \mathcal{H}^m$ ,  $i = 0, \dots, 2M$  are Hermitian matrices that are not necessarily definite,  $\mathbf{h}_i : \mathcal{D}_{h_i} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 0, \dots, M$  are convex differentiable functions, and  $\alpha_i \in \{0, 1\}$ ,  $i = 1, \dots, 2M$  takes value 1 or 0, respectively, depending whether the function  $f_i(\mathbf{x}^H \mathbf{A}_i \mathbf{x})$  is present or not. Similar to the function  $f_0$ , the convex function  $h_0$  is also assumed to be lower-bounded.

As it was mentioned earlier in Chapter 1, this type of optimization programming is frequently encountered in the fields of communications and signal processing due to the fact that such important properties as the rate, SNR, SINR are generally a composition of one-dimensional and quadratic functions.

By defining the additional variables  $\delta_i$ ,  $i = 1, 2, \dots, 2M$  and the set  $C \triangleq \{k \mid \alpha_k =$

$1, 1 \leq k \leq 2M\}$ , the problem above can be equivalently expressed as

$$\begin{aligned} & \min_{\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}} \quad f_0(\mathbf{x}^H \mathbf{A}_0 \mathbf{x}) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M \\ & \mathbf{x}^H \mathbf{A}_i \mathbf{x} = \delta_i, \quad i \in C \end{aligned} \quad (3.2)$$

where  $\boldsymbol{\delta}$  is the set of the variables  $\delta_i$ . Let us rewrite problem (3.2) as

$$\begin{aligned} & \min_{\mathbf{y}, \boldsymbol{\delta}} \quad \min_{\mathbf{x}} \quad f_0(\mathbf{x}^H \mathbf{A}_0 \mathbf{x}) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \mathbf{x}^H \mathbf{A}_i \mathbf{x} = \delta_i, \quad i \in C \\ & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M. \end{aligned} \quad (3.3)$$

Since the function  $h_0(\mathbf{y})$  as well as the constraint functions  $\alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, i = 1, \dots, M$  do not depend on  $\mathbf{x}$ , the problem (3.3) can be further recast as

$$\begin{aligned} & \min_{\mathbf{y}, \boldsymbol{\delta}} \quad h_0(\mathbf{y}) + \left[ \overbrace{\min_{\mathbf{x} \mid \mathbf{x}^H \mathbf{A}_i \mathbf{x} = \delta_i, i \in C} f_0(\mathbf{x}^H \mathbf{A}_0 \mathbf{x})}^{\text{Inner Optimization Problem}} \right] \\ \text{subject to} \quad & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M. \end{aligned} \quad (3.4)$$

Then the problem (3.4) can be further expressed as

$$\begin{aligned} & \min_{\mathbf{y}, \boldsymbol{\delta}} \quad f_0(k(\boldsymbol{\delta})) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M \end{aligned} \quad (3.5)$$

where  $k(\boldsymbol{\delta})$  is an optimal value function which is defined based on the inner optimization problem in (3.4) for a fixed value of  $\boldsymbol{\delta} \in \mathcal{D}$ . Here,  $\mathcal{D}$  is defined as the set of all  $\boldsymbol{\delta}$  such that the corresponding optimization problem obtained from  $k(\boldsymbol{\delta})$  for fixed  $\boldsymbol{\delta}$  is feasible. If the function  $f_0$  is increasing, i.e.,  $f_0(x_1) \leq f_0(x_2), x_1 \leq x_2$ , the optimal value function  $k(\boldsymbol{\delta})$  is defined as

$$k(\boldsymbol{\delta}) \triangleq \left\{ \min_{\mathbf{x}} \mathbf{x}^H \mathbf{A}_0 \mathbf{x} \mid \mathbf{x}^H \mathbf{A}_i \mathbf{x} = \delta_i, i \in C \right\} \quad (3.6)$$

while for a decreasing  $f_0$ , i.e.,  $f_0(x_1) \geq f_0(x_2), x_1 \leq x_2$ , the definition of the optimal value function is as follows

$$k(\boldsymbol{\delta}) \triangleq \left\{ \max_{\mathbf{x}} \mathbf{x}^H \mathbf{A}_0 \mathbf{x} \mid \mathbf{x}^H \mathbf{A}_i \mathbf{x} = \delta_i, i \in C \right\}. \quad (3.7)$$

In the rest of this chapter, we assume that the function  $f_0$  is increasing. However, all the discussion that follows can be similarly concluded for the case where the

function  $f_0$  is a decreasing function. We first consider the case of  $\text{card}\{C\} \leq 3$  where  $\text{card}\{\cdot\}$  is the cardinality operator. In other words, we first consider the case when the total number of composite functions  $f_i(\mathbf{x}^H \mathbf{A}_i \mathbf{x})$  in the constraints of the problem (3.1) does not exceed three or, equivalently, the case when the number of quadratic constraints in the optimization problem of  $k(\boldsymbol{\delta})$  (3.6) is less than or equal to three.

### 3.1 Number of quadratic functions in the constraints less than or equal to three

In this case by introducing the matrix  $\mathbf{X} \triangleq \mathbf{x}\mathbf{x}^H$  and observing that for any arbitrary matrix  $\mathbf{Y}$ , the relationship  $\mathbf{x}^H \mathbf{Y} \mathbf{x} = \text{tr}\{\mathbf{Y}\mathbf{x}\mathbf{x}^H\}$  holds, the optimal value function  $k(\boldsymbol{\delta})$  (3.6) can be equivalently recast as [70]

$$k(\boldsymbol{\delta}) = \left\{ \min_{\mathbf{X}} \text{tr}\{\mathbf{A}_0 \mathbf{X}\} \mid \text{tr}\{\mathbf{A}_i \mathbf{X}\} = \delta_i, i \in C, \text{rank}\{\mathbf{X}\} = 1, \mathbf{X} \succeq \mathbf{0} \right\}, \quad \boldsymbol{\delta} \in \mathcal{D}. \quad (3.8)$$

In the optimization problem obtained from the optimal value function  $k(\boldsymbol{\delta})$  (3.8) by fixing  $\boldsymbol{\delta}$ , the rank-one constraint  $\text{rank}\{\mathbf{X}\} = 1$  is the only non-convex constraint with respect to the new optimization variable  $\mathbf{X}$ . Using the SDP relaxation, the corresponding optimization problem can be relaxed by dropping the rank-one constraint, and the following new optimal value function  $h(\boldsymbol{\delta})$  can be defined

$$h(\boldsymbol{\delta}) \triangleq \left\{ \min_{\mathbf{X}} \text{tr}\{\mathbf{A}_0 \mathbf{X}\} \mid \text{tr}\{\mathbf{A}_i \mathbf{X}\} = \delta_i, i \in C, \mathbf{X} \succeq \mathbf{0} \right\}, \boldsymbol{\delta} \in \mathcal{D}' \quad (3.9)$$

where  $\mathcal{D}'$  is the set of all  $\boldsymbol{\delta}$  such that the optimization problem which corresponds to  $h(\boldsymbol{\delta})$  for a fixed  $\boldsymbol{\delta}$  is feasible. For brevity, we will refer to the optimization problems corresponding to the functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  when  $\boldsymbol{\delta}$  is fixed simply as the optimization problems of  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$ , respectively. The following lemma finds the relationship between the domains of the functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$ .

**Lemma 3.1.** *The domains of the functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  are the same, i.e.,  $\mathcal{D} = \mathcal{D}'$  if  $\text{card}\{C\} \leq 3$ .*

*Proof.* Let us assume that  $C = \{c_1, c_2, c_3\}$  which implies that  $\text{card}\{C\} = 3$ . Note that the other cases of  $\text{card}\{C\} = 2$  and  $\text{card}\{C\} = 1$  can be similarly proved. First, we prove that if  $\boldsymbol{\delta} \in \mathcal{D}$  then  $\boldsymbol{\delta} \in \mathcal{D}'$ . Let  $\boldsymbol{\delta} \in \mathcal{D}$ . It implies that there exists a vector  $\mathbf{x}_0$  such that the constraints  $\mathbf{x}_0^H \mathbf{A}_{c_1} \mathbf{x}_0 = \delta_{c_1}$ ,  $\mathbf{x}_0^H \mathbf{A}_{c_2} \mathbf{x}_0 = \delta_{c_2}$ , and  $\mathbf{x}_0^H \mathbf{A}_{c_3} \mathbf{x}_0 = \delta_{c_3}$

are satisfied. Defining the new matrix  $\mathbf{X}_0 \triangleq \mathbf{x}_0 \mathbf{x}_0^H$ , it is easy to verify that  $\mathbf{X}_0$  satisfies the constraints in the optimization problem of  $h(\boldsymbol{\delta})$  and, therefore,  $\boldsymbol{\delta} \in \mathcal{D}'$ . Now, let us assume that  $\boldsymbol{\delta} \in \mathcal{D}'$ . Therefore, there exists a positive semi-definite matrix  $\mathbf{X}_0 = \mathbf{V}_{m \times r} \mathbf{V}_{m \times r}^H$  with rank equal to  $r$  and  $\mathbf{V}_{m \times r}$  being a full rank matrix such that  $\text{tr}\{\mathbf{A}_{c_1} \mathbf{X}_0\} = \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_1} \mathbf{V}\} = \delta_{c_1}$ ,  $\text{tr}\{\mathbf{A}_{c_2} \mathbf{X}_0\} = \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_2} \mathbf{V}\} = \delta_{c_2}$ , and  $\text{tr}\{\mathbf{A}_{c_3} \mathbf{X}_0\} = \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_3} \mathbf{V}\} = \delta_{c_3}$ . If the rank of  $\mathbf{X}_0$  is one, then  $\mathbf{x}_0 = \mathbf{V}_{m \times 1}$  satisfies the constraints in the optimization problem of  $k(\boldsymbol{\delta})$  and trivially  $\boldsymbol{\delta} \in \mathcal{D}$ . Thus, we assume that  $r$  is greater than one and aim to show that based on  $\mathbf{X}_0$ , another rank-one feasible point for the optimization problem of  $h(\boldsymbol{\delta})$  can be constructed by following similar lines as in [106]. To this end, let us consider the following set of equations

$$\begin{aligned} \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_1} \mathbf{V} \boldsymbol{\Gamma}\} &= 0, \\ \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_2} \mathbf{V} \boldsymbol{\Gamma}\} &= 0, \\ \text{tr}\{\mathbf{V}^H \mathbf{A}_{c_3} \mathbf{V} \boldsymbol{\Gamma}\} &= 0 \end{aligned} \tag{3.10}$$

where the  $r \times r$  Hermitian matrix  $\boldsymbol{\Gamma}$  is an unknown variable. Due to the fact that  $\text{tr}\{\mathbf{V}^H \mathbf{A}_{c_1} \mathbf{V} \boldsymbol{\Gamma}\}$ ,  $\text{tr}\{\mathbf{V}^H \mathbf{A}_{c_2} \mathbf{V} \boldsymbol{\Gamma}\}$ , and  $\text{tr}\{\mathbf{V}^H \mathbf{A}_{c_3} \mathbf{V} \boldsymbol{\Gamma}\}$  are all real valued functions of  $\boldsymbol{\Gamma}$ , the set of equations (3.10) is a linear set of 3 equations with  $r^2$  variables, that is, the total number of real and imaginary variables in the matrix  $\boldsymbol{\Gamma}$ . Since the number of variables  $r^2$ , ( $r \geq 2$ ) is greater than the number of equations, there exists a nonzero solution denoted as  $\boldsymbol{\Gamma}_0$  for the linear set of equations (3.10). Let  $\delta_0$  denote the eigenvalue of the matrix  $\boldsymbol{\Gamma}_0$  which has the largest absolute value. Without loss of generality, we can assume that  $\delta_0 > 0$ , which is due to the fact that both  $\boldsymbol{\Gamma}_0$  and  $-\boldsymbol{\Gamma}_0$  are solutions of (3.10). Using  $\mathbf{X}_0$  and  $\boldsymbol{\Gamma}_0$ , we can construct a new matrix  $\mathbf{X}_0^{new} = \mathbf{V}(\mathbf{I}_r - \boldsymbol{\Gamma}_0/\delta_0)\mathbf{V}^H$ . It is then easy to verify that the expressions  $\text{tr}\{\mathbf{A}_{c_1} \mathbf{X}_0^{new}\} = \delta_{c_1}$ ,  $\text{tr}\{\mathbf{A}_{c_2} \mathbf{X}_0^{new}\} = \delta_{c_2}$ ,  $\text{tr}\{\mathbf{A}_{c_3} \mathbf{X}_0^{new}\} = \delta_{c_3}$ , and  $\mathbf{X}_0^{new} \succeq \mathbf{0}$  are valid and the rank of  $\mathbf{X}_0^{new}$  is less than or equal to  $r - 1$ . It is because the rank of the matrix  $(\mathbf{I}_r - \boldsymbol{\Gamma}_0/\delta_0)$  is less than or equal to  $r - 1$  and the fact that rank of any matrix product is less than or equal to the rank of each of the matrices. It means that  $\mathbf{X}_0^{new}$  is another feasible point of the optimization problem of  $h(\boldsymbol{\delta})$  and its rank has reduced at least by one. This process can be repeated until  $r^2 \leq 3$  or, equivalently, a rank-one feasible point is found. After a rank-one feasible point  $\mathbf{v}_{m \times 1} \mathbf{v}_{m \times 1}^H$  is constructed,  $\mathbf{x}_0 = \mathbf{v}_{m \times 1}$  is also a feasible point for the optimization problem of  $k(\boldsymbol{\delta})$ . Thus,  $\boldsymbol{\delta} \in \mathcal{D}$  which completes the proof.  $\square$

So far, we have shown that both optimal value functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  have the same domain. Since the feasible set of the optimization problem of  $k(\boldsymbol{\delta})$  is a subset of the feasible set of the optimization problem of  $h(\boldsymbol{\delta})$ , we expect that  $k(\boldsymbol{\delta})$  is greater than or equal to  $h(\boldsymbol{\delta})$  at every feasible point. However, due to the specific structure of the optimal value function  $k(\boldsymbol{\delta})$ , these two optimal value functions are equivalent as it is shown in the following theorem.

**Theorem 3.1.** *The optimal value functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  are equivalent, i.e.,  $k(\boldsymbol{\delta}) = h(\boldsymbol{\delta})$ ,  $\boldsymbol{\delta} \in \mathcal{D}$  if  $\text{card}\{C\} \leq 3$  and some mild conditions are satisfied. Additionally, based on the optimal solution of the problem  $h(\boldsymbol{\delta})$  the optimal solution of the problem  $k(\boldsymbol{\delta})$  can be extracted.*

*Proof.* Without loss of the generality, let us assume that  $C = \{c_1, c_2, c_3\}$ . Note that the other cases of  $\text{card}\{C\} = 2$  and  $\text{card}\{C\} = 1$  can be addressed similarly. In order to show that these optimal value functions are equal, we use the dual problem of the optimization problems of  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$ . It is easy to verify that both of the optimization problems of  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  have the following same dual problem

$$\begin{aligned} & \max_{\gamma_1, \gamma_2, \gamma_3} \gamma_1 \delta_{c_1} + \gamma_2 \delta_{c_2} + \gamma_3 \delta_{c_3} \\ & \text{subject to } \mathbf{A}_0 - \mathbf{A}_{\delta_{c_1}} \gamma_1 - \mathbf{A}_{\delta_{c_2}} \gamma_2 - \mathbf{A}_{\delta_{c_3}} \gamma_3 \succeq \mathbf{0}. \end{aligned} \quad (3.11)$$

The following optimal value function can be defined based on (3.11)

$$l(\boldsymbol{\delta}) \triangleq \left\{ \max_{\gamma_1, \gamma_2, \gamma_3} \gamma_1 \delta_{c_1} + \gamma_2 \delta_{c_2} + \gamma_3 \delta_{c_3} \mid \mathbf{A}_0 - \mathbf{A}_{\delta_{c_1}} \gamma_1 - \mathbf{A}_{\delta_{c_2}} \gamma_2 - \mathbf{A}_{\delta_{c_3}} \gamma_3 \succeq \mathbf{0} \right\}, \quad \boldsymbol{\delta} \in \mathcal{D} \quad (3.12)$$

Since the dual problem (3.11) gives a lower-bound for the optimization problems of  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$ , consequently, the function (3.12) is less than or equal to  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  for every  $\boldsymbol{\delta} \in \mathcal{D}$ . The optimization problem of  $h(\boldsymbol{\delta})$  is convex and satisfies the Slater's condition [70] for every  $\boldsymbol{\delta} \in \mathcal{D}$ , if there exists a strictly feasible point for its dual problem (3.11). Specifically if there exists a triple  $(\theta_1, \theta_2, \theta_3)$  such that the matrix  $\mathbf{A}_0 - \mathbf{A}_{\delta_{c_1}} \theta_1 - \mathbf{A}_{\delta_{c_2}} \theta_2 - \mathbf{A}_{\delta_{c_3}} \theta_3$  is positive definite, then the point  $(\gamma_1 = \theta_1, \delta_2 = \theta_2, \delta_3 = \theta_3)$  is a strictly feasible point for the dual problem (3.11). It is assumed that such triple exists. Therefore, the duality gap between the optimization problem of  $h(\boldsymbol{\delta})$  and its dual problem (3.11) is zero [70] which implies that for every  $\boldsymbol{\delta} \in \mathcal{D}$ ,  $h(\boldsymbol{\delta}) = l(\boldsymbol{\delta})$ .

Regarding the optimization problem of  $k(\boldsymbol{\delta})$  which is a QCQP [17]–[22], [106] it is known that the duality gap between a QCQP problem with three or less constraints and its dual problem is zero [106]. Specifically, Corollary 3.3 of [106, Section 3] implies that the duality gap between the optimization problem of  $k(\boldsymbol{\delta})$  and its dual optimization problem (3.11), is zero and hence,  $k(\boldsymbol{\delta}) = l(\boldsymbol{\delta})$ . Since both of the optimization problems of  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  have zero duality gap with their dual problem (3.11), it can be concluded that in addition to having the same domain, the functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  have the same optimal values, i.e.,  $k(\boldsymbol{\delta}) = h(\boldsymbol{\delta})$  for every feasible  $\boldsymbol{\delta}$ .

For any feasible point of the optimization problem of  $k(\boldsymbol{\delta})$  denoted as  $\mathbf{x}_0$ , the matrix  $\mathbf{x}_0\mathbf{x}_0^H$  is also a feasible point of the optimization problem of  $h(\boldsymbol{\delta})$  and their corresponding objective values are the same. Based on the later fact and also the fact that the functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  have the same optimal values, it can be concluded that if  $x_{\boldsymbol{\delta}}^{\text{opt}}$  denotes the optimal solution of the optimization problem of  $k(\boldsymbol{\delta})$ , then  $x_{\boldsymbol{\delta}}^{\text{opt}}(x_{\boldsymbol{\delta}}^{\text{opt}})^H$  is also the optimal solution of the optimization problem of  $h(\boldsymbol{\delta})$ . Therefore, for every  $\boldsymbol{\delta} \in \mathcal{D}$ , there exists a rank-one solution for the optimization problem of  $h(\boldsymbol{\delta})$ . The algorithm for constructing such a rank-one solution from a general-rank solution of the optimization problem of  $h(\boldsymbol{\delta})$  has been explained in [106].  $\square$

Although the optimal value functions  $k(\boldsymbol{\delta})$  and  $h(\boldsymbol{\delta})$  are equal, however, compared to the optimization problem of  $k(\boldsymbol{\delta})$  which is non-convex, the optimization problem of  $h(\boldsymbol{\delta})$  is convex. Using this fact and replacing  $k(\boldsymbol{\delta})$  by  $h(\boldsymbol{\delta})$  in the original optimization problem (3.3), this problem can be simplified as

$$\begin{aligned} & \min_{\mathbf{y}, \boldsymbol{\delta}, \mathbf{X}} && f_0(\text{tr}\{\mathbf{A}_0\mathbf{X}\}) + h_0(\mathbf{y}) \\ \text{subject to} &&& \text{tr}\{\mathbf{A}_i\mathbf{X}\} = \delta_i, \quad i \in C, \quad \mathbf{X} \succeq \mathbf{0}, \\ &&& \alpha_{2i-1}f_{2i-1}(\delta_{2i-1}) - \alpha_{2i}f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M. \end{aligned} \quad (3.13)$$

Therefore, instead of the original optimization problem (3.3), we can solve the simplified problem (3.13) in which the quadratic functions have been replaced with their corresponding linear functions. It is noteworthy to mention that in the simplified problem, the non-convex functions  $f_i(\mathbf{x}^H\mathbf{A}_i\mathbf{x})$ ,  $i = \{0\} \cup C$  are replaced by the convex functions  $f_i(\text{tr}\{\mathbf{A}_i\mathbf{X}\})$ ,  $i = \{0\} \cup C$ . The latter is due to the fact that the composition of a convex function with a linear function is also a convex function. Based on the optimal solution of the simplified problem, denoted as  $\mathbf{X}_{\text{opt}}$ ,  $\boldsymbol{\delta}_{\text{opt}}$ , and

$\mathbf{y}_{\text{opt}}$ , the optimal solution of the original problem can be found. The optimal values of  $\boldsymbol{\delta}$  and  $\mathbf{y}$  are equal to the corresponding optimal values of the simplified problem, while, the optimal value of  $\mathbf{x}$  can be constructed based on  $\mathbf{X}_{\text{opt}}$  using rank-reduction techniques [106] mentioned in the proof of Theorem 3.1.

If the corresponding coefficients of the functions  $f_{2i}, i = 1, \dots, M$ , i.e.,  $\alpha_{2i}$ , are all zero, then the problem (3.13) is convex and it can be easily solved. Particular, in this case, the objective function and the constraint functions of the simplified problem (3.13) are all convex. Once this problem is solved, the optimal  $\mathbf{x}$  can be extracted using Theorem 3.1. However if any of such coefficients is non-zero, the problem (3.13) is no longer convex and there exists a constraint which is the difference of two convex function and, therefore, the problem (3.13) is a DC programming problem. In this case, although the problem (3.13) boils down to the known family of DC programming problems, still there exists no solution for such DC programming problems with guaranteed polynomial time complexity. The typical approach for solving such problems is the various modifications of the branch-and-bound method [13], [31]–[36] which is an effective global optimization method. However, it does not have any worst-case polynomial time complexity guarantees [32] and [33], which significantly limits or even prohibits its applicability in practical communication systems. Thus, methods with guaranteed polynomial time complexity that can find at least a suboptimal solution for different types of DC programming problems are of a great importance. In what follows, we establish an iterative method for solving the problem (3.13) when at least one of the coefficients  $\alpha_{2i}, i = 1, \dots, M$  is non-zero and therefore the relaxed problem is DC.

### 3.1.1 Polynomial time DC algorithm

We develop an iterative method for solving the DC problem (3.13) at least suboptimally. The essence of the proposed method is to linearize the non-convex one-dimensional functions  $-f_{2i}(\delta_{2i})$  appearing in the constraints  $\alpha_{2i-1}f_{2i-1}(\delta_{2i-1}) - \alpha_{2i}f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0$  around suitably selected points in different iterations. This new proposed method will be referred to as the Polynomial time DC (POTDC). As it will be discussed shortly, it is guaranteed that POTDC finds at least a KKT point, i.e., a point which satisfies the KKT optimality conditions. In order to explain the intuition behind this method, let us replace the non-convex functions  $-f_{2i}(\delta_{2i}), i \in \mathcal{K} \triangleq \{i \mid \alpha_{2i} = 1\}$  by their corresponding linear approximations

around the points  $\delta_{2i,\text{Lin}}, i \in \mathcal{K}$ , i.e.,

$$-f_{2i}(\delta_{2i}) \approx -f_{2i}(\delta_{2i,\text{Lin}}) - \frac{df_{2i}(\delta_{2i})}{d\delta_{2i}} \Big|_{\delta_{2i}=\delta_{2i,\text{Lin}}} (\delta_{2i} - \delta_{2i,\text{Lin}}). \quad (3.14)$$

Performing such replacement results in the following optimization problem

$$\begin{aligned} \min_{\mathbf{y}, \delta, \mathbf{X}} \quad & f_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}\}) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}_i \mathbf{X}\} = \delta_i, \quad i \in C, \quad \mathbf{X} \succeq \mathbf{0}, \\ & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M, \quad i \notin \mathcal{K} \\ & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - f_{2i}(\delta_{2i,\text{Lin}}) - \frac{df_{2i}(\delta_{2i})}{d\delta_{2i}} \Big|_{\delta_{2i}=\delta_{2i,\text{Lin}}} (\delta_{2i} - \delta_{2i,\text{Lin}}) \\ & \quad + h_i(\mathbf{y}) \leq 0, \quad i \in \mathcal{K}. \end{aligned} \quad (3.15)$$

As compared to the original problem (3.13), the relaxed problem (3.15) is convex and can be efficiently solved up to a desired accuracy using the interior point-based numerical methods. For the fixed values of  $\delta_{2i}, i \in \mathcal{K}$  denoted as  $\Delta$ , let us define the optimal value functions  $f(\Delta)$  and  $g(\Delta, \Delta_{\text{Lin}})$  as the optimal value of the optimization problems (3.13) and (3.15), respectively, in which  $\Delta_{\text{Lin}}$  denotes the set of linearizing points, i.e.,  $\delta_{2i,\text{Lin}}, i \in \mathcal{K}$ . Since the optimization problem (3.15) is convex, its corresponding optimal value function  $g(\Delta, \Delta_{\text{Lin}})$  is also convex with respect to  $\Delta$  [108]. Furthermore, the optimal value function  $g(\Delta, \Delta_{\text{Lin}})$  provides an upper-bound for the optimal value function  $f(\Delta)$ , i.e.,  $f(\Delta) \leq g(\Delta, \Delta_{\text{Lin}})$ . The latter is due to the fact that the feasible set of the optimization problem (3.15) is a subset of the feasible set of the problem (3.13). Besides, with the assumption that the aforementioned optimal value functions are differentiable, it will be formally proved in Chapter 6, that the optimal value function  $g(\Delta, \Delta_{\text{Lin}})$  is tangent to  $f(\Delta)$  at  $\Delta = \Delta_{\text{Lin}}$ . The proof in Chapter 6 is for the case that the optimal value functions  $f(\Delta)$  and  $g(\Delta, \Delta_{\text{Lin}})$  are one-dimensional. However, the proof can be easily generalized for the multi-dimensional optimal value functions in a straightforward way. Since the aforementioned optimal value functions are tangent at the linearizing point, i.e.,  $\Delta_{\text{Lin}}$ , and  $f(\Delta)$  is upper-bounded by  $g(\Delta, \Delta_{\text{Lin}})$ , the optimal minimizer of the function  $g(\Delta, \Delta_{\text{Lin}})$  denoted as  $\Delta_{\text{opt}}$  is a decreasing point for  $f(\Delta)$ , that is,  $f(\Delta_{\text{Lin}}) \geq f(\Delta_{\text{opt}})$ . Based on this observation, the POTDC method first solves the problem (3.15) for the arbitrary chosen initial point. Once the optimal solution of this problem, denoted in the first iteration as  $\mathbf{y}_{\text{opt}}^{(1)}$ ,  $\mathbf{X}_{\text{opt}}^{(1)}$ , and  $\delta_{\text{opt}}^{(1)}$  is found, the algorithm proceeds to the second iteration by replacing the functions  $-f_{2i}(\delta_{2i}), i \in \mathcal{K}$  by their linear approximation around  $\delta_{\text{opt},2i}^{(1)}, i \in \mathcal{K}$ , respectively, found from the

previous (initially first) iteration. In the second iteration, the resulting optimization problem has the same structure as the problem (3.15) in which  $\delta_{2i,\text{Lin}}, i \in \mathcal{K}$  has to be set to  $\delta_{\text{opt},2i}^{(1)}, i \in \mathcal{K}$  obtained from the first iteration. This process continues, and  $k$ th iteration is obtained by replacing  $-f_{2i}(\delta_{2i}), i \in \mathcal{K}$  by its linearization of type (7.2) using  $\delta_{\text{opt}}^{(k-1)}$  found at the iteration  $k - 1$ . The POTDC algorithm for solving the problem (3.13) is summarized in Algorithm 3.1.

---

**Algorithm 3.1** The POTDC algorithm for solving the optimization problem (3.13)

---

**Initialize:** Arbitrarily select feasible points denoted as  $\delta_{2i,\text{Lin}}, i \in \mathcal{K}$ , set the counter  $k$  to be equal to 1.

**while** The termination condition is satisfied. **do**

Use the linearization of type (3.14) and solve the optimization problem (3.15) to obtain  $\mathbf{y}_{\text{opt}}^{(k)}, \mathbf{X}_{\text{opt}}^{(k)}$ , and  $\delta_{\text{opt}}^{(k)}$ .

Set  $\mathbf{X}_{\text{opt}} = \mathbf{X}_{\text{opt}}^{(k)}, \mathbf{y}_{\text{opt}} = \mathbf{y}_{\text{opt}}^{(k)}$ , and  $\delta_{2i,\text{Lin}} = \delta_{\text{opt},2i}^{(k)}, i \in \mathcal{K}$ .  
 $k = k + 1$ .

**end while**

**Output:**  $\mathbf{X}_{\text{opt}}$  and  $\mathbf{y}_{\text{opt}}$ .

---

The following lemma about the convergence of the proposed POTDC algorithm and the optimality of the point obtained by this algorithm is in order. Note that this lemma makes no assumptions about the differentiability of the optimal value function  $f(\Delta)$ .

**Lemma 3.2.** *The following statements regarding Algorithm 3.1 are true:*

- i) *The optimal value of the optimization problem in Algorithm 3.1 is non-increasing over iterations, i.e.,*

$$f_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}_{\text{opt}}^{(k+1)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k+1)}) \leq f_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}_{\text{opt}}^{(k)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k)}), \quad k \geq 1.$$

- ii) *The sequence of the optimal values which are generated by Algorithm 3.1, i.e.,  $f_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}_{\text{opt}}^{(k)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k)}), k = 1, 2, 3, \dots$ , are convergent. Termination condition is not considered.*

- iii) *If the Algorithm 3.1 (without considering termination condition) converges to a regular point, that point is a KKT point, i.e., a point which satisfies the KKT optimality conditions.*

*Proof.* i) Considering the linearized problem (3.15) at the iteration  $k + 1$ . It is easy to verify that  $\mathbf{X}_{\text{opt}}^{(k)}$ ,  $\mathbf{y}_{\text{opt}}^{(k)}$ , and  $\boldsymbol{\delta}_{\text{opt}}^{(k)}$  is a feasible point for this problem. Therefore, the optimal value at the iteration  $k + 1$  must be less than or equal to the optimal value at the iteration  $k$  which completes the proof.

ii) Since the sequence of the optimal values which are generated by Algorithm 3.1, i.e.,  $f_0(\text{tr}\{\mathbf{A}_0\mathbf{X}_{\text{opt}}^{(k)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k)})$ ,  $k \geq 1$ , is non-increasing and bounded from below, it is convergent [107]. Note that the sequence of the optimal values is bounded from below because the functions  $f_0$  and  $h_0$  are assumed to be bounded from below over the feasible set.

iii) The proof follows straightforwardly from Proposition 3.2 of [15, Section 3]. The essence of the proof in the aforementioned proposition is to show that if the sequence of the points generated by Algorithm 3.1, i.e.,  $\mathbf{X}_{\text{opt}}^{(k)}$ ,  $\mathbf{y}_{\text{opt}}^{(k)}$ , and  $\boldsymbol{\delta}_{\text{opt}}^{(k)}$ , converges to a regular point, then the sequence of the gradients of the objective and constraint functions at the points  $\mathbf{X}_{\text{opt}}^{(k)}$ ,  $\mathbf{y}_{\text{opt}}^{(k)}$ , and  $\boldsymbol{\delta}_{\text{opt}}^{(k)}$  also converges to the corresponding gradients at the limiting point. Moreover, with such assumption, the convergence of the Lagrangian multipliers is also proved. It is shown as well that, the limiting point of the Lagrangian multipliers and the achieved point satisfy the KKT optimality conditions.  $\square$

The termination condition is needed to stop the iterative Algorithm 3.1 when the value achieved by the algorithm is deemed close enough to the optimal solution. The fact that the sequence of optimal values generated by Algorithm 3.1 is non-increasing and convergent (see Lemma 3.2) can be used for choosing the termination condition. Based on the latter fact, the algorithm can be terminated if an improvement to the value of the objective function is less than a certain desired threshold, i.e.,  $f_0(\text{tr}\{\mathbf{A}_0\mathbf{X}_{\text{opt}}^{(k)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k)}) - f_0(\text{tr}\{\mathbf{A}_0\mathbf{X}_{\text{opt}}^{(k+1)}\}) + h_0(\mathbf{y}_{\text{opt}}^{(k+1)}) \leq \zeta$ , where  $\zeta$  is said to be the progress parameter. However, such termination condition may stop the iterative algorithm prematurely. Specifically, if the value of the objective function does not change on two consecutive iterations, but the last achieved point is not close enough to a KKT point, such termination criterion stops the iterations. In order to avoid this situation, one can define the termination condition based on approximate satisfaction of the KKT optimality conditions [15]. Specifically, the point  $\mathbf{X}_{\text{opt}}^{(k)}$ ,  $\mathbf{y}_{\text{opt}}^{(k)}$ , and  $\boldsymbol{\delta}_{\text{opt}}^{(k)}$  achieved at iteration  $k$  satisfies the KKT optimality

conditions approximately if  $\epsilon_k \leq \epsilon$  where  $\epsilon_k$  is defined as

$$\begin{aligned} \epsilon_k &\triangleq \min_{\lambda_i, \mathbf{Z}} \quad \|\nabla_{\mathbf{X}} L(\mathbf{X}_{\text{opt}}^{(k)}, \mathbf{y}_{\text{opt}}^{(k)}, \lambda_i, \mathbf{Z})\|^2 + \|\nabla_{\mathbf{y}} L(\mathbf{X}_{\text{opt}}^{(k)}, \mathbf{y}_{\text{opt}}^{(k)}, \lambda_i, \mathbf{Z})\|^2 \\ \text{subject to} \quad &\lambda_i \geq 0, i = 1, \dots, M, \quad \mathbf{Z} \succeq \mathbf{0}, \quad \text{tr}\{\mathbf{Z} \mathbf{X}_{\text{opt}}^{(k)}\} = 0 \\ &\lambda_i = 0 \quad \text{if} \quad \alpha_{2i-1} f_{2i-1}(\text{tr}\{\mathbf{A}_{2i-1} \mathbf{X}_{\text{opt}}^{(k)}\}) - \alpha_{2i} f_{2i}(\text{tr}\{\mathbf{A}_{2i} \mathbf{X}_{\text{opt}}^{(k)}\}) \\ &\quad + h_i(\mathbf{y}_{\text{opt}}^{(k)}) < 0, i = 1, \dots, M \end{aligned} \quad (3.16)$$

and  $L(\mathbf{X}, \mathbf{y}, \lambda_i, \mathbf{Z})$  is the Lagrangian function of the optimization problem (3.13) defined as

$$\begin{aligned} L(\mathbf{X}, \mathbf{y}, \lambda_i, \mathbf{Z}) &\triangleq f_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}\}) + h_0(\mathbf{y}) + \sum_{i=1}^M \lambda_i \left( \alpha_{2i-1} f_{2i-1}(\text{tr}\{\mathbf{A}_{2i-1} \mathbf{X}\}) \right. \\ &\quad \left. - \alpha_{2i} f_{2i}(\text{tr}\{\mathbf{A}_{2i} \mathbf{X}\}) + h_i(\mathbf{y}) \right) - \text{tr}\{\mathbf{Z} \mathbf{X}\}. \end{aligned} \quad (3.17)$$

Here  $\epsilon$  is the accuracy parameter and  $\lambda_i$ ,  $i = 1, \dots, M$  is the Lagrangian multiplier associated with the constraint

$$\alpha_{2i-1} f_{2i-1}(\text{tr}\{\mathbf{A}_{2i-1} \mathbf{X}\}) - \alpha_{2i} f_{2i}(\text{tr}\{\mathbf{A}_{2i} \mathbf{X}\}) + h_i(\mathbf{y}) \leq 0. \quad (3.18)$$

Gradient of the Lagrangian function which appears in the optimization problem (3.17) can be simply derived as

$$\begin{aligned} \nabla_{\mathbf{X}} L(\mathbf{X}, \mathbf{y}, \lambda_i, \mathbf{Z}) &= \mathbf{A}_0 f'_0(\text{tr}\{\mathbf{A}_0 \mathbf{X}\}) + \sum_{i=1}^M \lambda_i \left( \alpha_{2i-1} \mathbf{A}_{2i-1} f'_{2i-1}(\text{tr}\{\mathbf{A}_{2i-1} \mathbf{X}\}) \right. \\ &\quad \left. - \alpha_{2i} \mathbf{A}_{2i} f'_{2i}(\text{tr}\{\mathbf{A}_{2i} \mathbf{X}\}) \right) - \mathbf{Z} \end{aligned} \quad (3.19)$$

and

$$\nabla_{\mathbf{y}} L(\mathbf{X}, \mathbf{y}, \lambda_i, \mathbf{Z}) = h'_0(\mathbf{y}) + \sum_{i=1}^M \lambda_i h'_i(\mathbf{y}) \quad (3.20)$$

where  $(\cdot)'$  denotes the derivative operator. Please note that the optimization problem (3.17) is a convex problem which can be easily solved.

### 3.2 Number of quadratic functions in the constraints greater than three

When the total number of the composite functions  $f_i(\mathbf{x}^H \mathbf{A}_i \mathbf{x})$  in the constraints of the optimization problem (3.1) or equivalently  $\text{card}\{C\}$  is greater than or equal to four, the optimal value functions  $k(\boldsymbol{\delta})$  (3.6) and  $h(\boldsymbol{\delta})$  (3.9) may not be equal

in general. In this case, the optimal value function  $h(\boldsymbol{\delta})$  is a lower-bound of  $k(\boldsymbol{\delta})$ , i.e.,  $h(\boldsymbol{\delta}) \leq k(\boldsymbol{\delta})$  ( $h(\boldsymbol{\delta})$  is an upper-bound of  $k(\boldsymbol{\delta})$  if  $f_0$  is decreasing). This can be easily verified by considering the fact that if  $\mathbf{x}$  is a feasible point of the corresponding optimization problem of the optimal value function  $k(\boldsymbol{\delta})$ , then  $\mathbf{x}\mathbf{x}^H$  is also a feasible point of the corresponding optimal value function  $h(\boldsymbol{\delta})$  which implies that  $h(\boldsymbol{\delta}) \leq k(\boldsymbol{\delta})$ . By replacing the optimal value function  $k(\boldsymbol{\delta})$  by  $h(\boldsymbol{\delta})$  in the original optimization problem (3.5), this problem can be approximated by the following DC problem

$$\begin{aligned} \min_{\mathbf{y}, \boldsymbol{\delta}} \quad & f_0(h(\boldsymbol{\delta})) + h_0(\mathbf{y}) \\ \text{subject to} \quad & \alpha_{2i-1} f_{2i-1}(\delta_{2i-1}) - \alpha_{2i} f_{2i}(\delta_{2i}) + h_i(\mathbf{y}) \leq 0, \quad i = 1, \dots, M. \end{aligned} \tag{3.21}$$

Since  $f_0$  is a monotonic function, it can be concluded that the objective function of the optimization problem (3.21) is a lower-bound of the original problem. Thus, instead of the original objective function, the problem (3.21) aims at minimizing a lower-bound of the objective function. The problem (3.21) can be similarly solved by using the proposed POTDC method and the results of Lemma 3.2 hold true. Indeed, it is guaranteed that the POTDC finds a KKT point of the problem (3.21). The extraction of the in general suboptimal solution of the original problem from the optimal solution of the approximate problem (3.21) is done through randomization techniques.

## Chapter 4

# Transmit Beamspace Design for DOA Estimation in MIMO Radar

In array processing applications, the DOA parameter estimation problem is most fundamental [85]. Many DOA estimation techniques have been developed for the classical array processing single-input multiple output (SIMO) setup which is referred to as the phased-array radar [82]–[90]. The development of the MIMO radar [37], [75] has opened new opportunities in parameter estimation. The virtual array with a larger number of virtual antenna elements in a colocated MIMO radar can be used for improved DOA estimation performance as compared to traditional radar systems [39], [109] for relatively high SNRs, i.e., when the benefits of increased virtual aperture start to show up. The SNR gain for fully MIMO radar, i.e., the case when all the transmit signals are orthogonal (see Section 2.2), however, decreases as compared to the phased-array radar where the transmit array radiates a single waveform coherently from all antenna elements [38], [110]. A trade-off between the phased-array and fully MIMO radar system can be achieved [38], [111], [112] which gives the best of both configurations, i.e., the increased number of virtual antenna elements due to the use of waveform diversity together with SNR gain due to subaperture based coherent transmission.

Several transmit beamforming techniques have been developed in the literature to achieve transmit coherent gain in MIMO radar under the assumption that the general angular locations of the targets are known a priori to be located within a certain spacial sector. The increased number of degrees of freedom for MIMO radar due to the use of multiple waveforms is used for the purpose of synthesizing

a desired transmit beampattern based on optimizing the correlation matrix of the transmitted waveforms [37], [42], [43], [113]. To apply the designs obtained using the aforementioned methods, the actual waveforms still have to be found which is a relatively difficult and computationally demanding problem [114].

One of the major motivations for designing the transmit beampattern is realizing the possibility of achieving SNR gain together with increased aperture for improved DOA estimation in a wide range of SNRs [40], [41]. In particular, it has been shown in [41] that the performance of a MIMO radar system with a number of waveforms less than the number of transmit antennas associated with using transmit beamforming gain is better than the performance of a MIMO radar system with full waveform diversity with no transmit beamforming gain. Remarkably, using MIMO radar with proper transmit beamspace design, it is possible to achieve and guarantee the satisfaction of such desired property for DOA estimation as RIP at the receiver [41]. This is somewhat similar in effect to the property of orthogonal space-time block codes in that the shape of the transmitted constellation does not change at the receiver independent on a channel. The latter allows for simple decoder [115]. Similarly, here the RIP allows for simple DOA estimation techniques at the receiver although the RIP is actually enforced at the transmitter, and the propagation media cannot break it thanks to the proper design of beamspace. Since the RIP holds at the receiver independent on the propagation media and receive antenna array configuration, the receive antenna array can be an arbitrary array. However, the methods developed in [40], [41] suffer from the shortcomings that the transmit power distribution across the array elements is not uniform and the achieved phase rotations comes with variations in the magnitude of different transmit beams which may affect the performance of DOA estimation at the receiver.

In this chapter, we consider the problem of transmit beamspace design for DOA estimation in MIMO radar with colocated antennas. We propose a new method for designing the transmit beamspace that enables the use of search-free DOA estimation techniques at the receiver. The essence of the proposed method is to design the transmit beamspace matrix based on minimizing the difference between a desired transmit beampattern and the actual one while enforcing the uniform power distribution constraint across the transmit array elements. The desired transmit beampattern can be of arbitrary shape and is allowed to consist of one or more spatial sectors.

The case of even but otherwise arbitrary number of transmit waveforms is considered. To allow for a simple search-free DOA estimation algorithms at the receiver, the RIP is established at the transmit array by imposing a specific structure on the transmit beamspace matrix. The proposed structure is based on designing the transmit beams in pairs where the transmit weight vector associated with a certain transmit beam is the conjugate flipped version of the weight vector associated with another beam, i.e., one transmit weight vector is designed for each pair of transmit beams. All pairs are designed jointly while satisfying the requirements that the two transmit beams associated with each pair enjoy rotational invariance with respect to each other. Semi-definite relaxation is used to relax the proposed formulations into a convex problem that can be solved efficiently using, for example, interior point methods.

In comparison to previous methods that achieve phase rotation between two transmit beams, the proposed method enjoys the following advantages: (i) It ensures that the magnitude response of the two transmit beams associated with any pair is exactly the same at all spatial directions, a property that improves the DOA estimation performance; (ii) It ensures uniform power distribution across transmit antenna elements; (iii) It enables estimating the DOAs via estimating the accumulated phase rotations over all transmit beams instead of only two beams; (iv) It only involves optimization over half the entries of the transmit beamspace matrix which decreases the computational load.

We also propose an alternative formulation based on splitting the overall transmit beamspace design problem into several smaller problems. The alternative formulation is referred to as the spatial-division based design (SDD) which involves dividing the spatial domain into several subsectors and assigning a subset of transmit beamspace pairs to each subsector. The SDD method enables post processing of data associated with different subsectors independently with estimation performance comparable to the performance of the joint transmit beamspace design. Simulation results demonstrate the improvement in the DOA estimation performance that can be achieved by using the proposed joint transmit beamspace design and SDD methods as compared to the traditional MIMO radar technique.

The rest of this chapter is organized as follows. Section 4.1 introduces the system model for mono-static MIMO radar system with transmit beamspace. The problem formulations in terms of the unknown transmit beamspace matrix is derived in

Section 4.2. The transmit beamspace design problem for the case of two transmit waveforms as well as for the general case of even but otherwise arbitrary number of transmit waveforms is considered in Section 4.3. Section 4.4 gives simulation examples for the DOA estimation in a MIMO radar with the proposed transmit beamspace and the concluding remarks are drawn in Section 4.5.

## 4.1 System model

Consider a mono-static MIMO radar system equipped with a ULA of  $M$  colocated antennas with inter-element spacing  $d$  measured in wavelength and a receive array of  $N$  antennas configured in a random shape. The transmit and receive arrays are assumed to be close enough to each other such that the spatial angle of a target in the far-field remains the same with respect to both arrays. Let  $\boldsymbol{\phi}(t) = [\phi_1(t), \dots, \phi_K(t)]^T$  be the  $K \times 1$  vector that contains the complex envelopes of the waveforms  $\phi_k(t)$ ,  $k = 1, \dots, K$  which are assumed to be orthogonal, i.e.,

$$\int_0^{T_p} \phi_i(t)\phi_j^*(t) dt = \delta\{i - j\}, \quad i, j = 1, 2, \dots, K \quad (4.1)$$

where  $T_p$  is the pulse duration,  $(\cdot)^*$  stands for the conjugate, and  $\delta\{\cdot\}$  is the Kronecker delta. The actual transmitted signals are taken as linear combinations of the orthogonal waveforms. Therefore, the  $M \times 1$  vector of the baseband representation of the transmitted signals can be written as [41]

$$\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^T = \mathbf{W}\boldsymbol{\phi}(t) \quad (4.2)$$

where  $s_i(t)$  is the signal transmitted from antenna  $i$  and

$$\mathbf{W} = \begin{pmatrix} w_{1,1} & w_{2,1} & \cdots & w_{K,1} \\ w_{1,2} & w_{2,2} & \cdots & w_{K,2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,M} & w_{2,M} & \cdots & w_{K,M} \end{pmatrix} \quad (4.3)$$

is the  $M \times K$  transmit beamspace matrix. It is worth noting that each of the orthogonal waveforms  $\phi_k(t)$ ,  $k = 1, \dots, K$  is transmitted over one transmit beam where the  $k$ th column of the matrix  $\mathbf{W}$  corresponds to the transmit beamforming weight vector used to form the  $k$ th beam.

Let  $\mathbf{a}(\theta) \triangleq [1, e^{-j2\pi d \sin(\theta)}, \dots, e^{-j2\pi d(M-1) \sin(\theta)}]^T$  be the  $M \times 1$  transmit array steering vector. The transmit power distribution pattern can be expressed as [43]

$$G(\theta) = \frac{1}{4\pi} \mathbf{d}^H(\theta) \mathbf{R} \mathbf{d}(\theta), \quad -\pi/2 \leq \theta \leq \pi/2 \quad (4.4)$$

where  $\mathbf{d}(\theta) = \mathbf{a}^*(\theta)$ , and

$$\mathbf{R} = \int_0^{T_p} \mathbf{s}(t)\mathbf{s}^H(t)dt \quad (4.5)$$

is the cross-correlation matrix of the transmitted signals (4.2). One way to achieve a certain desired transmit beampattern is to optimize over the cross-correlation matrix  $\mathbf{R}$  such as in [43], [113]. In this case, a complementary problem has to be solved after obtaining  $\mathbf{R}$  in order to find appropriate signal vector  $\mathbf{s}(t)$  that satisfies (4.5). Solving such a complementary problem is in general difficult and computationally demanding [114]. However, in this chapter, we extend our approach of optimizing the transmit beampattern via designing the transmit beamspace matrix. According to this approach, the cross-correlation matrix is expressed as

$$\mathbf{R} = \mathbf{W}\mathbf{W}^H \quad (4.6)$$

that holds due to the orthogonality of the waveforms (see (4.1) and (4.2)). Then the transmit beamspace matrix  $\mathbf{W}$  can be designed to achieve the desired beampattern while satisfying many other requirements mandated by practical considerations such as equal transmit power distribution across the transmit array antenna elements, achieving a desired radar ambiguity function, etc. Moreover, this approach enables enforcing the RIP which facilitates subsequent processing steps at the receive antenna array, e.g., it enables applying accurate computationally efficient DOA estimation using search-free direction finding techniques such as ESPRIT.

Similar to the Section 2.2, the signal measured at the output of the receive array due to echoes from  $L$  narrowband far-field targets can be modeled as

$$\mathbf{x}(t, \varrho) = \sum_{l=1}^L \beta_l(\varrho) [\mathbf{d}^H(\theta_l)\mathbf{W}\phi(t)] \mathbf{b}(\theta_l) + \mathbf{z}(t, \varrho) \quad (4.7)$$

where  $t$  is the time index within the radar pulse,  $\varrho$  is the slow time index, i.e., the pulse number,  $\beta_l(\varrho)$  is the reflection coefficient of the target located at the unknown spatial angle  $\theta_l$ ,  $\mathbf{b}(\theta_l)$  is the receive array steering vector, and  $\mathbf{z}(t, \varrho)$  is the  $N \times 1$  vector of zero-mean white Gaussian noise with variance  $\sigma_z^2$ . In (4.7), the target reflection coefficients  $\beta_l(\varrho)$ ,  $l = 1, \dots, L$  are assumed to obey the Swerling II model [109], i.e., they remain constant during the duration of one radar pulse but change from pulse to pulse. Moreover, they are assumed to be drawn from a normal distribution with zero mean and variance  $\sigma_\beta^2$ .

By matched filtering  $\mathbf{x}(t, \varrho)$  to each of the orthogonal basis waveforms  $\phi_k(t)$ ,  $k = 1, \dots, K$ , the  $N \times 1$  virtual data vectors can be obtained as<sup>1</sup>

$$\begin{aligned} \mathbf{y}_k(\varrho) &\triangleq \int_{\mathbb{T}_p} \mathbf{x}(t, \varrho) \phi_k^*(t) dt \\ &\triangleq \sum_{l=1}^L \beta_l(\varrho) (\mathbf{d}^H(\theta_l) \mathbf{w}_k) \mathbf{b}(\theta_l) + \mathbf{z}_k(\varrho) \end{aligned} \quad (4.8)$$

where  $\mathbf{w}_k$  is the  $k$ th column of the transmit beamspace matrix  $\mathbf{W}$  and  $\mathbf{z}_k(\varrho) \triangleq \int_{\mathbb{T}_p} \mathbf{z}(t, \varrho) \phi_k^*(t) dt$  is the  $N \times 1$  noise term whose covariance is  $\sigma_z^2 \mathbf{I}_N$ .

Let  $\check{\mathbf{y}}_{l,k}(\varrho)$  be the noise free component of the virtual data vector (4.8) associated with the  $l$ th target, i.e.,  $\check{\mathbf{y}}_{l,k}(\varrho) \triangleq \beta_l(\varrho) (\mathbf{d}^H(\theta_l) \mathbf{w}_k) \mathbf{b}(\theta_l)$ . Then, one can easily observe that the  $k$ th and the  $k'$ th components associated with the  $l$ th target are related to each other through the following relationship

$$\begin{aligned} \check{\mathbf{y}}_{l,k'}(\varrho) &= \beta_l(\varrho) (\mathbf{d}^H(\theta_l) \mathbf{w}_{k'}) \mathbf{b}(\theta_l) \\ &= \frac{\mathbf{d}^H(\theta_l) \mathbf{w}_{k'}}{\mathbf{d}^H(\theta_l) \mathbf{w}_k} \cdot \check{\mathbf{y}}_{l,k}(\varrho) \\ &= e^{j(\psi_{k'}(\theta_l) - \psi_k(\theta_l))} \frac{|\mathbf{d}^H(\theta_l) \mathbf{w}_{k'}|}{|\mathbf{d}^H(\theta_l) \mathbf{w}_k|} \cdot \check{\mathbf{y}}_{l,k}(\varrho) \end{aligned} \quad (4.9)$$

where  $\psi_k(\theta)$  is the phase of the inner product  $\mathbf{d}^H(\theta) \mathbf{w}_k$ . The expression (4.9) means that the signal component  $\mathbf{y}_k(\varrho)$  corresponding to a given target is the same as the signal component  $\mathbf{y}_{k'}(\varrho)$  corresponding to the same target up to a phase rotation and a gain factor.

The RIP can be enforced by imposing the constraint  $|\mathbf{d}^H(\theta) \mathbf{w}_k| = |\mathbf{d}^H(\theta) \mathbf{w}_{k'}|$  while designing the transmit beamspace matrix  $\mathbf{W}$ . The main advantage of enforcing the RIP is that it allows us to estimate DOAs via estimating the phase rotation associated with the  $k$ th and  $k'$ th pair of the virtual data vectors using search-free techniques, e.g., ESPRIT. Moreover, if the number of transmit waveforms is more than two, the DOA estimation can be carried out via estimating the phase difference

$$\angle \sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i - \angle \sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i \quad (4.10)$$

and comparing it to a precalculated phase profile for the given spatial sector in which we have concentrated the transmit power. However, in the latter case, precautions

---

<sup>1</sup>Practically, this matched filtering step is performed for each Doppler-range bin, i.e., the received data  $\mathbf{x}(t, \varrho)$  is matched filtered to a time-delayed Doppler shifted version of the waveforms  $\phi_k(t)$ ,  $k = 1, \dots, K$ .

should be taken to assure the coherent accumulation of the  $K/2$  components in (4.10), i.e., to avoid gain loss as will be shown later in the chapter.

## 4.2 Problem formulation

The main goal is to design a transmit beamspace matrix  $\mathbf{W}$  which achieves a spatial beampattern that is as close as possible to a certain desired one. Substituting (4.6) in (4.4), the spatial beampattern can be rewritten as

$$\begin{aligned} G(\theta) &= \frac{1}{4\pi} \mathbf{d}^H(\theta) \mathbf{W} \mathbf{W}^H \mathbf{d}(\theta) \\ &= \frac{1}{4\pi} \sum_{i=1}^K \mathbf{w}_i^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{w}_i. \end{aligned} \quad (4.11)$$

Therefore, we design the transmit beamspace matrix  $\mathbf{W}$  based on minimizing the difference between the desired beampattern and the actual beampattern given by (4.11). Using the minimax criterion, the transmit beamspace matrix design problem can be formulated as

$$\begin{aligned} \min_{\mathbf{W}} \max_{\theta} & \left| G_d(\theta) - \frac{1}{4\pi} \sum_{i=1}^K \mathbf{w}_i^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{w}_i \right| \\ \text{subject to} & \quad \sum_{i=1}^K |\mathbf{w}_i(j)|^2 = \frac{P_t}{M}, \quad j = 1, \dots, M \end{aligned} \quad (4.12)$$

where  $G_d(\theta)$ ,  $\theta \in [-\pi/2, \pi/2]$  is the desired beampattern and  $P_t$  is the total transmit power. The  $M$  constraints enforced in (4.12) are used to ensure that individual antennas transmit equal powers given by  $P_t/M$ . It is equivalent to having the norms of the rows of  $\mathbf{W}$  to be equal to  $P_t/M$ . The uniform power distribution across the array antenna elements given by (4.12) is necessary from a practical point of view. In practice, each antenna in the transmit array typically uses the same power amplifier, and thus has the same dynamic power range. If the power used by different antenna elements is allowed to vary widely, this can severely degrade the performance of the system due to the nonlinear characteristics of the power amplifier.

Another goal that we wish to achieve is to enforce the RIP to enable for search-free DOA estimation. Enforcing the RIP between the  $k$ th and  $(K/2+k)$ th transmit beams is equivalent to ensuring that the following relationship holds

$$|\mathbf{w}_k^H \mathbf{d}(\theta)| = \left| \mathbf{w}_{\frac{K}{2}+k}^H \mathbf{d}(\theta) \right|, \quad \theta \in [-\pi/2, \pi/2]. \quad (4.13)$$

Ensuring (4.13), the optimization problem (4.12) can be reformulated as

$$\begin{aligned}
\min_{\mathbf{W}} \max_{\theta} & \quad \left| G_d(\theta) - \frac{1}{4\pi} \sum_{i=1}^K \mathbf{w}_i^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{w}_i \right| \\
\text{subject to} & \quad \sum_{i=1}^K |\mathbf{w}_i(j)|^2 = \frac{P_t}{M}, \quad j = 1, \dots, M \\
& \quad |\mathbf{w}_k^H \mathbf{d}(\theta)| = \left| \mathbf{w}_{\frac{K}{2}+k}^H \mathbf{d}(\theta) \right|, \quad \theta \in [-\pi/2, \pi/2], \quad k = 1, \dots, \frac{K}{2}.
\end{aligned} \tag{4.14}$$

It is worth noting that the constraints  $\sum_{i=1}^K |\mathbf{w}_i(j)|^2 = P_t/M$  as well as the constraints  $|\mathbf{w}_k^H \mathbf{d}(\theta)| = \left| \mathbf{w}_{\frac{K}{2}+k}^H \mathbf{d}(\theta) \right|$  correspond to non-convex sets and, therefore, the optimization problem (4.14) is a non-convex problem which is difficult to solve in a computationally efficient manner. Moreover, the fact that that the last constraint in (4.14) should be enforced for every direction  $\theta \in [-\pi/2, \pi/2]$ , i.e., the number of equations in (4.14) is significantly larger than the number of the variables, makes it unlikely to satisfy (4.14) unless a specific structure on the transmit beamspace matrix  $\mathbf{W}$  is imposed.

In the following section we propose a specific structure to  $\mathbf{W}$  to overcome the difficulties caused by (4.14) and show how to use SDP relaxation to overcome the difficulties caused by the non-convexity of (4.14).

## 4.3 Transmit beamspace design

### 4.3.1 Two transmit waveforms

We first consider a special, but practically important case of two orthonormal waveforms. Thus, the dimension of  $\mathbf{W}$  is  $M \times 2$ . Then under the aforementioned assumption of ULA at the MIMO radar transmitter, the RIP can be satisfied by choosing the transmit beamspace matrix to take the form

$$\mathbf{W} = [\mathbf{w}, \tilde{\mathbf{w}}^*] \tag{4.15}$$

where  $\tilde{\mathbf{w}}$  is the flipped version of vector  $\mathbf{w}$ , i.e.,  $\tilde{\mathbf{w}}(i) = \mathbf{w}(M - i + 1)$ ,  $i = 1, \dots, M$  and  $(\tilde{\cdot})$  denotes the flipping operator. Indeed, in this case,  $|\mathbf{d}^H(\theta) \mathbf{w}| = |\mathbf{d}^H(\theta) \tilde{\mathbf{w}}^*|$  and the RIP is clearly satisfied.

To prove that the specific structure (4.15) achieves the RIP, let us represent the vector  $\mathbf{w}$  as a vector of complex numbers

$$\mathbf{w} = [z_1 \ z_2 \ \dots \ z_M]^T \tag{4.16}$$

where  $z_m$ ,  $m = 1, \dots, M$  are complex numbers. Then the flipped-conjugate version of  $\mathbf{w}$  has the structure  $\tilde{\mathbf{w}}^* = [z_M^* \ z_{M-1}^* \ \dots \ z_1^*]^T$ . Examining the inner products  $\mathbf{d}^H(\theta)\mathbf{w}$  and  $\mathbf{d}^H(\theta)\tilde{\mathbf{w}}^*$  we see that the first inner product produces the sum

$$\mathbf{d}^H(\theta)\mathbf{w} = z_1 + z_2 e^{-j2\pi d \sin(\theta)} + \dots + z_M e^{-j2\pi d \sin(\theta)(M-1)} \quad (4.17)$$

and the second produces the sum

$$\mathbf{d}^H(\theta)\tilde{\mathbf{w}}_i^* = z_M^* + z_{M-1}^* e^{-j2\pi d \sin(\theta)} + \dots + z_1^* e^{-j2\pi d \sin(\theta)(M-1)}. \quad (4.18)$$

Factoring out the term  $e^{-j2\pi d \sin(\theta)(M-1)}$  from (4.18) and conjugating, we can see that the sums are identical in magnitude and indeed are the same up to a phase rotation  $\psi$ . This relationship is precisely the RIP, and it is enforced at the transmit antenna array by the structure imposed on the transmit beamspace matrix  $\mathbf{W}$ .

Substituting (4.15) in (4.14), the optimization problem can be reformulated for the case of two transmit waveforms as follows

$$\begin{aligned} & \min_{\mathbf{w}} \max_{\theta} |G_d(\theta) - \|[\mathbf{w} \ \tilde{\mathbf{w}}^*]^H \mathbf{d}(\theta)\|^2| \\ & \text{subject to } |\mathbf{w}(i)|^2 + |\tilde{\mathbf{w}}(i)|^2 = \frac{P_t}{M}, \quad i = 1, \dots, M. \end{aligned} \quad (4.19)$$

It is worth noting that the last constraints in (4.14) are not shown in the optimization problem (4.19) because they are inherently enforced due to the use of the specific structure of  $\mathbf{W}$  given in (4.15).

Introducing the auxiliary variable  $\delta$  and assuming that the number of transmit antenna elements  $M$  is even<sup>2</sup>, the optimization problem (4.19) can be approximately rewritten on a grid over a certain interval as

$$\begin{aligned} & \min_{\mathbf{w}, \delta} \delta \\ & \text{subject to } \frac{G_d(\theta_q)}{2} - |\mathbf{w}^H \mathbf{d}(\theta_q)|^2 \leq \delta, \quad q = 1, \dots, Q \\ & \quad \frac{G_d(\theta_q)}{2} - |\mathbf{w}^H \mathbf{d}(\theta_q)|^2 \geq -\delta, \quad q = 1, \dots, Q \\ & \quad |\mathbf{w}(i)|^2 + |\mathbf{w}(M-i+1)|^2 = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2}. \end{aligned} \quad (4.20)$$

where  $\theta_q \in [-\pi/2, \pi/2]$ ,  $q = 1, \dots, Q$  is a continuum of directions that are properly chosen (uniform or nonuniform) to approximate the spatial domain  $[-\pi/2, \pi/2]$ . In our numerical results, we will consider a uniform grid over the interval  $[-\pi/2, \pi/2]$

---

<sup>2</sup>The case of odd  $M$  can be considered in a completely similar way.

excluding the transition regions. It is worth noting that the optimization problem (4.20) has significantly larger number of degrees of freedom than the beamforming problem for the phased-array case where the magnitudes of  $\mathbf{w}(i)$ ,  $i = 1, \dots, M$  are fixed.

The problem (4.20) is a non-convex QCQP which is NP-hard in general. However, a well developed SDP relaxation technique can be used to solve it [17]–[22]. Indeed, using the facts that  $|\mathbf{w}^H \mathbf{d}(\theta_q)|^2 = \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{w} \mathbf{w}^H\}$  and  $|\mathbf{w}(i)|^2 + |\mathbf{w}(M - i + 1)|^2 = \text{tr}\{\mathbf{w} \mathbf{w}^H \mathbf{A}_i\}$ ,  $i = 1, \dots, M/2$ , where  $\mathbf{A}_i$  is an  $M \times M$  matrix such that  $\mathbf{A}_i(i, i) = \mathbf{A}_i(M - (i - 1), M - (i - 1)) = 1$  and the rest of the elements are equal to zero, the problem (4.20) can be cast as

$$\begin{aligned} & \min_{\mathbf{w}, \delta} \delta \\ & \text{subject to} \quad \frac{G_d(\theta_q)}{2} - \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{w} \mathbf{w}^H\} \leq \delta, \quad q = 1, \dots, Q \\ & \quad \quad \quad \frac{G_d(\theta_q)}{2} - \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{w} \mathbf{w}^H\} \geq -\delta, \quad q = 1, \dots, Q \\ & \quad \quad \quad \text{tr}\{\mathbf{w} \mathbf{w}^H \mathbf{A}_i\} = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2}. \end{aligned} \quad (4.21)$$

Introducing the new variable  $\mathbf{X} \triangleq \mathbf{w} \mathbf{w}^H$  and by following the same steps as in Subsection 2.1.7, the problem (4.21) can be equivalently written as

$$\begin{aligned} & \min_{\mathbf{X}, \delta} \delta \\ & \text{subject to} \quad \frac{G_d(\theta_q)}{2} - \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{X}\} \leq \delta, \quad q = 1, \dots, Q \\ & \quad \quad \quad \frac{G_d(\theta_q)}{2} - \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{X}\} \geq -\delta, \quad q = 1, \dots, Q \\ & \quad \quad \quad \text{tr}\{\mathbf{X} \mathbf{A}_i\} = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2}; \quad \text{rank}\{\mathbf{X}\} = 1 \end{aligned} \quad (4.22)$$

where  $\mathbf{X}$  is a Hermitian matrix. Note that the last two constraints in (4.22) imply that the matrix  $\mathbf{X}$  is positive semi-definite. The problem (4.22) is non-convex with respect to  $\mathbf{X}$  because the last constraint is not convex. However, by means of the SDP relaxation technique, this constraint can be replaced by another constraint, that is,  $\mathbf{X} \succeq \mathbf{0}$ . The resulting problem is the relaxed version of (4.22) and it is a convex SDP problem which can be efficiently solved using, for example, interior point methods. As it was explained in Section 2.1, extraction of the solution of the original problem is typically done via the so-called randomization techniques.

In order to explain the randomization technique used in this work, let  $\mathbf{X}_{\text{opt}}$  denote the optimal solution of the relaxed problem. If the rank of  $\mathbf{X}_{\text{opt}}$  is one,

the optimal solution of the original problem (4.20) can be obtained by simply finding the principal eigenvector of  $\mathbf{X}_{\text{opt}}$ . However, if the rank of the matrix  $\mathbf{X}_{\text{opt}}$  is higher than one, the randomization approach can be used. Various randomization techniques have been developed and are generally based on generating a set of candidate vectors and then choosing the candidate which gives the minimum of the objective function of the original problem. Our randomization procedure can be described as follows. Let  $\mathbf{X}_{\text{opt}} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$  denote the eigendecomposition of  $\mathbf{X}_{\text{opt}}$ . The candidate vector  $k$  can be chosen as  $\mathbf{w}_{\text{can},k} = \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{v}_k$  where  $\mathbf{v}_k$  is a random vector whose elements are random variables uniformly distributed on the unit circle in the complex plane. Candidate vectors are not always feasible and should be mapped to a nearby feasible point. This mapping is problem dependent [14]. In our case, if the condition  $|\mathbf{w}_{\text{can},k}(i)|^2 + |\mathbf{w}_{\text{can},k}(M - i + 1)|^2 = P_t/M$  does not hold, we can map this vector to a nearby feasible point by scaling  $\mathbf{w}_{\text{can},k}(i)$  and  $\mathbf{w}_{\text{can},k}(M - i + 1)$  to satisfy this constraint. Among the candidate vectors we then choose the one which gives the minimum objective function, i.e., the one with minimum  $\max_{\theta_q} \left| G_d(\theta_q)/2 - |\mathbf{w}_{\text{can},k}^H \mathbf{d}(\theta_q)|^2 \right|$ .

### 4.3.2 Even number of transmit waveforms

Let us consider now the  $M \times K$  transmit beamspace matrix  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$  where  $K \leq M$  and  $K$  is an even number. For convenience, the virtual received signal vector matched to the basis waveform  $\phi_k(t)$  is rewritten as

$$\begin{aligned} \mathbf{y}_k(\varrho) &\triangleq \int_{T_p} \mathbf{x}(t, \varrho) \phi_k^*(t) dt \\ &\triangleq \sum_{l=1}^L \beta_l(\varrho) e^{j\psi_k(\theta_l)} |\mathbf{d}^H(\theta_l) \mathbf{w}_k| \mathbf{b}(\theta_l) + \mathbf{z}_k(\varrho). \end{aligned} \quad (4.23)$$

From (4.23), it can be seen that the RIP between  $\mathbf{y}_k(\varrho)$  and  $\mathbf{y}_{k'}(\varrho)$ ,  $k \neq k'$  holds if

$$|\mathbf{d}^H(\theta) \mathbf{w}_k| = |\mathbf{d}^H(\theta) \mathbf{w}_{k'}|, \quad \theta \in [-\pi/2, \pi/2]. \quad (4.24)$$

In the previous subsection, we saw that by considering the following specific structure  $[\mathbf{w} \ \tilde{\mathbf{w}}^*]$  for the transmit beamspace matrix with only two waveforms, the RIP is guaranteed at the receive antenna array. In this part, we obtain the RIP for more general case of more than two waveforms which provides more degrees of freedom for obtaining a better performance. For this goal, we first show that if for

some  $k'$  the following relation holds

$$\left| \sum_{i=1}^{k'} \mathbf{d}^H(\theta) \mathbf{w}_i \right| = \left| \sum_{i=k'+1}^K \mathbf{d}^H(\theta) \mathbf{w}_i \right|, \quad \forall \theta \in [-\pi/2, \pi/2] \quad (4.25)$$

then the two new set of vectors defined as the summation of the first  $k'$  data vectors  $\mathbf{y}_i(\varrho)$ ,  $i = 1, \dots, k'$  and the last  $K - k'$  data vectors  $\mathbf{y}_i(\varrho)$ ,  $i = k' + 1, \dots, K$  will satisfy the RIP. More specifically, by defining the following vectors

$$\begin{aligned} \mathbf{g}_1(\varrho) &\triangleq \sum_{i=1}^{k'} \mathbf{y}_i(\varrho) \\ &= \sum_{l=1}^L \beta_l(\varrho) \left( \sum_{i=1}^{k'} \mathbf{d}^H(\theta_l) \mathbf{w}_i \right) \mathbf{b}(\theta_l) + \sum_{i=1}^{k'} \mathbf{z}_i(\varrho) \end{aligned} \quad (4.26)$$

$$\begin{aligned} \mathbf{g}_2(\varrho) &\triangleq \sum_{i=k'+1}^K \mathbf{y}_i(\varrho) \\ &= \sum_{l=1}^L \beta_l(\varrho) \left( \sum_{i=k'+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i \right) \mathbf{b}(\theta_l) + \sum_{i=k'+1}^K \mathbf{z}_i(\varrho) \end{aligned} \quad (4.27)$$

the corresponding signal component of target  $l$  in the vector  $\mathbf{g}_1(\varrho)$  has the same magnitude as in the vector  $\mathbf{g}_2(\varrho)$  if the equation (4.25) holds. In this case, the only difference between the signal components of the target  $l$  in the vectors  $\mathbf{g}_1(\varrho)$  and  $\mathbf{g}_2(\varrho)$  is the phase which can be used for DOA estimation. Based on this fact, for ensuring the RIP between the vectors  $\mathbf{g}_1(\varrho)$  and  $\mathbf{g}_2(\varrho)$ , equation (4.25) needs to be satisfied for every angle  $\theta \in [-\pi/2, \pi/2]$ . By noting that the equation  $|\mathbf{d}^H(\theta) \mathbf{w}| = |\mathbf{d}^H(\theta) \tilde{\mathbf{w}}^*|$  holds for any arbitrary  $\theta$ , it can be shown that the equation (4.25) holds for any arbitrary  $\theta$  only if the following structure on the matrix  $\mathbf{W}$  is imposed:

- $K$  is an even number,
- $k'$  equals to  $K/2$ ,
- $\mathbf{w}_i = \tilde{\mathbf{w}}_{K/2+i}^*$ ,  $i = 1, \dots, K/2$ .

More specifically, if the transmit beamspace matrix has the following structure

$$\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{K/2}, \tilde{\mathbf{w}}_1^*, \dots, \tilde{\mathbf{w}}_{K/2}^*] \quad (4.28)$$

then the signal component of  $\mathbf{g}_1(\varrho)$  associated with the  $l$ th target is the same as the corresponding signal component of  $\mathbf{g}_2(\varrho)$  up to phase rotation of

$$\angle \sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i - \angle \sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i \quad (4.29)$$

which can be used as a look-up table for finding DOA of a target. By considering the aforementioned structure for the transmit beamspace matrix  $\mathbf{W}$ , it is guaranteed that the RIP is satisfied and other additional design requirements can be satisfied through the proper design of  $\mathbf{w}_1, \dots, \mathbf{w}_{K/2}$ .

Substituting (4.28) in (4.14), the optimization problem of transmit beamspace matrix design can be reformulated as

$$\begin{aligned} & \min_{\mathbf{w}_k} \max_{\theta_q} \left| G_d(\theta_q) - \sum_{k=1}^{K/2} \|[\mathbf{w}_k \tilde{\mathbf{w}}_k^*]^H \mathbf{d}(\theta_q)\|^2 \right| \\ & \text{subject to } \sum_{k=1}^{K/2} |\mathbf{w}_k(i)|^2 + |\tilde{\mathbf{w}}_k(i)|^2 = \frac{P_t}{M}, \quad i = 1, \dots, M. \end{aligned} \quad (4.30)$$

For the case when the number of transmit antennas is even<sup>3</sup> and using the facts that

$$\|[\mathbf{w}_k \tilde{\mathbf{w}}_k^*]^H \mathbf{d}(\theta_q)\|^2 = 2|\mathbf{w}_k^H \mathbf{d}(\theta_q)|^2 \quad (4.31)$$

$$|\mathbf{w}_k^H \mathbf{d}(\theta_q)|^2 = \text{tr}\{\mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{w}_k \mathbf{w}_k^H\} \quad (4.32)$$

$$|\mathbf{w}_k(i)|^2 + |\mathbf{w}_k(M - i + 1)|^2 = \text{tr}\{\mathbf{w}_k \mathbf{w}_k^H \mathbf{A}_i\}, \quad i = 1, \dots, M/2 \quad (4.33)$$

the problem (4.30) can be recast as

$$\begin{aligned} & \min_{\mathbf{w}_k} \max_{\theta_q} \left| G_d(\theta_q)/2 - \sum_{k=1}^{K/2} |\mathbf{d}^H(\theta_q) \mathbf{w}_k|^2 \right| \\ & \text{subject to } \sum_{k=1}^{K/2} \text{tr}\{\mathbf{w}_k \mathbf{w}_k^H \mathbf{A}_i\} = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2} \end{aligned} \quad (4.34)$$

where as it was introduced earlier,  $\mathbf{A}_i$  is an  $M \times M$  matrix such that  $\mathbf{A}_i(i, i) = \mathbf{A}_i(M - (i - 1), M - (i - 1)) = 1$  and the rest of the elements are equal to zero. Introducing the new variables  $\mathbf{X}_k \triangleq \mathbf{w}_k \mathbf{w}_k^H$ ,  $k = 1, \dots, K/2$  and following similar steps as in the case of two transmit waveforms, the problem above can be equivalently rewritten as

$$\begin{aligned} & \min_{\mathbf{X}_k} \max_{\theta_q} \left| G_d(\theta_q)/2 - \sum_{k=1}^{K/2} \text{tr}\left\{ \mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{X}_k \right\} \right| \\ & \text{subject to } \sum_{k=1}^{K/2} \text{tr}\{\mathbf{X}_k \mathbf{A}_i\} = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2} \\ & \quad \text{rank}\{\mathbf{X}_k\} = 1, \quad k = 1, \dots, K/2 \end{aligned} \quad (4.35)$$

---

<sup>3</sup>The case when the number of transmit antennas is odd can be carried out in a straightforward manner.

where  $\mathbf{X}_k, k = 1, \dots, K/2$  are Hermitian matrices. The problem (4.35) can be solved in a similar way as the problem (4.22). Specifically, the optimal solution of the problem (4.35) can be approximated using the SDP relaxation [14], [22], i.e., dropping the rank-one constraints and solving the resulting convex problem.

By relaxing the rank-one constraints, the optimization problem (4.35) can be approximated as

$$\begin{aligned} & \min_{\mathbf{X}_k} \max_{\theta_q} \left| G_d(\theta_q)/2 - \sum_{k=1}^{K/2} \text{tr} \left\{ \mathbf{d}(\theta_q) \mathbf{d}^H(\theta_q) \mathbf{X}_k \right\} \right| \\ & \text{subject to } \sum_{k=1}^{K/2} \text{tr} \{ \mathbf{X}_k \mathbf{A}_i \} = \frac{P_t}{M}, \quad i = 1, \dots, \frac{M}{2} \\ & \quad \mathbf{X}_k \succeq \mathbf{0}, \quad k = 1, \dots, K/2. \end{aligned} \quad (4.36)$$

The problem (4.36) is convex and, therefore, it can be solved efficiently using interior point methods. Once the matrices  $\mathbf{X}_k \succeq \mathbf{0}, k = 1, \dots, K/2$  are obtained, the corresponding weight vectors  $\mathbf{w}_k, k = 1, \dots, K/2$  can be obtained using randomization techniques. Specifically, we use the randomization method introduced in Subsection 4.3.1 over every  $\mathbf{X}_k, k = 1, \dots, K/2$  separately and then map the resulted rank-one solutions to the closest feasible points. Among the candidate solutions, the best one is then selected.

### 4.3.3 Optimal rotation of the transmit beamspace matrix

The solution of the optimization problem (4.34) is not unique and as it will be explained shortly in details, any spatial rotation of the optimal transmit beamspace matrix is also optimal. Among the set of the optimal solutions of the problem (4.34), the one with better energy preservation is favorable. As a result, after the approximate optimal solution of the problem (4.34) is obtained, we still need to find the optimal rotation which results in the best possible transmit beamspace matrix in terms of the energy preservation.

More specifically, since the DOA of the target at  $\theta_l$  is estimated based on the phase difference between the signal components of this target in the newly defined vectors, i.e.,  $\mathbf{g}_1(\varrho) = \sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i$  and  $\mathbf{g}_2(\varrho) = \sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i$ , to obtain the best performance,  $\mathbf{W}$  should be designed in a way that the magnitudes of the summations  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i$  and  $\sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i$  take their largest values.

Since the phase of the product term  $\mathbf{d}^H(\theta_l) \mathbf{w}_i$  in  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i$  (or equivalently in  $\sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i$ ) may be different for different waveforms, the terms in the

summation  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i$  (or equivalently in the summation  $\sum_{i=K/2+1}^K \mathbf{d}^H(\theta_l) \mathbf{w}_i$ ) may add incoherently and, therefore, it may result in a small magnitude which in turn degrades the DOA estimation performance. In order to avoid this problem, we use the property that any arbitrary rotation of the transmit beamspace matrix does not change the transmit beampattern. Specifically, if

$$\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{K/2}, \tilde{\mathbf{w}}_1^*, \dots, \tilde{\mathbf{w}}_{K/2}^*] \quad (4.37)$$

is a transmit beamspace matrix with the introduced structure, then the new beamspace matrix defined as

$$\mathbf{W}_{\text{rot}} = [\mathbf{w}_{\text{rot},1}, \dots, \mathbf{w}_{\text{rot},K/2}, \tilde{\mathbf{w}}_{\text{rot},1}^*, \dots, \tilde{\mathbf{w}}_{\text{rot},K/2}^*] \quad (4.38)$$

has the same beampattern and the same power distribution across the antenna elements. Here  $[\mathbf{w}_{\text{rot},1}, \dots, \mathbf{w}_{\text{rot},K/2}] = [\mathbf{w}_1, \dots, \mathbf{w}_{K/2}] \mathbf{U}_{K/2 \times K/2}$  and  $\mathbf{U}_{K/2 \times K/2}$  is a unitary matrix. Based on this property, after proper design of the beamspace matrix with a desired beampattern and the RIP, we can rotate the beams so that the magnitude of the summation  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta_l) \mathbf{w}_i$  is increased as much as possible.

Since the actual locations of the targets are not known a priori, we design a unitary rotation matrix so that the integration of the squared magnitude of the summation  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta) \mathbf{w}_i$  over the desired sector is maximized. As an illustrating example, we consider the case when  $K$  is 4. In this case,

$$[\mathbf{w}_{\text{rot},1}, \mathbf{w}_{\text{rot},2}] = [\mathbf{w}_1, \mathbf{w}_2] \mathbf{U}_{2 \times 2} \quad (4.39)$$

and the integration of the squared magnitude of the summation  $\sum_{i=1}^2 \mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},i}$  over the desired sectors can be expressed as

$$\begin{aligned} \int_{\Theta} |\mathbf{w}_{\text{rot},1}^H \mathbf{d}(\theta) + \mathbf{w}_{\text{rot},2}^H \mathbf{d}(\theta)|^2 d\theta &= \int_{\Theta} \left( \mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},1} \mathbf{w}_{\text{rot},1}^H \mathbf{d}(\theta) + \mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},2} \mathbf{w}_{\text{rot},2}^H \mathbf{d}(\theta) \right. \\ &\quad \left. + 2\text{Re}\{\mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},1} \mathbf{w}_{\text{rot},2}^H \mathbf{d}(\theta)\} \right) d\theta \\ &= \int_{\Theta} \left( \mathbf{d}^H(\theta) \mathbf{w}_1 \mathbf{w}_1^H \mathbf{d}(\theta) + \mathbf{d}^H(\theta) \mathbf{w}_2 \mathbf{w}_2^H \mathbf{d}(\theta) \right. \\ &\quad \left. + 2\text{Re}\{\mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},1} \mathbf{w}_{\text{rot},2}^H \mathbf{d}(\theta)\} \right) d\theta \end{aligned} \quad (4.40)$$

where  $\Theta$  denotes the desired sectors and  $\text{Re}\{\cdot\}$  stands for the real part of a complex number. The last line follows from the equation (4.39). Defining the new vector

$\mathbf{e} = [1, -1]^T$ , the expression in (4.40) can be equivalently recast as

$$\int_{\Theta} \left( \mathbf{d}^H(\theta) \mathbf{w}_1 \mathbf{w}_1^H \mathbf{d}(\theta) + \mathbf{d}^H(\theta) \mathbf{w}_2 \mathbf{w}_2^H \mathbf{d}(\theta) + 2\text{Re}\{\mathbf{d}^H(\theta) \mathbf{w}_{\text{rot},1} \mathbf{w}_{\text{rot},2}^H \mathbf{d}(\theta)\} \right) d\theta = \int_{\Theta} \left( 2\mathbf{d}^H(\theta) \mathbf{w}_1 \mathbf{w}_1^H \mathbf{d}(\theta) + 2\mathbf{d}^H(\theta) \mathbf{w}_2 \mathbf{w}_2^H \mathbf{d}(\theta) - |\mathbf{d}^H(\theta) \mathbf{W} \mathbf{U} \mathbf{e}|^2 \right) d\theta. \quad (4.41)$$

We aim at maximizing the expression (4.41) with respect to the unitary rotation matrix  $\mathbf{U}$ . Since the first two terms inside the integral in (4.41) are independent of the unitary matrix, it only suffices to minimize the integration of the last term. Using the property that  $\|\mathbf{X}\|^2 = \text{tr}\{\mathbf{X}\mathbf{X}^H\}$ , and the cyclical property of the trace, i.e.,  $\text{tr}\{\mathbf{X}\mathbf{X}^H\} = \text{tr}\{\mathbf{X}^H\mathbf{X}\}$ , the integral of the last term in (4.41) can be equivalently expressed as

$$\int_{\Theta} \text{tr} \{ \mathbf{U} \mathbf{e} \mathbf{e}^H \mathbf{U}^H \mathbf{W}^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{W} \} d\theta. \quad (4.42)$$

The only term in the integral (4.42) which depends on  $\theta$  is  $\mathbf{W}^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{W}$ . Therefore, the minimization of the integration of the last term in (4.41) over the sector  $\Theta$  can be stated as the following optimization problem

$$\begin{aligned} \min_{\mathbf{U}} \quad & \text{tr}\{\mathbf{U}\mathbf{E}\mathbf{U}^H\mathbf{D}\} \\ \text{subject to} \quad & \mathbf{U}\mathbf{U}^H = \mathbf{I} \end{aligned} \quad (4.43)$$

where  $\mathbf{E} \triangleq \mathbf{e}\mathbf{e}^H$  and  $\mathbf{D} \triangleq \int_{\Theta} \mathbf{W}^H \mathbf{d}(\theta) \mathbf{d}^H(\theta) \mathbf{W} d\theta$ . Because of the unitary constraint, the optimization problem (4.43) is an optimization problem over the Stiefel manifold [116]–[119]. Note that since the objective function in the optimization problem (4.43) depends not only on the subspace spanned by  $\mathbf{U}$ , but on the basis as well, the corresponding manifold is Steifel manifold, in contrast to a more common Grassmannian manifold. In order to address this problem, we have adopted the steepest decent algorithm on the Lie group of unitary matrices which has been developed in [118]. This algorithm moves towards the optimal point over the iterations.

#### 4.3.4 Spatial-division based design

It is worth noting that instead of designing all transmit beams jointly, an easy alternative for designing  $\mathbf{W}$  is to design different pairs of beamforming vectors  $\{\mathbf{w}_k, \tilde{\mathbf{w}}_k^*\}$ ,  $k = 1, \dots, K/2$  separately. Specifically, in order to avoid the incoherent summation of the terms in  $\sum_{i=1}^{K/2} \mathbf{d}^H(\theta_i) \mathbf{w}_i$  (or equivalently in  $\sum_{i=K/2+1}^K \mathbf{d}^H(\theta_i) \mathbf{w}_i$ ), the matrix  $\mathbf{W}$  can be designed in such a way that the corresponding transmit beampatterns of the beamforming vectors  $\mathbf{w}_1, \dots, \mathbf{w}_{K/2}$  do not overlap and they cover different

parts of the desired sector with equal energy. This alternative design is referred to as the SDD method. The design of one pair  $\{\mathbf{w}_k, \tilde{\mathbf{w}}_k^*\}$  has been already explained in Subsection 4.3.1.

## 4.4 Simulation results

Throughout our simulations, we assume a uniform linear transmit array with  $M = 10$  antennas spaced half a wavelength apart, and a non-uniform linear receive array of  $N = 10$  elements. The locations of the receive antennas are randomly drawn from the set  $[0, 9]$  measured in half a wavelength. Noise signals are assumed to be Gaussian, zero-mean, and white both temporally and spatially. In each example, targets are assumed to lie within a given spatial sector. From example to example the sector widths in which transmit energy is focussed is changed, and, as a result, so does the optimal number of waveforms to be used in the optimization of the transmit beamspace matrix. The optimal number of waveforms is calculated based on the number of dominant eigenvalues of the positive definite matrix  $\mathbf{A} = \int_{\Theta} \mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta$  (see [41] for explanations and corresponding Cramer-Rao bound derivations and analysis). We assume that the number of dominant eigenvalues is even; otherwise, we round it up to the nearest even number. The reason that an odd number of dominant eigenvalues is rounded up, as opposed to down, is that overusing waveforms is less detrimental to the performance of DOA estimation than underusing, as it is shown in [41]. Four examples are chosen to test the performance of our algorithm. In Example 1, a single centrally located sector of width  $20^\circ$  is chosen to verify the importance of the uniform power distribution across the orthogonal waveforms. In Example 2, two separated sectors each with a width of  $20^\circ$  degrees are chosen. In Example 3, a single, centrally located sector of width  $10^\circ$  degrees is chosen. Finally, in Example 4, a single, centrally located sector of width  $30^\circ$  degrees is chosen. The optimal number of waveforms used for each example is two, four, two, and four, respectively. The methods tested by the examples are traditional MIMO radar with uniform transmit power density and  $K = M$  and the proposed jointly optimum transmit beamspace design method. In Example 3, we also consider the SSD method which is an easier alternative to the jointly optimal method. Throughout the simulations, we refer to the proposed transmit beamspace method as the optimal transmit beamspace design (although the solution obtained through SDP relaxation and randomization is suboptimal in general) to distinguish it from the

SDD method in which different pairs of the transmit beamspace matrix columns are designed separately. In Examples 1 and 3, the SDD is not considered since there is no need for more than two waveforms. We also do not apply the SDD method in the last example due to the fact that the corresponding spatially divided sectors in this case are adjacent and their sidelobes result in energy loss and performance degradation as opposed to Example 2. For the traditional MIMO radar, the following set of orthogonal baseband waveforms is used

$$\phi_m(t) = \sqrt{\frac{1}{T_p}} e^{j2\pi \frac{m}{T_p} t}, \quad m = 1, \dots, M \quad (4.44)$$

while for the proposed transmit beamspace-based method, the first  $K$  waveforms of (4.44) are employed.

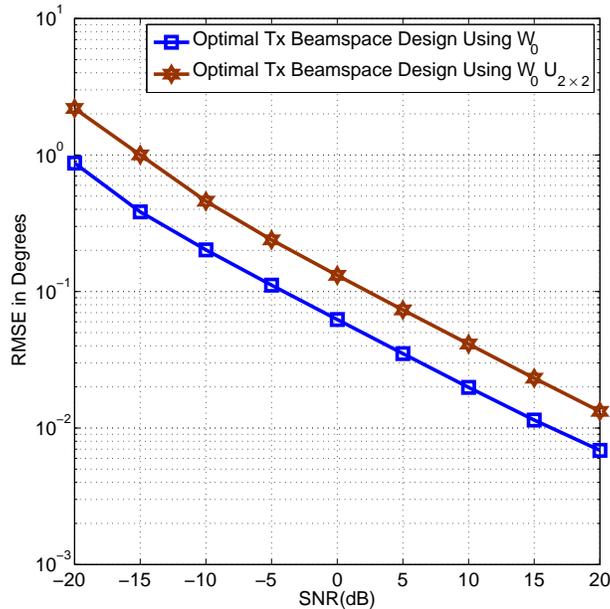
Throughout all simulations, the total transmit power remains constant at  $P_t = M$ . The root mean square error (RMSE) and probability of target resolution are calculated based on 500 independent Monte-Carlo runs.

#### 4.4.1 Example 1 : Effect of the uniform power distribution across the waveforms

In this example, we aim at studying how the lack of uniform transmission power across the transmit waveforms affects the performance of the new proposed method. For this goal, we consider two targets that are located in the directions  $-5^\circ$  and  $5^\circ$  and the desired sector is chosen as  $\theta = [-10^\circ, 10^\circ]$ . Two orthogonal waveforms are considered and optimal transmit beamspace matrix denoted as  $\mathbf{W}_0$  is obtained by solving the optimization problem (4.19). To simulate the case of non-uniform power distribution across the waveforms while preserving the same transmit beampattern of  $\mathbf{W}_0$ , we use the rotated transmit beamspace matrix  $\mathbf{W}_0 \mathbf{U}_{2 \times 2}$ , where  $\mathbf{U}_{2 \times 2}$  is a unitary matrix defined as

$$\mathbf{U}_{2 \times 2} = \begin{bmatrix} 0.6925 + j0.3994 & 0.4903 + j0.3468 \\ -0.4755 + j0.3669 & 0.6753 - j0.4279 \end{bmatrix}.$$

Note that  $\mathbf{W}_0$  and  $\mathbf{W}_0 \mathbf{U}_{2 \times 2}$  lead to the same transmit beampattern and as a result the same transmit power within the desired sector, however, compared to the former, the latter one does not have uniform transmit power across the waveforms and therefore it does not satisfy RIP. The RMSE curves of the proposed DOA estimation method for both  $\mathbf{W}_0$  and  $\mathbf{W}_0 \mathbf{U}_{2 \times 2}$  versus SNR are shown in Fig. 4.1. It can be seen from this figure that the lack of uniform transmission power across the waveforms can degrade the performance of DOA estimation significantly.



**Figure 4.1:** Example 1: Performance of the new proposed method with and without uniform power distribution across transmit waveforms.

#### 4.4.2 Example 2 : Two separated sectors of width $20^\circ$ degrees each

In the second example, two targets are assumed to lie within two spatial sectors: one from  $\theta = [-40^\circ, -20^\circ]$  and the other from  $\theta = [30^\circ, 50^\circ]$ . The targets are located at  $\theta_1 = -33^\circ$  and  $\theta_2 = 41^\circ$ . Fig. 4.2 shows the transmit beampatterns of the traditional MIMO with uniform transmit power distribution and both the optimal and SDD designs for  $\mathbf{W}$ . It can be seen in the figure that the optimal transmit beamspace method provides the most even concentration of power in the desired sectors. The SDD technique provides concentration of power in the desired sectors above and beyond traditional MIMO; however, the energy is not evenly distributed with one sector having a peak beampattern strength of 15 dB, while the other has a peak of no more than 12 dB. Fig. 4.3 shows the individual beampatterns associated with individual waveforms as well as the overall beampattern.

The performance of all three methods is compared in terms of the corresponding RMSEs versus SNR as shown in Fig. 4.4. As we can see in the figure, the jointly optimal transmit beamspace and the SDD methods have lower RMSEs as compared to the RMSE of the traditional MIMO radar. It is also observed from the figure that the performance of the SDD method is very close to the performance of the

jointly optimal one.

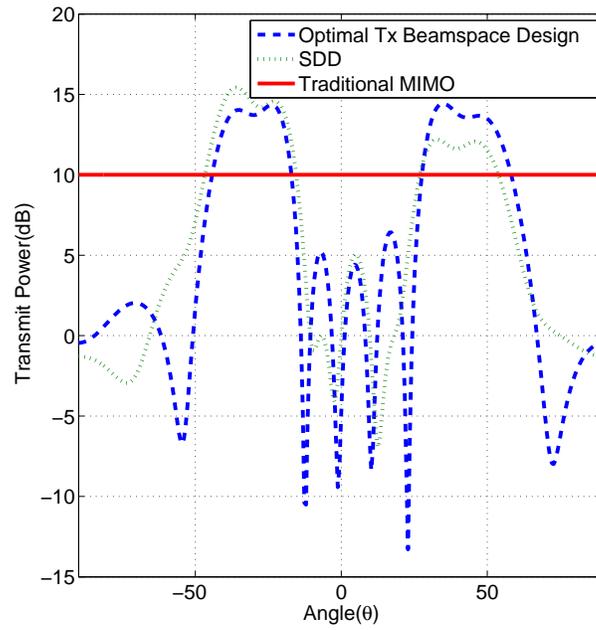
To assess the proposed method's ability to resolve closely located targets, we move both targets to the locations  $\theta_1 = 38^\circ$  and  $\theta_2 = 40^\circ$ . The performance of all three methods tested is given in terms of the probability of target resolution. Note that the targets are considered to be resolved if there are at least two peaks in the MUSIC spectrum and the following is satisfied [90]

$$\left| \hat{\theta}_l - \theta_l \right| \leq \frac{\Delta\theta}{2}, \quad l = 1, 2$$

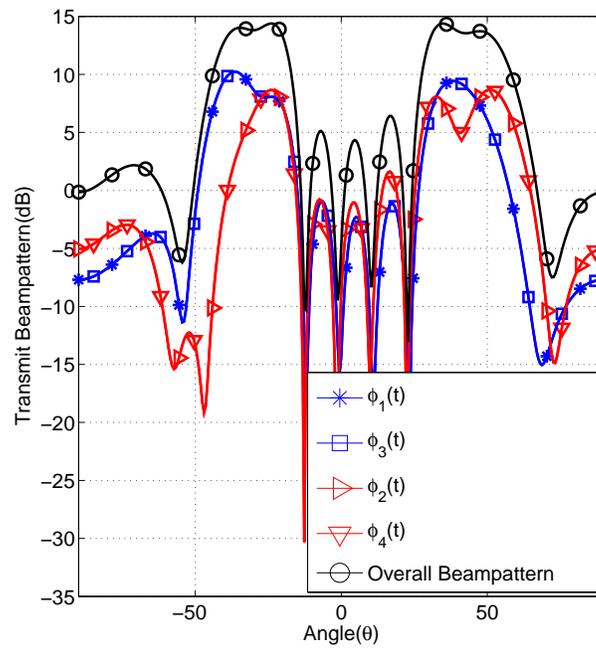
where  $\Delta\theta = |\theta_2 - \theta_1|$ . The probability of source resolution versus SNR for all methods tested are shown in Fig. 4.5. It can be seen from the figure that the SNR threshold at which the probability of target resolution transitions from very low values (i.e., resolution fail) to values close to one (i.e., resolution success) is lowest for the jointly optimal transmit beamspace design-based method, second lowest for the SDD method, and finally, highest for the traditional MIMO radar method. In other words, the figure shows that the jointly optimal transmit beamspace design-based method has a higher probability of target resolution at lower values of SNR than the SDD method, while the traditional MIMO radar method has the worst resolution performance.

#### 4.4.3 Example 3 : Single and centrally located sector of width $10^\circ$ degrees

In the third example, the targets are assumed to lie within a single thin sector of  $\theta = [-10^\circ, 0^\circ]$ . Due to the choice of the width of the sector, the optimal number of waveforms to use is only two. For this reason, only two methods are tested: the proposed transmit beamspace method and the traditional MIMO radar. The beampatterns for these two methods are shown in Fig. 4.6. It can be observed from the figure that our method offers a transmit power gain that is 5 dB higher than the traditional MIMO radar. In order to test the RMSE performance of both methods, targets are assumed to be located at  $\theta_1 = -7^\circ$  and  $\theta_2 = -2^\circ$ . The RMSE's are plotted versus SNR in Fig. 4.7. It can be observed from this figure that the proposed transmit beamspace method yields lower RMSE as compared to the traditional MIMO radar based method at moderate and high SNR values. In order to test the resolution capabilities of both methods tested, the targets are moved to  $\theta_1 = -3^\circ$  and  $\theta_2 = -1^\circ$ . The same criterion as in Example 2 is then used to determine the



**Figure 4.2:** Example 2: Transmit beampatterns of the traditional MIMO and the proposed transmit beamspace design-based methods.



**Figure 4.3:** Example 2: Individual beampatterns associated with individual waveforms and the overall beampattern.

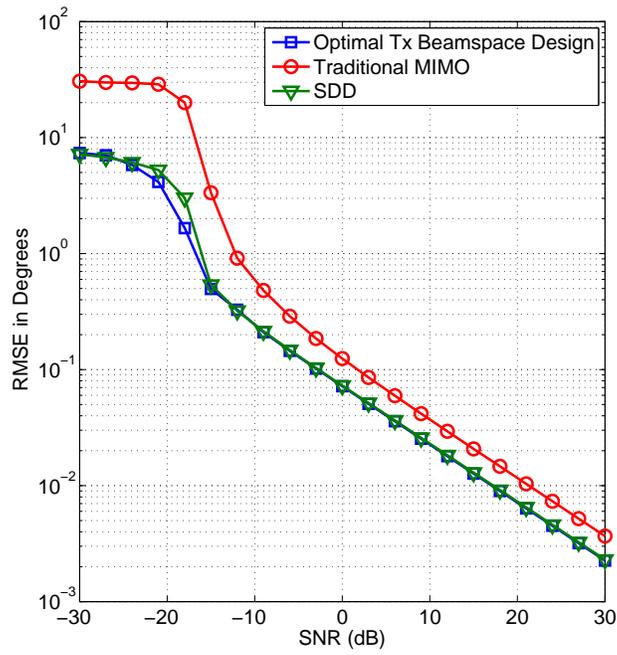


Figure 4.4: Example 2: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods.

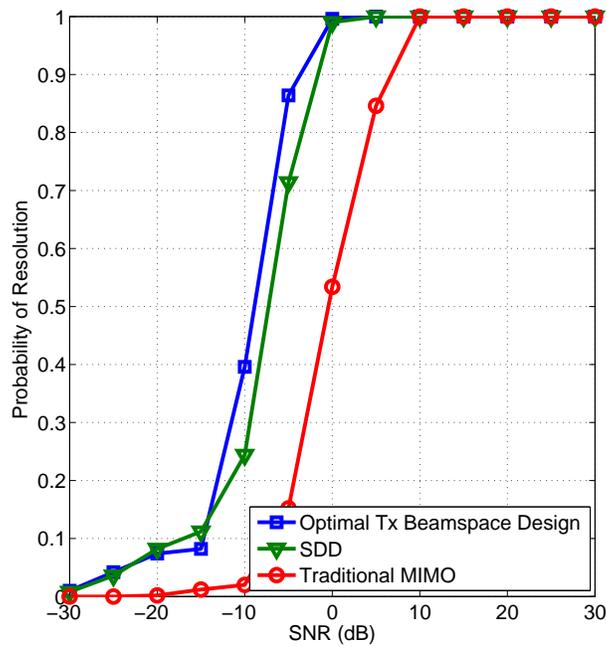
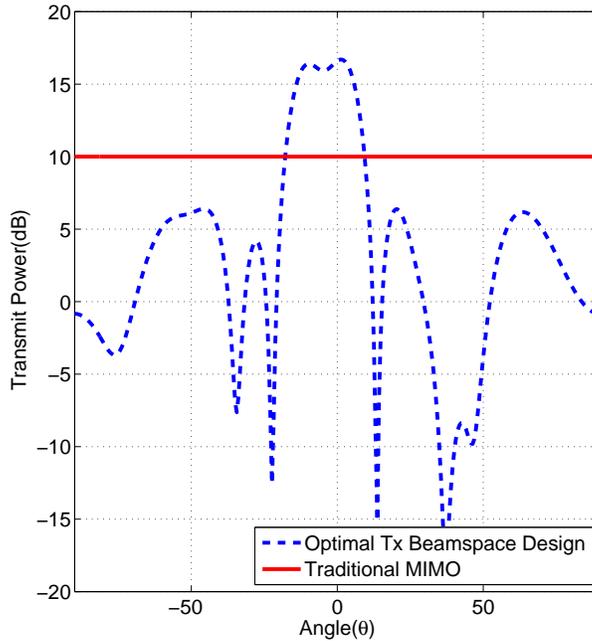


Figure 4.5: Example 2: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods.



**Figure 4.6:** Example 3: Transmit beampatterns of the traditional MIMO and the proposed transmit beamspace design-based method.

target resolution. The results of this test are displayed in Fig. 4.8 and agrees with the similar results in Example 2.

#### 4.4.4 Example 4 : Single and centrally located sector of width $30^\circ$ degrees

In the last example, a single wide sector is chosen as  $\theta = [-15^\circ, 15^\circ]$ . The optimal number of waveforms for such a sector is found to be four. Similar to the previous Example 3, we compare the performance of the proposed method to that of the traditional MIMO radar. Four transmit beams are used to simulate the optimal transmit beamspace design-based method. Fig. 4.9 shows the transmit beampatterns for the methods tested. In order to test the RMSE performance of the methods tested, two targets are assumed to be located at  $\theta_1 = -12^\circ$  and  $\theta_2 = 9^\circ$ . Fig. 4.10 shows the RMSEs versus SNR for the methods tested. As we can see in the figure, the RMSE for the jointly optimal transmit beamspace design-based method is lower than the RMSE for the traditional MIMO radar based method. Moreover, in order to test resolution, the targets are moved to  $\theta_1 = -3^\circ$  and  $\theta_2 = -1^\circ$ . The same criterion as in Example 2 is used to determine the target resolution. The results of

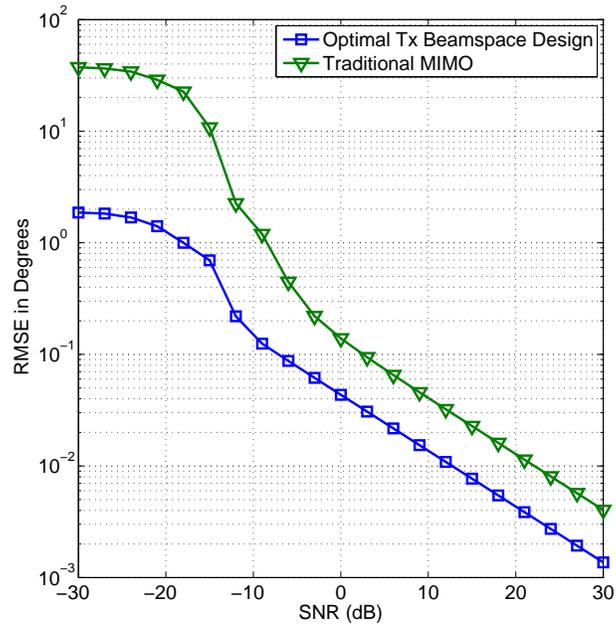


Figure 4.7: Example 3: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based method.

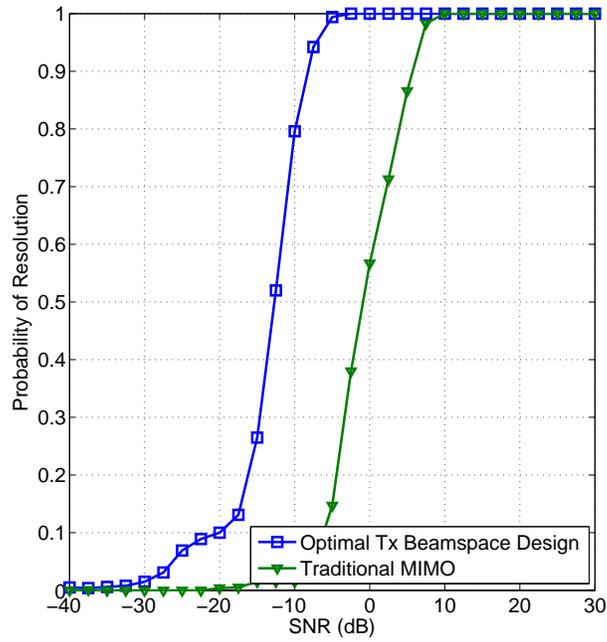
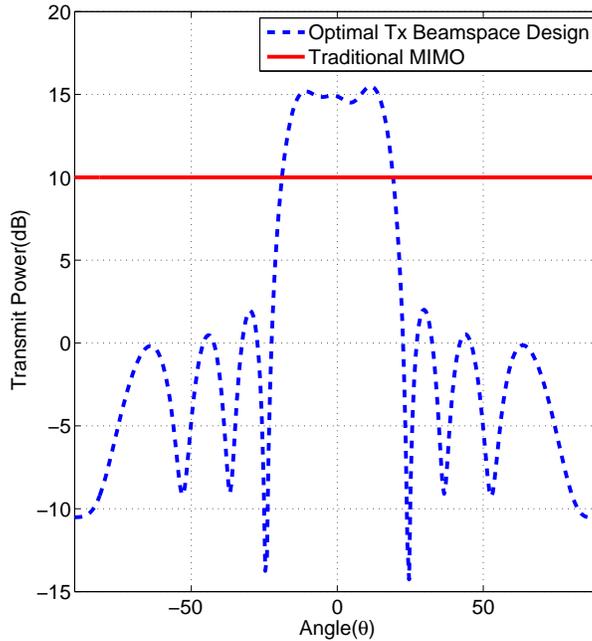


Figure 4.8: Example 3: Performance comparison between the traditional MIMO and the proposed transmit beamspace design-based methods.

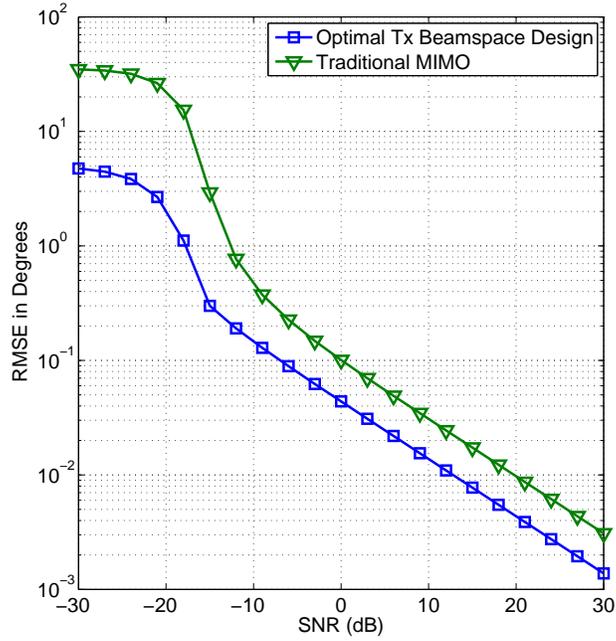


**Figure 4.9:** Example 4: Transmit beampatterns of the traditional MIMO and the proposed methods.

this test are similar to those displayed in Fig. 4.5, and, therefore, are not displayed here.

## 4.5 Chapter summary

The problem of transmit beamspace design for MIMO radar with colocated antennas with application to DOA estimation has been considered. A new method for designing the transmit beamspace matrix that enables the use of search-free DOA estimation techniques at the receiver has been introduced. The essence of the proposed method is to design the transmit beamspace matrix based on minimizing the difference between a desired transmit beampattern and the actual one. The case of even but otherwise arbitrary number of transmit waveforms has been considered. The transmit beams are designed in pairs where all pairs are designed jointly while satisfying the requirements that the two transmit beams associated with each pair enjoy rotational invariance with respect to each other. Unlike previous methods that achieve phase rotation between two transmit beams while allowing the magnitude to be different, a specific beamspace matrix structure achieves phase rotation while ensuring that the magnitude response of the two transmit beams is exactly the same



**Figure 4.10:** Example 4: Performance comparison between the traditional MIMO and the proposed methods.

at all spatial directions has been proposed. The SDP relaxation technique has been used to relax the proposed formulation into a convex optimization problem that can be solved efficiently using interior point methods. An alternative SDD method that divides the spatial domain into several subsectors and assigns a subset of the transmit beamspace pairs to each subsector has been also developed. The SDD method enables post processing of data associated with different subsectors independently with DOA estimation performance comparable to the performance of the joint transmit beamspace design-based method. Simulation results have been used to demonstrate the improvement in the DOA estimation performance offered by using the proposed joint and SDD transmit beamspace design methods as compared to the traditional MIMO radar.

## Chapter 5

# A New Robust Adaptive Beamforming

RAB has been an intensive research topic over several decades due to, on one hand, its importance in wireless communications, radar, sonar, microphone array speech processing, radio astronomy, medical imaging, and other fields; and on the other hand, because of the challenges related to the practical applications manifesting themselves in the robustness requirements. As it was explained in Section 2.3, the presence of the desired signal component in the training data, small sample size, and imprecise knowledge of the desired signal steering vector are the main causes of performance degradation in adaptive beamforming. The main goal of any RAB method is to provide robustness against such imperfections.

The most popular RAB methods that are based on the MVDR principle including (i) the worst-case-based adaptive beamforming techniques [7], [49], [50]–[51]; (ii) the doubly constrained robust Capon beamforming methods [53], [54]; (iii) the probabilistically constrained RAB technique [52]; (iv) the RAB technique based on steering vector estimation [55] have been clearly explained in Section 2.3.

Although the relationships between these MVDR RAB techniques have been established in the literature, the general notion of robustness and a unified principle for all MVDR RAB techniques have been missing. For example, the worst-case-based design is related to the diagonal loading principle [7], [50], while the probabilistically constrained design can be approximated by the worst-cases-based design [52]. Moreover, the worst-cases-based and doubly-constrained RAB techniques can be derived via steering vector estimation approach [50], [53].

In this chapter, we rethink the notion of robustness and present a unified principle to MVDR RAB design, that is, to use standard MVDR beamformer in tandem

with steering vector estimation based on some prior information and data covariance matrix estimation. This unified principle motivates us to develop a new technique which uses as little as possible, imprecise, and easy to obtain prior information about the desired signal/source, the antenna array, and the propagation media. We develop such a new RAB technique in which the steering vector is estimated through the beamformer output power maximization under the requirement that the estimate does not converge to any of the interference steering vectors and their linear combinations. The only prior information used is the imprecise knowledge of the angular sector of the desired signal and antenna array geometry, while the knowledge of the presumed steering vector is not needed. Such MVDR RAB technique can be mathematically formulated as a non-convex (due to an additional steering vector normalization condition) QCQP problem.

Moreover, our specific optimization problem allows for an exact solution using, for example, the duality theory [16], [54] or the iterative rank reduction technique [106]. In the optimization context, we develop some new results when answering the questions of (i) how to obtain a rank-one solution from a general-rank solution of the relaxed problem algebraically and (ii) when it is guaranteed that the solution of the relaxed problem is rank-one. The latter question, for example, is important because it had been observed that the probability of obtaining a rank-one solution for the class of problems similar to the one considered in this chapter is close to 1, while the theoretical upper-bound suggests a significantly smaller probability [14], [130]. Our result proves the correctness of the experimental observations about the high probability of a rank-one solution for the relaxed problem.

This chapter is organized as follows. A general notion of robustness and a unified principle for MVDR RAB design are given in Section 5.1 and the existing MVDR RAB techniques are summarized and shown to satisfy the general principle under their specific notions of robustness. In Section 5.2, we formulate a new MVDR RAB technique which uses as little as possible, imprecise, and easy to obtain prior information. An analysis of the problem as well as some new optimization related results are given in Section 5.3. Simulation results comparing the performance of the proposed method to the existing methods are shown in Section 5.4. Finally, Section 5.5 presents our conclusions.

## 5.1 The unified principle to MVDR RAB design

The MVDR-SMI beamformer is known to be not robust to any imperfect knowledge of the desired signal steering vector. Different RAB techniques have been developed which use different specific notions of robustness. However, the general meaning of robustness for any RAB technique is the ability to compute the beamforming vector so that the SINR is maximized despite possibly imperfect and little knowledge of prior information. Specifically, the main cause of performance degradation of MVDR-SMI beamformer is the situation when the desired signal steering vector is mixed with any of the interference steering vectors. Thus, if with little and imperfect prior information, an adaptive beamforming technique is able to estimate the desired signal steering vector so that it does not converge to any of the interferences and their linear combinations, such technique is called robust. Using this notion of robustness, the unified principle to MVDR RAB design can be formulated as follows. Use the standard MVDR-SMI beamformer (2.39) in tandem with steering vector estimation based on some prior information. Difference between various MVDR RAB techniques can be then shown to boil down to the differences in the assumed prior information, the specific notions of robustness used, and the corresponding steering vector estimation techniques used.

In this chapter, we consider the system model that was introduced in Subsection 2.3.1. As it was mentioned in Subsection 2.3.2, in the case of steering vector mismatch  $\boldsymbol{\delta}$ , the beamformer output power can be also written as a function of  $\boldsymbol{\delta}$  as

$$P(\boldsymbol{\delta}) = \frac{1}{(\mathbf{p} + \boldsymbol{\delta})^H \hat{\mathbf{R}}^{-1} (\mathbf{p} + \boldsymbol{\delta})}. \quad (5.1)$$

where  $\mathbf{a} = \mathbf{p} + \boldsymbol{\delta}$  is the actual steering vector and  $\mathbf{p}$  is the presumed one. Then the best estimate of  $\boldsymbol{\delta}$ , denoted as  $\hat{\boldsymbol{\delta}}$  (the estimate of  $\mathbf{a}$  is  $\hat{\mathbf{a}} = \mathbf{p} + \hat{\boldsymbol{\delta}}$ ), is the one which maximizes (5.1) under the constraint that  $\hat{\mathbf{a}}$  does not converge to any interference steering vectors and their linear combinations.

The well known MVDR RAB techniques are summarized in Table 5.1 where the corresponding notions of robustness and prior information used by the techniques are listed. In this table, the multi-rank beamformer matrix of the eigenvalue beamforming method of [97] is computed as  $\mathbf{W} = \hat{\mathbf{R}}^{-1} \boldsymbol{\Psi} (\boldsymbol{\Psi}^H \hat{\mathbf{R}}^{-1} \boldsymbol{\Psi})^{-1} \mathbf{Q}$ , where  $\mathbf{Q}$  is a data dependent left-orthogonal matrix, i.e.,  $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$ ,  $\boldsymbol{\Psi}$  is the linear subspace in which the desired signal is located. For resolving a signal with a rank-one covariance matrix and an unknown but fixed angle of arrival, the columns of  $\mathbf{Q}$

should be selected as the dominant eigenvectors of the error covariance matrix, i.e.,  $\mathbf{R}_e = (\Psi^H \mathbf{R}^{-1} \Psi)^{-1}$ . If it is assumed that the signal lies in a known subspace, but the angle of arrival is unknown and unfixed (for example, it randomly changes from snapshot to snapshot), it is the subdominant eigenvectors of the error covariance matrix that should be used as the columns of the matrix  $\mathbf{Q}$ .

It can be seen from Table 5.1 that the main problem for any MVDR RAB technique is to estimate the steering vector, while avoiding its convergence to any of the interferences and their linear combinations. It is achieved in different techniques by exploiting different prior information and solving different optimization problems. The complexity of the corresponding steering vector estimation problems can vary from the complexity of eigenvalue decomposition to the complexity of solving QCQP programming problem. All known MVDR RAB techniques require the knowledge of the presumed steering vector that, in turn, implies that the antenna array geometry, propagation media, and desired source characteristics such as the presumed angle of arrival are known. Therefore, it is of great importance to develop an MVDR RAB technique that requires as little as possible and easy to obtain prior information.

## 5.2 New beamforming problem formulation

For estimating the actual steering vector  $\mathbf{a}$ , we first observe that the maximization of the output power (5.1) is equivalent to the minimization of the denominator of (5.1). One obvious constraint that must be imposed on the estimate  $\hat{\mathbf{a}}$  is the norm constraint  $\|\hat{\mathbf{a}}\|^2 = M$ . To avoid the convergence of the estimate  $\hat{\mathbf{a}}$  to any of the interference steering vectors and their linear combinations, we introduce a new constraint in what follows.

In order to establish such a new constraint, we assume that the desired source is located in the known angular sector of  $\Theta = [\theta_{\min}, \theta_{\max}]$  which can be obtained, for example, using low resolution direction finding methods. This angular sector is assumed to be distinguishable from general locations of the interfering signals. Let  $\tilde{\mathbf{C}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$  denote the eigenvalue decomposition of the matrix  $\tilde{\mathbf{C}} \triangleq \int_{\tilde{\Theta}} \mathbf{d}(\tilde{\theta}) \mathbf{d}^H(\theta) d\theta$  where  $\mathbf{d}(\theta)$  is the steering vector associated with direction  $\theta$  and having the structure defined by the antenna geometry, the sector  $\tilde{\Theta}$  is the complement of the sector  $\Theta$ , and  $\mathbf{U}$  and  $\mathbf{\Lambda}$  are unitary and diagonal matrices, respectively. Column  $i$  of the unitary matrix  $\mathbf{U}$  denoted as  $\mathbf{u}_i$  and the  $i$ th diagonal element of the diagonal matrix  $\mathbf{\Lambda}$  denoted as  $\lambda_i$  are, respectively, the  $i$ th eigenvector and the  $i$ th eigenvalue of

**Table 5.1:** Different robust adaptive beamforming methods.

RAB	Notion of robustness	Assumptions/ Principles	Method	Prior Information	Disadvantages
Eigenspace-based beamformer [94]	Projection of the presumed steering vector to the signal-plus-interference subspace	Signal-plus-interference subspace denoted as $\mathbf{E}$ is obtained through eigendecomposition of $\hat{\mathbf{R}}$	$\hat{\mathbf{a}} = \mathbf{E}\mathbf{E}^H \mathbf{p}$ $\mathbf{w}_{\text{eig}} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{a}}$	The presumed steering vector $\mathbf{p}$ and the number of interfering signals	High probability of subspace swap and incorrect estimation of the signal-plus-interference subspace dimension at low SNR [95]
Worst-case-based and doubly constrained RAB techniques [7], [50], [53]	Presumed steering vector and its uncertainty region	The actual steering vector $\mathbf{a}$ is modeled as $\mathbf{a} = \mathbf{p} + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is an unknown deterministic norm-bounded ( $\ \boldsymbol{\delta}\  \leq \varepsilon$ , $\varepsilon$ is some bound) mismatch vector	$\max_{\sigma^2, \hat{\mathbf{a}}} \sigma^2$ subject to $\hat{\mathbf{R}} - \sigma^2 \hat{\mathbf{a}} \hat{\mathbf{a}}^H \geq 0,$ $\ \hat{\mathbf{a}} - \mathbf{p}\  \leq \varepsilon$ and also $\ \hat{\mathbf{a}}\ ^2 = M$ for doubly-constrained method	The presumed steering vector $\mathbf{p}$ and the uncertainty bounding value $\varepsilon$	Difficult to obtain $\varepsilon$ in practice
Probabilistically constrained RAB [52]	The non-outage probability	The steering vector $\mathbf{a}$ is modeled as $\mathbf{a} = \mathbf{p} + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a random mismatch vector with known or worst-case distribution	$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w}$ subject to $\Pr\{ \mathbf{w}^H \mathbf{a}  \geq 1\} \geq p_0$	The presumed steering vector $\mathbf{p}$ , the preselected non-outage probability value $p_0$ , and possibly distribution of $\boldsymbol{\delta}$	The value $p_0$ and the mismatch distributions may not be known
Eigenvalue RAB of [97]	General as in [55]	Steering vector lies in a known signal subspace and the rank of the signal correlation matrix is known	$\min_{\mathbf{w}} \text{tr}\{\mathbf{W}^H \hat{\mathbf{R}} \mathbf{W}\}$ subject to $\mathbf{W}^H \boldsymbol{\Psi} = \mathbf{Q}^H$	The linear subspace in which the desired signal is located and the rank of the desired signal covariance matrix	Very specific modeling of the covariance matrix. The signal subspace has to be known

$\tilde{\mathbf{C}}$ . It is assumed without loss of generality that the eigenvalues  $\lambda_i$ ,  $i = 1, \dots, M$  are ordered in the descending order, i.e.,  $\lambda_i \geq \lambda_{i+1}$ ,  $i = 1, \dots, M - 1$ . By splitting matrix  $\mathbf{U}$  to the  $M \times K$  matrix  $\mathbf{U}_1$  and the  $M \times (M - K)$  matrix  $\mathbf{U}_2$  as  $\mathbf{U} = [\mathbf{U}_1 \ \mathbf{U}_2]$ , the matrix  $\tilde{\mathbf{C}}$  can be decomposed as  $\tilde{\mathbf{C}} = \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{U}_1^H + \mathbf{U}_2 \mathbf{\Lambda}_2 \mathbf{U}_2^H$ , where the  $K \times K$  diagonal matrix  $\mathbf{\Lambda}_1$  contains the  $K$  dominant eigenvalues, while the other  $(M - K) \times (M - K)$  diagonal matrix  $\mathbf{\Lambda}_2$  contains the  $M - K$  subdominant eigenvalues. Since the matrix  $\tilde{\mathbf{C}}$  is computed by the integration over the complement of the desired sector, it can be concluded that for the properly chosen  $K$ , the steering vector in the desired sector and its complement can be approximately expressed as linear combinations of the columns of  $\mathbf{U}_2$  and  $\mathbf{U}_1$ , respectively, that is,

$$\mathbf{d}(\theta) \cong \mathbf{U}_2 \mathbf{v}_2, \quad \theta \in \Theta \quad (5.2)$$

$$\mathbf{d}(\theta) \cong \mathbf{U}_1 \mathbf{v}_1, \quad \theta \in \tilde{\Theta} \quad (5.3)$$

where  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are some coefficient vectors.

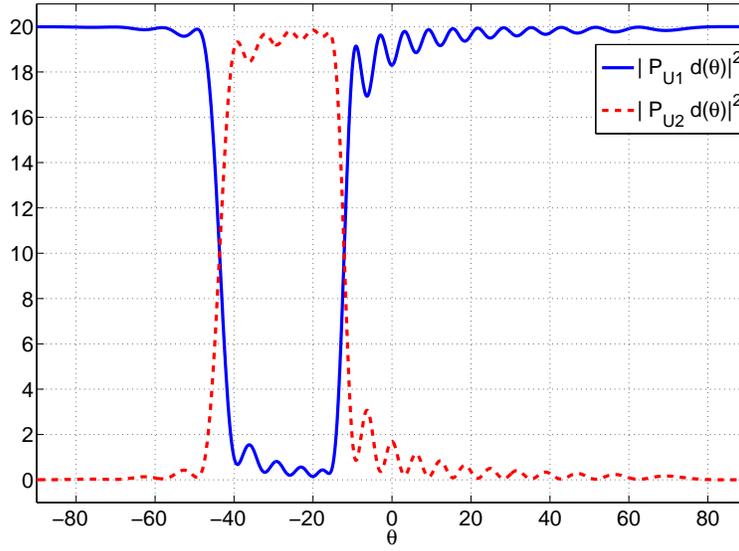
*Example 1:* As an illustrative example, let us consider the squared norm of the vectors that are obtained by projecting the vector  $\mathbf{d}(\theta)$  onto the subspaces spanned by  $\mathbf{U}_1$  and  $\mathbf{U}_2$ . Fig. 5.1 depicts such squared norms, i.e.,  $\|\mathbf{P}_{\mathbf{U}_1} \mathbf{d}(\theta)\|^2 = \|\mathbf{U}_1 \mathbf{U}_1^H \mathbf{d}(\theta)\|^2$  and  $\|\mathbf{P}_{\mathbf{U}_2} \mathbf{d}(\theta)\|^2 = \|\mathbf{U}_2 \mathbf{U}_2^H \mathbf{d}(\theta)\|^2$ , versus  $\theta$ . The example set up is the following. The angular sector is  $\Theta = [-35^\circ, -15^\circ]$ ,  $K = 5$ , and  $M = 20$ . It can be observed from the figure that the approximations (5.2) and (5.3) are accurate and the squared norms of the projections of the vector  $\mathbf{d}(\theta)$  on  $\mathbf{U}_2$  and  $\mathbf{U}_1$ , i.e., inside and outside of the desired sector, respectively, are almost equal to  $M$ .

Using (5.2) and (5.3), we know that  $\|\mathbf{v}_1\|^2 = M$  and  $\|\mathbf{v}_2\|^2 = M$ , and the following equations are in order

$$\begin{aligned} \mathbf{d}^H(\theta) \tilde{\mathbf{C}} \mathbf{d}(\theta) &\cong (\mathbf{U}_1 \mathbf{v}_1)^H \tilde{\mathbf{C}} (\mathbf{U}_1 \mathbf{v}_1) \\ &= \mathbf{v}_1^H \mathbf{\Lambda}_1 \mathbf{v}_1, \quad \theta \in \tilde{\Theta} \end{aligned} \quad (5.4)$$

$$\begin{aligned} \mathbf{d}^H(\theta) \tilde{\mathbf{C}} \mathbf{d}(\theta) &\cong (\mathbf{U}_2 \mathbf{v}_2)^H \tilde{\mathbf{C}} (\mathbf{U}_2 \mathbf{v}_2) \\ &= \mathbf{v}_2^H \mathbf{\Lambda}_2 \mathbf{v}_2, \quad \theta \in \Theta. \end{aligned} \quad (5.5)$$

Note that since  $\mathbf{\Lambda}_1$  is the diagonal matrix of  $K$  dominant eigenvalues of  $\tilde{\mathbf{C}}$  and  $\mathbf{\Lambda}_2$  contains the remaining eigenvalues, the quadratic form  $\mathbf{d}^H(\theta) \tilde{\mathbf{C}} \mathbf{d}(\theta)$  takes larger values outside of the desired sector. Based on this observation, the estimate  $\hat{\mathbf{a}}$  can be forced not to converge to any vector with direction located within the complement



**Figure 5.1:** Squared norm of the projection of vector  $\mathbf{d}(\theta)$  on the linear subspaces spanned by the columns of  $\mathbf{U}_1$  and  $\mathbf{U}_2$  versus  $\theta$ .

of  $\Theta$  including the interference steering vectors and their linear combinations by means of the following new constraint

$$\hat{\mathbf{a}}^H \tilde{\mathbf{C}} \hat{\mathbf{a}} \leq \Delta_0 \quad (5.6)$$

where  $\Delta_0$  is a uniquely selected value for a given angular sector  $\Theta$ , that is,

$$\Delta_0 \triangleq \max_{\theta \in \Theta} \mathbf{d}^H(\theta) \tilde{\mathbf{C}} \mathbf{d}(\theta). \quad (5.7)$$

Using the definition of  $\Delta_0$  (5.7) together with (5.5), we can find that

$$\Delta_0 = \max_{\theta \in \Theta} \mathbf{v}^H(\theta) \mathbf{\Lambda}_2 \mathbf{v}(\theta) \leq M \lambda_{K+1}. \quad (5.8)$$

It is interesting that as compared to [55], where the number of  $K$  is required, the proposed constraint (5.6) does not depend on  $K$ . In order to explain how the constraint (5.6) avoids the convergence of the steering vector estimate to any linear combination of the steering vectors of the interferences, let  $\mathbf{A} \triangleq [\mathbf{a}_1, \dots, \mathbf{a}_J]$  denote a set of  $J$  plane wave steering vectors corresponding to the interferences that are located outside of the desired sector. Using (5.2) and (5.3), these steering vectors can be approximated as  $\mathbf{a}_j = \mathbf{U}_1 \mathbf{r}_j$ ,  $j = 1, \dots, J$  where  $\mathbf{r}_j$ ,  $j = 1, \dots, J$  are some  $K \times 1$  coefficient vectors. Then every linear combination of these steering vectors

which has norm squared equal to  $M$  can be expressed as

$$\mathbf{f} = \mathbf{A}\boldsymbol{\eta} = \mathbf{U}_1\mathbf{z} \quad (5.9)$$

where  $\boldsymbol{\eta}$  is a  $J \times 1$  coefficient vector,  $\mathbf{z} \triangleq [\mathbf{r}_1, \dots, \mathbf{r}_J]\boldsymbol{\eta}$ ,  $[\mathbf{r}_1, \dots, \mathbf{r}_J]$  is  $K \times J$  matrix, and  $\|\mathbf{z}\|^2 = \|\mathbf{a}\|^2 = M$ . Then the following is true

$$\begin{aligned} \mathbf{f}^H \tilde{\mathbf{C}}\mathbf{f} &= (\mathbf{U}_1\mathbf{z})^H \tilde{\mathbf{C}}\mathbf{U}_1\mathbf{z} = \mathbf{z}^H \boldsymbol{\Lambda}_1\mathbf{z} \\ &\geq M\lambda_K \geq M\lambda_{K+1} \geq \Delta_0 \end{aligned} \quad (5.10)$$

where the last inequality follows from (5.8). In fact, (5.10) implies that  $\mathbf{f}$  obtained as a linear combination of all interference steering vectors does not satisfy the constraint (5.6). Thus, (5.6) prevents the convergence of the estimate  $\hat{\mathbf{a}}$  to any of the interference steering vectors or their linear combination.

It is worth stressing that no restrictions/assumptions on the structure of the interferences are needed. Moreover, the interferences do not need to have the same structure as the desired signal. The constraint (5.6) does not use any of such information. Indeed, interferences may have a rank-one or multi-rank covariance matrix. Specifically, an interference source with a rank-one covariance matrix corresponds to either a plane wave source or a locally coherent scattered source [120]. An interference source with a multi-rank covariance matrix corresponds, for example, to a locally incoherent scattered source [97].

Steering vector of an interference with a locally coherent scattered source can be expressed as [120]

$$\tilde{\mathbf{a}}_i = \mathbf{a}_i + \sum_{l=1}^T e^{j\psi_l} \mathbf{b}(\theta_l) \quad (5.11)$$

where  $\mathbf{a}_i$  corresponds to the direct path of the interference and  $\mathbf{b}(\theta_l)$ ,  $l = 1, \dots, T$  correspond to the coherently scattered paths. Here  $\mathbf{b}(\theta_l)$  is a plane wave impinging on the array from the direction  $\theta_l$  which is fixed for different snapshots. The parameters  $\psi_l \in [0, 2\pi]$ ,  $l = 1, \dots, T$  denote the phase shift of different paths that are also fixed for different snapshots. Thus, the spatial signature of a locally coherent scattered source (5.11) is fixed over different snapshots and is a linear combination of plane wave steering vectors. A normalized linear combination of plane wave steering vectors that all lie outside of the desired sector does not satisfy the quadratic constraint (5.10). Thus, the interferences with rank-one covariance matrix will be avoided by means of the new constraint (5.6).

Steering vector of a locally incoherent scattered source is time-varying and can be modeled as [97]

$$\tilde{\mathbf{a}}_i(k) = s_0(k)\mathbf{a}_i + \sum_{l=1}^T s_l(k)\mathbf{b}(\theta_l) \quad (5.12)$$

where  $s_l(k)$ ,  $l = 0, \dots, T$  are independently and identically distributed (i.i.d.) zero-mean random variables with variances  $\nu_l$ ,  $l = 0, \dots, T$ . The random variables  $s_l(k)$  change from snapshot to snapshot. The correlation matrix of a locally incoherent scattered source (5.12) can be written as

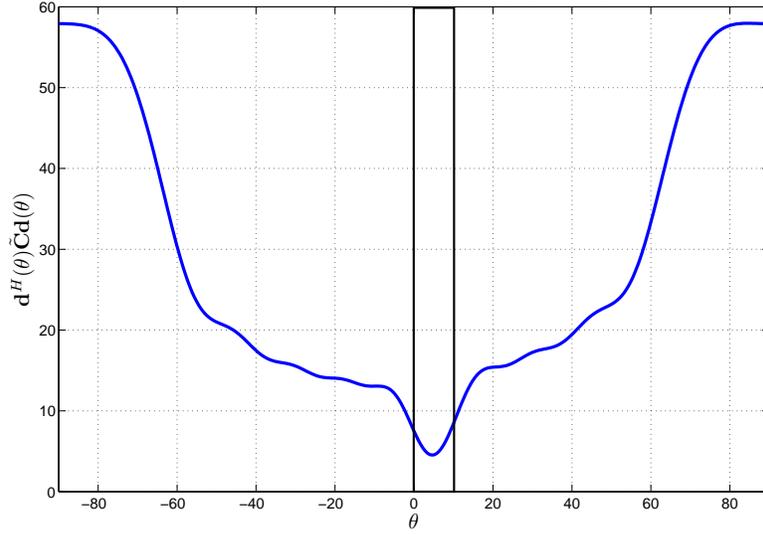
$$\mathbf{R}_i = \nu_0\mathbf{a}_i\mathbf{a}_i^H + \sum_{l=1}^T \nu_l\mathbf{b}(\theta_l)\mathbf{b}^H(\theta_l). \quad (5.13)$$

Based on (5.13), it can be concluded that a locally incoherent scattered source is equivalent to  $T + 1$  independent plane wave sources that impinge on the array from  $T + 1$  different directions. Thus, as long as all such plane waves lie outside of the desired sector  $\Theta$ , they do not satisfy the quadratic constraint (5.6) and the interferences with multi-rank covariance matrix will be avoided as well by means of the constraint (5.6).

To further illustrate how the constraint (5.6) works, let us consider the following example.

*Example 2:* Consider a ULA of 10 omni-directional antenna elements spaced half wavelength apart from each other. Let the range of the desired signal angular locations be  $\Theta = [0^\circ, 10^\circ]$ . Fig. 5.2 depicts the values of the quadratic term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  for different angles. The rectangular bar in the figure marks the directions within the presumed angular sector  $\Theta$ . It can be observed from this figure that the term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  has the smallest values within the angular sector  $\Theta$  and increases outside of the sector. Therefore, if  $\Delta_0$  is selected to be equal to the maximum value of the term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  within the presumed angular sector  $\Theta$ , the constraint (5.6) guarantees that the estimate of the desired signal steering vector does not converge to any of the interference steering vectors and their linear combinations. It is worth noting that  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta) = \Delta_0$  must occur at one of the edges of  $\Theta$ . However, the value of the quadratic term at the other edge of  $\Theta$  may be smaller than  $\Delta_0$ . Therefore, we define another sector  $\Theta_a \geq \Theta$  at which the equality  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta) = \Delta_0$  holds at both edges, i.e., the sector  $\Theta_a$  is the actual sector at which the constraint (5.6) must be satisfied.

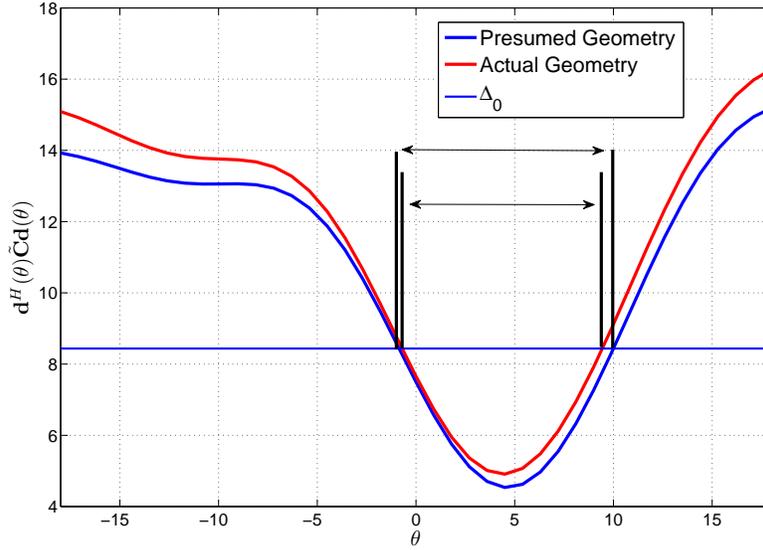
In order to compute the matrix  $\tilde{\mathbf{C}}$ , the presumed knowledge of the antenna



**Figure 5.2:** Values of the term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  in the constraint (5.6) for different angles.

array geometry is required. Due to the imperfect array calibration, the precise knowledge of the antenna array geometry may be unavailable. If array perturbation is present, the curve for the quadratic form  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  versus  $\theta$  can deviate from the one drawn under the assumption of no array perturbation. This situation is demonstrated in Fig. 5.3 which depicts the quadratic term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  versus  $\theta$  in the presence and in the absence of array perturbations. Although the angular sector  $\Theta_a$  computed for a given  $\Delta_0$  using (5.7) under the assumption of no array perturbation may change if the array is perturbed, the constraint (5.6) still remains precise as long as  $\Theta_a$  contains the desired signal and does not contain any interfering sources. Therefore, an inaccurate information about the antenna array geometry is sufficiently good. In fact, in one of our simulation examples we show that even if the perturbations of the antenna array geometry are the highest possible, the constraint (5.6) remains precise.

Taking into account the normalization constraint and the constraint (5.6), the problem of estimating the desired signal steering vector based on the knowledge of



**Figure 5.3:** Comparison between the quadratic term  $\mathbf{d}^H(\theta)\tilde{\mathbf{C}}\mathbf{d}(\theta)$  with and without array perturbations.

the sector  $\Theta$  can be formulated as the following optimization problem

$$\begin{aligned}
 \min_{\hat{\mathbf{a}}} \quad & \hat{\mathbf{a}}^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{a}} \\
 \text{subject to} \quad & \|\hat{\mathbf{a}}\|^2 = M \\
 & \hat{\mathbf{a}}^H \tilde{\mathbf{C}} \hat{\mathbf{a}} \leq \Delta_0.
 \end{aligned} \tag{5.14}$$

Unlike the last constraint in the problem (2.46) which was used for avoiding the noise power magnification, the second constraint in (5.14) is enforcing  $\hat{\mathbf{a}}$  not to converge to any steering vector associated with the interferences and their linear combinations. Compared to the other MVDR RAB methods, which require the knowledge of the presumed steering vector and, thus, the knowledge of the presumed antenna array geometry, propagation media, and source characteristics, only imprecise knowledge of the antenna array geometry and approximate knowledge of the angular sector  $\Theta$  are needed for the proposed method.

Due to the non-convex equality constraint in problem (5.14), the QCQP problem (5.14) is non-convex. Although the problem (5.14) is non-convex, the strong duality holds for the problems of this type and, thus, the solution based on SDR is the exact one. In the following section we develop some new results regarding this problem while looking for a new algebraic way of finding the rank-one solution from the

general-rank solution of the relaxed problem. We also obtain the condition under which the solution of the relaxed problem is guaranteed to be rank-one.

It is interesting that the steering vector estimation problem in [55] can be also expressed as a QCQP problem that makes it possible to find a much simpler solution than the sequential quadratic programming of [55] and draw some insightful connections to the newly proposed problem (5.14). Let us first find the set of vectors satisfying the constraint  $\mathbf{P}^\perp \hat{\mathbf{a}} = \mathbf{0}$  in problem (2.46). Note that  $\mathbf{P}^\perp \hat{\mathbf{a}} = \mathbf{0}$  implies that  $\hat{\mathbf{a}} = \mathbf{L}\mathbf{L}^H \hat{\mathbf{a}}$  where  $\mathbf{L}$  has been defined after equation (2.45) and, therefore, we can write that

$$\hat{\mathbf{a}} = \mathbf{L}\mathbf{b} \quad (5.15)$$

where  $\mathbf{b}$  is an  $L \times 1$  complex valued vector. Using (5.15), the optimization problem for estimating the steering vector in [55], i.e., (2.46), can be equivalently rewritten in terms of  $\mathbf{b}$  as

$$\begin{aligned} \min_{\mathbf{b}} \quad & \mathbf{b}^H \mathbf{L}^H \hat{\mathbf{R}}^{-1} \mathbf{L} \mathbf{b} \\ \text{subject to} \quad & \|\mathbf{b}\|^2 = M \\ & \mathbf{b}^H \mathbf{L}^H \tilde{\mathbf{C}} \mathbf{L} \mathbf{b} \leq \mathbf{p}^H \tilde{\mathbf{C}} \mathbf{p} \end{aligned} \quad (5.16)$$

which is a QCQP problem. Thus, as compared to (5.16), where the constraint  $\mathbf{P}^\perp \hat{\mathbf{a}} = \mathbf{0}$  enforces the estimated steering vector to be a linear combination of  $L$  dominant eigenvectors  $[\mathbf{l}_1, \dots, \mathbf{l}_L]$ , the steering vector in (5.14) is not restricted by such requirement, while the convergence to any of the interference steering vectors and their linear combinations is avoided by means of the constraint (5.6). As a result, the problem (5.14) has more degrees of freedom. Thus, it is expected that the new RAB method will outperform the one of [55].

As it has been explained in [55], the solution of the problem (2.46) leads to a better performance for the corresponding RAB compared to the other RAB techniques and, particularly, the worst-case-based and probabilistically constrained techniques. This performance improvement is the result of forming the beam toward a single corrected steering vector yielding maximum output power, while the worst-case-based method maximizes the output power for all steering vectors in its corresponding uncertainty set. Thus, despite a significantly more relaxed assumptions on the prior information, the performance of the new MVDR RAB technique based on (5.14) is expected to be superior to that of the other RAB techniques.

### 5.3 Steering vector estimation via SDP relaxation

The first step is to make sure that the problem (5.14) is feasible. Fortunately, it can be easily verified that (5.14) is feasible if and only if  $\Delta_0/M$  is greater than or equal to the smallest eigenvalue of the matrix  $\tilde{\mathbf{C}}$ . Indeed, if the smallest eigenvalue of  $\tilde{\mathbf{C}}$  is larger than  $\Delta_0/M$ , then the second constraint of (5.14) can not be satisfied for any estimate  $\hat{\mathbf{a}}$ . The selection of  $\Delta_0$  according to (5.7) satisfies the feasibility condition that guarantees the feasibility of (5.14).

#### 5.3.1 SDP relaxation

If the problem (5.14) is feasible, the equalities  $\hat{\mathbf{a}}^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{a}} = \text{tr}\{\hat{\mathbf{R}}^{-1} \hat{\mathbf{a}} \hat{\mathbf{a}}^H\}$  and  $\hat{\mathbf{a}}^H \tilde{\mathbf{C}} \hat{\mathbf{a}} = \text{tr}\{\tilde{\mathbf{C}} \hat{\mathbf{a}} \hat{\mathbf{a}}^H\}$  can be used to rewrite it as

$$\begin{aligned} \min_{\hat{\mathbf{a}}} \quad & \text{tr}\{\hat{\mathbf{R}}^{-1} \hat{\mathbf{a}} \hat{\mathbf{a}}^H\} \\ \text{subject to} \quad & \text{tr}\{\hat{\mathbf{a}} \hat{\mathbf{a}}^H\} = M \\ & \text{tr}\{\tilde{\mathbf{C}} \hat{\mathbf{a}} \hat{\mathbf{a}}^H\} \leq \Delta_0. \end{aligned} \quad (5.17)$$

Introducing the following positive semi-definite matrix variable  $\mathbf{A} \triangleq \hat{\mathbf{a}} \hat{\mathbf{a}}^H$ ,  $\mathbf{A} \succeq \mathbf{0}$ , the problem (5.17) can be recast as

$$\begin{aligned} \min_{\mathbf{A}} \quad & \text{tr}\{\hat{\mathbf{R}}^{-1} \mathbf{A}\} \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}\} = M \\ & \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}\} \leq \Delta_0 \\ & \text{rank}\{\mathbf{A}\} = 1. \end{aligned} \quad (5.18)$$

The only non-convex constraint in (5.18) is the rank-one constraint while all other constraint functions and the objective function are linear with respect to  $\mathbf{A}$ . Using the SDP relaxation technique, the relaxed problem can be obtained by dropping the non-convex rank-one constraint and requiring that  $\mathbf{A} \succeq \mathbf{0}$ . Thus, the problem (5.18) is replaced by the following relaxed convex problem

$$\begin{aligned} \min_{\mathbf{A}} \quad & \text{tr}\{\hat{\mathbf{R}}^{-1} \mathbf{A}\} \\ \text{subject to} \quad & \text{tr}\{\mathbf{A}\} = M \\ & \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}\} \leq \Delta_0 \\ & \mathbf{A} \succeq \mathbf{0}. \end{aligned} \quad (5.19)$$

### 5.3.2 Rank of the optimal solution

If  $\Delta_0$  is selected differently from (5.7), the original problem may be infeasible, while the relaxed one is feasible. However, by using the Lemma 3.1, the exact equivalence between the feasibility of the original and relaxed problem is concluded. In other words, the problem (5.19) is feasible if and only if the problem (5.14) is feasible.

If the optimal solution  $\mathbf{A}$  of the relaxed problem (5.19) is a rank-one matrix, then the principal eigenvector of  $\mathbf{A}$  scaled by the square root of the largest eigenvalue is the exact solution of the problem (5.14). However, even if  $\mathbf{A}$  is not rank-one, it has been shown in [16], [54] that the rank-one solution for the problems of type (5.14) can be found using the duality theory based on the fact that the strong duality holds for such problems. Moreover, the rank-one solution can be found using the well known rank reduction technique [106]. A new algebraic way of extracting the rank-one optimal solution of the problem (5.14) from the non-rank-one optimal solution of the problem (5.19) is summarized by means of the following new constructive theorem.

**Theorem 5.1.** *Let  $\mathbf{A}^*$  be the rank  $r$  optimal minimizer of the relaxed problem (5.19), i.e.,  $\mathbf{A}^* = \mathbf{Y}\mathbf{Y}^H$  where  $\mathbf{Y}$  is an  $M \times r$  full rank matrix. If  $r = 1$ , the optimal solution of the original problem simply equals  $\mathbf{Y}$ . Otherwise, it equals  $\mathbf{Y}\mathbf{v}$  where  $\mathbf{v}$  is an  $r \times 1$  vector such that  $\|\mathbf{Y}\mathbf{v}\| = \sqrt{M}$  and  $\mathbf{v}^H \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{v} = \text{tr}\{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}\}$ . One possible solution for the vector  $\mathbf{v}$  is proportional to the sum of the eigenvectors of the following  $r \times r$  matrix*

$$\mathbf{D} = \frac{1}{M} \mathbf{Y}^H \mathbf{Y} - \frac{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}}{\text{tr}\{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}\}}. \quad (5.20)$$

**Proof:** See Appendix A.

One more important question is under which condition the solution of the relaxed problem (5.19) is always rank-one. The importance of this question also follows from the fact that it has been observed that the probability of obtaining a rank-one solution for the class of considered problems is close to 1, while the theoretical upper-bound suggests a significantly smaller probability [14]. Our next result precisely explains and approves the correctness of the experimental observation about the high probability of the rank-one solution for the relaxed problem (5.19).

It is worth noting that any phase rotation of  $\hat{\mathbf{a}}$  does not change the SINR at the output of the corresponding RAB. Therefore, we say that the optimal solution  $\hat{\mathbf{a}}$  is

unique when the value of the output SINR or output power (5.1) is the same for any  $\hat{\mathbf{a}}' = \hat{\mathbf{a}}e^{j\phi}$ . Then the following lemma holds.

**Lemma 5.1.** *Under the condition that the solution of the original problem (5.14) is unique in the sense mentioned above, the solution of the relaxed problem (5.19) always has rank one.*

**Proof:** See Appendix B

Under the condition of Lemma 5.1, the solution of (5.19) is rank-one and the solution of (5.14) can be found as a scaled version of the dominant eigenvector of the solution of (5.19). If the uniqueness condition of Lemma 5.1 is not satisfied for (5.14), we resort to the constructive result of Theorem 5.1 for finding the rank-one solution of (5.14) algebraically. An example of a situation when the condition of Lemma 5.1 is not satisfied is given next. In general, such situations are rare that, in fact, has been also observed by means of simulations in other works.

*Example 3:* Let us consider a ULA with 10 omni-directional antenna elements. The presumed direction of arrival of the desired user is assumed to be  $\theta_p = 3^\circ$  with no interfering sources and the range of the desired signal angular locations is equal to  $\Theta = [\theta_p - 12^\circ, \theta_p + 12^\circ]$ . The actual steering vector of the desired user is perturbed due to the incoherent local scattering effect and it can be expressed as  $\tilde{\mathbf{a}}(k) = v_0(k)\mathbf{p} + v_1(k)\mathbf{b}$ , where  $\mathbf{p} = \mathbf{d}(3^\circ)$  is the steering vector of the direct path,  $\mathbf{b}$  is the steering vector of the scattered path, and  $v_0(k)$  and  $v_1(k)$  are i.i.d. zero mean complex Gaussian random variables with unit variance which change from snapshot to snapshot. If  $\mathbf{b}$  is orthogonal to  $\mathbf{p}$ , that is the case when  $\mathbf{b}$  is selected as  $\mathbf{d}(-8.4916^\circ)$ , both  $\mathbf{p}$  and  $\mathbf{b}$  are the eigenvectors of the matrix  $\mathbf{R}^{-1}$  which correspond to the smallest eigenvalue. Since, these vectors satisfy the constraints the constraints of the problem (5.14) and correspond to the minimum eigenvalue, both of them are optimal solutions of (5.14). Thus, the solution of (5.14) is not unique.

## 5.4 Simulation results

Throughout the simulations, a ULA of 10 omni-directional antenna elements with the inter-element spacing of half wavelength is considered unless otherwise is specified. Additive noise in antenna elements is modeled as spatially and temporally independent complex Gaussian noise with zero mean and unit variance. Two interfering sources are assumed to impinge on the antenna array from the directions  $30^\circ$

and  $50^\circ$ , while the presumed direction towards the desired signal is assumed to be  $\theta_p = 3^\circ$  unless otherwise is specified. In all simulation examples, the interference-to-noise ratio (INR) equals 30 dB and the desired signal is always present in the training data. For obtaining each point in the curves, 100 independent runs are used.

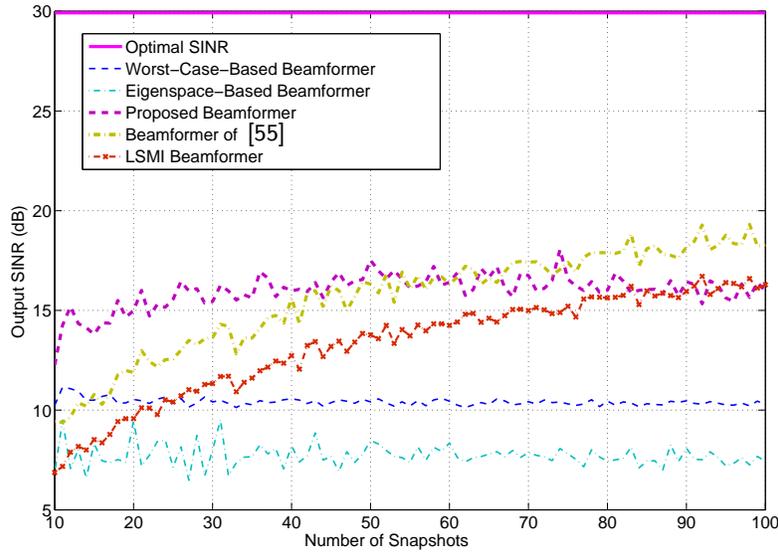
The proposed beamformer is compared with the following four methods in terms of the output SINR: (i) the eigenspace-based beamformer of [94], (ii) the worst-case-based RAB of [7], (iii) the beamformer of [55], and (iv) the diagonally loaded SMI (LSMI) beamformer [46]. Moreover, in the last simulation example, a comparison is made with the multi-rank eigenvalue beamformer of [97]. For the proposed beamformer and the beamformer of [55], the angular sector of interest  $\Theta$  is assumed to be  $\Theta = [\theta_p - 5^\circ, \theta_p + 5^\circ]$ . The CVX MATLAB toolbox is used for solving the optimization problem (5.19). The value  $\delta = 0.1$  and 8 dominant eigenvectors of the matrix  $\mathbf{C}$  are used in the beamformer of [55] and the value  $\varepsilon = 0.3M$  is used for the worst-case-based beamformer as it has been recommended in [7]. The dimension of the signal-plus-interference subspace is assumed to be always estimated correctly for the eigenspace-based beamformer of [94]. Diagonal loading factor of the SMI beamformer is selected as twice the noise power as recommended by Cox *et al.* in [46].

#### 5.4.1 Example 1 : Exactly known signal steering vector

In this example, we consider the case when the actual steering vector is known exactly. Even in this case, the presence of the desired signal in the training data can substantially reduce the convergence rates of adaptive beamforming algorithms as compared to the signal-free training data case [48].

In Fig. 5.4, the mean output SINRs for the four methods tested are illustrated versus the number of training snapshots for the fixed single-sensor SNR = 20 dB. Fig. 5.5 displays the mean output SINR of the same methods versus the SNR for fixed training data size of  $K = 30$ . It can be seen from these figures that the proposed beamforming technique outperforms the other techniques. Only in the situation when the number of snapshots is larger than 70, the technique of [55] results in slightly better performance. It is because if the desired signal steering vector is known precisely, the only source of error is the finite sample size used, but the difference between the sample data covariance matrix and the theoretical data covariance matrix due to finite sample size can be equivalently transferred to

the error in the steering vector [48]. Therefore, as the number of samples increases and the variance of the steering vector mismatch decreases, the estimator with more restrictive constraints, that is the one of [55], may indeed result in a better performance.

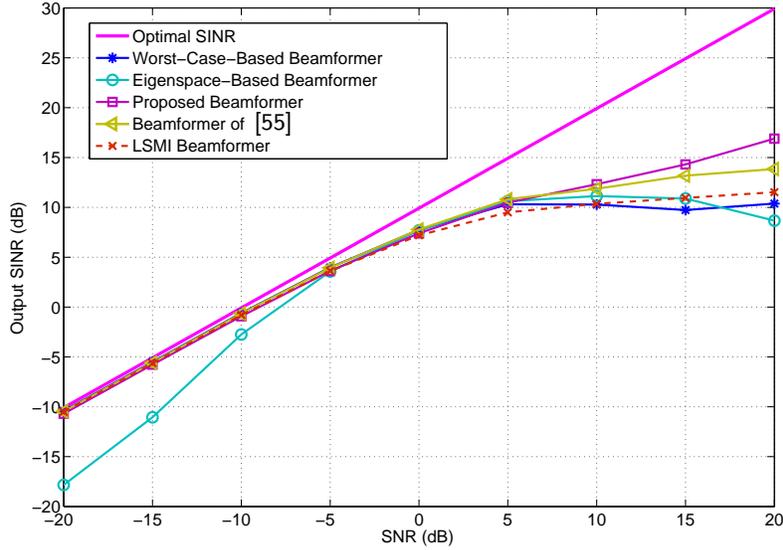


**Figure 5.4:** Example 1: Output SINR versus training sample size  $K$  for fixed SNR = 20 dB and INR = 30 dB.

#### 5.4.2 Example 2 : Desired signal steering vector mismatch due to wavefront distortion

In the second example, we consider the situation when the signal steering vector is distorted by wave propagation effects in an inhomogeneous medium. Specifically, independent-increment phase distortions are accumulated by the components of the presumed steering vector. It is assumed that the phase increments remain fixed in each simulation run and are independently chosen from a Gaussian random generator with zero mean and standard deviation 0.04.

The output SINR curves for the proposed, SMI, and worst-case-based methods are shown versus the SNR for fixed training data size  $K = 30$  in Fig. 5.6. It can be seen that the proposed beamforming technique outperforms the worst-case-based one at low and moderate SNRs. However, when SNR is much larger than INR so that signal-to-interference ratio goes to infinity, the proposed and the worst-case-based MVDR RAB techniques perform almost equivalently. Indeed, in the latter case, it is

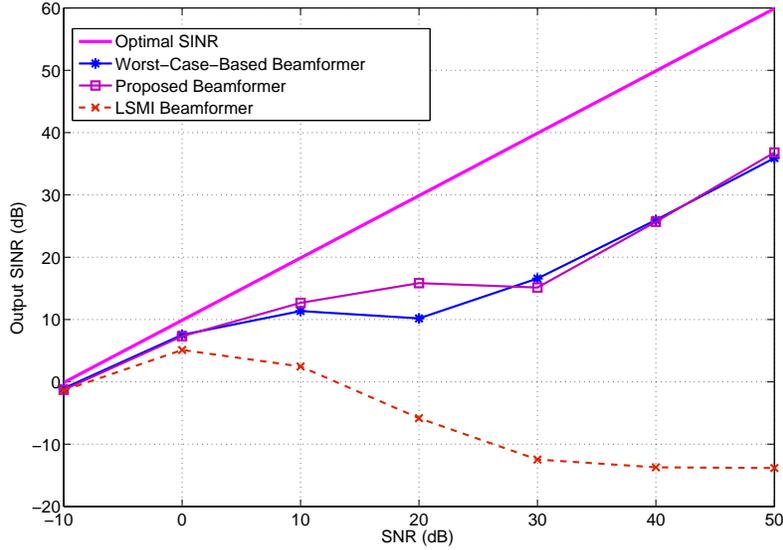


**Figure 5.5:** Example 1: Output SINR versus SNR for training data size of  $K = 30$  and  $\text{INR} = 30$  dB.

guaranteed that the estimate of the desired signal steering vector does not converge to an interference steering vector and, thus, the second constraint of (5.14) is never active and can be dropped. The solution of the problem (5.14) without considering the second constraint is  $\hat{\mathbf{a}} = \sqrt{M}\mathcal{P}(\hat{\mathbf{R}}^{-1})$ . However, the problem (5.14) without the second constraint coincides with the problem [53, (39)] obtained after dropping the constraint  $\|\delta\| \leq \varepsilon$ . Thus, the proposed and the worst-case-based MVRD RAB techniques should indeed yield the same solution  $\mathbf{w}_1 = \hat{\mathbf{R}}^{-1}\mathcal{P}\{\hat{\mathbf{R}}^{-1}\}$ .

### 5.4.3 Example 3 : Effect of the error in the knowledge of the antenna array geometry

In this example, we first aim at checking how the presence of antenna array perturbations affect the sector  $\Theta_a$ . Specifically, we want to characterize quantitatively the dependence of  $\Theta_a$  on the level of antenna array perturbations, which grows from zero to its maximum value, for a given  $\Delta_0$ . The presumed angular sector is assumed to be  $[0^\circ, 10^\circ]$  for  $\theta_p = 5^\circ$ . Let the antenna array perturbations be caused by errors in the antenna element positions which are drawn uniformly from  $[-\alpha, \alpha]$  where  $\alpha$  is the level of perturbations measured in wavelength. Table 5.2 illustrates the average width of  $\Theta_a$  for different values of  $\alpha$ . It can be seen from this table that the deviation of  $\Theta_a$  due to array perturbations compared to the case of no perturbations is 0% for



**Figure 5.6:** Example 2: Output SINR versus SNR for training data size of  $K = 30$  and  $\text{INR} = 30$  dB.

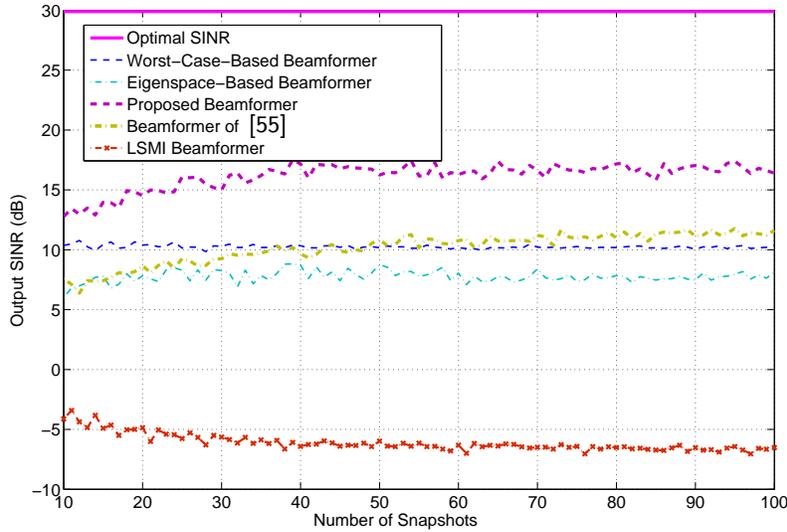
small perturbations and near 0% even for the highest levels of perturbations. Here the term ‘deviation’ stands for the percent of non-overlap between angular sectors corresponding to the cases of no perturbations and perturbations present. Based on this observation, one can conclude that the matrix  $\tilde{\mathbf{C}}$  required for implementing the constraint (5.6) can be computed using the presumed array geometry which is not required to be precise and can be, in fact, very approximate. It is worth noting that we also observed throughout extensive simulations that if the sector  $\Theta$  gets away from the broadside (it is near  $-90^\circ$  or  $90^\circ$ ), then the length of the associated sector  $\Theta_a$  increases. Such increase can be noticeable especially when the number of antenna elements in the antenna array is small. To avoid the situation when  $\Theta_a$  may contain interference sources because of its bigger size, the presteering filter-type technique [121] can be used, for example, to ensure that the center of the sector  $\Theta$  for the signal at the output of the presteering filter is around the broadside.

We also study the effect of the error in the knowledge of antenna array geometry used for the computation of the matrix  $\tilde{\mathbf{C}}$  on the performance of the proposed MVDR RAB technique. The difference between the presumed and actual positions of each antenna element is modeled as a uniform random variable distributed in the interval  $[-0.05, 0.05]$  measured in wavelength. In addition to the antenna element

**Table 5.2:** Width of the feasible set  $\Theta_a$  versus the level of perturbations  $\alpha$ .

Level of perturbations $\alpha$	0	0.05	0.1	0.15	0.2	0.25
Width of $\Theta_a$	10.80°	10.80°	10.72°	10.57°	10.53°	10.42°
Deviation(%)	0%	0%	0.7%	2.06%	2.48%	3.48%

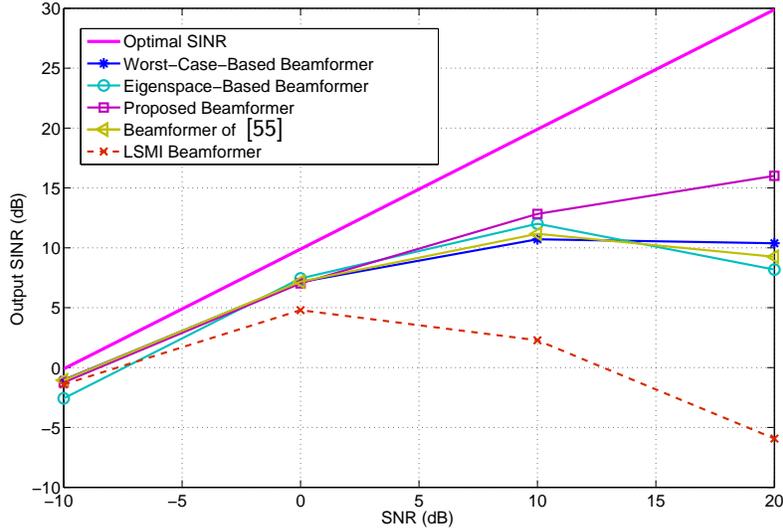
displacements, the signal steering vector is distorted as in our Simulation Example 2. Figs. 5.7 and 5.8 depict the output SINR performance of the RAB techniques tested versus the number of training snapshots for fixed single-sensor SNR= 20 dB and versus the SNR for fixed training data size  $K = 30$ , respectively. As it can be observed from the figures, the proposed method has a better performance even if there is an error in the knowledge of the antenna array geometry.



**Figure 5.7:** Example 3: Output SINR versus training sample size  $K$  for fixed SNR = 20 dB and INR = 30 dB for the case of perturbations in antenna array geometry.

#### 5.4.4 Example 4 : Desired signal steering vector mismatch due to coherent local scattering [120]

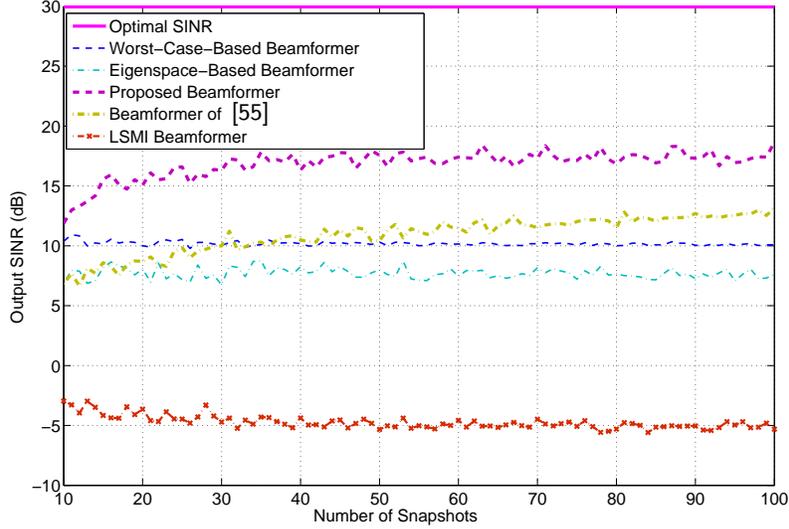
In this example, the desired signal steering vector is distorted by local scattering effects so that the actual steering vector is formed by five signal paths as  $\mathbf{a} = \mathbf{p} + \sum_{i=1}^4 e^{j\psi_i} \mathbf{b}(\theta_i)$  where  $\mathbf{p}$  corresponds to the direct path and  $\mathbf{b}(\theta_i)$ ,  $i = 1, 2, 3, 4$  correspond to the coherently scattered paths. The  $i$ th path  $\mathbf{b}(\theta_i)$  is modeled as a



**Figure 5.8:** Example 3: Output SINR versus SNR for training data size of  $K = 30$  and INR = 30 dB for the case of perturbations in antenna array geometry.

plane wave impinging on the array from the direction  $\theta_i$ . The angles  $\theta_i$ ,  $i = 1, 2, 3, 4$  are independently drawn in each simulation run from a uniform random generator with mean  $3^\circ$  and standard deviation  $1^\circ$ . The parameters  $\psi_i$ ,  $i = 1, 2, 3, 4$  represent path phases that are independently and uniformly drawn from the interval  $[0, 2\pi]$  in each simulation run. Note that  $\theta_i$  and  $\psi_i$ ,  $i = 1, 2, 3, 4$  change from run to run but do not change from snapshot to snapshot.

Fig. 5.9 displays the output SINR performance of all four methods tested versus the number of training snapshots  $K$  for fixed single-sensor SNR = 20 dB. Note that the SNR in this example is defined by taking into account all signal paths. The output SINR performance of the same methods versus SNR for the fixed training data size  $K = 30$  is displayed in Fig. 5.10. Similar to the previous example, the proposed beamformer significantly outperforms other beamformers due to its ability to estimate the desired signal steering vector with higher accuracy than other methods. As compared to the eigenspace-based method, the proposed technique does not suffer from the subspace swap phenomenon at low SNRs since it does not use eigenvalue decomposition of the sample covariance matrix.



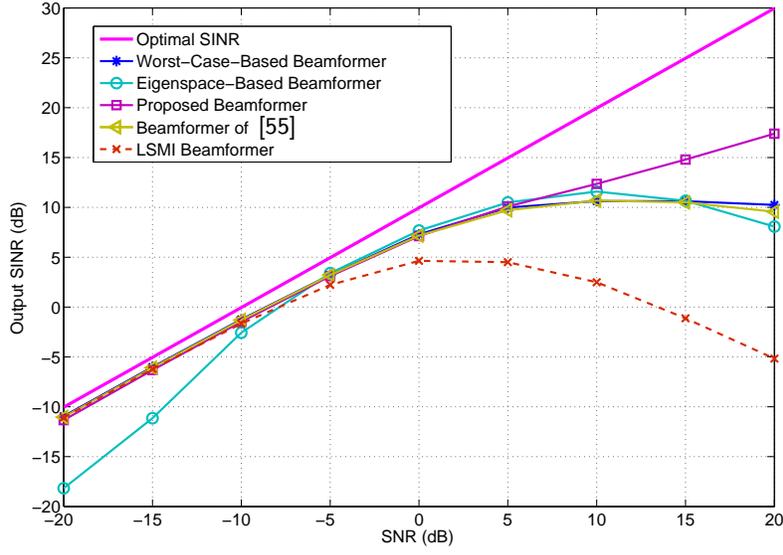
**Figure 5.9:** Example 4: Output SINR versus training sample size  $K$  for fixed SNR = 20 dB and INR = 30 dB.

#### 5.4.5 Example 5 : Comparison with eigenvalue beamforming-based methods of [97]

In this example, we consider the case where the desired and interference signals have the same structure and are modeled as signals with a rank-one covariance matrix from a  $p$ -dimensional subspace. Specifically, the model introduced in [97] for the desired and interference signals is adopted. The corresponding steering vector of the desired and interference signals are all modeled as  $\mathbf{s} = \mathbf{\Psi}\mathbf{b}_0s$  where  $\mathbf{\Psi}$  is an  $M \times p$  ( $p < M$ ) matrix whose columns are orthogonal ( $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_{p \times p}$ ) and  $\mathbf{b}_0$  is an unknown but fixed vector from one snapshot to another. The matrix  $\mathbf{\Psi}$  (different for each signal) is obtained by choosing  $p = 3$  dominant eigenvectors of the matrix  $\int_{\phi_p - \Delta\phi}^{\phi_p + \Delta\phi} \mathbf{d}(\theta)\mathbf{d}^H(\theta)d\theta$  as the columns of  $\mathbf{\Psi}$  where  $\phi_p$  denotes the presumed location of the source and  $\Delta\phi$  equals to  $5^\circ$  for all the signals. In order to satisfy the assumption of the proposed RAB which requires the desired user to lie inside the desired sector, the number of the antenna elements is taken to be equal to  $M = 30$ . Note that if  $M = 10$  the matrix  $\int_{\phi_p - \Delta\phi}^{\phi_p + \Delta\phi} \mathbf{d}(\theta)\mathbf{d}^H(\theta)d\theta$  does not have three dominant eigenvectors.

To obtain the signal subspace which corresponds to the desired signal, we find the maximum of the *bearing pattern response* defined as

$$P_0(\theta) = \text{tr} \{ \mathbf{Q}^H (\mathbf{\Psi}^H(\theta)\mathbf{R}^{-1}\mathbf{\Psi}(\theta))^{-1}\mathbf{Q} \} \quad (5.21)$$

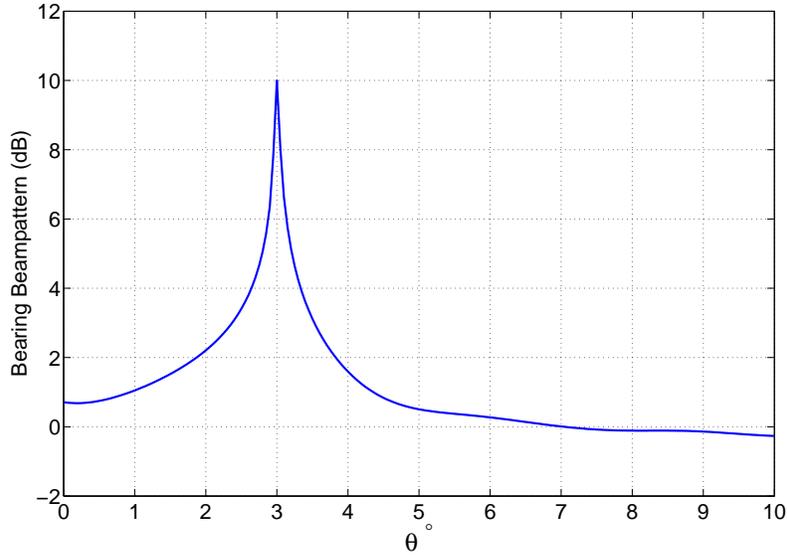


**Figure 5.10:** Example 4: Output SINR versus SNR for training data size of  $K = 30$  and  $\text{INR} = 30$  dB.

in the desired sector where  $\Psi(\theta)$  is the orthogonal matrix whose columns are equal to the three dominant eigenvectors of the  $M \times p$  matrix  $\int_{\theta-\Delta\phi}^{\theta+\Delta\phi} \mathbf{d}(\theta)\mathbf{d}^H(\theta)d\theta$  [97]. Fig. 5.11 shows the sample bearing pattern response. As it is expected, the maximum occurs exactly around  $3^\circ$ .

It is noteworthy to mention that the proposed RAB, the RAB of [55], and the eigenvalue beamformer of [97] all use the knowledge of approximate antenna array geometry. In order to evaluate how the error in the knowledge of antenna array geometry affects these methods, two different cases are considered. In the first case, knowledge of antenna array geometry is accurate while in the second case, antenna array perturbations are considered. Similar to our Simulation Example 3, antenna array perturbations are modeled as errors in the antenna element positions which are drawn uniformly from the interval  $[-0.05, 0.05]$  measured in wavelength. For the multi-rank beamformer of [97], we use  $\text{tr}\{\mathbf{W}^H \mathbf{R}_s \mathbf{W}\} / \text{tr}\{\mathbf{W}^H \mathbf{R}_{i+n} \mathbf{W}\}$  as the output SINR where  $\mathbf{R}_s$  denotes the correlation matrix of the desired signal. Figs. 5.12 and 5.13 illustrate the performance of the aforementioned methods for the cases when the knowledge of antenna array geometry is accurate and approximate, respectively. The best performance by the eigenvalue beamformer of [97] is obtained when  $\mathbf{Q}$  only contains the most dominant eigenvector of the error covariance matrix.

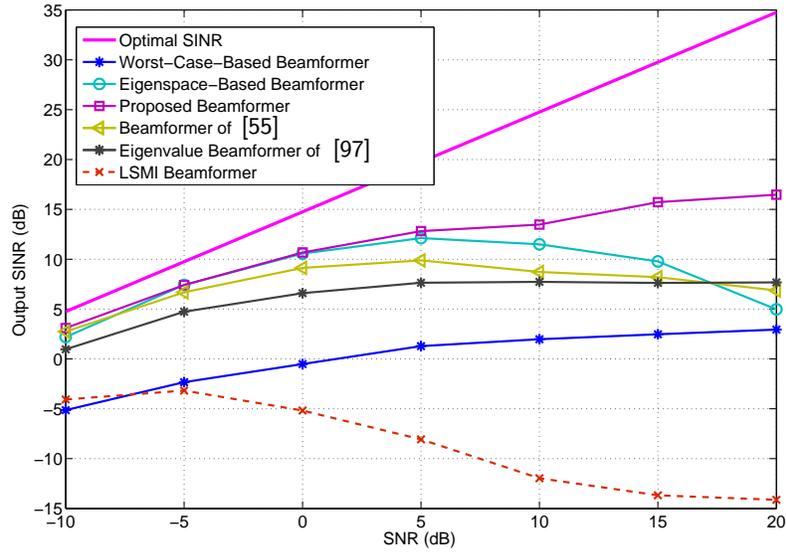
As it can be seen from the figures, the proposed RAB method outperforms all other RAB methods.



**Figure 5.11:** Example 5: Bearing beampattern corresponding to the eigenvalue beamforming method of [97].

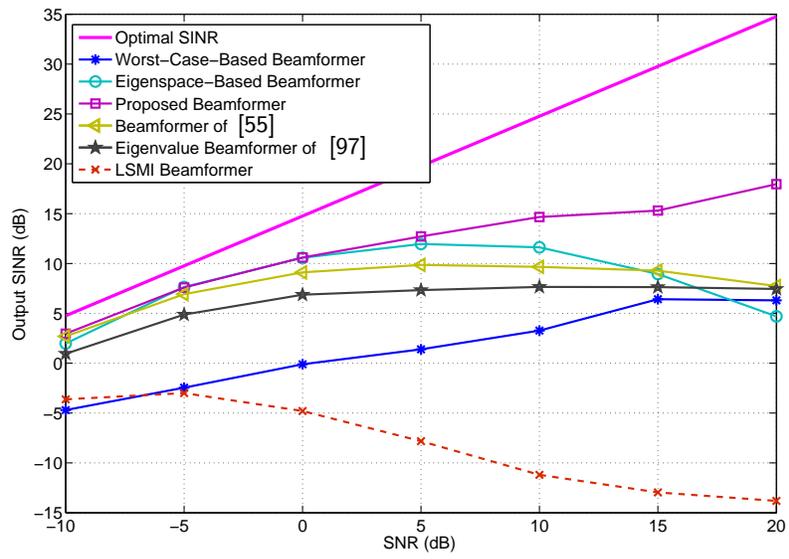
## 5.5 Chapter summary

The MVDR RAB techniques have been considered from the viewpoint of a single unified principle, that is, to use standard MVDR beamforming in tandem with an estimate of the signal steering vector found based on some prior information. It has been demonstrated that differences between various MVDR RAB techniques occur only because of the differences in the assumed prior information and the corresponding signal steering vector estimation techniques. The latter fact has motivated us to develop a new MVDR RAB technique that uses as little as possible, imprecise, and easy to obtain prior information. The new MVDR RAB technique, which assumes only an imprecise knowledge of antenna array geometry and angular sector in which the actual steering vector lies, has been developed. It is mathematically expressed as the well known non-convex QCQP problem with one convex quadratic inequality constraint and one non-convex quadratic equality constraint. A number of methods for finding efficiently the exact optimal solution for such problem is known. In addition to the existing methods, we have developed a new algebraic



**Figure 5.12:** Example 5: Output SINR versus SNR for training data size of  $K = 30$  and  $\text{INR} = 30$  dB.

method of finding the rank-one solution from the general-rank solution of the relaxed problem. The condition under which the solution of the relaxed problem is guaranteed to be rank-one has been also derived. Our simulation results demonstrate the superior performance for the proposed MVDR RAB technique over the existing state-of-the-art RAB methods.



**Figure 5.13:** Example 5: Output SINR versus SNR for training data size of  $K = 30$  and  $\text{INR} = 30$  dB.

## Chapter 6

# Robust Adaptive Beamforming for General-Rank Signal Model with Positive Semi-Definite Constraint

In the previous chapter, we introduced a unified principle for the MVDR RAB design problem. Based on the newly introduced unified principle, we then developed a new approach to RAB problem in the presence of signal steering vector mismatches. The newly developed approach had the advantage of using as little as possible and easy to obtain prior information. The RAB methods that were discussed and established in the previous chapter are based on the point source signal model assumption, i.e., when the rank of the desired signal covariance matrix is equal to one [49]–[52]. Despite the excellent robustness of these methods against mismatches of the underlying point source assumption, they do not provide enough robustness when the rank of the desired signal covariance matrix is higher than one. Therefore in this chapter, we consider the RAB problem when the rank of the source covariance matrix is higher than one.

The RAB for the general-rank signal model based on the explicit modeling of the error mismatches has been developed in [8] based on the worst-case performance optimization principle. Although the RAB of [8] has a simple closed-form solution, it is overly conservative because the worst-case correlation matrix of the desired signal may be negative definite [10], [56] (see also Subsection 2.3.3). Thus, less conservative approaches have been developed in [10], [56] by considering an additional PSD constraint to the worst-case signal covariance matrix. The major shortcoming of the RAB methods of [10], [56] is that they find only a suboptimal solution and there may

be a significant gap to the globally optimal solution. For example, the RAB of [10] finds a suboptimal solution in an iterative way, but there is no guarantee that such iterative method converges [56]. A closed-form approximate suboptimal solution is proposed in [122], however, this solution may be quite far from the globally optimal one as well. All these shortcomings motivate us to look for new efficient ways to solve the aforementioned non-convex problem hopefully globally optimally.

Interestingly, the resulted non-convex optimization problem is a special case of the general form of the QCQP problems (3.1) that was introduced in Chapter 3 of this thesis. As a result, we can solve the corresponding non-convex optimization problem by means of the POTDC algorithm (see Subsection 3.1.1). Specifically, the original non-convex optimization problem is first recast as the minimization of a one-dimensional optimal value function. Although the corresponding optimization problem of the newly introduced optimal value function is non-convex, it is then replaced with another equivalent function. The optimization problem that corresponds to such new optimal value function is convex and can be solved efficiently. The new one-dimensional optimal value function is then minimized using the POTDC algorithm. The point found by the POTDC algorithm for RAB of general-rank signal model with positive semi-definite constraint is guaranteed to be a KKT point. We prove a number of results that lead us to the equivalence between the claim of global optimality for the POTDC algorithm as applied to the problem under consideration and the convexity of the newly obtained one-dimensional optimal value function. The latter convexity of the newly obtained one-dimensional optimal value function can be checked numerically by using the convexity on lines property of convex functions. The fact that enables such numerical check is that the argument of such optimal value function is proved to take values only in a closed interval. In addition, we also develop tight lower-bound for such optimal value function that is used in the simulations for further confirming global optimality of the POTDC method. The rest of the chapter is organized as follows. Problem is formulated in Section 6.1. A new method is developed in Section 6.2 followed by simulation results in Section 6.3. Finally, Section 6.4 presents our conclusions.

## 6.1 Problem formulation

Decomposing  $\mathbf{R}_s$  as  $\mathbf{R}_s = \mathbf{Q}^H \mathbf{Q}$ , the RAB problem for a norm-bounded mismatch  $\|\Delta\| \leq \eta$  to the matrix  $\mathbf{Q}$  is given as [10]

$$\begin{aligned} & \min_{\mathbf{w}} \quad \max_{\|\Delta_2\| \leq \gamma} \mathbf{w}^H (\hat{\mathbf{R}} + \Delta_2) \mathbf{w} \\ \text{subject to} \quad & \min_{\|\Delta\| \leq \eta} \mathbf{w}^H (\mathbf{Q} + \Delta)^H (\mathbf{Q} + \Delta) \mathbf{w} \geq 1. \end{aligned} \quad (6.1)$$

Note that the norm-bounded mismatch to the matrix  $\mathbf{Q}$  has been first adopted in [10]. For every  $\Delta$  in the optimization problem (6.1) whose norm is less than or equal to  $\eta$ , the expression  $\mathbf{w}^H (\mathbf{Q} + \Delta)^H (\mathbf{Q} + \Delta) \mathbf{w} \geq 1$  represents a non-convex quadratic constraint with respect to  $\mathbf{w}$ . Because there exists infinite number of mismatches  $\Delta$ , there also exists infinite number of such non-convex quadratic constraints. By finding the minimum possible value of the quadratic term  $\mathbf{w}^H (\mathbf{Q} + \Delta)^H (\mathbf{Q} + \Delta) \mathbf{w}$  with respect to  $\Delta$  for a fixed  $\mathbf{w}$ , the infinite number of such non-convex quadratic constraints can be replaced with a single constraint. For this goal, we consider the following optimization problem

$$\begin{aligned} & \min_{\Delta} \quad \mathbf{w}^H (\mathbf{Q} + \Delta)^H (\mathbf{Q} + \Delta) \mathbf{w} \\ \text{subject to} \quad & \|\Delta\|^2 \leq \eta^2. \end{aligned} \quad (6.2)$$

This problem is convex and its optimal value can be expressed as a function of  $\mathbf{w}$  as given by the following lemma.

**Lemma 6.1.** *The optimal value of the optimization problem (6.2) as a function of  $\mathbf{w}$  is equal to*

$$\min_{\|\Delta\|^2 \leq \eta^2} \mathbf{w}^H (\mathbf{Q} + \Delta)^H (\mathbf{Q} + \Delta) \mathbf{w} = \begin{cases} (\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\|)^2, & \|\mathbf{Q}\mathbf{w}\| \geq \eta\|\mathbf{w}\| \\ 0, & \text{otherwise.} \end{cases} \quad (6.3)$$

**Proof:** See Appendix C.

Maximum of the quadratic term  $\mathbf{w}^H (\hat{\mathbf{R}} + \Delta_2) \mathbf{w}$  with respect to  $\Delta_2$ ,  $\|\Delta_2\| \leq \gamma$  that appears in the objective of the problem (6.1) can be easily derived as  $\mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w}$ . It is obvious from (6.3) that the desired signal can be totally removed from the beamformer output if  $\|\mathbf{Q}\mathbf{w}\| < \eta\|\mathbf{w}\|$ . Based on the later fact,  $\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\|$  should be greater than or equal to zero. For any such  $\mathbf{w}$ , the new constraint in the optimization problem (6.1) can be expressed as  $\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1$ . Since

$\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1$  also implies that  $\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 0$ , the RAB problem (6.1) can be equivalently rewritten as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w} \\ \text{subject to} \quad & \|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1. \end{aligned} \tag{6.4}$$

Due to the non-convex DC constraint, the problem (6.4) is non-convex DC programming problem. DC optimization problems are believed to be NP-hard in general [32], [33]. There is a number of methods that can be applied to address DC problem of type (6.4) in the literature. Among these methods are the generalized polyblock algorithm, the extended general power iterative algorithm [28], DC iteration-based method [123], etc. However, the existing methods do not guarantee to find the solution of (6.4), i.e., to converge to the global optimum of (6.4), in polynomial time. This means that the problem (6.4) may be NP-hard. The best what is possible to show, for example, for the DC iteration-based method is that it can find a KKT point. The overall computational complexity of the DC iteration-based method can be, however, quite high because the number of iterations required to converge grows dramatically with the dimension of the problem.

Recently, the problem (6.4) has also been suboptimally solved using an iterative SDR-based algorithm in [10] which also does not result in the globally optimal solution and for which the convergence even to a KKT point is not guaranteed. A closed-form suboptimal solution for the aforementioned non-convex DC problem has been also derived in [122]. Despite its computational simplicity, the performance of the method of [122] may be far from the global optimum and even the KKT point. Another iterative algorithm has been proposed in [56], but it modifies the problem (6.4) and solves the modified problem instead which again gives no guarantees for finding the globally optimal solution of the original problem (6.4). In what follows, we develop a new polynomial time algorithm for addressing the DC programming problem of (6.4) by means of the POTDC algorithm.

## 6.2 New proposed method

By introducing the auxiliary optimization variable  $\alpha \geq 1$  and setting  $\|\mathbf{Q}\mathbf{w}\| = \sqrt{\alpha}$ , the problem (6.4) can be equivalently rewritten as

$$\begin{aligned} \min_{\mathbf{w}, \alpha} \quad & \mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w} = \alpha \\ & \mathbf{w}^H \mathbf{w} \leq \frac{(\sqrt{\alpha} - 1)^2}{\eta^2}, \quad \alpha \geq 1. \end{aligned} \quad (6.5)$$

Note that  $\alpha$  is restricted to be greater than or equal to one because  $\|\mathbf{Q}\mathbf{w}\|$  is greater than or equal to one due to the constraint of the problem (6.4). For future needs, we find the set of all  $\alpha$ 's for which the optimization problem (6.5) is feasible. Let us define the following set for a fixed value of  $\alpha \geq 1$ ,

$$S(\alpha) \triangleq \{\mathbf{w} \mid \mathbf{w}^H \mathbf{w} \leq (\sqrt{\alpha} - 1)^2 / \eta^2\}. \quad (6.6)$$

It is trivial that for every  $\mathbf{w} \in S(\alpha)$ , the quadratic term  $\mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w}$  is non-negative as  $\mathbf{Q}^H \mathbf{Q}$  is a positive semi-definite matrix. Using the minimax theorem [124], it can be easily verified that the maximum value of the quadratic term  $\mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w}$  over  $\mathbf{w} \in S(\alpha)$  is equal to  $((\sqrt{\alpha} - 1)^2 / \eta^2) \cdot \lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\}$  and this value is achieved by

$$\mathbf{w}_\alpha = \frac{\sqrt{\alpha} - 1}{\eta} \mathcal{P}\{\mathbf{Q}^H \mathbf{Q}\} \in S(\alpha). \quad (6.7)$$

Here  $\lambda_{\max}\{\cdot\}$  stands for the largest eigenvalue operator. Due to the fact that for any  $0 \leq \beta \leq 1$ , the scaled vector  $\beta \mathbf{w}_\alpha$  lies inside the set  $S(\alpha)$  (6.6), the quadratic term  $\mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w}$  can take values only in the interval  $[0, ((\sqrt{\alpha} - 1)^2 / \eta^2) \cdot \lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\}]$  over  $\mathbf{w} \in S(\alpha)$ .

Considering the later fact and also the optimization problem (6.5), it can be concluded that  $\alpha$  is feasible if and only if  $\alpha \in [0, ((\sqrt{\alpha} - 1)^2 / \eta^2) \cdot \lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\}]$  which implies that

$$\frac{(\sqrt{\alpha} - 1)^2}{\eta^2} \cdot \lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\} \geq \alpha \quad (6.8)$$

or, equivalently, that

$$\frac{(\sqrt{\alpha} - 1)^2}{\alpha} \geq \frac{\eta^2}{\lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\}}. \quad (6.9)$$

The function  $(\sqrt{\alpha} - 1)^2 / \alpha$  is strictly increasing and it is also less than or equal to one for  $\alpha \geq 1$ . Therefore, it can be immediately found that the problem (6.5) is infeasible for any  $\alpha \geq 1$  if  $\lambda_{\max}\{\mathbf{Q}^H \mathbf{Q}\} \leq \eta^2$ . Thus, hereafter, it is assumed that

$\lambda_{\max}\{\mathbf{Q}^H\mathbf{Q}\} > \eta^2$ . Moreover, using (6.9) and the fact that the function  $(\sqrt{\alpha}-1)^2/\alpha$  is strictly increasing, it can be found that the feasible set of the problem (6.5) corresponds to

$$\alpha \geq \frac{1}{\left(1 - \frac{\eta}{\sqrt{\lambda_{\max}\{\mathbf{Q}^H\mathbf{Q}\}}}\right)^2} \geq 1. \quad (6.10)$$

As we will see in the following sections, for developing a lower-bound for the problem (6.5), an upper-bound for the optimal value of  $\alpha$  in (6.5) is needed. Such upper-bound is obtained in terms of the following lemma.

**Lemma 6.2.** *The optimal value of the optimization variable  $\alpha$  in the problem (6.5) is upper-bounded by  $\lambda_{\max}\left\{(\hat{\mathbf{R}} + \gamma\mathbf{I})^{-1}\mathbf{Q}^H\mathbf{Q}\right\} \mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0$ , where  $\mathbf{w}_0$  is any arbitrary feasible point of the problem (6.5).*

**Proof:** See Appendix D.

Using Lemma 6.2, the problem (6.5) can be equivalently stated as

$$\begin{aligned} \min_{\theta_1 \leq \alpha \leq \theta_2} \quad & \overbrace{\min_{\mathbf{w}} \mathbf{w}^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}}^{\text{Inner Problem}} \\ \text{subject to} \quad & \mathbf{w}^H\mathbf{Q}^H\mathbf{Q}\mathbf{w} = \alpha, \\ & \mathbf{w}^H\mathbf{w} \leq \frac{(\sqrt{\alpha}-1)^2}{\eta^2} \end{aligned} \quad (6.11)$$

where

$$\theta_1 = \frac{1}{\left(1 - \frac{\eta}{\sqrt{\lambda_{\max}\{\mathbf{Q}^H\mathbf{Q}\}}}\right)^2} \quad (6.12)$$

and

$$\theta_2 = \lambda_{\max}\left\{(\hat{\mathbf{R}} + \gamma\mathbf{I})^{-1}\mathbf{Q}^H\mathbf{Q}\right\} \mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0. \quad (6.13)$$

For a fixed value of  $\alpha$ , the inner optimization problem in (6.11) is non-convex with respect to  $\mathbf{w}$ . Based on the inner optimization problem in (6.11) when  $\alpha$  is fixed, we define the following optimal value function

$$h(\alpha) \triangleq \left\{ \min_{\mathbf{w}} \mathbf{w}^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w} \mid \mathbf{w}^H\mathbf{Q}^H\mathbf{Q}\mathbf{w} = \alpha, \mathbf{w}^H\mathbf{w} \leq \frac{(\sqrt{\alpha}-1)^2}{\eta^2} \right\}, \quad \theta_1 \leq \alpha \leq \theta_2. \quad (6.14)$$

Using the optimal value function (6.14), the problem (6.11) can be equivalently expressed as

$$\min_{\alpha} \quad h(\alpha) \quad \text{subject to} \quad \theta_1 \leq \alpha \leq \theta_2. \quad (6.15)$$

The corresponding optimization problem of  $h(\alpha)$  for a fixed value of  $\alpha$  is non-convex. In what follows, we aim at replacing  $h(\alpha)$  with an equivalent optimal value function whose corresponding optimization problem is convex.

Introducing the matrix  $\mathbf{W} \triangleq \mathbf{w}\mathbf{w}^H$  and using the fact that for any arbitrary matrix  $\mathbf{A}$ ,  $\mathbf{w}^H \mathbf{A} \mathbf{w} = \text{tr}\{\mathbf{A}\mathbf{w}\mathbf{w}^H\}$ , the optimal value function (6.14) can be equivalently recast as

$$h(\alpha) = \left\{ \min_{\mathbf{W}} \text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}\} \mid \text{tr}\{\mathbf{Q}^H \mathbf{Q} \mathbf{W}\} = \alpha, \right. \\ \left. \text{tr}\{\mathbf{W}\} \leq \frac{(\sqrt{\alpha}-1)^2}{\eta^2}, \mathbf{W} \succeq \mathbf{0}, \text{rank}\{\mathbf{W}\} = 1 \right\}, \theta_1 \leq \alpha \leq \theta_2 \quad (6.16)$$

By dropping the rank-one constraint in the corresponding optimization problem of  $h(\alpha)$  when  $\alpha$  is fixed, ( $\theta_1 \leq \alpha \leq \theta_2$ ), a new optimal value function denoted as  $k(\alpha)$  can be defined as

$$k(\alpha) \triangleq \left\{ \min_{\mathbf{W}} \text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}\} \mid \text{tr}\{\mathbf{Q}^H \mathbf{Q} \mathbf{W}\} = \alpha, \right. \\ \left. \text{tr}\{\mathbf{W}\} \leq \frac{(\sqrt{\alpha}-1)^2}{\eta^2}, \mathbf{W} \succeq \mathbf{0} \right\}, \theta_1 \leq \alpha \leq \theta_2. \quad (6.17)$$

For brevity, we will refer to the optimization problems that correspond to the optimal value functions  $h(\alpha)$  and  $k(\alpha)$  when  $\alpha$  is fixed, as the optimization problems of  $h(\alpha)$  and  $k(\alpha)$ , respectively. Although the corresponding optimization problem of  $h(\alpha)$  and  $k(\alpha)$  have different convexity properties, they are indeed equivalent, i.e.,  $h(\alpha) = k(\alpha)$  for any  $\alpha \in [\theta_1, \theta_2]$  according to the Lemma 3.1 and Theorem 3.1. Furthermore, based on the optimal solution of the optimization problem of  $k(\alpha)$  when  $\alpha$  is fixed, the optimal solution of the optimization problem of  $h(\alpha)$  can be constructed.

Using the equivalence between the aforementioned optimal value functions, the original problem (6.15) can be expressed as

$$\min_{\alpha} \quad k(\alpha) \quad \text{subject to} \quad \theta_1 \leq \alpha \leq \theta_2. \quad (6.18)$$

It is noteworthy to mention that based on the optimal solution of (6.18) denoted as  $\alpha_{\text{opt}}$ , we can easily obtain the optimal solution of the original problem (6.15) or, equivalently, the optimal solution of the problem (6.11). Specifically, since the optimal value functions  $h(\alpha)$  and  $k(\alpha)$  are equivalent,  $\alpha_{\text{opt}}$  is also the optimal solution

of the problem (6.15) and, thus, also the problem (6.11). Moreover, the optimization problem of  $k(\alpha_{\text{opt}})$  is convex and can be easily solved. In addition, using the results in Theorem 3.1, based on the optimal solution of the optimization problem of  $k(\alpha_{\text{opt}})$ , the optimal solution of the optimization problem of  $h(\alpha_{\text{opt}})$  can be constructed. Therefore, in the rest of the chapter, we concentrate on the problem (6.18).

Since for every fixed value of  $\alpha$ , the corresponding optimization problem of  $k(\alpha)$  is a convex SDP problem, one possible approach for solving (6.18) is based on exhaustive search over  $\alpha$ . In other words,  $\alpha$  can be found by using an exhaustive search over a fine grid on the interval of  $[\theta_1, \theta_2]$ . Although this search method is inefficient, it can be used as a benchmark.

Using the definition of the optimal value function  $k(\alpha)$ , the problem (6.18) can be equivalently expressed as

$$\begin{aligned} \min_{\mathbf{W}, \alpha} \quad & \text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \right\} \\ \text{subject to} \quad & \text{tr} \{ \mathbf{Q}^H \mathbf{Q} \mathbf{W} \} = \alpha \\ & \eta^2 \text{tr} \{ \mathbf{W} \} \leq (\sqrt{\alpha} - 1)^2 \\ & \mathbf{W} \succeq 0, \theta_1 \leq \alpha \leq \theta_2. \end{aligned} \tag{6.19}$$

Note that replacing  $h(\alpha)$  by  $k(\alpha)$  results in a much simpler problem. Indeed, compared to the original problem (6.11), in which the first constraint is non-convex, the corresponding first constraint of (6.19) is convex. All the constraints and the objective function of the problem (6.19) are convex except for the constraint  $\text{tr} \{ \mathbf{W} \} \leq (\sqrt{\alpha} - 1)^2 / \eta^2$  which is non-convex only in a single variable  $\alpha$  and which makes the problem (6.19) non-convex overall. This single non-convex constraint can be rewritten equivalently as  $\eta^2 \text{tr} \{ \mathbf{W} \} - (\alpha + 1) + 2\sqrt{\alpha} \leq 0$  where all the terms are linear with respect to  $\mathbf{W}$  and  $\alpha$  except for the concave term of  $\sqrt{\alpha}$ . The latter constraint can be handled iteratively by building a POTDC-type algorithm (see Section 3.1.1) based on the iterative linear approximation of the non-convex term  $\sqrt{\alpha}$  around suitably selected points. It is interesting to mention that this iterative linear approximation can be also interpreted in terms of DC iteration over a single non-convex term  $\sqrt{\alpha}$ . The fact that iterations are needed only over a single variable helps to reduce dramatically the number of iterations of the algorithm and allows for very simple interpretations shown below.

### 6.2.1 Iterative POTDC algorithm

Let us consider the optimization problem (6.19) and replace the term  $\sqrt{\alpha}$  by its linear approximation around  $\alpha_c$ , i.e.,  $\sqrt{\alpha} \approx \sqrt{\alpha_c} + (\alpha - \alpha_c)/(2\sqrt{\alpha_c})$ . It leads to the following SDP problem

$$\begin{aligned} \min_{\mathbf{W}, \alpha} \quad & \text{tr} \{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \} \\ \text{subject to} \quad & \text{tr} \{ \mathbf{Q}^H \mathbf{Q} \mathbf{W} \} = \alpha \\ & \eta^2 \text{tr} \{ \mathbf{W} \} + (\sqrt{\alpha_c} - 1) + \alpha \left( \frac{1}{\sqrt{\alpha_c}} - 1 \right) \leq 0 \\ & \mathbf{W} \succeq 0, \quad \theta_1 \leq \alpha \leq \theta_2. \end{aligned} \quad (6.20)$$

To understand the POTDC algorithm intuitively and also to see how the linearization points are selected in different iterations, let us define the following optimal value function based on the optimization problem (6.20)

$$\begin{aligned} l(\alpha, \alpha_c) \triangleq \left\{ \min_{\mathbf{W}} \text{tr} \{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \} \mid \text{tr} \{ \mathbf{Q}^H \mathbf{Q} \mathbf{W} \} = \alpha, \right. \\ \left. \eta^2 \text{tr} \{ \mathbf{W} \} + (\sqrt{\alpha_c} - 1) + \alpha \left( \frac{1}{\sqrt{\alpha_c}} - 1 \right) \leq 0, \mathbf{W} \succeq \mathbf{0} \right\}, \quad \theta_1 \leq \alpha \leq \theta_2 \end{aligned} \quad (6.21)$$

where  $\alpha_c$  in  $l(\alpha, \alpha_c)$  denotes the linearization point. The optimal value function  $l(\alpha, \alpha_c)$  can be also obtained through  $k(\alpha)$  in (6.17) by replacing the term  $\sqrt{\alpha}$  in  $\eta^2 \text{tr} \{ \mathbf{W} \} - (\alpha + 1) + 2\sqrt{\alpha} \leq 0$  with its linear approximation around  $\alpha_c$ . Since  $\sqrt{\alpha}$  and its linear approximation have the same values at  $\alpha_c$ ,  $l(\alpha, \alpha_c)$  and  $k(\alpha)$  take the same values at this point. The following lemma establishes the relationship between the optimal value functions  $k(\alpha)$  and  $l(\alpha, \alpha_c)$ .

**Lemma 6.3.** *The optimal value function  $l(\alpha, \alpha_c)$  is a convex upper-bound of  $k(\alpha)$  for any arbitrary  $\alpha_c \in [\theta_1, \theta_2]$ , i.e.,  $l(\alpha, \alpha_c) \geq k(\alpha)$ ,  $\forall \alpha \in [\theta_1, \theta_2]$  and  $l(\alpha, \alpha_c)$  is convex with respect to  $\alpha$ . Furthermore, the values of the optimal value functions  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  as well as their right and left derivatives are equal at the point  $\alpha = \alpha_c$ . In other words, under the condition that  $k(\alpha)$  is differentiable at  $\alpha_c$ ,  $l(\alpha, \alpha_c)$  is tangent to  $k(\alpha)$  at this point.*

**Proof:** See Appendix E.

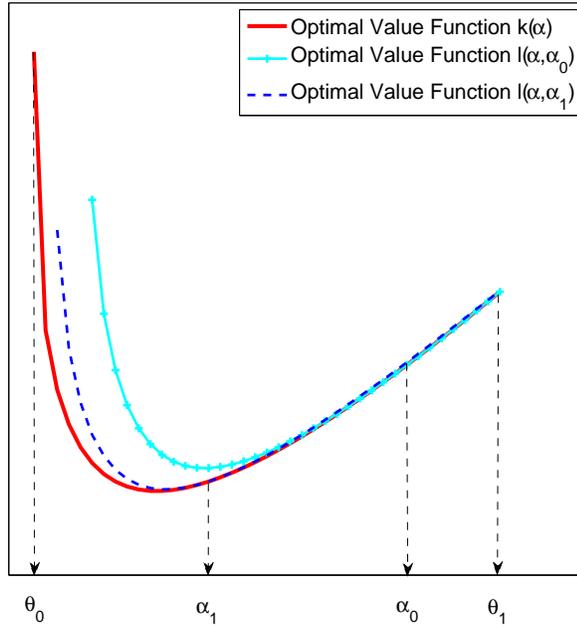
In what follows, for the sake of clarity of the explanations, it is assumed that the function  $k(\alpha)$  is differentiable over the interval of  $(\theta_1, \theta_2)$ , however, this property is not generally required as we will see later. Let us consider an arbitrary point,

denoted as  $\alpha_0$ ,  $\alpha_0 \in (\theta_1, \theta_2)$  as the initial linearization point, i.e.,  $\alpha_c = \alpha_0$ . Based on Lemma 6.3,  $l(\alpha, \alpha_0)$  is a convex function with respect to  $\alpha$  which is tangent to  $k(\alpha)$  at the linearization point  $\alpha = \alpha_0$ , and it is also an upper-bound to  $k(\alpha)$ . Let  $\alpha_1$  denote the global minimizer of  $l(\alpha, \alpha_0)$  that can be easily obtained due to the convexity of  $l(\alpha, \alpha_0)$  with polynomial time complexity.

Since  $l(\alpha, \alpha_0)$  is tangent to  $k(\alpha)$  at  $\alpha = \alpha_0$  and it is also an upper-bound for  $k(\alpha)$ , it can be concluded that  $\alpha_1$  is a descend point for  $k(\alpha)$ , i.e.,  $k(\alpha_1) \leq k(\alpha_0)$  as it is shown in Fig. 6.1. Specifically, the fact that  $l(\alpha, \alpha_0)$  is tangent to  $k(\alpha)$  at  $\alpha = \alpha_0$  and  $\alpha_1$  is the global minimizer of  $l(\alpha, \alpha_0)$  implies that

$$l(\alpha_1, \alpha_0) \leq l(\alpha_0, \alpha_0) = k(\alpha_0). \quad (6.22)$$

Furthermore, since  $l(\alpha, \alpha_0)$  is an upper-bound for  $k(\alpha)$ , it can be found that  $k(\alpha_1) \leq l(\alpha_1, \alpha_0)$ . Due to the later fact and also the equation (6.22), it is concluded that  $k(\alpha_1) \leq k(\alpha_0)$ .



**Figure 6.1:** Iterative method for minimizing the optimal value function  $k(\alpha)$ . The convex optimal value function  $l(\alpha, \alpha_0)$  is an upper-bound to  $k(\alpha)$  which is tangent to it at  $\alpha = \alpha_0$ , and its minimum is denoted as  $\alpha_1$ . The point  $\alpha_1$  is used to establish another convex upper-bound function denoted as  $l(\alpha, \alpha_1)$  and this process continues.

Choosing  $\alpha_1$  as the linearization point in the second iteration, and finding the global minimizer of  $l(\alpha, \alpha_1)$  over the interval  $[\theta_1, \theta_2]$  denoted as  $\alpha_2$ , another descend

point can be obtained, i.e.,  $k(\alpha_2) \leq k(\alpha_1)$ . This process can be continued until convergence.

Then the proposed iterative descend method can be described as shown in Algorithm 6.1. Since the optimization problem (6.5) is an special case of the problem (3.1) with two only quadratic functions in the constraints with a non-negative objective function, the corresponding results holds true. Indeed, according to Lemma 3.2, the Algorithm 6.1 converges to a KKT point, i.e., a point which satisfies the KKT optimality conditions. In our simulations, we choose the stopping criteria in Algorithm 6.1 based on the difference between two consecutive optimal values, however, the stopping criteria can be also defined based on the approximate satisfaction of the KKT optimality conditions (see Subsection 3.1.1 ).

---

**Algorithm 6.1** The iterative POTDC algorithm

---

**Require:** An arbitrary  $\alpha_c \in [\theta_1, \theta_2]$ ,  
the progress parameter  $\zeta$ ,  
set  $i$  equal to 1.

**repeat**

Solve the following optimization problem using  $\alpha_c$  to obtain  $\mathbf{W}_{\text{opt}}$  and  $\alpha_{\text{opt}}$

$$\begin{aligned} \min_{\mathbf{W}, \alpha} \quad & \text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \right\} \\ \text{subject to} \quad & \text{tr} \{ \mathbf{Q}^H \mathbf{Q} \mathbf{W} \} = \alpha \\ & \eta^2 \text{tr} \{ \mathbf{W} \} + (\sqrt{\alpha_c} - 1) + \alpha \left( \frac{1}{\sqrt{\alpha_c}} - 1 \right) \leq 0 \\ & \mathbf{W} \succeq 0, \quad \theta_1 \leq \alpha \leq \theta_2 \end{aligned}$$

and set

$$\begin{aligned} \mathbf{W}_{\text{opt}, i} &\leftarrow \mathbf{W}_{\text{opt}}, & \alpha_{\text{opt}, i} &\leftarrow \alpha_{\text{opt}} \\ \alpha_c &\leftarrow \alpha_{\text{opt}}, & i &\leftarrow i + 1 \end{aligned}$$

**until**

$$\text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W}_{\text{opt}, i-1} \right\} - \text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W}_{\text{opt}, i} \right\} \leq \zeta \text{ for } i \geq 2 .$$


---

Moreover, the point obtained by Algorithm 6.1 is guaranteed to be the global optimum of the problem considered if the optimal value function  $k(\alpha)$  is a convex function of  $\alpha$ . The convexity of  $k(\alpha)$  under certain conditions is established in the following theorem.

**Theorem 6.1.** *Let the covariance matrices of the desired and interference sources have, respectively, the following structures  $\mathbf{R}_s = \mathbf{U}_1 \mathbf{V}_s \mathbf{U}_1^H$  and  $\mathbf{R}_i = \mathbf{U}_2 \mathbf{V}_i \mathbf{U}_2^H$ ,  $i = 1, \dots, N$ , where  $\mathbf{U}_1$  and  $\mathbf{U}_2$  span orthogonal subspaces, and  $\mathbf{V}_s$  and  $\mathbf{V}_i$ ,  $i = 1, \dots, N$  are some positive semi-definite matrices. Then the optimal value function  $k(\alpha)$  is*

convex assuming that the received signal covariance matrix is known perfectly.

**Proof:** See Appendix F.

The conditions of Theorem 6.1 are somewhat idealistic in real applications since the received signal covariance matrix is not precisely known in reality and the subspaces spanned by  $\mathbf{U}_1$  and  $\mathbf{U}_2$  may overlap in general. However, there have been developed various covariance matrix estimation techniques, which improve the estimation of  $\mathbf{R}$  even for the small sample size case [125]–[127]. Moreover, under the condition that the desired and interfering sources are locally incoherently scattered and the desired source is well-separated from the interfering sources, i.e., the angular power density of the interfering sources equals zero over the angles in which the desired source is spread, it is guaranteed that  $\mathbf{U}_1$  and  $\mathbf{U}_2$  span orthogonal subspaces. The later is not uncommon in practical applications or can be achieved by using well known de-correlation techniques for the desired and interference sources such as forward-backward averaging and/or spatial smoothing.

Let us show for completeness that if the desired and interfering sources are locally incoherently scattered and the desired source is well-separated from the interfering sources, then  $\mathbf{U}_1$  and  $\mathbf{U}_2$  span orthogonal subspaces. The covariance matrices of the desired and interference sources can be expressed, respectively, as

$$\mathbf{R}_s = \sigma_s^2 \int_{-\pi/2}^{\pi/2} \zeta_s(\theta) \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta \quad (6.23)$$

and

$$\mathbf{R}_i = \sigma_i^2 \int_{-\pi/2}^{\pi/2} \zeta_i(\theta) \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta, \quad i = 1, \dots, N \quad (6.24)$$

where  $\zeta_s(\theta)$  and  $\zeta_i(\theta)$  denote the normalized angular power densities of the desired source and the  $i$ th interference, and  $\sigma_s^2$  and  $\sigma_i^2$  are the desired source and  $i$ th interference source powers, respectively. Let  $\Theta_s$  denote the angular sector in which the desired source is spread. Then the condition that the desired source is well-separated from other interfering sources can be mathematically expressed as  $\zeta_i(\theta) = 0, \theta \in \Theta_s$ . Defining the matrix  $\mathbf{C} \triangleq \int_{\tilde{\Theta}_s} \mathbf{a}(\theta) \mathbf{a}^H(\theta) d\theta$  where  $\tilde{\Theta}_s$  denotes the complement of  $\Theta_s$ , it was shown in Section 5.2 that for the properly chosen  $K$ , the steering vector in  $\Theta_s$  and its complement  $\tilde{\Theta}_s$  can be approximately expressed as the linear combination of the column of  $\mathbf{U}_2$  and  $\mathbf{U}_1$ , respectively, that is,

$$\mathbf{d}(\theta) \cong \mathbf{U}_2 \mathbf{v}_2(\theta), \quad \theta \in \tilde{\Theta}_s \quad (6.25)$$

$$\mathbf{d}(\theta) \cong \mathbf{U}_1 \mathbf{v}_1(\theta), \quad \theta \in \Theta_s \quad (6.26)$$

where  $\mathbf{U}_2$  and  $\mathbf{U}_1$  denote the corresponding eigenvectors of the  $K$  largest eigenvalues and the rest of the eigenvalues of the matrix  $\mathbf{C}$ , respectively, and  $\mathbf{v}_1(\theta)$  and  $\mathbf{v}_2(\theta)$  are some coefficient vectors. Based on (6.25) and (6.26), the covariance matrices of the desired and the inference sources have the structure used in Theorem 6.1 and  $\mathbf{U}_1$  and  $\mathbf{U}_2$  are approximately orthogonal.

It is also worth noting that even a more relaxed property of the optimal value function  $k(\alpha)$  would be sufficient to guarantee global optimality for the POTDC algorithm. Specifically, if  $k(\alpha)$  defined in (6.17) is a strictly quasi-convex function of  $\alpha \in [\theta_1, \theta_2]$ , then the point found by the POTDC algorithm is still guaranteed to be the global optimum of the optimization problem (6.4) [128]. Thus, the conditions of Theorem 6.1 can be possibly further relaxed.

The worst-case computational complexity of a general SDP problem can be expressed as  $\mathcal{O}(n_c^2 n_v^{2.5} + n_c n_v^{3.5})$  where  $n_c$  and  $n_v$  denote, respectively, the number of constraints and the number of variables of the primal problem (2.18) [129] and  $\mathcal{O}(\cdot)$  stands for the big-O (the highest order of complexity). Therefore, the computational complexity of Algorithm 6.1 is equal to that of the SDP optimization problem in Algorithm 6.1, that is,  $\mathcal{O}((M^2 + 1)^{3.5})$  times the number of iterations (see also Simulation Example 1 in Section 6.3). The RAB algorithm of [10] is iterative as well and its computational complexity is equal to  $\mathcal{O}(M^{2 \times 3.5})$  times the number of iterations. The complexity of the RABs of [8] and [122] is  $\mathcal{O}(M^3)$ . The comparison of the overall complexity of the proposed POTDC algorithm with that of the DC iteration-based method will be explicitly performed in Simulation Example 4. Although the computational complexity of the new proposed method may be slightly higher than that of the other RABs, it finds the global optimum and results in superior performance as it is shown in the next section.

### 6.2.2 Lower-bounds for the optimal value

We also aim at developing a tight lower-bound for the optimal value of the optimization problem (6.19). Such lower-bound can be used for assessing the performance of the proposed iterative algorithm.

As it was mentioned earlier, although the objective function of the optimization problem (6.19) is convex, its feasible set is non-convex due to the second constraint in problem (6.19). A lower-bound for the optimal value of the optimization problem (6.19) can be achieved by replacing the second constraint in (6.19) by its correspond-

ing convex-hull. However, such lower-bound may not be tight. In order to obtain a tight lower-bound, we can divide the sector  $[\theta_1, \theta_2]$  into  $N$  subsectors and solve the optimization problem (6.19) over each subsector in which the second constraint of (6.19) has been replaced with the corresponding convex hull. The minimum of the optimal values of such optimization problem over the subsectors is the lower-bound for the problem (6.19). It is obvious that by increasing  $N$ , the lower-bound becomes tighter.

### 6.3 Simulation results

Let us consider a ULA of 10 omni-directional antenna elements with the inter-element spacing of half wavelength. Additive noise in antenna elements is modeled as spatially and temporally independent complex Gaussian noise with zero mean and unit variance. Throughout all simulation examples, it is assumed that in addition to the desired source, an interference source with the INR of 30 dB impinges on the antenna array. For obtaining each point in the simulation examples, 100 independent runs are used unless otherwise is specified and the sample data covariance matrix is estimated using  $K = 50$  snapshots.

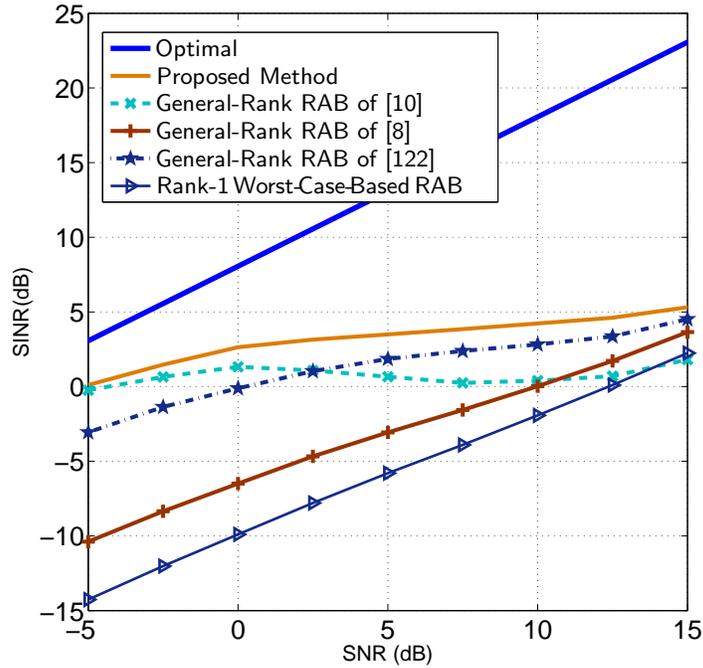
The new proposed method is compared in terms of the output SINR to the general-rank RAB methods of [8], [10], [122] and to the rank-one worst-case-based RAB of [7]. Moreover, the proposed method and the aforementioned general-rank RAB methods are also compared in terms of the achieved values for the objective function of the problem (6.4). The diagonal loading parameters of  $\gamma = 10$  and  $\eta = 0.5\sqrt{\text{tr}\{\mathbf{R}_s\}}$  are chosen for the proposed RAB and the RAB methods of [10] and [122], and the parameters of  $\gamma = 10$  and  $\epsilon = 8\sigma_s^2$  are chosen for the RAB of [8]. The initial point  $\alpha_0$  in the first iteration of the proposed method equals to  $(\theta_1 + \theta_2)/2$  unless otherwise is specified. The progress parameter  $\zeta$  for the proposed method is chosen to be equal to  $10^{-6}$ .

#### 6.3.1 Example 1 : Parameter mismatch

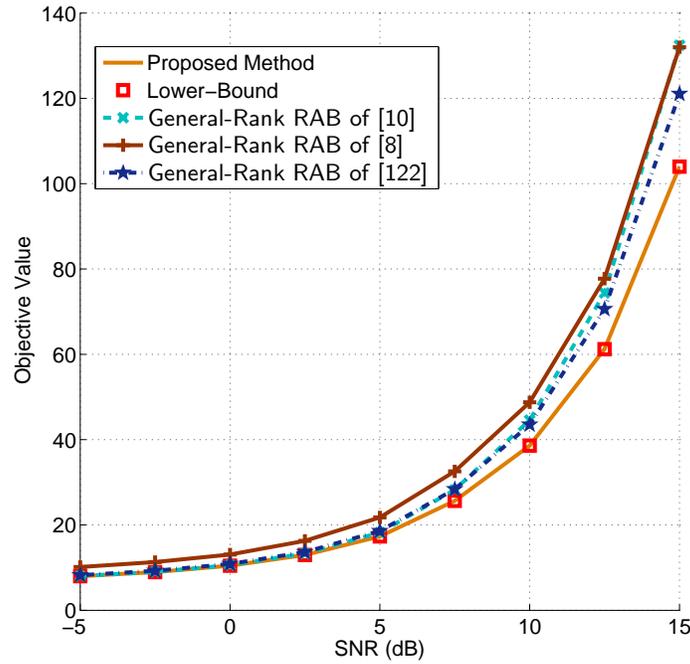
In this example, the desired and interference sources are locally incoherently scattered with Gaussian and uniform angular power densities with central angles of  $30^\circ$  and  $10^\circ$ , respectively. The angular spreads of the desired and the interfering sources are assumed to be  $4^\circ$  and  $10^\circ$ , respectively. The presumed knowledge of the desired source is different from the actual one and is characterized by an incoherently

scattered source with Gaussian angular power density whose central angle and angular spread are  $34^\circ$  and  $6^\circ$ , respectively. Note that, the presumed knowledge about the shape of the angular power density of the desired source is correct while the presumed central angle and angular spread deviate from the actual one.

In Figs. 6.2 and 6.3, the output SINR and the objective function values of the problem (6.4), respectively, are plotted versus SNR. It can be observed from the figures that the new proposed method based on the POTDC algorithm has superior performance over the other RABs. Although the method of [10] does not have a guaranteed convergence, it results in a better average performance as compared to the methods of [8] and [122] at low SNR values while the method of [122] outperforms [10] at higher SNRs. Moreover, the Fig. 6.3 confirms that the new proposed method achieves the global minimum of the optimization problem (6.4) since the corresponding objective value coincides with the lower-bound on the objective function of the problem (6.4). Fig. 6.4 shows the convergence of the iterative POTDC method. It shows the average of the optimal value found by the algorithm over iterations for SNR=15 dB. It can be observed that the proposed algorithm converges to the global optimum in about 4 iterations.



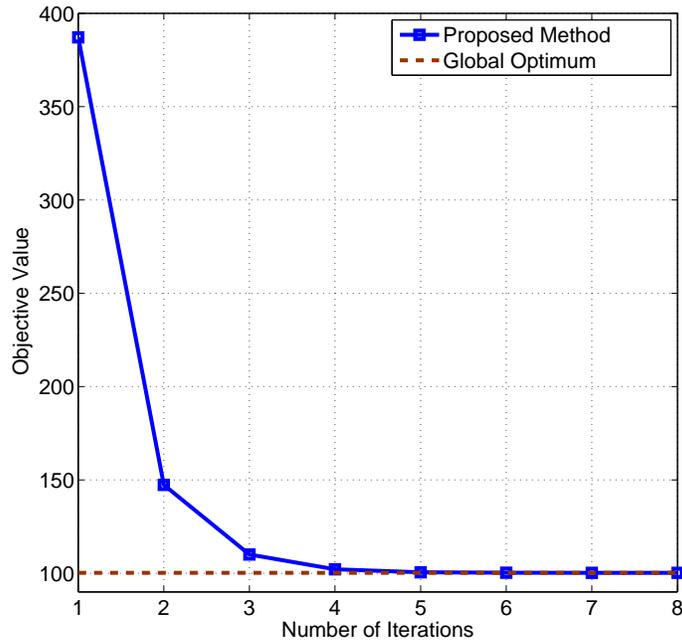
**Figure 6.2:** Example 1: Output SINR versus SNR, INR=30 dB and  $K = 50$ .



**Figure 6.3:** Example 1: Objective function value of the problem (6.4) versus SNR, INR=30 dB and  $K = 50$ .

### 6.3.2 Example 2 : Effect of the rank of desired signal covariance matrix

In the second example, we study how the rank of the actual correlation matrix of the desired source  $\mathbf{R}_s$  affects the performance of the proposed general-rank RAB and other methods tested. The same simulation set up as in the previous example is considered. The only difference is that the actual angular spread of the desired source varies and so does the actual rank of the desired source covariance matrix. The angular spread of the desired user is chosen to be  $1^\circ$ ,  $2^\circ$ ,  $5^\circ$ ,  $9^\circ$ , and  $14^\circ$ . Figs. 6.5 and 6.6 show, respectively, the output SINR and the objective function values of the problem (6.4) versus the rank of the actual correlation matrix of the desired source when SNR=10 dB. It can be seen from the figures that the proposed method outperforms the methods tested in all rank in terms of the objective value of the optimization problem (6.4) and it achieves the globally optimal solution as it coincides with the lower-bound.



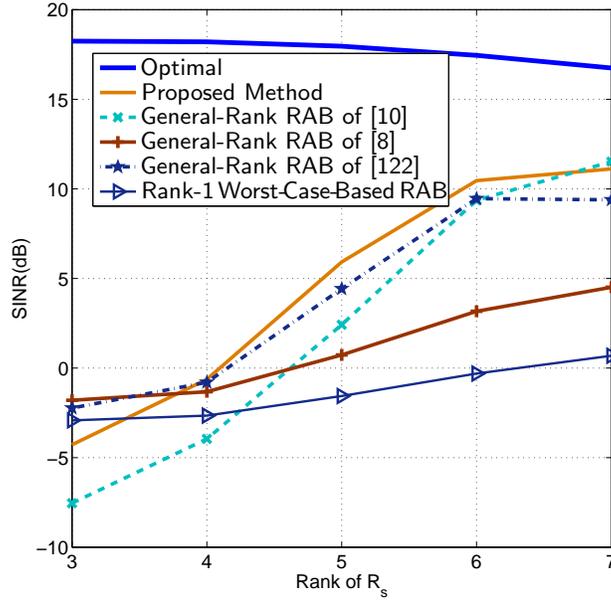
**Figure 6.4:** Example 1: Objective function value of the problem (6.4) versus the number of iterations, SNR=15 dB, INR=30 dB and  $K = 50$ .

### 6.3.3 Example 3 : Distribution mismatch

In this example, we also consider the locally incoherently scattered desired and interference sources. However, compared to the previous example, there is a substantial error in the knowledge of the desired source angular power density.

The interference source is modeled as in the previous example, while the angular power density of the desired source is assumed to be a truncated Laplacian function distorted by severe fluctuations. The central angle and the scale parameter of the Laplacian distribution is assumed to be  $30^\circ$  and 0.1, respectively, and it is assumed to be equal to zero outside of the interval  $[15^\circ, 45^\circ]$  as it has been shown in Fig. 6.7. The presumed knowledge of the desired source is different from the actual one and is characterized by an incoherently scattered source with Gaussian angular power density whose central angle and angular spread are  $32^\circ$  and  $6^\circ$ , respectively.

Fig. 6.8 depicts the corresponding output SINR of the problem (6.4) obtained by the beamforming methods tested versus SNR. It can be concluded from the figure that the proposed method has superior performance over the other methods.



**Figure 6.5:** Example 2: Output SINR versus the actual rank of  $\mathbf{R}_s$ , SNR=10 dB, INR=30 dB and  $K = 50$ .

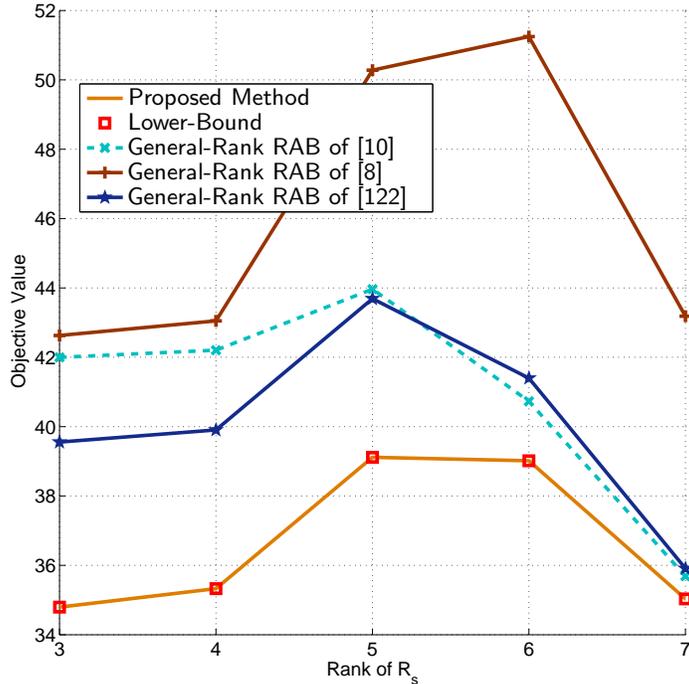
### 6.3.4 Example 4 : Complexity Comparison

In this example, we compare the efficiency of the proposed POTDC method to that of the DC iteration-based method that can be written for the problem under consideration as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w} \\ \text{subject to} \quad & f(\mathbf{w}^{(k)}) + \langle \nabla f(\mathbf{w}^{(k)}), \mathbf{w} - \mathbf{w}^{(k)} \rangle - \eta \|\mathbf{w}\| \geq 1 \end{aligned} \quad (6.27)$$

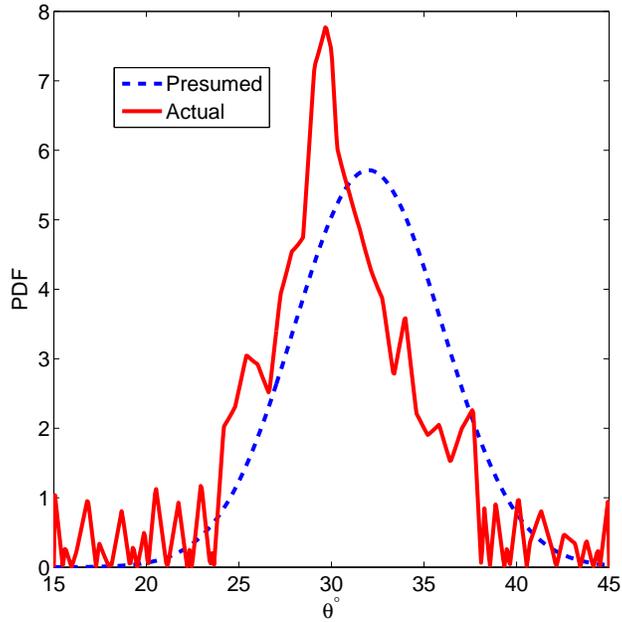
where  $\langle \cdot, \cdot \rangle$  denotes the inner product and the function  $f(\mathbf{w}) = \|\mathbf{Q}\mathbf{w}\|$  is replaced with the first two terms of the Taylor expansion of  $f(\mathbf{w})$  around  $\mathbf{w}^{(k)}$ . At the first iteration  $\mathbf{w}^{(0)}$  is initialized and in the next iterations  $\mathbf{w}^{(k)}$  is selected as the optimal  $\mathbf{w}$  obtained from solving (6.27) in the previous iteration. Thus, the iteration are performed over the whole vector of variables of the problem.

The simulation set up is the same as in our Simulation Example 1 except that different number of antennas are used. For a fair comparison, the initial point  $\alpha_0$  in the proposed POTDC method and  $\mathbf{w}^{(0)}$  in (6.27) are chosen randomly. Particularly, the initialization point for the proposed POTDC method is chosen uniformly over the interval  $[\theta_1, \theta_2]$  while the imaginary and real parts of the initial vector  $\mathbf{w}^{(0)}$



**Figure 6.6:** Example 2: Objective function value of the problem (6.4) versus the actual rank of  $\mathbf{R}_s$ , SNR=10 dB, INR=30 dB and  $K = 50$ .

in (6.27) are chosen independently as zero mean, unit variance, Gaussian random variables. If the so-generated  $\mathbf{w}^{(0)}$  is not feasible, another initialization point is generated and this process continues until a feasible point is resulted. Note that the time which is consumed during the generation of a feasible point is negligible and it has not been considered in the average CPU time comparison. Table 6.1 shows the average number of the iterations for the aforementioned methods versus the size of the antenna array. The accuracy is set to  $10^{-6}$ , SNR= -5 dB, and each number in the table is obtained by averaging the results from 200 runs. From this table, it can be seen that the number of the iterations for the POTDC method is almost fixed while it increases for the DC-iteration method as the size of the array increases. The latter phenomenon can be justified by considering the DC iteration-type interpretation of the POTDC method over the one-dimensional function of  $k(\alpha)$ . The dimension of  $k(\alpha)$  is independent of the size of the array (thus, the size of the optimization problem), while the size of search space for the DC iteration-based method (6.27), that is,  $2M$ , increases as the size of the array increases. The



**Figure 6.7:** Example 3: Actual and presumed angular power densities of general-rank source.

**Table 6.1:** Average number of the iterations

Array size	8	10	12	14	16	18	20
POTDC	3.29	3.24	3.03	3.12	3.20	3.28	3.11
DC iteration	4.87	5.82	5.98	6.82	7.62	8.24	8.95

average (over 200 runs) CPU time for the aforementioned methods is also shown in Table 2. Both methods have been implemented in Matlab using CVX software and run on the same Notebook with AMD A8-4500m APU CPU 1.90 GHz. Table 6.2

**Table 6.2:** Average CPU time

Array size	8	10	12	14	16	18	20
POTDC	0.591	0.582	0.578	0.682	0.838	1.076	1.183
DC iteration	1.880	2.243	2.344	2.752	3.323	4.012	4.762

confirms that the proposed method is more efficient than the DC iteration-based one in terms of the time which is spent for solving the same problem. Note that although the number of variables in the matrix  $\mathbf{W}$  of the optimization problem (6.20) is in general  $M^2 + 1$  (since  $\mathbf{W}$  has to be a Hermitian matrix, after the rank-one constraint is relaxed), the probability that the optimal  $\mathbf{W}$  is rank-one has been

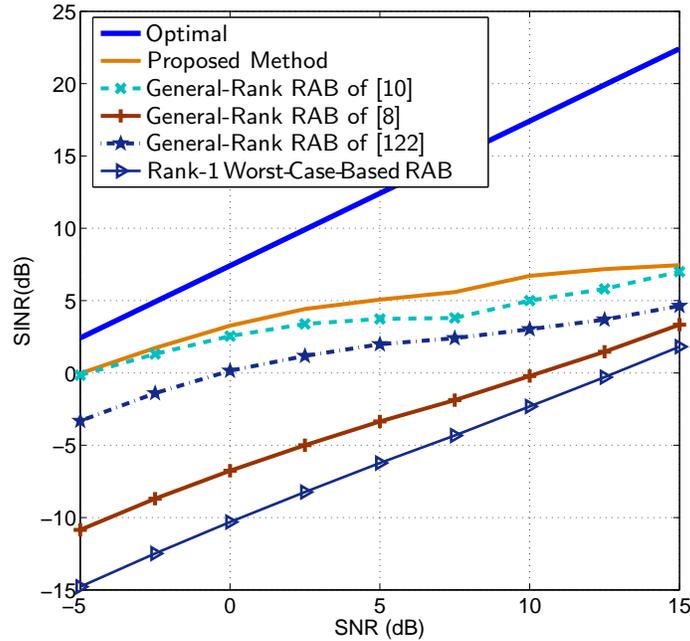


Figure 6.8: Example 3: Output SINR versus SNR, INR=30 dB and  $K = 50$ .

shown to be very high [14], [22], [130], and [131]. It is also approved by these simulations. Thus, in almost all cases, for different data sets, the actual dimension of the problem (6.20) is  $2M + 1$ . As a result, the average complexity of solving (6.20) is significantly smaller than the worst-case complexity. The latter is due to the fact that the worst-case computational complexity of a general SDP problem is a function of the number of the variables and also the number of constraints in the primal problem (see Subsection 6.2.1).

## 6.4 Chapter summary

We have considered the RAB problem for general-rank signal model with additional positive semi-definite constraint. Such RAB problem corresponds to the class of generalized QCQP problems which can be precisely recast as a non-convex DC optimization problem. We have studied this non-convex DC problem and designed the POTDC-type algorithm for solving it. It has been proved that the point found by the POTDC algorithm for RAB for general-rank signal model with positive semi-definite constraint is a KKT point. Moreover, the problem considered can

be solved globally optimally under certain conditions. Specifically, we have proved a number of results that lead us to the equivalence between the claim of global optimality and convexity of the one-dimensional optimal value function (6.17). The latter convexity has been then proved under the conditions of Theorem 6.1. The resulted RAB method shows superior performance compared to the other existing methods in terms of the output SINR. It also has complexity that is guaranteed to be polynomial.

## Chapter 7

# Two-Way AF MIMO Relay Amplification Matrix Design

TWR has been a very attractive research topic recently due to its ability to overcome the rate loss associated with the conventional one-way relaying systems. Moreover, most of communications are actually two-way communications. Thus, TWR has a wide range of applications. As it was explained in Section 2.4, the spectral efficiency of the TWR can be attributed to the relaxation of the requirement of ‘orthogonal/non-interfering’ transmissions between the terminals and the relay.

One fundamental problem associated with TWR systems is the relay transmit strategy design based on the available CSI [61]–[69]. It is usually designed in such a way that a particular performance criterion is optimized, while keeping some constraints on the available resources. The main difficulty of the TWR transmit strategy design is the non-convex nature of its corresponding optimization problem. In this chapter, we consider the AF TWR system with two terminals equipped with a single antenna and one relay with multiple antennas. Note that the relay transmit strategy design of an AF relay is equivalent to the optimal design of its amplification matrix. We study the optimal amplification matrix design when the maximum sum-rate, PF and the MMRF are used as the design criteria. This is a basic model which can be extended in many ways. The significant advantage of considering this basic model is that the corresponding achievable rate is analyzed in the existing literature [62]. It enables us to concentrate on the mathematical issues of the corresponding optimization problem which are of significant and ubiquitous interest. We first concentrate on the design of the relay transmit strategy that maximizes the sum-rate of both terminals when there is a constraint on the total relay transmit power.

It is shown that the optimization problem of finding the relay amplification matrix for the sum-rate maximization in the considered AF two-way relaying system is equivalent to finding the maximum of the product of quadratic fractional functions under a quadratic power constraint on the available power at the relay. Such a problem is a specific realization of the generalized QCQP problem which was introduced in Chapter 3, and therefore, it can be precisely recast as a DC optimization problem. The typical approach for solving DC problems is the use of various modifications of the branch-and-bound method [13], and [31]–[36] that is an effective global optimization method. However as it was mentioned earlier, it does not have any worst-case polynomial time computational complexity guarantees [13], and [31]–[36]. Thus, in order to address this problem, we resort to the proposed POTDC algorithm that can be adopted for corresponding DC problem and that has polynomial time complexity.

It is noteworthy to mention that, the problem considered in [26] is somehow related to our problem. Specifically, the work in [26] considers the fractional QCQP problem that is closest mathematically to the one addressed in this chapter with the significant difference though that the objective in [26] contains only a single quadratic fractional function that simplifies the problem dramatically.

By means of the POTDC method, we aim at developing a polynomial time algorithm for solving the non-convex problem of maximizing a product of quadratic fractional functions under a quadratic constraint, which precisely corresponds to the sum-rate maximization in AF MIMO TWR. Specifically and similar to the previous chapter, we first rewrite the original non-convex problem as the maximization of an optimal value function. Despite the fact that the corresponding optimization problem of the newly defined optimal value function is non-convex, it is replaced with another equivalent function. The optimization problem that corresponds to such new optimal value function is convex and it simplifies the problem significantly. The new resulted optimal value function is then maximized recursively using the POTDC method. The proposed algorithm is guaranteed to find at least a KKT point, i.e., a point which satisfies the KKT optimality conditions. Moreover, according to our numerous numerical simulations, the obtained point is also globally optimal. The latter is confirmed by the fact that the point found by the algorithm coincides with the newly developed upper-bound for the optimal value of the problem. Similar to the previous chapter, the global optimality of the point found by

the POTDC method is equivalent to the concavity of an optimal value function of single parameter. Such Concavity can be checked numerically as it was explained in Section 6.3.

We also consider the relay amplification matrix design when the max-min rate and PF are used as the design criteria. Similar to the maximum sum-rate criterion, it is shown that the corresponding optimization problems for the max-min rate and PF criteria are also in the form of the generalized QCQP problems that can be precisely recast as the corresponding DC programming problems. These problems are then addressed similarly using the POTDC algorithm. Our numerical results confirm the superiority of the proposed relay transmit strategy design over other the state-of-the-art methods. Moreover, they confirm the global optimality of the proposed method.

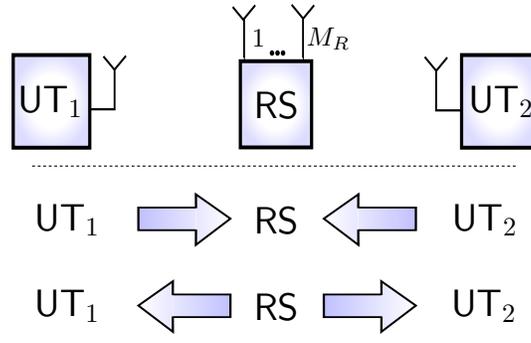
The rest of this chapter is organized as follows. The AF MIMO TWR system model is given in Section 7.1 while the sum-rate maximization problem for the corresponding system is formulated in Section 7.2. The POTDC algorithm for the sum-rate maximization is developed in Section 7.3 and an upper-bound for the optimal value of this maximization problem is found in Section 7.4. In Section 7.5, the relay amplification matrix design problems based on PF and MMRF criteria are formulated and then they are solved using the POTDC method. Simulation results are reported in Sections 7.6 and 7.7. Finally, Section 7.8 presents our conclusions and discussions.

## 7.1 System model

We consider a TWR system with two single antenna terminals and an AF relay equipped with  $M_R$  antennas. Fig. 7.1 shows the system we study in this chapter. In the first transmission phase, both terminals transmit to the relay. Assuming frequency-flat quasi-static block fading, the received signal at the relay can be expressed as

$$\mathbf{r} = \mathbf{h}_1^{(f)} x_1 + \mathbf{h}_2^{(f)} x_2 + \mathbf{n}_R \quad (7.1)$$

where  $\mathbf{h}_i^{(f)} \triangleq [h_{i,1}, \dots, h_{i,M_R}]^T \in \mathbb{C}^{M_R}$  represents the (forward) channel vector between terminal  $i$  and the relay,  $x_i$  is the transmitted symbol from terminal  $i$ ,  $\mathbf{n}_R \in \mathbb{C}^{M_R}$  denotes the additive noise component at the relay. Let  $P_{T,i} \triangleq \mathbb{E}\{|x_i|^2\}$  be the



**Figure 7.1:** Two-way relaying system model.

average transmit power of terminal  $i$  and  $\mathbf{R}_{N,R} \triangleq \mathbb{E}\{\mathbf{n}_R \mathbf{n}_R^H\}$  be the noise covariance matrix at the relay. For the special case of white noise we have  $\mathbf{R}_{N,R} = P_{N,R} \mathbf{I}_{M_R}$  where  $P_{N,R} = \text{tr}\{\mathbf{R}_{N,R}\}/M_R$ ,  $\mathbf{I}_{M_R}$  is the identity matrix of size  $M_R \times M_R$ . The relay amplifies the received signal by multiplying it with a relay amplification matrix  $\mathbf{G} \in \mathbb{C}^{M_R \times M_R}$ , i.e., it transmits the signal

$$\bar{\mathbf{r}} = \mathbf{G} \mathbf{r}. \quad (7.2)$$

The transmit power used by the relay can be expressed as

$$\mathbb{E}\{\|\bar{\mathbf{r}}\|_2^2\} = \mathbb{E}\{\text{tr}\{\mathbf{G} \mathbf{r} \mathbf{r}^H \mathbf{G}^H\}\} = \text{tr}\{\mathbf{G} \mathbf{R}_R \mathbf{G}^H\} = \text{tr}\{\mathbf{G}^H \mathbf{G} \mathbf{R}_R\} \quad (7.3)$$

where  $\mathbf{R}_R \triangleq \mathbb{E}\{\mathbf{r} \mathbf{r}^H\}$  is the covariance matrix of  $\mathbf{r}$  which is given by

$$\mathbf{R}_R = \mathbf{h}_1^{(f)} \left(\mathbf{h}_1^{(f)}\right)^H P_{T,1} + \mathbf{h}_2^{(f)} \left(\mathbf{h}_2^{(f)}\right)^H P_{T,2} + \mathbf{R}_{N,R}. \quad (7.4)$$

The covariance matrix  $\mathbf{R}_R$  is assumed to be full rank which is true under the common practical assumption that the noise covariance matrix  $\mathbf{R}_{N,R}$  is full rank. However, the case of rank deficient  $\mathbf{R}_R$  is considered for completeness in Appendix I as well.

Using the equality

$$\text{tr}\{\mathbf{A}^H \mathbf{B}\} = \text{vec}\{\mathbf{A}\}^H \text{vec}\{\mathbf{B}\} \quad (7.5)$$

which holds for any arbitrary square matrices  $\mathbf{A}$  and  $\mathbf{B}$ , the total transmit power of the relay (7.3) can be equivalently expressed as

$$\mathbb{E}\{\|\bar{\mathbf{r}}\|_2^2\} = \text{vec}\{\mathbf{G}\}^H \text{vec}\{\mathbf{G} \mathbf{R}_R\}, \quad (7.6)$$

where  $\text{vec}\{\cdot\}$  stands for the vectorization operation that transforms a matrix into a long vector stacking the columns of the matrix one after another. Finally, using the equality  $\text{vec}\{\mathbf{A} \mathbf{B}\} = (\mathbf{B}^T \otimes \mathbf{I}) \text{vec}\{\mathbf{A}\}$ , which is valid for any arbitrary square

matrices  $\mathbf{A}$  and  $\mathbf{B}$ , (7.6) can be equivalently rewritten as the following quadratic form

$$\mathbb{E}\{\|\bar{\mathbf{r}}\|_2^2\} = \mathbf{g}^H \underbrace{(\mathbf{R}_R^T \otimes \mathbf{I}_{M_R})}_{\mathbf{Q}} \mathbf{g} = \mathbf{g}^H \mathbf{Q} \mathbf{g} \quad (7.7)$$

where  $\mathbf{g} \triangleq \text{vec}\{\mathbf{G}\}$ ,  $\mathbf{Q} \triangleq \mathbf{R}_R^T \otimes \mathbf{I}_{M_R}$ . Since  $\mathbf{Q}$  is the Kronecker product of the two full rank positive definite matrices  $\mathbf{R}_R^T$  and  $\mathbf{I}_{M_R}$ , it is also full rank and positive definite [132].

In the second phase, the terminals receive the relay's transmission via the (backward) channels  $(\mathbf{h}_1^{(b)})^T$  and  $(\mathbf{h}_2^{(b)})^T$  (in the special case when reciprocity holds we have  $\mathbf{h}_i^{(b)} = \mathbf{h}_i^{(f)}$  for  $i = 1, 2$ ). Consequently, the received signals  $y_i$ ,  $i = 1, 2$  at both terminals can be expressed, respectively, as

$$y_1 = h_{1,1}^{(e)} x_1 + h_{1,2}^{(e)} x_2 + \tilde{n}_1 \quad (7.8)$$

$$y_2 = h_{2,2}^{(e)} x_2 + h_{2,1}^{(e)} x_1 + \tilde{n}_2 \quad (7.9)$$

where  $h_{i,j}^{(e)} \triangleq (\mathbf{h}_i^{(b)})^T \mathbf{G} \mathbf{h}_j^{(f)}$  is the effective channel between terminals  $i$  and  $j$  for  $i, j = 1, 2$  and  $\tilde{n}_i \triangleq (\mathbf{h}_i^{(b)})^T \mathbf{G} \mathbf{n}_R + n_i$  represents the effective noise contribution at terminal  $i$  which comprises the terminal's own noise as well as the forwarded relay noise. The first term in the received signal of each terminal represents the self-interference, which can be subtracted by the terminal since its own transmitted signal is known. The required channel knowledge for this step can be easily obtained, for example, via the LS compound channel estimator of [133].

After the cancelation of the self-interference, the two-way relaying system is decoupled into two parallel single-user single-input single-output (SISO) systems. Consequently, the rate  $r_i$  of terminal  $i$  can be expressed as

$$r_i = \frac{1}{2} \log_2 \left( 1 + \frac{P_{R,i}}{\tilde{P}_{N,i}} \right) = \frac{1}{2} \log_2 \left( \frac{\tilde{P}_{R,i}}{\tilde{P}_{N,i}} \right) \quad (7.10)$$

where  $P_{R,i}$  and  $\tilde{P}_{N,i}$  are the powers of the desired signal and the effective noise term at terminal  $i$ , respectively, and  $\tilde{P}_{R,i} \triangleq P_{R,i} + \tilde{P}_{N,i}$ . Specifically,  $P_{R,1} \triangleq \mathbb{E} \left\{ \left| h_{1,2}^{(e)} x_2 \right|^2 \right\}$ ,  $P_{R,2} \triangleq \mathbb{E} \left\{ \left| h_{2,1}^{(e)} x_1 \right|^2 \right\}$ , and  $\tilde{P}_{N,i} \triangleq \mathbb{E} \{ |\tilde{n}_i|^2 \}$  for  $i = 1, 2$ . Note that the factor  $1/2$  results from the two time slots needed for the bidirectional transmission. The powers of the desired signal and the effective noise term at terminal  $i$  can

be equivalently expressed as

$$P_{R,1} = P_{T,2} \left| \left( \mathbf{h}_1^{(b)} \right)^T \mathbf{G} \mathbf{h}_2^{(f)} \right|^2 \quad (7.11)$$

$$P_{R,2} = P_{T,1} \left| \left( \mathbf{h}_2^{(b)} \right)^T \mathbf{G} \mathbf{h}_1^{(f)} \right|^2 \quad (7.12)$$

$$\tilde{P}_{N,i} = \mathbb{E} \left\{ \left| \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{n}_R + n_i \right|^2 \right\} = \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{R}_{N,R} \mathbf{G}^H \left( \mathbf{h}_i^{(b)} \right)^* + P_{N,i} \quad (7.13)$$

where the expectation is taken with respect to the transmit signals and also the additional noise terms,  $P_{N,i}$  denotes the variance of the additive noise at terminal  $i$ , i.e.,  $n_i$ . Moreover, these powers can be further expressed as quadratic forms in  $\mathbf{g}$ . For this goal, first note that by using the following equality

$$\text{vec}\{\mathbf{ABC}\} = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}\{\mathbf{B}\} \quad (7.14)$$

which is valid for any arbitrary matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  of compatible dimensions, the term  $\left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{h}_j^{(f)}$  can be modified as follows

$$\left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{h}_j^{(f)} = \text{vec} \left\{ \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{h}_j^{(f)} \right\} = \left( \left( \mathbf{h}_j^{(f)} \right)^T \otimes \left( \mathbf{h}_i^{(b)} \right)^T \right) \text{vec}\{\mathbf{G}\}. \quad (7.15)$$

Using (7.15), the power of the desired signal at the first terminal can be expressed as

$$P_{R,1} = \mathbf{g}^H \left( \left( \mathbf{h}_2^{(f)} \right)^T \otimes \left( \mathbf{h}_1^{(b)} \right)^T \right)^H \left( \left( \mathbf{h}_2^{(f)} \right)^T \otimes \left( \mathbf{h}_1^{(b)} \right)^T \right) \mathbf{g} \cdot P_{T,2}. \quad (7.16)$$

Applying also the equality  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$  to (7.16) which is valid for any arbitrary matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  of compatible dimensions,  $P_{R,1}$  can be expressed as the following quadratic form

$$P_{R,1} = \mathbf{g}^H \left[ \left( \mathbf{h}_2^{(f)} \left( \mathbf{h}_2^{(f)} \right)^H \right) \otimes \left( \mathbf{h}_1^{(b)} \left( \mathbf{h}_1^{(b)} \right)^H \right) \right]^T \mathbf{g} \cdot P_{T,2}. \quad (7.17)$$

The power of the desired signal at the second terminal, i.e.,  $P_{R,2}$ , can be obtained similarly. By defining the matrices  $\mathbf{K}_{2,1}$  and  $\mathbf{K}_{1,2}$  as follows

$$\mathbf{K}_{2,1} \triangleq \left[ \left( \mathbf{h}_2^{(f)} \left( \mathbf{h}_2^{(f)} \right)^H \right) \otimes \left( \mathbf{h}_1^{(b)} \left( \mathbf{h}_1^{(b)} \right)^H \right) \right]^T \quad (7.18)$$

$$\mathbf{K}_{1,2} \triangleq \left[ \left( \mathbf{h}_1^{(f)} \left( \mathbf{h}_1^{(f)} \right)^H \right) \otimes \left( \mathbf{h}_2^{(b)} \left( \mathbf{h}_2^{(b)} \right)^H \right) \right]^T \quad (7.19)$$

the powers of the desired signal can be expressed as

$$P_{R,1} = \mathbf{g}^H \mathbf{K}_{2,1} \mathbf{g} \cdot P_{T,2} \quad (7.20)$$

$$P_{R,2} = \mathbf{g}^H \mathbf{K}_{1,2} \mathbf{g} \cdot P_{T,1}. \quad (7.21)$$

Since the matrices  $\mathbf{h}_i^{(f)} \left( \mathbf{h}_i^{(f)} \right)^H$ ,  $i = 1, 2$  and  $\mathbf{h}_i^{(b)} \left( \mathbf{h}_i^{(b)} \right)^H$ ,  $i = 1, 2$  are all positive semi-definite and the Kronecker product of positive semi-definite matrices is a positive semi-definite matrix [132], the matrices  $\mathbf{K}_{2,1}$  and  $\mathbf{K}_{1,2}$  are also positive semi-definite. As the last step, the effective noise  $\tilde{P}_{N,i}$  can be converted into a quadratic form of  $\mathbf{g}$  through the following train of equalities

$$\begin{aligned} \tilde{P}_{N,i} &= \mathbb{E} \left\{ \left| \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{n}_R + n_i \right|^2 \right\} \\ &= \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{R}_{N,R} \mathbf{G}^H \left( \mathbf{h}_i^{(b)} \right)^* + P_{N,i} \\ &= \text{tr} \left\{ \mathbf{G}^H \left( \mathbf{h}_i^{(b)} \right)^* \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{R}_{N,R} \right\} + P_{N,i} \end{aligned} \quad (7.22)$$

$$= \text{vec} \{ \mathbf{G} \}^H \text{vec} \left\{ \left( \mathbf{h}_i^{(b)} \right)^* \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G} \mathbf{R}_{N,R} \right\} + P_{N,i} \quad (7.23)$$

$$= \text{vec} \{ \mathbf{G} \}^H \left[ \mathbf{R}_{N,R} \otimes \left( \mathbf{h}_i^{(b)} \right) \left( \mathbf{h}_i^{(b)} \right)^H \right]^T \text{vec} \{ \mathbf{G} \} + P_{N,i} \quad (7.24)$$

$$= \mathbf{g}^H \mathbf{J}_i \mathbf{g} + P_{N,i} \quad (7.25)$$

where (7.23) is obtained from (7.22) by applying the equality (7.5), (7.24) is obtained from (7.23) by applying the equality (7.14), and the matrix  $\mathbf{J}_i$  in (7.25) is defined as

$$\mathbf{J}_i \triangleq \left[ \mathbf{R}_{N,R} \otimes \left( \mathbf{h}_i^{(b)} \left( \mathbf{h}_i^{(b)} \right)^H \right) \right]^T. \quad (7.26)$$

Note that,  $\mathbf{J}_i, i = 1, 2$  are positive semi-definite matrices because the matrices  $\mathbf{R}_{N,R}$  and  $\mathbf{h}_i^{(b)} \left( \mathbf{h}_i^{(b)} \right)^H$ ,  $i = 1, 2$  are positive semi-definite.

## 7.2 Problem statement for sum-rate maximization

Our goal is to find the relay amplification matrix  $\mathbf{G}$  which maximizes the sum-rate  $r_1 + r_2$  subject to a power constraint at the relay. For convenience we express the objective function and its solution in terms of  $\mathbf{g}$ . Then the power constrained sum-rate maximization problem can be expressed as

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} (r_1 + r_2) \quad (7.27)$$

where  $P_{T,R}$  is the total available transmit power at the relay. Using the definitions from the previous section, this optimization problem can be rewritten as

$$\begin{aligned} \mathbf{g}_{\text{opt}} &= \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} \frac{1}{2} \log_2 \left[ \left( 1 + \frac{P_{R,1}}{\tilde{P}_{N,1}} \right) \cdot \left( 1 + \frac{P_{R,2}}{\tilde{P}_{N,2}} \right) \right] \\ &= \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} \left( 1 + \frac{P_{R,1}}{\tilde{P}_{N,1}} \right) \cdot \left( 1 + \frac{P_{R,2}}{\tilde{P}_{N,2}} \right) \end{aligned} \quad (7.28)$$

$$= \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} \frac{\tilde{P}_{R,1}}{\tilde{P}_{N,1}} \cdot \frac{\tilde{P}_{R,2}}{\tilde{P}_{N,2}} \quad (7.29)$$

where we have used the fact that  $0.5 \cdot \log_2(x)$  is a monotonic function in  $x \in \mathbb{R}^+$  where  $\mathbb{R}^+$  is the set of positive real numbers, and  $\tilde{P}_{R,i}, i = 1, 2$  are defined after (7.10).

It is worth noting that the inequality constraint in the optimization problem (7.29) has to be active at the optimal point. This can be easily shown by contradiction. Assume that  $\mathbf{g}_{\text{opt}}$  satisfies  $\mathbf{g}_{\text{opt}}^H \mathbf{Q} \mathbf{g}_{\text{opt}} < P_{T,R}$ . Then we can find a constant  $c > 1$  such that  $\bar{\mathbf{g}}_{\text{opt}} = c \cdot \mathbf{g}_{\text{opt}}$  satisfies  $\bar{\mathbf{g}}_{\text{opt}}^H \mathbf{Q} \bar{\mathbf{g}}_{\text{opt}} = P_{T,R}$ . The latter follows from the fact that  $\mathbf{Q}$  is positive definite and, therefore,  $\mathbf{g}_{\text{opt}}^H \mathbf{Q} \mathbf{g}_{\text{opt}}$  is positive. However, inserting  $\bar{\mathbf{g}}_{\text{opt}}$  in the objective function of (7.28), we obtain

$$\begin{aligned} & \left( 1 + \frac{c^2 \cdot \mathbf{g}_{\text{opt}}^H \mathbf{K}_{2,1} \mathbf{g}_{\text{opt}} P_{T,2}}{c^2 \cdot \mathbf{g}_{\text{opt}}^H \mathbf{J}_1 \mathbf{g}_{\text{opt}} + P_{N,1}} \right) \cdot \left( 1 + \frac{c^2 \cdot \mathbf{g}_{\text{opt}}^H \mathbf{K}_{1,2} \mathbf{g}_{\text{opt}} P_{T,1}}{c^2 \cdot \mathbf{g}_{\text{opt}}^H \mathbf{J}_2 \mathbf{g}_{\text{opt}} + P_{N,2}} \right) \\ &= \left( 1 + \frac{\mathbf{g}_{\text{opt}}^H \mathbf{K}_{2,1} \mathbf{g}_{\text{opt}} P_{T,2}}{\mathbf{g}_{\text{opt}}^H \mathbf{J}_1 \mathbf{g}_{\text{opt}} + \frac{P_{N,1}}{c^2}} \right) \cdot \left( 1 + \frac{\mathbf{g}_{\text{opt}}^H \mathbf{K}_{1,2} \mathbf{g}_{\text{opt}} P_{T,1}}{\mathbf{g}_{\text{opt}}^H \mathbf{J}_2 \mathbf{g}_{\text{opt}} + \frac{P_{N,2}}{c^2}} \right) \end{aligned} \quad (7.30)$$

which is monotonically increasing in  $c$ . Since we have  $c > 1$ , the vector  $\bar{\mathbf{g}}_{\text{opt}}$  provides a larger value of the objective functions than  $\mathbf{g}_{\text{opt}}$  which contradicts the assumption that  $\mathbf{g}_{\text{opt}}$  was optimal.

As a result, we have shown that the optimal vector  $\mathbf{g}_{\text{opt}}$  must satisfy the total power constraint of the problem (7.29) with equality, i.e.,  $\mathbf{g}_{\text{opt}}^H \mathbf{Q} \mathbf{g}_{\text{opt}} = P_{T,R}$ . Using this fact, the inequality constraint in the problem (7.29) can be replaced by the constraint  $\mathbf{g}^H \mathbf{Q} \mathbf{g} = P_{T,R}$ . This enables us to substitute the constant term  $P_{N,i}$ , which appears in the effective noise power at terminal  $i$  (7.25), with the quadratic term of  $(P_{N,i}/P_{T,R}) \cdot \mathbf{g}_{\text{opt}}^H \mathbf{Q} \mathbf{g}_{\text{opt}}$ . This leads to an equivalent homogeneous expression for the ratio of  $\tilde{P}_{R,i}/\tilde{P}_{N,i}, i = 1, 2$ . Thus, by using such substitution,  $\tilde{P}_{N,i}, i = 1, 2$  from (7.25) can be equivalently written as

$$\tilde{P}_{N,i} = \mathbf{g}^H \mathbf{B}_i \mathbf{g}, \quad i = 1, 2 \quad (7.31)$$

where  $\mathbf{B}_i$  is defined as

$$\mathbf{B}_i \triangleq \mathbf{J}_i + \frac{P_{N,i}}{P_{T,R}} \cdot \mathbf{Q}. \quad (7.32)$$

Inserting (7.20), (7.21), and (7.32) into (7.29), the optimization problem becomes

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} = P_{T,R}} \frac{\mathbf{g}^H \mathbf{A}_1 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_1 \mathbf{g}} \cdot \frac{\mathbf{g}^H \mathbf{A}_2 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_2 \mathbf{g}} \quad (7.33)$$

where we have defined the new matrices  $\mathbf{A}_1 \triangleq \mathbf{K}_{2,1} \cdot P_{T,2} + \mathbf{B}_1$  and  $\mathbf{A}_2 \triangleq \mathbf{K}_{1,2} \cdot P_{T,1} + \mathbf{B}_2$ . Since the matrices  $\mathbf{J}_i$ ,  $i = 1, 2$ ,  $\mathbf{K}_{1,2}$ , and  $\mathbf{K}_{2,1}$  are positive semi-definite and  $\mathbf{Q}$  is a full rank positive definite matrix, the matrices  $\mathbf{A}_1$ ,  $\mathbf{A}_2$ ,  $\mathbf{B}_1$ , and  $\mathbf{B}_2$  are all full rank positive definite matrices and hence invertible. Moreover,  $\mathbf{A}_1$ ,  $\mathbf{A}_2$ ,  $\mathbf{B}_1$ , and  $\mathbf{B}_2$  are all  $M_R^2 \times M_R^2$  matrices.

As a final simplifying step we observe that the objective function of (7.33) is homogeneous in  $\mathbf{g}$ , meaning that an arbitrary rescaling of  $\mathbf{g}$  has no effect on the value of the objective function. Consequently, the equality constraint can be dropped since any solution to the unconstrained problem can be rescaled to meet the equality constraint without any loss in terms of the objective function. Therefore, the final form of our problem statement is given by

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g}} \frac{\mathbf{g}^H \mathbf{A}_1 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_1 \mathbf{g}} \cdot \frac{\mathbf{g}^H \mathbf{A}_2 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_2 \mathbf{g}}. \quad (7.34)$$

The optimization problem (7.34) can be interpreted as the product of two Rayleigh quotients (quadratic fractional functions). Moreover, it can be expressed as a DC programming problem. Indeed, as we will show later the objective function of the problem (7.34) can be written as a summation of two concave functions with positive sign and one concave function with negative sign. Thus, the objective of the equivalent problem is, in fact, the difference of convex functions which is in general non-convex. As it was mentioned earlier, the available algorithms for solving such DC programming problems are based on the branch-and-bound method that does not have any polynomial time computational complexity guarantees [13], and [31]–[36]. However, as we show next by means of the POTDC, at least a KKT point of the problem (7.34) can be found in polynomial time with a great evidence that such a solution is also globally optimal.

### 7.3 POTDC algorithm in the sum-rate maximization

Since the problem (7.34) is homogeneous, without loss of generality, we can fix the quadratic term  $\mathbf{g}^H \mathbf{B}_1 \mathbf{g}$  to be equal to one at the optimal point. By doing so and also

by defining the additional variables  $\tau$  and  $\beta$ , the problem (7.34) can be equivalently recast as

$$\begin{aligned} & \max_{\mathbf{g}, \tau, \beta} \quad \mathbf{g}^H \mathbf{A}_1 \mathbf{g} \cdot \frac{\tau}{\beta} \\ & \text{subject to} \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \tau, \quad \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta. \end{aligned} \quad (7.35)$$

For future reference, we need the range of the variable  $\beta$ . Due to the fact that the quadratic function  $\mathbf{g}^H \mathbf{B}_1 \mathbf{g}$  is set to one, this range can be easily obtained. Specifically, the smallest value of  $\beta$  for which the problem (7.35) is still feasible can be obtained by solving the following problem

$$\min_{\mathbf{g}} \mathbf{g}^H \mathbf{B}_2 \mathbf{g} \quad \text{subject to} \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1. \quad (7.36)$$

Note that  $\tau$  does not impose any restriction on the smallest possible value of  $\beta$ , because if  $\mathbf{g}_{\min, \beta}$  denotes the optimal solution of the problem (7.36), then  $\tau$  can be chosen as  $\tau = \mathbf{g}_{\min, \beta}^H \mathbf{A}_2 \mathbf{g}_{\min, \beta}$ . Since the matrix  $\mathbf{B}_1$  is positive definite, it can be decomposed as  $\mathbf{B}_1 = \mathbf{B}_1^{1/2} (\mathbf{B}_1^{1/2})^H$  where  $\mathbf{B}_1^{1/2}$  is a square root of  $\mathbf{B}_1$  and it is invertible due to the properties in [134]. By defining the new vector  $\mathbf{y} \triangleq (\mathbf{B}_1^{1/2})^H \mathbf{g}$ , i.e.,  $\mathbf{g} = (\mathbf{B}_1^{-1/2})^H \mathbf{y}$  the problem (7.36) is equivalent to

$$\min_{\mathbf{y}} \mathbf{y}^H \mathbf{B}_1^{-1/2} \mathbf{B}_2 (\mathbf{B}_1^{-1/2})^H \mathbf{y} \quad \text{subject to} \quad \mathbf{y}^H \mathbf{y} = 1. \quad (7.37)$$

It is well known that according to the minimax Theorem [124], the optimal value of (7.37) is the smallest eigenvalue of the matrix  $\mathbf{B}_1^{-1/2} \mathbf{B}_2 (\mathbf{B}_1^{-1/2})^H$ . Using the fact that for any arbitrary square matrices  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$ , the eigenvalues of the matrix products  $\mathbf{Z}_1 \mathbf{Z}_2$  and  $\mathbf{Z}_2 \mathbf{Z}_1$  are the same [135], it can be concluded that the smallest eigenvalue of  $\mathbf{B}_1^{-1/2} \mathbf{B}_2 (\mathbf{B}_1^{-1/2})^H$  is the same as the smallest eigenvalue of  $(\mathbf{B}_1^{-1/2})^H \mathbf{B}_1^{-1/2} \mathbf{B}_2$  or, equivalently,  $\mathbf{B}_1^{-1} \mathbf{B}_2$ .

The largest value of  $\beta$  for which the problem (7.35) is still feasible can be obtained in a similar way, and it is equal to the largest eigenvalue of the matrix  $\mathbf{B}_1^{-1} \mathbf{B}_2$ . As a result, the range of  $\beta$  is  $[\lambda_{\min}\{\mathbf{B}_1^{-1} \mathbf{B}_2\}, \lambda_{\max}\{\mathbf{B}_1^{-1} \mathbf{B}_2\}]$  where  $\lambda_{\min}\{\cdot\}$  denotes the smallest eigenvalue operator. Note that, since the matrices  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are positive definite (hence  $\mathbf{B}_1^{-1}$  is also positive definite), the eigenvalues of the product  $\mathbf{B}_1^{-1} \mathbf{B}_2$  are all positive due to the properties in [134] including  $\lambda_{\min}\{\mathbf{B}_1^{-1} \mathbf{B}_2\}$ .

For future reference, we define the following optimal value function function of  $\tau$  and  $\beta$

$$\begin{aligned} g(\tau, \beta) \triangleq & \left\{ \max_{\mathbf{g}} \mathbf{g}^H \mathbf{A}_1 \mathbf{g} \mid \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \tau, \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta \right\}, \\ & (\tau, \beta) \in \mathcal{D}(7.38) \end{aligned}$$

where  $\mathcal{D} \subset \mathbb{R}^2$  is the set of all pairs  $(\tau, \beta)$  such that the corresponding optimization problem obtained from  $g(\tau, \beta)$  for fixed  $\tau$  and  $\beta$  is feasible. Therefore, using the optimal value function  $g(\tau, \beta)$ , the original optimization problem (7.35) can be equivalently recast as

$$\max_{\tau, \beta} g(\tau, \beta) \cdot \frac{\tau}{\beta}. \quad (7.39)$$

Introducing the matrix  $\mathbf{X} \triangleq \mathbf{g}\mathbf{g}^H$  and observing that for any arbitrary matrix  $\mathbf{Y}$ , the relationship  $\mathbf{g}^H \mathbf{Y} \mathbf{g} = \text{tr}\{\mathbf{Y}\mathbf{g}\mathbf{g}^H\}$  holds, the optimal value function  $g(\tau, \beta)$  can be equivalently recast as

$$g(\tau, \beta) = \left\{ \max_{\mathbf{X}} \text{tr}\{\mathbf{A}_1 \mathbf{X}\} \mid \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau, \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \right. \\ \left. \text{rank}\{\mathbf{X}\} = 1, \mathbf{X} \succeq \mathbf{0} \right\}, (\tau, \beta) \in \mathcal{D}. \quad (7.40)$$

In the optimization problem obtained from the optimal value function  $g(\tau, \beta)$  (7.40) by fixing  $\tau$  and  $\beta$ , the rank-one constraint  $\text{rank}\{\mathbf{X}\} = 1$  is the only non-convex constraint with respect to the new optimization variable  $\mathbf{X}$ . Using SDR, the corresponding optimization problem can be relaxed by dropping the rank-one constraint, and the following new optimal value function  $h(\tau, \beta)$  can be defined

$$h(\tau, \beta) \triangleq \left\{ \max_{\mathbf{X}} \text{tr}\{\mathbf{A}_1 \mathbf{X}\} \mid \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau, \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \right. \\ \left. \mathbf{X} \succeq \mathbf{0} \right\}, (\tau, \beta) \in \mathcal{D}' \quad (7.41)$$

where  $\mathcal{D}' \subset \mathbb{R}^2$  is the set of all pairs  $(\tau, \beta)$  such that the optimization problem corresponding to  $h(\tau, \beta)$  for fixed  $\tau$  and  $\beta$  is feasible. For brevity, we will refer to the optimization problems corresponding to the functions  $g(\tau, \beta)$  and  $h(\tau, \beta)$  when  $\tau$  and  $\beta$  are fixed simply as the optimization problems of  $g(\tau, \beta)$  and  $h(\tau, \beta)$ , respectively. The optimal value functions  $g(\tau, \beta)$  and  $h(\tau, \beta)$  are special cases of the optimal value functions (3.6) and (3.9) that were defined in Chapter 3. Therefore, according to the Lemma 3.1, the domains of these optimal value functions are the same, i.e.,  $\mathcal{D} = \mathcal{D}'$ . Moreover, the Theorem 3.1 implies that these optimal value functions are equivalent, i.e.,  $g(\tau, \beta) = h(\tau, \beta)$ ,  $(\tau, \beta) \in \mathcal{D}$ . Furthermore, based on the optimal solution of the optimization problem of  $h(\tau, \beta)$  when  $\tau$  and  $\beta$  are fixed, the optimal solution of the optimization problem of  $g(\tau, \beta)$  can be constructed by means of the Theorem 3.1.

Although the optimal value functions  $g(\tau, \beta)$  and  $h(\tau, \beta)$  are equal, however, compared to the optimization problem of  $g(\tau, \beta)$  which is non-convex, the optimization problem of  $h(\tau, \beta)$  is convex. Using this fact and replacing  $g(\tau, \beta)$  by  $h(\tau, \beta)$  in the original optimization problem (7.39), the problem (7.35) can be simplified as

$$\begin{aligned} & \max_{\mathbf{X}, \tau, \beta} \quad \text{tr}\{\mathbf{A}_1 \mathbf{X}\} \cdot \frac{\tau}{\beta} \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \quad (7.42)$$

Therefore, instead of the original optimization problem (7.35), we can solve the simplified problem (7.42). Based on the optimal solution of the simplified problem, denoted as  $\mathbf{X}_{\text{opt}}$ ,  $\tau_{\text{opt}}$ , and  $\beta_{\text{opt}}$ , the optimal solution of the original problem can be found. The optimal values of  $\tau$  and  $\beta$  are equal to the corresponding optimal values of the simplified problem, while the optimal  $\mathbf{g}$  can be constructed based on  $\mathbf{X}_{\text{opt}}$  using rank-reduction techniques [106].

It is worth stressing that for every feasible point of the optimization problem (7.42) denoted as  $\mathbf{X}$ , the terms  $\text{tr}\{\mathbf{A}_1 \mathbf{X}\}$ ,  $\text{tr}\{\mathbf{A}_2 \mathbf{X}\}$ , and  $\text{tr}\{\mathbf{B}_2 \mathbf{X}\}$  are positive, and therefore, the corresponding objective value is positive as well. The latter can be easily verified by applying Lemma 1 of [137, Section 2] which states that for every Hermitian matrix  $\mathbf{A}$  and Hermitian positive semi-definite matrix  $\mathbf{B}$ ,  $\text{tr}\{\mathbf{A}\mathbf{B}\}$  is greater than or equal to  $\lambda_{\min}\{\mathbf{A}\}\text{tr}\{\mathbf{B}\}$ . Applying this lemma, it can be found that

$$\begin{aligned} \text{tr}\{\mathbf{A}_1 \mathbf{X}\} &= \text{tr}\{\mathbf{X}\mathbf{A}_1\} = \text{tr}\left\{\mathbf{X}\mathbf{B}_1^{1/2}\mathbf{B}_1^{-1/2}\mathbf{A}_1(\mathbf{B}_1^{1/2}\mathbf{B}_1^{-1/2})^H\right\} \\ &= \text{tr}\left\{(\mathbf{B}_1^{1/2})^H\mathbf{X}\mathbf{B}_1^{1/2}\mathbf{B}_1^{-1/2}\mathbf{A}_1(\mathbf{B}_1^{-1/2})^H\right\} \\ &\geq \text{tr}\{\mathbf{B}_1 \mathbf{X}\}\lambda_{\min}\{\mathbf{B}_1^{-1}\mathbf{A}_1\} \\ &= \lambda_{\min}\{\mathbf{B}_1^{-1}\mathbf{A}_1\} \end{aligned} \quad (7.43)$$

where Lemma 1 of [137, Section 2] has been applied in the second line of (7.43) and the last equality follows from the fact that  $\text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1$  as  $\mathbf{X}$  is a feasible point. Since  $\mathbf{B}_1^{-1}$  and  $\mathbf{A}_1$  are positive definite, all the eigenvalues of the product  $\mathbf{B}_1^{-1}\mathbf{A}_1$  are positive [134], and therefore,  $\lambda_{\min}\{\mathbf{A}_1\mathbf{B}_1^{-1}\}$  is positive. In a similar way, it can be proved that  $\text{tr}\{\mathbf{A}_2 \mathbf{X}\}$  and  $\text{tr}\{\mathbf{B}_2 \mathbf{X}\}$  are necessarily positive, and therefore, the variables  $\tau$  and  $\beta$  are also positive. Thus, the task of maximizing the objective function in the problem (7.42) is equivalent to maximizing the logarithm

of this objective function because  $\ln(x)$  is a strictly increasing function and the objective function in (7.42) is positive. Then, the optimization problem (7.42) can be equivalently rewritten as

$$\begin{aligned} & \max_{\mathbf{X}, \tau, \beta} \ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}) + \ln(\tau) - \ln(\beta) \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \tag{7.44}$$

In summary, by replacing  $g(\tau, \beta)$  by  $h(\tau, \beta)$ , we are able to write our optimization problem as a DC programming problem, where the fact that  $\ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\})$  in the objective of (7.44) is a concave function is also considered. Although the problem (7.44) boils down to the known family of DC programming problems, still there exists no solution for such DC programming problems with guaranteed polynomial time complexity. However, the problem (7.44) has a very particular structure, such as, all the constraints are convex and the terms  $\ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\})$  and  $\ln(\tau)$  in the objective are concave. Thus, the only term that makes the problem overall non-convex is the term  $-\ln(\beta)$  in the objective. If  $-\ln(\beta)$  is piece-wise linearized over a finite number of intervals<sup>1</sup>, then the objective function becomes concave on these intervals and the whole problem (7.44) becomes convex. The resulting convex problems over different linearization intervals for  $-\ln(\beta)$  can be solved efficiently in polynomial time, and then, the suboptimal solution of the problem (7.44) can be found. The fact that such a solution is suboptimal follows from the linearization, which has a finite accuracy. The smaller the intervals are, the more accurate the solution of (7.44) becomes. However, such a solution procedure is not the most efficient in terms of computational complexity. Thus, we resort to the POTDC method which was introduced in Chapter 3. For this goal, let us introduce a new additional variable  $t$  and then express the problem (7.44) equivalently as

$$\begin{aligned} & \max_{\mathbf{X}, \tau, \beta, t} \ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}) + \ln(\tau) - t \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \quad \ln(\beta) \leq t, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \tag{7.45}$$

The objective function of the optimization problem (7.45) is concave and all the constraints except the constraint  $\ln(\beta) \leq t$  are convex. The POTDC is based on

---

<sup>1</sup>As explained before, the parameter  $\beta$  can take values only in a finite interval. Thus, a finite number of linearization intervals for  $-\ln(\beta)$  is needed.

linearizing the non-convex term  $\ln(\beta)$  in the constraint  $\ln(\beta) \leq t$  suitably selected points in different iterations. More specifically, the linearizing point in each iteration is selected so that the objective function increases in every iteration of the iterative algorithm. In the first iteration, we start with an arbitrary point selected in the interval  $[\lambda_{\min}\{\mathbf{B}_1^{-1}\mathbf{B}_2\}, \lambda_{\max}\{\mathbf{B}_1^{-1}\mathbf{B}_2\}]$  and denoted as  $\beta_c$ . Then the non-convex function  $\ln(\beta)$  is replaced by its linear approximation around this point  $\beta_c$ , that is,

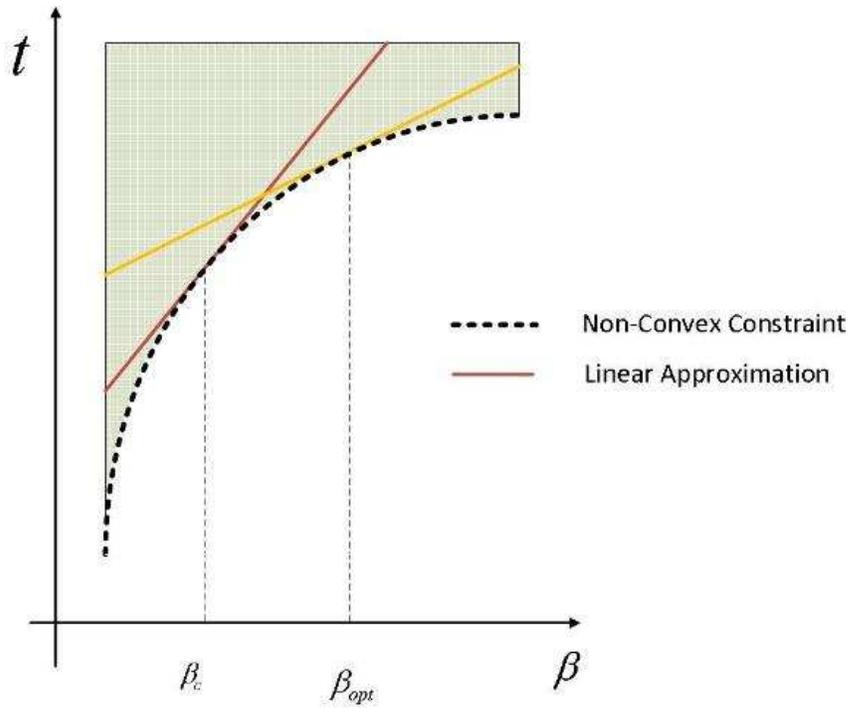
$$\ln(\beta) \approx \ln(\beta_c) + \frac{1}{\beta_c}(\beta - \beta_c) \quad (7.46)$$

which results in the following convex optimization problem

$$\begin{aligned} & \max_{\mathbf{X}, \tau, \beta, t} \quad \ln(\text{tr}\{\mathbf{A}_1\mathbf{X}\}) + \ln(\tau) - t \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1\mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2\mathbf{X}\} = \tau, \quad \text{tr}\{\mathbf{B}_2\mathbf{X}\} = \beta \\ & \ln(\beta_c) + \frac{1}{\beta_c}(\beta - \beta_c) \leq t, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \quad (7.47)$$

The problem (7.47) can be efficiently solved using the interior point-based numerical methods. As it is mentioned in Chapter 6, the worst-case computational complexity of solving a general SDP problem is equal to  $\mathcal{O}(n_c^2 n_v^{2.5} + n_c n_v^{3.5})$  [129] where  $n_c$  and  $n_v$  denote, respectively, the number of the constraints and the number of the variables of the primal problem (2.18). Therefore, the worst-case computational complexity of solving the problem (7.47) is  $\mathcal{O}((M_R^4 + 3)^{3.5})$  where  $M_R^4 + 3$  is the total number of optimization variables in the problem (7.47) including the real and imaginary parts of the elements of  $\mathbf{X}$  as well as  $\tau$ ,  $\beta$ , and  $t$ . Once the optimal solution of this problem, denoted in the first iteration as  $\mathbf{X}_{\text{opt}}^{(1)}$ ,  $\tau_{\text{opt}}^{(1)}$ ,  $\beta_{\text{opt}}^{(1)}$ , and  $t_{\text{opt}}^{(1)}$ , is found, the algorithm proceeds to the second iteration by replacing the function  $\ln(\beta)$  by its linear approximation around  $\beta_{\text{opt}}^{(1)}$  found from the previous (first) iteration. Fig. 7.2 shows how  $\ln(\beta)$  is replaced by its linear approximation around  $\beta_c$  where  $\beta_{\text{opt}}$  is the optimal value of  $\beta$  obtained through solving (7.47) using such a linear approximation. In the second iteration, the resulting optimization problem has the same structure as the problem (7.47) in which  $\beta_c$  has to be set to  $\beta_{\text{opt}}^{(1)}$  obtained from the first iteration. This process continues, and  $k$ th iteration is obtained by replacing  $\ln(\beta)$  by its linearization of type (7.46) around  $\beta_{\text{opt}}^{(k-1)}$  found at the iteration  $k - 1$ . The POTDC algorithm for solving the problem (7.45) is summarized in Algorithm 7.1.

According to the Lemma 3.2, the optimal values of the objective function of the optimization problem (7.47) obtained over the iterations of the POTDC algorithm



**Figure 7.2:** Linear approximation of  $\ln(\beta)$  around  $\beta_c$ . The region above the dashed curve is non-convex.

are non-decreasing. Moreover, this Lemma implies that the solution obtained using the Algorithm 7.1 satisfies the KKT conditions.

As soon as the solution of the relaxed problem (7.45) is found, the solution of the original problem (7.35), which is equivalent to the solution of the sum-rate maximization problem (7.34), can be found using one of the existing methods for extracting a rank-one solution [106], [136]. The rank reduction-based technique of [106] and the algebraic technique of [136] have been mentioned earlier, while the method based on solving the dual problem (see [16]) exploits the fact that a QCQP with only two constraints has zero duality gap. Note that the computational complexity of the POTDC algorithm is equivalent to that of solving an SDP problem, i.e.,  $\mathcal{O}((M_R^4 + 3)^{3.5})$ , times the number of iterations. It is noteworthy to mention that the computational complexity comparison between the new proposed method and the branch-and-bound method confirms the superiority of the proposed method [28].

Although the POTDC algorithm finds a KKT point for the considered sum-rate maximization problem, we also aim at showing the evidence that such a point is the globally optimal point. Toward this end, we will need an upper-bound for the

---

**Algorithm 7.1** The POTDC algorithm for solving the optimization problem (7.45)

---

**Initialize:** Select an arbitrary  $\beta_c$  from the interval  $[\lambda_{\min}\{\mathbf{B}_1^{-1}\mathbf{B}_2\}, \lambda_{\max}\{\mathbf{B}_1^{-1}\mathbf{B}_2\}]$ , set the counter  $k$  to be equal to 1 and choose an progress parameter  $\epsilon$ .

**while** The difference between the values of the objective function in two consecutive iterations is larger than  $\epsilon$ . **do**

Use the linearization of type (7.46) and solve the optimization problem (7.47) to obtain  $\mathbf{X}_{\text{opt}}^{(k)}$ ,  $\tau_{\text{opt}}^{(k)}$ ,  $\beta_{\text{opt}}^{(k)}$  and  $t_{\text{opt}}^{(k)}$ .

Set  $\mathbf{X}_{\text{opt}} = \mathbf{X}_{\text{opt}}^{(k)}$ , and  $\beta_c = \beta_{\text{opt}}^{(k)}$ .  
 $k = k + 1$ .

**end while**

**Output:**  $\mathbf{X}_{\text{opt}}$ .

---

optimal value.

## 7.4 An upper-bound for the optimal value

Through extensive simulations we have observed that regardless of the initial value chosen for  $\beta_c$  in the first iteration of the POTDC algorithm, the proposed iterative method always converges to the global optimum of the problem (7.45). However, since the original problem is not convex, this can not be easily verified analytically. A comparison between the optimal value obtained by using the proposed iterative method and also the globally optimal value can be, however, done by developing a tight upper-bound for the optimal value of the problem and comparing the solution to such an upper-bound. Thus, in this section, we find such an upper-bound for the optimal value of the optimization problem (7.45). For this goal, we first consider the following lemma which gives an upper-bound for the optimal value of the variable  $\beta$  in the problem (7.45). This lemma will further be used for obtaining the desired upper-bound for our problem.

**Lemma 7.1.** *The optimal value of the variable  $\beta$  in (7.45) or equivalently (7.44), denoted as  $\beta_{\text{opt}}$  is upper-bounded by  $e^{(q^* - p^*)}$ , where  $p^*$  is the value of the objective function in the problem (7.45) or equivalently (7.44) corresponding to any arbitrary feasible point and  $q^*$  is the optimal value of the following convex optimization prob-*

$$\begin{aligned}
 q^* &= \max_{\mathbf{X}, \tau, \beta} \ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}) + \ln(\tau) \\
 \text{subject to } & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \tau, \quad \mathbf{X} \succeq \mathbf{0}.
 \end{aligned} \tag{7.48}$$

**Proof:** See Appendix G.

Note that as mentioned earlier,  $p^*$  is the objective value of the problem (7.44) that corresponds to an arbitrary feasible point. In order to obtain the tightest possible upper-bound for  $\beta_{\text{opt}}$ , we choose  $p^*$  to be the largest possible value that we already know. A suitable choice for  $p^*$  is then the one which is obtained using the POTDC algorithm. In other words, we choose  $p^*$  as the corresponding objective value of the problem (7.44) at the optimal point which is resulted from the POTDC algorithm. Thus, we have obtained an upper-bound for  $\beta_{\text{opt}}$  which makes it further possible to develop an upper-bound for the optimal value of the optimization problem (7.44). To this end, we consider the only non-convex constraint of this problem, i.e.,  $\ln(\beta) \leq t$ . Fig. 7.3 illustrates a subset of the feasible region corresponding to the non-convex constraint  $\ln(\beta) \leq t$  where  $\beta_{\text{min}}$  equals  $\lambda_{\text{min}}\{\mathbf{B}_1^{-1} \mathbf{B}_2\}$ , i.e., the smallest value of  $\beta$  for which the problem (7.45) is feasible, and  $\beta_{\text{max}}$  is the upper-bound for the optimal value  $\beta_{\text{opt}}$  given by Lemma 7.1 ( $\beta_{\text{max}}$  is equal to  $\lambda_{\text{max}}\{\mathbf{B}_1^{-1} \mathbf{B}_2\}$  if it is smaller than the upper-bound of  $\beta_{\text{opt}}$  obtained using Lemma 7.1). For obtaining an upper-bound for the optimal value of the problem (7.45), we divide the interval  $[\beta_{\text{min}}, \beta_{\text{max}}]$  into  $N$  sections as it is shown in Fig. 7.3. Then, each section is considered separately. In each such section, the corresponding non-convex feasible set is replaced by its convex-hull and each corresponding optimization problem is solved separately as well. The maximum optimal value of such  $N$  convex optimization problems is then the upper-bound. Indeed, solving the resulting  $N$  convex optimization problems and choosing the maximum optimum value among them is equivalent to replacing the constraint  $\ln(\beta) \leq t$  with the feasible set which is described by the region above the solid line in Fig. 7.3. The upper-bound becomes more and more accurate when the number of the intervals, i.e.,  $N$  increases.

---

<sup>2</sup>Note that this optimization problem can be solved efficiently using numerical methods, for example, interior point methods.

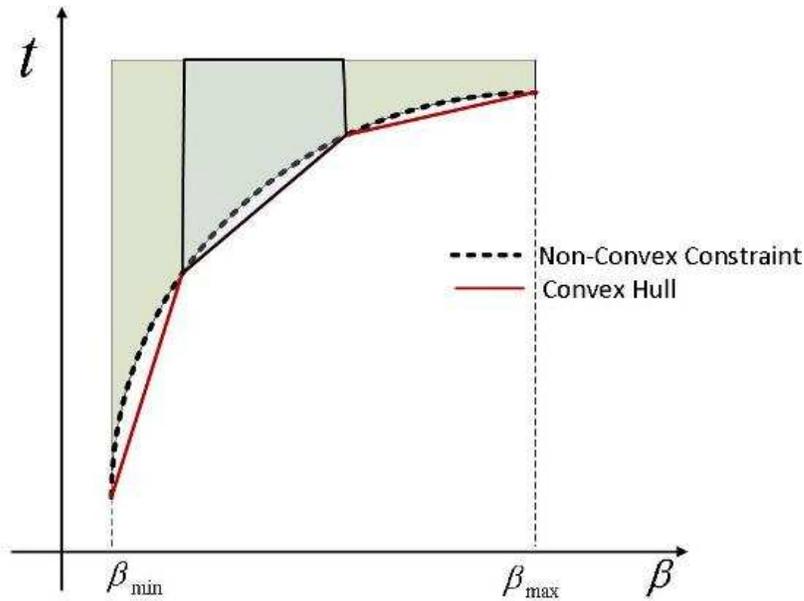


Figure 7.3: Feasible region of the constraint  $\ln(\beta) \leq t$  and the convex hull in each sub-division.

## 7.5 Proportional fair and max-min rate fair approaches

In the previous section, we considered the relay amplification matrix design to maximize the overall sum-rate. Although, the maximum sum-rate is a common performance criterion in relay transmit strategy design, it does not result in a fair resource distribution among different users. Specifically, it assigns the majority of the available resources to the users with better channel qualities so that the highest possible overall sum-rate can be achieved. Such resource allocation is not fair as the users with lower channel qualities are not able to communicate properly. The importance of the user fairness in asymmetric TWR systems has been recently demonstrated in [67], [68] and [69]. The authors of [67] study the optimal power allocation problem for single antenna users and single antenna relay where the sum-rate is maximized under the fairness constraint. Relay beamforming and optimal power allocation for a pair of single antenna users and several single antenna relays based on max-min SNR (data rate) has been also considered in [68] and [69].

In this section, we study the relay amplification matrix design for the same setup that was introduced in Section 7.1 when the MMRF and PF are used as the design criteria. It will be shown that the corresponding optimization problems of such design problems are also specific realizations of the introduced generalized QCQP

problem which can be precisely recast as DC programming problems. Since the approach for solving the corresponding optimization problems is so similar to the sum-rate maximization problem, the following steps will be only explained briefly.

### 7.5.1 Proportional fairness

PF has been initially introduced and applied in game theory [138]. In application to the resource allocation problem, it is known to provide a good trade-off between the maximum sum-rate and the user fairness [139]. It is also well known that a proportionally fair resource allocation/beamformer maximizes the sum of the logarithmic average sum-rate [140]. The relay amplification matrix design problem based on the PF criterion and subject to the total power constraint of the relay can be expressed as the following optimization problem

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} \frac{1}{4} \ln \left( 1 + \frac{P_{R,1}}{\tilde{P}_{N,1}} \right) \ln \left( 1 + \frac{P_{R,2}}{\tilde{P}_{N,2}} \right) \quad (7.49)$$

By following the same steps as in Section 7.2, the optimization problem (7.49) can be expressed as the following homogeneous problem

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g}} \ln \left( \frac{\mathbf{g}^H \mathbf{A}_1 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_1 \mathbf{g}} \right) \ln \left( \frac{\mathbf{g}^H \mathbf{A}_2 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_2 \mathbf{g}} \right) \quad (7.50)$$

Note that since the objective function of the optimization problem (7.50) is homogeneous, the equality constraint has been dropped. Moreover, since logarithm is a strictly increasing function and the objective function of the optimization problem (7.49) or, equivalently, (7.50) is positive, by taking logarithm of (7.50), the problem can be equivalently recast as

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g}} \ln \left( \ln \left( \frac{\mathbf{g}^H \mathbf{A}_1 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_1 \mathbf{g}} \right) \right) + \ln \left( \ln \left( \frac{\mathbf{g}^H \mathbf{A}_2 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_2 \mathbf{g}} \right) \right). \quad (7.51)$$

Using the fact that the problem (7.51) is homogeneous and by defining additional variables  $\alpha$  and  $\beta$ , (7.51) can be equivalently rewritten as

$$\begin{aligned} & \max_{\mathbf{g}, \alpha, \beta} \quad \ln \left( \ln \left( \mathbf{g}^H \mathbf{A}_1 \mathbf{g} \right) \right) + \ln \left( \ln(\alpha) - \ln(\beta) \right) \\ & \text{subject to} \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \\ & \quad \quad \quad \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta. \end{aligned} \quad (7.52)$$

Problem (7.52) is an special case of the generalized QCQP problem (3.1) that was introduced in Chapter 3 with only three quadratic constraints. Similar to the maximum sum-rate design problem and by defining the matrix  $\mathbf{X} = \mathbf{g} \mathbf{g}^H$  and based on

the Lemma 3.1 and Theorem 3.1, the problem (7.52) can be equivalently expressed as

$$\begin{aligned} & \max_{\mathbf{X}, \alpha, \beta} \quad \ln\left(\ln\left(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}\right)\right) + \ln\left(\ln(\alpha) - \ln(\beta)\right) \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \alpha, \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta. \end{aligned} \tag{7.53}$$

The optimal solution of the problem (7.52) can be easily extracted from the optimal solution of the problem (7.53) using Theorem 3.1. By defining the additional variable  $\gamma$ , the problem (7.53) can be finally recast as

$$\begin{aligned} & \max_{\mathbf{X}, \alpha, \beta, \gamma} \quad \ln\left(\ln\left(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}\right)\right) + \ln\left(\ln(\alpha) - \gamma\right) \\ \text{subject to} \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \alpha, \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \quad \ln(\beta) \leq \gamma. \end{aligned} \tag{7.54}$$

This is a DC programming problem and it has similar mathematical structure to the sum-rate maximization problem (7.45). It can be similarly handled using the POTDC algorithm. Specifically, the objective is concave and all the constraints except the last one are convex where the last one is a DC constraint. The point found by the POTDC algorithm is guaranteed to be a KKT point. However, based on our extensive simulation results for different scenarios, the POTDC algorithm actually finds the globally optimal solution of the problem (7.54). Indeed similar to the Subsection (6.2.1), the global optimality of this method reduces to the concavity of the one-dimensional optimal value function that is defined using the optimization problem (7.54) when  $\beta$  is fixed. The concavity of the optimal value function denoted as  $k(\beta)$  can be numerically checked. Note that similar to the sum-rate maximization problem, it can be shown that  $\beta$  takes values in a closed interval. Therefore,  $k(\beta)$  is defined only in a closed interval. The numerical check for the concavity of  $k(\beta)$  is based on a search for a counterexample. Specifically, since  $k(\beta)$  is defined over a closed interval  $[\theta_1, \theta_2]$ , the concavity can be checked probabilistically by generating three random points  $\beta_1$ ,  $\beta_2$ , and  $\theta$  over the intervals  $[\theta_1, \theta_2]$ ,  $[\theta_1, \theta_2]$ , and  $[0, 1]$ , respectively, and checking the validity of the inequality  $f(\theta\beta_1 + (1-\theta)\beta_2) < \theta f(\beta_1) + (1-\theta)f(\beta_2)$  in order to find a counterexample to concavity of  $k(\alpha)$ . If such a point can not be found over numerous realization of  $\beta_1$ ,  $\beta_2$ , and  $\theta$ , then it can be concluded that  $k(\beta)$  is most probably a concave function.

### 7.5.2 Max-min rate fairness

MMRF resource allocation/berfomer aims at maximizing the minimum received rate for each terminal subject to the total power constraint at the relay. The corresponding optimization problem can be expressed as

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g}} \min_{\mathbf{Q}} \left\{ \frac{1}{2} \ln \left( 1 + \frac{P_{R,1}}{\tilde{P}_{N,1}} \right), \frac{1}{2} \ln \left( 1 + \frac{P_{R,2}}{\tilde{P}_{N,2}} \right) \right\}. \quad (7.55)$$

Similar to the PF beamforming problem (7.49), this problem can be equivalently expressed as the following homogeneous problem

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g}} \min \left\{ \ln \left( \frac{\mathbf{g}^H \mathbf{A}_1 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_1 \mathbf{g}} \right), \ln \left( \frac{\mathbf{g}^H \mathbf{A}_2 \mathbf{g}}{\mathbf{g}^H \mathbf{B}_2 \mathbf{g}} \right) \right\}. \quad (7.56)$$

By defining the additional variables  $\alpha$  and  $\beta$  and using the fact that the problem (7.56) is homogeneous, (7.56) can be equivalently recast as

$$\begin{aligned} \max_{\alpha, \beta} \max_{\mathbf{g}} \min & \left\{ \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}), \ln(\alpha) - \ln(\beta) \right\} \\ \text{subject to} & \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \\ & \quad \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta. \end{aligned} \quad (7.57)$$

Exchanging the order of maximum and minimum in the objective of (7.57) can simplify this problem so that POTDC algorithm can then be directly applied to it. The following lemma considers the possibility of such exchange in the order of maximum and minimum.

**Lemma 7.2.** *For fixed values of  $\alpha$  and  $\beta$ , the following optimization problems have the same optimal values, i.e.,  $p_1 = p_2$ ,*

$$\begin{aligned} p_1 \triangleq & \max_{\mathbf{g}} \min \left\{ \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}), \ln(\alpha) - \ln(\beta) \right\} \\ \text{subject to} & \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \\ & \quad \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta \end{aligned} \quad (7.58)$$

and

$$\begin{aligned} p_2 \triangleq & \min \left\{ \max_{\mathbf{g}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}), \ln(\alpha) - \ln(\beta) \right\} \\ \text{subject to} & \quad \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \\ & \quad \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta. \end{aligned} \quad (7.59)$$

**Proof:** See Appendix H.

Using Lemma 7.2, the optimization problem (7.57) can be equivalently rewritten as

$$\begin{aligned} \max_{\alpha, \beta} \min \left\{ \max_{\mathbf{g}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}), \ln(\alpha) - \ln(\beta) \right\} \\ \text{subject to } \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \quad \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \\ \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta. \end{aligned} \quad (7.60)$$

The inner optimization problem in (7.60) is also a realization of the generalized QCQP problem (3.1) with only three quadratic constraints. Therefore, according to the Lemma 3.1 and Theorem 3.1 and by defining the matrix  $\mathbf{X} = \mathbf{g}\mathbf{g}^H$ , the problem (7.60) can be equivalently recast as

$$\begin{aligned} \max_{\alpha, \beta} \min \left\{ \max_{\mathbf{X}} \ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}), \ln(\alpha) - \ln(\beta) \right\} \\ \text{subject to } \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \alpha, \\ \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta \end{aligned} \quad (7.61)$$

where the optimal solution of the problem (7.60) can be extracted from the optimal solution of the problem (7.61) as it has been explained in Theorem 3.1. Therefore, these two problems are equivalent. Eventually by defining the additional variable  $t$ , the problem (7.61) can be recast as

$$\begin{aligned} \max_{\alpha, \beta, \mathbf{X}, t} \quad & t \\ \text{subject to } \quad & \text{tr}\{\mathbf{B}_1 \mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2 \mathbf{X}\} = \alpha, \\ & \text{tr}\{\mathbf{B}_2 \mathbf{X}\} = \beta, \quad \ln(\text{tr}\{\mathbf{A}_1 \mathbf{X}\}) \geq t, \\ & \ln(\alpha) - \ln(\beta) \geq t. \end{aligned} \quad (7.62)$$

The objective function is concave and all the constraints of the problem (7.62) except the last constraint are convex. The last constraint is a DC constraint. Thus, (7.62) can be handled by using POTDC algorithm as well.

## 7.6 Simulation results for sum-rate maximization problem

In this section, we evaluate the performance of the new proposed methods via numerical simulations. Consider a communication system consisting of two single antenna terminals and an AF MIMO relay with  $M_R$  antenna elements. The communication

between the terminals is bidirectional, i.e., it is performed based on the two-way relaying scheme. It is assumed that perfect channel knowledge is available at the terminals and at the relay, while the terminals use only effective channels (scalars), but the relay needs full channel vectors. The relay estimates the corresponding channel coefficients between the relay antenna elements and the terminals based on the pilots which are transmitted from the terminals. Then based on these channel vectors, the relay computes the relay amplification matrix  $\mathbf{G}$  and then uses it for forwarding the pilot signals to the terminals. After receiving the forwarded pilot signals from the relay via the effective channels, the terminals can estimate the effective channels using a suitable pilot-based channel estimation scheme, e.g., the LS.

The noise powers of the relays and the terminals  $P_{N,R}$ ,  $P_{N,1}$ , and  $P_{N,2}$  are assumed to be equal to  $\sigma^2$  unless otherwise specified. Uncorrelated Rayleigh fading channels are considered and it is assumed that reciprocity holds, i.e.,  $\mathbf{h}_i^{(f)} = \mathbf{h}_i^{(b)}$  for  $i = 1, 2$ . The relay is assumed to be located over a line of unit length which connects the terminals to each other and the variances of the channel coefficients between terminal  $i$  and the relay antenna elements are all assumed to be proportional to  $1/d_i^\nu$ , where  $d_i \in (0, 1)$  is the distance between the relay and the terminal  $i$  and  $\nu$  is the path-loss exponent which is assumed to be equal to 3 throughout the simulations.<sup>3</sup> For obtaining each point, 100 independent simulation runs are used unless otherwise is specified.

In order to design the relay amplification matrix  $\mathbf{G}$ , five different methods are considered including the proposed POTDC method for sum-rate maximization, 2-D RATE-maximization via Generalized Eigenvectors (RAGES) and 1-D RAGES algorithms [65], the algebraic norm-maximizing (ANOMAX) transmit strategy of [66] and the discrete Fourier transform (DFT) method that chooses the relay precoding matrix as a scaled DFT matrix. Note that the ANOMAX strategy provides a closed-form solution for the problem. Also note that for the DFT method no channel knowledge is needed. Thus, the DFT method serves as a benchmark for evaluating the gain achieved by using channel knowledge. The upper-bound is also shown in Simulation Examples 1 and 2. For obtaining the upper-bound, the interval  $[\beta_{\min}, \beta_{\max}]$  is divided in 30 segments. In addition, the proposed techniques are

---

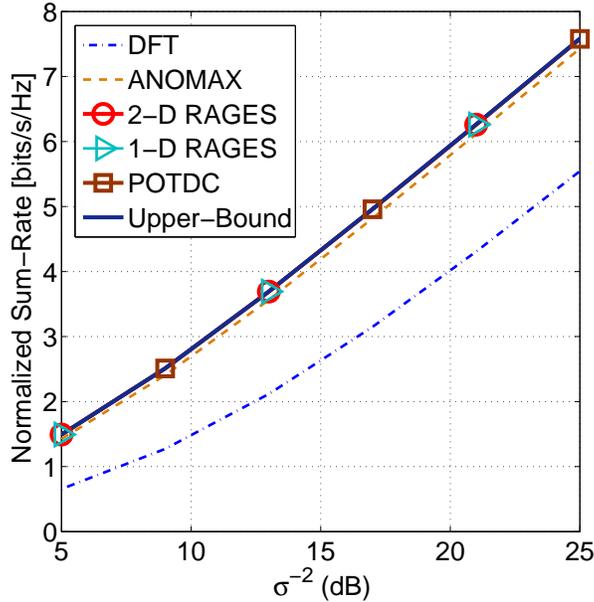
<sup>3</sup>It is experimentally found that typically  $2 \leq \nu \leq 6$  (see [81, p. 46–48] and references therein). However,  $\nu$  can be smaller than 2 when we have a wave-guide effect, i.e., indoors in corridors or in urban scenarios with narrow street canyons.

compared to the SNR-balancing technique of [69] for the scenario when multiple single antenna relay nodes are used and the method of [69] is applicable.

### 7.6.1 Example 1 : Symmetric channel conditions

In our first example, we consider the case when the channels between the relay antenna elements and both terminals have the same channel quality. More specifically, it is assumed that the relay is located in the middle of the connecting line between the terminals and the transmit powers  $P_{T,1}$  and  $P_{T,2}$  and the total transmit power of the MIMO relay  $P_{T,R}$  are all assumed to be equal to 1.

Fig. 7.4 shows the sum-rate achieved by different aforementioned methods versus  $\sigma^{-2}$  for the case of  $M_R = 3$ . It can be seen in this figure that the performance of the proposed method, 1-D RAGES, and 2-D RAGES coincide with the upper-bound. Thus, the proposed method, 1-D RAGES, and 2-D RAGES perform globally optimally in terms of providing the maximum sum-rate. The ANOMAX technique performs close to the optimal, while the DFT method gives a significantly lower sum-rate. It is noteworthy to mention that although the RAGES-based methods achieve the globally optimal solution, they are heuristic and can not be generalized.

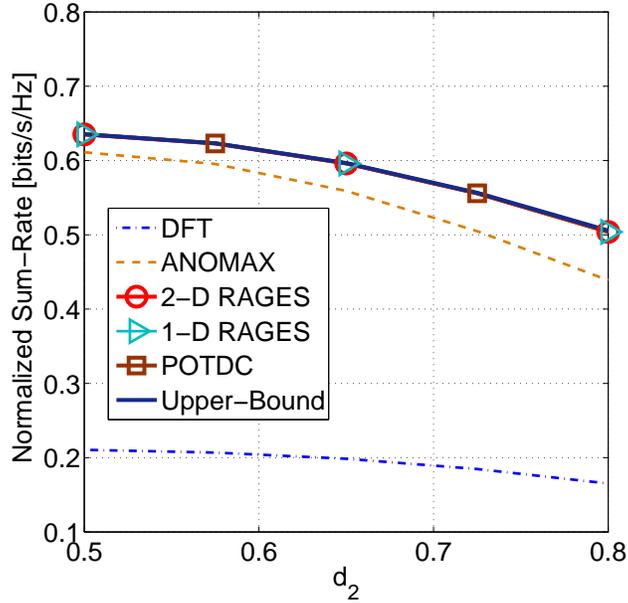


**Figure 7.4:** Example 1: The case of symmetric channel conditions. Sum-rate  $r_1 + r_2$  versus  $\sigma^{-2}$  for  $M_R = 3$  antennas.

### 7.6.2 Example 2 : Asymmetric channel conditions

In the second example, we consider the case when the channels between the relay antenna elements and the second terminal have better channel quality than the channels between the relay antenna elements and the first terminal. Thus, we evaluate the effect of the relay location on the achievable sum-rate. Particularly, we consider the case when the distance between the relay and the second terminal  $d_2$  is less than or equal to the distance between the relay and the first terminal  $d_1$ . The total transmit power of the terminals, i.e.,  $P_{T,1}$  and  $P_{T,2}$  and the total transmit power of the MIMO relay  $P_{T,R}$  all are assumed to be equal to 1 and the noise power in the relay antenna elements and the terminals all are assumed to be equal to 1.

Fig. 7.5 shows the sum-rate achieved in this scenario by different methods tested versus the distance between the relay and the second terminal  $d_2$ , for the case of  $M_R = 3$ . It can be seen in this figure that the proposed method and the RAGES-based methods perform optimally, while the performance (sum-rate) of ANOMAX is slightly worse.



**Figure 7.5:** Example 2: The case of asymmetric channel conditions. Sum-rate  $r_1 + r_2$  versus the distance between the relay and the second terminal  $d_2$  for  $M_R = 3$  antennas.

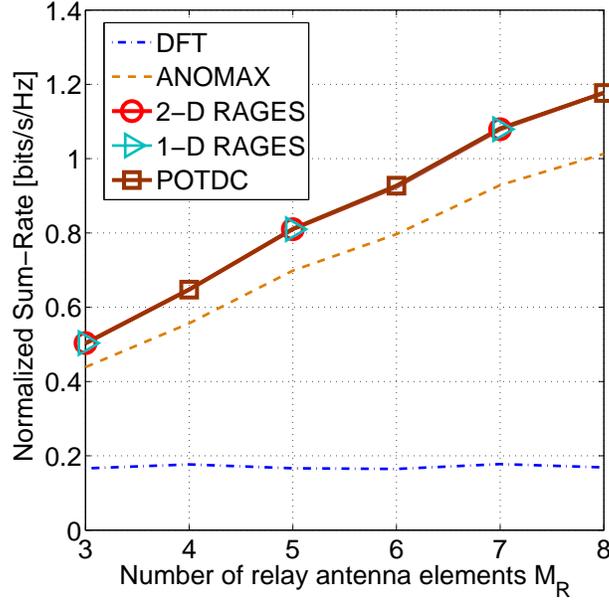
As mentioned earlier, it is guaranteed that the POTDC algorithm converges to at least a KKT point of the sum-rate maximization problem. However, our

extensive simulation results confirm that the POTDC algorithm converges to the global maximum of the problem in all simulation runs. It is approved by the fact that the performance of the POTDC algorithm coincides with the upper-bound. Moreover, the 2-D RAGES and 1-D RAGES are, in fact, globally optimal, too. Note that although the 2-D RAGES and 1-D RAGES achieve the globally optimal solution, they are largely heuristic methods and are applicable only to the sum-rate maximization in a two-way relaying system with two single antenna users. In other words, they can not be extended to problems of a more general form. For example, in contrast to our proposed method which can be easily applied to the multi-operator two-way relay networks, the RAGES is not applicable in that case [141]. The ANOMAX and DFT methods, however, do not achieve the maximum sum-rate. The loss in sum-rate related to the DFT method is quite significant while the loss in sum-rate related to the ANOMAX method grows from small in the case of symmetric channel conditions to significant in the case of asymmetric channel conditions. Although ANOMAX enjoys a closed-form solution and it is even applicable in the case when terminals have multiple antennas, it is not a good substitute for the proposed methods because of the significant gap in performance in the asymmetric case.

### 7.6.3 Example 3 : Effect of the number of relay antenna elements

In this example, we consider the effect of the number of relay antenna elements  $M_R$  on the achievable sum-rate for the aforementioned methods. The powers assigned to the first and second terminals as well as to the relay are all equal to 1. The relay is assumed to be located at the distance of  $1/5$  from the second user. Moreover, the noise powers at the terminals and at the relay antenna elements are all assumed to be equal to 1. For obtaining each point in this simulation example, 200 independent runs are used.

Fig. 7.6 depicts the sum-rates achieved by different methods versus the number of relay antenna elements  $M_R$ . As it is expected, by increasing  $M_R$  (thus, increasing the number of degrees of freedom), the sum-rate increases. For the DFT method, the sum-rate does not increase with the increase of  $M_R$  because of the lack of channel knowledge for this method. The proposed methods achieve higher sum-rate compared to ANOMAX.

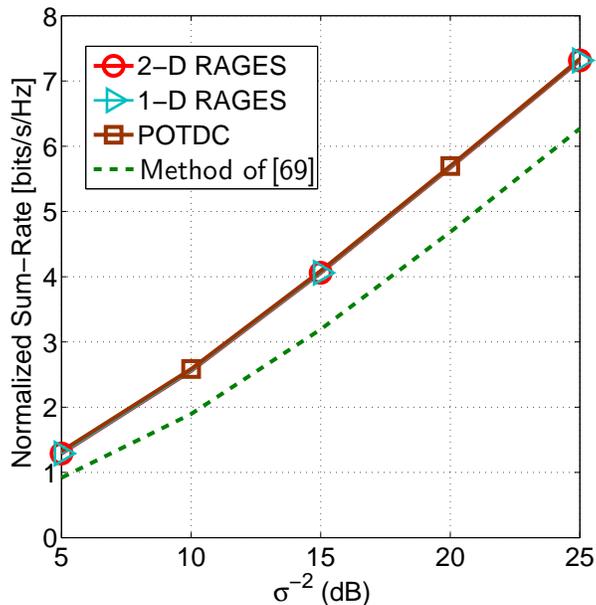


**Figure 7.6:** Example 3: The case of asymmetric channel conditions. Sum-rate  $r_1 + r_2$  versus the number of the relay antenna elements  $M_R$ .

#### 7.6.4 Example 4 : Performance comparison for the scenario of two-way relaying via multiple single antenna relays

In our last example, we compare the proposed POTDC method with the SNR balancing-based approach of [69]. The method of [69] is developed for two-way relaying systems which consist of two single antenna terminals and multiple single antenna relay nodes. Subject to the constraint on the total transmit power of the relay nodes and the terminals, the method of [69] designs a beamforming vector for the relay nodes and the transmit powers of the terminals to maximize the minimum received SNR at the terminals. In order to make a fair comparison, we consider a diagonal structure for the relay amplification matrix  $\mathbf{G}$  that corresponds to the case of [69] when multiple single antenna nodes are used for relaying. It is worth mentioning that for imposing such a diagonal structure for the relay amplification matrix  $\mathbf{G}$  in POTDC and RAGES, the vector  $\mathbf{g}_{M_R^2 \times 1} = \text{vec}\{\mathbf{G}\}$  is replaced with  $\mathbf{g}_{M_R \times 1} = \text{diag}\{\mathbf{G}\}$  and the matrices  $\mathbf{A}_i$  and  $\mathbf{B}_i, i = 1, 2$  are replaced with new square matrices  $\tilde{\mathbf{A}}_i$  and  $\tilde{\mathbf{B}}_i, i = 1, 2$  of size  $M_R \times M_R$  such that  $[\tilde{\mathbf{A}}_i]_{m,n} = [\mathbf{A}_i]_{(m-1) \cdot M_R + m, (n-1) \cdot M_R + n}$  and  $[\tilde{\mathbf{B}}_i]_{m,n} = [\mathbf{B}_i]_{(m-1) \cdot M_R + m, (n-1) \cdot M_R + n}, m, n = 1, \dots, M_R$ . Moreover, for the proposed POTDC method and the RAGES-based methods, we assume fixed transmit powers at the terminals and fixed total transmit power at the relay nodes that

are all equal to 1, while for the method of [69], the total transmit power at the relay nodes and the terminals is assumed to be equal to 3. Thus, the overall consumed power by the proposed POTDC method, the RAGES-based methods and the method of [69] is the same, however, compared to [69], which also optimizes the power usage of the terminals, the transmit powers of the terminals in the proposed POTDC method and the RAGES-based methods are fixed. The relay is assumed to lie in the middle in between the terminals. Fig. 7.7 shows the corresponding performance of the methods tested. From this figure it can be seen that the proposed POTDC method and the RAGES-based methods demonstrate a better performance compared to the method of [69] as it may be expected even though they use a fixed transmit power for the terminals.



**Figure 7.7:** Example 4: The case of symmetric channel conditions and distributed single antenna relays. Sum-rate  $r_1 + r_2$  versus  $\sigma^{-2}$  for  $M_R = 3$  antennas.

## 7.7 Simulation results for PF and MMRF relay amplification matrix design

For evaluating the performance of the proportionally fair and max-min rate fair relay amplification matrix design, we consider a similar simulation set up as in Section 7.6.

### 7.7.1 Example 1 : Symmetric channel conditions

More specifically, the same single antenna terminals are considered that are communicating through the AF MIMO relay equipped with  $M_R = 3$  antennas. The transmit powers of the terminals  $P_{T,1}$  and  $P_{T,2}$  and the total transmit power of the MIMO relay  $P_{T,R}$  are all assumed to be equal to 1. The relay is assumed to be located at the middle of the terminals while the path-loss parameter is selected to be equal to 2. The noise power for all the antenna elements is assumed to be the same, and it is denoted as  $\sigma^2$  except the second terminal whose noise power is 20 dB larger than for the rest of the noises. The difference in noise power is used for modeling the asymmetric environmental conditions for the users.

We compare the proposed PF and the max-min fairness beamforming methods with other relay beamforming methods in terms of the fairness index defined as  $(r_1^2 + r_2^2)/(2 \cdot (r_1 + r_2)^2)$  [142], [143] and the minimum user data rate, respectively. The methods used for performance comparison are the sum-rate maximization based relay amplification matrix design which was developed in Section 7.3, ANOMAX and DFT that were introduced in the previous section. Furthermore, the proposed relay beamforming methods are compared with upper-bound of the optimization problems (7.49) and (7.55) found in a similar way as in Section 7.4. Fig. 7.8 shows the fairness index for the PF beamforming method in comparison to that of the other methods tested versus  $\sigma^{-2}$ .

Fig. 7.9 shows the minimum data rate of the max-min fairness beamformer in comparison to that of the other aforementioned methods also versus  $\sigma^{-2}$ . From these figures, it can be seen that the proposed methods outperform the other state-of-art relay beamforming methods in the scenario with the noise power asymmetry at the terminals. Moreover, the POTDC algorithm is able to find global optimums of the corresponding optimization problems.

## 7.8 Chapter summery

We have shown that the sum-rate maximization problem in two-way AF MIMO relaying can be cast as a generalized QCQP problem. By means of the SDR, the corresponding optimization problem is recast precisely as a DC programming problem. Although DC programming problems can be solved using the branch-and-bound method, this method does not have any polynomial time guarantees for its worst-

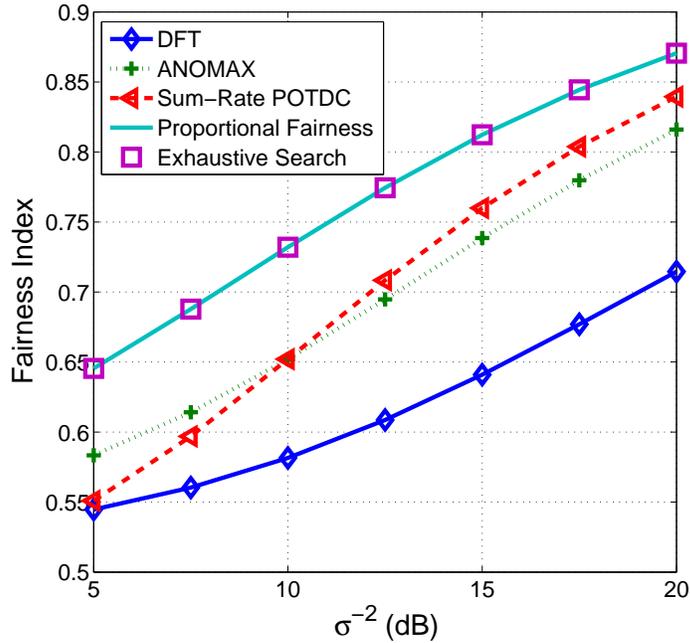


Figure 7.8: Example 1: Fairness index versus  $\sigma^{-2}$ .

case complexity. Therefore, we have applied the proposed POTDC algorithm for finding a KKT point for the aforementioned problem with polynomial time worst-case complexity. There is, however, a great evidence that the globally optimal solution is also achieved. Moreover, a tight upper-bound for the maximum possible sum-rate is developed and it is demonstrated by simulations that the solution obtained by the POTDC method achieves the upper-bound and is, thus, globally optimal.

Next, we considered the relay amplification matrix design problem based on the max-min rate and PF criteria. It is shown that these design problems can also be recast as DC programming problems as well which can be addressed using the POTDC algorithm. In application to the corresponding optimization problems, the POTDC algorithm finds the KKT point. Moreover, its global optimality in each specific case can be easily checked by the means of a simple numerical global optimality test that aims at ensuring that a certain one-dimensional optimal value function is convex.

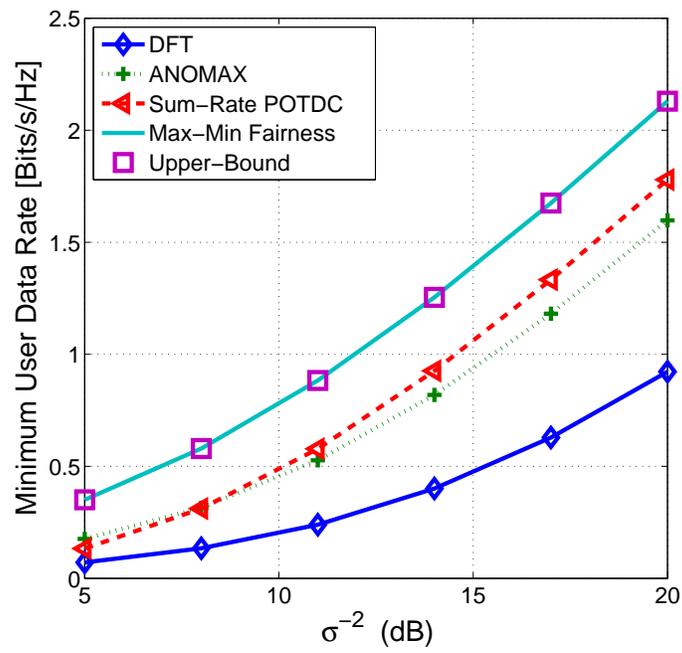


Figure 7.9: Example 1: Minimum data rate versus  $\sigma^{-2}$ .

# Chapter 8

## Conclusion

The contributions of this dissertation are briefly summarized in this chapter. Moreover, the potential future research directions are discussed at the end of the chapter.

### 8.1 Summary of contributions

In this thesis, we introduced a generalization of the QCQP optimization problem which appears in many signal processing and communications applications. More specifically, such generalization involves the composition of one-dimensional convex and quadratic functions in the constraint and the objective functions. It is shown that this generalization of the QCQP problem can be precisely or approximately recast as a DC optimization problem by using the semi-definite relaxation technique. Although the DC programming problems can be solved globally optimally using various modification of the so-called branch-and-bound methods, there is no guarantee of worst-case polynomial time complexity for these methods. For addressing this problem, which may prohibit the practical applicability of the branch-and-bound methods in the real-time applications, we proposed a new method for solving the resulted DC optimization problems at least suboptimally. Particularly, the new proposed method is based on the linearization and an iterative Newton-type search over a small set of optimization variables. The global optimality of the solution by the proposed method is discussed while a geometric interpenetration of the iterations is given. Next, we consider the application of the introduced generalized QCQP and the pure QCQP optimization problems for four different problems in array processing and wireless communications. The new contributions in the tackled problems comes in order.

### 8.1.1 Transmit beamspace design for DOA estimation in MIMO radar

We studied the MIMO radar transmit beamspace design problem which allows to achieve a virtual array with a large number of elements and at the same time to enjoy large SNR gains in different virtual elements. Specifically, we studied the design of the transmit beamspace matrix in such a way that its actual transmit beampattern is as close as possible to a desired one and the power is uniformly distributed across the transmit antenna elements which is important due to practical consideration. By a proper choice of the desired beampattern, this way of designing can result in large SNR gains through the focus of the transmit energy in the desired sectors where the targets are likely to be located. In addition to these design requirement and for allowing for a simple search-free DOA estimation algorithm at the receive array with arbitrary shape, the rotational invariance property (RIP) for the virtual array was also established by a proper design of transmit beamspace at the transmit array. More specifically, it was shown that by imposing a specific structure on the beamspace matrix the RIP can be enforced at the transmit array. For the case of even but otherwise arbitrary number of transmit waveforms, the design problem of such transmit beamspace matrix is formulated as non-convex QCQP problem. Then, this non-convex problem is approximately solved by using the SDR techniques. Numerical results confirm the superiority of the proposed transmit beamspace design in DOA estimation as compared to the traditional MIMO radar techniques.

### 8.1.2 A new robust adaptive beamforming

We studied the MVDR RAB design problem from a new perspective. Specifically, we showed that all the MVDR RAB methods can be viewed based on a single unified principle, i.e., to use standard MVDR beamforming in tandem with an estimate of the signal steering vector found based on some prior information. According to this unified principle, the differences between different MVDR RAB methods can be attributed to the difference in the available prior information as well as the estimation method which is applied. By means of this unified principle, we developed a new MVDR RAB technique which uses only imprecise knowledge of antenna array geometry and angular sector in which the actual steering vector is located. The new proposed RAB design is then expressed as a non-convex QCQP problem with only

two constraints which can be precisely solved by using some known results. However, we have developed a new algebraic method of finding the rank-one solution from the general-rank solution of the relaxed problem. Moreover, the condition under which the solution of the relaxed problem is guaranteed to be rank-one has been discussed. The numerical results confirm the superior performance of the new proposed method as opposed to the existing state-of-the-art RAB methods.

### **8.1.3 Robust adaptive beamforming for general-rank signal model with positive semi-definite constraint**

The RAB problem for the general-rank signal model with the additional PSD constraint was considered. The existing approaches for solving the corresponding non-convex optimization problem are iterative methods for which the convergence is not guaranteed and they solve the problem only suboptimally. These shortcomings motivated us to find a better method. Specifically, we were able to show that the aforementioned non-convex problem belongs to the class of generalized QCQP optimization problems which was introduced in Chapter 3. Using the newly developed POTDC method, we designed a polynomial time algorithm for addressing the corresponding optimization problem. It is guaranteed that the point found by the POTDC algorithm for this problem is a KKT point. Moreover, there is a strong evidence that the POTDC method actually finds the globally optimal point of the problem under consideration, which is shown in terms of a number of lemmas and one theorem. Furthermore, we demonstrated the equivalence between the claim of global optimality for the POTDC algorithm as applied to the problem under consideration and the convexity of the one-dimensional optimal value function (6.17). As opposed to the other existing methods, the proposed RAB method shows superior performance in terms of the output SINR.

### **8.1.4 Two-way AF MIMO relay transmit strategy design**

We considered the two-way relaying transmit strategy design problems based on the maximum sum-rate, MMRF, and the PF criteria. Such design problems were shown to belong to the class of generalized QCQP optimization problems which can be precisely recast as DC programming problems. Although DC programming problems can be solved by the branch-and-bound method, this method does not have any polynomial time complexity (worst-case) guarantees. Therefore, we have

applied the proposed POTDC algorithm for finding a KKT point of the aforementioned problems with polynomial time worst-case complexity. There is, however, a great evidence that the globally optimal solution is also achieved. Moreover, tight upper-bounds for the corresponding optimization problems have been developed and it is demonstrated by simulations that the solutions obtained by the POTDC method achieve the upper-bounds and are, thus, globally optimal for each criterion considered. The claim of global optimality in application to the these problems is equivalent to the concavity of one-dimensional optimal value functions which can be checked numerically.

### **8.1.5 Probable future research**

#### **Global optimality of the POTDC method**

In application to the general-rank RAB and the relay amplification matrix design, we have always observed that the proposed POTDC method results in the globally optimal solution of the corresponding optimization problems. Moreover, it was demonstrated that the claim of global optimality for the POTDC method is equivalent to the convexity/concavity of a certain optimal value function. Despite its very significant effect on the related fields, a rigorous proof of the convexity/concavity for the aforementioned optimal value function remains in general still an open problem of significant interest.

#### **Other applications of the POTDC method**

We studied the optimal relay amplification matrix design for an AF TWR system with two terminals equipped with a single antenna and one relay with multiple antennas. An interesting research direction is the extension of this basic model in different ways. One particular example of such an extension is the optimal amplification matrix design when the users are allowed to have multiple antennas and there is a direct link between the users.

# Appendix A

## Proof of Theorem 5.1

Let  $\mathbf{A}^*$  be the optimal minimizer of the relaxed problem (5.19), and its rank be  $r$ . Consider the following decomposition of  $\mathbf{A}^*$

$$\mathbf{A}^* = \mathbf{Y}\mathbf{Y}^H \quad (\text{A.1})$$

where  $\mathbf{Y}$  is an  $M \times r$  complex valued full rank matrix. It is trivial to see that if the rank of the optimal minimizer of the relaxed problem (5.19), i.e.,  $\mathbf{A}^*$ , is one, then  $\mathbf{Y}$  is also the optimal minimizer of the original problem (5.14). Thus, it is assumed in the following that  $r > 1$ .

Following similar steps as in [144], we start by considering the following auxiliary optimization problem

$$\begin{aligned} \min_{\mathbf{G}} \quad & \text{tr}\{\hat{\mathbf{R}}^{-1}\mathbf{Y}\mathbf{G}\mathbf{Y}^H\} \\ \text{subject to} \quad & \text{tr}\{\mathbf{Y}\mathbf{G}\mathbf{Y}^H\} = M \\ & \text{tr}\{\tilde{\mathbf{C}}\mathbf{Y}\mathbf{G}\mathbf{Y}^H\} = \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\} \\ & \mathbf{G} \succeq 0 \end{aligned} \quad (\text{A.2})$$

where  $\mathbf{G}$  is an  $r \times r$  Hermitian matrix. The matrix  $\mathbf{A}$  in (5.19) can be expressed as a function of the matrix  $\mathbf{G}$  in (A.2) as  $\mathbf{A}(\mathbf{G}) = \mathbf{Y}\mathbf{G}\mathbf{Y}^H$ . Moreover, it can be easily shown that if  $\mathbf{G}$  is a positive semi-definite matrix, then  $\mathbf{A}(\mathbf{G})$  is also a positive semi-definite matrix. In addition, if  $\mathbf{G}$  is a feasible point of the problem (A.2),  $\mathbf{A}(\mathbf{G})$  is also a feasible point of (5.19). The latter is true because  $\mathbf{A}(\mathbf{G})$  is a positive semi-definite matrix and it satisfies the constraints  $\text{tr}\{\mathbf{A}(\mathbf{G})\} = M$  and  $\text{tr}\{\tilde{\mathbf{C}}\mathbf{A}(\mathbf{G})\} = \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\} \leq \Delta_0$ . This implies that the minimum value of the problem (A.2) is greater than or equal to the minimum value of the problem (5.19).

It is then easy to verify that  $\mathbf{G}^* = \mathbf{I}_{r \times r}$  is a feasible point of the auxiliary optimization problem (A.2). Moreover, the fact that  $\text{tr}\{\hat{\mathbf{R}}^{-1}\mathbf{Y}\mathbf{G}^*\mathbf{Y}^H\} = \text{tr}\{\hat{\mathbf{R}}^{-1}\mathbf{A}^*\} \triangleq$

$\beta$  (here  $\beta$  denotes the minimum value of the relaxed problem (5.19)) together with the fact that the minimum value of the auxiliary problem (A.2) is greater than or equal to  $\beta$ , implies that  $\mathbf{G}^* = \mathbf{I}_{r \times r}$  is the optimal solution of the auxiliary problem (A.2).

In what follows, we show that every feasible point of the problem (A.2) denoted as  $\mathbf{G}'$  is also an optimal minimizer of the same problem. Thus,  $\mathbf{A}(\mathbf{G}') = \mathbf{Y}\mathbf{G}'\mathbf{Y}^H$  is an optimal minimizer of (5.19).

Let us start by considering the following dual problem of (A.2)

$$\begin{aligned} & \max_{\nu_1, \nu_2, \mathbf{Z}} && \nu_1 M + \nu_2 \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\} \\ \text{subject to} &&& \mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} - \nu_1 \mathbf{Y}^H \mathbf{Y} - \nu_2 \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \succeq \mathbf{Z} \\ &&& \mathbf{Z} \succeq \mathbf{0} \end{aligned} \tag{A.3}$$

where  $\nu_1$  and  $\nu_2$  are the Lagrangian multipliers associated with the first and second constraints in (A.2), respectively, and  $\mathbf{Z}$  is an  $r \times r$  Hermitian matrix of Lagrangian multipliers associated with the constraint  $\mathbf{G} \succeq \mathbf{0}$ . The problem (A.2) is convex, and it satisfies the Slater's condition because, as it was mentioned, the positive definite matrix  $\mathbf{G} = \mathbf{I}_{r \times r}$  is a strictly feasible point for (A.2). Thus, the strong duality between (A.2) and (A.3) holds.

Let  $\nu_1^*$ ,  $\nu_2^*$ , and  $\mathbf{Z}^*$  be one possible optimal solution of the dual problem (A.3). Since strong duality holds, we can state that  $\nu_1^* M + \nu_2^* \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\} = \beta$ . Considering the fact that  $\mathbf{I}_{r \times r}$  is the optimal solution of the primal problem (A.2), the complementary slackness condition implies that

$$\text{tr}\{\mathbf{G}^* \mathbf{Z}^*\} = \text{tr}\{\mathbf{Z}^*\} = 0. \tag{A.4}$$

Since  $\mathbf{Z}^*$  is a positive semi-definite matrix, the condition (A.4) implies that  $\mathbf{Z}^* = \mathbf{0}$ . Then it follows from the first constraint of (A.3) that

$$\mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} - \nu_1^* \mathbf{Y}^H \mathbf{Y} - \nu_2^* \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \succeq \mathbf{0}. \tag{A.5}$$

The fact that  $\beta - \nu_1^* M - \nu_2^* \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\} = 0$  implies that  $\text{tr}\{\mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} - \nu_1^* \mathbf{Y}^H \mathbf{Y} - \nu_2^* \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}\} = 0$ . As a result, it can be easily verified that the constraint (A.5) is active, i.e., it is satisfied as equality for optimal  $\nu_1^*$  and  $\nu_2^*$ . Therefore, we can write that

$$\mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} = \nu_1^* \mathbf{Y}^H \mathbf{Y} + \nu_2^* \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}. \tag{A.6}$$

Let  $\mathbf{G}'$  be another feasible point of (A.2) different from  $\mathbf{I}_{r \times r}$ . Then the following conditions must hold

$$\text{tr}\{\mathbf{Y}^H \mathbf{Y} \mathbf{G}'\} = M \quad (\text{A.7})$$

$$\text{tr}\{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{G}'\} = \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\} \quad (\text{A.8})$$

$$\mathbf{G}' \succeq 0. \quad (\text{A.9})$$

Multiplying both sides of the equation (A.6) by  $\mathbf{G}'$ , we obtain

$$\mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} \mathbf{G}' = \nu_1^* \mathbf{Y}^H \mathbf{Y} \mathbf{G}' + \nu_2^* \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{G}'. \quad (\text{A.10})$$

Moreover, taking the trace of the right hand and left hand sides of (A.10), we have

$$\begin{aligned} \text{tr}\{\mathbf{Y}^H \hat{\mathbf{R}}^{-1} \mathbf{Y} \mathbf{G}'\} &= \nu_1^* \text{tr}\{\mathbf{Y}^H \mathbf{Y} \mathbf{G}'\} + \nu_2^* \text{tr}\{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{G}'\} \\ &= \nu_1^* M + \nu_2^* \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\} = \beta. \end{aligned} \quad (\text{A.11})$$

This implies that  $\mathbf{G}'$  is also a possible optimal solution of (A.2). Therefore, every feasible point of (A.2) is also a possible optimal solution.

Finally, we show that there exists a feasible point of (A.2) whose rank is one. As it has been proved above, such feasible point is also a possible optimal solution. Let  $\mathbf{H} \triangleq \mathbf{v} \mathbf{v}^H$ , and we are interested in finding such  $\mathbf{v}$  that

$$\text{tr}\{\mathbf{Y}^H \mathbf{Y} \mathbf{H}\} = \mathbf{v}^H \mathbf{Y}^H \mathbf{Y} \mathbf{v} = M \quad (\text{A.12})$$

$$\text{tr}\{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{H}\} = \mathbf{v}^H \mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y} \mathbf{v} = \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\}. \quad (\text{A.13})$$

Equivalently, the conditions (A.12) and (A.13) can be rewritten as

$$\frac{1}{M} \mathbf{v}^H \mathbf{Y}^H \mathbf{Y} \mathbf{v} = 1 \quad (\text{A.14})$$

$$\mathbf{v}^H \frac{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}}{\text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\}} \mathbf{v} = 1. \quad (\text{A.15})$$

Moreover, equating the left hand side of (A.14) to the left hand side of (A.15), we obtain that

$$\frac{1}{M} \mathbf{v}^H \mathbf{Y}^H \mathbf{Y} \mathbf{v} = \mathbf{v}^H \frac{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}}{\text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\}} \mathbf{v}. \quad (\text{A.16})$$

Finding the difference between the left and right hand sides of (A.16), we also obtain that

$$\mathbf{v}^H \left( \frac{1}{M} \mathbf{Y}^H \mathbf{Y} - \frac{\mathbf{Y}^H \tilde{\mathbf{C}} \mathbf{Y}}{\text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\}} \right) \mathbf{v} = \mathbf{v}^H \mathbf{D} \mathbf{v} = 0. \quad (\text{A.17})$$

Considering the fact that  $\text{tr}\{\mathbf{Y}^H\mathbf{Y}\} = M$  and  $\text{tr}\{\mathbf{Y}^H\tilde{\mathbf{C}}\mathbf{Y}\} = \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\}$ , we find that  $\text{tr}\{\mathbf{Y}^H\mathbf{Y}/M - \mathbf{Y}^H\tilde{\mathbf{C}}\mathbf{Y}/\text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\}\} = \text{tr}\{\mathbf{D}\} = 0$ . Therefore, the vector  $\mathbf{v}$  can be chosen proportional to the sum of the eigenvectors of the matrix  $\mathbf{D}$  and it can be scaled so that  $\mathbf{v}^H\mathbf{Y}^H\tilde{\mathbf{C}}\mathbf{Y}\mathbf{v} = \text{tr}\{\tilde{\mathbf{C}}\mathbf{A}^*\}$  and, thus, (A.12) and (A.13) are satisfied.

So far we have found a rank-one solution for the auxiliary optimization problem (A.2), that is,  $\mathbf{G} = \mathbf{v}\mathbf{v}^H$ . Since  $\mathbf{G} = \mathbf{v}\mathbf{v}^H$  is a possible optimal solution of the auxiliary problem (A.2), then  $\mathbf{Y}\mathbf{G}\mathbf{Y}^H = (\mathbf{Y}\mathbf{v})(\mathbf{Y}\mathbf{v})^H$  is a possible optimal solution of the relaxed problem (5.19). Moreover, since the solution  $(\mathbf{Y}\mathbf{v})(\mathbf{Y}\mathbf{v})^H$  is rank-one,  $\mathbf{Y}\mathbf{v}$  is a possible optimal solution of the original optimization problem (5.14). This completes the proof.  $\square$

# Appendix B

## Proof of Lemma 5.1

Let  $\mathbf{A}^*$  be one possible optimal solution of the problem (5.19) whose rank  $r$  is greater than one. Using the rank-one decomposition of Hermitian matrices [145], the matrix  $\mathbf{A}^*$  can be written as

$$\mathbf{A}^* = \sum_{j=1}^r \mathbf{z}_j \mathbf{z}_j^H \quad (\text{B.1})$$

where

$$\mathbf{z}_j^H \mathbf{z}_j = \frac{1}{r} \text{tr}\{\mathbf{A}^*\} = \frac{M}{r}, \quad j = 1, \dots, r \quad (\text{B.2})$$

$$\mathbf{z}_j^H \tilde{\mathbf{C}} \mathbf{z}_j = \frac{1}{r} \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\}, \quad j = 1, \dots, r. \quad (\text{B.3})$$

Let us show that the terms  $\mathbf{z}_j^H \hat{\mathbf{R}}^{-1} \mathbf{z}_j$ ,  $j = 1, \dots, r$  are equal to each other for all  $j = 1, \dots, r$ . We prove it by contradiction assuming first that there exist such  $\mathbf{z}_m$  and  $\mathbf{z}_n$ ,  $m \neq n$  that  $\mathbf{z}_m^H \hat{\mathbf{R}}^{-1} \mathbf{z}_m < \mathbf{z}_n^H \hat{\mathbf{R}}^{-1} \mathbf{z}_n$ . Let the matrix  $\mathbf{A}_0^*$  be constructed as  $\mathbf{A}_0^* = \mathbf{A}^* - \mathbf{z}_n \mathbf{z}_n^H + \mathbf{z}_m \mathbf{z}_m^H$ . It is easy to see that  $\text{tr}\{\mathbf{A}^*\} = \text{tr}\{\mathbf{A}_0^*\}$  and  $\text{tr}\{\tilde{\mathbf{C}} \mathbf{A}^*\} = \text{tr}\{\tilde{\mathbf{C}} \mathbf{A}_0^*\}$ , which means that  $\mathbf{A}_0^*$  is also a feasible point of the problem (5.19). However, based on our assumption that  $\mathbf{z}_m^H \hat{\mathbf{R}}^{-1} \mathbf{z}_m < \mathbf{z}_n^H \hat{\mathbf{R}}^{-1} \mathbf{z}_n$ , it can be concluded that  $\text{tr}\{\hat{\mathbf{R}}^{-1} \mathbf{A}_0^*\} < \text{tr}\{\hat{\mathbf{R}}^{-1} \mathbf{A}^*\}$ , which is obviously a contradiction. Thus, all terms  $\mathbf{z}_j^H \hat{\mathbf{R}}^{-1} \mathbf{z}_j$ ,  $j = 1, \dots, r$  must take the same value. Using this fact together with the equations (B.2) and (B.3), we can conclude that  $r \mathbf{z}_j \mathbf{z}_j^H$  for any  $j = 1, \dots, r$  is a possible optimal solution of the relaxed problem (5.19) which has rank one. Thus, the optimal solution of the original problem (5.14) is  $\sqrt{r} \mathbf{z}_j$  for any  $j = 1, \dots, r$ . Since the vectors  $\mathbf{z}_j$ ,  $j = 1, \dots, r$  in (B.1) are linearly independent and each of them represents a possible optimal solution of (5.14), we conclude that there are many possible optimal solution of (5.14). However, it contradicts the assumption that the optimal solution of (5.14) is unique (regardless possible phase rotation

as explained earlier in the Subsection 5.3.2). Thus, the only optimal solution  $\mathbf{A}^*$  of the relaxed problem (5.19) has to be rank-one. This completes the proof.  $\square$

# Appendix C

## Proof of Lemma 6.1

Since the objective function as well as the constraints of the optimization problem (6.2) are all quadratic functions of  $\Delta$ , this problem is convex. It is easy to verify that this problem satisfies the Slater's constraint qualification and as a result the KKT conditions are necessary and sufficient optimality conditions. Let us introduce the Lagrangian as

$$L(\mathbf{Q}, \Delta, \mu) \triangleq \mathbf{w}^H (\mathbf{Q}^H \mathbf{Q} + \mathbf{Q}^H \Delta + \Delta^H \mathbf{Q} + \Delta^H \Delta) \mathbf{w} + \mu(\|\Delta\|^2 - \eta^2) \quad (\text{C.1})$$

where  $\mu$  is the non-negative Lagrangian multiplier. The KKT optimality conditions are

$$\nabla_{\Delta} L(\mathbf{Q}, \Delta, \mu) = \mathbf{0} \quad (\text{C.2a})$$

$$\|\Delta\|^2 \leq \eta^2 \quad (\text{C.2b})$$

$$\mu(\|\Delta\|^2 - \eta^2) = 0 \quad (\text{C.2c})$$

$$\mu \geq 0 \quad (\text{C.2d})$$

where  $\mathbf{0}$  is the vector of zeros. Using the matrix differentiation, the zero gradient condition (C.2a) can be expressed as  $\mathbf{Q}\mathbf{w}\mathbf{w}^H + \Delta\mathbf{w}\mathbf{w}^H + \mu\Delta = \mathbf{0}$  or, equivalently, as

$$\Delta = -\mathbf{Q}\mathbf{w}\mathbf{w}^H(\mathbf{w}\mathbf{w}^H + \mu\mathbf{I})^{-1}. \quad (\text{C.3})$$

Moreover, using the matrix inversion lemma [132], the expression (C.3) can be simplified as

$$\Delta = -\frac{\mathbf{Q}\mathbf{w}\mathbf{w}^H}{\|\mathbf{w}\|^2 + \mu}. \quad (\text{C.4})$$

The Lagrangian multiplier  $\mu$  can be determined based on the conditions (C.2b)–(C.2d). For this goal, we find a simpler expression for the norm of the matrix

$\mathbf{Q}\mathbf{w}\mathbf{w}^H$  as follows

$$\begin{aligned}
\|\mathbf{Q}\mathbf{w}\mathbf{w}^H\|^2 &= \text{tr}\{\mathbf{Q}\mathbf{w}\mathbf{w}^H\mathbf{w}\mathbf{w}^H\mathbf{Q}^H\} \\
&= \text{tr}\{\mathbf{Q}\mathbf{w}\mathbf{w}^H\mathbf{Q}^H\} \cdot \mathbf{w}^H\mathbf{w} \\
&= \|\mathbf{Q}\mathbf{w}\|^2\|\mathbf{w}\|^2
\end{aligned} \tag{C.5}$$

Using (C.5), it can be obtained that

$$\delta(\mu) \triangleq \|\Delta\| = \frac{\|\mathbf{Q}\mathbf{w}\|\|\mathbf{w}\|}{\|\mathbf{w}\|^2 + \mu}. \tag{C.6}$$

where the new function  $\delta(\mu)$  is defined for notation simplicity. It is easy to verify that  $\delta(\mu)$  is a strictly decreasing function with respect to  $\mu \geq 0$ . Consequently, for any arbitrary  $\mu \geq 0$ , it is true that  $\delta(\mu) \leq \delta(0) = \|\mathbf{Q}\mathbf{w}\|/\|\mathbf{w}\|$ . Depending on whether  $\delta(0)$  is less than or equal to  $\eta$  or not, the following two cases are possible. If  $\delta(0) \leq \eta$ , then  $\mu$  and  $\Delta$  can be found as  $\mu = 0$  and  $\Delta = -\mathbf{Q}\mathbf{w}\mathbf{w}^H/\|\mathbf{w}\|^2$ , which is obtained by simply substituting  $\mu = 0$  in (C.4). In this case, the KKT conditions (C.2b)–(C.2d) are obviously satisfied. In the other case, when  $\delta(0) > \eta$ , the above obtained  $\Delta$  for  $\mu = 0$  does not satisfy the condition (C.2b) because  $\|\Delta\| = \delta(0) > \eta$ . Since,  $\delta(\mu)$  is a strictly decreasing function with respect to  $\mu \geq 0$ , for satisfying (C.2b), the value of  $\mu$  must be strictly larger than zero and as a result the condition (C.2c) implies that  $\|\Delta\| = \eta$ . Note that if  $\mu > 0$  and  $\|\Delta\|$  obtained by substituting such  $\mu$  in (C.4) is equal to  $\eta$ , then the KKT conditions (C.2b)–(C.2d) are all satisfied. Thus, we need to find the value of  $\mu$  such that the corresponding  $\|\Delta\|$  is equal to  $\eta$ . By equating  $\delta(\mu)$  to  $\eta$ , it can be resulted that  $\mu_0 = \|\mathbf{w}\|/\eta \cdot (\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\|)$ . Considering the above two cases together, the optimal  $\Delta$  can be expressed as

$$\Delta = \begin{cases} -\eta \frac{\mathbf{Q}\mathbf{w}\mathbf{w}^H}{\|\mathbf{Q}\mathbf{w}\|\|\mathbf{w}\|}, & \|\mathbf{Q}\mathbf{w}\| \geq \eta\|\mathbf{w}\| \\ -\frac{\mathbf{Q}\mathbf{w}\mathbf{w}^H}{\|\mathbf{w}\|^2}, & \text{otherwise.} \end{cases} \tag{C.7}$$

Finally, substituting (C.7) in the objective function of the problem (6.2), the worst-case signal power for a fixed beamforming vector  $\mathbf{w}$  can be found as shown in (6.3).

□

# Appendix D

## Proof of Lemma 6.2

First, we verify whether the optimal solution of the optimization problem (6.5), or equivalently, the following problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w} \\ \text{subject to} \quad & \|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1. \end{aligned} \quad (\text{D.1})$$

is achievable or not.

Let  $\mathbf{w}_0$  denote any arbitrary feasible point of the problem (D.1). It is easy to see that if  $\mathbf{w}^H\mathbf{w} \geq \mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0/\lambda_{\min}\{\hat{\mathbf{R}} + \gamma\mathbf{I}\}$  then  $\mathbf{w}^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}$  is greater than or equal to  $\mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0$ . The latter implies that if the optimal solution is achievable, it lies inside the sphere of  $\mathbf{w}^H\mathbf{w} \leq \mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0/\lambda_{\min}\{\hat{\mathbf{R}} + \gamma\mathbf{I}\}$ . Based on this fact, the optimization problem (D.1) can be recast as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w} \\ \text{subject to} \quad & \|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1, \\ & \mathbf{w}^H\mathbf{w} \leq \mathbf{w}_0^H(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{w}_0/\lambda_{\min}\{\hat{\mathbf{R}} + \gamma\mathbf{I}\}. \end{aligned} \quad (\text{D.2})$$

Feasible set of the new added constraint of the problem (D.2) is bounded and closed. In what follows, we show that feasible set of the constraint  $\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1$  is also closed. To show it, it suffices to consider the complement of this constraint as follows

$$\sqrt{\mathbf{w}^H\mathbf{R}_s\mathbf{w}} - \eta\sqrt{\mathbf{w}^H\mathbf{w}} < 1 \quad (\text{D.3})$$

and prove that it is an open set. Let  $\mathbf{w}_1$  denote an arbitrary point which satisfies (D.3) and let us consider a sphere whose center is  $\mathbf{w}_1$  and with radius of  $\epsilon$ . The points inside this sphere form a set which is denoted as  $\mathcal{A} \triangleq \{\mathbf{w} \mid \|\mathbf{w} - \mathbf{w}_1\| \leq \epsilon\}$ . We aim at showing that for any  $\mathbf{w}_1$ , if  $\epsilon$  is chosen sufficiently small then the set

$\mathcal{A}$  lies inside the set defined by the constraint  $\sqrt{\mathbf{w}^H \mathbf{R}_s \mathbf{w}} - \eta \sqrt{\mathbf{w}^H \mathbf{w}} < 1$ . To this end, let us first find the minimum of the quadratic term  $\mathbf{w}^H \mathbf{w}$  over the set  $\mathcal{A}$ . By choosing  $\epsilon$  in such a way that  $\epsilon < \|\mathbf{w}_1\|$ , the minimum of this quadratic term over  $\mathcal{A}$  can be easily shown to be  $\mathbf{w}_1^H \mathbf{w}_1 (1 - \epsilon / \|\mathbf{w}_1\|)^2$ . Next, we also find an upper-bound for the maximum value of the quadratic term  $\mathbf{w}^H \mathbf{R}_s \mathbf{w}$  over the set  $\mathcal{A}$ . For this goal, let  $\mathbf{R}_s = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$  denote the eigenvalue decomposition of the matrix  $\mathbf{R}_s$  where  $\mathbf{U}$  is a unitary matrix and the diagonal matrix  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_M\}$  contains the non-negative eigenvalues of  $\mathbf{R}_s$ . Using the eigenvalue decomposition, the term  $(\mathbf{w}_1 + \mathbf{w})^H \mathbf{R}_s (\mathbf{w}_1 + \mathbf{w})$  can be expressed as

$$(\mathbf{w}_1 + \mathbf{w})^H \mathbf{R}_s (\mathbf{w}_1 + \mathbf{w}) = \sum_{i=1}^M \lambda_i |\mathbf{u}_i^H (\mathbf{w}_1 + \mathbf{w})|^2. \quad (\text{D.4})$$

Based on the triangular and Cauchy-Schwartz inequalities, it can be concluded that  $|\mathbf{u}_i^H (\mathbf{w}_1 + \mathbf{w})| \leq |\mathbf{u}_i^H \mathbf{w}_1| + \epsilon \|\mathbf{u}_i\|^2 = |\mathbf{u}_i^H \mathbf{w}_1| + \epsilon$ . Using the latter fact and also the equation (D.4), it is resulted that

$$\begin{aligned} (\mathbf{w}_1 + \mathbf{w})^H \mathbf{R}_s (\mathbf{w}_1 + \mathbf{w}) &= \sum_{i=1}^M \lambda_i |\mathbf{u}_i^H (\mathbf{w}_1 + \mathbf{w})|^2 \\ &\leq \sum_{i=1}^M \lambda_i (|\mathbf{u}_i^H \mathbf{w}_1| + \epsilon)^2 \\ &\leq \mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1 (1 + 2\epsilon \frac{\sum_{i=1}^M \lambda_i |\mathbf{u}_i^H \mathbf{w}_1|}{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1}) + \mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1 \cdot \epsilon^2 \frac{\sum_{i=1}^M \lambda_i}{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1} \\ &\leq \mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1 (1 + \epsilon a_{\max})^2 \end{aligned} \quad (\text{D.5})$$

where  $a_{\max}$  is defined as

$$a_{\max} \triangleq \max \left( \sqrt{\frac{\sum_{i=1}^M \lambda_i}{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1}}, \frac{\sum_{i=1}^M \lambda_i |\mathbf{u}_i^H \mathbf{w}_1|}{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1} \right). \quad (\text{D.6})$$

Considering the so-obtained upper-bound and lower-bound, it can be concluded that

$$\begin{aligned} \sqrt{\mathbf{w}^H \mathbf{R}_s \mathbf{w}} - \eta \sqrt{\mathbf{w}^H \mathbf{w}} &\leq \\ (1 + \epsilon a_{\max}) \sqrt{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1} - (1 - \epsilon / \|\mathbf{w}_1\|) \eta \sqrt{\mathbf{w}_1^H \mathbf{w}_1}. \end{aligned} \quad (\text{D.7})$$

By choosing  $\epsilon$  so that

$$\epsilon < \frac{1 - (\sqrt{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1} - \eta \sqrt{\mathbf{w}_1^H \mathbf{w}_1})}{\sqrt{\mathbf{w}_1^H \mathbf{R}_s \mathbf{w}_1} a_{\max} + \eta \sqrt{\mathbf{w}_1^H \mathbf{w}_1} / \|\mathbf{w}_1\|} \quad (\text{D.8})$$

the right hand side of the inequality (D.7) becomes less than one and therefore the sphere with center of  $\mathbf{w}_1$  and radius given by (D.8) lies inside the set specified by the constraint  $\sqrt{\mathbf{w}^H \mathbf{R}_s \mathbf{w}} - \eta \sqrt{\mathbf{w}^H \mathbf{w}} < 1$ . Therefore, the feasible set of the constraint  $\|\mathbf{Q}\mathbf{w}\| - \eta\|\mathbf{w}\| \geq 1$  is closed. Since the feasible sets of both of the constraints are closed and one of them is bounded, the feasible set of the problem (D.2), which is the intersection of these two sets, is also closed and bounded. The latter implies that the feasible set of the problem (D.2) is compact. Therefore, also based on the fact that the objective function of (D.2) is continuous, the optimal solution of (D.2), or equivalently (6.5), is achievable.

Let  $(\mathbf{w}_{\text{opt}}, \alpha_{\text{opt}})$  denote the optimal solution of the problem (6.5). Let us define the following auxiliary optimization problem based on the problem (6.5)

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w} = \alpha_{\text{opt}} \\ & \mathbf{w}^H \mathbf{w} \leq \frac{(\sqrt{\alpha_{\text{opt}}} - 1)^2}{\eta^2}. \end{aligned} \quad (\text{D.9})$$

It can be seen that if  $\mathbf{w}$  is a feasible point of the problem (D.9), then the pair  $(\mathbf{w}, \alpha_{\text{opt}})$  is also a feasible point of the problem (6.5) which implies that the optimal value of the problem (D.9) is greater than or equal to that of (6.5). However, since  $\mathbf{w}_{\text{opt}}$  is a feasible point of the problem (D.9) and the value of the objective function at this feasible point is equal to the optimal value of the problem (6.5), i.e., it is equivalent to  $\mathbf{w}_{\text{opt}}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w}_{\text{opt}}$ , it can be concluded that both of the optimization problems (6.5) and (D.9) have the same optimal value. Let us define another auxiliary optimization problem based on the problem (D.9) as

$$\begin{aligned} g \triangleq \min_{\mathbf{w}} \quad & \mathbf{w}^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{Q}^H \mathbf{Q} \mathbf{w} = \alpha_{\text{opt}} \end{aligned} \quad (\text{D.10})$$

which is obtained from (D.9) by dropping the last constraint of the problem (D.9). The feasible set of the optimization problem (D.9) is a subset of the feasible set of the optimization problem (D.10). As a result, the optimal value  $g$  of the problem (D.10) is smaller than or equal to the optimal value of the problem (D.9), and thus also, the optimal value of the problem (6.5). Using the minimax Theorem [124], it is easy to verify that  $g = \alpha_{\text{opt}} / \lambda_{\max} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1} \mathbf{Q}^H \mathbf{Q} \right\}$ . Since  $g$  is smaller than or equal to the optimal value of the problem (6.5), it is upper-bounded by

$\mathbf{w}_0^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w}_0$ , where  $\mathbf{w}_0$  is an arbitrary feasible point of (6.5). The latter implies that  $\alpha_{\text{opt}} \leq \lambda_{\max} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1} \mathbf{Q}^H \mathbf{Q} \right\} \mathbf{w}_0^H (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{w}_0$ .  $\square$

# Appendix E

## Proof of Lemma 6.3

First, we prove that  $l(\alpha, \alpha_c)$  (6.21) is a convex function with respect to  $\alpha$ . For this goal, let  $\mathbf{W}_{\alpha_1}$  and  $\mathbf{W}_{\alpha_2}$  denote the optimal solutions of the optimization problems of  $l(\alpha_1, \alpha_c)$  and  $l(\alpha_2, \alpha_c)$ , respectively, i.e.,  $l(\alpha_1, \alpha_c) = \text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}_{\alpha_1}\}$  and  $l(\alpha_2, \alpha_c) = \text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}_{\alpha_2}\}$ , where  $\alpha_1$  and  $\alpha_2$  are any two arbitrary points in the interval  $[\theta_1, \theta_2]$ . It is trivial to verify that  $\theta\mathbf{W}_{\alpha_1} + (1 - \theta)\mathbf{W}_{\alpha_2}$  is a feasible point of the corresponding optimization problem of  $l(\theta\alpha_1 + (1 - \theta)\alpha_2, \alpha_c)$  (see the definition (6.21)). Therefore,

$$\begin{aligned}
 l(\theta\alpha_1 + (1 - \theta)\alpha_2, \alpha_c) & \\
 & \leq \text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})(\theta\mathbf{W}_{\alpha_1} + (1 - \theta)\mathbf{W}_{\alpha_2})\} \\
 & = \theta\text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}_{\alpha_1}\} \\
 & \quad + (1 - \theta)\text{tr}\{(\hat{\mathbf{R}} + \gamma\mathbf{I})\mathbf{W}_{\alpha_2}\} \\
 & = \theta l(\alpha_1, \alpha_c) + (1 - \theta)l(\alpha_2, \alpha_c)
 \end{aligned} \tag{E.1}$$

which proves that  $l(\alpha, \alpha_c)$  is a convex function with respect to  $\alpha$ .

In order to show that  $l(\alpha, \alpha_c)$  is greater than or equal to  $k(\alpha)$ , it suffices to show that the feasible set of the optimization problem of  $l(\alpha, \alpha_c)$  is a subset of the feasible set of the optimization problem of  $k(\alpha)$ . Let  $\mathbf{W}_1$  denote a feasible point of the optimization problem of  $l(\alpha, \alpha_c)$ , it is easy to verify that  $\mathbf{W}_1$  is also a feasible point of the optimization problem of  $k(\alpha)$  if the inequality  $\sqrt{\alpha} \leq \sqrt{\alpha_c} + \frac{\alpha - \alpha_c}{2\sqrt{\alpha_c}}$  holds. This inequality can be rearranged as

$$(\sqrt{\alpha} - \sqrt{\alpha_c})^2 \geq 0 \tag{E.2}$$

and it is valid for any arbitrary  $\alpha$ . Therefore,  $\mathbf{W}_1$  is also a feasible point of the optimization problem of  $k(\alpha)$  which implies that  $l(\alpha, \alpha_c) \geq k(\alpha)$ .

In order to show that the right and left derivatives are equal, we use the result of [146, Theorem 10] which gives expressions for the directional derivatives of a parametric SDP. Specifically, in [146, Theorem 10] the directional derivatives for the following optimal value function

$$\psi(\mathbf{u}) \triangleq \{\min_{\mathbf{y}} f(\mathbf{y}, \mathbf{u}) \mid \mathbf{G}(\mathbf{y}, \mathbf{u}) \preceq \mathbf{0}_{n \times n}\} \quad (\text{E.3})$$

are derived, where  $f(\mathbf{y}, \mathbf{u})$  and  $\mathbf{G}(\mathbf{y}, \mathbf{u})$  are a scalar and an  $n \times n$  matrix, respectively,  $\mathbf{y} \in \mathbb{R}^m$  is the optimization variables and  $\mathbf{u} \in \mathbb{R}^k$  is the optimization parameters. Let  $\mathbf{u}_c$  be an arbitrary fixed point. If the optimization problem of  $\psi(\mathbf{u}_c)$  poses certain properties, then according to [146, Theorem 10] it is directionally differentiable at  $\mathbf{u}_c$ . These properties are (i) the functions  $f(\mathbf{y}, \mathbf{u})$  and  $\mathbf{G}(\mathbf{y}, \mathbf{u})$  are continuously differentiable, (ii) the optimization problem of  $\psi(\mathbf{u}_c)$  is convex, (iii) the set of optimal solutions of the optimization problem of  $\psi(\mathbf{u}_c)$  denoted as  $\mathcal{M}$  is nonempty and bounded, (iv) the Slater condition for the optimization problem of  $\psi(\mathbf{u}_c)$  holds true, and (v) the *inf-compactness* condition is satisfied. Here inf-compactness condition refers to the condition of the existence of  $\alpha > \psi(\mathbf{u}_c)$  and a compact set  $S \subset \mathbb{R}^m$  such that  $\{\mathbf{y} \mid f(\mathbf{y}, \mathbf{u}) \leq \alpha, \mathbf{G}(\mathbf{y}, \mathbf{u}) \preceq \mathbf{0}\} \subset S$  for all  $\mathbf{u}$  in a neighborhood of  $\mathbf{u}_c$ . If for all  $\mathbf{u}$  the optimization problem of  $\psi(\mathbf{u})$  is convex and the set of optimal solutions of  $\psi(\mathbf{u})$  is non-empty and bounded, then the inf-compactness conditions holds automatically.

The directional derivative of  $\psi(\mathbf{u})$  at  $\mathbf{u}_c$  in a direction  $\mathbf{d} \in \mathbb{R}^k$  is given by

$$\psi'(\mathbf{u}_c, \mathbf{d}) = \min_{\mathbf{y} \in \mathcal{M}} \max_{\Omega \in \mathcal{Z}} \mathbf{d}^T \nabla_{\mathbf{u}} L(\mathbf{y}, \Omega, \mathbf{u}_c), \quad (\text{E.4})$$

where  $\mathcal{Z}$  is the set of optimal solutions of the dual problem of the optimization problem of  $\psi(\mathbf{u}_c)$  and  $L(\mathbf{y}, \Omega, \mathbf{u})$  denotes the Lagrangian defined as

$$L(\mathbf{y}, \Omega, \mathbf{u}) \triangleq f(\mathbf{y}, \mathbf{u}) + \text{tr}\{\Omega \mathbf{G}(\mathbf{y}, \mathbf{u})\} \quad (\text{E.5})$$

where  $\Omega$  denotes the Lagrangian multiplier matrix.

Let us look again to the definitions of the optimal value functions  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  (6.17) and (6.21), respectively, and define the following block diagonal matrix

$$\mathbf{G}_1(\mathbf{W}, \alpha) \triangleq \begin{pmatrix} -\mathbf{W} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \eta^2 \text{tr}\{\mathbf{W}\} - (\sqrt{\alpha} - 1)^2 & 0 & 0 \\ \mathbf{0} & 0 & \text{tr}\{\mathbf{Q}^H \mathbf{Q} \mathbf{W}\} - \alpha & 0 \\ \mathbf{0} & 0 & 0 & \alpha - \text{tr}\{\mathbf{Q}^H \mathbf{Q} \mathbf{W}\} \end{pmatrix}$$

as well as another block diagonal matrix denoted as  $\mathbf{G}_2(\mathbf{W}, \alpha)$  which has exactly same structure as the matrix  $\mathbf{G}_1(\mathbf{W}, \alpha)$  with difference that the element  $\eta^2 \text{tr}\{\mathbf{W}\} - (\sqrt{\alpha} - 1)^2$  in  $\mathbf{G}_1(\mathbf{W}, \alpha)$  is replaced by  $\eta^2 \cdot \text{tr}\{\mathbf{W}\} + (\sqrt{\alpha_c} - 1) + \alpha (1/\sqrt{\alpha_c} - 1)$  in  $\mathbf{G}_2(\mathbf{W}, \alpha)$ . Then the optimal value functions  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  can be equivalently recast as

$$k(\alpha) = \left\{ \min_{\mathbf{W}} \text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \right\} \mid \mathbf{G}_1(\mathbf{W}, \alpha) \preceq \mathbf{0} \right\}, \quad \theta_1 \leq \alpha \leq \theta_2 \quad (\text{E.6})$$

and

$$l(\alpha, \alpha_c) = \left\{ \min_{\mathbf{W}} \text{tr} \left\{ (\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W} \right\} \mid \mathbf{G}_2(\mathbf{W}, \alpha) \preceq \mathbf{0} \right\}, \quad \theta_1 \leq \alpha \leq \theta_2. \quad (\text{E.7})$$

It is trivial to verify that the optimization problems of  $k(\alpha_c)$  and  $l(\alpha_c, \alpha_c)$  can be expressed as

$$\begin{aligned} & \min_{\mathbf{W}} \quad \text{tr}\{(\hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{W}\} \\ & \text{subject to} \quad \text{tr}\{\mathbf{Q}^H \mathbf{Q} \mathbf{W}\} = \alpha_c \\ & \quad \quad \quad \text{tr}\{\mathbf{W}\} \leq \frac{(\sqrt{\alpha_c} - 1)^2}{\eta^2} \\ & \quad \quad \quad \mathbf{W} \succeq \mathbf{0}. \end{aligned} \quad (\text{E.8})$$

The problem (E.8) is convex and its solution set is non-empty and bounded. Indeed, let  $\mathbf{W}_1$  and  $\mathbf{W}_2$  denote two optimal solutions of the problem above. The Euclidean distance between  $\mathbf{W}_1$  and  $\mathbf{W}_2$  can be expressed as

$$\begin{aligned} \|\mathbf{W}_1 - \mathbf{W}_2\| &= \sqrt{\text{tr}\{\mathbf{W}_1^2\} + \text{tr}\{\mathbf{W}_2^2\} - 2\text{tr}\{\mathbf{W}_1 \mathbf{W}_2\}} \\ &\leq \sqrt{2 \frac{(\sqrt{\alpha_c} - 1)^4}{\eta^4}} \end{aligned} \quad (\text{E.9})$$

where the last line is due to the fact that the matrix product  $\mathbf{W}_1 \mathbf{W}_2$  is positive semi-definite and, therefore,  $\text{tr}\{\mathbf{W}_1 \mathbf{W}_2\} \geq 0$ , and also the fact that for any arbitrary positive semi-definite matrix  $\text{tr}\{\mathbf{A}^2\} \leq \text{tr}\{\mathbf{A}\}^2$ . From the equation above, it can be seen that the distance between any two arbitrary optimal solutions of (E.8) is finite and, therefore, the solution set is bounded. It is easy to verify that the optimization problem (E.8) satisfies the strong duality. It can be shown that the inf-compactness condition is satisfied by verifying that the optimization problems of  $k(\alpha)$  and  $l(\alpha, \alpha_c)$

are convex and their corresponding solution sets are bounded for any  $\alpha$ . Therefore, both of the optimal value functions  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  are directionally differentiable at  $\alpha_c$ .

Using the result of [146, Theorem 10], the directional derivatives of  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  at direction  $d$  can be respectively computed as

$$k'(\alpha, d) = \min_{\mathbf{W} \in \mathcal{M}} \max_{\mathbf{\Omega} \in \mathcal{Z}} d \left( \text{tr} \left\{ \mathbf{\Omega} \cdot \frac{d}{d\alpha} \mathbf{G}_1(\mathbf{W}, \alpha) \Big|_{\alpha=\alpha_c} \right\} \right) \quad (\text{E.10})$$

and

$$l'(\alpha, \alpha_c, d) = \min_{\mathbf{W} \in \mathcal{M}} \max_{\mathbf{\Omega} \in \mathcal{Z}} d \left( \text{tr} \left\{ \mathbf{\Omega} \cdot \frac{d}{d\alpha} \mathbf{G}_2(\mathbf{W}, \alpha) \Big|_{\alpha=\alpha_c} \right\} \right) \quad (\text{E.11})$$

where  $\mathcal{M}$  and  $\mathcal{Z}$  denote the optimal solution sets of the optimization problem of (E.8) and its dual problem, respectively. Using the definitions of  $\mathbf{G}_1(\mathbf{W}, \alpha)$  and  $\mathbf{G}_2(\mathbf{W}, \alpha)$ , it can be seen that the terms  $d\mathbf{G}_1(\mathbf{W}, \alpha)/d\alpha$  and  $d\mathbf{G}_2(\mathbf{W}, \alpha)/d\alpha$  are equal at  $\alpha = \alpha_c$  and, therefore, the directional derivatives are equivalent. The latter implies that the left and right derivatives of  $k(\alpha)$  and  $l(\alpha, \alpha_c)$  are equal at  $\alpha = \alpha_c$ .

□

## Appendix F

### Proof of Theorem 6.1

Using the expressions for the covariance matrices of the desired and interference sources defined in the conditions of the Theorem 6.1 and based on the assumption that the actual received signal covariance matrix is known, the optimization problem (6.5) can be expressed as

$$\begin{aligned}
 \min_{\mathbf{w}, \alpha} \quad & \mathbf{w}^H \left( \mathbf{U}_1 \mathbf{V}_s \mathbf{U}_1^H + \mathbf{U}_2 \left( \sum_{i=1}^N \mathbf{V}_i \right) \mathbf{U}_2^H + \sigma_n^2 \mathbf{I} \right) \mathbf{w} \\
 \text{subject to} \quad & \mathbf{w}^H \mathbf{U}_1 \mathbf{V}_s \mathbf{U}_1^H \mathbf{w} = \alpha \\
 & \mathbf{w}^H \mathbf{w} \leq \frac{(\sqrt{\alpha} - 1)^2}{\eta^2}, \quad \alpha \geq 1
 \end{aligned} \tag{F.1}$$

where  $\sigma_n^2$  is the noise power in each antenna element and  $\gamma = 0$  since  $\mathbf{R}$  is precisely known. By replacing the term  $\mathbf{w}^H \mathbf{U}_1 \mathbf{V}_s \mathbf{U}_1^H \mathbf{w}$  with  $\alpha$  in the objective function of the problem (F.1) due to the first constraint and then changing the variable  $\bar{\mathbf{w}} = \eta \mathbf{w} / (\sqrt{\alpha} - 1)$ , the optimization problem (F.1) can be equivalently expressed as

$$\begin{aligned}
 \min_{\bar{\mathbf{w}}, \alpha} \quad & \alpha + \frac{(\sqrt{\alpha} - 1)^2}{\eta^2} \cdot \bar{\mathbf{w}}^H \left( \mathbf{U}_2 \left( \sum_{i=1}^N \mathbf{V}_i \right) \mathbf{U}_2^H + \sigma_n^2 \mathbf{I} \right) \bar{\mathbf{w}} \\
 \text{subject to} \quad & \bar{\mathbf{w}}^H \mathbf{U}_1 \mathbf{V}_s \mathbf{U}_1^H \bar{\mathbf{w}} = \eta^2 \frac{\alpha}{(\sqrt{\alpha} - 1)^2} \\
 & \bar{\mathbf{w}}^H \bar{\mathbf{w}} \leq 1.
 \end{aligned} \tag{F.2}$$

Since  $\mathbf{U}_1$  and  $\mathbf{U}_2$  span orthogonal subspaces, the vector  $\bar{\mathbf{w}}$  can be expressed as  $\bar{\mathbf{w}} = \mathbf{U}_1 \mathbf{v}_1 + \mathbf{U}_2 \mathbf{v}_2 + \mathbf{E} \mathbf{v}_3$ , where  $\mathbf{E}$  is the complement subspace to the subspaces spanned by  $\mathbf{U}_1$  and  $\mathbf{U}_2$ . Then, the optimization problem (F.2) can be further

expressed as

$$\begin{aligned}
& \min_{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \alpha} \alpha + \frac{(\sqrt{\alpha}-1)^2}{\eta^2} \cdot \left( \mathbf{v}_2^H \left( \sum_{i=1}^N \mathbf{V}_i \right) \mathbf{v}_2 + \sigma_n^2 \left( \sum_{i=1}^3 \mathbf{v}_i^H \mathbf{v}_i \right) \right) \\
& \text{subject to } \mathbf{v}_1^H \mathbf{V}_s \mathbf{v}_1 = \eta^2 \frac{\alpha}{(\sqrt{\alpha}-1)^2} \\
& \sum_{i=1}^3 \mathbf{v}_i^H \mathbf{v}_i \leq 1.
\end{aligned} \tag{F.3}$$

From (F.3), it can be immediately concluded that the optimal  $\mathbf{v}_2$  and  $\mathbf{v}_3$  are both zero. Indeed, if  $\bar{\mathbf{w}}_{\text{opt}} = \mathbf{U}_1 \mathbf{v}_{1,\text{opt}} + \mathbf{U}_2 \mathbf{v}_{2,\text{opt}} + \mathbf{E} \mathbf{v}_{3,\text{opt}}$  and  $\alpha_{\text{opt}}$  is the optimal solution, then  $\bar{\mathbf{w}}^* = \mathbf{U}_1 \mathbf{v}_{1,\text{opt}}$  is also a feasible point of the problem (F.3) whose corresponding objective value is less than or equal to that of  $\bar{\mathbf{w}}_{\text{opt}}$  which contradicts the optimality of  $\bar{\mathbf{w}}_{\text{opt}}$ , and thus,  $\mathbf{v}_{2,\text{opt}}$  and  $\mathbf{v}_{3,\text{opt}}$  are both zero. Based on this fact the problem (F.3) can be simplified as

$$\begin{aligned}
& \min_{\mathbf{v}_1, \alpha} \alpha + \frac{(\sqrt{\alpha}-1)^2}{\eta^2} \cdot \left( \sigma_n^2 \mathbf{v}_1^H \mathbf{v}_1 \right) \\
& \text{subject to } \mathbf{v}_1^H \mathbf{V}_s \mathbf{v}_1 = \eta^2 \frac{\alpha}{(\sqrt{\alpha}-1)^2} \\
& \mathbf{v}_1^H \mathbf{v}_1 \leq 1.
\end{aligned} \tag{F.4}$$

For a fixed value of  $\alpha$ , the problem (F.4) is feasible if the minimum value of the quadratic term  $\mathbf{v}_1^H \mathbf{v}_1$  subject to the constraint  $\mathbf{v}_1^H \mathbf{V}_s \mathbf{v}_1 = \eta^2 \alpha / (\sqrt{\alpha}-1)^2$  is less than or equal to one. Such minimum is equal to  $\eta^2 \alpha / ((\sqrt{\alpha}-1)^2 \lambda_{\max}\{\mathbf{V}_s\})$ , and it is achieved by  $\mathbf{v}_1 = \eta \sqrt{\alpha} / (\sqrt{\alpha}-1) \mathcal{P}\{\mathbf{V}_s\}$ . Then, the problem (F.4) can be equivalently recast as

$$\begin{aligned}
& \min_{\alpha} \alpha(1 + \sigma_n^2) \\
& \text{subject to } \frac{(\sqrt{\alpha}-1)^2}{\alpha} \geq \frac{\eta^2}{\lambda_{\max}\{\mathbf{V}_s\}}.
\end{aligned} \tag{F.5}$$

Since the function  $(\sqrt{\alpha}-1)^2/\alpha$  is strictly increasing and the smallest value of  $\alpha$  for which the problem (F.5) is feasible is equal to  $1/\left(1 - \eta/\sqrt{\lambda_{\max}\{\mathbf{V}_s\}}\right)^2$ , the optimization problem (F.1) can be finally expressed as

$$\begin{aligned}
& \min_{\alpha} \alpha(1 + \sigma_n^2) \\
& \text{subject to } \alpha \geq \frac{1}{\left(1 - \frac{\eta}{\sqrt{\lambda_{\max}\{\mathbf{V}_s\}}}\right)^2}.
\end{aligned} \tag{F.6}$$

Therefore, the optimal value function  $k(\alpha)$  equals to  $\alpha(1 + \sigma_n^2)$  under the conditions of the Theorem 6.1, which is a linear function. Moreover, the optimization problem (F.6) has the trivial solution of  $\alpha_{\text{opt}} = 1 / \left(1 - \eta / \sqrt{\lambda_{\max}\{\mathbf{V}_s\}}\right)^2$  whose corresponding optimal beamforming vector is as

$$\mathbf{w}_{\text{opt}} = \mathbf{U}_1 \eta \sqrt{\alpha} / (\sqrt{\alpha} - 1) \mathcal{P}\{\mathbf{V}_s\}. \quad (\text{F.7})$$

This completes the proof. □

# Appendix G

## Proof of Lemma 7.1

Let  $\mathbf{X}_{\text{opt}}$ ,  $\tau_{\text{opt}}$ , and  $\beta_{\text{opt}}$  denote the optimal solution of the optimization problem (7.44). We define another auxiliary optimization problem based on the problem (7.44) by fixing the variable  $\beta$  to  $\beta_{\text{opt}}$  as

$$\begin{aligned} & \max_{\mathbf{X}, \tau} \ln(\text{tr}\{\mathbf{A}_1\mathbf{X}\}) + \ln(\tau) - \ln(\beta_{\text{opt}}) \\ & \text{subject to } \text{tr}\{\mathbf{B}_1\mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2\mathbf{X}\} = \tau \\ & \text{tr}\{\mathbf{B}_2\mathbf{X}\} = \beta_{\text{opt}}, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \tag{G.1}$$

For every feasible point of the problem (G.1), denoted as  $(\mathbf{X}, \tau)$ , it is easy to verify that  $(\mathbf{X}, \tau, \beta_{\text{opt}})$  is also a feasible point of the problem (7.44). Based on this fact, it can be concluded that the optimal value of the problem (G.1) is less than or equal to the optimal value of the problem (7.44). However, since  $(\mathbf{X}_{\text{opt}}, \tau_{\text{opt}})$  is a feasible point of the problem (G.1) and the value of the objective function at this feasible point is equal to the optimal value of the problem (7.44), that is,  $\ln(\text{tr}\{\mathbf{A}_1\mathbf{X}_{\text{opt}}\}) + \ln(\tau_{\text{opt}}) - \ln(\beta_{\text{opt}})$ , we find that both optimization problems (G.1) and (7.44) have the same optimal value. To find an upper-bound for  $\beta_{\text{opt}}$ , we make the feasible set of the problem (G.1) independent of  $\beta_{\text{opt}}$  which can be done by dropping the constraint  $\text{tr}\{\mathbf{B}_2\mathbf{X}\} = \beta_{\text{opt}}$  in the problem (G.1). Then the following problem is obtained:

$$\begin{aligned} & \max_{\mathbf{X}, \tau, \beta} \ln(\text{tr}\{\mathbf{A}_1\mathbf{X}\}) + \ln(\tau) - \ln(\beta_{\text{opt}}) \\ & \text{subject to } \text{tr}\{\mathbf{B}_1\mathbf{X}\} = 1, \quad \text{tr}\{\mathbf{A}_2\mathbf{X}\} = \tau, \quad \mathbf{X} \succeq \mathbf{0}. \end{aligned} \tag{G.2}$$

Noticing that the feasible set of the optimization problem (G.1) is a subset of the feasible set of the problem (G.2), it is straightforward to conclude that the optimal value of the problem (G.2) is equal to or greater than the optimal value of the

problem (G.1) and, thus, also the optimal value of the problem (7.44). On the other hand,  $p^*$  is the value of the objective function of the problem (7.44) which corresponds to an arbitrary feasible point and as a result is less than or equal to the optimal value of the problem (7.44). Since the optimal value of the problem (G.2) is greater than or equal to the optimal value of the problem (7.44) and the optimal value of the problem (7.44) is greater than or equal to  $p^*$ , the optimal value of the problem (G.2), denoted as  $q^* - \ln(\beta_{\text{opt}})$ , is greater than or equal to  $p^*$ , and therefore,  $\beta_{\text{opt}} \leq e^{(q^* - p^*)}$  which completes the proof.  $\square$

# Appendix H

## Proof of Lemma 7.2

We define first the feasible set of the optimization problems (7.58) and (7.59) as

$$\mathcal{S} \triangleq \{\mathbf{g} \mid \mathbf{g}^H \mathbf{B}_1 \mathbf{g} = 1, \mathbf{g}^H \mathbf{A}_2 \mathbf{g} = \alpha, \mathbf{g}^H \mathbf{B}_2 \mathbf{g} = \beta\}.$$

Two different cases are possible. If for every  $\mathbf{g} \in \mathcal{S}$ ,  $\ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) \leq \ln(\alpha) - \ln(\beta)$ , then it can be easily verified that

$$p_1 = \max_{\mathbf{g} \in \mathcal{S}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}). \quad (\text{H.1})$$

Furthermore, since for  $\mathbf{g} \in \mathcal{S}$ ,  $\ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) \leq \ln(\alpha) - \ln(\beta)$ , it is also true that  $\max_{\mathbf{g} \in \mathcal{S}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) \leq \ln(\alpha) - \ln(\beta)$  and therefore

$$p_2 = \max_{\mathbf{g} \in \mathcal{S}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}). \quad (\text{H.2})$$

Hence, trivially  $p_1 = p_2$ .

In the other case, let  $\mathcal{D}$  denote the set of all vectors  $\mathbf{g} \in \mathcal{S}$  such that  $\ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) > \ln(\alpha) - \ln(\beta)$  and let  $\tilde{\mathcal{D}}$  denote its complement. Considering the inner minimization problem of the problem (7.58), it can be simply concluded that  $p_1$  is the maximum of the following function over  $\mathbf{g} \in \mathcal{S}$

$$k(\mathbf{g}) \triangleq \begin{cases} \ln(\alpha) - \ln(\beta), & \mathbf{g} \in \mathcal{D} \\ \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}), & \mathbf{g} \in \tilde{\mathcal{D}}. \end{cases} \quad (\text{H.3})$$

Since for  $\mathbf{g} \in \tilde{\mathcal{D}}$ ,  $k(\mathbf{g}) = \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) \leq \ln(\alpha) - \ln(\beta)$ , it is resulted that  $p_1 = \ln(\alpha) - \ln(\beta)$ .

Moreover, since for  $\mathbf{g} \in \mathcal{D}$ ,  $\ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) > \ln(\alpha) - \ln(\beta)$ , it is also true that  $\max_{\mathbf{g} \in \mathcal{S}} \ln(\mathbf{g}^H \mathbf{A}_1 \mathbf{g}) > \ln(\alpha) - \ln(\beta)$ . Therefore  $p_2 = \ln(\alpha) - \ln(\beta)$  that completes the proof.  $\square$

# Appendix I

## Rank deficient received signal covariance matrix

The matrix  $\mathbf{R}_R$  (7.4) is rank deficient only if the noise is spatially correlated. Let the rank of  $\mathbf{R}_R$  be denoted as  $N$ ,  $N < M_R$ . Since the matrix  $\mathbf{Q} \triangleq \mathbf{R}_R^T \otimes \mathbf{I}_{M_R}$ , its rank is equal to the rank of the matrix  $\mathbf{R}_R$  multiplied by the rank of the matrix  $\mathbf{I}_{M_R}$ , i.e.,  $\text{rank}\{\mathbf{Q}\} = M_R N$ . Then if  $\mathbf{R}_R$  is rank deficient,  $\mathbf{Q}$  is also rank deficient.

For convenience, we restate the sum-rate maximization problem (7.28) as

$$\mathbf{g}_{\text{opt}} = \arg \max_{\mathbf{g} | \mathbf{g}^H \mathbf{Q} \mathbf{g} \leq P_{T,R}} s(\mathbf{g}) \quad (\text{I.1})$$

where

$$s(\mathbf{g}) \triangleq \frac{1}{2} \log_2 \left( \left( 1 + \frac{\mathbf{g}^H \mathbf{K}_{2,1} \mathbf{g} \cdot P_{T,2}}{\mathbf{g}^H \mathbf{J}_1 \mathbf{g} + P_{N,1}} \right) \cdot \left( 1 + \frac{\mathbf{g}^H \mathbf{K}_{1,2} \mathbf{g} \cdot P_{T,1}}{\mathbf{g}^H \mathbf{J}_2 \mathbf{g} + P_{N,2}} \right) \right). \quad (\text{I.2})$$

If  $\mathbf{Q}$  is rank deficient, then for every  $\mathbf{g} \in \text{Null}\{\mathbf{Q}\}$  where  $\text{Null}\{\mathbf{Q}\}$  is the null space of  $\mathbf{Q}$ , the total transmit power from the relay is zero, i.e.,  $\mathbf{g}^H \mathbf{Q} \mathbf{g} = 0$ . Moreover, the corresponding sum-rate  $s(\mathbf{g})$  for any vector  $\mathbf{g} \in \text{Null}\{\mathbf{Q}\}$  is equal to zero. To show this, let us consider  $\mathbf{g}_0 = \text{vec}\{\mathbf{G}_0\} \in \text{Null}\{\mathbf{Q}\}$  that straightforwardly implies that

$$\mathbf{g}_0^H \mathbf{Q} \mathbf{g}_0 = \text{tr} \{ \mathbf{G}_0 \mathbf{R}_R \mathbf{G}_0^H \} = 0. \quad (\text{I.3})$$

Note that the first equality in (I.3) follows from (7.3). Substituting (7.4) in (I.3), we obtain

$$\begin{aligned} \text{tr} \left\{ \left( \mathbf{G}_0 \mathbf{h}_1^{(f)} \right) \left( \mathbf{G}_0 \mathbf{h}_1^{(f)} \right)^H \cdot P_{T,1} + \left( \mathbf{G}_0 \mathbf{h}_2^{(f)} \right) \left( \mathbf{G}_0 \mathbf{h}_2^{(f)} \right)^H \cdot P_{T,2} \right. \\ \left. + \left( \mathbf{G}_0 \mathbf{R}_{N,R}^{1/2} \right) \left( \mathbf{G}_0 \mathbf{R}_{N,R}^{1/2} \right)^H \right\} = 0. \end{aligned} \quad (\text{I.4})$$

Since the following matrices  $(\mathbf{G}_0 \mathbf{h}_1^{(f)}) (\mathbf{G}_0 \mathbf{h}_1^{(f)})^H$ ,  $(\mathbf{G}_0 \mathbf{h}_2^{(f)}) (\mathbf{G}_0 \mathbf{h}_2^{(f)})^H$ , and  $(\mathbf{G}_0 \mathbf{R}_{N,R}^{1/2}) (\mathbf{G}_0 \mathbf{R}_{N,R}^{1/2})^H$  in (I.4) are all positive semi-definite and the powers  $P_{T,1}$  and  $P_{T,2}$  are strictly positive, (I.4) is satisfied only if  $\mathbf{G}_0 \mathbf{h}_1^{(f)} = \mathbf{0}$ ,  $\mathbf{G}_0 \mathbf{h}_2^{(f)} = \mathbf{0}$ , and  $\mathbf{G}_0 \mathbf{R}_{N,R}^{1/2} = \mathbf{0}$ . Therefore, the following equations are in order

$$\mathbf{g}_0^H \mathbf{K}_{2,1} \mathbf{g}_0 = \left| \left( \mathbf{h}_1^{(b)} \right)^T \mathbf{G}_0 \mathbf{h}_2^{(f)} \right|^2 = 0 \quad (\text{I.5})$$

$$\mathbf{g}_0^H \mathbf{K}_{1,2} \mathbf{g}_0 = \left| \left( \mathbf{h}_2^{(b)} \right)^T \mathbf{G}_0 \mathbf{h}_1^{(f)} \right|^2 = 0 \quad (\text{I.6})$$

$$\mathbf{g}_0^H \mathbf{J}_i \mathbf{g}_0 = \left( \mathbf{h}_i^{(b)} \right)^T \mathbf{G}_0 \mathbf{R}_{N,R} \mathbf{G}_0^H \left( \mathbf{h}_i^{(b)} \right)^* = 0, \quad i = 1, 2. \quad (\text{I.7})$$

Substituting (I.5)–(I.7) in (I.2), we conclude that the sum-rate is indeed zero for any  $\mathbf{g}_0 \in \text{Null}\{\mathbf{Q}\}$ . Furthermore, in a similar way, it can be shown that, for any  $\mathbf{g} = \mathbf{g}_0 + \mathbf{g}_1$  such that  $\mathbf{g}_0 \in \text{Null}\{\mathbf{Q}\}$ ,  $s(\mathbf{g}) = s(\mathbf{g}_1)$ , and  $(\mathbf{g}_0 + \mathbf{g}_1)^H \mathbf{Q} (\mathbf{g}_0 + \mathbf{g}_1) = \mathbf{g}_1^H \mathbf{Q} \mathbf{g}_1$ , which means that  $\mathbf{g}_0$  does not have any contribution in the transmit power as well as the sum-rate.

Using the above observations, it is easy to see that if  $\mathbf{R}_R$  is rank deficient, the only thing required to do is reformulating the rate function (I.2) in the following manner. Denote the eigenvalue decomposition of the matrix  $\mathbf{Q}$  as  $\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$  where  $\mathbf{U}_{M_R^2 \times M_R^2}$  and  $\mathbf{\Lambda}_{M_R^2 \times M_R^2}$  are unitary and diagonal matrices of eigenvectors and eigenvalues, respectively. The  $i$ th eigenvector and the  $i$ th eigenvalue of  $\mathbf{Q}$  denoted as  $\mathbf{u}_i$  and  $\lambda_i$ , respectively, constitutes the column  $i$  of  $\mathbf{U}$  and  $i$ th diagonal element of  $\mathbf{\Lambda}$ . It is assumed without loss of generality that the eigenvalues  $\lambda_i$ ,  $i = 1, \dots, M_R^2$  are ordered in the descending order, i.e.,  $\lambda_i \geq \lambda_{i+1}$ ,  $i = 1, \dots, M_R^2 - 1$ . Since in the case of rank deficient  $\mathbf{R}_R$ , the rank of  $\mathbf{Q}$  is equal to  $M_R N$ , the last  $M_R(M_R - N)$  eigenvalues of  $\mathbf{Q}$  are zero. By splitting  $\mathbf{U}$  to the  $M_R^2 \times (M_R N)$  matrix  $\mathbf{U}_1$  and the  $M_R^2 \times (M_R(M_R - N))$  matrix  $\mathbf{U}_2$  as  $\mathbf{U} = [\mathbf{U}_1 \ \mathbf{U}_2]$ , the matrix  $\mathbf{Q}$  can be decomposed as  $\mathbf{Q} = \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{U}_1^H + \mathbf{U}_2 \mathbf{\Lambda}_2 \mathbf{U}_2^H$  where the  $(M_R N) \times (M_R N)$  diagonal matrix  $\mathbf{\Lambda}_1$  contains the  $M_R N$  dominant eigenvalues, while the other  $(M_R(M_R - N)) \times (M_R(M_R - N))$  diagonal matrix  $\mathbf{\Lambda}_2$  contains the  $M_R(M_R - N)$  zero eigenvalues. Since  $\mathbf{U}$  is unitary, any arbitrary vector  $\mathbf{g}$  can be expressed as  $\mathbf{g} = \mathbf{U}_1 \boldsymbol{\alpha} + \mathbf{U}_2 \boldsymbol{\beta}$  where  $\boldsymbol{\alpha}_{M_R N}$  and  $\boldsymbol{\beta}_{M_R(M_R - N)}$  are the coefficient vectors. It is easy to verify that the component  $\mathbf{U}_2 \boldsymbol{\beta}$  lies inside  $\text{Null}\{\mathbf{Q}\}$  and as a result  $s(\mathbf{U}_1 \boldsymbol{\alpha}_1 + \mathbf{U}_2 \boldsymbol{\alpha}_2) = s(\mathbf{U}_1 \boldsymbol{\alpha}_1)$  and  $(\mathbf{U}_1 \boldsymbol{\alpha}_1 + \mathbf{U}_2 \boldsymbol{\alpha}_2)^H \mathbf{Q} (\mathbf{U}_1 \boldsymbol{\alpha}_1 + \mathbf{U}_2 \boldsymbol{\alpha}_2) = (\mathbf{U}_1 \boldsymbol{\alpha}_1)^H \mathbf{Q} (\mathbf{U}_1 \boldsymbol{\alpha}_1)$ . Therefore,  $\boldsymbol{\beta}$  can be any arbitrary vector and we only need to find the optimal  $\boldsymbol{\alpha}$ . Substituting  $\mathbf{U}_1 \boldsymbol{\alpha}$  in

(I.2), the sum-rate can be expressed only as a function of  $\boldsymbol{\alpha}$  as follows

$$s(\mathbf{U}_1 \boldsymbol{\alpha}) = \frac{1}{2} \log_2 \left( \left( 1 + \frac{\boldsymbol{\alpha}^H (\mathbf{U}_1^H \mathbf{K}_{2,1} \mathbf{U}_1) \boldsymbol{\alpha} \cdot P_{T,2}}{\boldsymbol{\alpha}^H (\mathbf{U}_1^H \mathbf{J}_1 \mathbf{U}_1) \boldsymbol{\alpha} + P_{N,1}} \right) \right). \quad (\text{I.8})$$

$$\left( 1 + \frac{\boldsymbol{\alpha}^H (\mathbf{U}_1^H \mathbf{K}_{1,2} \mathbf{U}_1) \boldsymbol{\alpha} \cdot P_{T,1}}{\boldsymbol{\alpha}^H (\mathbf{U}_1^H \mathbf{J}_2 \mathbf{U}_1) \boldsymbol{\alpha} + P_{N,2}} \right). \quad (\text{I.9})$$

Then the optimization problem (I.1) is equivalent to the maximization of (I.9) under the constraint

$$\boldsymbol{\alpha}^H \boldsymbol{\Lambda}_1 \boldsymbol{\alpha} \leq P_{T,R}. \quad (\text{I.10})$$

Since the matrix  $\boldsymbol{\Lambda}_1$  is full rank, the corresponding optimization problem can be solved by the methods that we develop in the Chapter 7.

# Bibliography

- [1] O. Sallent, J. Perez-Romero, J. Sanchez-Gonzalez, R. Agusti, M. A. Diaz-Guerra, D. Henche, and D. Paul, "A roadmap from UMTS optimization to LTE self-optimization," *IEEE Mag. Commun.*, vol. 49, no. 6, pp. 172–182, Jun. 2011.
- [2] J. Fan, G. Y. Li, Y. Qinye, B. Peng, and X. Zhu, "Joint user pairing and resource allocation for LTE uplink transmission," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2838–2847, Aug. 2012.
- [3] D. R. Fuhrmann, "A geometric approach to subspace tracking," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1997, pp. 783–787.
- [4] A. Hassaniien, S. A. Vorobyov, and A. B. Gershman, "Moving target parameters estimation in non-coherent MIMO radar systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2354–2361, May 2012.
- [5] C. Shen, T.-H. Chang, K. -Y. Wang, Z. Qiu, and C. -Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: an ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, Jun. 2012.
- [6] Z. -Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.
- [7] S. A. Vorobyov, A. B. Gershman, and Z. -Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, pp. 313–324, Feb. 2003.
- [8] S. Shahbazpanahi, A. B. Gershman, Z. -Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general rank signal models," *IEEE Trans. Signal Process.*, vol. 51, pp. 2257–2269, Sept. 2003.
- [9] A. Khabbazzibasmenj and S. A. Vorobyov, "Power allocation based on SEP minimization in two-hop decode-and-forward relay networks," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3954–3963, Aug. 2011.
- [10] H. H. Chen and A. B. Gershman, "Robust adaptive beamforming for general-rank signal models with positive semi-definite constraints," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, USA, Apr. 2008, pp. 2341–2344.
- [11] T. K. Phan, S. A. Vorobyov, C. Tellambura, and T. Le-Ngoc, "Power control for wireless cellular systems via D.C. programming," in *Proc. IEEE Workshop Statistical Signal Process. (SSP)*, Madison, WI, USA, Aug. 2007, pp. 507–511.
- [12] Y. Xu, S. Panigrahi, and T. Le-Ngoc, "A concave minimization approach to dynamic spectrum management for digital subscriber lines," in *Proc. IEEE Inter. Conf. Commun. (ICC)*, Istanbul, Turkey, Jun. 2006, pp. 84–89.

- [13] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, pp. 2159–2168, Jul. 2010.
- [14] Z. -Q. Luo, W. -K. Ma, M. -C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Mag. Signal Process.*, vol. 27, no. 3, pp. 20–34, 2010.
- [15] A. Beck, A. Ben-Tal and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, 2010.
- [16] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 3, pp. 844–860, 2006.
- [17] H. Wolkowicz, "Relaxations of Q2P," in *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, H. Wolkowicz, R. Saigal, and L. Venberghe, Eds. Norwell, MA: Kluwer, 2000, ch. 13.4.
- [18] S. Zhang and Y. Huang, "Complex quadratic optimization and semidefinite programming," *SIAM J. Optim.*, vol. 16, no. 3, pp. 871–890, 2006.
- [19] N. D. Sidiropoulos, T. N. Davidson, and Z. -Q. Luo "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.
- [20] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optim. Methods Softw.*, vol. 9, no. 1–3, pp. 141–160, 1998.
- [21] S. Zhang, "Quadratic maximization and semidefinite relaxation," *Math. Program.*, vol. 87, pp. 453–465, 2000.
- [22] K. T. Phan, S. A. Vorobyov, N. D. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, pp. 2323–2335, Jun. 2009.
- [23] G. Wang, "A semidefinite relaxation method for energy-based source localization in sensor networks." *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2293–2301, Jun. 2011.
- [24] A. De Maio, S. De Nicola, Y. Huang, S. Zhang, and A. Farina, "Code design to optimize radar detection performance under accuracy and similarity constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5618–5629, Nov. 2008.
- [25] Z. Mao, X. Wang, and X. Wang, "Semidefinite programming relaxation approach for multiuser detection of QAM signals," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4275–4279, Dec. 2007.
- [26] A. De Maio, Y. Huang, D. P. Palomar, S. Zhang, and A. Farina, "Fractional QCQP with application in ML steering direction estimation for radar detection," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 172–185, Jan. 2011.
- [27] S. K. Joshi, P. C. Weeraddana, M. Codreanu, and M. Latva-aho, "Weighted sum-rate maximization for MISO downlink cellular networks via branch and bound," *IEEE Trans. Signal Process.*, vol. 60, pp. 2090–2095, Apr. 2012.
- [28] J. Zhang, F. Roemer, M. Haardt, A. Khabbazi-basmenj, and S. A. Vorobyov, "Sum rate maximization for multi-pair two-way relaying with single-antenna amplify and forward relays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 2477–2480.

- [29] A. Khabbazibasmenj and S. A. Vorobyov, "Two-way relay beamforming design: Proportional fair and max-min rate fair approaches using POTDC," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 26–31, 2013, pp. 4997–5001.
- [30] A. Khabbazibasmenj and S. A. Vorobyov, "A computationally efficient robust adaptive beamforming for general-rank signal models with positive semi-definiteness constraint," in *Proc. IEEE Workshop Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, San Juan, Puerto Rico, Dec. 2011, pp. 185–188.
- [31] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Weighted sum-rate maximization for a set of interfering links via branch-and-bound," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3977–3996, Aug. 2011.
- [32] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.
- [33] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*. Springer, 1996.
- [34] H. Tuy, *Convex Analysis and Global Optimization*. Kluwer Academic Publishers, 1998.
- [35] A. Rubinov, H. Tuy, and H. Mays, "An algorithm for monotonic global optimization problems," *Optim.*, vol. 49, pp. 205–221, 2001.
- [36] H. Tuy, F. Al-Khayyal, and P. T. Thach, "Monotonic optimization: Branch and cut methods," in *Essays and Surveys in Global Optim.* Ed.: C. Audet, P. Hansen, G. Savard, Springer, 2005, pp. 39–78.
- [37] J. Li and P. Stoica, *MIMO Radar Signal Processing*. New Jersey: Wiley, 2009.
- [38] A. Hassanien and S. A. Vorobyov, "Phased-MIMO radar: A tradeoff between phased-array and MIMO radars," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3137–3151, Jun. 2010.
- [39] C. Duofang, C. Baixiao, and Q. Guodong, "Angle estimation using ESPRIT in MIMO radar," *Electron. Lett.*, vol. 44, no. 12, pp. 770–771, Jun. 2008.
- [40] A. Hassanien and S. A. Vorobyov, "Direction finding for MIMO radar with colocated antennas using transmit beamspace preprocessing," in *Proc. IEEE Workshop Computational Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009, pp. 181–184.
- [41] A. Hassanien and S. A. Vorobyov, "Transmit energy focusing for DOA estimation in MIMO radar with colocated antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2669–2682, Jun. 2011.
- [42] D. R. Fuhrmann and G. S. Antonio, "Transmit beamforming for MIMO radar systems using partial signal correlation," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2004, pp. 295–299.
- [43] D. R. Fuhrmann and G. S. Antonio, "Transmit beamforming for MIMO radar systems using signal cross-correlation," *Trans. Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 171–186, Jan. 2008.
- [44] A. B. Gershman, "Robust adaptive beamforming in sensor arrays," *Int. J. Electron. Commun.*, vol. 53, pp. 305–314, Dec. 1999.
- [45] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Elsevier Signal Process.*, Special Issue: *Advances in Sensor Array Process.*, to be published in Jul. 2013.

- [46] H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1365-1376, Oct. 1987.
- [47] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Eng. Electron. Phys.*, vol. 26, pp. 87-95, Mar. 1981.
- [48] D. D. Feldman and L. J. Griffiths, "A projection approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, pp. 867-876, Apr. 1994.
- [49] S. A. Vorobyov, A. B. Gershman, Z. -Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Process. Lett.*, vol. 11, pp. 108-111, Feb. 2004.
- [50] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, pp. 1702-1715, Jul. 2003.
- [51] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, pp. 1684-1696, May 2005.
- [52] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, pp. 5719-5724, Nov. 2008.
- [53] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, pp. 2407-2423, Sept. 2004.
- [54] A. Beck and Y. C. Eldar, "Doubly constrained robust capon beamformer with ellipsoidal uncertainty sets," *IEEE Trans. Signal Process.*, vol. 55, pp. 753-758, Feb. 2007.
- [55] A. Hassaniien, S. A. Vorobyov, and K. M. Wong, "Robust adaptive beamforming using sequential programming: An iterative solution to the mismatch problem," *IEEE Signal Process. Lett.*, vol. 15, pp. 733-736, 2008.
- [56] H. H. Chen and A. B. Gershman, "Worst-case based robust adaptive beamforming for general-rank signal models using positive semi-definite covariance constraint," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prauge, Czech Republic, May 2011, pp. 2628-2631.
- [57] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.
- [58] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939-1948, Nov. 2003.
- [59] J. N. Laneman and G. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 1, no. 11, Taipei, Taiwan, Nov. 2002, pp. 77-81.
- [60] L. Sanguinetti, A. A. D. Amico, Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331-1346, Sept. 2012.
- [61] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimization linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837-4852, Dec. 2009.

- [62] 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 J. Sel. Areas of Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.
- [63] N. Lee, H. J. Yang, and J. Chun, "Achievable sum-rate maximizing AF relay beamforming scheme in two-way relay channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 300–305.
- [64] C. Y. Leow, Z. Ding, K. K. Leung, and D. L. Goeckel, "On the study of analogue network coding for multi-pair, bidirectional relay channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 670–681, Feb. 2011.
- [65] F. Roemer and M. Haardt, "Sum-rate maximization in two-way relaying systems with MIMO amplify and forward relays via generalized eigenvectors," in *Proc. 18-th European Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 377–381.
- [66] F. Roemer and M. Haardt, "Algebraic norm-maximizing (ANOMAX) transmit strategy for two-way relaying with MIMO amplify and forward relays," *IEEE Signal Proc. Lett.*, vol. 16, no. 10, pp. 909–912, Oct. 2009.
- [67] M. Pischella and D. Le Ruyet, "Optimal power allocation for the two-way relay channel with data rate fairness," *IEEE Commun. Lett.*, vol. 15, no. 9, pp. 959–961, Sept. 2011.
- [68] Y. Jing and S. Shahbazpanahi, "Max-min optimal joint power control and distributed beamforming for two-way relay networks under per-node power constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 12, Dec. 2012.
- [69] S. Shahbazpanahi and M. Dong, "A semi-closed form solution to the SNR balancing problem of two-way relay network beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, Texas, USA, Mar. 2010, pp. 2514–2517.
- [70] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 1994.
- [71] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed. New York: Mc- Graw-Hill, 2001.
- [72] S. Haykin, J. Litva, and T. J. Shepherd, *Radar Array Processing*. New York: Springer-Verlag, 1993.
- [73] R. Klemm, *Applications of Space-Time Adaptive Processing*. London, U.K.: IEE Press, 2004.
- [74] L. Xu, J. Li, and P. Stoica, "Target detection and parameter estimation for MIMO radar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 927–939, Jul. 2008.
- [75] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO radar: An idea whose time has come," in *Proc. IEEE Radar Conf.*, Honolulu, HI, Apr. 2004, vol. 2, pp. 71–78.
- [76] A. Haimovich, R. Blum, and L. Cimini, "MIMO radar with widely separated antennas," *IEEE Signal Process. Mag.*, vol. 25, pp. 116–129, Jan. 2008.
- [77] A. De Maio, M. Lops, and L. Venturino, "Diversity-integration tradeoffs in MIMO detection," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5051–5061, Oct. 2008.
- [78] M. Akcakaya and A. Nehorai, "MIMO radar sensitivity analysis for target detection," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3241–3250, Jul. 2011.

- [79] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, pp. 106-114, Sept. 2007.
- [80] F. Daum and J. Huang, "MIMO radar: Snake oil or good idea," *IEEE Aerosp. Electron. Syst. Mag.*, pp. 8-12, May 2009.
- [81] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [82] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Rome, NY, 1979, pp. 234-258.
- [83] F. Li and R. J. Vaccaro, "Analysis of min-norm and MUSIC with arbitrary array geometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 6, pp. 976-985, Nov. 1990.
- [84] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110-1121, May 1991.
- [85] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67-94, Aug. 1996.
- [86] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276-280, Mar. 1986.
- [87] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984-995, Jul. 1989.
- [88] M. Haardt, "Efficient one-, two-, and multidimensional high-resolution array signal processing," Ph.D. dissertation, Shaker Verlag, Aachen, Germany, 1997.
- [89] B. D. Van Veen, and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Process. Mag.*, no. 2, pp. 4-24, Apr. 1988.
- [90] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
- [91] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, pp. 853-863, Nov. 1974.
- [92] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, pp. 27-34, Jan. 1982.
- [93] E. K. Hung and R. M. Turner, "A fast beamforming algorithm for large arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 19, pp. 598-607, Jul. 1983.
- [94] L. Chang and C. C. Yeh, "Performance of DMI and eigenspace-based beamformers," *IEEE Trans. Antennas Propag.*, vol. 40, pp. 1336-1347, Nov. 1992.
- [95] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Process.*, vol. 43, pp. 730-736, Mar. 1995.
- [96] L. Lei, J. P. Lie, A. B. Gershman, and C. M. S. See, "Robust adaptive beamforming in partly calibrated sparse sensor arrays," *IEEE Trans. Signal Process.*, vol. 58, pp. 1661-1667, Mar. 2010.
- [97] A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. Lundberg, "Eigenvalue beamforming using a multi-rank MVDR beamformer and subspace selection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1954-1967, May 2008.

- [98] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.
- [99] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379-389, Feb. 2007.
- [100] T. J. Oechterding, I. Bjelakovic, C. Schnurr, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 454-458, Jan. 2008.
- [101] R. Ahlswede, N. Cai, S. -Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204-1216, Jul. 2000.
- [102] I. Hammerstrom, M. Kuhn, C. Esli, J. Zhao, A. Wittneben, and G. Bauch, "MIMO two-way relaying with transmit CSI at the relay," in *Proc. IEEE Signal Process. Advances Wireless Commun.*, Helsinki, Finland, Jun. 2007, pp. 1-5.
- [103] J. Joung and A. H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1833-1846, Mar. 2010.
- [104] A. U. T. Amah and A. Klein, "Pair-aware transceive beamforming for non-regenerative multi-user two-way relaying," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2506-2509.
- [105] H. Q. Ngo, T. Q. S. Quek, and H. Shin, "Amplify-and-forward two-way relay networks: Error exponents and resource allocation," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2653-2666, Sept. 2010.
- [106] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, pp. 664-678, Feb. 2010.
- [107] S. Ponnusamy, *Foundations of Mathematical Analysis*. Springer, 2012.
- [108] A. Khabbazi-basmenj and S. A. Vorobyov, "Robust adaptive beamforming for general-rank signal model with positive semi-definite constraint via POTDC," submitted to *IEEE Trans. Signal Process.*. (see <http://arxiv.org/abs/1212.3624>)
- [109] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693-5705, Nov. 2010.
- [110] A. Hassanien and S. A. Vorobyov, "Why the phased-MIMO radar outperforms the phased-array and MIMO radars," in *Proc. 18th European Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1234-1238.
- [111] D. R. Fuhrmann, J. Browning, and M. Rangaswamy, "Signaling strategies for the hybrid MIMO phased-array radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 66-78, Feb. 2010.
- [112] D. Wilcox and M. Sellathurai, "On MIMO radar subarrayed transmit beamforming," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2076-2081, Apr. 2012.
- [113] T. Aittomaki and V. Koivunen, "Beampattern optimization by minimization of quartic polynomial," in *Proc. IEEE Workshop Statistical Signal Process. (SSP)*, Cardiff, U.K., Sept. 2009, pp. 437-440.

- [114] H. He, P. Stoica, and J. Li, “Designing unimodular sequence sets with good correlations—Including an application to MIMO radar,” *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4391–4405, Nov. 2009.
- [115] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 1456–1467, Jul. 1999.
- [116] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthonormality constraints,” *SIAM J. Matrix Anal. Applic.*, vol. 20, no. 2, pp. 303–353, 1998.
- [117] H. Manton, “Optimization algorithms exploiting unitary constraints,” *IEEE Trans. Signal Process.*, vol. 50, pp. 635–650, Mar. 2002.
- [118] T. E. Abrudan, J. Eriksson, and V. Koivunen, “Steepest descent algorithms for optimization under unitary matrix constraint,” *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1134–1147, Mar. 2008.
- [119] P. -A. Absil, R. Mahony, and R. Sepulchre, “Riemannian geometry of Grassmann manifolds with a view on algorithmic computation,” *Acta Applicandae Mathematicae*, vol. 80, no. 2, pp. 199–220, 2004.
- [120] J. Goldberg and H. Messer, “Inherent limitations in the localization of a coherently scattered source,” *IEEE Trans. Signal Process.*, vol. 46, pp. 3441–3444, Dec. 1998.
- [121] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” *Proc. of the IEEE*, vol. 60, pp. 926–935, 1972.
- [122] C. W. Xing, S. D. Ma, and Y. C. Wu, “On low complexity robust beamforming with positive semidefinite constraints,” *IEEE Trans. Signal Process.*, vol. 57, pp. 4942–4945, Dec. 2009.
- [123] A. L. Yuille and A. Rangarajan, “The concave-convex procedure,” *Neural Computation*, vol. 15, pp. 915–936, 2003.
- [124] S. Haykin, *Adaptive Filter Theory*. (3rd Edition). Prentice Hall, 1995.
- [125] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, “Shrinkage algorithms for MMSE covariance estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.
- [126] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, “Maximum likelihood estimation of a structured covariance matrix with a condition number constraint,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, Jun. 2012.
- [127] M. Steiner and K. Gerlach, “Fast converging adaptive processor for a structured covariance matrix,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, pp. 1115–1126, Oct. 2000.
- [128] M. S. Bazaraa, C. M. Shetty, and H. D. Sherali, *Nonlinear Programming: Theory and Algorithms*. 2nd ed. New York: Wiley, 1993.
- [129] Y. Nesterov and A. Nemirovsky, *Interior Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [130] Z. -Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang, “Approximation bounds for quadratic optimization with homogeneous quadratic constraints,” *SIAM J. Optim.*, vol. 18, no. 1, pp. 1–28, Feb. 2007.

- [131] A. Khabbazibasmenj, S. A. Vorobyov, and A. Hassanien, “Robust adaptive beamforming based on steering vector estimation with as little as possible prior information,” *IEEE Trans. Signal Process.*, vol. 60, pp. 2974–2987, Jun. 2012.
- [132] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. SIAM, U.S., 2004.
- [133] F. Roemer and M. Haardt, “Tensor-based channel estimation (TENICE) and iterative refinements for two-way relaying with multiple antennas and spatial reuse,” *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5720–5735, Nov. 2010.
- [134] L. Hogben, *Handbook of Linear Algebra*. CRC Press, 2007.
- [135] J. M. Ortega, *Matrix Theory: A Second Course*. Springer, 1987.
  
- [136] A. Khabbazibasmenj, S. A. Vorobyov and A. Hassanien, “Robust adaptive beamforming via estimating steering vector based on semidefinite relaxation,” in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 1102–1106.
- [137] S. -D. Wang, T. -S. Kuo, and C. -F. Hsu, “Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation,” *IEEE Trans. Automatic Control*, vol. 31, no. 7, pp. 654–656, Jul. 1986.
- [138] F. P. Kelly, A. K. Maulloo and D. K. H. Tan., “Rate control in communication networks: Shadow prices, proportional fairness and stability,” *J. Operational Research Society*, vol. 49, pp. 237–252, Apr. 1998.
- [139] T. Bonald, L. Massoulié, A. Proutiere and J. Virtamo, “A queueing analysis of max-min fairness, proportional fairness and balanced fairness,” *Queueing Systems*, no. 1, pp. 65–84, 2006.
- [140] H. Kim and Y. Han, “A proportional fair scheduling for multicarrier transmission systems,” *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 210–212, Mar. 2005.
- [141] J. Zhang, S. A. Vorobyov, A. Khabbazibasmenj, and M. Haardt, “Sum rate maximization in multi-operator two-way relay networks with a MIMO AF relay via POTDC,” *21st European Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, Sept. 9–13, 2013, accepted.
- [142] R. Jain, A. Durrezi, and G. Babic, “Throughput fairness index: An explanation,” Document no. ATM Forum/99-0045, Feb. 1999.
- [143] K. T. Phan, L. Le, S. A. Vorobyov, and T. Le-Ngoc, “Power allocation and admission control in multi-user relay networks via convex programming: Centralized and distributed schemes,” *EURASIP J. Wireless Commun. and Networking*, Special Issue: *Optim. Techniques in Wireless Commun.*, vol. 2009, 2009, Article ID 901965, 12 pages.
- [144] Y. Ye, *Conic Linear Optimization*, Course notes for CME336, Department of Management Science and Engineering, Stanford University.
- [145] Y. Huang and S. Zhang, “Complex matrix decomposition and quadratic programming,” *Math. Operations Research*, vol. 32, no. 3, pp. 758–768, Aug. 2007.
- [146] A. Shapiro, “First and second order analysis of nonlinear semidefinite programs,” *Math. Programming Ser. B*, vol. 77, pp. 301–320, 1997.