

Robust Minimum Dispersion Beamforming

ROBUST MINIMUM DISPERSION BEAMFORMING

BY

XUE JIANG, M.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright by Xue Jiang, August 2014

All Rights Reserved

Doctor of Philosophy (2014)
(Electrical & Computer Engineering)

McMaster University
Hamilton, Ontario, Canada

TITLE: Robust Minimum Dispersion Beamforming

AUTHOR: Xue Jiang
M.Sc., (Acoustics)
Institute of Acoustics, Chinese Academy of Sciences, Bei-
jing, China

SUPERVISOR: Dr. T. Kirubarajan

NUMBER OF PAGES: xv, 173

Dedicate to my husband, my parents and my grandparents

Abstract

Beamforming is a spatial filtering technique using a sensor array to enhance the signal of interest (SOI) and suppress interferences and noise. It is widely used in radar, sonar, wireless communications, Global Positioning System (GPS) navigation, microphone array speech processing and many other areas. Most existing beamforming approaches are based on the minimum variance (MV) criterion. The MV approach is statistically optimal only when the desired signal, interferences and the noise are Gaussian-distributed. However, many real-world signals are non-Gaussian. For non-Gaussian signals, the higher-order statistics or fractional lower-order statistics contain useful information and can be utilized to improve the beamformer performance. In this thesis, a family of the minimum dispersion (MD) criterion-based robust beamforming algorithms, which minimize the ℓ_p -norm ($p \geq 1$) of the array output subject to linear or nonlinear constraints, are proposed for non-Gaussian signals. The dispersion, which is a generalization of variance, implicitly exploits the higher-order statistics for $p > 2$ or fractional lower-order statistics for $p < 2$.

We utilize the MD criterion with a single linear constraint and multiple linear constraints, which gives us the minimum dispersion distortionless response (MDDR)

beamformer and linearly constrained minimum dispersion (LCMD) beamformer, respectively. The MDDR and LCMD beamformers can be tailored to Gaussian, sub-Gaussian or super-Gaussian signals and noise by adjusting the value of p . Three efficient iterative algorithms, namely, the iteratively reweighted MVDR (IR-MVDR), complex-valued full Newton's and partial Newton's methods, are devised to solve the resulting convex optimization problems.

We extend the LCMD beamformer to the quadratically constrained minimum dispersion (QCMD) beamformer. The robustness against model uncertainty of the QCMD beamformer is significantly enhanced compared with the LCMD beamformer. A gradient projection algorithmic framework is developed to efficiently solve the resulting convex optimization problem. Furthermore, we derive a closed-form expression of the projection onto the constraint set.

Note that sub-Gaussian signals that are frequently encountered in practical applications. Therefore, a minimum ℓ_∞ -norm criterion is then adopted by the robust linear programming beamformer (RLPB). In this way, the sub-Gaussianity of the signals can be fully exploited. We model the uncertainty region as a rhombus in which the ℓ_1 -norm of the steering vector error is bounded. As a result, the proposed RLPB beamformer can be obtained by solving a linear programming (LP) problem. We also present the theoretical explanation to the reason why the RLPB can implicitly exploit the high-order statistics from the statistical perspective.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Prof. T. Kirubarajan for his guidance and support during my stay in McMaster University. He led me to the field of robust signal processing and gave me the freedom to explore new ideas. He has taught me a lot of things I need to be a great researcher. He is also a perfect gentleman who is always nice, considerate, and polite.

I would like to thank Dr. A. Yasotharan at Defence Research and Development Canada for his support, comments and encouragement. I do appreciate his help to my research works. I am also grateful to Prof. H. C. So at City University of Hong Kong for his numerous helpful comments to my papers. I would like to thank the members of my Ph.D. defence and supervisory committee: Prof. D. Cassidy, Prof. A. Anpalagan, Prof. I. Bruce, and Prof. S. Sirouspour for their comments and kind suggestions to my research works and thesis.

I have learned a lot from the excited graduate courses I have taken and audited at McMaster offering by Prof. J. P. Reilly, Prof. M. Bakr, Prof. T. N. Davidson, and Prof. S. Haykin. I will keep in mind and never forget the encouraging words Prof. S. Haykin told me. Thank you. I also would like to express my gratitude to Prof. X. Wu, Prof. T. R. Field, Prof. Y. Haddara and Prof. T. Todd. It was a happy and rewarding experience to work with you as a teaching assistant. Many thanks to Prof.

C.-H. Chen and Prof. T. N. Davidson for their comments to my presentation and my research work during the poster seminar series for graduate students in ECE. My special thanks go to Prof. K. M. Wong and Prof. J. P. Reilly for providing strong supports for my future career. I am grateful to Prof. K. M. Wong for sharing his research experience with me, which is invaluable to me and my academic career. I want to thank my labmates and friends at McMaster. I will remember and miss all the time we spent together. I also would like to thank Ms. Cheryl Gies, our lovely secretary, for her kind assistance and professional support.

Last but not least, I would like to thank my parents and my grandparents for their unselfish love during my life. I hope to make them proud. I also would like to thank my husband for his endless love, encouragement, support, and tolerance. Without him I cannot be what I am today. I hope that we can accompany each other to enjoy the wonderful journey through our lives.

Notations

Bold upper case letter	Matrix
Bold lower case letter	Vector
<i>I</i>	Identity matrix
1	All-ones matrix or vector
0	All-zeros matrix or vector
$(\cdot)^*$	Complex conjugate
$(\cdot)^T$	Transpose
$(\cdot)^H$	Hermitian transpose
$j = \sqrt{-1}$	Imaginary unit
$\text{Re}(\cdot)$ or $(\cdot)_R$	Real part of a complex-valued number
$\text{Im}(\cdot)$ or $(\cdot)_I$	Imaginary part of a complex-valued number
$ \cdot $	Absolute value of a real number or the modulus of a complex number
$\ \cdot\ $	Euclidean norm (i.e., ℓ_2 -norm) of a vector
$\ \cdot\ _p$	ℓ_p -norm of a vector
$E\{\cdot\}$	Expectation operator
$\text{Pr}(\cdot)$	Probability operator

\mathbb{R}	Set of real numbers
\mathbb{R}^+	Set of nonnegative real numbers
\mathbb{C}	Set of complex number
$\lambda_{\max}(\cdot)$	Maximum eigenvalue of a square matrix
$\text{tr}(\cdot)$	Trace of a square matrix

Contents

Abstract	iv
Acknowledgements	vi
Notations	viii
1 Introduction	1
1.1 Background of Beamforming	1
1.2 Contributions of Our Works	4
1.3 Signal Model and Beamforming Techniques	6
1.3.1 Signal Model	6
1.3.2 Beamforming Techniques	8
1.4 Outlines of the Thesis	12
2 Linearly Constrained Minimum Dispersion Beamforming	14
2.1 Introduction	15
2.2 Minimum Dispersion Beamformer Via ℓ_p -Norm Minimization	19
2.2.1 Motivation by Non-Gaussianity	19
2.2.2 Minimum Dispersion Criterion	20

2.2.3	Iteratively Reweighted MVDR Algorithm	22
2.2.4	Complex-Valued Newton's Methods with Equality Constraint	24
2.2.5	MDDR Beamforming via ℓ_∞ -Norm Minimization	35
2.3	Extension to Multiple Linear Constraints	39
2.4	Simulation Results	42
2.4.1	Results with Perfect Steering Vector	43
2.4.2	Results with AOA Mismatch	55
2.4.3	Results with Sensor Position Mismatch	58
2.5	Conclusion	59

3	Gradient Projection for Quadratically Constrained Minimum Dispersion Beamforming	62
3.1	Introduction	63
3.2	MDDR and LCMD Beamformers	66
3.3	QCMD Beamformer	68
3.3.1	Formulation of QCMD	68
3.3.2	Fast PGM	71
3.3.3	ℓ_∞ -Norm QCMD Beamformer	81
3.4	Robust Constant Modulus Beamformer	83
3.5	Simulation Results	84
3.5.1	Convergence Behavior of PGM	86
3.5.2	Running Time Comparison	89
3.5.3	Random Steering Vector Mismatch	90
3.5.4	AOA Mismatch	95
3.6	Conclusion	96

4	Robust Beamforming by Linear Programming	99
4.1	Introduction	100
4.2	Robust Linear Programming Beamformer	104
4.2.1	ℓ_p -Modulus and New Definition of ℓ_p -Norm	105
4.2.2	Formulation	107
4.2.3	Solution Based on Linear Programming	109
4.2.4	Selection of Uncertainty Set Size ε_1	114
4.2.5	Performance Measures	118
4.3	Analysis From Statistical Viewpoint	121
4.4	Simulation Results	125
4.4.1	AOA Mismatch	127
4.4.2	Random Steering Vector Mismatch	135
4.4.3	Robustness to Parameter Selection	137
4.4.4	Robustness to Signal Properties	143
4.5	Conclusion	151
5	Conclusion	155

List of Figures

1.1	A uniform linear antenna array (ULA).	7
1.2	The diagram of narrowband beamforming.	8
2.1	Convergence rate versus number of iterations of the proposed iterative methods and gradient descent method	36
2.2	Convergence rate versus number of iterations using different initial values	37
2.3	Output SINR versus SNR for QPSK signals and additive Gaussian noise.	44
2.4	Output SINR versus number of snapshots for QPSK signals and additive Gaussian noise.	45
2.5	Output SINR versus number of sensors for QPSK signals and additive Gaussian noise.	46
2.6	Output SINR versus SNR for GGD signals and noise with $\beta = 0.4$	49
2.7	Output SINR versus number of snapshots for GGD signals and noise with $\beta = 0.4$	50
2.8	Output SINR versus number of sensors for GGD signals and noise with $\beta = 0.4$	51
2.9	Output SINR versus SNR for Gaussian signals and noise.	52
2.10	Output SINR versus SNR for QPSK signals and Gaussian noise with six interferers.	53

2.11	Output SINR versus interference number for QPSK signals and Gaussian noise.	54
2.12	Output SINR versus SNR for QPSK signals in the presence of 2° AOA mismatch.	56
2.13	Output SINR versus number of snapshots for QPSK signals in the presence of 2° AOA mismatch.	57
2.14	Output SINR versus AOA of SOI for 2° AOA mismatch.	58
2.15	Output SINR versus relative error of inter-sensor spacing.	60
3.1	Objective function versus number of iterations.	87
3.2	Output SINR versus number of iterations.	88
3.3	Output SINR versus SNR with random steering vector mismatch.	92
3.4	Output SINR versus number of samples with random steering vector mismatch.	93
3.5	Output SINR versus perturbation amount of steering vector.	94
3.6	Output SINR versus number of interferers.	95
3.7	Output SINR versus AOA mismatch.	97
4.1	Scatter plots of output constellations of the six beamformers	128
4.2	Output SINR versus SNR with 2° AOA mismatch.	129
4.3	NMSE versus SNR with 2° AOA mismatch.	130
4.4	Output SINR versus number of snapshots with 2° AOA mismatch.	131
4.5	NMSE versus number of snapshots with 2° AOA mismatch.	132
4.6	Output SINR versus AOA mismatch.	133
4.7	NMSE versus AOA mismatch.	134
4.8	Output SINR versus SNR with random steering vector mismatch.	137

4.9	NMSE versus SNR with random steering vector mismatch.	138
4.10	Output SINR versus number of snapshots with random steering vector mismatch.	139
4.11	NMSE versus number of snapshots with random steering vector mismatch.	140
4.12	Output SINR versus perturbation amount of steering vector.	141
4.13	Output SINR versus perturbation amount of steering vector.	142
4.14	Output SINR of RMVB versus ε_2 and RLPB versus ε_1	144
4.15	Output SINR versus modulation type	145
4.16	NMSE versus modulation type	146
4.17	Output SINR versus normalized CFO.	147
4.18	NMSE versus normalized CFO.	148
4.19	Output SINR versus SNR with hybrid modulation.	149
4.20	NMSE versus SNR with hybrid modulation.	150
4.21	Output SINR versus SNR with Gaussian signals and random steering vector mismatch.	152
4.22	NMSE versus SNR with Gaussian signals and random steering vector mismatch.	153

Chapter 1

Introduction

In this chapter, we review the basic concepts of beamforming technique and briefly describe the major results of each chapter. The chapter is organized as follows: Section 1.1 gives the background of beamforming. The contributions of our works are described in Section 1.2. In Section 1.3, the signal model is given and several representative beamforming techniques are briefly reviewed. Section 1.4 gives the outline of this thesis.

1.1 Background of Beamforming

Sensors may have different gains for signals transmitted to different angles and signals received from different angles. The directivity pattern of a sensor is a function of angles, which is called the beampattern. It is governed by the sensor's physical construction and cannot be changed once the sensor is built [1]. Consider a sensor with the beampattern that has a large gain around angle 45° but has small gains at other angles. We can use this sensor to receive the signal of interest (SOI) from 45° .

However, to receive the signals from other angles, we need to rotate the sensor to the desired direction mechanically, which is costly and usually slow.

To avoid mechanical rotation, we can use a technology called beamforming, which allows us to change the beampattern electronically. Beamforming is a spatial filtering technique [2] using a sensor array to enhance the SOI and suppress interferences and noise. Usually, we assume that all the sensors have omnidirectional beampatterns. Beamforming is widely used in radar, sonar, wireless communications, microphone array speech processing, Global Positioning System (GPS) navigation, biomedicine and many other areas [3–8]. It is applicable to either transmission or reception of signals. In this thesis, we consider the receive beamforming.

The conventional beamforming technique, i.e., Bartlett beamformer, dates back to the second world war [9]. It maximizes the power of the beamformer output of the desired signal, which yields a spatial matched filtering to the SOI. Since it is independent of the received signals, its capability for interference suppression is quite limited. The modern beamforming techniques are data-dependent, which obtain the weights vector by optimizing the performance of the beamformer based on the received data [10]. The data-dependent beamformers can achieve better resolution and much better interference mitigation capability than the data-independent ones. A classical data-dependent beamforming method is the minimum variance distortionless response (MVDR) beamformer [11]. It minimizes the output variance while keeping the response of the desired signal to unity. However, the performance of the MVDR beamformer degrades when the sample size is too small. Furthermore, it is quite sensitive to the steering vector mismatch, which is an unavoidable problem in practical applications because of a variety of reasons, such as angle-of-arrival

(AOA) mismatch, imperfect array calibration, source wavefront distortion, and distorted antenna shape [8,12]. The errors between the assumed steering vector and the true one may lead to a severe undesired attenuation of the SOI. This effect is commonly referred to as signal cancellation or signal self-nulling [13]. In [14], the linearly constrained minimum variance (LCMV) beamformer has been proposed by adding additional linear constraints to the MVDR beamformer to broaden the coverage around the nominal AOA. In this way, its robustness to AOA mismatch is enhanced. The drawback of the LCMV beamformer is that the degrees of freedom for interference suppression are reduced as more linear constraints are added [8]. Meanwhile, it cannot handle the case of arbitrary steering vector mismatch.

In order to improve the robustness against general steering vector mismatch, much effort has been made over the past three decades [10,15–36]. The eigenspace-based beamformer [15] is a powerful technique which is applicable to any type of mismatch. However, it loses efficiency as the signal-to-noise ratio (SNR) decreases or number of interferences increases. Robust beamforming techniques based on worst-case performance optimization are proposed in [18,20,24], where nonlinear constraints are employed instead of the linear ones. The key idea of [18,20,24] is to model the actual steering vector as the sum of the nominal steering vector and an uncertainty term, where the Euclidean norm of the uncertainty is upper-bounded. A spherical uncertainty region is used in [18] and a more general ellipsoidal uncertainty region is considered in robust minimum variance beamformer (RMVB) [24] and robust Capon beamformer (RCB) [20]. The problems of worst-case performance optimization with infinitely many nonconvex constraints are finally converted to a second-order cone programming (SOCP), which can be solved by the standard interior-point methods

(IPM). In addition, using the Lagrangian multiplier method, the RMVB and the RCB can be solved with a low computational complexity. Using the idea of worst-case performance optimization, a general-rank model has been suggested in [21] and [36], which is applicable to both rank-one and higher-rank SOI models.

1.2 Contributions of Our Works

Most existing works are based on second-order statistics [10, 15–30, 32, 34–36], which is only optimal for Gaussian signals and noises because a zero-mean Gaussian distribution is completely characterized by its second-order statistics. However, many real world signals are non-Gaussian [37, 38]. Based on the kurtosis of a distribution, non-Gaussian distributions can be classified into two categories, namely, sub-Gaussian with kurtosis smaller than three and super-Gaussian with kurtosis larger than three [39, 40]. Many signals that arise in wireless communications, radar, sonar, and GPS navigation are sub-Gaussian [37, 41]. On the other hand, common super-Gaussian signals include speech and biomedical data [40, 42]. Non-Gaussian noise is also frequently encountered in practice [43]. For non-Gaussian signals and noise, the higher- and lower-order statistics contain useful information and can be utilized to improve the beamformer performance. In this thesis, we focus on the robust beamforming techniques for non-Gaussian signals.

In this thesis, a family of the MD criterion-based beamforming algorithms are proposed. In statistics, $E\{|\cdot|^p\}$ is referred to as the dispersion, which is a generalization of variance. It implicitly exploits the higher-order statistics for $p > 2$ or fractional lower-order statistics for $p < 2$ [38]. Hence, the MD criterion can be tailored to Gaussian, sub-Gaussian or super-Gaussian signals and noise by choosing different values

of p . Combining the MD criterion with linear constraints or nonlinear constraints gives us different robust minimum dispersion beamforming techniques. To our best knowledge, there is no computationally simple and efficient numerical algorithm for the MD optimization problem. Therefore, several efficient algorithms, which converge fast and are computationally efficient, are also proposed.

Now we briefly summarize the contributions of our work as follows.

- i) The minimum dispersion distortionless response (MDDR) beamformer and linearly constrained minimum dispersion (LCMD) beamformer are proposed by using the MD criterion with a single linear constraint and multiple linear constraints, respectively. The LCMD beamformer is robust against AOA mismatch. The MDDR and LCMD beamformers can be considered as the extensions of the MVDR and LCMV beamformers from the Hilbert space to ℓ_p -space, respectively. Three efficient iterative algorithms, namely, the iteratively reweighted MVDR (IR-MVDR), complex-valued full Newton's and partial Newton's methods, are devised to solve the resulting ℓ_p -norm minimization problems with a linear constraint and multiple linear constraints.
- ii) We extend the LCMD beamformer to the quadratically constrained minimum dispersion (QCMD) beamformer. The robustness against model uncertainty of the QCMD beamformer is significantly enhanced compared with the LCMD beamformer. A gradient projection algorithmic framework is developed to efficiently solve the resulting convex optimization problem. Furthermore, we derive a closed-form expression of the projection onto the constraint set.
- iii) The minimum ℓ_∞ -norm criterion is then adopted by the robust linear programming beamformer (RLPB), which is proposed for sub-Gaussian signals that are

frequently encountered in many practical applications. We model the uncertainty region as a rhombus in which the ℓ_1 -norm of the steering vector error is bounded. As a result, the proposed RLPB beamformer can be obtained by solving a linear programming (LP) problem.

1.3 Signal Model and Beamforming Techniques

1.3.1 Signal Model

Consider an array of M receiving sensors. The narrowband far-field sources impinge on this array from different directions. The complex baseband signal received by the m th ($1 \leq m \leq M$) sensor at time n is denoted as $x_m(n)$. The vector of the array output $\mathbf{x}(n) = [x_1(n), \dots, x_M(n)]^T$ is expressed as

$$\mathbf{x}(n) = s(n)\mathbf{a} + \sum_{i=1}^I s_i(n)\mathbf{a}_i + \mathbf{v}(n) \quad (1.1)$$

where $s(n)$ is the SOI, $\{s_i(n)\}_{i=1}^I$ are the I interferences, $\mathbf{a} \in \mathbb{C}^M$ and $\{\mathbf{a}_i\}_{i=1}^I$ are the steering vectors of the SOI and interferences, respectively, and $\mathbf{v}(n)$ is the additive noise. The SOI is assumed to be uncorrelated with the interferences and noise. This assumption holds in many practical applications and it is also widely assumed in array processing, e.g., see [8], [15], [20] and [44]. Certainly, there exist cases where the signals are mutually correlated or coherent due to multipath propagation [45]. This case, which may be handled using spatial smoothing techniques [45], is beyond the scope of this thesis. We collect all the I interferences into a term $\mathbf{i}(n) = \sum_i s_i(n)\mathbf{a}_i$. Depending on the array configuration, the steering vector has different forms. For

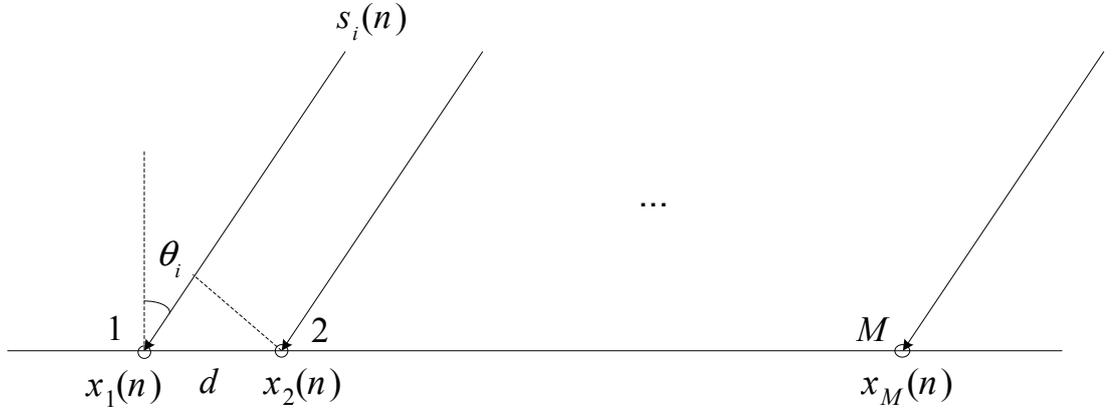


Figure 1.1: A uniform linear antenna array (ULA).

example, \mathbf{a} has the following form for a ULA depicted in Fig. 1.1

$$\mathbf{a}(\theta) = [1, e^{j(2\pi/\zeta)d\sin\theta}, \dots, e^{j(M-1)(2\pi/\zeta)d\sin\theta}]^T \quad (1.2)$$

where θ is the AOA, d is the inter-sensor spacing, and ζ is the wavelength.

The task of data-dependent beamforming is to design a beamformer $\mathbf{w} \in \mathbb{C}^M$ to enhance the SOI and suppress interference and noise using the observed data $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(N)] \in \mathbb{C}^{M \times N}$ with N being the number of snapshots. A general structure for narrowband beamforming is shown in Fig. 1.2. The output of the beamformer is expressed as

$$y(n) = \mathbf{w}^H \mathbf{x}(n). \quad (1.3)$$

It is desired that the output $y(n)$ preserves the desired signal component and mitigates the interference and noise.

The output signal-to-interferences-plus-noise ratio (SINR), which is taken as the

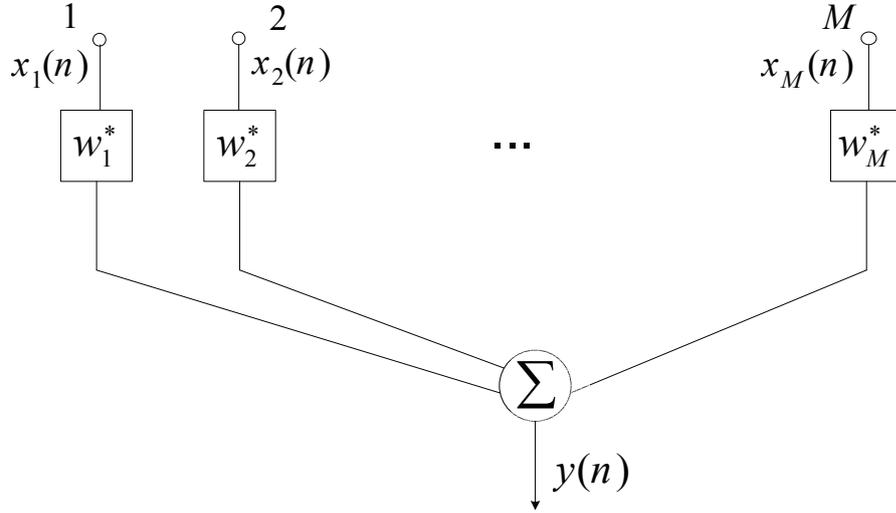


Figure 1.2: The diagram of narrowband beamforming.

performance measure of a beamformer, is defined as

$$\text{SINR} = \frac{\text{E} \left\{ |s(n)\mathbf{w}^H \mathbf{a}|^2 \right\}}{\text{E} \left\{ |\mathbf{w}^H (\mathbf{i}(n) + \mathbf{v}(n))|^2 \right\}} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{a}|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}} \quad (1.4)$$

where $\sigma_s^2 = \text{E}\{|s(n)|^2\}$ is the power of the SOI and \mathbf{R}_{i+n} is the interferences-plus-noise covariance matrix.

1.3.2 Beamforming Techniques

In this section, we review several representative data-dependent beamforming techniques, namely, MVDR beamformer [11], LCMV beamformer [14], subspace beamformer [15] and robust MV beamformer [18, 24]. This will help us clarify the relationships and differences between these well-known beamforming techniques and the proposed ones in Chapters 2–4.

1) *MVDR Beamformer*: The MVDR beamformer [11] maximizes the output SINR by minimizing the total output variance while constraining the SOI response to be unity, i.e.,

$$\begin{aligned} \min_{\mathbf{w}} \quad & (\mathbb{E}\{|y(n)|^2\} = \mathbf{w}^H \mathbf{R} \mathbf{w}) \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \quad (1.5)$$

where $\mathbf{R} = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ is the covariance matrix of $\mathbf{x}(n)$. Note that (1.5) is equivalent to minimizing $\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}$ subject to $\mathbf{a}^H \mathbf{w} = 1$ because

$$\mathbf{w}^H \mathbf{R} \mathbf{w} = \mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w} + \sigma_s^2 |\mathbf{w}^H \mathbf{a}|^2 = \mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w} + \sigma_s^2. \quad (1.6)$$

The closed-form solution of (1.5) is known as MVDR or Capon beamformer [11] and is given by

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}. \quad (1.7)$$

In practice, the true covariance matrix \mathbf{R} is unknown and it is estimated using N snapshots as

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^H(n) = \frac{1}{N} \mathbf{X} \mathbf{X}^H. \quad (1.8)$$

By replacing \mathbf{R} in (1.7) with its estimate $\hat{\mathbf{R}}$, we obtain the so-called sample matrix inversion (SMI) beamformer. In other words, the SMI beamformer is a practical implementation of the MVDR beamformer. When the sample size N is small, the covariance matrix \mathbf{R} cannot be estimated accurately and the performance of the SMI beamformer will degrade [8, 15, 46]. Moreover, it is sensitive to steering vector mismatches and its performance deteriorates in the presence of array mismatches [7, 8].

2) *LCMV Beamformer*: One cause for steering vector mismatch is the AOA estimation error. A remedy to address the AOA mismatch is to impose multiple linear constraints for a small spread of angles around the nominal AOA [14]. That is,

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{C}^H \mathbf{w} = \mathbf{g} \end{aligned} \quad (1.9)$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K] \in \mathbb{C}^{M \times K}$ contains K steering vectors or the derivatives of the steering vectors and $\mathbf{g} = [g_1, \dots, g_K]^T$ is usually taken as the vector with all elements being unity. The closed-form solution of the optimization problem in (1.9) is called the LCMV beamformer and is given by [14]

$$\mathbf{w}_{\text{LCMV}} = \mathbf{R}^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{g}. \quad (1.10)$$

When the number of linear constraints is $K = 1$, the LCMV beamformer reduces to the MVDR beamformer. The disadvantages of the LCMV beamformer are twofold. On one hand, the degrees of freedom for interference suppression are reduced as the linear constraints are added. On the other hand, it cannot handle the case of arbitrary steering vector mismatch [8].

3) *Subspace Beamformer*: To mitigate the adverse effects induced by the noise subspace disturbance, the subspace beamformer that uses only the signal-plus-interference subspace component of the sample correlation matrix is proposed in [15]. The eigenvalue decomposition (EVD) of \mathbf{R} is given by

$$\mathbf{R} = \mathbf{U} \mathbf{\Gamma} \mathbf{U}^H = \mathbf{U}_s \mathbf{\Gamma}_s \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Gamma}_n \mathbf{U}_n^H \quad (1.11)$$

where $\mathbf{U} = [\mathbf{U}_s, \mathbf{U}_n]$, $\mathbf{\Gamma}_s = \text{diag} \{\gamma_1, \dots, \gamma_{I+1}\}$ is a diagonal matrix containing the $I+1$ non-increasing principal eigenvalues and \mathbf{U}_s contains the corresponding orthonormal eigenvectors. Note that \mathbf{U}_n contains the $M - (I+1)$ orthonormal eigenvectors associated with the eigenvalues listed in the diagonal elements of $\mathbf{\Gamma}_n = \text{diag} \{\gamma_{I+2}, \dots, \gamma_M\}$. The range space spanned by \mathbf{U}_s is the signal-plus-interference subspace and its orthogonal complement, spanned by \mathbf{U}_n , is the noise subspace. The MVDR beamformer of (1.7) can be expressed in terms of eigenspace representation as [15]

$$\mathbf{w}_{\text{eig}} = c (\mathbf{U}_s \mathbf{\Gamma}_s^{-1} \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Gamma}_n^{-1} \mathbf{U}_n^H) \mathbf{a} \quad (1.12)$$

where c is a constant and does not affect the performance of a beamformer. The subspace beamformer further assumes that the noise is spatially white. This requires the noise covariance matrix to be expressed as $\sigma_v^2 \mathbf{I}$ with σ_v^2 being the noise variance. In this case, the eigenvalues of the noise subspace satisfies $\gamma_{I+2} = \dots = \gamma_M = \sigma_v^2$. Under this assumption, one can conclude that the steering vector \mathbf{a} is orthogonal to \mathbf{U}_n , i.e., $\mathbf{U}_n^H \mathbf{a} = \mathbf{0}$ [47]. Therefore (1.12) can be simplified as

$$\mathbf{w}_{\text{eig}} = \frac{\mathbf{U}_s \mathbf{\Gamma}_s^{-1} \mathbf{U}_s^H \mathbf{a}}{\mathbf{a}^H \mathbf{U}_s \mathbf{\Gamma}_s^{-1} \mathbf{U}_s^H \mathbf{a}}. \quad (1.13)$$

Note that \mathbf{w}_{eig} in (1.13) is referred to as the subspace beamformer [15]. A key issue with the subspace beamformer is the need to determine the dimension of signal-plus-interference subspace, which is equal to the value of the number of SOI plus interferences, i.e., $I+1$. The Akaike information criterion (AIC) or the minimum description length (MDL) criterion [48] can be applied to this source enumeration problem. The subspace beamformer is a powerful technique which is applicable to

any steering vector mismatch. However, it loses efficiency as SNR decreases or the number of interferences increases.

4) *Robust MV Beamformer*: Due to a variety of mismatches, the steering vector \mathbf{a} is not known exactly. The actual steering vector is expressed as

$$\mathbf{c} = \mathbf{a} + \mathbf{e} \quad (1.14)$$

where $\mathbf{e} \in \mathbb{C}^M$ is the steering vector error. It is assumed that the Euclidean norm of \mathbf{e} is upper-bounded, which corresponds to the case of spherical uncertainty set [18]. A more general ellipsoidal uncertainty region is considered in [20] and [24]. The goal of robust MV beamforming is to ensure that the magnitude response does not attenuate in the uncertainty set while minimizing the variance of the output, i.e.,

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R} \mathbf{w} \\ \text{s.t.} \quad & |(\mathbf{a} + \mathbf{e})^H \mathbf{w}| \geq 1, \text{ for all } \mathbf{e} \in \mathcal{E} \end{aligned} \quad (1.15)$$

where \mathcal{E} is the uncertainty region. Under the spherical or ellipsoidal model, the robust MV beamformer can be obtained by solving an SOCP problem [18].

1.4 Outlines of the Thesis

The remainder of this thesis is organized as follows. The MDDR beamformer and LCMD beamformer are proposed in Chapter 2. Three efficient iterative algorithms, namely, the IR-MVDR, complex-valued full Newton's and partial Newton's methods are also derived for solving the resultant ℓ_p -minimization problem with linear constraints. In Chapter 3, the QCMD beamforming technique is exploited and three

fast projected gradient methods (PGMs) are developed for solving the resulting ℓ_p -minimization problem with quadratic constraints. The RLPB for mismatch-robust beamforming is presented in Chapter 4. Finally, conclusions are drawn in Chapter 5. The main body of the thesis has been presented in the following journal papers:

- **X. Jiang**, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, “Minimum dispersion beamforming for non-Gaussian signals,” *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1879–1893, April 2014.
- **X. Jiang**, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, “Robust beamforming by linear programming,” *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1834–1849, April 2014.
- **X. Jiang**, W.-J. Zeng, A. Yasotharan, H. C. So, and T. Kirubarajan, “Gradient projection for robust minimum dispersion beamforming,” Submitted to *IEEE Transactions on Signal Processing*.

Chapter 2

Linearly Constrained Minimum Dispersion Beamforming

Most of the existing beamforming methods are based on the minimum variance (MV) criterion. The MV approach is statistically optimal only when the signal, interferences and the noise are Gaussian-distributed. However, non-Gaussian signals arise in a variety of practical applications. In this chapter, minimum dispersion distortionless response (MDDR) beamforming, which minimizes the ℓ_p -norm of the output while constraining the desired signal response to be unity, is devised for non-Gaussian signals. It is shown that the MDDR beamformer, which implicitly exploits non-Gaussianity, can improve the performance significantly if $p > 2$ for sub-Gaussian signals or $p < 2$ for super-Gaussian signals. Three efficient algorithms, the iteratively reweighted minimum variance distortionless response (IR-MVDR), complex-valued full Newton's and partial Newton's methods, are developed to solve the resulting ℓ_p -norm minimization with a linear constraint. Furthermore, the MDDR beamformer

with a single constraint is generalized to the linearly constrained minimum dispersion (LCMD) beamformer with multiple linear constraints, which exhibits robustness against steering vector mismatch. The LCMD beamformer yields significant performance improvement over the conventional linearly constrained minimum variance (LCMV) beamformer. Simulation results are provided to demonstrate the superior performance of the proposed minimum dispersion beamforming approaches. Most of the results of this chapter have been reported in our recent journal paper [49].

2.1 Introduction

Many existing data-dependent beamforming methods are based on the MV criterion [8,24]. In particular, the MVDR beamformer constrains the response of the SOI to be unity and minimizes the variance of the output [11]. It is generally recognized as an optimal beamformer since minimizing the output variance with distortionless response constraint is equivalent to maximizing the output signal-to-interferences-plus-noise ratio (SINR). However, the MVDR approach can achieve optimality only when the true covariance matrix is available [8]. This requires an infinite number of snapshots, which is impractical. In fact, the performance of the MVDR beamformer degrades significantly with short data length [8,46]. On the other hand, the eigenspace-based beamformer [15], which uses only the signal-plus-interference subspace component of the sample correlation matrix, can mitigate the adverse effects induced by the noise subspace disturbance, and hence performs better than the MVDR beamformer.

The MVDR and subspace beamformers exploit only the second-order statistics of the array output. The MV criterion is statistically optimal for Gaussian signals and noise because the first- and second-order statistics of a Gaussian distribution

contain all necessary statistical information. Nevertheless, many real-world signals are non-Gaussian [43, 52]. Based on the kurtosis of a distribution, non-Gaussian distributions can be classified into two categories, namely, sub-Gaussian with kurtosis smaller than three and super-Gaussian with kurtosis larger than three [39, 40]. Many signals that arise in wireless communications, radar, sonar, and GPS navigation are sub-Gaussian [37, 41]. On the other hand, common super-Gaussian signals include speech and biomedical data [40, 42]. Non-Gaussian noise is also frequently encountered in practice [43]. For non-Gaussian signals and noise, the higher-order and lower fractional order statistics contain useful information and can be utilized to improve the beamformer performance. In [44], a blind beamforming method was proposed for non-Gaussian signals using fourth-order cumulants. However, it uses only the fourth-order statistics and other statistical information is discarded. The ℓ_p -norm minimization criterion was proposed for beamforming in [53]. However, the goal of [53] was to suppress impulsive noise only with $1 \leq p < 2$. A smaller value of p was suggested in [53] although there is no theoretical justification. However, it will be shown in this chapter that this recommendation is ambiguous. Moreover, there is no efficient algorithm for solving the resulting ℓ_p -norm minimization. It uses the gradient descent method, but there is no discussion on the selection of the step size parameter. The special case with $p = 1$ was reconsidered in [31] while taking into account the steering vector error. Again, the gradient descent method was adopted to solve the optimization problem. In fact, the gradient descent scheme is quite slow and it may not even converge unless the step size is appropriate.

In this chapter, we do not explicitly construct any higher or lower-order statistics but adopt the minimum dispersion (MD) criterion for beamforming. The proposed

MDDR beamformer minimizes the ℓ_p -norm ($p \geq 1$) of the output while constraining the desired signal response to be unity. The dispersion, which is a generalization of variance, implicitly exploits the higher-order statistics for $p > 2$ or fractional lower-order statistics for $p < 2$ [38]. Compared with [53] and [31], the proposed MDDR beamformer can be tailored to Gaussian, sub-Gaussian or super-Gaussian signals and noise by choosing different values of p . We fully analyze the selection of p for signals and noise with different statistical properties. To our best knowledge, there is no computationally simple and efficient numerical algorithm for the MD beamforming problem. Therefore, three efficient iterative algorithms, which converge fast and are computationally efficient, are also proposed.

Another drawback of the standard MVDR beamformer is that it is too sensitive to steering vector mismatch [8, 18, 20, 24, 26]. The SOI will be considered as an interference and hence attenuated by the MVDR beamformer if the steering vector of the SOI is imprecise. Therefore its performance dramatically degrades under these conditions. One common cause of steering vector mismatch is due to the angle-of-arrival (AOA) estimation error. Several mismatch-robust beamforming approaches have been proposed [14, 18, 20, 24]. The LCMV beamformer [14] is a direct extension of the MVDR beamformer. It tries to cope with the AOA mismatch by imposing multiple linear constraints for a small spread of angles around the nominal AOA. Analogous to this, we derive the LCMD beamformer, which uses multiple linear constraints to enhance the robustness against steering vector uncertainty.

We briefly summarize the contributions of our work on MD beamforming as follows.

- i) It is pointed out that the MVDR approach is not statistically optimal in the

presence of non-Gaussian signals. We show that the proposed MDDR beamformer can effectively exploit the non-Gaussianity and hence considerably improve the performance if $p > 2$ for sub-Gaussian signals or $p < 2$ for super-Gaussian signals. We also discuss the case of $p \rightarrow \infty$, i.e., the ℓ_∞ -norm MDDR beamformer.

- ii) Three efficient iterative algorithms, the IR-MVDR and the complex-valued full Newton's and partial Newton's methods, are devised to solve the resulting ℓ_p -norm minimization problem with a linear constraint. The three algorithms converge fast and are computationally efficient. They have the same complexity as the MVDR beamformer. It is also shown that the IR-MVDR is a special case of the partial Newton's method with a fixed step size.
- iii) Note that many conventional optimization methods cannot directly handle complex-valued variables. They work by splitting the complex variables into real and imaginary parts, which may distort the special data structures used for implementation. Unlike these conventional methods, the proposed algorithms directly handle complex variables as a single entity.
- iv) We extend the MDDR beamformer with a single constraint to the LCMD beamformer with multiple linear constraints, which exhibits robustness against steering vector mismatch. The three algorithms for ℓ_p -norm minimization are accordingly generalized to the case of multiple constraints.

The remainder of this chapter is organized as follows. In Section 2.2, we present the minimum dispersion beamformer based on ℓ_p -norm minimization. Three iterative algorithms, namely, the IR-MVDR and two complex-valued Newton's methods,

are developed for the efficient computation of the MDDR beamformer. In Section 2.3, the minimum dispersion beamformer is extended to multiple linear constraints. Computer simulations are performed to demonstrate the effectiveness of the proposed beamformer for non-Gaussian signals in Section 2.4. Finally, conclusions are presented in Section 2.5.

2.2 Minimum Dispersion Beamformer Via ℓ_p -Norm Minimization

2.2.1 Motivation by Non-Gaussianity

The MVDR and subspace beamformers utilize only the second-order statistics. For Gaussian signals and noise, the MV criterion is statistically optimal because the first- and second-order statistics of a Gaussian distribution contain all necessary and sufficient statistical information. However, many signals in practice are non-Gaussian distributed. Random signals can be classified into three classes according to the kurtosis [40]. The kurtosis of a random stationary signal $s(n)$ with zero-mean is defined as

$$\kappa(s(n)) = \frac{\mathbb{E}\{|s(n)|^4\}}{(\mathbb{E}\{|s(n)|^2\})^2}. \quad (2.1)$$

If $s(n)$ is Gaussian, then $\kappa(s(n)) = 3$. If $\kappa(s(n)) < 3$, $s(n)$ is sub-Gaussian. There are a number of sub-Gaussian distributions such as uniform distribution and Bernoulli distribution. If $\kappa(s(n)) > 3$, $s(n)$ is super-Gaussian. Super-Gaussian distributions, which include Laplace distribution and α -stable distribution [54], are also common. For example, the phase shift keying (PSK) and quadrature amplitude modulation

(QAM), radar, sonar, and GPS navigation signals are sub-Gaussian [37,41]. A common example of super-Gaussian signal is speech [40]. In addition to non-Gaussian signals, non-Gaussian noise is also frequently encountered [43]. For non-Gaussian signals, the higher-order (higher than 2) and lower fractional order statistics contain useful information and can be exploited to improve the performance of beamforming. In [44], a blind beamforming technique was proposed for non-Gaussian signals using the fourth-order cumulants. In the following, we will introduce the minimum dispersion beamforming, which implicitly uses the higher-order or fractional lower-order statistics.

2.2.2 Minimum Dispersion Criterion

The proposed MDDR beamformer is obtained by solving the following linearly constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \text{E} \{ |\mathbf{w}^H \mathbf{x}(n)|^p \} \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \tag{2.2}$$

where $p \geq 1$. Clearly, the MDDR beamformer is reduced to the MVDR beamformer for $p = 2$. In statistics, $\text{E}\{|y(n)|^p\}$ is referred to as dispersion, which is a generalization of variance [38]. Therefore, we call the solution of (2.2) as the minimum dispersion beamformer. It should be pointed out that the criterion of (2.2) has been proposed for beamforming in [53]. Furthermore, $p = 1$ is adopted in the beamformer of [31]. However, no insightful guideline on how to choose an appropriate p is given in [53] and it advocates to use small value of p , which is not technically accurate because the choice of p is not only related to statistical characteristics of the noise but also the

signal sources. Also at present, there is no efficient numerical algorithms to solve the problem in (2.2). In the following, we will develop three low complexity algorithms with fast convergence to calculate the MDDR beamformer and discuss how to choose an appropriate p .

Replacing the expectation with the sample mean and ignoring the constant $1/N$, (2.2) can be rewritten as

$$\begin{aligned} \min_{\mathbf{w}} \quad & (f_p(\mathbf{w}) = \|\mathbf{X}^H \mathbf{w}\|_p^p) \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \quad (2.3)$$

where $\mathbf{X}^H \mathbf{w} = \mathbf{y}^*$ is the conjugate of the beamformer output $\mathbf{y} = [y(1), \dots, y(N)]^T$, and the ℓ_p -norm is defined as

$$\|\mathbf{y}\|_p = \left(\sum_{n=1}^N |y(n)|^p \right)^{1/p}. \quad (2.4)$$

However, (2.3) has no closed-form solution except for $p = 2$. The optimization problem in (2.3) is convex for $p \geq 1$ and the global optimum is guaranteed using the standard interior point method for convex optimization [55]. However, we propose simpler and more efficient algorithms for solving it. Note that the ℓ_p -norm minimization of (2.3) is a constrained optimization problem, but different from the one encountered in robust linear regression using least ℓ_p -norm [56, 57], which is unconstrained. Before discussing the optimization algorithms, we give general guidelines for selecting p . For Gaussian signals, the optimal p is 2. For sub-Gaussian signals, $p > 2$ will achieve better performance, whereas $p < 2$ is preferred for super-Gaussian signals. The optimal value of p depends on the probability density function (PDF) of the signals. In the simulation examples, we will further investigate the selection of p .

For signals with very strong super-Gaussianity, e.g., heavy-tailed distributions, like Laplacian [58] or α -stable distribution [54], one may require $0 < p < 1$. However, $0 < p < 1$ leads to a non-differential and nonconvex ℓ_p -norm minimization problem. It has been pointed out that (2.3) with $0 < p < 1$ is strongly NP-hard and that the global minimum is difficult to obtain [59]. Due to the mathematical difficulty, we do not consider the choice of $0 < p < 1$ in this work.

2.2.3 Iteratively Reweighted MVDR Algorithm

We rewrite the objective function in (2.3) as

$$\begin{aligned} f_p(\mathbf{w}) &= \|\mathbf{y}^*\|_p^p = \sum_{n=1}^N |y(n)|^p = \sum_{n=1}^N |y(n)|^{p-2} |y(n)|^2 \\ &= \|\Phi \mathbf{y}\|^2 = \|\Phi \mathbf{y}^*\|^2 \end{aligned} \quad (2.5)$$

where Φ is a diagonal weighting matrix

$$\Phi = \text{diag} \{ |y(1)|^{(p-2)/2}, \dots, |y(N)|^{(p-2)/2} \} \quad (2.6)$$

with its diagonal elements being real and positive numbers. This means that the ℓ_p -norm minimization problem can be converted into an ℓ_2 -norm minimization one.

Equation (2.5) can be further expressed as

$$f_p(\mathbf{w}) = \mathbf{y}^T \Phi^H \Phi \mathbf{y}^* = \mathbf{y}^T \mathbf{D}(\mathbf{w}) \mathbf{y}^* = \mathbf{w}^H \mathbf{X} \mathbf{D}(\mathbf{w}) \mathbf{X}^H \mathbf{w} \quad (2.7)$$

where

$$\mathbf{D}(\mathbf{w}) = \text{diag} \{ |y(1)|^{p-2}, \dots, |y(N)|^{p-2} \}. \quad (2.8)$$

Note that \mathbf{D} depends on the unknown \mathbf{w} because it is related to \mathbf{y} . Therefore, it is a function of \mathbf{w} , which is written as $\mathbf{D}(\mathbf{w})$. Now we rewrite (2.3) as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H (\mathbf{X}\mathbf{D}(\mathbf{w})\mathbf{X}^H) \mathbf{w} \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \quad (2.9)$$

whose optimal solution is given by

$$\mathbf{w} = \frac{(\mathbf{X}\mathbf{D}(\mathbf{w})\mathbf{X}^H)^{-1} \mathbf{a}}{\mathbf{a}^H (\mathbf{X}\mathbf{D}(\mathbf{w})\mathbf{X}^H)^{-1} \mathbf{a}}. \quad (2.10)$$

Equation (2.10) has a structure similar to that of the MVDR beamformer but the covariance matrix has been reweighted using the weighting matrix $\mathbf{D}(\mathbf{w})$. However, we cannot obtain a closed-form expression for the optimal \mathbf{w} since $\mathbf{D}(\mathbf{w})$ is related to the unknown \mathbf{w} . An alternative is to use the following fixed-point iteration

$$\mathbf{w}^{k+1} = \frac{(\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1} \mathbf{a}}{\mathbf{a}^H (\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1} \mathbf{a}} \quad (2.11)$$

to find the optimal solution, where the superscript $(\cdot)^k$ is used to denote the result at the k th ($k = 0, 1, \dots$) iteration. In each iteration, the MVDR beamformer with a reweighted covariance matrix is computed. Therefore, we refer to this algorithm as IR-MVDR, which is summarized in Algorithm 1. The initial value can be taken as the data-independent beamformer

$$\mathbf{w}^0 = \mathbf{a} / \|\mathbf{a}\|^2. \quad (2.12)$$

Next, we analyze the computational complexity of the IR-MVDR method. The com-

Algorithm 1 IR-MVDR

Given the received data \mathbf{X} , error tolerance ϵ , and maximum iteration number K_{\max} .

Initialize: $\mathbf{w}^0 = \mathbf{a}/\|\mathbf{a}\|^2$.

for $k = 0, 1, 2, \dots$ **do**

 Compute output $\mathbf{y} = (\mathbf{X}^H \mathbf{w}^k)^*$ and weighting matrix

$$\mathbf{D}(\mathbf{w}^k) = \text{diag} \{ |y(1)|^{p-2}, \dots, |y(N)|^{p-2} \}.$$

 Update beamformer: $\mathbf{w}^{k+1} = \frac{(\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1}\mathbf{a}}{\mathbf{a}^H(\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1}\mathbf{a}}$.

Stop if $\|\mathbf{w}^{k+1} - \mathbf{w}^k\|/\|\mathbf{w}^{k+1}\| < \epsilon$ or $k > K_{\max}$ is satisfied.

end for

plexity of matrix multiplication $\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H$ is $\mathcal{O}(NM^2)$ because $\mathbf{D}(\mathbf{w}^k)$ is diagonal. The computational cost for calculating $(\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1}\mathbf{a}$ is $\mathcal{O}(M^3)$. Hence the complexity of IR-MVDR is $\max(\mathcal{O}(NM^2), \mathcal{O}(M^3))$ in each iteration. Since it is always assumed that the sample size is larger than the number of sensors, i.e., $N > M$, then the complexity is $\mathcal{O}(NM^2)$ of each iteration.

Remark 2.1: The convergence behavior of the IR-MVDR is similar to that of the iteratively reweighted least-squares (IRLS) algorithm for unconstrained ℓ_p -norm minimization [56, 57]. Their convergence is guaranteed only for some specific values of p . We find that it does not converge when $p \geq 3.4$. This means that 3.4 is the critical value of the IR-MVDR scheme.

2.2.4 Complex-Valued Newton's Methods with Equality Constraint

Despite the simplicity of the IR-MVDR algorithm, it may not converge [60]. In this subsection, we propose two Newton's methods, whose convergence is guaranteed for $p > 1$, to efficiently solve the complex-valued ℓ_p -norm minimization of (2.3). It is

worth pointing out that the proposed Newton's methods are different from those in [61]. The Newton's methods of [61] can deal only with the unconstrained ℓ_p -norm minimization problem, but the proposed schemes can handle the problem with equality constraints.

The optimization problem of (2.3) involves complex-valued variables. We first give the definition of the gradient with respect to (w.r.t.) complex-valued variables. The gradient of the objective $f_p(\mathbf{w})$ w.r.t. the complex vector $\mathbf{w} \in \mathbb{C}^M$ is defined as

$$\nabla f_p(\mathbf{w}) = \frac{\partial f_p(\mathbf{w})}{\partial \mathbf{w}^*} = \left[\frac{\partial f_p}{\partial w_1^*}, \dots, \frac{\partial f_p}{\partial w_M^*} \right]^T \quad (2.13)$$

where

$$\frac{\partial f_p}{\partial w_i^*} = \frac{1}{2} \left(\frac{\partial f_p}{\partial \text{Re}(w_i)} + j \frac{\partial f_p}{\partial \text{Im}(w_i)} \right), \quad i = 1, \dots, M.$$

The Newton's method uses the following iteration to find the minimizer of (2.3):

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w} \quad (2.14)$$

where \mathbf{w} is the point of current iteration, and $\Delta \mathbf{w}$ is the update direction, which is called the Newton direction [62]. It requires that \mathbf{w} be feasible in each iteration, which is equivalent to requiring

$$\mathbf{a}^H \Delta \mathbf{w} = 0. \quad (2.15)$$

The second-order Taylor expansion of $f_p(\mathbf{w} + \Delta \mathbf{w})$ around the complex vector \mathbf{w} is given by

$$f_p(\mathbf{w} + \Delta \mathbf{w}) = f_p(\mathbf{w}) + q_{\mathbf{w}}(\Delta \mathbf{w}) + o(\|\Delta \mathbf{w}\|^2). \quad (2.16)$$

Here, $q_{\mathbf{w}}(\Delta \mathbf{w})$ is a quadratic function w.r.t. $\Delta \mathbf{w}$, which contains the first two order

expansion terms and can be expressed as

$$\begin{aligned}
q_{\mathbf{w}}(\Delta\mathbf{w}) &= \nabla f_p(\mathbf{w})^H \Delta\mathbf{w} + \nabla f_p(\mathbf{w})^T \Delta\mathbf{w}^* \\
&+ \frac{1}{2} \Delta\mathbf{w}^H \mathbf{H}_{\mathbf{w}^*\mathbf{w}} \Delta\mathbf{w} + \frac{1}{2} \Delta\mathbf{w}^T \mathbf{H}_{\mathbf{w}\mathbf{w}^*} \Delta\mathbf{w}^* \\
&+ \frac{1}{2} \Delta\mathbf{w}^H \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} \Delta\mathbf{w}^* + \frac{1}{2} \Delta\mathbf{w}^T \mathbf{H}_{\mathbf{w}\mathbf{w}} \Delta\mathbf{w}
\end{aligned} \tag{2.17}$$

where

$$\nabla f_p(\mathbf{w}) = \frac{\partial f_p(\mathbf{w})}{\partial \mathbf{w}^*} = \frac{p}{2} \mathbf{X} \mathbf{D}(\mathbf{w}) \mathbf{X}^H \mathbf{w} \tag{2.18}$$

is the gradient of $f_p(\mathbf{w})$, and the four $M \times M$ partial Hessian matrices are expressed as

$$\begin{aligned}
\mathbf{H}_{\mathbf{w}^*\mathbf{w}} &= \frac{\partial^2 f_p(\mathbf{w})}{\partial \mathbf{w}^* \partial \mathbf{w}^T} = \frac{p^2}{4} \mathbf{X} \mathbf{D}(\mathbf{w}) \mathbf{X}^H, \\
\mathbf{H}_{\mathbf{w}\mathbf{w}} &= \frac{\partial^2 f_p(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T} = \frac{p(p-2)}{4} \mathbf{X}^* \mathbf{E}(\mathbf{w}) \mathbf{X}^H, \\
\mathbf{H}_{\mathbf{w}\mathbf{w}^*} &= \mathbf{H}_{\mathbf{w}^*\mathbf{w}}, \quad \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} = \mathbf{H}_{\mathbf{w}\mathbf{w}}
\end{aligned} \tag{2.19}$$

where $\mathbf{E}(\mathbf{w})$ is a diagonal matrix of the form

$$\mathbf{E}(\mathbf{w}) = \text{diag} \{ |y(1)|^{p-4} y^2(1), \dots, |y(N)|^{p-4} y^2(N) \}. \tag{2.20}$$

The partial Hessians $\mathbf{H}_{\mathbf{w}^*\mathbf{w}}$ and $\mathbf{H}_{\mathbf{w}\mathbf{w}^*}$ are positive definite because $\mathbf{D}(\mathbf{w})$ in (2.8) is positive definite. Then, $q_{\mathbf{w}}(\Delta\mathbf{w})$ can be written more compactly as

$$\begin{aligned}
q_{\mathbf{w}}(\Delta\mathbf{w}) &= [\nabla f_p(\mathbf{w})^H, \nabla f_p(\mathbf{w})^T] \begin{bmatrix} \Delta\mathbf{w} \\ \Delta\mathbf{w}^* \end{bmatrix} + \\
&[\Delta\mathbf{w}^H, \Delta\mathbf{w}^T] \begin{bmatrix} \mathbf{H}_{\mathbf{w}^*\mathbf{w}} & \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} \\ \mathbf{H}_{\mathbf{w}\mathbf{w}} & \mathbf{H}_{\mathbf{w}\mathbf{w}^*} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{w} \\ \Delta\mathbf{w}^* \end{bmatrix}
\end{aligned} \tag{2.21}$$

where the $2M \times 2M$ full Hessian matrix is denoted as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\mathbf{w}^*\mathbf{w}} & \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} \\ \mathbf{H}_{\mathbf{w}\mathbf{w}} & \mathbf{H}_{\mathbf{w}\mathbf{w}^*} \end{bmatrix}. \quad (2.22)$$

The full Hessian matrix is positive definite when $p > 1$. It is noticed that the two off-diagonal block matrices $\mathbf{H}_{\mathbf{w}^*\mathbf{w}^*}$ and $\mathbf{H}_{\mathbf{w}\mathbf{w}}$ become zero if $p = 2$. In this case, these two partial Hessian matrices contain no information. When $p \neq 2$, these two matrices do not vanish and contain useful information for optimization.

1) *Full Newton's Method:* For fixed \mathbf{w} in the current iteration, the Newton's method aims to find an update direction $\Delta\mathbf{w}$ that minimizes the quadratic function $q_{\mathbf{w}}(\Delta\mathbf{w})$ under the linear constraint. That is,

$$\begin{aligned} \min_{\Delta\mathbf{w}} \quad & q_{\mathbf{w}}(\Delta\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{a}^H \Delta\mathbf{w} = 0. \end{aligned} \quad (2.23)$$

We use the method of Lagrangian multipliers [62] to solve the constrained optimization problem. The Lagrangian function of (2.23) is

$$\mathcal{L}(\Delta\mathbf{w}, \lambda) = q_{\mathbf{w}}(\Delta\mathbf{w}) + \lambda \mathbf{a}^H \Delta\mathbf{w} \quad (2.24)$$

where λ is the Lagrangian multiplier. In optimization with complex-valued variables, the unknown $\Delta\mathbf{w}$ and its conjugate $\Delta\mathbf{w}^*$ are jointly considered. According to the

optimal condition and using (2.17), we obtain

$$\begin{aligned}\frac{\partial \mathcal{L}(\Delta \mathbf{w}, \lambda)}{\partial \Delta \mathbf{w}^*} &= \nabla f_p(\mathbf{w}) + \mathbf{H}_{\mathbf{w}^* \mathbf{w}} \Delta \mathbf{w} + \mathbf{H}_{\mathbf{w}^* \mathbf{w}^*} \Delta \mathbf{w}^* + \lambda \mathbf{a} = \mathbf{0}, \\ \frac{\partial \mathcal{L}(\Delta \mathbf{w}, \lambda)}{\partial \Delta \mathbf{w}} &= \nabla f_p(\mathbf{w})^* + \mathbf{H}_{\mathbf{w} \mathbf{w}^*} \Delta \mathbf{w}^* + \mathbf{H}_{\mathbf{w} \mathbf{w}} \Delta \mathbf{w} + \lambda \mathbf{a}^* = \mathbf{0}, \\ \frac{\partial \mathcal{L}(\Delta \mathbf{w}, \lambda)}{\partial \lambda} &= \mathbf{a}^H \Delta \mathbf{w} = 0\end{aligned}\quad (2.25)$$

which can be written compactly as

$$\begin{bmatrix} \mathbf{H}_{\mathbf{w}^* \mathbf{w}} & \mathbf{H}_{\mathbf{w}^* \mathbf{w}^*} & \mathbf{a} \\ \mathbf{H}_{\mathbf{w} \mathbf{w}} & \mathbf{H}_{\mathbf{w} \mathbf{w}^*} & \mathbf{a}^* \\ \mathbf{a}^H & \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \mathbf{w}^* \\ \lambda \end{bmatrix} = - \begin{bmatrix} \nabla f_p(\mathbf{w}) \\ \nabla f_p(\mathbf{w})^* \\ 0 \end{bmatrix}. \quad (2.26)$$

By solving the linear system of (2.26), the Newton direction $\Delta \mathbf{w}$ is obtained. Proposition 2.1 guarantees the uniqueness of the solution of (2.26).

Proposition 2.1: There is a unique solution for (2.25) or (2.26) with $p > 1$. Hence, the update direction of the full Newton's method given by (2.25) is unique. In other words, the coefficient matrix of (2.26) must be nonsingular.

Proof: It follows from (2.25) that

$$\begin{bmatrix} \mathbf{H}_{\mathbf{w}^* \mathbf{w}} & \mathbf{H}_{\mathbf{w}^* \mathbf{w}^*} \\ \mathbf{H}_{\mathbf{w} \mathbf{w}} & \mathbf{H}_{\mathbf{w} \mathbf{w}^*} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \mathbf{w}^* \end{bmatrix} = - \begin{bmatrix} \nabla f_p(\mathbf{w}) + \lambda \mathbf{a} \\ \nabla f_p(\mathbf{w})^* + \lambda \mathbf{a}^* \end{bmatrix}. \quad (2.27)$$

For $p > 1$, the full Hessian matrix in (2.22) is positive and hence nonsingular since the objective function $f_p(\mathbf{w})$ is strictly convex. Then, the linear system of equations

of (2.27) has a unique solution:

$$\begin{bmatrix} \Delta \mathbf{w} \\ \Delta \mathbf{w}^* \end{bmatrix} = -\mathbf{H}^{-1} \left(\begin{bmatrix} \nabla f_p(\mathbf{w}) \\ \nabla f_p(\mathbf{w})^* \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{a} \\ \mathbf{a}^* \end{bmatrix} \right). \quad (2.28)$$

Taking the complex conjugate of both sides of the constraint $\mathbf{a}^H \Delta \mathbf{w} = 0$ yields $\mathbf{a}^T \Delta \mathbf{w}^* = 0$. Hence, we obtain the equivalent constraint

$$[\mathbf{a}, \mathbf{a}^*]^H \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \mathbf{w}^* \end{bmatrix} = 0. \quad (2.29)$$

Substituting (2.28) into (2.29) leads to a unique solution of the Lagrangian multiplier

$$\lambda = \frac{[\mathbf{a}, \mathbf{a}^*]^H \mathbf{H}^{-1} \begin{bmatrix} \nabla f_p(\mathbf{w}) \\ \nabla f_p(\mathbf{w})^* \end{bmatrix}}{[\mathbf{a}, \mathbf{a}^*]^H \mathbf{H}^{-1} \begin{bmatrix} \mathbf{a} \\ \mathbf{a}^* \end{bmatrix}}. \quad (2.30)$$

Note that the denominator of (2.30), which is a quadratic form associated with \mathbf{H}^{-1} , must be positive and hence impossible to be zero because the inverse of the full Hessian \mathbf{H}^{-1} is positive definite and $\mathbf{a} \neq \mathbf{0}$. This guarantees the uniqueness of λ . Now it is clear that $(\Delta \mathbf{w}, \Delta \mathbf{w}^*, \lambda)$ has a unique solution and the update direction is uniquely determined. This also means that the coefficient matrix of (2.26) is nonsingular. \square

The full Newton's method exploits all the partial Hessian matrices, which is the reason for its name. The complexity of solving (2.26) is $\mathcal{O}(2M + 1)^3 = \mathcal{O}(M^3)$ since the size of the system is $(2M + 1) \times (2M + 1)$. Recalling that the cost of calculating

the partial Hessian matrices is $\mathcal{O}(NM^2)$, the complexity of the full Newton's method is thus $\max(\mathcal{O}(NM^2), \mathcal{O}(M^3))$ in each iteration, which is the same as that of the IR-MVDR algorithm. After obtaining the Newton direction $\Delta\mathbf{w}^k$ in the k th iteration, the beamformer is updated as

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \mu_k \Delta\mathbf{w}^k \quad (2.31)$$

where $\mu_k \geq 0$ is the step size. The selection of μ_k is an important issue. In the conventional Newton's method [62], the fixed step size $\mu_k = 1$ is adopted, which is clearly not optimal. For a given Newton direction $\Delta\mathbf{w}^k$, the optimal step size is given by solving the line search

$$\mu_k = \arg \min_{\mu \geq 0} \|\mathbf{X}^H (\mathbf{w}^k + \mu \Delta\mathbf{w}^k)\|_p^p. \quad (2.32)$$

Recalling that $\mathbf{X}^H \mathbf{w}^k = \mathbf{y}^*$ is the conjugate of the array output at the k th iteration and denoting the output increment as $\mathbf{X}^H \Delta\mathbf{w}^k = \Delta\mathbf{y}^*$, the objective function w.r.t. the step size μ can be written as

$$\min_{\mu \geq 0} f_p(\mu) = \|\mathbf{y}^* + \mu \Delta\mathbf{y}^*\|_p^p \quad (2.33)$$

which is a simple one-dimensional optimization problem and can be easily solved by traditional line search techniques such as the golden section search or the tangential method [63]. The global optimality of μ is guaranteed since $f_p(\mu)$ is unimodal w.r.t. μ if $p \geq 1$. The line search has a marginal computational cost of $\mathcal{O}(N)$.

If the initial value \mathbf{w}^0 is feasible, then the solution of each iteration \mathbf{w}^k is feasible

due to the constraint $\mathbf{a}^H \Delta \mathbf{w} = 0$. We can initialize the beamformer using the sample matrix inversion (SMI) beamformer [15]

$$\mathbf{w}_{\text{SMI}} = \frac{(\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{a}}{\mathbf{a}^H (\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{a}} \quad (2.34)$$

or just initialize it as the data-independent beamformer of (2.12). These two initializations ensure that \mathbf{w}^0 is feasible. The full Newton's method for MDDR beamforming is summarized in Algorithm 2. The convergence of the Newton's method has been already proved [62–64]. In particular, it converges to the global minimum with a quadratic convergence rate if the point is sufficiently close to the optimum since the problem of (2.3) is convex [62].

2) *Partial Newton's Method*: As a simplification of the full Newton's method, the partial Newton's method ignores $\mathbf{H}_{\mathbf{w}\mathbf{w}}$ and $\mathbf{H}_{\mathbf{w}^*\mathbf{w}^*}$ by assuming them as $\mathbf{0}$. Then the linear system of (2.25) reduces to

$$\mathbf{H}_{\mathbf{w}^*\mathbf{w}} \Delta \mathbf{w} + \lambda \mathbf{a} = -\nabla f_p(\mathbf{w}) \quad (2.35)$$

$$\mathbf{a}^H \Delta \mathbf{w} = 0 \quad (2.36)$$

From (2.35), we obtain

$$\Delta \mathbf{w} = -\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1} (\nabla f_p(\mathbf{w}) + \lambda \mathbf{a}). \quad (2.37)$$

Substituting (2.37) into (2.36), λ is solved as

$$\lambda = -\frac{\mathbf{a}^H \mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1} \nabla f_p(\mathbf{w})}{\mathbf{a}^H \mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1} \mathbf{a}}. \quad (2.38)$$

Substituting (2.38) back into (2.37) leads to a closed-form solution of the update

Algorithm 2 Full Newton's method for MDDR beamforming

Given the received data \mathbf{X} and error tolerance ϵ .

Initialize: $\mathbf{w}^0 = \mathbf{a}/\|\mathbf{a}\|^2$ or set \mathbf{w}^0 using SMI beamformer of (2.34).

for $k = 0, 1, 2, \dots$ **do**

 Compute output $\mathbf{y} = (\mathbf{X}^H \mathbf{w}^k)^*$ and construct two diagonal matrices

$$\mathbf{D}(\mathbf{w}^k) = \text{diag} \{ |y(1)|^{p-2}, \dots, |y(N)|^{p-2} \}$$

$$\mathbf{E}(\mathbf{w}^k) = \text{diag} \{ |y(1)|^{p-4} y^2(1), \dots, |y(N)|^{p-4} y^2(N) \}.$$

 Calculate gradient and Hessian matrices

$$\nabla f_p(\mathbf{w}^k) = \frac{p}{2} \mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H \mathbf{w}^k$$

$$\mathbf{H}_{\mathbf{w}^* \mathbf{w}} = \frac{p^2}{4} \mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H$$

$$\mathbf{H}_{\mathbf{w} \mathbf{w}} = \frac{p(p-2)}{4} \mathbf{X}^* \mathbf{E}(\mathbf{w}^k) \mathbf{X}^H.$$

 Solve (2.26) to obtain Newton direction $\Delta \mathbf{w}^k$.

 Determine optimal step size μ_k by (2.33).

 Update beamformer:

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \mu_k \Delta \mathbf{w}^k.$$

Stop if $|\text{Re}(\nabla f_p(\mathbf{w}^k)^H \Delta \mathbf{w}^k)| < \epsilon$.

end for

direction as

$$\Delta \mathbf{w} = -\mathbf{H}_{\mathbf{w}^* \mathbf{w}}^{-1} \left(\nabla f_p(\mathbf{w}) - \frac{\mathbf{a}^H \mathbf{H}_{\mathbf{w}^* \mathbf{w}}^{-1} \nabla f_p(\mathbf{w})}{\mathbf{a}^H \mathbf{H}_{\mathbf{w}^* \mathbf{w}}^{-1} \mathbf{a}} \mathbf{a} \right). \quad (2.39)$$

Once the Newton direction is determined, the beamformer is updated using (2.31). Again, a line search procedure in (2.33) can be applied to obtain the optimal step size. We refer to this algorithm as the partial Newton's method because it utilizes only the partial Hessian matrix $\mathbf{H}_{\mathbf{w}^* \mathbf{w}}$. Since the size of the linear system of the partial Newton's method is only half of the full Newton's method, the cost for computing the Newton's direction for the partial Newton's method is approximately 1/8 of that of the full Newton's method. The computational simplicity results from ignoring the off-diagonal Hessian matrices $\mathbf{H}_{\mathbf{w}^* \mathbf{w}^*}$ and $\mathbf{H}_{\mathbf{w} \mathbf{w}}$. These two partial Hessian matrices are not null and contain useful information for $p \neq 2$. Hence the performance of the partial Newton's method is inferior to that of the full Newton's method — its convergence rate is slower than that of the full Newton's method.

An interesting relationship between the IR-MVDR and the partial Newton's method for ℓ_p -norm minimization is described in the following proposition.

Proposition 2.2: The IR-MVDR algorithm is a special case of the partial Newton's method using a fixed step size of $p/2$.

Proof: The update formula of the partial Newton's method with step size of $p/2$ is

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \frac{p}{2} \Delta \mathbf{w}^k. \quad (2.40)$$

Substituting (2.37) into (2.40) leads to

$$\begin{aligned}
\mathbf{w}^{k+1} &= \mathbf{w}^k - \frac{p}{2} \mathbf{H}_{\mathbf{w}^k}^{-1} (\nabla f_p(\mathbf{w}^k) + \lambda \mathbf{a}) \\
&= \mathbf{w}^k - \frac{p}{2} \cdot \frac{4}{p^2} (\mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H)^{-1} \\
&\quad \times \left(\frac{p}{2} \mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H \mathbf{w}^k + \lambda \mathbf{a} \right) \\
&= \frac{-2\lambda}{p} (\mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H)^{-1} \mathbf{a}.
\end{aligned} \tag{2.41}$$

Since \mathbf{w}^{k+1} is feasible, i.e., it satisfies the constraint $\mathbf{a}^H \mathbf{w}^{k+1} = 1$, we have

$$\frac{-2\lambda}{p} = \frac{1}{\mathbf{a}^H (\mathbf{X} \mathbf{D}(\mathbf{w}^k) \mathbf{X}^H)^{-1} \mathbf{a}}. \tag{2.42}$$

Plugging (2.42) into (2.41) yields (2.11), i.e., the update formula of the IR-MVDR algorithm. \square

Clearly, the fixed step size of $p/2$ is not optimal. Therefore the IR-MVDR is inferior to the partial Newton's method in terms of convergence rate.

The convergence rates of the IR-MVDR, two Newton's methods, and the gradient descent method [53] with optimal step sizes for different values of p are compared here. Note that [53] does not discuss how to select a step size for the gradient descent method although it is an important issue. Here we use the optimal step size for this gradient method, which yields the best performance. Six values of p , namely, $p \in \{1.2, 1.5, 3.2, 3.4, 4, 8\}$, are tried. In this numerical example, we use the experimental settings in example 1 of Section 2.4. We are primarily interested in the behavior, as a function of the number of iterations, of the relative error $|f_p(\mathbf{w}^k) - f_p(\bar{\mathbf{w}})|/f_p(\bar{\mathbf{w}})$, where $\bar{\mathbf{w}}$ and $f_p(\bar{\mathbf{w}})$ are the minimizer and the global minimum of (2.3), respectively. This global minimum can be calculated exactly (in practice, up to the computer

round-off precision) with a finite number of steps using the full Newton's method or any optimization software package in advance. For fair comparison, all the methods use the same initial value of $\mathbf{w}^0 = \mathbf{a}/\|\mathbf{a}\|^2$. Fig. 2.1 shows the convergence rates of the three methods. We can see that the IR-MVDR algorithm does not converge for $p \geq 3.4$ while the two Newton's methods converge in all cases. The gradient method [53] converges very slowly. When the IR-MVDR algorithm converges, it has a linear convergence rate. The partial Newton's method also has a linear convergence rate but it converges faster than the IR-MVDR. The full Newton's method has a quadratic convergence rate and converges very fast. It only needs several iterations for convergence. Ignoring the two off-diagonal Hessian matrices makes the partial Newton's method lose the property of quadratic convergence.

Remark 2.2: We can use the SMI beamformer of (2.34) as the initial value to speed up the convergence of the IR-MVDR and the two Newton's methods because the SMI beamformer may be closer to the true solution than the data-independent beamformer of (2.12). Fig. 2.2 compares the convergence rates using these two different initializations. It is observed that the SMI beamformer is a better initial value that can accelerate the convergence rate.

2.2.5 MDDR Beamforming via ℓ_∞ -Norm Minimization

The MDDR beamformer is applicable to the case of $p \rightarrow \infty$, where the ℓ_p -norm becomes

$$\|\mathbf{y}\|_\infty = \max_{1 \leq n \leq N} |y(n)|. \quad (2.43)$$

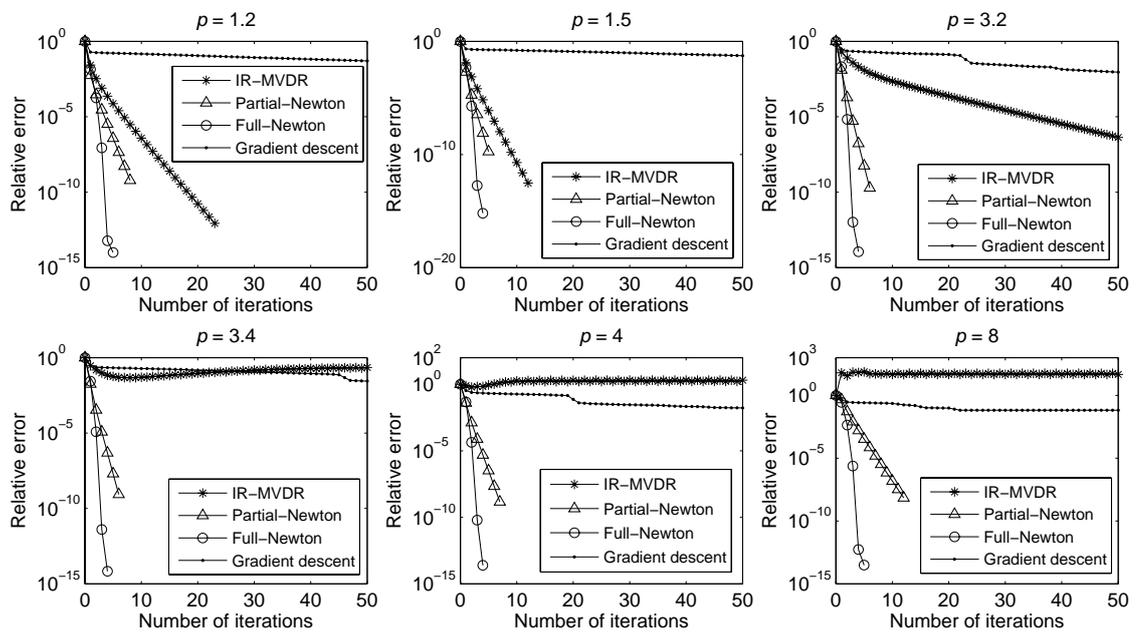


Figure 2.1: Convergence rate versus number of iterations of the IR-MVDR, two Newton's methods, and gradient descent method [53] with optimal step sizes for $\rho = 1.2, 1.5, 3.2, 3.4, 4,$ and 8 .

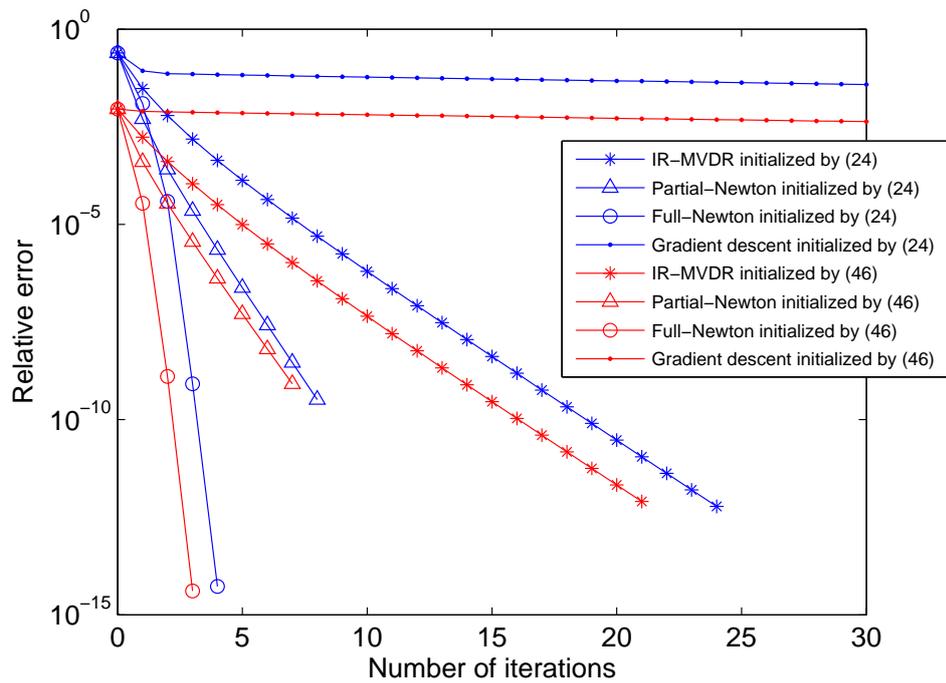


Figure 2.2: Convergence rate versus number of iterations for $p = 1.2$ using different initial values. The blue curves are results using data-independent beamformer as initialization while the red curves are obtained by use of SMI beamformer as initial value.

Accordingly, the ℓ_∞ -norm MMDR beamforming corresponds to

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_\infty \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \quad (2.44)$$

Since the function of ℓ_∞ -norm is non-differentiable, the Newton's method cannot be applied. However, we show that (2.44) can be converted into a second-order cone programming (SOCP). First, the complex-valued variables are split into real-valued ones, that is, $\mathbf{w} = \mathbf{w}_R + j\mathbf{w}_I$, $\mathbf{a} = \mathbf{a}_R + j\mathbf{a}_I$, $\mathbf{y} = \mathbf{y}_R + j\mathbf{y}_I$, and $\mathbf{X} = \mathbf{X}_R + j\mathbf{X}_I$ with $\mathbf{w}_R, \mathbf{w}_I, \mathbf{a}_R, \mathbf{a}_I \in \mathbb{R}^M$, $\mathbf{y}_R, \mathbf{y}_I \in \mathbb{R}^N$, and $\mathbf{X}_R, \mathbf{X}_I \in \mathbb{R}^{M \times N}$. Then we have $|y(n)| = \sqrt{y_R^2(n) + y_I^2(n)}$. The problem of (2.44) is reformulated as the following SOCP:

$$\begin{aligned} \min_{\mathbf{w}_R, \mathbf{w}_I, \mathbf{y}_R, \mathbf{y}_I, u} \quad & u \\ \text{s.t.} \quad & \sqrt{y_R^2(n) + y_I^2(n)} \leq u, \quad n = 1, \dots, N \\ & \begin{bmatrix} \mathbf{X}_R & \mathbf{X}_I \\ \mathbf{X}_I & -\mathbf{X}_R \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} \\ & \begin{bmatrix} \mathbf{a}_R & -\mathbf{a}_I \\ \mathbf{a}_I & \mathbf{a}_R \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned} \quad (2.45)$$

where $u \in \mathbb{R}^+$ is an auxiliary variable participating in the optimization. The Newton's method cannot handle this SOCP problem because it contains an inequality constraint. A more sophisticated interior point method [55] is needed to solve it. This results in a computational complexity of $\mathcal{O}((N+M)^{3.5})$ in each iteration, which is higher than those of the IR-MVDR and the Newton's methods.

In order to avoid solving the SOCP, one can use a large enough value for p to

approximate the ℓ_∞ -norm. From the simulation results, we find that $p = 20$ provides performance similar to that of the ℓ_∞ -norm. It is not advisable to adopt too large values for p since it may result in overflow during calculating.

2.3 Extension to Multiple Linear Constraints

It is known that the MVDR beamformer suffers significant performance degradation due to the uncertainty or mismatch in the steering vector [8, 18, 24]. One cause for steering vector mismatch is the AOA estimation error. When the steering vector of the SOI is imprecise, the SOI will be mistaken as interference and attenuated by the MVDR beamformer [26]. A remedy to address the AOA mismatch is to impose multiple linear constraints for a small spread of angles around the nominal AOA [14]. That is,

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{C}^H \mathbf{w} = \mathbf{g} \end{aligned} \tag{2.46}$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K] \in \mathbb{C}^{M \times K}$ contains K steering vectors or the derivatives of the steering vectors and $\mathbf{g} = [g_1, \dots, g_K]^T$ is usually taken as the vector with all elements being unity. The closed-form solution of the optimization problem in (2.46) is called LCMV beamformer and is given by [14]

$$\mathbf{w}_{\text{LCMV}} = \mathbf{R}^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{g}. \tag{2.47}$$

When the number of linear constraints is $K = 1$, the LCMV beamformer reduces to the MVDR beamformer. Similar to the MDDR beamformer, we can generalize the

LCMV method to the following LCMD beamforming

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_p^p \\ \text{s.t.} \quad & \mathbf{C}^H \mathbf{w} = \mathbf{g}. \end{aligned} \quad (2.48)$$

Equation (2.48) is an extension of (2.3) from a single linear constraint to multiple constraints. Meanwhile, the three algorithms for solving (2.3) can be extended to multiple constraints in (2.48).

Replacing the covariance matrix \mathbf{R} using the reweighted version $\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H$, the update rule of the Iteratively Reweighted LCMD (IR-LCMD) algorithm is given by

$$\mathbf{w}^{k+1} = (\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1} \mathbf{C} \left(\mathbf{C}^H (\mathbf{X}\mathbf{D}(\mathbf{w}^k)\mathbf{X}^H)^{-1} \mathbf{C} \right)^{-1} \mathbf{g} \quad (2.49)$$

which is an extension of the IR-MVDR algorithm of (2.11).

The full Newton's method can also be applied to solve the LCMD problem of (2.48). The Newton direction $\Delta\mathbf{w}$ is calculated by

$$\begin{aligned} \min_{\Delta\mathbf{w}} \quad & q_{\mathbf{w}}(\Delta\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{C}^H \Delta\mathbf{w} = 0. \end{aligned} \quad (2.50)$$

The Lagrangian function of (2.50) is

$$\mathcal{L}_c(\Delta\mathbf{w}, \boldsymbol{\lambda}) = q_{\mathbf{w}}(\Delta\mathbf{w}) + \boldsymbol{\lambda}^T \mathbf{C}^H \Delta\mathbf{w} \quad (2.51)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_K]^T$ is the Lagrangian multiplier vector corresponding to the K

linear constraints. Applying the optimal condition to (2.51) leads to

$$\begin{aligned}\frac{\mathcal{L}_c(\Delta\mathbf{w}, \boldsymbol{\lambda})}{\partial\Delta\mathbf{w}^*} &= \nabla f_p(\mathbf{w}) + \mathbf{H}_{\mathbf{w}^*\mathbf{w}}\Delta\mathbf{w} + \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*}\Delta\mathbf{w}^* + \mathbf{C}\boldsymbol{\lambda} = \mathbf{0}, \\ \frac{\mathcal{L}_c(\Delta\mathbf{w}, \boldsymbol{\lambda})}{\partial\Delta\mathbf{w}} &= \nabla f_p(\mathbf{w})^* + \mathbf{H}_{\mathbf{w}\mathbf{w}^*}\Delta\mathbf{w}^* + \mathbf{H}_{\mathbf{w}\mathbf{w}}\Delta\mathbf{w} + \mathbf{C}^*\boldsymbol{\lambda} = \mathbf{0}, \\ \frac{\partial\mathcal{L}_c(\Delta\mathbf{w}, \boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} &= \mathbf{C}^H\Delta\mathbf{w} = \mathbf{0}\end{aligned}\quad (2.52)$$

which can be rewritten as

$$\begin{bmatrix} \mathbf{H}_{\mathbf{w}^*\mathbf{w}} & \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} & \mathbf{C} \\ \mathbf{H}_{\mathbf{w}\mathbf{w}} & \mathbf{H}_{\mathbf{w}\mathbf{w}^*} & \mathbf{C}^* \\ \mathbf{C}^H & \mathbf{0}_{K\times M} & \mathbf{0}_{K\times K} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{w} \\ \Delta\mathbf{w}^* \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla f_p(\mathbf{w}) \\ \nabla f_p(\mathbf{w})^* \\ \mathbf{0} \end{bmatrix}. \quad (2.53)$$

The Newton direction $\Delta\mathbf{w}$ is obtained by solving the linear system of (2.53), which requires a complexity of $\mathcal{O}((2M + K)^3)$.

Again, assigning the two Hessian matrices $\mathbf{H}_{\mathbf{w}\mathbf{w}} = \mathbf{H}_{\mathbf{w}^*\mathbf{w}^*} = \mathbf{0}$ gives the update direction of the partial Newton's method

$$\begin{aligned}\mathbf{H}_{\mathbf{w}^*\mathbf{w}}\Delta\mathbf{w} + \mathbf{C}\boldsymbol{\lambda} &= -\nabla f_p(\mathbf{w}) \\ \mathbf{C}^H\Delta\mathbf{w} &= \mathbf{0}.\end{aligned}\quad (2.54)$$

Solving (2.54) gives

$$\boldsymbol{\lambda} = -(\mathbf{C}^H\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1}\nabla f_p(\mathbf{w}) \quad (2.55)$$

and

$$\Delta\mathbf{w} = -\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1}\left(\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{H}_{\mathbf{w}^*\mathbf{w}}^{-1}\right)\nabla f_p(\mathbf{w}). \quad (2.56)$$

After obtaining the Newton direction, the remaining steps of the full and partial Newton's methods are the same as the single constraint algorithms in Section 2.2.4.

The Newton's method exploits the second-order derivatives of the objective function, which are contained in the Hessian matrix. The gradient descent method in [53] just utilizes the first-order derivatives. Assuming that the Hessian matrix $\mathbf{H}_{\mathbf{w}^*\mathbf{w}} = \mathbf{I}$, (2.56) is simplified to

$$\Delta\mathbf{w} = -\left(\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\right)\nabla f_p(\mathbf{w}) \quad (2.57)$$

which is the update direction of the gradient descent method for constrained optimization problems. Note that $\left(\mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\right) = \mathbf{P}_{\mathbf{C}}^\perp$ is the projector onto the orthogonal complementary space of $\text{range}(\mathbf{C})$. Therefore the update direction $\Delta\mathbf{w} = -\mathbf{P}_{\mathbf{C}}^\perp\nabla f_p(\mathbf{w})$ represents the *projected gradient* or *gradient projection* [64]. In [53], the gradient projection method is adopted to compute the LCMD beamformer. However, it does not mention how to select a step size. The step size is determined empirically. For the proposed full and partial Newton's methods, the optimal step size is solved using line search method for a given Newton direction. Furthermore, the convergence rate of the gradient projection method is slower than that of the Newton's method and even much slower than that of the IR-LCMV algorithm.

2.4 Simulation Results

To facilitate a fair comparison, experimental parameters used in [24] and [26] are adopted in our computer simulations as well. A uniform linear array (ULA) of M

omnidirectional sensors spaced half a wavelength apart is considered. The steering vector is computed using (1.2). Three zero-mean signals, namely, the SOI $s(n)$ and two uncorrelated interferences $s_1(n)$ and $s_2(n)$, impinge on the array. The AOA of the SOI is $\theta = 43^\circ$ and the AOAs of the two interferences are $\theta_1 = 30^\circ$ and $\theta_2 = 75^\circ$. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_v^2}. \quad (2.58)$$

The two interferences are stronger than the SOI with variances being $\sigma_1^2 = 4\sigma_s^2$ and $\sigma_2^2 = 9\sigma_s^2$. That is, they are 6 dB and 9.5 dB above the SOI, respectively.

2.4.1 Results with Perfect Steering Vector

We first present the simulation results with a perfect steering vector. The performance of the MDDR with a variety of p , MVDR (i.e., $p = 2$), and subspace beamformers, as well as the optimal SINR bound, are compared. According to (3.53), the upper bound on the SINR is the maximum generalized eigenvalue of the matrix pair $(\sigma_s^2 \mathbf{a}\mathbf{a}^H, \mathbf{R}_{i+n})$. Note that the subspace beamformer requires the dimension of the signal-plus-interference subspace. The minimum description length (MDL) [48] is applied to estimate this quantity. Monte Carlo trials are conducted to evaluate the output SINR performance of the beamforming algorithms. When plotting the SINR curves, 200 Monte Carlo trials are performed to calculate the average output SINR.

Example 1: Sub-Gaussian signals

In the first example, the SOI and two interferences are QPSK signals, which are frequently encountered in communications and are sub-Gaussian distributed. The additive noise is a white Gaussian process. Fig. 2.3 shows the output SINR versus

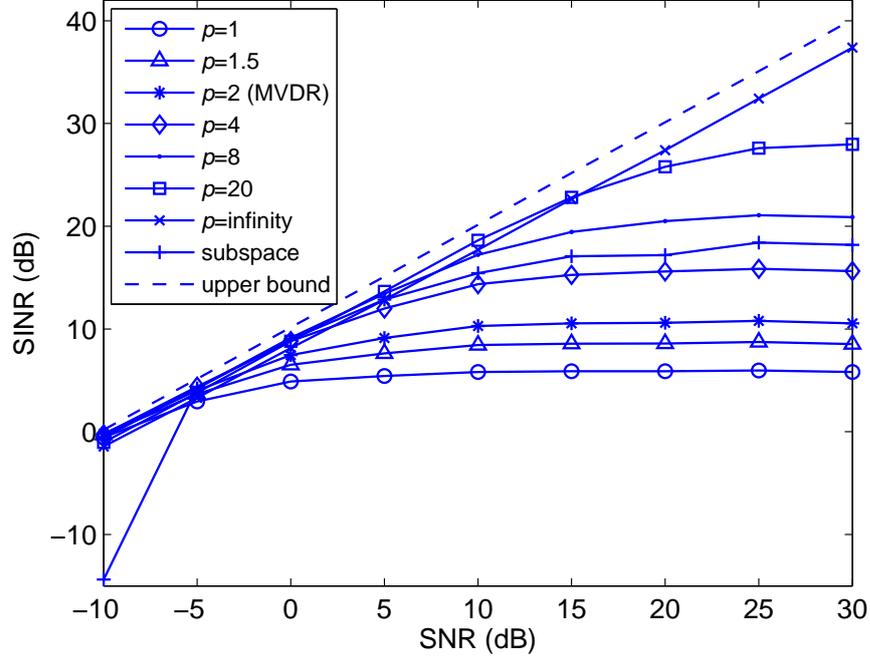


Figure 2.3: Output SINR versus SNR for QPSK signals and additive Gaussian noise.

SNR when the number of sensors $M = 10$ and number of snapshots $N = 100$. Fig. 2.4 displays the output SINR versus N at $M = 10$ and SNR = 10 dB. Fig. 2.5 shows the output SINR versus M at $N = 100$ and SNR = 10 dB.

From Figs. 2.3 to 2.5, it can be seen that the MDDR beamforming of $p > 2$ leads to an improved performance compared with the MVDR beamformer ($p = 2$) for QPSK signals. It can be seen from Fig. 2.3, when SNR ≥ 10 dB, the output SINR for $p = 20$ is 10–20 dB higher than that of the MVDR beamformer. In Figs. 2.4 and 2.5, this performance gain is still about 10 dB when the number of snapshots is larger than 50 or the number of sensors is more than 8. In particular, only the ℓ_∞ -MDDR beamformer ($p = \infty$) approaches the upper bound of the SINR as the SNR

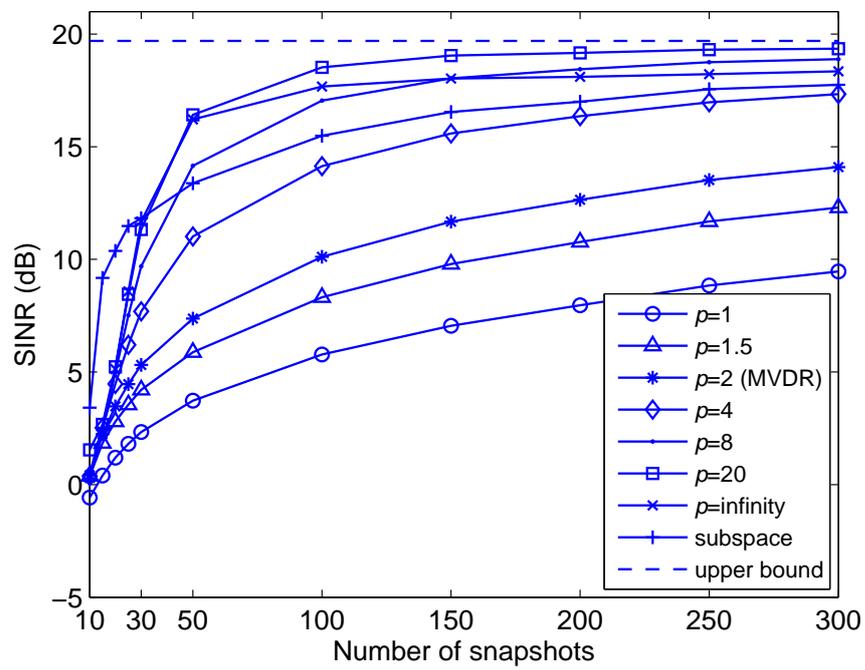


Figure 2.4: Output SINR versus number of snapshots for QPSK signals and additive Gaussian noise.

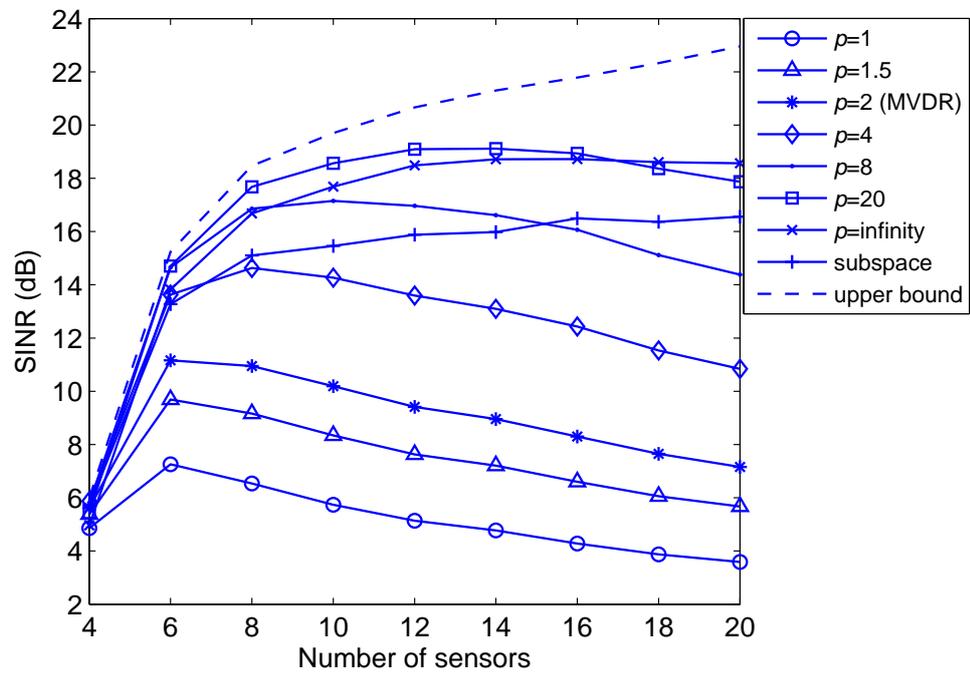


Figure 2.5: Output SINR versus number of sensors for QPSK signals and additive Gaussian noise.

increases. The results demonstrate that the MDDR beamformer with larger value of p has better performance. The MDDR beamformer with $p < 2$ is not recommended for sub-Gaussian signals because its performance is worse than that of the MVDR beamformer. The subspace beamformer is also superior to the MVDR beamformer and it has similar performance as the MDDR beamformer with $p = 4$. However, its performance degrades severely at low SNR.

Example 2: Super-Gaussian signals

In the second example, the beamforming algorithms are tested using super-Gaussian signals. The SOI, two interferences and noise are modeled as random processes satisfying a generalized Gaussian distribution (GGD) [58]. The PDF of the circular zero-mean GGD with variance σ_s^2 is

$$p_s(s) = \frac{\beta\Gamma(4/\beta)}{2\pi\sigma_s^2\Gamma^2(2/\beta)} \exp\left(-\frac{|s|^\beta}{c\sigma_s^\beta}\right) \quad (2.59)$$

where $\beta > 0$ is the shape parameter, $\Gamma(\cdot)$ is the Gamma function, and $c = (\Gamma(2/\beta)/\Gamma(4/\beta))^{\beta/2}$ [58]. When $\beta = 2$, the GGD reduces to the circular Gaussian distribution. Note that $\beta > 2$ models sub-Gaussian signals while $\beta < 2$ models super-Gaussian ones. Especially, $\beta = 1$ corresponds to the Laplacian distribution [58], which is widely used to model speech signal [42]. The smaller the value of β , the more impulsive the signal is. We take $\beta = 0.4$ in this example.

Fig. 2.6 displays the output SINR versus SNR for $M = 10$ and $N = 100$. Fig. 2.7 illustrates the output SINR versus N for $M = 10$ and SNR = 10 dB. Fig. 2.8 shows the output SINR versus M for $N = 200$ and SNR = 0 dB.

The results of Figs. 2.6, 2.7, and 2.8 are opposite to those given by Figs. 2.3, 2.4, and 2.5, respectively. This is not surprising because the signals used in this example

are super-Gaussian distributed, which are different from the sub-Gaussian signals used in the first example. These results illustrate that the MDDR beamformer with $p < 2$ will yield a SINR gain for super-Gaussian signals. The performance improvement with $p = 1$ is about 5–7 dB when $\text{SNR} \geq 10$ dB, which is optimal in this example. In the presence of super-Gaussian signals, the MDDR beamformer with smaller p has better performance. Contrary to the case of sub-Gaussian signal, $p > 2$ is not recommended for super-Gaussian signals. The subspace beamformer has a performance comparable to that of the MDDR beamformer with $p = 1$.

As SNR increases, even the optimal setting $p = 1$ cannot approach the upper bound of SINR. It still has a large gap from the bound. Hence we infer that $p < 1$ may achieve better performance for GGD signals with $\beta = 0.4$. However, as mentioned above, we do not consider the choice of $0 < p < 1$ since the resulting optimization problem is difficult to solve [59].

Example 3: Gaussian signals

In the third example, all the signals and noise are Gaussian. Fig. 2.9 illustrates the output SINR versus SNR for $M = 10$ and $N = 100$. We can see that $p = 2$ is the optimal value for Gaussian signals. It is demonstrated that all the MDDR methods have similar performance if p is not far away from 2. The subspace method has the best performance for Gaussian signal model if the SNR is not very low. However, the performance of the subspace beamformer degrades at smaller SNRs.

Example 4: Impact of interference number

In the fourth example, we investigate the impact of the interference number on beamforming performance. The SOI and the interferences are QPSK modulation and noise values are Gaussian. The snapshot number is $N = 100$. The AOA of the SOI is

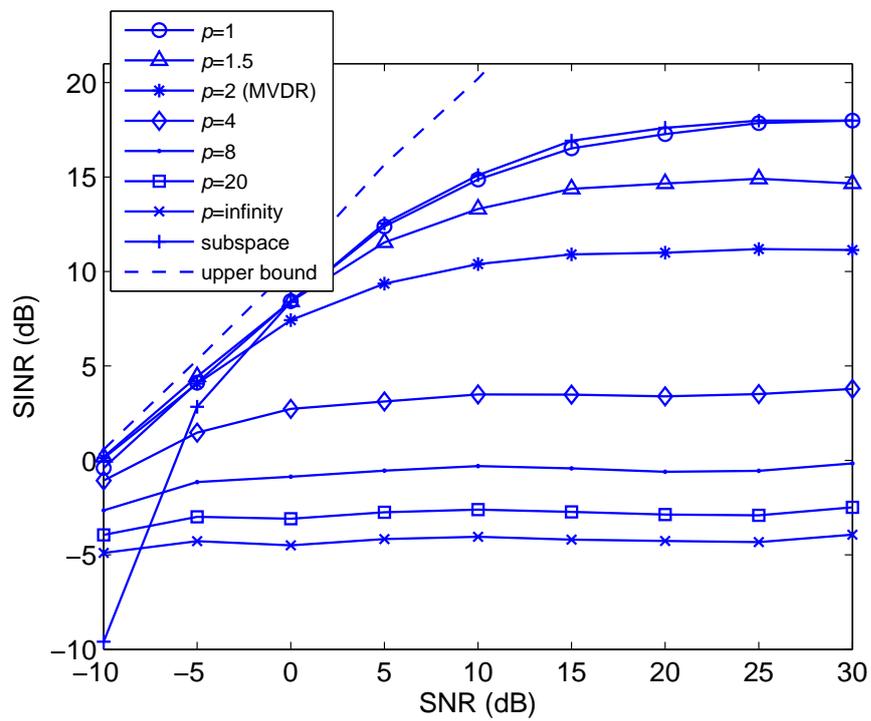


Figure 2.6: Output SINR versus SNR for GGD signals and noise with $\beta = 0.4$.

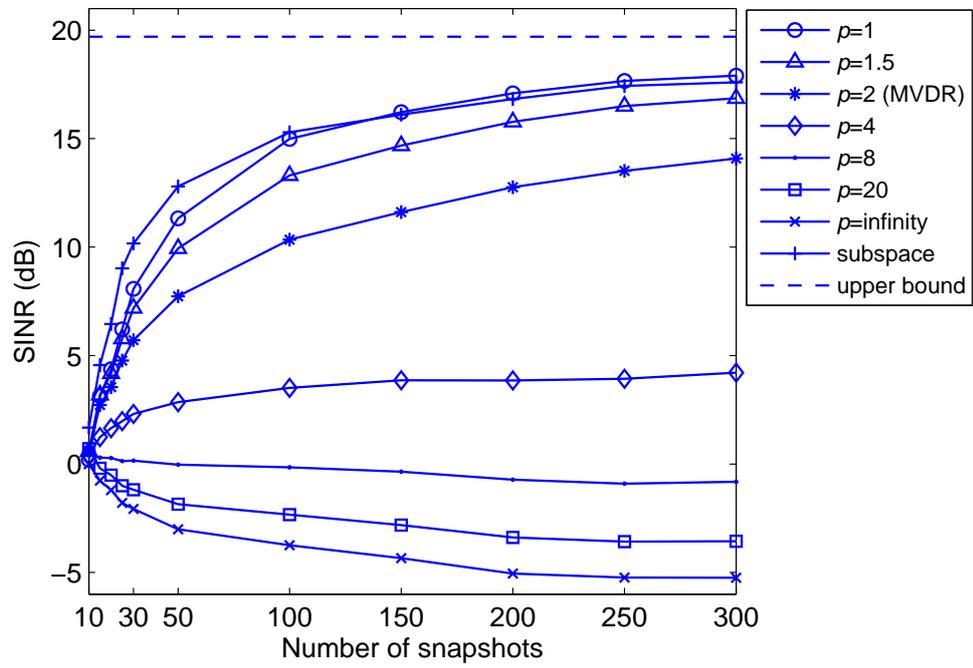


Figure 2.7: Output SINR versus number of snapshots for GGD signals and noise with $\beta = 0.4$.

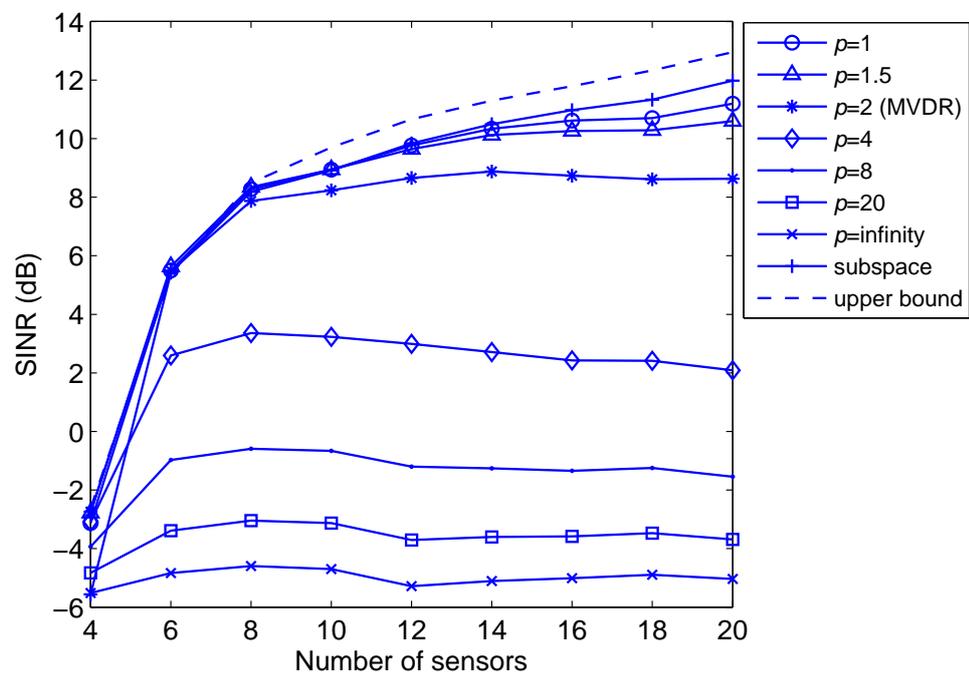


Figure 2.8: Output SINR versus number of sensors for GGD signals and noise with $\beta = 0.4$.

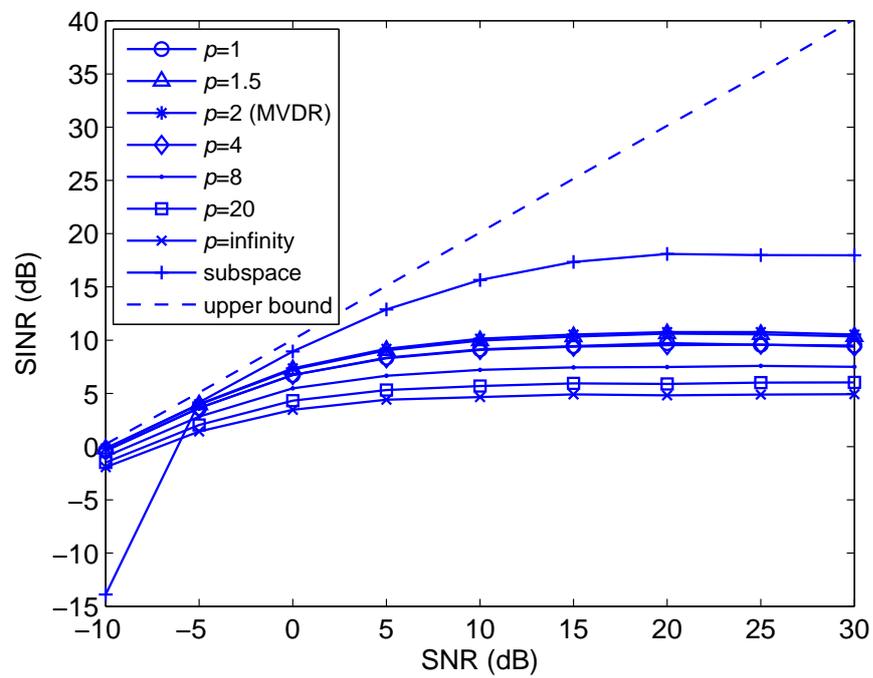


Figure 2.9: Output SINR versus SNR for Gaussian signals and noise.

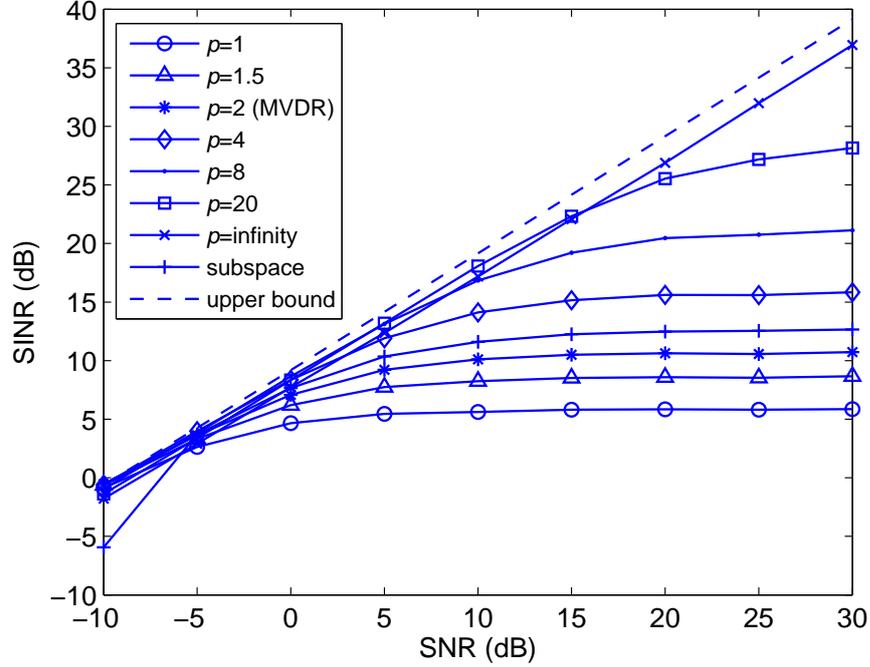


Figure 2.10: Output SINR versus SNR for QPSK signals and Gaussian noise with six interferers.

fixed to 43° while that of the i th ($1 \leq i \leq I$) interference is $\theta_i = -30^\circ + (i-1)10^\circ$. The power of all interferers is the same and 10 dB higher than the SOI. We first consider $I = 6$ interferers. Fig. 2.10 plots the output SINR versus SNR. By comparing Fig. 2.3, which has only two interferers, with Fig. 2.10, we observe that the performance of the subspace beamformer substantially degrades for larger signal number values. Fig. 2.11 shows the output SINR versus interference number at SNR = 20 dB. It can be seen that the subspace beamformer loses its efficiency as the interference number increases while the MVDR and MDDR beamformers maintain their performance.

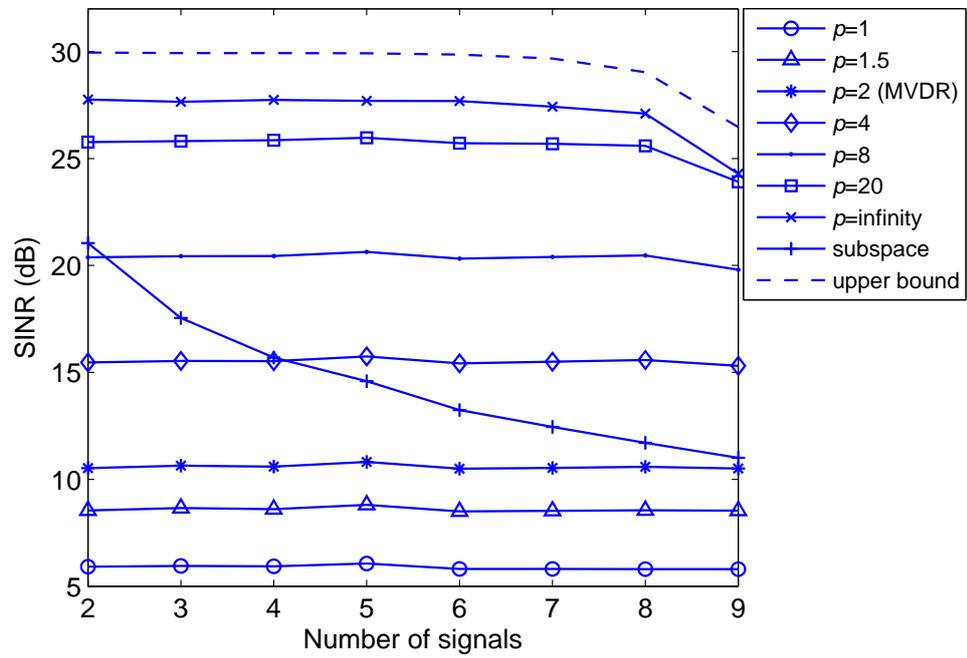


Figure 2.11: Output SINR versus interference number for QPSK signals and Gaussian noise.

2.4.2 Results with AOA Mismatch

In this subsection, we investigate the performances of the LCMD, LCMV, and MVDR beamformers in the presence of steering vector mismatch. The steering vector mismatch is first simulated using imprecise AOA. We focus on the sub-Gaussian signal case where the QPSK signal is used. The experimental settings are the same as that of Example 1 in Section 2.4.1. The true AOA of the SOI is 43° but the assumed AOA is 45° . For LCMV and LCMD beamformers, two linear constraints which force the responses of the signals from 42° and 48° to be unity.

Figs. 2.12 and 2.13 show the output SINR versus SNR and number of snapshots, respectively. It can be seen that the performance of the MVDR beamformer significantly degrades due to the 2° AOA error. It is not surprising that the output SINR of the MVDR beamformer decreases as the SNR increases because the signal cancellation phenomenon caused by AOA mismatch is more severe at high SNRs. The LCMV beamformer with two linear constraints enhances the robustness against AOA mismatch. The LCMD beamformer with $p > 2$, especially for relatively large values of p , significantly improves the performance the LCMV for sub-Gaussian QPSK signals. For example, the SINR of the proposed LCMD beamformer with $p = 20$ is about 10 dB higher than that of the LCMV beamformer when the number of snapshots is greater than 50.

We then investigate the robustness against the look direction of the SOI. The AOA of the SOI, i.e., θ , is varied from -20° to 20° while the AOAs of the two interferers are fixed to 30° and 75° , respectively. It is assumed that there is a 2° error in the AOA estimate. That is, the AOA estimate is $\theta + 2^\circ$. Hence we impose two linear constraints as $\theta - 1^\circ$ and $\theta + 3^\circ$, which are centered around the AOA estimate $\theta + 2^\circ$,

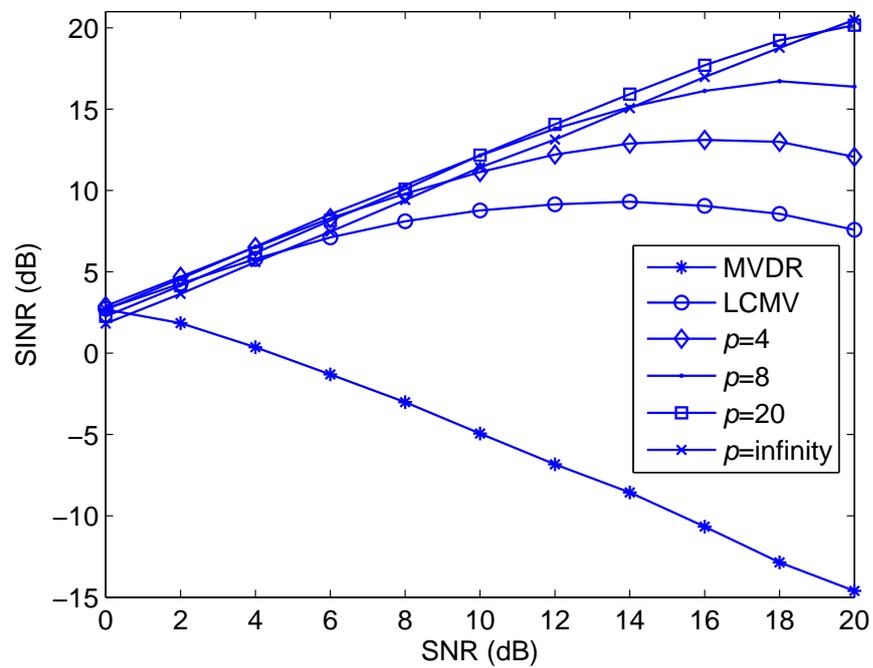


Figure 2.12: Output SINR versus SNR for QPSK signals in the presence of 2° AOA mismatch.

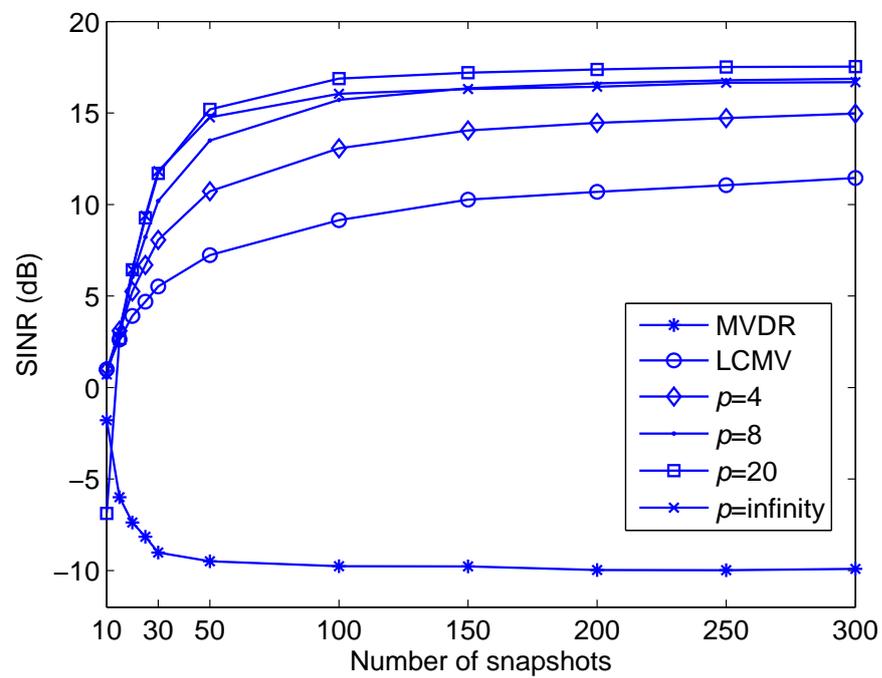


Figure 2.13: Output SINR versus number of snapshots for QPSK signals in the presence of 2° AOA mismatch.

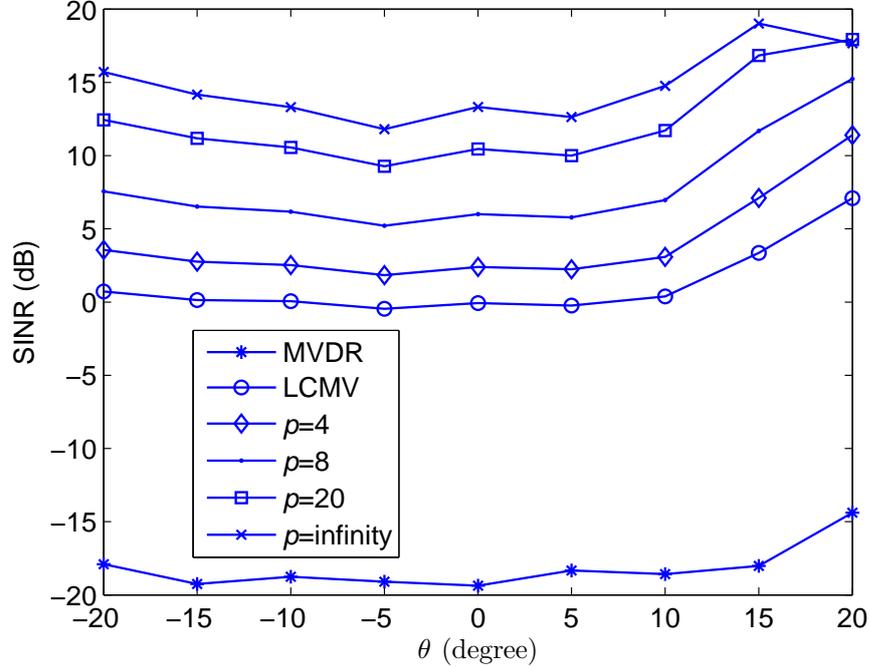


Figure 2.14: Output SINR versus AOA of SOI for 2° AOA mismatch.

for LCMV and LCMD beamformers. Fig. 2.14 shows the output SINR versus θ , from which we see that the LCMD beamformer is not sensitive to the look direction.

2.4.3 Results with Sensor Position Mismatch

In addition to AOA estimation error, the steering vector mismatch may be caused by a variety of reasons, such as uncertainties in array response and sensor geometry position. Similar to LCMV, the LCMD beamformer was originally designed for AOA mismatch. However, it can be adapted to handle other types of mismatch. In this simulation example, the error in sensor geometry position is considered. From (1.2), it is observed that the steering vector is a function of the inter-sensor spacing d

and hence denoted as $\mathbf{a}(d)$. In practice, the inter-sensor spacing with error can be expressed as $d = d_0 + \Delta d$, where d_0 is the nominal value and Δd is the error. We investigate the effect of spacing error on beamforming performance. The nominal spacing is equal to half the wavelength, i.e., $d_0 = 0.5\zeta$. The maximal relative error $\Delta d/d_0$ is $\pm 10\%$, which corresponds $0.9d_0 \leq d \leq 1.1d_0$. Three linear constraints, namely, $\{\mathbf{c}_k^H \mathbf{w} = 1\}_{k=1}^3$, are used with the LCMV and LCMD beamformers. We set $\mathbf{c}_1 = \mathbf{a}(0.9d_0)$, $\mathbf{c}_2 = \mathbf{a}(d_0)$, and $\mathbf{c}_3 = \mathbf{a}(1.1d_0)$ in (2.48). The SNR is 20 dB and snapshot number $N = 100$. The other experimental settings are the same as those in Example 1 in Section 2.4.1.

Figs. 2.15 shows the output SINR versus the relative error that varies from -10% to 10% . The performance of the MVDR beamformer significantly degrades as the spacing error increases. The LCMV and LCMD beamformers with three linear constraints improve the robustness against sensor position. Again, the LCMD beamformer with $p > 2$ yields better performance than the LCMV.

2.5 Conclusion

By recognizing the fact that the minimum variance criterion is not statistically optimal for non-Gaussian signals, this chapter investigated the minimum dispersion beamforming with either single or multiple linear constraints. The use of a single linear constraint resulted in the MDDR beamformer while the use of multiple constraints led to the LCMD beamformer. The LCMD beamformer is robust against steering vector mismatches. The MDDR and LCMD beamformers outperform their respective standard counterparts based on minimum variance, namely, the MVDR and LCMV beamformers. Three computationally efficient algorithms, i.e., the IR-MVDR, full

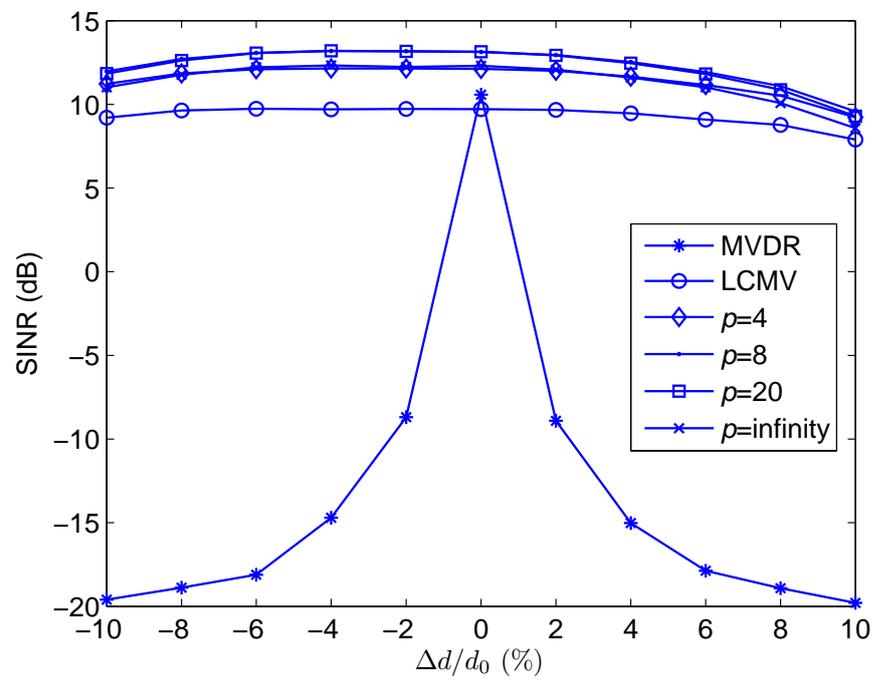


Figure 2.15: Output SINR versus relative error of inter-sensor spacing.

Newton's and partial Newton's methods, were developed to efficiently solve the resultant ℓ_p -minimization problem with linear equality constraints. It was shown that the IR-MVDR is a special case of the partial Newton's method. Simulation results demonstrated the superior performance of the minimum dispersion beamformers. An important future work is combining the minimum dispersion criterion with the nonlinear constraints, such as those in [18, 24]. These nonlinear constraints force the magnitude responses of the steering vectors in an uncertainty set to exceed unity. This will generate new beamforming techniques that are more robust against steering vector mismatch as well as enhance the SINR performance.

Chapter 3

Gradient Projection for Quadratically Constrained Minimum Dispersion Beamforming

A quadratically constrained minimum dispersion (QCMD) beamformer that is robust against model uncertainties is devised for non-Gaussian signals. Different from the minimum variance based beamformers, the QCMD beamformer minimizes the ℓ_p -norm ($p \geq 1$) of the output while constraining the magnitude response of any steering vector within a spherical uncertainty set to exceed unity. A gradient projection algorithmic framework is proposed to efficiently solve the resulting convex optimization problem instead of directly applying the standard optimization algorithm which has a high computational complexity. In each iteration, the gradient projection updates the solution along the gradient direction and projects it back to the constraint set. Importantly, a closed-form expression of the projection onto the constraint set is derived, which only needs a low complexity of $\mathcal{O}(M)$ with M being the sensor number.

Therefore, the proposed algorithm is much faster and simpler to implement compared with the standard method. In addition, the robust constant modulus beamformer (RCMB) is also discussed as a special case of the QCMD beamformer. Simulation results demonstrate the efficiency of the gradient projection algorithm and superiority of the QCMD beamformer over several representative robust beamformers, indicating that it can approach the optimal performance bound. Most of the results of this chapter have been reported in our recent journal paper [51].

3.1 Introduction

Beamforming is an important technique used to enhance the desired signal and alleviate the interference and noise in array processing. Its applications can be found in radar, sonar, wireless communications, audio processing, biomedicine and many other areas [5, 7, 8].

The minimum variance distortionless response (MVDR) beamformer [11], which is a classical data dependent beamforming method, minimizes the variance of the array output while fixing the response to the direction of the signal of interest (SOI) to unity. However, the MVDR beamformer is quite sensitive to the steering vector mismatch, which is an unavoidable problem in practical applications [8]. The angle-of-arrival (AOA) estimation error is a common cause for the steering vector mismatch. In [14], the linearly constrained minimum variance (LCMV) beamformer has been proposed by adding additional linear constraints to the MVDR beamformer to broaden the coverage around the nominal AOA. In this way, its robustness to AOA mismatch is enhanced. The drawback of the LCMV beamformer is that the degrees of freedom for interference suppression are reduced as more linear constraints are added [8]. There

are also several methods developed for AOA mismatch [17, 26, 28]. Nonetheless, they cannot handle the case of arbitrary steering vector mismatch.

In order to improve the robustness for general steering vector mismatch, a number of robust beamforming techniques have been proposed [10, 36]. The eigenspace-based beamformer [15] is a powerful technique which is applicable to any type of mismatch. However, it loses efficiency as the signal-to-noise ratio (SNR) decreases or number of interferences increases. Robust beamforming techniques based on worst-case performance optimization are proposed in [18]– [20], where nonlinear constraints are employed instead of the linear ones. The key idea of [18]– [20] is to model the actual steering vector as the sum of the nominal steering vector and an uncertainty term, where the Euclidean norm of the uncertainty is upper-bounded. A spherical uncertainty region is used in [18] and a more general ellipsoidal uncertainty region is considered in robust minimum variance beamformer (RMVB) [24] and robust Capon beamformer (RCB) [20]. The problems of worst-case performance optimization with infinitely many nonconvex constraints are finally converted to a second-order cone programming (SOCP), which can be solved by the standard interior-point methods (IPM). In addition, using the Lagrangian multiplier method, the RMVB and the RCB can be solved with a low computational complexity. Using the idea of worst-case performance optimization, a general-rank model has been suggested in [21] and [36], which is applicable to both rank-one and higher-rank SOI models.

Note that most existing robust beamforming techniques are based on the second-order statistics (variance) of the array output [8, 11, 14, 15, 17, 18, 20, 21, 24, 26–28, 35, 36], which is optimal for Gaussian signals and noise. However, many real world signals and noise are non-Gaussian. For example, the digitally modulated signals

encountered in wireless communication systems are found to be sub-Gaussian [37,41]. In this case, the higher-order statistics can be utilized to improve the performance of the beamformer. In [49], a minimum dispersion (MD) criterion is proposed by minimizing the ℓ_p -norm ($p \geq 1$) of the array output, which implicitly utilizes the higher-order statistics with $p \geq 2$ or lower-order statistics with $p \leq 2$. The MD criterion with multiple linear constraints leads to the linearly constrained minimum dispersion (LCMD) beamformer [49]. Although the LCMD beamformer achieves a performance gain compared with the LCMV beamformer for non-Gaussian signals, it inherits the disadvantages from the LCMV beamformer.

In this chapter, by exploiting the MD criterion combined with magnitude response constraints, we develop a quadratically constrained minimum dispersion (QCMD) beamforming technique for non-Gaussian signals that is robust to arbitrary steering vector mismatch. Instead of resorting to the standard optimization algorithm with a high computational complexity, a gradient projection algorithmic framework is proposed to solve the resultant constrained ℓ_p -norm minimization problem. We briefly summarize the contributions of our work on QCMD beamforming as follows.

- i) We extend the LCMD beamformer [49] with multiple linear constraints to the QCMD beamformer with nonlinear constraints. The QCMD beamformer significantly enhances the robustness against model mismatch compared with the LCMD beamformer.
- ii) The property, including the condition of nonemptiness and closeness, of the convex constraint set induced by the quadratic magnitude constraint is investigated. In addition, we derive a closed-form expression of the projection onto the constraint set. Computation of this projection only requires a low complexity

of $\mathcal{O}(M)$ with M being the number of sensors.

- iii) A gradient projection algorithmic framework is developed to efficiently solve the resulting convex optimization problem instead of directly employing the standard algorithm that has a high computational complexity. There are three variants of this method depending on different line search strategies. All the three projected gradient methods (PGMs) are fast and simple to implement due to the low cost for computing the projection.
- iv) For communication signals, the robust constant modulus beamformer (RCMB) is proposed by exploiting the property of constant modulus or finite alphabet. The RCMB, which minimizes a combination of the ℓ_4 - and ℓ_2 -norm of the output, can be viewed a special case of the QCMD beamformer.

The remainder of this chapter is organized as follows. In Section 3.2, the signal model for beamforming is given and the MVDR, minimum dispersion distortionless response (MDDR), and LCMD beamformers are briefly reviewed. In Section 3.3, we present the fast PGMs for mismatch-robust beamforming. Simulation results are provided in Section 3.5. Finally, conclusions are drawn in Section 3.6.

3.2 MDDR and LCMD Beamformers

Instead of the minimum variance (MV) criterion, the MDDR beamformer [49] solves the following linearly constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \text{E} \{ |\mathbf{w}^H \mathbf{x}(n)|^p \} \\ \text{s.t.} \quad & \mathbf{a}^H \mathbf{w} = 1 \end{aligned} \tag{3.1}$$

where $p \geq 1$. Clearly, the MDDR beamformer reduces to the MVDR beamformer for $p = 2$. The quantity $E\{|y(n)|^p\}$ is called dispersion of $y(n)$ in statistic literature, which is a generalization of variance [49]. Therefore, we call the solution of (3.1) as the minimum dispersion beamformer. Similar to the MVDR beamformer, the MDDR beamformer is not robust to steering vector mismatch. The LCMD beamformer [49], which is an extension of the MDDR beamformer in (3.1) from a single linear constraint to multiple linear constraints, is obtained by solving

$$\begin{aligned} \min_{\mathbf{w}} \quad & E\{|\mathbf{w}^H \mathbf{x}(n)|^p\} \\ \text{s.t.} \quad & \mathbf{C}^H \mathbf{w} = \mathbf{g} \end{aligned} \tag{3.2}$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K] \in \mathbb{C}^{M \times K}$ contains K steering vectors and $\mathbf{g} = [g_1, \dots, g_K]^T$ is usually taken as the vector with all elements being unit. By selecting, as the columns of \mathbf{C} , the steering vectors corresponding to a small spread of angles around the nominal AOA, the LCMD beamformer can handle the AOA mismatch. When the number of linear constraints is $K = 1$, the LCMD beamformer reduces to the MDDR beamformer. Efficient numerical algorithms for solving (3.1) and (3.2) have been developed in [49].

The LCMD beamformer broadens the mainbeam and hence enhances the robustness against the AOA mismatch. However, there are two drawbacks with the LCMD beamformer. On one hand, as more linear constraints are added, the beamformer loses more degrees of freedom for interference suppression although the robustness is enhanced. On the other hand, the LCMD beamformer can only handle AOA mismatch but not random mismatch.

3.3 QCMD Beamformer

In this section, we formulate the QCMD beamformer and derive the fast PGM for the resulting constrained optimization problem.

3.3.1 Formulation of QCMD

Most of the existing robust beamforming techniques are based on minimizing the variance of the output [8]. They belong to the class of second-order statistics methods. The MV criterion is statistically optimal for Gaussian signals and noise because the first- and second-order statistics of a Gaussian distribution contain all necessary statistical information. Nevertheless, many real-world signals encountered in radar, sonar, wireless communications, and navigation are sub-Gaussian distributed [37,41]. For sub-Gaussian signals, the higher-order (larger than 2) statistics contain useful information and can be exploited to improve the beamforming performance [44]. To utilize the higher-order statistics, we suggest to minimize the ℓ_p -norm with $p > 2$ of the output instead of minimizing the variance.

Stacking N samples of the beamformer output $y(n) = \mathbf{w}^H \mathbf{x}(n)$ into a vector $\mathbf{y} = [y(1), \dots, y(N)]^T$, one has the following matrix-vector formulation:

$$\mathbf{y}^* = \mathbf{X}^H \mathbf{w}. \quad (3.3)$$

The minimum dispersion beamformer minimizes the p th power of the ℓ_p -norm of the output, i.e.,

$$\min_{\mathbf{w}} \|\mathbf{X}^H \mathbf{w}\|_p^p \quad (3.4)$$

where the ℓ_p -norm of the complex-valued vector \mathbf{y} is defined as

$$\|\mathbf{y}\|_p = \left(\sum_{n=1}^N |y(n)|^p \right)^{1/p} \quad (3.5)$$

with $|y(n)| = \sqrt{\text{Re}^2(y(n)) + \text{Im}^2(y(n))}$ being the modulus of $y(n)$. The guideline on selecting an appropriate value of p is given in [49]. For Gaussian signals, the optimal p is 2. For sub-Gaussian signals, $p > 2$ will achieve better performance, whereas $p < 2$ is preferred for super-Gaussian signals. The optimal value of p depends on the probability density function (PDF) of the signals. For a large number of modulated signals in radar, sonar, and wireless communications, which are considered in this chapter, $p > 2$ will lead to a performance improvement.

Due to a variety of mismatches, the steering vector \mathbf{a} is not known exactly. The actual steering vector is expressed as

$$\mathbf{c} = \mathbf{a} + \mathbf{e} \quad (3.6)$$

where $\mathbf{e} \in \mathbb{C}^M$ is the steering vector error. It is assumed that the error is in an uncertainty set. The goal of robust beamforming is to ensure that the magnitude response does not attenuate in the uncertainty set. In [18], the uncertainty region is modeled as a sphere, that is,

$$\mathcal{E} = \{\mathbf{e} \mid \|\mathbf{e}\| \leq \varepsilon\} \quad (3.7)$$

where ε is the radius of the sphere. In [24], the uncertainty region is modeled as an ellipsoid. Under the spherical or ellipsoidal model, the robust MV beamformer can be obtained by solving an SOCP problem [18].

The proposed robust beamformer minimizes the ℓ_p -norm of the output while requiring the magnitude of the array response to exceed unity for all $\mathbf{e} \in \mathcal{E}$. This can be written as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_p^p \\ \text{s.t.} \quad & |(\mathbf{a} + \mathbf{e})^H \mathbf{w}| \geq 1, \text{ for all } \mathbf{e} \in \mathcal{E}. \end{aligned} \quad (3.8)$$

The constraint $|(\mathbf{a} + \mathbf{e})^H \mathbf{w}|^2 = \mathbf{w}^H (\mathbf{a} + \mathbf{e})(\mathbf{a} + \mathbf{e})^H \mathbf{w} \geq 1$ is a quadratic constraint of \mathbf{w} . Therefore we refer to the solution of (3.8) as QCMD beamformer. However, directly solving (3.8) is not easy because there are infinitely many constraints and they are nonconvex as well.

Following the strategy of [18], we can seek a tight lower bound of $|(\mathbf{a} + \mathbf{e})^H \mathbf{w}|$ using the triangle inequality and Cauchy-Schwartz inequality

$$\begin{aligned} |(\mathbf{a} + \mathbf{e})^H \mathbf{w}| & \geq |\mathbf{a}^H \mathbf{w}| - |\mathbf{e}^H \mathbf{w}| \\ & \geq |\mathbf{a}^H \mathbf{w}| - \|\mathbf{e}\| \|\mathbf{w}\| \\ & \geq |\mathbf{a}^H \mathbf{w}| - \varepsilon \|\mathbf{w}\|. \end{aligned} \quad (3.9)$$

Replacing $|(\mathbf{a} + \mathbf{e})^H \mathbf{w}|$ in (3.8) with this lower bound, (3.8) can be converted into

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_p^p \\ \text{s.t.} \quad & |\mathbf{a}^H \mathbf{w}| - \varepsilon \|\mathbf{w}\| \geq 1. \end{aligned} \quad (3.10)$$

Note that the objective function of (3.10) is phase-invariant w.r.t. \mathbf{w} . Therefore, we can find \mathbf{w} which satisfies $|\mathbf{a}^H \mathbf{w}| = \text{Re}(\mathbf{a}^H \mathbf{w})$ and $\text{Im}(\mathbf{a}^H \mathbf{w}) = 0$ while the objective

function remains unchanged. That is, (3.10) is equivalent to

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_p^p \\ \text{s.t.} \quad & \text{Re}(\mathbf{a}^H \mathbf{w}) \geq \varepsilon \|\mathbf{w}\| + 1. \end{aligned} \quad (3.11)$$

It is clear that (3.11) is a convex optimization because the objective function is convex and the constraint constitutes a convex set. The solution of (3.11) gives the QCMD beamformer, which reduces to the RMVB of [18] for $p = 2$. The case of $p = 2$ will lead to an SOCP. Although there exist well-established IPM for solving (3.11) with arbitrary value of $p \geq 1$, it will result in a high computational complexity if the standard IPM is directly employed. In the next subsection, we will develop a gradient projection algorithmic framework for solving (3.11), which has a quite low complexity.

3.3.2 Fast PGM

1) *Property of Constraint Set:* The constraint in (3.11) is a convex set, which is denoted by

$$\mathcal{C} = \{\mathbf{w} \mid \text{Re}(\mathbf{a}^H \mathbf{w}) \geq \varepsilon \|\mathbf{w}\| + 1\} . \quad (3.12)$$

By this way, (3.11) is written as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \left(f_p(\mathbf{w}) \triangleq \|\mathbf{X}^H \mathbf{w}\|_p^p \right) \\ \text{s.t.} \quad & \mathbf{w} \in \mathcal{C}. \end{aligned} \quad (3.13)$$

Before discussing the algorithm for solving (3.13), we first need to investigate whether the constrained optimization problem of (3.13) has a solution. In other words, the

first task is to check if (3.13) is feasible. This problem is equivalent to determining whether the convex set \mathcal{C} is nonempty. The following proposition gives the answer to this question.

Proposition 1: If $\varepsilon \geq \|\mathbf{a}\|$, then \mathcal{C} is empty; otherwise \mathcal{C} is nonempty and has infinitely many elements.

Proof: By Cauchy-Schwartz inequality, we have

$$\operatorname{Re}(\mathbf{a}^H \mathbf{w}) \leq |\mathbf{a}^H \mathbf{w}| \leq \|\mathbf{w}\| \|\mathbf{a}\|. \quad (3.14)$$

If \mathcal{C} is nonempty, it requires

$$\|\mathbf{w}\| \|\mathbf{a}\| \geq \varepsilon \|\mathbf{w}\| + 1 \quad (3.15)$$

which yields

$$\varepsilon \leq \|\mathbf{a}\| - \frac{1}{\|\mathbf{w}\|} < \|\mathbf{a}\|. \quad (3.16)$$

If $\varepsilon \geq \|\mathbf{a}\|$, it contradicts with (3.16). This means that there is no such a \mathbf{w} satisfying (3.16) and hence \mathcal{C} is empty. If $\varepsilon < \|\mathbf{a}\|$ and the two inequalities in (3.14) are tight, then all the vectors satisfying

$$\|\mathbf{w}\| \geq \frac{1}{\|\mathbf{a}\| - \varepsilon} \quad (3.17)$$

are in \mathcal{C} . Therefore, \mathcal{C} is nonempty. To be more specific, if $\mathbf{w} = \beta \mathbf{a}$ with $\beta \in \mathbb{C}$ being a scalar, the second inequality in (3.14), i.e., the Cauchy-Schwartz inequality, is tight. In addition, the tightness of the inequality $\operatorname{Re}(\mathbf{a}^H \mathbf{w}) \leq |\mathbf{a}^H \mathbf{w}|$ requires $\mathbf{a}^H \mathbf{w}$ is a positive real-valued number. Therefore, we have $\mathbf{a}^H \mathbf{w} = \beta \|\mathbf{a}\|^2 \in \mathbb{R}^+$. Then the scalar β is a positive real number and hence $\|\mathbf{w}\| = \beta \|\mathbf{a}\|$. Substituting this result

into (3.17) yields

$$\beta \geq \frac{1}{\|\mathbf{a}\|(\|\mathbf{a}\| - \varepsilon)}. \quad (3.18)$$

It is clear that $\mathbf{w} = \beta\mathbf{a}$ with all β satisfying (3.18) are in \mathcal{C} . That is, \mathcal{C} is nonempty and it contains infinitely many elements. \square

In particular, the point

$$\mathbf{w} = \frac{\mathbf{a}}{\|\mathbf{a}\|(\|\mathbf{a}\| - \varepsilon)} \quad (3.19)$$

lies in the boundary of \mathcal{C} . Proposition 1 provides a guideline for selecting the parameter ε . It requires $\varepsilon < \|\mathbf{a}\|$ to make the optimization problem of (3.13) feasible. From now on, we assume that $\varepsilon < \|\mathbf{a}\|$ is satisfied.

The following proposition is a direct conclusion followed from Proposition 1 and the property of projection onto convex sets (POCS) [64].

Proposition 2: The convex set \mathcal{C} is nonempty, closed, and unbounded if $\varepsilon < \|\mathbf{a}\|$. Therefore, there exists a unique projection onto \mathcal{C} for any vector $\mathbf{z} \in \mathbb{C}^M$.

The operator of the projection onto the convex set \mathcal{C} is denoted by $P_{\mathcal{C}}(\cdot)$.

2) *Framework of PGM:* We propose to use the PGM to solve the constrained optimization problem of (3.13). The gradient of the objective $f_p(\mathbf{w})$ w.r.t. the complex vector $\mathbf{w} \in \mathbb{C}^M$ is defined as

$$\nabla f_p(\mathbf{w}) = \frac{\partial f_p(\mathbf{w})}{\partial \mathbf{w}^*} = \left[\frac{\partial f_p}{\partial w_1^*}, \dots, \frac{\partial f_p}{\partial w_M^*} \right]^T \quad (3.20)$$

where

$$\frac{\partial f_p}{\partial w_i^*} = \frac{1}{2} \left(\frac{\partial f_p}{\partial \text{Re}(w_i)} + j \frac{\partial f_p}{\partial \text{Im}(w_i)} \right), \quad i = 1, \dots, M. \quad (3.21)$$

By this definition, we compute the gradient of $f_p(\mathbf{w})$ as

$$\nabla f_p(\mathbf{w}) = \frac{p}{2} \mathbf{X} \mathbf{D}(\mathbf{w}) \mathbf{X}^H \mathbf{w} \quad (3.22)$$

where

$$\mathbf{D}(\mathbf{w}) = \text{diag} \{ |y(1)|^{p-2}, \dots, |y(N)|^{p-2} \}. \quad (3.23)$$

Note that \mathbf{D} depends on the unknown \mathbf{w} because it is related to \mathbf{y} . Therefore, it is a function of \mathbf{w} , which is written as $\mathbf{D}(\mathbf{w})$.

In order to solve (3.13), the PGM generates a sequence $\{\mathbf{w}^k\} \in \mathbb{C}^M$ ($k = 1, 2, \dots$) through the following iterative procedure:

Initialization: Take $\mathbf{w}^0 \in \mathcal{C}$, e.g., initialize \mathbf{w} according to (3.19).

Iterative step: If the convergence condition is satisfied, then stop. Otherwise, let

$$\mathbf{u}^k = P_{\mathcal{C}}(\mathbf{w}^k - \mu_k \nabla f_p(\mathbf{w}^k)) \quad (3.24)$$

$$\Delta \mathbf{w}^k = \mathbf{u}^k - \mathbf{w}^k \quad (3.25)$$

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha_k \Delta \mathbf{w}^k \quad (3.26)$$

where $\mu_k > 0$, $0 \leq \alpha_k \leq 1$ are positive step sizes, and $\Delta \mathbf{w}^k = \mathbf{u}^k - \mathbf{w}^k$ denotes the search direction in the k th iteration.

It is clear that \mathbf{w}^{k+1} belongs to the constrained set \mathcal{C} for all k due to $\alpha_k \in [0, 1]$. That is, the sequence $\{\mathbf{w}^k\}$ is feasible. The step sizes can be determined by line search. There are three different PGMs according to different selections of the step sizes μ_k and α_k , as follows.

- (a) PGM-1: The first step size is fixed as $\mu_k = 1$ and the second one is determined by the exact line search

$$\alpha_k = \arg \min_{\alpha \geq 0} f_p(\mathbf{w}^k + \alpha \Delta \mathbf{w}^k). \quad (3.27)$$

Note that the one-dimensional (1-D) function $\tilde{f}_p(\alpha) \triangleq f_p(\mathbf{w}^k + \alpha \Delta \mathbf{w}^k)$ is convex w.r.t. α . The global minimizer of α is guaranteed to be found. Classical exact line search techniques such as the golden section search or tangential method [63] can be applied to find this global minimizer. The line search has a computational cost of $\mathcal{O}(N)$. In optimization literature, the PGM-1 is referred to as Armijo search along the feasible direction [64, 65].

- (b) PGM-2: The second step size is fixed as $\alpha_k = 1$. Then the iteration is simplified to

$$\mathbf{w}^{k+1} = P_{\mathcal{C}}(\mathbf{w}^k - \mu_k \nabla f_p(\mathbf{w}^k)). \quad (3.28)$$

Note that the 1-D function $\tilde{\tilde{f}}_p(\mu) \triangleq f_p(P_{\mathcal{C}}(\mathbf{w}^k - \mu \nabla f_p(\mathbf{w}^k)))$ is not convex w.r.t. μ even though $f_p(\mathbf{w})$ is convex w.r.t. \mathbf{w} . Therefore, it is difficult to find the global minimum of $\tilde{\tilde{f}}_p(\mu)$ and to perform an exact line search to determine the optimal step size μ_k . The non-convexity of $\tilde{\tilde{f}}_p(\mu)$ is due to the nonlinearity of the projection operator $P_{\mathcal{C}}(\cdot)$. Instead, it chooses μ_k to sufficiently decrease the objective function by an inexact line search, e.g., backtracking line search [62]. The procedure of backtracking line search is listed in Algorithm 3. The parameter $\gamma \in (0, 1)$. Typical algorithmic parameters are $\mu^0 = 1$, $\gamma = 0.5$, and $\delta = 0.1$. The PGM-2 is referred to as Armijo search along the boundary of \mathcal{C} [64, 65].

- (c) PGM-3: It is just a combination of PGM-1 and PGM-2. To be specific, the PGM-3 determines the first step size μ_k using the backtracking line search of PGM-2 and the second one α_k via the exact line search of PGM-1.

Algorithm 3 Backtracking line search

Initialize: $\mu = \mu^0$.
for $i = 1, 2, \dots$ **do**
 Compute $\mathbf{w}^{k+1} = P_{\mathcal{C}}(\mathbf{w}^k - \mu \nabla f_p(\mathbf{w}^k))$;
 Break if $f_p(\mathbf{w}^{k+1}) < f_p(\mathbf{w}^k) + \delta \cdot 2\text{Re}(\nabla f_p(\mathbf{w}^k)^H (\mathbf{w}^{k+1} - \mathbf{w}^k))$
 $\mu \leftarrow \gamma\mu$;
end for
Output: $\mu_k = \mu$.

Some remarks on the PGMs are listed in the following.

Remark 1: It can be verified that $\text{Re}(\nabla f_p(\mathbf{w}^k)^H (\mathbf{w}^{k+1} - \mathbf{w}^k)) < 0$ using the property of the POCS [64]. Hence, the three PGMs strictly decrease the objective function in each iteration. In other words, the three algorithms are descent methods. In addition, the objective function is bounded below zero. Therefore, the three algorithms converge. Furthermore, it has been proved that all the three PGMs converge to the global minimum for convex optimization. For the proof of the global convergence of the PGM, the interested reader is referred to [65, 66].

Remark 2: From the numerical results shown in Fig. 3.1, we see that the PGM-1 has a much slower convergence rate compared with PGM-2 and PGM-3. PGM-2 and PGM-3 have similar convergence performance. This result implies that selection of μ_k is more important than that of α_k . As can be seen from Fig. 3.2, the three PGMs need less than 10 iterations to achieve a satisfactory signal-to-interferences-plus-noise ratio (SINR) performance.

Remark 3: PGM-1 only needs to compute the projection onto \mathcal{C} once while many

projections are possibly required for PGM-2 and PGM-3 in the k th iteration. In the next subsection, we will prove that the projection $P_{\mathcal{C}}(\cdot)$ has a closed-form expression and is easy to compute with a low complexity. PGM-2 and PGM-3 are competitive since $P_{\mathcal{C}}(\cdot)$ is very easy to calculate. 3) *Closed-Form of Projection*: The remaining problem is to efficiently compute the projection $P_{\mathcal{C}}(\cdot)$. For any vector $\mathbf{z} \notin \mathcal{C}$, the projection onto \mathcal{C} is the point in \mathcal{C} that is closest to \mathbf{z} , which is the solution of the optimization

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{w} - \mathbf{z}\|^2 \\ \text{s.t.} \quad & \text{Re}(\mathbf{a}^H \mathbf{w}) \geq \varepsilon \|\mathbf{w}\| + 1. \end{aligned} \quad (3.29)$$

Note that \mathbf{z} is not in \mathcal{C} . Its projection onto \mathcal{C} , i.e., the optimal solution of (3.29), must lie in the boundary of \mathcal{C} . Thus, (3.29) with the inequality constraint is equivalent to the following one with an equality constraint

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{w} - \mathbf{z}\|^2 \\ \text{s.t.} \quad & \text{Re}(\mathbf{a}^H \mathbf{w}) = \varepsilon \|\mathbf{w}\| + 1. \end{aligned} \quad (3.30)$$

Denote $\mathbf{w} = \mathbf{w}_R + j\mathbf{w}_I$, $\mathbf{z} = \mathbf{z}_R + j\mathbf{z}_I$, $\mathbf{a} = \mathbf{a}_R + j\mathbf{a}_I$, and the expanded real-valued vectors

$$\bar{\mathbf{w}} = \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix}, \quad \bar{\mathbf{z}} = \begin{bmatrix} \mathbf{z}_R \\ \mathbf{z}_I \end{bmatrix}, \quad \bar{\mathbf{a}} = \begin{bmatrix} \mathbf{a}_R \\ \mathbf{a}_I \end{bmatrix} \in \mathbb{R}^{2M}. \quad (3.31)$$

By exploiting $\|\mathbf{w} - \mathbf{z}\|^2 = \|\bar{\mathbf{w}} - \bar{\mathbf{z}}\|^2$, $\|\mathbf{w}\|^2 = \|\bar{\mathbf{w}}\|^2$, and $\text{Re}(\mathbf{a}^H \mathbf{w}) = \mathbf{a}_R^T \mathbf{w}_R + \mathbf{a}_I^T \mathbf{w}_I = \bar{\mathbf{a}}^T \bar{\mathbf{w}}$, the optimization problem of (3.30) with complex-valued variables is converted

into the one with real-valued variables

$$\begin{aligned} \min_{\bar{\mathbf{w}}} \quad & \|\bar{\mathbf{w}} - \bar{\mathbf{z}}\|^2 \\ \text{s.t.} \quad & \bar{\mathbf{a}}^T \bar{\mathbf{w}} = \varepsilon \|\bar{\mathbf{w}}\| + 1. \end{aligned} \quad (3.32)$$

If the optimal solution of (3.32) denoted by $\bar{\mathbf{w}}^*$ is obtained, we can at once construct the projection $\mathbf{w}^* = P_C(\mathbf{z})$. Note that the optimal point $\bar{\mathbf{w}}^*$ is unique because the projection is unique. The following proposition describes how to efficiently compute the optimal solution of (3.32).

Proposition 3: The optimal solution of (3.32) has the closed-form

$$\bar{\mathbf{w}}^* = \frac{1}{1 + \lambda^* \varepsilon^2} \left(\bar{\mathbf{z}} + \frac{\lambda^* (r - 1 - \lambda^* \varepsilon^2)}{1 + \lambda^* (\varepsilon^2 - \|\bar{\mathbf{a}}\|^2)} \bar{\mathbf{a}} \right) \quad (3.33)$$

where

$$r = \bar{\mathbf{a}}^T \bar{\mathbf{z}} \quad (3.34)$$

is the inner product of $\bar{\mathbf{a}}$ and $\bar{\mathbf{z}}$, and λ^* is the optimal Lagrangian multiplier (dual variable) of the optimization problem (3.32). The optimal λ^* is the unique positive real root of the following quartic equation

$$g(\lambda) \triangleq g_4 \lambda^4 + g_3 \lambda^3 + g_2 \lambda^2 + g_1 \lambda + g_0 = 0 \quad (3.35)$$

where the polynomial coefficients have the forms of

$$\begin{aligned}
g_4 &= \varepsilon^6 (\|\bar{\mathbf{a}}\|^2 - \varepsilon^2) \\
g_3 &= 2\varepsilon^4 (\|\bar{\mathbf{a}}\|^2 - 2\varepsilon^2) \\
g_2 &= \varepsilon^2 (\|\bar{\mathbf{a}}\|^2 + 2(r-3)\varepsilon^2 - (\|\bar{\mathbf{a}}\|^2 - \varepsilon^2) r^2 \\
&\quad + (\|\bar{\mathbf{a}}\|^2 - \varepsilon^2)^2 \|\bar{\mathbf{z}}\|^2) \\
g_1 &= 2\varepsilon^2 (2r - (\|\bar{\mathbf{a}}\|^2 - \varepsilon^2) \|\bar{\mathbf{z}}\|^2 - 2) \\
g_0 &= \varepsilon^2 \|\bar{\mathbf{z}}\|^2 - (r-1)^2.
\end{aligned} \tag{3.36}$$

It is well known that the solution of quartic equation has an analytic form [67]. Therefore λ^* can be obtained in closed-form with a low complexity of $\mathcal{O}(1)$. According to (3.33), the projection has a closed-form, whose computational complexity is only $\mathcal{O}(M)$.

Proof: For the purpose of computational simplicity, we take square on both sides of the constraint $\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1 = \varepsilon \|\bar{\mathbf{w}}\|$ and consider the more loosened problem

$$\begin{aligned}
&\min_{\bar{\mathbf{w}}} \|\bar{\mathbf{w}} - \bar{\mathbf{z}}\|^2 \\
&\text{s.t. } (\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1)^2 = \varepsilon^2 \|\bar{\mathbf{w}}\|^2.
\end{aligned} \tag{3.37}$$

It can be seen later that there are two real-valued Karush-Kuhn-Tucker (KKT) points of (3.37). One corresponds to the original constraint $\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1 = \varepsilon \|\bar{\mathbf{w}}\|$ and the other corresponds to $\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1 = -\varepsilon \|\bar{\mathbf{w}}\|$. The optimal solution of (3.32) $\bar{\mathbf{w}}^*$ is the KKT point associated with the original constraint. The Lagrangian function of the optimization

problem (3.37) is

$$\mathcal{L}(\bar{\mathbf{w}}, \lambda) = \|\bar{\mathbf{w}} - \bar{\mathbf{z}}\|^2 + \lambda (\varepsilon^2 \bar{\mathbf{w}}^T \bar{\mathbf{w}} - \bar{\mathbf{w}}^T \bar{\mathbf{a}} \bar{\mathbf{a}}^T \bar{\mathbf{w}} + 2\bar{\mathbf{w}}^T \bar{\mathbf{a}} - 1). \quad (3.38)$$

According to the KKT condition [62], the optimal point satisfies

$$\frac{\partial \mathcal{L}(\bar{\mathbf{w}}, \lambda)}{\partial \bar{\mathbf{w}}} = 2(\bar{\mathbf{w}} - \bar{\mathbf{z}}) + 2\lambda (\varepsilon^2 \bar{\mathbf{w}} - \bar{\mathbf{a}} \bar{\mathbf{a}}^T \bar{\mathbf{w}} + \bar{\mathbf{a}}) = \mathbf{0} \quad (3.39)$$

which leads to

$$\bar{\mathbf{w}} = ((1 + \lambda \varepsilon^2) \mathbf{I} - \lambda \bar{\mathbf{a}} \bar{\mathbf{a}}^T)^{-1} (\bar{\mathbf{z}} - \lambda \bar{\mathbf{a}}). \quad (3.40)$$

Applying the matrix inversion lemma [68] yields

$$\begin{aligned} & ((1 + \lambda \varepsilon^2) \mathbf{I} - \lambda \bar{\mathbf{a}} \bar{\mathbf{a}}^T)^{-1} \\ &= \frac{1}{1 + \lambda \varepsilon^2} \left(\mathbf{I} + \frac{\lambda \bar{\mathbf{a}} \bar{\mathbf{a}}^T}{1 + \lambda (\varepsilon^2 - \|\bar{\mathbf{a}}\|^2)} \right). \end{aligned} \quad (3.41)$$

Substituting (3.41) into (3.40), we obtain

$$\begin{aligned} \bar{\mathbf{w}} &= \frac{1}{1 + \lambda \varepsilon^2} \left(\bar{\mathbf{z}} - \lambda \bar{\mathbf{a}} + \frac{\lambda (r - \lambda \|\bar{\mathbf{a}}\|^2)}{1 + \lambda (\varepsilon^2 - \|\bar{\mathbf{a}}\|^2)} \bar{\mathbf{a}} \right) \\ &= \frac{1}{1 + \lambda \varepsilon^2} \left(\bar{\mathbf{z}} + \frac{\lambda (r - 1 - \lambda \varepsilon^2)}{1 + \lambda (\varepsilon^2 - \|\bar{\mathbf{a}}\|^2)} \bar{\mathbf{a}} \right). \end{aligned} \quad (3.42)$$

Substituting (3.42) back into the constraint in (3.37) and performing some manipulations, we can derive that the optimal Lagrangian multiplier satisfies the quartic equation of (3.35) with the polynomial coefficients shown in (3.36).

The quartic equation of (3.35) has a pair of complex conjugate roots, a positive real root and a negative real root. The two complex roots are not of interest since

the optimal solution is real-valued. The negative real root corresponds to $\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1 = -\varepsilon \|\bar{\mathbf{w}}\|$ while the positive real root corresponds to the original constraint $\bar{\mathbf{a}}^T \bar{\mathbf{w}} - 1 = \varepsilon \|\bar{\mathbf{w}}\|$. Thus, the optimal Lagrangian multiplier λ^* is taken as the unique positive real root. \square

It is emphasized that the optimal primal and dual variables $\bar{\mathbf{w}}^*$ and λ^* has a closed-form solution because the quartic equation has analytic roots. The proposed PGMs are very fast since the projection onto the constraint set is easy to compute.

Now it is clear that the dominant cost of the PGMs is to calculate the gradient of (3.22) and evaluate the objective function, which has a complexity of $\mathcal{O}(NM)$ in each iteration. However, the complexity of the IPM in each iteration is $\max(\mathcal{O}(N^3), \mathcal{O}(M^3))^1$, which is much higher than the PGMs.

3.3.3 ℓ_∞ -Norm QCMD Beamformer

The proposed gradient projection algorithm is applicable to any finite value of $1 < p < \infty$. However, it does not work for the case of $p = \infty$ because the ℓ_∞ -norm function is non-differentiable². The ℓ_∞ -norm QCMD beamformer solves the optimization problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{X}^H \mathbf{w}\|_\infty \\ \text{s.t.} \quad & \text{Re}(\mathbf{a}^H \mathbf{w}) \geq \varepsilon \|\mathbf{w}\| + 1 \end{aligned} \quad (3.43)$$

where the ℓ_∞ -norm of the complex-valued vector $\mathbf{y} = \mathbf{y}_R + \mathbf{j}\mathbf{y}_I$ is defined as

$$\|\mathbf{y}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{y}\|_p = \max_{1 \leq n \leq N} |y(n)| \quad (3.44)$$

¹Since the sample size N is usually larger than the sensor number M , the complexity of the IPM in each iteration is $\max(\mathcal{O}(N^3), \mathcal{O}(M^3)) = \mathcal{O}(N^3)$.

²There is no definition of gradient for the non-differentiable functions and hence the gradient projection cannot be applied.

with the modulus being $|y(n)| = \sqrt{y_R^2(n) + y_I^2(n)}$.

By splitting the complex-valued variables into real-valued ones, the problem of (3.43) can be converted into the following SOCP:

$$\begin{aligned}
& \min_{\bar{\mathbf{w}}, \mathbf{y}_R, \mathbf{y}_I, t} t \\
& \text{s.t. } \sqrt{y_R^2(n) + y_I^2(n)} \leq t, \quad n = 1, \dots, N \\
& \begin{bmatrix} \mathbf{X}_R & \mathbf{X}_I \\ \mathbf{X}_I & -\mathbf{X}_R \end{bmatrix}^T \bar{\mathbf{w}} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} \\
& \bar{\mathbf{a}}^T \bar{\mathbf{w}} \geq \varepsilon \|\bar{\mathbf{w}}\| + 1
\end{aligned} \tag{3.45}$$

where $t \in \mathbb{R}^+$ is an auxiliary variable, $\mathbf{X} = \mathbf{X}_R + j\mathbf{X}_I$, $\bar{\mathbf{w}}$ and $\bar{\mathbf{a}}$ are defined in (3.31). The IPM is required to solve the SOCP of (3.45), which results in a high computational complexity. One can use a large enough p to approximate the ℓ_∞ -norm to avoid solving the SOCP. Simulation results demonstrate that $p = 20$ provides performance similar to that of the ℓ_∞ -norm. It should be pointed out that too large values for p are not advisable because it may result in overflow when computing the p th power of a number.

Note that the ℓ_∞ -norm QCMD beamformer is different from the recently proposed robust linear programming beamformer (RLPB) [50], where an ℓ_∞ -norm minimization criterion is also used. The ℓ_∞ -norm adopted by the RLPB is based on the ℓ_∞ -modulus of the complex numbers but the QCMD beamformer uses the conventional modulus. Additionally, the QCMD beamformer makes use of a spherical uncertainty set while the RLPB adopts a rhombic uncertainty region.

3.4 Robust Constant Modulus Beamformer

The QCMD beamformer exploits the property of non-Gaussianity or finite alphabet of the signals. This property has also been widely used in blind equalization. By borrowing the idea of the most representative blind equalization techniques, i.e., the constant modulus algorithm (CMA) [37, 69], we propose to use the following constant modulus (CM) criterion

$$\min \mathbb{E} \left\{ (|y(n)|^2 - \kappa)^2 \right\} \quad (3.46)$$

for beamforming, where $\kappa > 0$ is the dispersion constant [37] defined as

$$\kappa = \frac{\mathbb{E} \{|s(n)|^4\}}{\mathbb{E} \{|s(n)|^2\}}. \quad (3.47)$$

Different constellations may have different dispersion constants. Expanding (4.30) and replacing the expectation by the sample mean, the CM criterion is equivalent to minimizing

$$f_{\text{CM}}(\mathbf{w}) = \|\mathbf{y}\|_4^4 - 2\kappa\|\mathbf{y}\|^2 = \|\mathbf{X}^H\mathbf{w}\|_4^4 - 2\kappa\|\mathbf{X}^H\mathbf{w}\|^2 \quad (3.48)$$

where $\|\mathbf{y}\|_4$ is the ℓ_4 -norm of \mathbf{y} . Therefore the CM beamformer minimizes a combination of the ℓ_4 -norm and ℓ_2 -norm, which exploits both the fourth- and second-order statistics. Since the objective function of the CM criterion also describes the dispersion of the recovered signal, the CM beamformer still belongs to the minimum dispersion beamforming. By imposing the constraint in (3.11), we formulate the

following optimization problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & f_{\text{CM}}(\mathbf{w}) \\ \text{s.t.} \quad & \text{Re}(\mathbf{a}^H \mathbf{w}) \geq \varepsilon \|\mathbf{w}\| + 1 \end{aligned} \quad (3.49)$$

whose solution corresponds to the RCMB. Note that $f_{\text{CM}}(\mathbf{w})$ is nonconvex and it is difficult to find the global minimum of (3.49). This constitutes a drawback of the RCMB. Instead of seeking the global minimum, we can employ the three PGMs to obtain a stationary point of (3.49). Simulation results demonstrate that a stationary point of (3.49) can provide a satisfactory performance. The PGMs for solving (3.49) are almost the same as those for solving (3.11) in Section 3.3. That is, we just replace the gradient of $f_p(\mathbf{w})$ by the gradient of $f_{\text{CM}}(\mathbf{w})$, which is

$$\nabla f_{\text{CM}}(\mathbf{w}) = 2\mathbf{X}(\mathbf{D}(\mathbf{w}) - \kappa\mathbf{I})\mathbf{X}^H\mathbf{w}. \quad (3.50)$$

Remark 4: For common digital modulated communication signals, such as PSK and QAM, we find that the performance of the RCMB is better than that of QCMD beamformer with $p = 4$ but is inferior to that of $p = 8$. This is because the RCMB exploits a combination of the fourth- and second-order statistics, which is helpful to improve the performance compared with the use of only fourth-order statistics.

3.5 Simulation Results

Similar parameter settings as in [26] and [24] are taken in the simulations. A uniform linear array (ULA) of $M = 10$ omnidirectional sensors with a half-wavelength spacing

is considered. For ULA, the steering vector has the following form:

$$\mathbf{a}(\theta) = [1, e^{j(2\pi/\zeta)d \sin \theta}, \dots, e^{j(M-1)(2\pi/\zeta)d \sin \theta}]^T \quad (3.51)$$

where θ is the AOA, d is the inter-sensor spacing, and ζ is the wavelength. Three zero-mean sub-Gaussian signals, namely, the desired source $s(n)$ and two uncorrelated interferences $s_1(n)$ and $s_2(n)$, impinge on the array. Unless stated otherwise, the AOA of the desired signal is $\theta = 43^\circ$ and the AOAs of the two interferences are $\theta_1 = 30^\circ$ and $\theta_2 = 75^\circ$. We consider applications in communications because most of the corresponding signals are sub-Gaussian. We take quadrature phase shift keying (QPSK) scheme as example. That is, the desired signal and interferences adopts QPSK modulation while the noise is Gaussian distributed. The SNR is defined as

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_v^2} \quad (3.52)$$

where σ_s^2 and σ_v^2 are the variances of the desired signal and additive noise, respectively. The two interferences are stronger than the desired signal with variances being $\sigma_1^2 = \sigma_2^2 = 10\sigma_s^2$. That is, they are 10 dB above the desired signal. We adopt the output SINR as the performance measure of beamforming, which is defined as

$$\text{SINR} = \frac{\text{E} \left\{ |s(n) \mathbf{w}^H \mathbf{a}|^2 \right\}}{\text{E} \left\{ |\mathbf{w}^H (\mathbf{i}(n) + \mathbf{v}(n))|^2 \right\}} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{a}|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}}. \quad (3.53)$$

The output SINRs of six robust beamformers, namely, the subspace [15], RMVB [18,24], general-rank [21], RCMB, LCMD [49] and QCMD beamformers with different values of p , are compared. Four values of $p = 4, 8, 20$, and ∞ are taken for the

LCMD and QCMD beamformers. Note that the results of $p < \infty$ are obtained by using the proposed PGM while those of $p = \infty$ are computed by solving the SOCP of (3.45) using the IPM. The upper bound of the SINR is the maximum eigenvalue of the matrix $\sigma_s^2 \mathbf{R}_{i+n}^{-1} \mathbf{a} \mathbf{a}^H$, which is also provided for comparison. Since the subspace beamformer requires the dimension of the signal-plus-interference subspace, the minimum description length (MDL) principle [48] is adopted to estimate this quantity. When plotting the SINR curves, 200 Monte Carlo trials are performed for their computation.

3.5.1 Convergence Behavior of PGM

We first investigate the convergence behavior of three PGMs. Figs. 3.1 and 3.2 plot the objective function and output SINR versus the number of iterations, respectively. The three PGMs take the same initial value of (3.19). Three typical values of $p = 4, 8, \text{ and } 20$ are tried. The global minimum of the objective function is also plotted in Fig. 3.1. The upper bound of the SINR and the output SINR of the RMVB are plotted in Fig. 3.2 for the purpose of comparison. As can be seen, the PGM-1 has a much slower convergence rate to the global minimum compared with PGM-2 and PGM-3. PGM-2 and PGM-3 have similar convergence rate. The three PGMs need less than 10 iterations to achieve a satisfactory SINR performance. Since the output SINR reflects the performance of a beamformer, it requires no more than 10 iterations of the PGMs for beamforming.

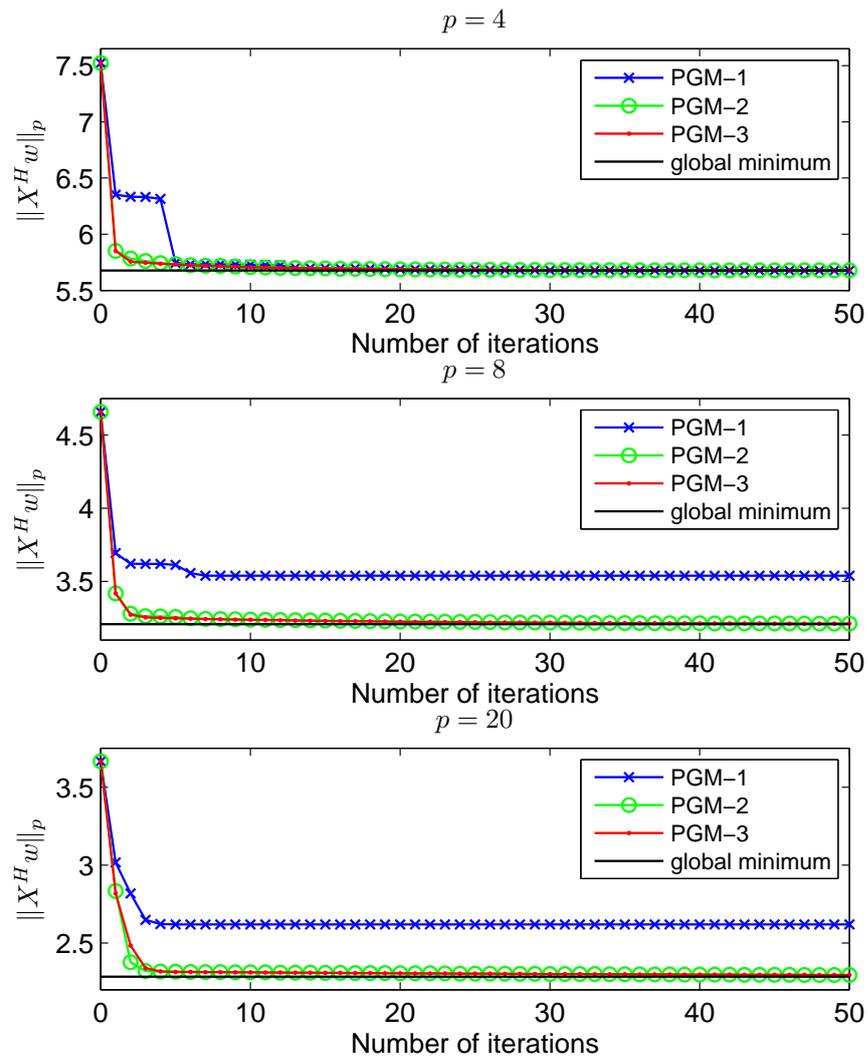


Figure 3.1: Objective function versus number of iterations.

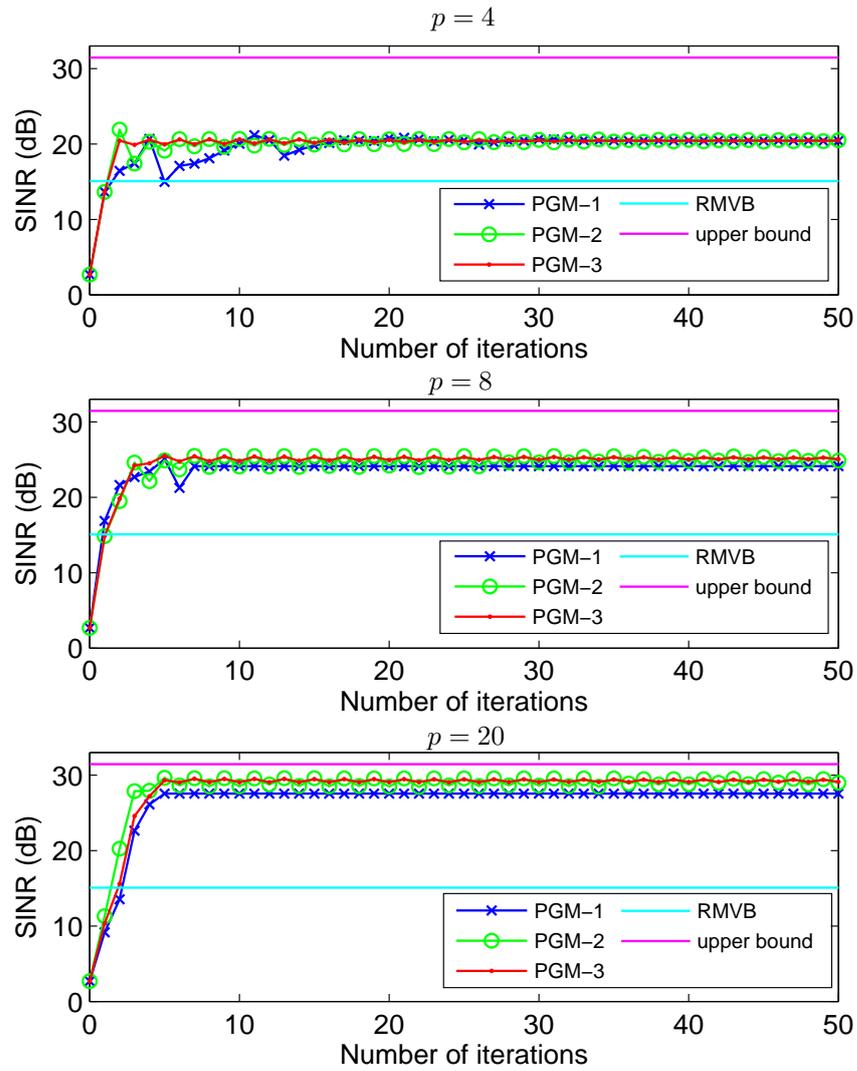


Figure 3.2: Output SINR versus number of iterations.

3.5.2 Running Time Comparison

In this simulation, we compare the running time of the proposed PGMs with a standard convex optimization package for solving (3.11). The package for convex optimization is CVX version 1.22, available online at [70]. Since the three PGMs have similar computational complexity, we only list the results of PGM-2 as example. We run the two methods in MATLAB on a computer with a 2.2 GHz CPU and 2 GB memory. The sensor number is fixed to $M = 10$ while the received signal length N takes values of 100, 500, 1000, 5000, and 10^4 . The running times (in seconds) of the two methods for $p = 4, 8, \text{ and } 20$, which are based on an average of 20 independent runs, are listed in Tables 3.1, 3.2 and 3.3, respectively.

It can be observed that the proposed PGM is much faster than the CVX package using standard IPM, especially for large N . The CVX package is too time-consuming when the problem size is large. The running time of the PGM is approximately linear with N while that of the CVX is cubic with N . This agrees with the earlier conclusion that the PGM and IPM have the computational complexities of $\mathcal{O}(NM)$ and $\mathcal{O}(N^3)$ in one iteration, respectively. Since the complexity of the proposed PGM is significantly smaller, it is very efficient for solving the QCMD beamforming problem with large sample size N .

Table 3.1: Running time of PGM-2 and CVX for $p = 4$.

N	100	500	1000	5000	10^4
PGM-2	0.0399	0.2324	0.3058	1.3884	2.7043
CVX	0.5321	15.4722	125.795	1.592×10^4	Out of memory

Table 3.2: Running time of PGM-2 and CVX for $p = 8$.

N	100	500	1000	5000	10^4
PGM-2	0.0416	0.1685	0.2855	1.4586	2.9593
CVX	0.5980	16.321	135.602	1.735×10^4	Out of memory

Table 3.3: Running time of PGM-2 and CVX for $p = 20$.

N	100	500	1000	5000	10^4
PGM-2	0.0572	0.1981	0.3869	1.9758	3.9577
CVX	0.7609	18.789	147.613	1.968×10^4	Out of memory

3.5.3 Random Steering Vector Mismatch

In this subsection, we consider the case of random steering vector mismatch, where the error \mathbf{e} is modeled as zero-mean circular Gaussian distributed variables with variance σ_e^2 , that is,

$$\mathbf{e} \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_M). \quad (3.54)$$

According to the central limit theorem [71], the Gaussian distribution can well model the arbitrary random model errors. The RMVB, QCMD, and RCMB take the same algorithmic parameter $\varepsilon = 5.6\sigma_e$ such that the probability of the steering vector lying in the uncertainty region is 95% [50]. The parameter for the general-rank beamformer is $\varepsilon_F = 2\sqrt{M}\varepsilon$. Note that the LCMV and LCMD beamformers are designed only for AOA mismatch and cannot handle random mismatch. Therefore, we do not include the results of the LCMV and LCMD beamformers in this experiment.

The relative perturbation

$$\text{Ptb} = \frac{\text{tr}(\sigma_e^2 \mathbf{I}_M)}{\|\mathbf{a}\|^2} = \sigma_e^2 \quad (3.55)$$

describes the quantitative steering vector error. The steering vector perturbation is

first fixed to $P_{tb} = -10$ dB while the SNR and N vary. Fig. 3.3 plots the output SINR versus SNR with $N = 100$. Fig. 3.4 displays the output SINR versus N at SNR = 15 dB. Then the SNR is fixed to 15 dB and $N = 100$ while the array perturbation varies from -30 dB to -6 dB. The -6 dB array perturbation corresponding to $\sigma_e = 0.5$ causes a large steering vector error. Fig. 3.5 shows the SINR versus steering vector perturbation.

It can be observed from Figs. 3.3 to 3.5 that the QCMD beamformer has the best performance for random steering vector mismatch. Generally speaking, the performance of QCMD beamformer improves as p increases. Therefore, a larger value of p will improve the beamforming performance for sub-Gaussian communication signals. The QCMD beamformer of $p = 20$ significantly outperforms other robust beamforming approaches. For example, as shown in Fig. 3.5, the output SINR of the QCMD beamformer is about 8 dB higher than those of robust minimum variance beamformers. It maintains satisfactory performance with large steering vector errors and smaller numbers of samples. Furthermore, only the performance of the QCMD beamformer can approach the optimal bounds with high SNR, large sample size, or small amount of perturbation.

Next, we investigate the impact of the interference number on beamforming performance with SNR = 15 dB, $P_{tb} = -10$ dB, and $N = 100$. The AOA of the desired signal is fixed to 43° while that of the i th ($1 \leq i \leq I$) interference is $\theta_i = -30^\circ + (i-1)10^\circ$. The power of all interferers is the same and 10 dB higher than that of the desired signal. We observe that the performance of the subspace beamformer substantially degrades for larger number of signals. Fig. 3.6 shows the output SINR versus interference number. It can be observed that the performance of subspace, RMVB, and

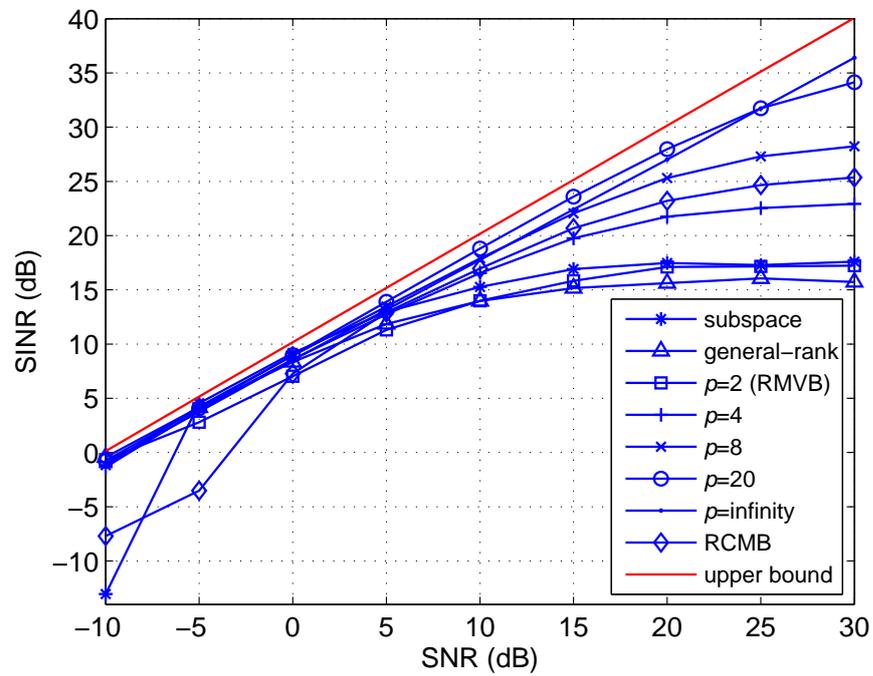


Figure 3.3: Output SINR versus SNR with random steering vector mismatch.

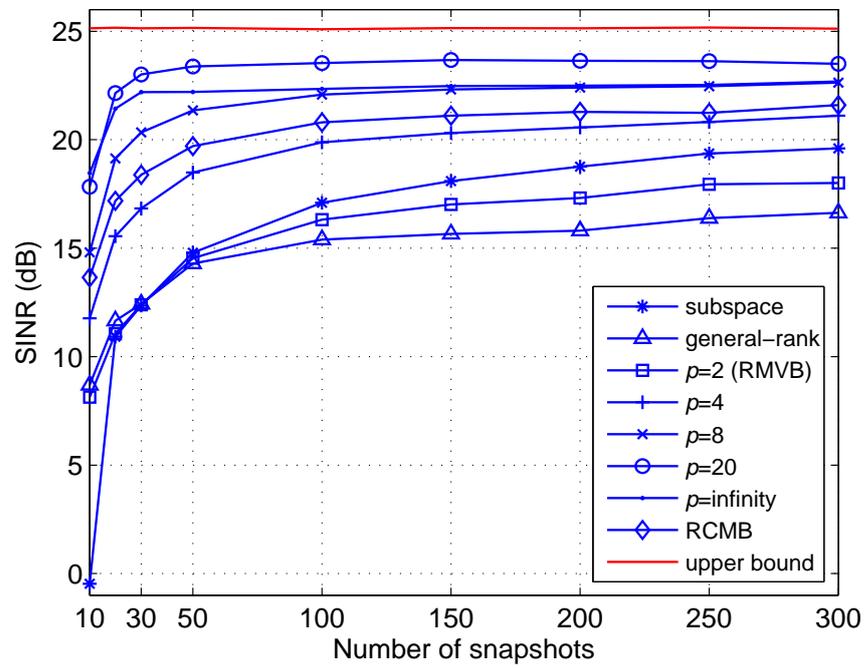


Figure 3.4: Output SINR versus number of samples with random steering vector mismatch.

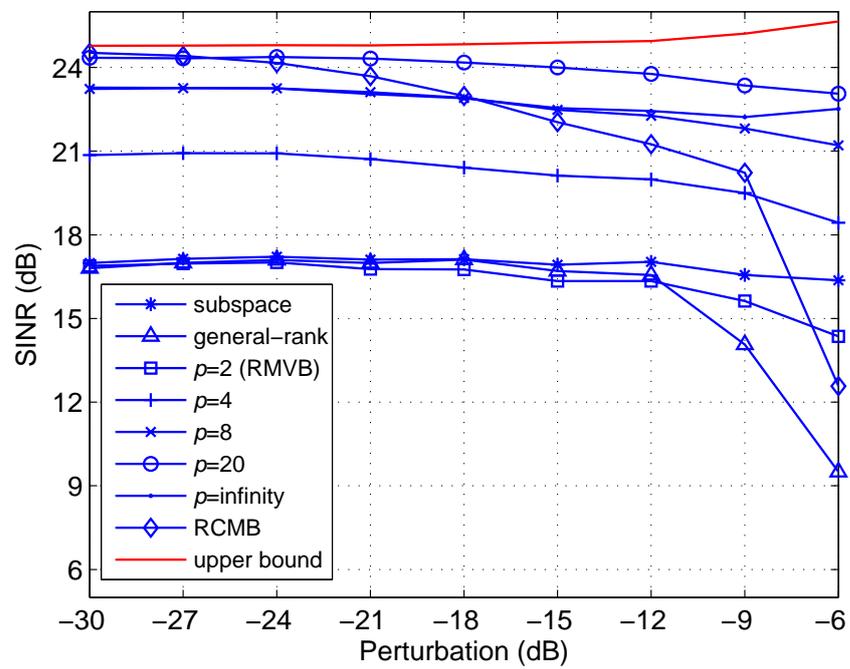


Figure 3.5: Output SINR versus perturbation amount of steering vector.

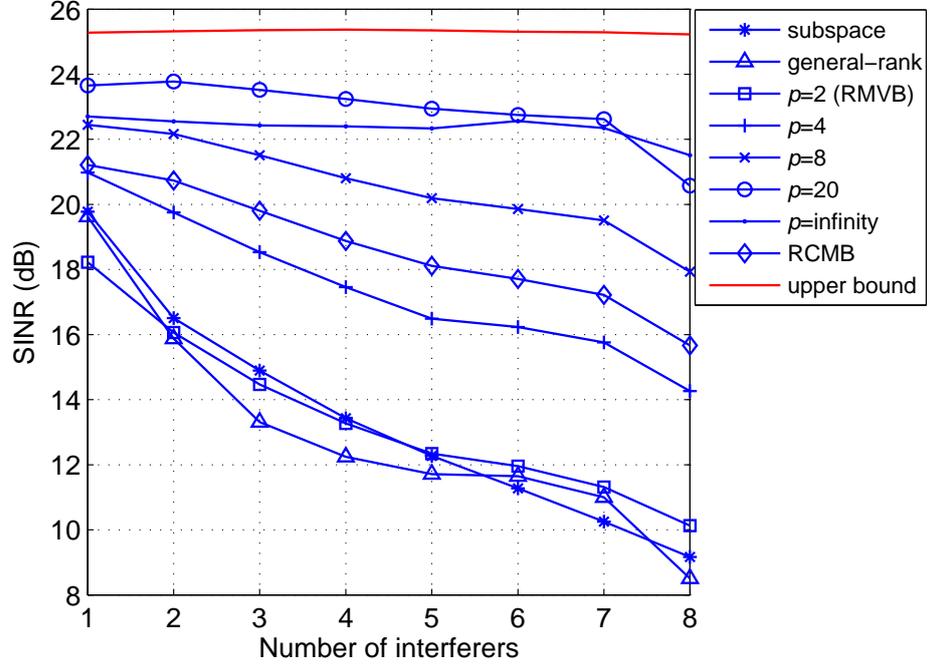


Figure 3.6: Output SINR versus number of interferers.

general-rank beamformers dramatically degrades as the number of interferences increases while the QCMD beamformer is less sensitive to the number of interferences.

3.5.4 AOA Mismatch

We then consider the steering vector mismatch induced by AOA estimation error. The assumed AOA of the desired signal is 45° but the true AOA is $45^\circ + \Delta\theta$ with $\Delta\theta$ being the AOA estimation error. The maximal possible AOA deviation for all beamformers is assigned as 3° . In other words, the true AOA lies in the interval $[42^\circ, 48^\circ]$. Hence for the LCMV and LCMD beamformers, two linear constraints force the responses of the

signals from 42° and 48° to be unity. For fair comparison, all beamforming methods use the same *a priori* information on the maximal possible AOA mismatch. Therefore the algorithm parameter for the RMVB and QCMD beamformer is determined as

$$\varepsilon = \max_{42^\circ \leq \theta \leq 48^\circ} \|\mathbf{a}(\theta) - \mathbf{a}(45^\circ)\| = 1.95 \quad (3.56)$$

while for the general-rank beamformer

$$\varepsilon_F = \max_{42^\circ \leq \theta \leq 48^\circ} \|\mathbf{a}(\theta)\mathbf{a}^H(\theta) - \mathbf{a}(45^\circ)\mathbf{a}^H(45^\circ)\|_F = 4.73. \quad (3.57)$$

The SNR is fixed to 15 dB and $N = 100$. Fig. 3.7 plots the SINR versus AOA mismatch, when the AOA deviation varies from -3° to 3° with an interval 0.5° . The results of Fig. 3.7 illustrate that there is a substantial performance improvement from the LCMD to QCMD by adopting the quadratic constraint instead of multiple linear constraints. Again, the QCMD beamformer with relative large value of p outperforms other robust beamformers and approaches the optimal bound.

3.6 Conclusion

A QCMD beamformer that minimizes the ℓ_p -norm of the output while constraining the magnitude response of any steering vector in a spherical uncertainty region to exceed unity is proposed. As a major contribution of this chapter, the property of the convex constraint set is analyzed and a generic gradient projection framework which includes three PGMs is developed for efficiently solving the resulting constrained ℓ_p -minimization problem. The proposed PGMs are quite suitable for the

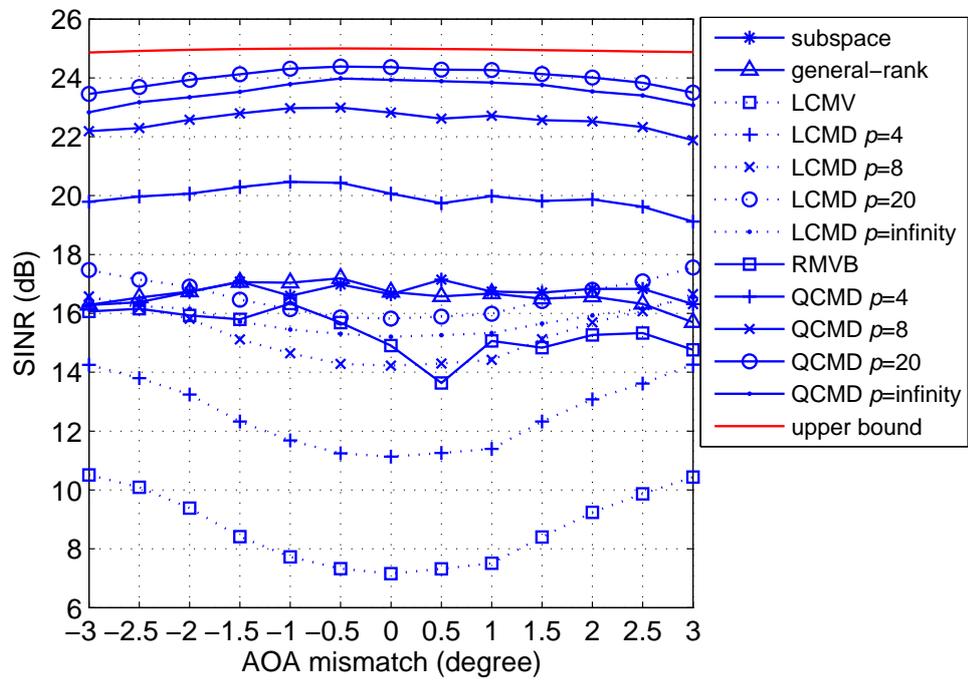


Figure 3.7: Output SINR versus AOA mismatch.

robust beamforming problem with large sample size or large sensor number because their computational complexity is linearly proportional to the number of samples and sensors. By implicitly exploiting the higher-order statistics, the QCMD beamformer substantially improves the SINR performance compared with the MV based robust beamformers.

Chapter 4

Robust Beamforming by Linear Programming

In this chapter, a robust linear programming beamformer (RLPB) is proposed for non-Gaussian signals in the presence of steering vector uncertainties. Unlike most of the existing beamforming techniques based on the minimum variance criterion, the proposed RLPB minimizes the ℓ_∞ -norm of the output to exploit the non-Gaussianity. We make use of a new definition of the ℓ_p -norm ($1 \leq p \leq \infty$) of a complex-valued vector, which is based on the ℓ_p -modulus of complex numbers. To achieve robustness against steering vector mismatch, the proposed method constrains the ℓ_∞ -modulus of the response of any steering vector within a specified uncertainty set to exceed unity. The uncertainty set is modeled as a rhombus, which differs from the spherical or ellipsoidal uncertainty region widely adopted in the literature. The resulting optimization problem is cast as a linear programming and hence can be solved efficiently. The proposed RLPB is computationally simpler than its robust counterparts requiring

solution to a second-order cone programming. We also address the issue of appropriately choosing the uncertainty region size. Simulation results demonstrate the superiority of the proposed RLPB over several state-of-the-art robust beamformers and show that its performance can approach the optimal performance bounds. Most of the results of this chapter have been reported in our recent journal paper [50].

4.1 Introduction

A classical data-dependent beamforming method is the minimum variance distortionless response (MVDR) beamformer [11]. It minimizes the output variance while keeping the response of the desired signal to unity. The MVDR beamformer is based on the second-order covariance matrix of the interference and noise components. However, the true covariance matrix is not available in practice and is always substituted by the sample covariance matrix [7, 8]. The minimum number of samples needed to achieve a satisfactory performance is given by the well-known Reed-Mallett-Brennan rule [46]. When the sample number is too small, the performance of the MVDR beamformer will degrade. Another drawback of the MVDR beamformer is that it is too sensitive to steering vector mismatches [7]. Therefore, it requires precise knowledge of the steering vector of the desired signal. Nevertheless, the steering vector is not known exactly due to a variety of reasons, such as angle-of-arrival (AOA) mismatch, imperfect array calibration, local scattering, and antenna shape distortion [8]. The steering vector uncertainty causes the phenomenon of signal cancellation [7, 26] and may lead to a dramatic performance degradation in the MVDR beamformer.

Several robust beamforming approaches have been proposed to mitigate the steering vector uncertainties [26]–[35]. The linearly constrained minimum variance (LCMV)

beamformer [14], which is a direct extension of the MVDR beamformer, handles the AOA mismatch by imposing multiple linear constraints for a small spread of angles around the nominal AOA. It broadens the mainbeam and hence enhances the robustness against the AOA mismatch. However, there are two disadvantages with the LCMV beamformer. On one hand, as more linear constraints are added, the beamformer loses more degrees of freedom. That is, although the robustness is enhanced, the remaining degrees of freedom for interference suppression is reduced [8]. On the other hand, the LCMV beamformer is designed only for AOA mismatch but not random mismatch. There are several other techniques developed for dealing with this type of mismatch (see [17, 26] and the references therein). Note that these methods are restricted to scenarios with AOA mismatch. The eigenspace-based beamformer [15] is a powerful robust scheme that can cope with arbitrary steering vector mismatches [18]. It uses only the signal-plus-interference subspace component of the sample covariance matrix, which can mitigate the adverse effect induced by the noise subspace disturbance. Hence the subspace scheme outperforms the MVDR and LCMV beamformers. Nonetheless, it is effective only under high signal-to-noise ratio (SNR) conditions and with sufficiently large data. Furthermore, it requires that the dimension of the signal-plus-interference subspace be exactly known and be much lower than the number of sensors [18]. The diagonal loading method based on regularization [16] is also widely used for robust beamforming. But it is not clear how to choose an appropriate diagonal loading factor.

In the last decade, many advanced mismatch-robust beamforming approaches based on convex optimization have been developed [7, 18, 24, 27, 28]. In these methods, the output variance is minimized while the magnitude responses of the steering vectors

in an uncertainty set are constrained to exceed unity. This idea is called worst-case performance optimization [18]. The uncertainty set of the steering vector is modeled as a sphere in [18] and as a more general ellipsoid in the robust minimum variance beamformer (RMVB) [24]. Employing the spherical or ellipsoidal uncertainty region, the original nonconvex optimization with infinitely many quadratic constraints is ultimately converted into a second-order cone programming (SOCP) problem, which can be efficiently solved [18]. Furthermore, it has been shown that the RMVB [24] and robust Capon beamformer (RCB) [20] can be obtained in a form of diagonal loading using the Lagrangian multiplier technique. Although there is no closed form expression for the optimal diagonal loading factor, it can be calculated by solving a nonlinear equation with low computational complexity. In [21], a general-rank signal model has been considered based on worst-case performance optimization. This model is applicable to both rank-one (point source) and higher-rank (spread source) signal models. Multiple quadratic constraints are employed instead of the linear ones of the LCMV beamformer in [28], which leads to a quadratically constrained quadratic program (QCQP). The resulting nonconvex QCQP is relaxed to a semidefinite programming (SDP) problem [28]. The SDP-based beamformer of [28] can only handle finite constraints and is only applicable to AOA mismatch. In [27], the robust beamforming via worst-case signal-to-interferences-plus-noise ratio (SINR) maximization can be cast as the SDP. It is more general and flexible in modeling uncertainty than prior works using a special or ellipsoidal form of uncertainty region.

All robust beamforming techniques mentioned above [14]– [28] are based on the minimum variance (MV) criterion. In fact, most of the existing robust beamformers use the MV beamforming framework [8]. The MV criterion is statistically optimal

under the Gaussian assumption because zero-mean Gaussian distributions are completely characterized by their second-order statistics. However, real-world signals often exhibit non-Gaussianity. Many digitally modulated signals arising in radar, sonar and wireless communications are demonstrated to be sub-Gaussian with kurtosis smaller than that of Gaussian distribution [39, 41]. In this case, the higher-order statistics, which contain useful statistical information, can be utilized to improve the performance of the beamformer. In [44], a blind beamforming method using the fourth-order cumulants of the array output is proposed for non-Gaussian signals. However, it only exploits the fourth-order components while other higher-order statistics are ignored. In this chapter, we focus on designing a robust beamformer for sub-Gaussian signals that are frequently encountered in many practical applications. We introduce the ℓ_p -modulus ($1 \leq p \leq \infty$) and a new concept of the ℓ_p -norm of complex numbers. The minimum ℓ_∞ -norm criterion is then adopted by the proposed beamformer, which implicitly exploits the higher-order statistics of the observed signal.

In contrast to the SOCP-based beamforming approaches that assume a spherical or ellipsoidal uncertainty set, we model the uncertainty region as a rhombus in which the ℓ_1 -norm of the steering vector error is bounded. The issue of determining the size of the rhombic uncertainty region is discussed for both AOA and random mismatches. The proposed beamformer minimizes the ℓ_∞ -norm of the output while the ℓ_∞ -modulus response¹ of any steering vector in a rhombic set is constrained to exceed unity to guarantee robustness against mismatch. As a result of the newly defined ℓ_∞ -modulus and ℓ_p -norm as well as the rhombic uncertainty set, the proposed beamformer can

¹All existing beamforming methods use the magnitude (modulus) response while the proposed beamformer adopts the ℓ_∞ -modulus response.

be obtained by solving a linear programming (LP) problem. Hence our proposed technique is referred to as the robust linear programming beamformer (RLPB). The LP is the simplest convex optimization problem and, thus, can be efficiently solved by the well-established simplex method [72] or interior point method [55]. Furthermore, it is known that the computational effort per iteration required by interior point methods to solve the LP problem is much less than that required to solve SOCP or SDP problem with similar size and structure [73, 74]. Therefore the proposed RLPB is computationally simpler than the SOCP and SDP based beamformers.

The remainder of this chapter is organized as follows. In Section 4.2, we present the RLPB for mismatch-robust beamforming. The issue of algorithm parameter selection is also discussed. A theoretical analysis of the RLPB from the statistical viewpoint is given in Section 4.3. Simulation results are provided in Section 4.4 to demonstrate the effectiveness of the proposed beamformer. Finally, conclusions are drawn in Section 4.5.

4.2 Robust Linear Programming Beamformer

As stated in Chapter 1, the performance of the sample matrix inversion (SMI) beamformer will degrade when the sample size N is small. Moreover, it is sensitive to steering vector mismatches and its performance deteriorates in the presence of array mismatches [7, 8]. In this section, we derive the new robust beamformer based on linear programming. First, we introduce the definitions of the ℓ_p -modulus of a complex number and the ℓ_p -norm of a complex-valued vector, which is used by the proposed beamforming technique.

4.2.1 ℓ_p -Modulus and New Definition of ℓ_p -Norm

The standard modulus of a complex number $z = z_R + jz_I$ is defined as

$$|z| = \sqrt{z_R^2 + z_I^2}. \quad (4.1)$$

As a generalization of the standard modulus, the ℓ_p -modulus of z , which is denoted by $|z|_p$, is defined as [75, 76]

$$|z|_p = (|z_R|^p + |z_I|^p)^{\frac{1}{p}}. \quad (4.2)$$

where $p \geq 1$. The standard modulus is a special case of the ℓ_p -modulus for $p = 2$. By the Minkowski inequality [77], the triangle inequality

$$|z_1 + z_2|_p \leq |z_1|_p + |z_2|_p \quad (4.3)$$

holds for any z_1 and z_2 . Hence the ℓ_p -modulus can be viewed as a distance (modulus) metric of a complex number. The limit of (4.2) as $p \rightarrow \infty$ yields the ℓ_∞ -modulus

$$|z|_\infty = \lim_{p \rightarrow \infty} |z|_p = \max(|z_R|, |z_I|). \quad (4.4)$$

The ℓ_p -modulus is sensitive to phase rotation for $p \neq 2$. That is, $|z|_p = |ze^{j\phi}|_p$ does not hold for any z and $\phi \in [0, 2\pi)$. Due to this property, the ℓ_p -modulus has been used for automatic phase recovery in blind equalization [75, 76].

The standard definition of the ℓ_p -norm of the complex vector $\mathbf{z} = [z_1, \dots, z_M]^T \in$

\mathbb{C}^M is given by

$$\|\mathbf{z}\|_{\ell_p} = \left(\sum_{i=1}^M |z_i|^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^M (\operatorname{Re}^2(z_i) + \operatorname{Im}^2(z_i))^{p/2} \right)^{\frac{1}{p}}. \quad (4.5)$$

The standard ℓ_p -norm $\|\mathbf{z}\|_{\ell_p}$ is based on the conventional modulus. The proposed beamforming technique uses a new definition of the ℓ_p -norm, which is based on the ℓ_p -modulus of (4.2) and has the form of

$$\|\mathbf{z}\|_p = \left(\sum_{i=1}^M |z_i|_p^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^M |\operatorname{Re}(z_i)|^p + |\operatorname{Im}(z_i)|^p \right)^{\frac{1}{p}}. \quad (4.6)$$

Note that the new ℓ_p -norm is denoted as $\|\mathbf{z}\|_p$ to distinguish it from the conventional ℓ_p -norm $\|\mathbf{z}\|_{\ell_p}$. The two norms are equivalent only for $p = 2$. In this case, the two norms equal the Euclidean norm, that is, $\|\mathbf{z}\|_2 = \|\mathbf{z}\|_{\ell_2} = \sqrt{\mathbf{z}^H \mathbf{z}}$. The Euclidean norm will be denoted as $\|\mathbf{z}\|$ for convenience. The new ℓ_∞ -norm is obtained by taking the limit

$$\begin{aligned} \|\mathbf{z}\|_\infty &= \lim_{p \rightarrow \infty} \|\mathbf{z}\|_p = \max_{1 \leq i \leq M} |z_i|_\infty \\ &= \max_{1 \leq i \leq M} (\max(|\operatorname{Re}(z_i)|, |\operatorname{Im}(z_i)|)). \end{aligned} \quad (4.7)$$

The ℓ_p -norm $\|\mathbf{z}\|_p$ of the M -dimensional complex vector $\mathbf{z} = \mathbf{z}_R + j\mathbf{z}_I$ can be viewed as the ℓ_p -norm of the $2M$ -dimensional real vector $\bar{\mathbf{z}} = [\mathbf{z}_R^T, \mathbf{z}_I^T]^T$, namely,

$$\|\mathbf{z}\|_p = \|\bar{\mathbf{z}}\|_p. \quad (4.8)$$

In other words, the new ℓ_p -norm in \mathbb{C}^M is equivalent to the standard ℓ_p -norm in \mathbb{R}^{2M} . Therefore all the properties of a norm such as triangle inequality hold for the new

ℓ_p -norm. In particular, we have

$$\|\mathbf{z}\|_1 = \left\| \begin{bmatrix} \mathbf{z}_R^T \\ \mathbf{z}_I^T \end{bmatrix}^T \right\|_1 \quad (4.9)$$

and

$$\|\mathbf{z}\|_\infty = \left\| \begin{bmatrix} \mathbf{z}_R^T \\ \mathbf{z}_I^T \end{bmatrix}^T \right\|_\infty. \quad (4.10)$$

4.2.2 Formulation

Most of the existing robust beamforming techniques are based on minimizing the variance of the output [8]. They belong to the class of second-order statistics methods. The MV criterion is statistically optimal for Gaussian signals and noise because the first- and second-order statistics of a Gaussian distribution contain all necessary statistical information. Nevertheless, many real-world signals encountered in radar, sonar, wireless communications, and navigation are sub-Gaussian distributed [37, 41]. For sub-Gaussian signals, the higher-order (larger than 2) statistics contain useful information and can be exploited to improve the performance of beamforming [44]. To utilize the higher-order statistics, we suggest to minimize the ℓ_p -norm with $p > 2$ of the output instead of minimizing the variance. In particular, the proposed beamformer is based on the minimum of ℓ_∞ -norm.

Stacking N snapshots of the beamformer output $y(n) = \mathbf{w}^H \mathbf{x}(n)$ into a vector $\mathbf{y} = [y(1), \dots, y(N)]^T$, one has the following matrix-vector formulation:

$$\mathbf{y}^* = \mathbf{X}^H \mathbf{w} \quad (4.11)$$

where \mathbf{y}^* is the conjugate of \mathbf{y} . The proposed beamformer minimizes the ℓ_∞ -norm of

the output, namely,

$$\min_{\mathbf{w}} \|\mathbf{X}^H \mathbf{w}\|_{\infty}. \quad (4.12)$$

Denoting $\mathbf{w} = \mathbf{w}_R + j\mathbf{w}_I$, $\mathbf{X} = \mathbf{X}_R + j\mathbf{X}_I$, and the expanded real-valued matrix and vector

$$\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_R & -\mathbf{X}_I \\ \mathbf{X}_I & \mathbf{X}_R \end{bmatrix} \in \mathbb{R}^{2M \times 2N}, \quad \bar{\mathbf{w}} = \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix} \in \mathbb{R}^{2M} \quad (4.13)$$

and using (4.10), we convert the ℓ_{∞} -norm of complex variables into that of real variables

$$\|\mathbf{X}^H \mathbf{w}\|_{\infty} = \|\bar{\mathbf{X}}^T \bar{\mathbf{w}}\|_{\infty}. \quad (4.14)$$

In practical applications, due to a variety of mismatches, the steering vector \mathbf{a} is not known exactly. The actual steering vector is expressed as

$$\mathbf{c} = \mathbf{a} + \mathbf{e} \quad (4.15)$$

where $\mathbf{e} \in \mathbb{C}^M$ is the steering vector error. It is assumed that the error is in an uncertainty set. The goal of robust beamforming is to ensure that the magnitude response does not attenuate in the uncertainty set. In [18], the uncertainty region is modeled as a sphere, that is, $\|\mathbf{e}\| \leq \varepsilon_2$, where ε_2 is the radius of the sphere. In [24], the uncertainty region is modeled as an ellipsoid. Under the spherical or ellipsoidal model, the robust beamforming is finally cast as an SOCP problem. Here we propose a new uncertainty set model, which is given by

$$\mathcal{E} = \{\mathbf{e} \mid \|\mathbf{e}\|_1 \leq \varepsilon_1\} \quad (4.16)$$

where $\|\mathbf{e}\|_1 = \left\| \begin{bmatrix} \mathbf{e}_R^T \\ \mathbf{e}_I^T \end{bmatrix} \right\|_1$ with $\mathbf{e} = \mathbf{e}_R + j\mathbf{e}_I$ being the new ℓ_1 -norm of the complex vector \mathbf{e} and ε_1 describes the size of the uncertainty region. Geometrically, \mathcal{E} is a rhombus.

The robust beamformer minimizes the ℓ_∞ -norm of the output while requiring the ℓ_∞ -modulus of the array response to exceed unity for all $\mathbf{e} \in \mathcal{E}$. This can be written as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \left(\|\mathbf{X}^H \mathbf{w}\|_\infty = \|\bar{\mathbf{X}}^T \bar{\mathbf{w}}\|_\infty \right) \\ \text{s.t.} \quad & |(\mathbf{a} + \mathbf{e})^H \mathbf{w}|_\infty \geq 1, \text{ for all } \mathbf{e} \in \mathcal{E} \end{aligned} \quad (4.17)$$

where the ℓ_∞ -modulus of a complex number is defined in (4.4). However, it is not easy to directly solve (4.17) due to the following two difficulties.

- 1) The constraint is nonconvex.
- 2) There are infinitely many constraints.

In the next subsection, we will reformulate (4.17) into a linear programming that can be solved efficiently.

4.2.3 Solution Based on Linear Programming

We first find a tight lower bound of the ℓ_∞ -modulus of the array response in (4.17). By the triangle inequality in (4.3), we obtain

$$|\mathbf{a}^H \mathbf{w} + \mathbf{e}^H \mathbf{w}|_\infty \geq |\mathbf{a}^H \mathbf{w}|_\infty - |\mathbf{e}^H \mathbf{w}|_\infty. \quad (4.18)$$

The term $\mathbf{e}^H \mathbf{w}$ can be expressed using real-valued variables as

$$\mathbf{e}^H \mathbf{w} = \mathbf{e}_R^T \mathbf{w}_R + \mathbf{e}_I^T \mathbf{w}_I + j(-\mathbf{e}_I^T \mathbf{w}_R + \mathbf{e}_R^T \mathbf{w}_I). \quad (4.19)$$

Therefore,

$$|\mathbf{e}^H \mathbf{w}|_\infty = \left\| \begin{bmatrix} \mathbf{e}_R^T \mathbf{w}_R + \mathbf{e}_I^T \mathbf{w}_I \\ -\mathbf{e}_I^T \mathbf{w}_R + \mathbf{e}_R^T \mathbf{w}_I \end{bmatrix} \right\|_\infty = \|\mathbf{E}^T \bar{\mathbf{w}}\|_\infty \quad (4.20)$$

where

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_R & -\mathbf{e}_I \\ \mathbf{e}_I & \mathbf{e}_R \end{bmatrix} \in \mathbb{R}^{2M \times 2}. \quad (4.21)$$

Applying the matrix norm inequality [68] yields

$$|\mathbf{e}^H \mathbf{w}|_\infty = \|\mathbf{E}^T \bar{\mathbf{w}}\|_\infty \leq \|\mathbf{E}^T\|_\infty \|\bar{\mathbf{w}}\|_\infty = \|\mathbf{E}\|_1 \|\bar{\mathbf{w}}\|_\infty \quad (4.22)$$

where $\|\mathbf{E}^T\|_\infty$ is the maximum row sum matrix norm of \mathbf{E}^T , and $\|\mathbf{E}\|_1$ is the maximum column sum matrix norm of \mathbf{E} [68]. They are related by $\|\mathbf{E}\|_1 = \|\mathbf{E}^T\|_\infty$. We can further compute and bound $\|\mathbf{E}\|_1$ as

$$\begin{aligned} \|\mathbf{E}\|_1 &= \max \left(\left\| \begin{bmatrix} \mathbf{e}_R \\ \mathbf{e}_I \end{bmatrix} \right\|_1, \left\| \begin{bmatrix} -\mathbf{e}_I \\ \mathbf{e}_R \end{bmatrix} \right\|_1 \right) \\ &= \left\| \begin{bmatrix} \mathbf{e}_R \\ \mathbf{e}_I \end{bmatrix} \right\|_1 = \|\bar{\mathbf{e}}\|_1 = \|\mathbf{e}\|_1 \leq \varepsilon_1 \end{aligned} \quad (4.23)$$

where $\bar{\mathbf{e}} = [\mathbf{e}_R^T, \mathbf{e}_I^T]^T \in \mathbb{R}^{2M}$. Using (4.22) and (4.23), we obtain

$$|\mathbf{e}^H \mathbf{w}|_\infty \leq \varepsilon_1 \|\bar{\mathbf{w}}\|_\infty. \quad (4.24)$$

On the other hand, denoting $\mathbf{a} = \mathbf{a}_R + j\mathbf{a}_I$, $|\mathbf{a}^H \mathbf{w}|_\infty$ is bounded from below by

$$\begin{aligned} |\mathbf{a}^H \mathbf{w}|_\infty &= \max(|\operatorname{Re}(\mathbf{a}^H \mathbf{w})|, |\operatorname{Im}(\mathbf{a}^H \mathbf{w})|) \\ &\geq \operatorname{Re}(\mathbf{a}^H \mathbf{w}) = \mathbf{a}_R^T \mathbf{w}_R + \mathbf{a}_I^T \mathbf{w}_I = \bar{\mathbf{a}}^T \bar{\mathbf{w}} \end{aligned} \quad (4.25)$$

where $\bar{\mathbf{a}} = [\mathbf{a}_R^T, \mathbf{a}_I^T]^T$. According to (4.18), (4.24) and (4.25), we have

$$|\mathbf{a}^H \mathbf{w} + \mathbf{e}^H \mathbf{w}|_\infty \geq \bar{\mathbf{a}}^T \bar{\mathbf{w}} - \varepsilon_1 \|\bar{\mathbf{w}}\|_\infty. \quad (4.26)$$

Substituting this inequality into the constraint in (4.17) leads to the following optimization problem:

$$\begin{aligned} \min_{\bar{\mathbf{w}} \in \mathbb{R}^{2M}} \quad & \|\bar{\mathbf{X}}^T \bar{\mathbf{w}}\|_\infty \\ \text{s.t.} \quad & \bar{\mathbf{a}}^T \bar{\mathbf{w}} - \varepsilon_1 \|\bar{\mathbf{w}}\|_\infty \geq 1. \end{aligned} \quad (4.27)$$

which is much simpler than (4.17) since it only has a single constraint. In addition, (4.27) is a convex optimization problem because the objective is a convex function and the constraint is a convex set. Note that the original optimization problem of (4.17) gives the physical meanings of the proposed beamformer. However, it is difficult to directly solve such a nonconvex problem with infinitely many constraints. Hence we finally relax it into a convex one with a single constraint in (4.27). It should also be pointed out that (4.27) is not equivalent to the original optimization problem of (4.17) since the two inequalities of (4.24) and (4.25) cannot simultaneously be tight

in general. It is clear that the constraint of (4.27) is stricter than that of (4.17). That is, the feasible region of (4.17) is larger than that of (4.27). Therefore any feasible solutions of (4.27) is feasible for (4.17). This means that the robustness of the beamformer given by (4.27) is guaranteed. In addition, the original problem of (4.17) is nonconvex and difficult to handle but (4.27) is convex and hence easier to be solved.

Furthermore, by introducing two auxiliary variables $u, r \in \mathbb{R}$, (4.27) can be converted into an LP problem as follows:

$$\begin{aligned}
& \min_{\bar{\mathbf{w}}, u, r} u \\
& \text{s.t.} \quad -u\mathbf{1}_{2N} \leq \bar{\mathbf{X}}^T \bar{\mathbf{w}} \leq u\mathbf{1}_{2N} \\
& \quad \quad \bar{\mathbf{a}}^T \bar{\mathbf{w}} \geq \varepsilon_1 r + 1 \\
& \quad \quad -r\mathbf{1}_{2M} \leq \bar{\mathbf{w}} \leq r\mathbf{1}_{2M}.
\end{aligned} \tag{4.28}$$

Therefore, the proposed beamforming method is referred to as the RLPB. The LP problem in (4.28) is compactly written in a standard form as

$$\begin{aligned}
& \min_{\bar{\mathbf{w}}, u, r} [\mathbf{0}_{2M}^T, 1, 0] \begin{bmatrix} \bar{\mathbf{w}} \\ u \\ r \end{bmatrix} \\
& \text{s.t.} \quad \begin{bmatrix} \bar{\mathbf{X}}^T & -\mathbf{1}_{2N} & \mathbf{0}_{2N} \\ -\bar{\mathbf{X}}^T & -\mathbf{1}_{2N} & \mathbf{0}_{2N} \\ \mathbf{I}_{2M \times 2M} & \mathbf{0}_{2M} & -\mathbf{1}_{2M} \\ -\mathbf{I}_{2M \times 2M} & \mathbf{0}_{2M} & -\mathbf{1}_{2M} \\ -\bar{\mathbf{a}}^T & 0 & \varepsilon_1 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{w}} \\ u \\ r \end{bmatrix} \leq \begin{bmatrix} \mathbf{0}_{2N} \\ \mathbf{0}_{2N} \\ \mathbf{0}_{2M} \\ \mathbf{0}_{2M} \\ -1 \end{bmatrix}.
\end{aligned} \tag{4.29}$$

The problem of (4.29) can be efficiently solved using the simplex method [72] or interior point method [55]. The computational complexity per iteration using the primal-dual interior point method [55] to solve the LP problem is much less than that required to solve SOCP and SDP problems with the similar size [73, 74]. Therefore, compared with the SOCP- or SDP-based robust beamforming approaches [18, 28], the proposed RLPB is computationally simpler. It should be pointed out that the RMVB [24] and RCB [20] can be calculated using the Lagrangian multiplier technique instead of directly solving an SOCP. It just requires the solution to a low complexity nonlinear equation, which can be faster than the RLPB.

Remark 1: There are two reasons why the proposed beamformer makes use of rhombic uncertainty set rather than the conventional sphere or ellipsoid. First, it is a natural result of the fact that the ℓ_∞ -modulus of the array response is used as the constraint instead of the conventional modulus. From (4.18) – (4.24), it can be seen that the ℓ_∞ -modulus constraint elicits the rhombic region $\|\mathbf{e}\|_1 \leq \varepsilon_1$, where the relation $\|\mathbf{E}^T\|_\infty = \|\mathbf{E}\|_1 = \|\mathbf{e}\|_1$ plays the dominant role in converting the ℓ_∞ -norm into ℓ_1 -norm. Second, due to the rhombic uncertainty region, we can cast the optimization problem into an LP. The use of spherical or ellipsoidal uncertainty set will lead to an SOCP rather than LP.

Remark 2: The proposed RLPB utilizes the property of non-Gaussianity or finite alphabet of the signals. This property has also been widely used in blind equalization. One of the most representative blind equalization techniques is the constant modulus algorithm (CMA) [37, 69]. It is of interest to discuss the difference between the objective functions of the RLPB and CMA. The CMA solves the following minimization

problem:

$$\min \mathbb{E} \left\{ (|y(n)|^2 - \kappa)^2 \right\} \quad (4.30)$$

where $\kappa > 0$ is the dispersion constant [37] defined as

$$\kappa = \frac{\mathbb{E} \{|s(n)|^4\}}{\mathbb{E} \{|s(n)|^2\}}. \quad (4.31)$$

Different constellations may have different dispersion constants. Expanding (4.30) and replacing the expectation by the sample mean, the CMA is equivalent to minimizing

$$f_{\text{CMA}}(\mathbf{w}) = \|\mathbf{y}\|_{\ell_4}^4 - 2\kappa\|\mathbf{y}\|^2 = \|\mathbf{X}^H\mathbf{w}\|_{\ell_4}^4 - 2\kappa\|\mathbf{X}^H\mathbf{w}\|^2 \quad (4.32)$$

where $\|\mathbf{y}\|_{\ell_4}$ is the ℓ_4 -norm based on the conventional modulus. Therefore the CMA minimizes a combination of the ℓ_4 -norm and ℓ_2 -norm, which exploits both the fourth- and second-order statistics. The objective function of the CMA is different from the newly defined ℓ_∞ -norm adopted by the RLPB, which is based on the ℓ_∞ -modulus. Note that $f_{\text{CMA}}(\mathbf{w})$ is nonconvex and it is difficult to find the global minimum. This constitutes a drawback of the CMA.

4.2.4 Selection of Uncertainty Set Size ε_1

The size of the uncertainty region ε_1 is a crucial parameter of the proposed RLPB. Larger values of ε_1 means that the beamformer can tolerate larger steering vector mismatches. However, interference suppression capability will be weakened if ε_1 is too large. Therefore, it is important to select an appropriate value for ε_1 for the proposed RLPB according to the level of steering vector mismatch.

First we assign a maximal possible value that ε_1 can take. To guarantee the

feasibility of (4.27), one requires

$$|\bar{\mathbf{a}}^T \bar{\mathbf{w}}| \geq \bar{\mathbf{a}}^T \bar{\mathbf{w}} \geq \varepsilon_1 \|\bar{\mathbf{w}}\|_\infty + 1. \quad (4.33)$$

On the other hand, by Hölder inequality² [77], we have

$$|\bar{\mathbf{a}}^T \bar{\mathbf{w}}| \leq \|\bar{\mathbf{a}}\|_1 \|\bar{\mathbf{w}}\|_\infty. \quad (4.34)$$

Therefore,

$$\varepsilon_1 \|\bar{\mathbf{w}}\|_\infty + 1 \leq \|\bar{\mathbf{a}}\|_1 \|\bar{\mathbf{w}}\|_\infty \quad (4.35)$$

which yields

$$\varepsilon_1 \leq \|\bar{\mathbf{a}}\|_1 - \frac{1}{\|\bar{\mathbf{w}}\|_\infty} < \|\bar{\mathbf{a}}\|_1 = \|\mathbf{a}\|_1. \quad (4.36)$$

To obtain a solution for (4.27), we require $\varepsilon_1 < \|\mathbf{a}\|_1$. Otherwise the optimization problem of (4.27) will be infeasible. Although $\varepsilon_1 < \|\mathbf{a}\|_1$ gives a general guideline for selecting the parameter ε_1 with any type of steering vector mismatch, $\|\mathbf{a}\|_1$ is only a coarse bound of ε_1 and it is always too large for an appropriate ε_1 . If some *a priori* information on the cause of the steering vector uncertainty is available, a better ε_1 can be selected by exploiting this *a priori* knowledge.

As mentioned in Section 4.1, there are many causes giving rise to steering vector mismatch. Here we consider two types of steering vector errors. The first error type is AOA mismatch, where the steering vector error \mathbf{e} is parameterized by the AOA θ , denoted by $\mathbf{e}(\theta) = \mathbf{c}(\theta) - \mathbf{a}$ with \mathbf{a} the assumed steering vector and $\mathbf{c}(\theta)$ the actual steering vector. If the beamformer performance is guaranteed for all $\theta \in [\theta_l, \theta_u]$, ε_1

²The Hölder inequality $|\bar{\mathbf{a}}^T \bar{\mathbf{w}}| \leq \|\bar{\mathbf{a}}\|_p \|\bar{\mathbf{w}}\|_q$ satisfies for all $1/p + 1/q = 1$ with $1 \leq p, q \leq \infty$ [77]. In particular, it reduces to $|\bar{\mathbf{a}}^T \bar{\mathbf{w}}| \leq \|\bar{\mathbf{a}}\|_1 \|\bar{\mathbf{w}}\|_\infty$ for $p = 1$ and $q = \infty$.

can be determined by

$$\varepsilon_1 = \max_{\theta_l \leq \theta \leq \theta_u} \|\mathbf{c}(\theta) - \mathbf{a}\|_1. \quad (4.37)$$

This maximum is easily computed in advance using grid search. If \mathbf{a} and $\mathbf{c}(\theta)$ are not far away from each other, the maximum usually appears at the boundary θ_l or θ_u .

The second error type is random mismatch, where the steering vector \mathbf{e} is modeled as zero-mean circular Gaussian distributed variables with variance $2\sigma_e^2$, that is,

$$\mathbf{e} \sim \mathcal{CN}(\mathbf{0}, 2\sigma_e^2 \mathbf{I}_M). \quad (4.38)$$

By the central limit theorem [71], the Gaussian distribution can model the arbitrary random model errors well. The real-valued variables are distributed as $\bar{\mathbf{e}} = [\mathbf{e}_R^T, \mathbf{e}_I^T]^T \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{2M})$. The quantity

$$\text{Ptb} = \frac{\text{tr}(2\sigma_e^2 \mathbf{I}_M)}{\|\mathbf{a}\|^2} = 2\sigma_e^2 \quad (4.39)$$

defines the relative perturbation of the steering vector, which is exploited from the fact of $\|\mathbf{a}\|^2 = M$. The following proposition guides us to the selection of ε_2 for the RMVB and ε_1 for the RLPB such that the steering vector error \mathbf{e} will be within the uncertainty region with a high probability.

Proposition: For a given probability $P_\varepsilon \in [0, 1]$, let

$$\varepsilon_2 = \sqrt{\tau} \sigma_e, \text{ and } \varepsilon_1 = \sqrt{2M\tau} \sigma_e \quad (4.40)$$

with

$$\tau = F_{\chi^2}^{-1}(P_\varepsilon, 2M) \quad (4.41)$$

where $F_{\chi^2}^{-1}(P_\varepsilon, 2M)$ denotes the inverse of the cumulative distribution function (CDF) of the χ^2 -distribution of degree $2M$. Then we have

$$\Pr(\|\mathbf{e}\| \leq \varepsilon_2) = P_\varepsilon \quad (4.42)$$

and

$$\Pr(\|\mathbf{e}\|_1 \leq \varepsilon_1) \geq P_\varepsilon. \quad (4.43)$$

Proof: Since each component of $\bar{\mathbf{e}} = [\bar{e}_1, \dots, \bar{e}_{2M}]^T$ is i.i.d. Gaussian-distributed, the following quantity satisfies the χ^2 -distribution of degree $2M$ [78]:

$$\frac{\|\bar{\mathbf{e}}\|^2}{\sigma_e^2} = \sum_{i=1}^{2M} \bar{e}_i^2 \sim \chi^2(2M). \quad (4.44)$$

The CDF of the χ^2 -distribution with degree $2M$ is [78]

$$F_{\chi^2}(\beta, 2M) = \frac{\gamma(M, \beta/2)}{\Gamma(M)} \quad (4.45)$$

where $\Gamma(\cdot)$ is the Gamma function and $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function [79]. The following probability

$$P_\varepsilon = \Pr(\|\bar{\mathbf{e}}\| \leq \varepsilon_2) = \Pr\left(\frac{\|\bar{\mathbf{e}}\|^2}{\sigma_e^2} \leq \frac{\varepsilon_2^2}{\sigma_e^2}\right) = F_{\chi^2}\left(\frac{\varepsilon_2^2}{\sigma_e^2}, 2M\right) \quad (4.46)$$

yields the relation

$$\frac{\varepsilon_2^2}{\sigma_e^2} = F_{\chi^2}^{-1}(P_\varepsilon, 2M) = \tau. \quad (4.47)$$

Hence we can conclude that $\varepsilon_2 = \sqrt{\tau}\sigma_e$. Using the norm inequality $\|\bar{\mathbf{e}}\|_1 \leq \sqrt{2M}\|\bar{\mathbf{e}}\|$,

if $\varepsilon_1 = \sqrt{2M}\varepsilon_2 = \sqrt{2M}\tau\sigma_e$, we have

$$\begin{aligned} \Pr(\|\bar{\mathbf{e}}\|_1 \leq \varepsilon_1) &\geq \Pr\left(\sqrt{2M}\|\bar{\mathbf{e}}\| \leq \varepsilon_1\right) \\ &= \Pr\left(\sqrt{2M}\|\bar{\mathbf{e}}\| \leq \sqrt{2M}\varepsilon_2\right) \\ &= \Pr(\|\bar{\mathbf{e}}\| \leq \varepsilon_2) = P_\varepsilon. \end{aligned} \quad (4.48)$$

□

The inverse function of the CDF $F_{\chi^2}^{-1}(P_\varepsilon, 2M)$ is a standard function in statistics. Its value, which can be computed numerically, has been provided in standard percentile tables [78]. For example, $\tau = F_{\chi^2}^{-1}(P_\varepsilon, 2M) = 31.41$ for $P_\varepsilon = 95\%$ and $M = 10$.

Remark 3: From the simulation results in Section 4.4.3, we find that the RLPB is not sensitive to the selection of ε_1 . That is, it can work well for a wide range of ε_1 . This advantage makes the selection of an appropriate ε_1 rather flexible.

4.2.5 Performance Measures

The output SINR of (1.4) has been widely used as a performance measure of a beamformer. The upper bound of the SINR is the maximum eigenvalue of the matrix $\sigma_s^2 \mathbf{R}_{i+n}^{-1} \mathbf{a} \mathbf{a}^H$. The goal of beamforming is to recover the desired signal while mitigating interferences and noise. Therefore, it is expected that the beamformer output $y(n)$ recovers the desired signal $s(n)$. Note that there is an inherent complex-valued ambiguity factor α between $s(n)$ and $y(n)$. The normalized correlation coefficient can measure the similarity of \mathbf{y} and $\mathbf{s} = [s(1), \dots, s(N)]^T$ well, which is defined as

$$\rho(\mathbf{s}, \mathbf{y}) = \frac{|\mathbf{s}^H \mathbf{y}|}{\|\mathbf{s}\| \|\mathbf{y}\|}. \quad (4.49)$$

By Cauchy-Schwartz inequality, the correlation coefficient satisfies $0 \leq \rho(\mathbf{s}, \mathbf{y}) \leq 1$. In addition to the correlation coefficient, the following mean square error (MSE)

$$\text{MSE} = \min_{\alpha} \|\alpha \mathbf{y} - \mathbf{s}\|^2 \quad (4.50)$$

also describes the similarity, where α is the ambiguity factor. The minimizer of (4.50) is $\alpha = (\mathbf{y}^H \mathbf{s}) / \|\mathbf{y}\|^2$. Substituting this minimizer back into (4.50), the MSE is obtained as

$$\text{MSE} = \|\mathbf{s}\|^2 - \frac{|\mathbf{s}^H \mathbf{y}|^2}{\|\mathbf{y}\|^2}. \quad (4.51)$$

The normalized mean square error (NMSE) is calculated as

$$\text{NMSE} = \frac{\text{MSE}}{\|\mathbf{s}\|^2} = 1 - \frac{|\mathbf{s}^H \mathbf{y}|^2}{\|\mathbf{s}\|^2 \|\mathbf{y}\|^2} = 1 - \rho^2(\mathbf{s}, \mathbf{y}) \quad (4.52)$$

which reveals the relationship between the NMSE and correlation coefficient. For different beamformers, the corresponding NMSE can be calculated using (4.52) to compare their performance.

We derive a lower bound of the NMSE. Recalling that $\mathbf{y}^* = \mathbf{X}^H \mathbf{w}$, the quantity

$$\frac{|\mathbf{s}^H \mathbf{y}|^2}{\|\mathbf{y}\|^2} = \frac{\mathbf{w}^H (\mathbf{X} \mathbf{s}^* \mathbf{s}^T \mathbf{X}^H) \mathbf{w}}{\mathbf{w}^H (\mathbf{X} \mathbf{X}^H) \mathbf{w}} \quad (4.53)$$

is a generalized Rayleigh quotient [80]. Its maximum equals the largest eigenvalue of the matrix $(\mathbf{X} \mathbf{X}^H)^{-1} (\mathbf{X} \mathbf{s}^* \mathbf{s}^T \mathbf{X}^H)$ [80]. Since $\text{rank}((\mathbf{X} \mathbf{X}^H)^{-1} \mathbf{X} \mathbf{s}^* \mathbf{s}^T \mathbf{X}^H) = 1$, the maximum eigenvalue is its unique non-zero eigenvalue and equals its trace [68]. That

is,

$$\begin{aligned}
\max \frac{|\mathbf{s}^H \mathbf{y}|^2}{\|\mathbf{y}\|^2} &= \lambda_{\max} \left((\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}\mathbf{s}^* \mathbf{s}^T \mathbf{X}^H \right) \\
&= \text{tr} \left((\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}\mathbf{s}^* \mathbf{s}^T \mathbf{X}^H \right) \\
&= \text{tr} \left(\mathbf{s}^T \mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}\mathbf{s}^* \right) \\
&= \mathbf{s}^T \mathbf{P}_\mathbf{X} \mathbf{s}^*
\end{aligned} \tag{4.54}$$

where $\mathbf{P}_\mathbf{X} = \mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}$ is the projection matrix onto the range space of \mathbf{X}^H .

The lower bound of the NMSE is then given by

$$\text{NMSE}_{\text{opt}} = 1 - \frac{\mathbf{s}^T \mathbf{P}_\mathbf{X} \mathbf{s}^*}{\|\mathbf{s}\|^2}. \tag{4.55}$$

We further analyze the asymptotic expression of the NMSE lower bound as $N \rightarrow \infty$.

Since the signals in practical applications are ergodic, we can derive that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}\mathbf{X}^H = \mathbf{R}, \quad \lim_{N \rightarrow \infty} \frac{1}{N} \|\mathbf{s}\|^2 = \sigma_s^2 \tag{4.56}$$

and

$$\begin{aligned}
\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}\mathbf{s}^* &= \text{E} \{ \mathbf{x}(n) s^*(n) \} \\
&= \text{E} \left\{ \left(s(n) \mathbf{a} + \sum_{i=1}^I s_i(n) \mathbf{a}_i + \mathbf{v}(n) \right) s^*(n) \right\} \\
&= \text{E} \{ |s(n)|^2 \} \mathbf{a} = \sigma_s^2 \mathbf{a}
\end{aligned} \tag{4.57}$$

where we have used the fact that the desired signal is uncorrelated with the interferences and noise. Substituting (4.56) and (4.57) into (4.55), the asymptotic optimal NMSE is obtained as

$$\lim_{N \rightarrow \infty} \text{NMSE}_{\text{opt}} = 1 - \sigma_s^2 \mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}. \tag{4.58}$$

Recall that the upper bound of the output SINR is the maximum eigenvalue of $\mathbf{R}_{i+n}^{-1}\mathbf{R}_s$, where $\mathbf{R}_s \triangleq \sigma_s^2\mathbf{a}\mathbf{a}^H$. Since $\text{rank}(\mathbf{R}_{i+n}^{-1}\mathbf{R}_s) = 1$, its maximum eigenvalue is its unique non-zero eigenvalue and equals its trace. Hence, the upper bound of SINR can be simplified as

$$\text{SINR}_{\text{opt}} = \sigma_s^2\mathbf{a}^H\mathbf{R}_{i+n}^{-1}\mathbf{a}. \quad (4.59)$$

Using the fact that $\mathbf{R} = \mathbf{R}_s + \mathbf{R}_{i+n}$ and the matrix inversion lemma [68], we find the relationship between the optimal NMSE and SINR:

$$\lim_{N \rightarrow \infty} \text{NMSE}_{\text{opt}} = \frac{1}{1 + \text{SINR}_{\text{opt}}}. \quad (4.60)$$

Although the upper bound of SINR and the lower bound of the NMSE is related by (4.60), it should be pointed out that this relationship is valid only when the number of snapshots is infinite. In addition, the output SINR and NMSE of a beamformer do not have the relationship in (4.60) if it is not the optimal one. Therefore the output SINR and NMSE are two independent performance measures.

4.3 Analysis From Statistical Viewpoint

The RLPB is based on the minimization of the ℓ_∞ -norm of the output. This enables it to implicitly exploit higher-order statistics. Therefore the performance of the RLPB is improved compared with the MV beamformers when it is applied to sub-Gaussian signals. In this section, we will give a theoretical explanation to this point from a statistical perspective. The signal model of (1.1) is compactly rewritten as

$$\mathbf{x}(n) = \mathbf{A}\mathbf{b}(n) + \mathbf{v}(n) \quad (4.61)$$

where the array manifold matrix $\mathbf{A} = [\mathbf{a}, \mathbf{a}_1, \dots, \mathbf{a}_I]$ contains all steering vectors, $\mathbf{b}(n) = [s(n), s_1(n), \dots, s_I(n)]^T$ is the vector consisting of the desired signal and interferences. For convenience, we also denote $\mathbf{b}(n) = [b^1(n), \dots, b^{I+1}(n)]^T$. That is, the i th ($1 \leq i \leq I + 1$) element of $\mathbf{b}(n)$ is denoted as $b^i(n) = b_R^i(n) + \mathbf{j}b_I^i(n)$ with $b_R^i(n)$ and $b_I^i(n)$ representing the real and imaginary parts of $b^i(n)$, respectively. Then, it can be expressed as $\mathbf{b}(n) = \mathbf{b}_R(n) + \mathbf{j}\mathbf{b}_I(n)$. According to (1.3) and (4.61), the beamformer output is expressed as

$$y(n) = \mathbf{w}^H \mathbf{A} \mathbf{b}(n) + \mathbf{w}^H \mathbf{v}(n) = \mathbf{g}^H \mathbf{b}(n) + \nu(n) \quad (4.62)$$

where $\mathbf{g} = \mathbf{A}^H \mathbf{w} \in \mathbb{C}^{I+1}$ denotes the overall response of the array and beamformer, and $\nu(n)$ is the noise component after beamforming. Denoting $\mathbf{g} = \mathbf{g}_R + \mathbf{j}\mathbf{g}_I$, one gets

$$\begin{aligned} y_R(n) &= \mathbf{g}_R^T \mathbf{b}_R(n) + \mathbf{g}_I^T \mathbf{b}_I(n) + \nu_R(n) \\ &= \bar{\mathbf{g}}^T \bar{\mathbf{b}}(n) = \sum_{i=1}^{2I+3} \bar{g}_i \bar{b}_i(n) \end{aligned} \quad (4.63)$$

and

$$\begin{aligned} y_I(n) &= \mathbf{g}_R^T \mathbf{b}_I(n) - \mathbf{g}_I^T \mathbf{b}_R(n) + \nu_I(n) \\ &= \bar{\bar{\mathbf{g}}}^T \bar{\bar{\mathbf{b}}}(n) = \sum_{i=1}^{2I+3} \bar{\bar{g}}_i \bar{\bar{b}}_i(n) \end{aligned} \quad (4.64)$$

where $\bar{\mathbf{g}} = [\mathbf{g}_R^T, \mathbf{g}_I^T, 1]^T$, $\bar{\mathbf{b}}(n) = [\mathbf{b}_R^T(n), \mathbf{b}_I^T(n), \nu_R(n)]^T$, $\bar{\bar{\mathbf{g}}} = [-\mathbf{g}_I^T, \mathbf{g}_R^T, 1]^T$, and $\bar{\bar{\mathbf{b}}}(n) = [\mathbf{b}_R^T(n), \mathbf{b}_I^T(n), \nu_I(n)]^T$, with \bar{g}_i , $\bar{\bar{g}}_i$, $\bar{b}_i(n)$, and $\bar{\bar{b}}_i(n)$ being the corresponding i th elements, respectively. Note that $\nu_R(n)$ and $\nu_I(n)$ have been absorbed into $\bar{\mathbf{b}}(n)$ and $\bar{\bar{\mathbf{b}}}(n)$, respectively. In the following, we omit the time index n for simplicity, i.e., $\bar{\mathbf{b}}(n)$ and $\bar{\bar{\mathbf{b}}}(n)$ are abbreviated as $\bar{\mathbf{b}}$ and $\bar{\bar{\mathbf{b}}}$, respectively. In addition, we define $L = 2I + 3$.

Consider the following objective function

$$f_p(\mathbf{b}, \nu) = \mathbb{E}\{|y(n)|^p\} = \mathbb{E}\{|y_R(n)|^p\} + \mathbb{E}\{|y_I(n)|^p\}. \quad (4.65)$$

Note that the expectation is over the random variables \mathbf{b} and ν . Hence f_p is a function of \mathbf{b} and ν . Replacing the expectation by the sample mean based on the N available snapshots and letting $p \rightarrow \infty$, the minimization of $f_p(\mathbf{b}, \nu)$ reduces to minimizing $\|\mathbf{y}\|_\infty$, which is the objective function of the RLPB. Now we analyze $f_p(\mathbf{b}, \nu)$. The result is then applicable to RLPB as $p \rightarrow \infty$. First we consider the multivariate function

$$f_R^{(p)}(\bar{\mathbf{b}}) = |y_R(n)|^p = \left| \sum_{i=1}^L \bar{g}_i \bar{b}_i \right|^p. \quad (4.66)$$

Employing a Taylor series expansion of multiple variables $\bar{\mathbf{b}}$ around zeros, $f_R^{(p)}(\bar{\mathbf{b}})$ can be approximated by

$$\begin{aligned} f_R^{(p)}(\bar{\mathbf{b}}) \approx & \sum_{i_1=1}^L \frac{\partial f_R^{(p)}}{\partial \bar{b}_{i_1}} \Big|_{\bar{\mathbf{b}}=\mathbf{0}} \bar{b}_{i_1} + \frac{1}{2!} \sum_{i_1=1}^L \sum_{i_2=1}^L \frac{\partial^2 f_R^{(p)}}{\partial \bar{b}_{i_1} \partial \bar{b}_{i_2}} \Big|_{\bar{\mathbf{b}}=\mathbf{0}} \bar{b}_{i_1} \bar{b}_{i_2} \\ & + \cdots + \frac{1}{K!} \sum_{i_1=1}^L \cdots \sum_{i_K=1}^L \frac{\partial^K f_R^{(p)}}{\partial \bar{b}_{i_1} \cdots \partial \bar{b}_{i_K}} \Big|_{\bar{\mathbf{b}}=\mathbf{0}} \bar{b}_{i_1} \cdots \bar{b}_{i_K} \end{aligned} \quad (4.67)$$

where K is the approximation order and we have used the fact $f_R^{(p)}(\mathbf{0}) = \mathbf{0}$. The value of K required for achieving a high accuracy approximation depends on p . When p is an even number, the function can be exactly represented with $K = p$. For example, with $p = 2$, $f_R^{(2)}(\bar{\mathbf{b}}) = \sum_{i_1, i_2} \bar{g}_{i_1} \bar{g}_{i_2} \bar{b}_{i_1} \bar{b}_{i_2}$ is just a quadratic form and $K = 2$ is enough in (4.67). In general, a larger p requires a higher approximation order K . For the ℓ_∞ -norm adopted by the RLPB, the desired value of K approaches infinity as $p \rightarrow \infty$.

$$\mathbb{E} \{ \bar{b}_{i_1} \cdots \bar{b}_{i_k} \} = \begin{cases} \mathbb{E} \{ \bar{b}_i^k \}, & \text{if } i_1 = \cdots = i_k = i \\ \mathbb{E} \{ \bar{b}_{i_1}^2 \} \mathbb{E} \{ \bar{b}_{i_3}^2 \} \cdots \mathbb{E} \{ \bar{b}_{i_{k-1}}^2 \}, & \text{if } i_1 = i_2, i_3 = i_4, \cdots, i_{k-1} = i_k \\ \mathbb{E} \{ \bar{b}_{i_1}^4 \} \mathbb{E} \{ \bar{b}_{i_5}^4 \} \cdots \mathbb{E} \{ \bar{b}_{i_{k-3}}^4 \}, & \text{if } i_1 = \cdots = i_4, \cdots, i_{k-3} = \cdots = i_k \\ \vdots & \\ 0, & \text{if there is an } i_j \neq i_l, 1 \leq l \leq L, l \neq j \end{cases} \quad (4.69)$$

Therefore, we have

$$\begin{aligned} & \mathbb{E} \{ f_R^{(\infty)}(\bar{\mathbf{b}}) \} \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i_1=1}^L \cdots \sum_{i_k=1}^L \left. \frac{\partial^k f_R^{(p)}}{\partial \bar{b}_{i_1} \cdots \partial \bar{b}_{i_k}} \right|_{\bar{\mathbf{b}}=\mathbf{0}} \mathbb{E} \{ \bar{b}_{i_1} \cdots \bar{b}_{i_k} \}. \end{aligned} \quad (4.68)$$

Recall that all signals and noise components are zero-mean and mutually independent. In addition, we assume that the real and imaginary parts of each signal are also independent. Hence all the elements of $\bar{\mathbf{b}}$ are independent with one another. The $\mathbb{E} \{ \bar{b}_{i_1} \cdots \bar{b}_{i_k} \}$ takes different values for different $\{i_1, \cdots, i_k\}$. For example, it equals zero if k is an odd number. For the case that k is an even number, its representative values are shown in (4.69), which includes the k th-order moment of \bar{b}_i , i.e., $\mathbb{E} \{ \bar{b}_i^k \}$. Recall that \bar{b}_i denotes the real or imaginary part of the desired signal, interference or noise, which means that $\mathbb{E} \{ \bar{b}_i^k \}$ represents higher-order statistics of the signal and interferences. Furthermore, (4.68) contains the products of the even-order moments of $\{ \bar{b}_i \}_{i=1}^L$.

Similarly, we obtain the expectation of $f_I^{(p)}(\bar{\mathbf{b}}) = |y_I(n)|^p$ as $p \rightarrow \infty$ as

$$\begin{aligned} & \mathbb{E} \left\{ f_I^{(\infty)}(\bar{\mathbf{b}}) \right\} \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i_1=1}^L \cdots \sum_{i_k=1}^L \left. \frac{\partial^k f_I^{(p)}}{\partial \bar{b}_{i_1} \cdots \partial \bar{b}_{i_k}} \right|_{\bar{\mathbf{b}}=\mathbf{0}} \mathbb{E} \left\{ \bar{b}_{i_1} \cdots \bar{b}_{i_k} \right\} \end{aligned} \quad (4.70)$$

where $\mathbb{E} \left\{ \bar{b}_{i_1} \cdots \bar{b}_{i_k} \right\}$ has a similar structure as (4.69), which contains the k th-order moment $\mathbb{E} \left\{ \bar{b}_i^k \right\}$ as well as the products of the even-order moments of $\left\{ \bar{b}_i \right\}_{i=1}^L$.

According to (4.65), it is known that the RLPB minimizes the sum of (4.68) and (4.70), which includes all k th-order ($1 \leq k \leq \infty$) statistics. The contribution of the higher-order moments will decrease as the order k increases due to the attenuation factor of $1/k!$. Thus, it is clear that the RLPB implicitly exploits the higher-order statistics of the desired signal, interferences and noise. When the signal and interferences are non-Gaussian, these higher-order statistics that contain useful statistical information will substantially improve the performance of the RLPB. When $p = 2$, all higher-statistics vanish and only the second-order moments remain because the higher-order terms in (4.67) become zero. Therefore, $p = 2$ corresponding to the MV-based beamformers only utilizes the second-order statistics. The performance of such second-order statistic-based beamformers is inferior to that of the RLPB for sub-Gaussian signals.

4.4 Simulation Results

We follow similar experimental settings as in [24] and [26] in the simulations. A uniform linear array (ULA) of $M = 10$ omnidirectional sensors with a half-wavelength

spacing is considered. For ULA, the steering vector has the following form:

$$\mathbf{a}(\theta) = [1, e^{j(2\pi/\zeta)d\sin\theta}, \dots, e^{j(M-1)(2\pi/\zeta)d\sin\theta}]^T \quad (4.71)$$

where θ is the AOA, d is the inter-sensor spacing, and ζ is the wavelength. Three zero-mean sub-Gaussian signals, namely, the desired source $s(n)$ and two uncorrelated interferences $s_1(n)$ and $s_2(n)$, impinge on the array. Unless stated otherwise, the AOA of the desired signal is $\theta = 43^\circ$ and the AOAs of the two interferences are $\theta_1 = 30^\circ$ and $\theta_2 = 75^\circ$. We consider applications in communications because most of the corresponding signals are sub-Gaussian. We take different phase shift keying (PSK), quadrature amplitude modulation (QAM), frequency modulation (FM), and phase modulation (PM) schemes as example. The SNR is defined as

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_v^2} \quad (4.72)$$

where σ_s^2 and σ_v^2 are the variances of the desired signal and additive noise, respectively. The two interferences are stronger than the desired signal with variances being $\sigma_1^2 = 4\sigma_s^2$ and $\sigma_2^2 = 9\sigma_s^2$. That is, they are 6 dB and 9.5 dB above the desired signal, respectively. The output SINRs and NMSEs of six beamformers, namely, the MVDR [11], LCMV [14], subspace [15], RMVB [18, 24], general-rank [21], and the RLPB, are compared. The upper bound of the SINR and the lower bound of NMSE are also provided for comparison. Since the subspace beamformer requires the dimension of the signal-plus-interference subspace, the minimum description length (MDL) principle [48] is adopted to estimate this quantity. When plotting the SINR and NMSE curves, 200 Monte Carlo trials are performed for their computation.

4.4.1 AOA Mismatch

We first consider the steering vector mismatch induced by AOA estimation errors. The true AOA of the desired signal is 43° but the assumed AOA is 45° . The *a priori* maximum AOA deviation for all beamformers is assigned as 3° . Hence for the LCMV beamformer, two linear constraints force the responses of the signals from 42° and 48° to be unity. For fair comparison, all beamforming methods use the same *a priori* information on the maximum possible AOA mismatch. Therefore the algorithm parameter for the RMVB is determined as [26]

$$\varepsilon_2 = \max_{42^\circ \leq \theta \leq 48^\circ} \|\mathbf{a}(\theta) - \mathbf{a}(45^\circ)\|_2 = 1.95 \quad (4.73)$$

while for the RLPB

$$\varepsilon_1 = \max_{42^\circ \leq \theta \leq 48^\circ} \|\mathbf{a}(\theta) - \mathbf{a}(45^\circ)\|_1 = 6.65 \quad (4.74)$$

and for the general-rank beamformer

$$\varepsilon_F = \max_{42^\circ \leq \theta \leq 48^\circ} \|\mathbf{a}(\theta)\mathbf{a}^H(\theta) - \mathbf{a}(45^\circ)\mathbf{a}^H(45^\circ)\|_F = 4.73 \quad (4.75)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

In the first simulation, the desired signal and two interferences are 16-QAM signals, while the additive noise is a white Gaussian process. Fig. 4.1 shows the constellations of the outputs of the six beamformers at SNR = 25 dB and number of snapshots $N = 200$. The MVDR beamformer fails to recover the desired signal in the presence of 2° AOA mismatch. The LCMV beamformer also exhibits limited capability for

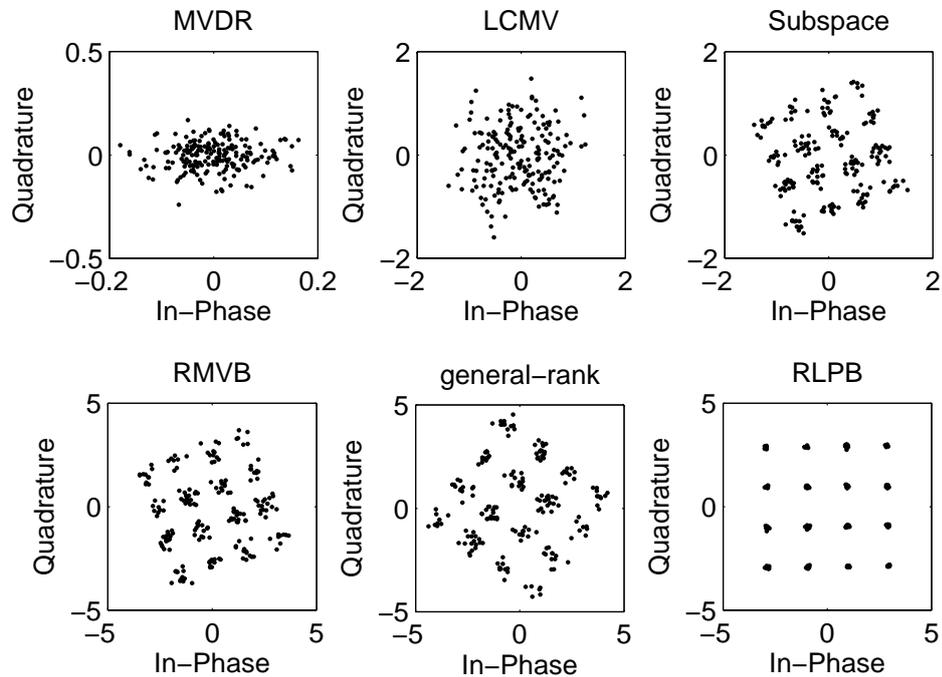
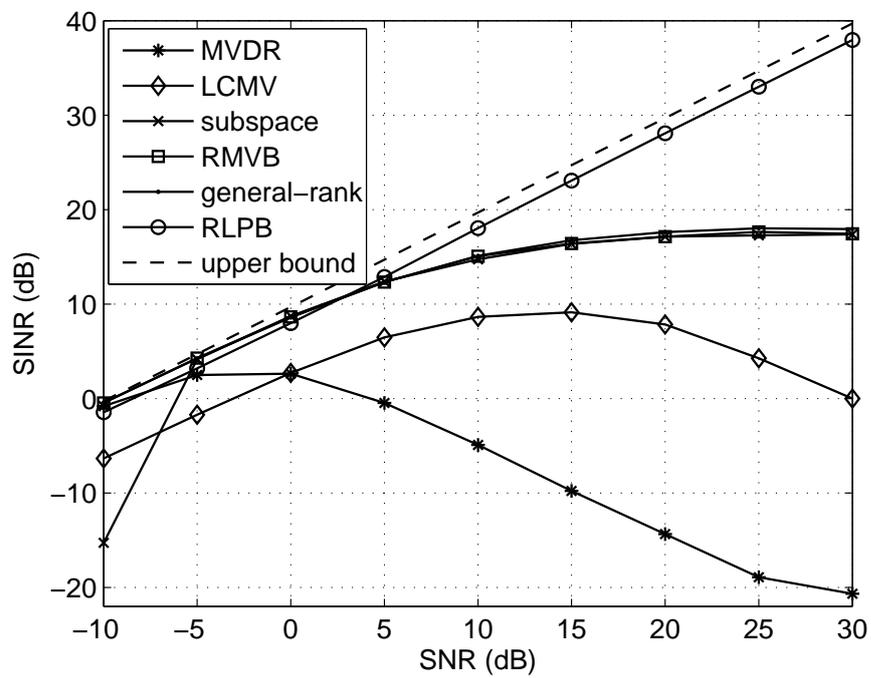
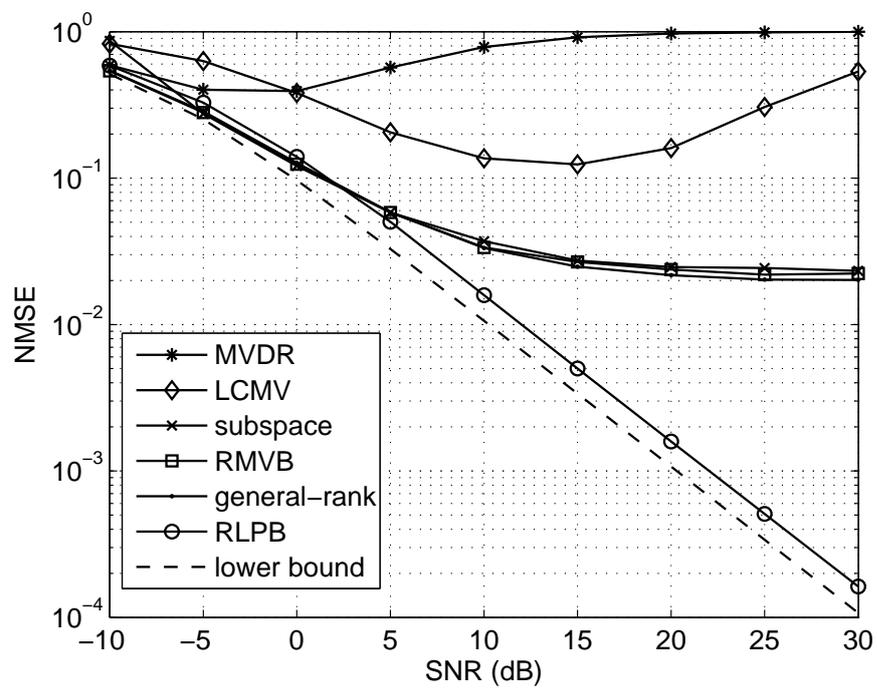


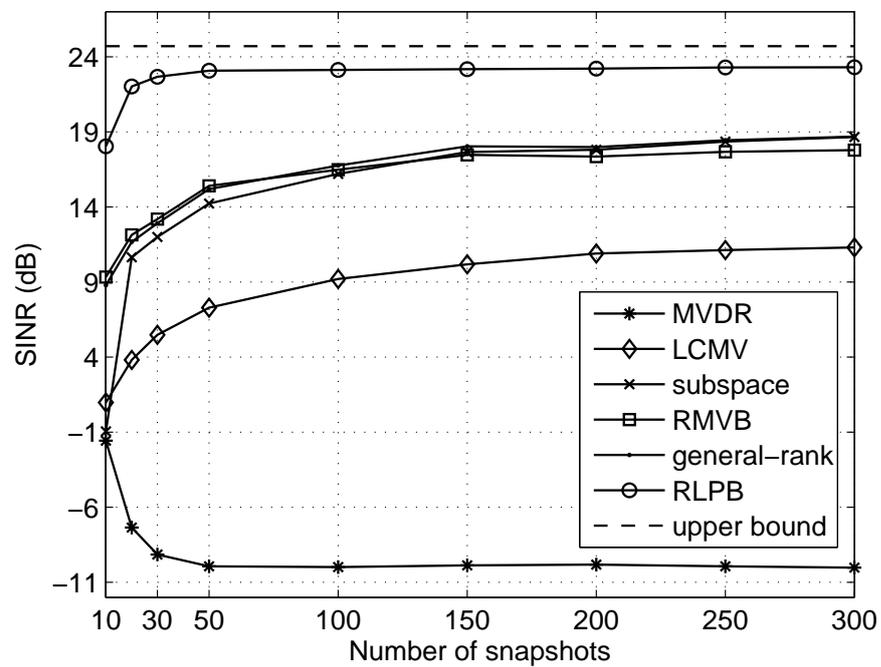
Figure 4.1: Scatter plots of output constellations of the six beamformers when the desired signal uses 16-QAM modulation.

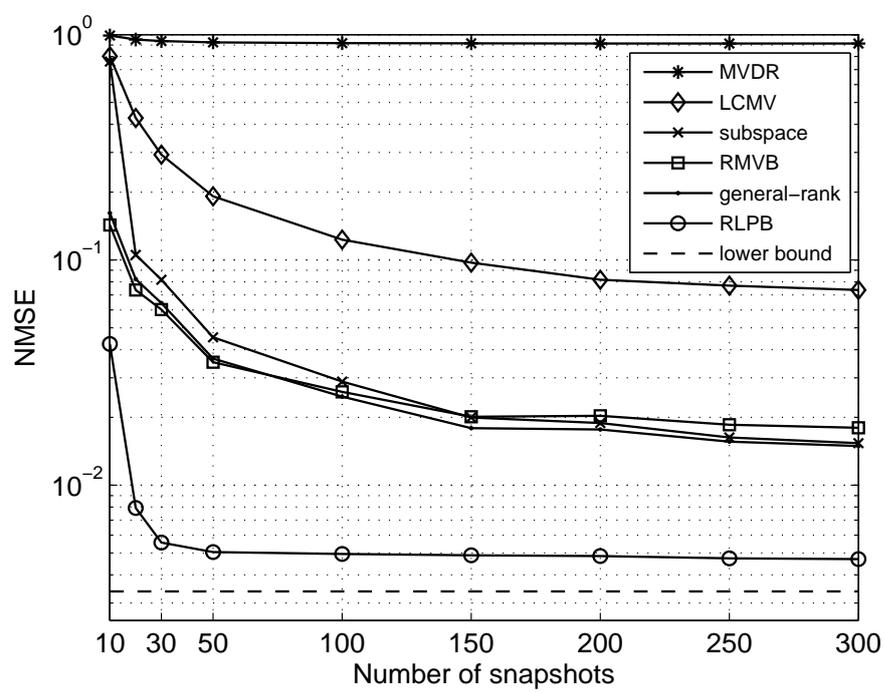
interference suppression. The output of the RLPB exhibits much tighter groupings than those of the RMVB, subspace, and general-rank beamformers.

The desired signal and two interferences are taken as quadrature PSK (QPSK) signals in the subsequent Monte Carlo trials. First, we fix the AOA mismatch to 2° and vary the SNR and N . Figs. 4.2 and 4.3 show the output SINR and NMSE performance versus versus SNR with $N = 100$, respectively. Figs. 4.4 and 4.5 show the SINR and NMSE results versus N at SNR = 15 dB, respectively. Then the SNR is fixed to 15 dB and $N = 100$. Figs. 4.6 and 4.7 plot the SINRs and NMSEs versus AOA mismatch, respectively, when the AOA deviation varies from -3° to 3° with an interval 0.5° . From Figs. 4.2 to 4.7, we can see that the MVDR is not robust

Figure 4.2: Output SINR versus SNR with 2° AOA mismatch.

Figure 4.3: NMSE versus SNR with 2° AOA mismatch.

Figure 4.4: Output SINR versus number of snapshots with 2° AOA mismatch.

Figure 4.5: NMSE versus number of snapshots with 2° AOA mismatch.

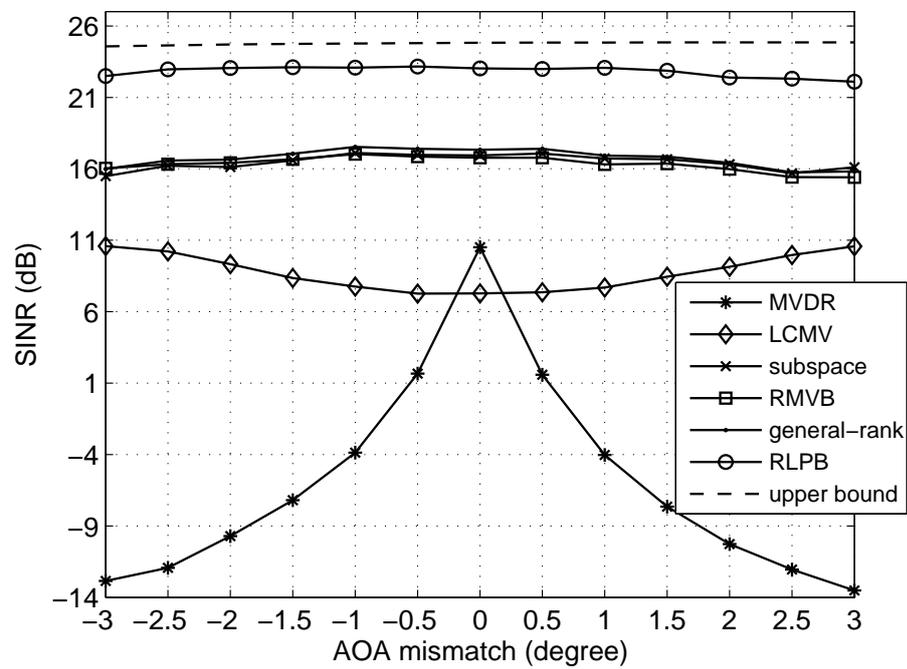


Figure 4.6: Output SINR versus AOA mismatch.

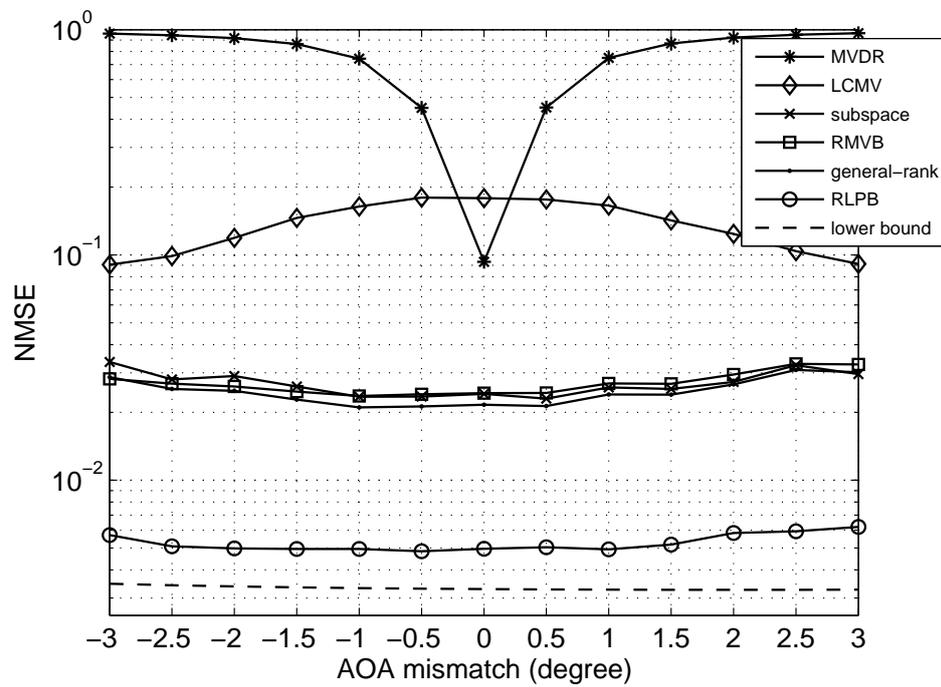


Figure 4.7: NMSE versus AOA mismatch.

against AOA mismatch. The LCMV enhances the robustness but it is inferior to the RMVB and general-rank beamformer, which adopt nonlinear constraints. The subspace beamformer has a comparable performance at high SNR levels and large numbers of snapshots. However, its performance severely degrades at low SNR and small snapshot numbers. The RLPB has the best performance among all beamformers. It leads to significant performance improvement compared with the robust minimum variance beamformers. For example, from Figs. 4.4 and 4.6, we see that the output SINR of the RLPB is about 5–10 dB higher than those of robust minimum variance beamformers. This is because more statistical information on the QPSK signals is exploited by using the minimum ℓ_∞ -norm criterion. From Figs. 4.4 and 4.5, we may find that the performance of the RLPB outperforms the other beamformers even when the number of snapshots is small. It is worth noting that only the RLPB can approach the upper bound of the SINR and the lower bound of NMSE as the SNR increases. The performances of other methods saturate with a large gap compared with the optimal bounds.

4.4.2 Random Steering Vector Mismatch

Now we investigate the performance of the six beamformers with random steering vector mismatch, where the error \mathbf{e} satisfies a complex circular Gaussian distribution of (4.38). The probability that the steering vector lies in the uncertainty region is adopted as $P_\varepsilon = 95\%$, which results in $\sqrt{\tau} = \sqrt{F_{\chi^2}^{-1}(P_\varepsilon, 2M)} = 5.6$. The algorithm parameters for the RMVB and RLPB are $\varepsilon_2 = \sqrt{\tau}\sigma_e$ and $\varepsilon_1 = \sqrt{2M}\varepsilon_2$. Using the

following inequality

$$\begin{aligned}
\|(\mathbf{a} + \mathbf{e})(\mathbf{a} + \mathbf{e})^H - \mathbf{a}\mathbf{a}^H\|_F &= \|\mathbf{a}\mathbf{e}^H + \mathbf{e}\mathbf{a}^H + \mathbf{e}\mathbf{e}^H\|_F \\
&\leq 2\|\mathbf{a}\mathbf{e}^H\|_F + \|\mathbf{e}\mathbf{e}^H\|_F \simeq 2\|\mathbf{a}\mathbf{e}^H\|_F \\
&= 2\sqrt{\text{tr}(\mathbf{a}\mathbf{e}^H\mathbf{e}\mathbf{a}^H)} = 2\|\mathbf{e}\|\|\mathbf{a}\| \leq 2\sqrt{M}\varepsilon_2
\end{aligned} \tag{4.76}$$

where we have omitted the second-order error term $\|\mathbf{e}\mathbf{e}^H\|_F$ and only kept the first-order one, the parameter for the general-rank beamformer can be approximately determined as $\varepsilon_F = 2\sqrt{M}\varepsilon_2$. Note that the LCMV beamformer is designed only for AOA mismatch and cannot handle random mismatch. It adopts the same parameter settings as in Section 4.4.1.

The steering vector perturbation is first fixed to $\text{Ptb} = -10$ dB while the SNR and N vary. Figs. 4.8 and 4.9 plot the output SINR and NMSE versus SNR with $N = 100$, respectively. Figs. 4.10 and 4.11 display the SINR and NMSE versus N at SNR = 15 dB, respectively. Then the SNR is fixed to 15 dB and $N = 100$ while the array perturbation varies from -30 dB to -6 dB. The -6 dB array perturbation corresponding to $\sigma_e = 0.35$ causes a large steering vector error. Figs. 4.12 and 4.13 show the SINR and NMSE versus steering vector perturbation, respectively.

It can be seen from Figs. 4.8 – 4.13 that the RLPB has the best performance for random steering vector mismatch. It significantly outperforms other robust beamforming schemes. For example, as shown in Fig. 4.12, the output SINR of the proposed RLPB is about 8 dB higher than those of robust minimum variance beamformers. It maintains satisfactory performance with large steering vector errors and

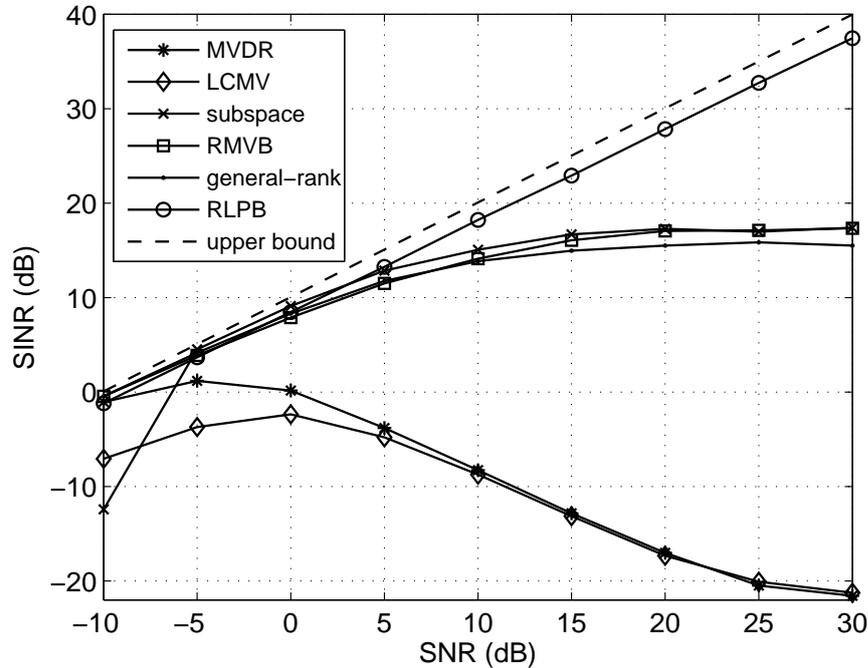


Figure 4.8: Output SINR versus SNR with random steering vector mismatch.

smaller numbers of snapshots. Again, only the performance of the RLPB can approach the optimal bounds with large numbers of snapshots or small amount of perturbation. Note that the performance of the LCMV beamformer designed for AOA mismatch dramatically degrades, which demonstrates that it cannot handle random mismatch.

4.4.3 Robustness to Parameter Selection

The sizes of the uncertainty sets, i.e., ε_2 and ε_1 , are crucial parameters of the RMVB and RLPB, respectively. We investigate the effect of parameter selection for Gaussian random steering vector mismatch. The SNR is fixed at 15 dB, steering vector

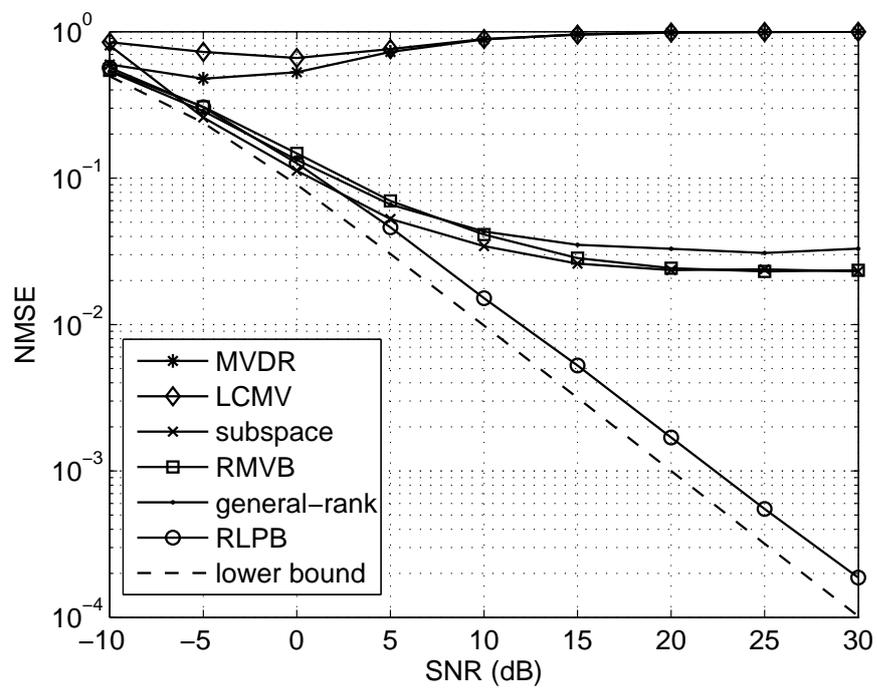


Figure 4.9: NMSE versus SNR with random steering vector mismatch.

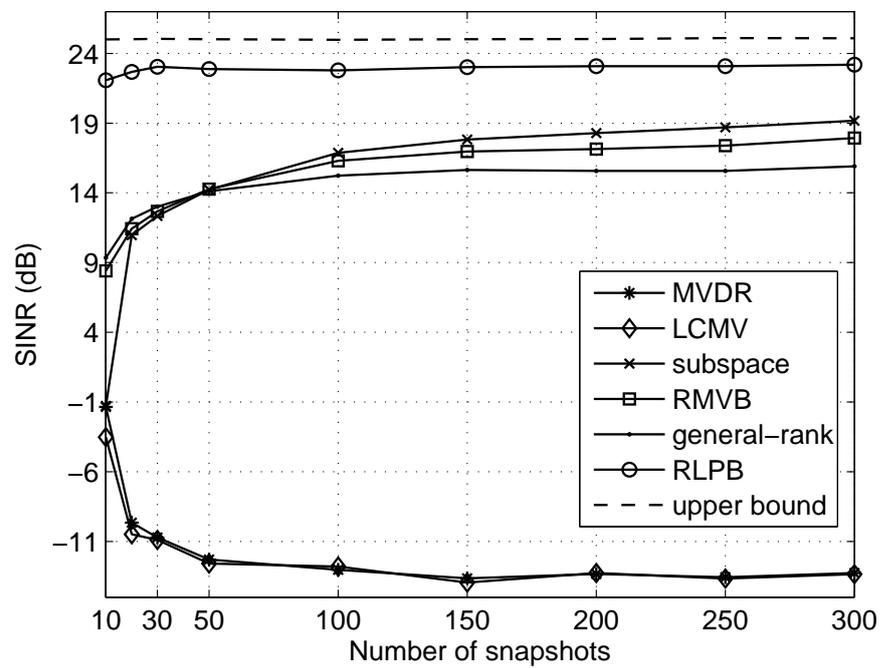


Figure 4.10: Output SINR versus number of snapshots with random steering vector mismatch.

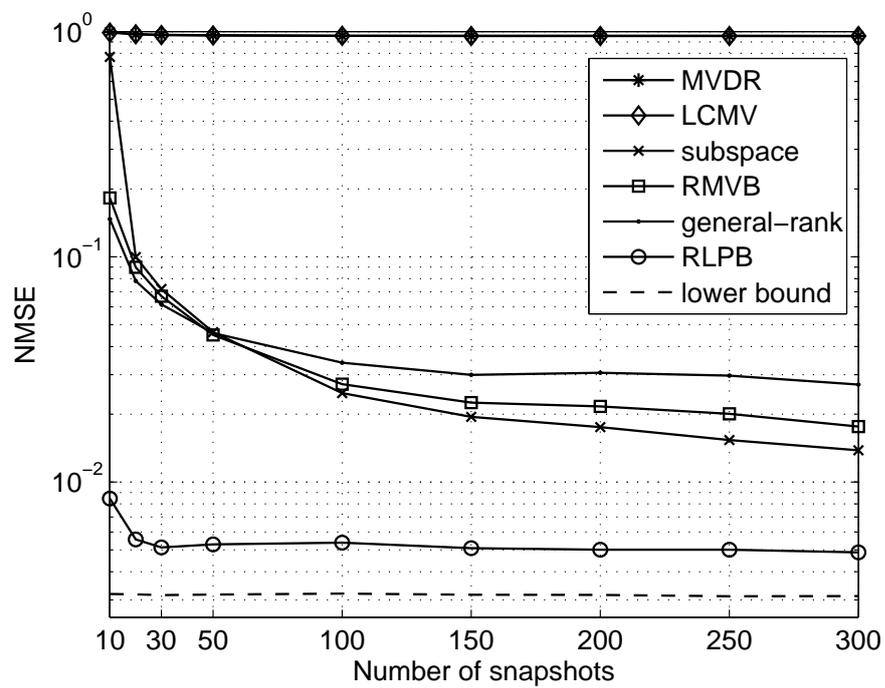


Figure 4.11: NMSE versus number of snapshots with random steering vector mismatch.

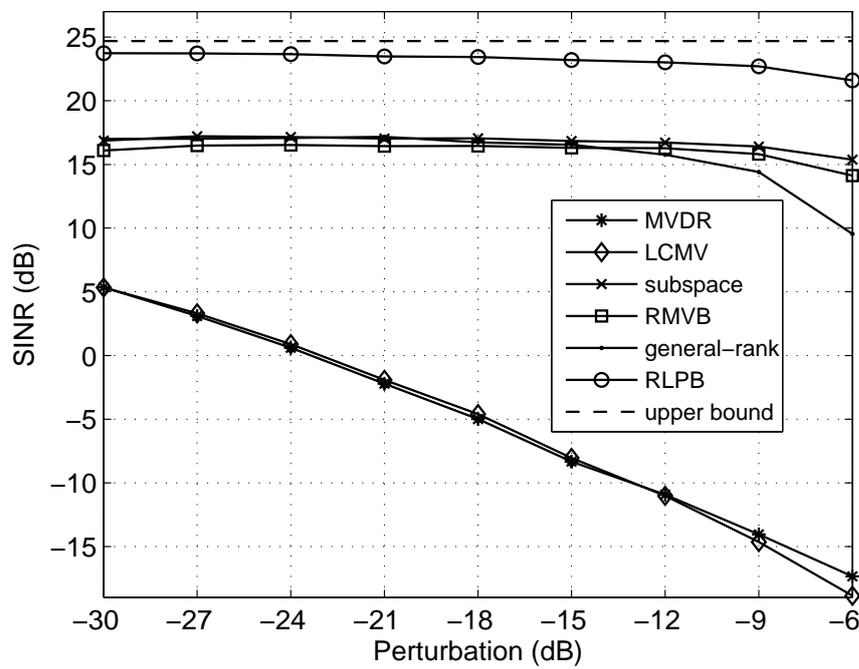


Figure 4.12: Output SINR versus perturbation amount of steering vector.

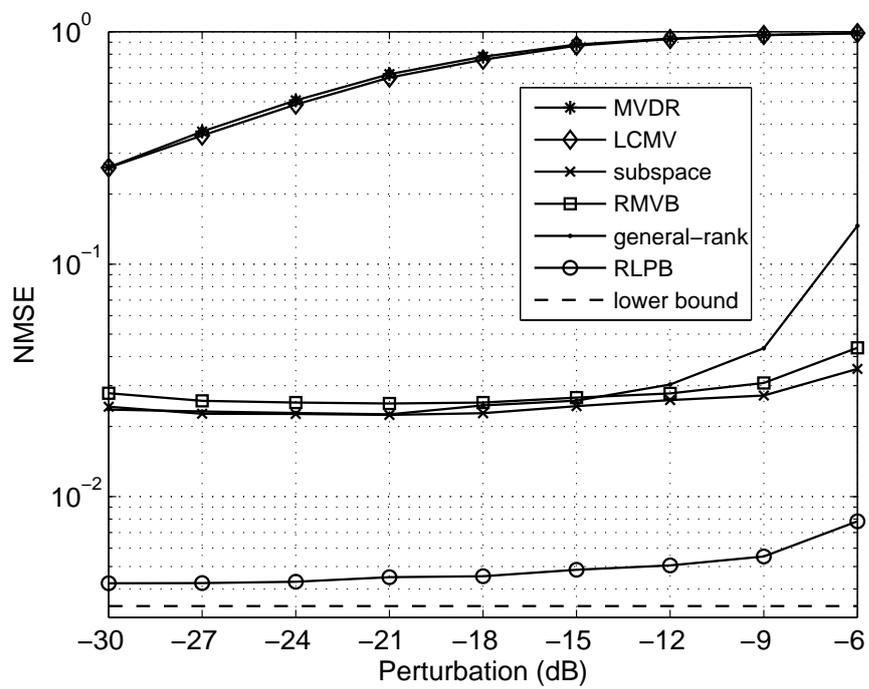


Figure 4.13: Output SINR versus perturbation amount of steering vector.

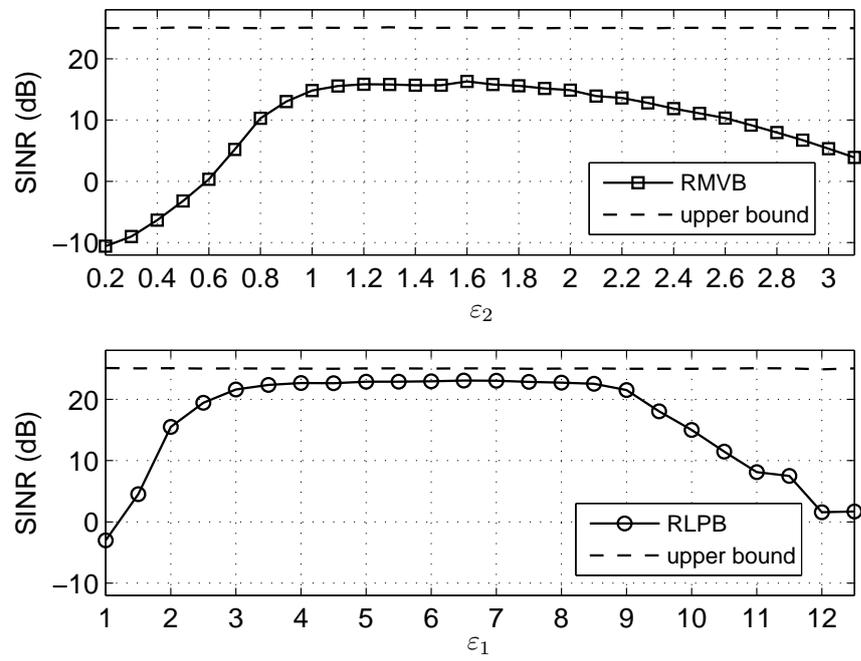
perturbation $P_{tb} = -10$ dB, and $N = 100$. The other experimental settings are the same as in Section 4.4.2. The maximal possible values of the algorithmic parameters satisfy $\varepsilon_2 < \|\mathbf{a}\| = \sqrt{M} = 3.16$ and $\varepsilon_1 < \|\mathbf{a}\|_1 = 12.6$.

The output SINR of the RMVB versus ε_2 and that of RLPB versus ε_1 are plotted in Fig. 4.14, which shows that $\varepsilon_2 \in [1, 2]$ and $\varepsilon_1 \in [3, 9]$ can provide satisfactory performance for the RMVB and RLPB, respectively. If the probability is set to $P_\varepsilon = 95\%$, the resulting parameters are $\varepsilon_2 = 1.25$ and $\varepsilon_1 = 5.6$, which are located in the intervals of $[1, 2]$ and $[3, 9]$, respectively. Since a wide range of ε_1 yields reliable performance, the requirement for the selection of an appropriate ε_1 is rather mild. Furthermore, the selection of ε_1 is more flexible than that of ε_2 because the interval $[3, 9]$ is wider than $[1, 2]$.

4.4.4 Robustness to Signal Properties

The performance improvement in the RLPB originates from fully exploiting the signal properties. In this subsection, the effect caused by the signal properties, including modulation type, constellation distortion due to carrier frequency offset (CFO), and the probability distribution of the signals is analyzed. The simulation settings are the same as in Section 4.4.2 unless stated otherwise. The steering vector perturbation is fixed at $P_{tb} = -10$ dB and $N = 100$.

First, the sensitivity to modulation type of the beamformer is investigated. Eight modulation types, i.e., binary PSK (BPSK), QPSK, 8PSK, 16-PSK, 4QAM, 8QAM, 16-QAM, and 32-QAM, are considered. The SNR is 20 dB. In each simulation, the desired signal and interferences take one modulation type. Figs. 4.15 and 4.16 show the output SINR and NMSE versus modulation type, respectively. It is observed that

Figure 4.14: Output SINR of RMVB versus ϵ_2 and RLPB versus ϵ_1 .

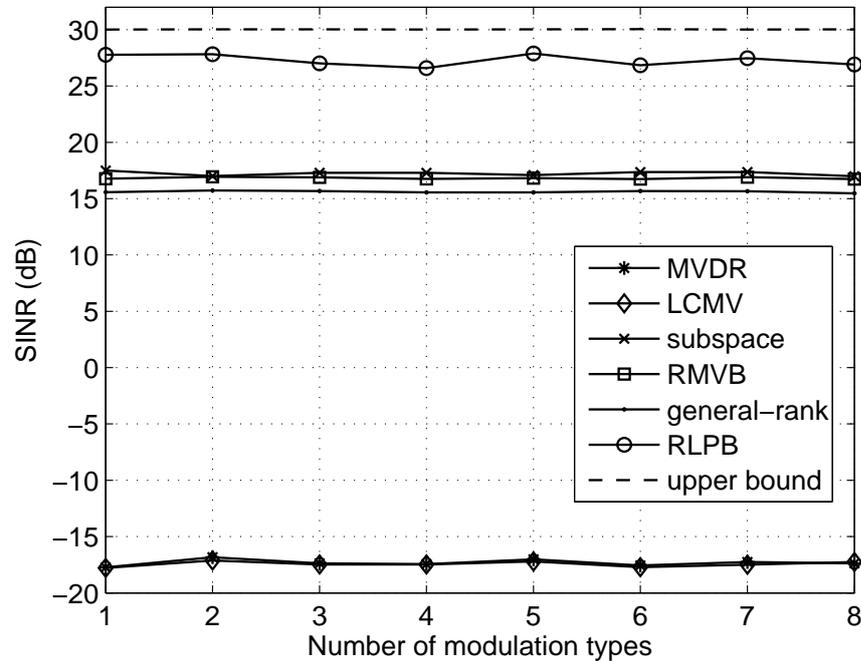


Figure 4.15: Output SINR versus modulation type where the numbers 1 to 8 denote BPSK, QPSK, 8PSK, 16-PSK, 4QAM, 8QAM, 16-QAM, and 32-QAM, respectively.

the RLPB performs well for a variety of modulation types and is better than the MV-based beamformers. Its performance is not sensitive with respect to modulation type.

Second, we study the effect of constellation distortion due to CFO, which is mainly induced by oscillator discrepancies between the transmitter and receiver and/or Doppler shifts. The desired signal adopts QPSK modulation while the two interferers use 16-QAM. Figs. 4.17 and 4.18 depict the output SINR and NMSE versus the CFO, respectively, at SNR = 15 dB. The CFO is normalized with respect to the sampling rate varying from 10^{-6} to 1, which covers small to large frequency offsets. We can see

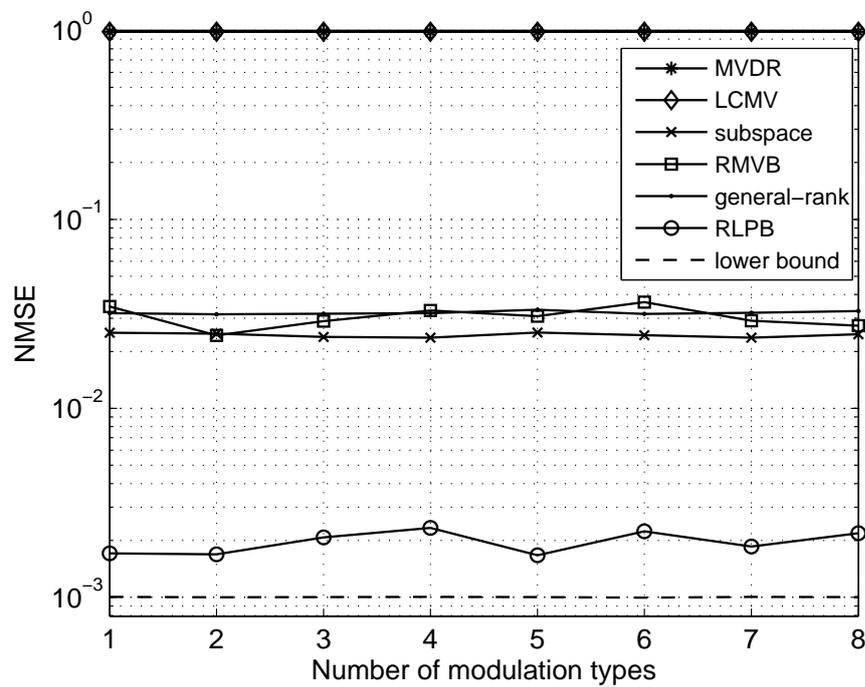


Figure 4.16: NMSE versus modulation type where the numbers 1 to 8 denote BPSK, QPSK, 8PSK, 16-PSK, 4QAM, 8QAM, 16-QAM, and 32-QAM, respectively.

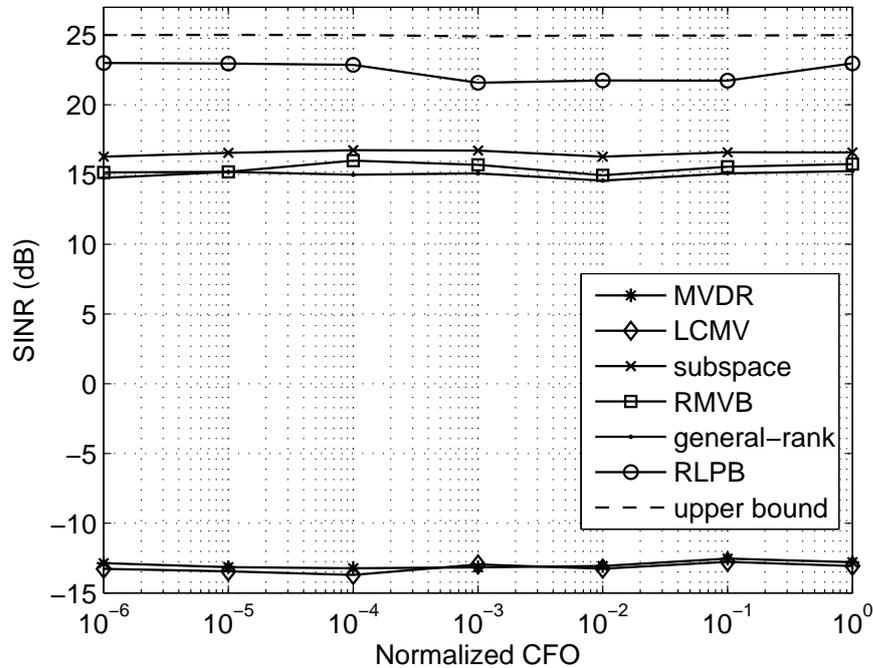


Figure 4.17: Output SINR versus normalized CFO.

that the RLPB is not sensitive to the CFO and it maintains good performance even in the presence of large frequency offset.

Third, we consider several signals with hybrid modulation types, including both digital and analog modulations. Two common analog modulations, namely, FM and PM, are taken into account. The AOA of the desired QPSK signal is 43° . The modulation types of the four interferers are 16-QAM, 4QAM, linear FM, and random PM. All four interferers are 10 dB stronger than the desired signal and their AOAs are 30° , 75° , -20° , and 10° , respectively. The complex baseband envelop of the random PM signal is expressed as $s_{\text{PM}}(n) = \exp(j\phi_n)$, where ϕ_n is a random variable satisfying a uniform distribution in $[0, 2\pi)$. That is, it is generated with uniformly distributed

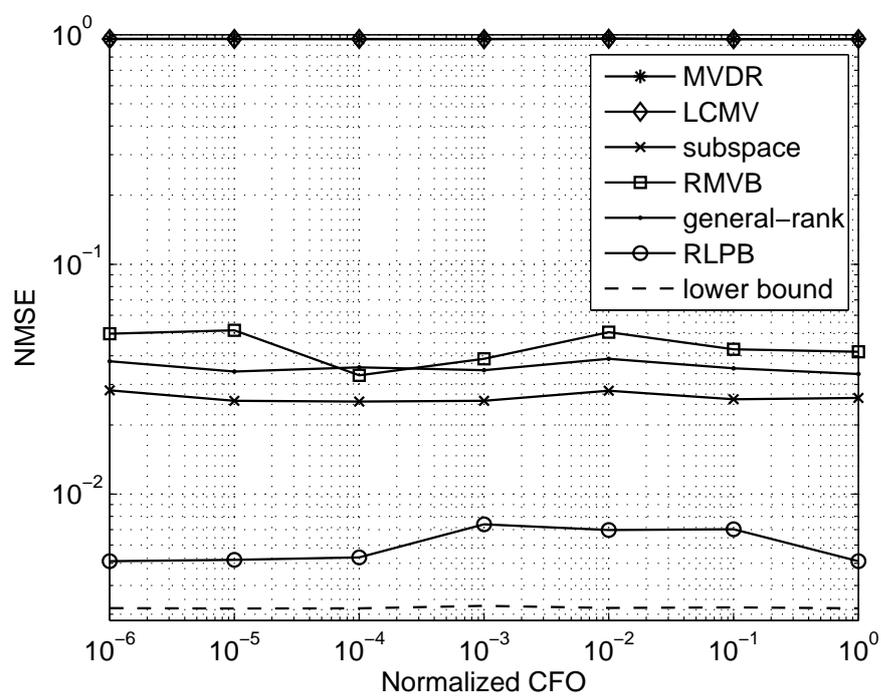


Figure 4.18: NMSE versus normalized CFO.

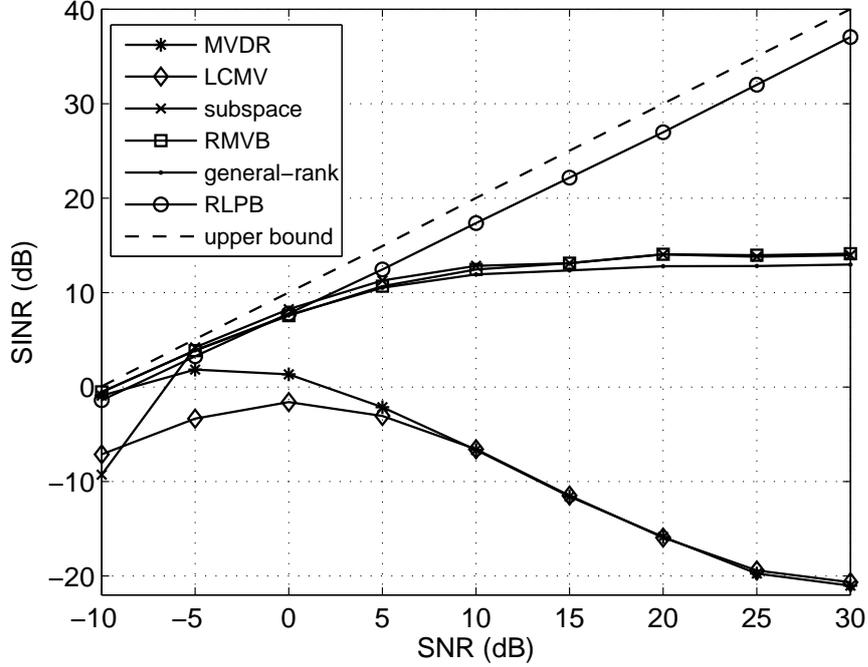


Figure 4.19: Output SINR versus SNR with hybrid modulation.

random phase in $[0, 2\pi)$. The complex baseband envelop of the linear FM signal has the following form:

$$s_{\text{FM}}(n) = \exp(j2\pi Bn^2/N), \quad n = 0, \dots, N - 1 \quad (4.77)$$

where $B \in (0, 1]$ is the normalized bandwidth. In the simulation, we set $B = 1/16$. Figs. 4.19 and 4.20 plot the output SINR and NMSE versus SNR with hybrid modulation, respectively. It can be seen that the RLPB can handle general interferences with hybrid modulation.

In the last simulation, we investigate the performance of the six beamformers when

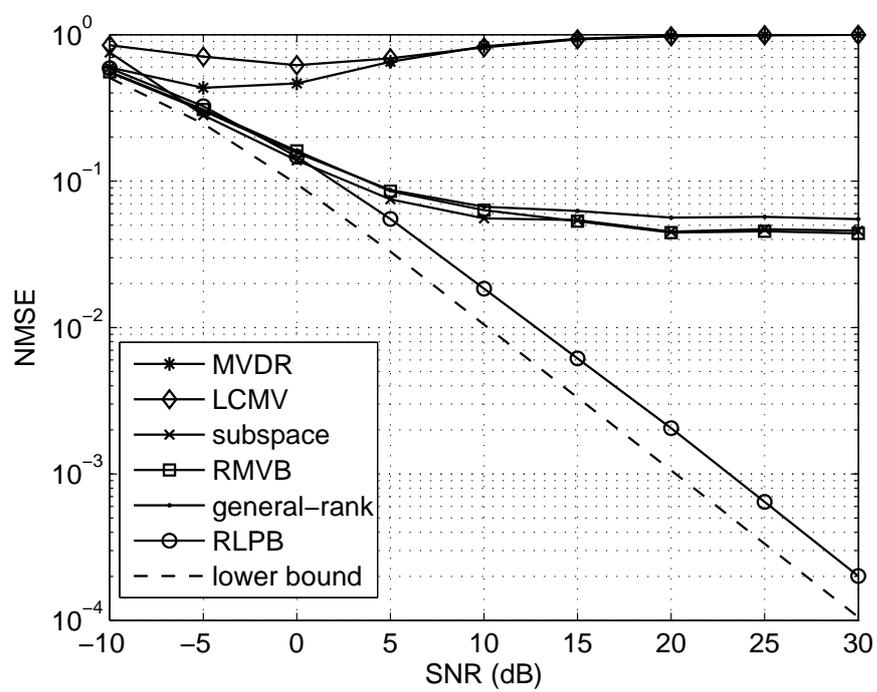


Figure 4.20: NMSE versus SNR with hybrid modulation.

the desired signal, interferences, and noise are Gaussian distributed. The simulation settings are the same as in Section 4.4.2 except that the desired signal and two interferences are now replaced by Gaussian signals. Figs. 4.21 and 4.22 show the output SINR and NMSE versus SNR, respectively. It is observed that the performance of the RLPB degrades and it is inferior to the three MV-based robust beamformers. This result is not surprising because all signals and noise are Gaussian distributed. Since all statistical information are included in the second-order statistics for Gaussian signals, the MV criterion is statistically optimal. In this case, the MV-based robust beamformers is superior to the RLPB using ℓ_∞ -norm minimization. The RLPB still significantly outperforms the MVDR and LCMV beamformers. Note that the blind beamforming technique based on fourth-order cumulants [44] and CMA [37,69] fail if all signals are Gaussian distributed, but the RLPB still works. It is worth pointing out that the scenario where all signals and interferences are Gaussian is rare in practice because the modulated radio signals are always sub-Gaussian [37,41].

4.5 Conclusion

Most of the advanced convex optimization approaches to robust beamforming are based on the minimum variance criterion and assume a spherical or ellipsoidal uncertainty region of the steering vector error. These robust beamforming schemes require solving an SOCP. In order to exploit higher-order statistics of sub-Gaussian signals, the proposed RLPB adopts the criterion of minimum ℓ_∞ -norm of the output, where a newly defined ℓ_p -norm is based on the ℓ_p -modulus of complex numbers. The uncertainty region of the RLPB is modeled as a rhombus in which the ℓ_∞ -modulus response of any steering vector is constrained to exceed unity. The issue of selecting the size of

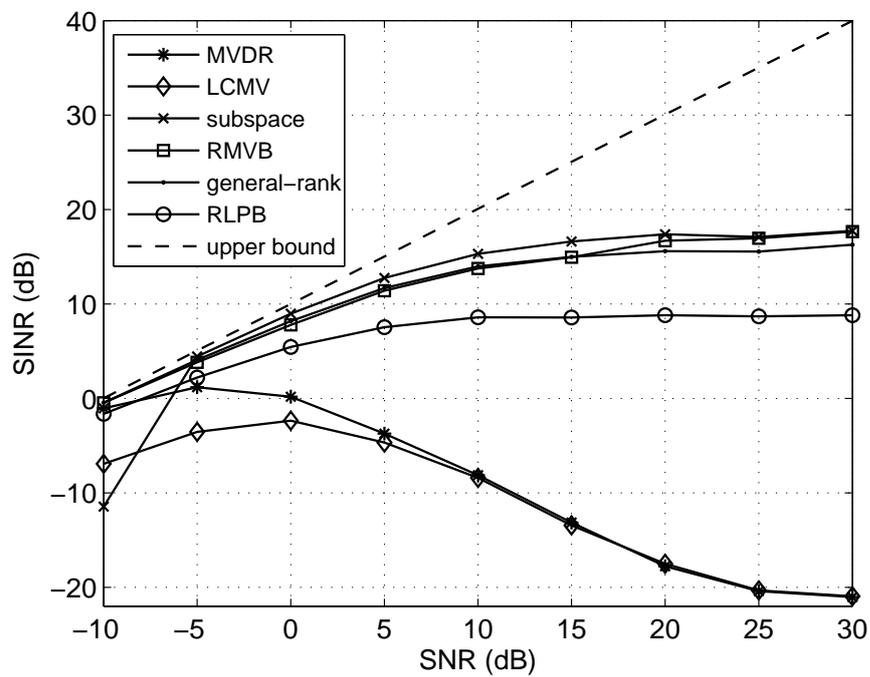


Figure 4.21: Output SINR versus SNR with Gaussian signals and random steering vector mismatch.

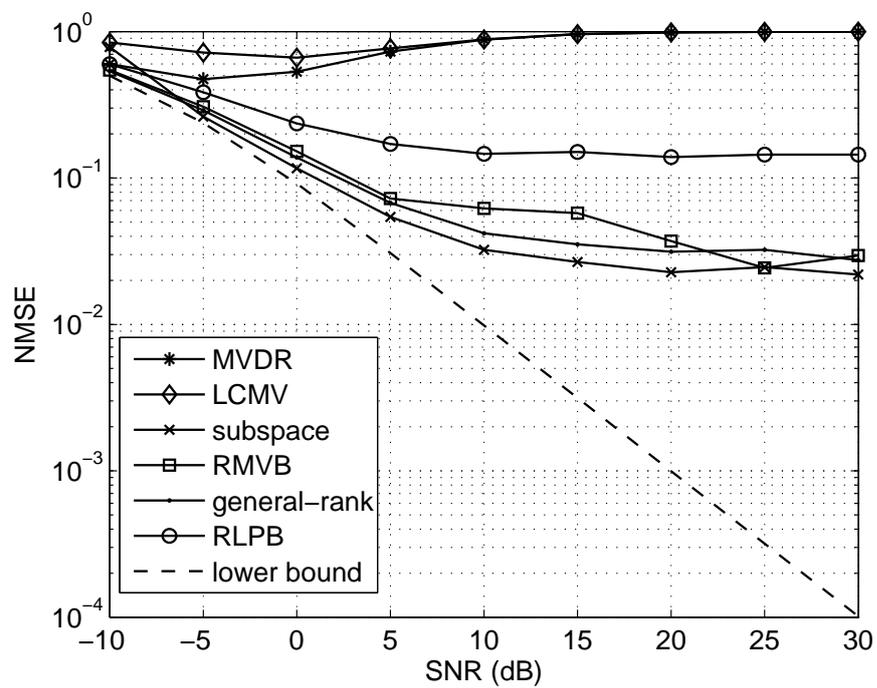


Figure 4.22: NMSE versus SNR with Gaussian signals and random steering vector mismatch.

the rhombic uncertainty region is also discussed for both AOA mismatch and random mismatch. The proposed beamformer is finally converted into a linear programming and hence is simpler to be solved than the SOCP based methods. A theoretical explanation to the reason why the RLPB can implicitly exploit the high-order statistics is given from the statistical perspective. Simulation results illustrate that the RLPB significantly outperforms other representative robust beamformers.

Chapter 5

Conclusion

In this thesis, we have presented the robust minimum dispersion beamforming techniques for non-Gaussian signals.

In Chapter 2, the minimum dispersion distortionless response (MDDR) beamformer and linearly constrained minimum dispersion (LCMD) beamformer are proposed for non-Gaussian signals by using the MD criterion with a single linear constraint and multiple linear constraints, respectively. The MDDR and LCMD beamformers outperform their respective standard counterparts based on minimum variance (MV), namely, the minimum variance distortionless response (MVDR) and linearly constrained minimum variance (LCMV) beamformers, for non-Gaussian signals. Three efficient iterative algorithms, namely, the iteratively reweighted MVDR (IR-MVDR), complex-valued full Newton's and partial Newton's methods, are devised to solve the resulting convex optimization problems.

In Chapter 3, we extend the LCMD beamformer to the quadratically constrained minimum dispersion (QCMD) beamformer, whose robustness against model uncertainty is significantly enhanced compared with the LCMD beamformer. A gradient

projection algorithmic framework is developed to efficiently solve the resulting convex optimization problem. Furthermore, we derive a closed-form expression of the projection onto the constraint set. The proposed projected gradient methods are quite suitable for the robust beamforming problem with large sample size or large sensor number because their computational complexity is linearly proportional to the number of samples and sensors. Simulation results demonstrate that the QCMD beamformer substantially improves the SINR performance compared with the MV-based robust beamformers.

In Chapter 4, the robust linear programming beamformer (RLPB) is proposed using a minimum ℓ_∞ -norm criterion for sub-Gaussian signals. We model the uncertainty region as a rhombus in which the ℓ_1 -norm of the steering vector error is bounded. As a result, the proposed RLPB beamformer can be obtained by solving a linear programming (LP) problem. A theoretical explanation to the reason why the RLPB can implicitly exploit the high-order statistics is given from the statistical perspective. Simulation results illustrate that the RLPB significantly outperforms other representative robust beamformers by using different sub-Gaussian signals.

There are various topics worthy of future research.

- *Wideband Beamforming:* The narrowband assumption are used throughout the thesis. However, when the desired signal or the interferences are wideband, the narrowband beamforming scheme no longer holds. In this case, we have to employ an additional processing dimension for effective operation, such as tapped delay-lines (or FIR/IIR filters), which lead to a wideband beamforming system [32, 81]. To our best knowledge, the utilization of higher- or lower-order statistics for non-Gaussian signals has not been considered in the existing

wideband beamforming methods. Therefore, an important future work is to develop an MD-based robust wideband beamforming approach. It will attract more attentions in the field of microphone array speech processing because the speech signals are known to be wideband and super-Gaussian [81].

- *Multiuser Receivers:* The MV-based beamforming approaches have been extended to the case of designing multiuser receivers for code-division multiple access (CDMA) [82–91] and space-time coded multiple-input multiple-output (MIMO) communication systems [92–94]. Since most communication signals are sub-Gaussian [37,41], the higher-order statistics contain useful information. Therefore, we may consider to incorporate the MD criterion into the designing of multiuser receivers.
- *Transmit Beamforming and Network Beamforming:* Classic beamforming is matched to a single steering vector of interest, which can be applied to both receive beamforming and unicast transmit beamforming towards a single receiver [7]. However, in the cellular multiuser downlink, multiple transmit beamforming weight vectors have to be jointly designed to balance the interference between different transmissions to different users [95–99]. Network beamforming is a rapidly developing research field over the last five years [7, 100–105], which belongs to the general field of cooperative communications [7, 106]. It can be considered as a certain combination of receive and transmit beamforming strategies. Both transmit beamforming and network beamforming are relative young and exciting research fields that we want to engage in the future. The robustness against channel state information is still a challenging problem for

network beamforming. In addition, the computationally efficient implementation methods are critical for real-time applications of beamforming [107].

Bibliography

- [1] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, 1993.
- [2] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.*, vol. 5 (no. 2): pp. 4–24, Apr. 1988.
- [3] T. J. Shepherd and J. G. McWhirter. Systolic adaptive beamforming. In S. Haykin, J. Litva, and T. J. Shepherd, editors, *Radar Array Processing*. Springer-Verlag, Berlin, 1993.
- [4] A. B. Gershman, E. Németh, and J. F. Böhme. Experimental performance of adaptive beamforming in a sonar environment with a towed array and moving interfering sources. *IEEE Trans. Signal Process.*, vol. 48 (no. 1): pp. 246–250, Jan. 2000.
- [5] L. C. Godara. Application of antenna arrays to mobile communications, part ii: Beamforming and direction-of-arrival considerations. *Proc. IEEE*, vol. 85 (no. 8): pp. 1195–1245, Aug. 1997.

- [6] J. Benesty, J. Chen, Y. A. Huang, and J. Dmochowski. On microphone-array beamforming from a MIMO acoustic signal processing perspective. *IEEE Audio, Speech, Language Process.*, vol. 15 (no. 3): pp. 1053–1065, Mar. 2007.
- [7] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten. Convex optimization-based beamforming. *IEEE Signal Process. Mag.*, vol. 27 (no. 4): pp. 62–75, May 2010.
- [8] S. A. Vorobyov. Principles of minimum variance robust adaptive beamforming design. *Signal Process.*, vol. 93 (no. 12): pp. 3264–3277, Dec. 2013.
- [9] 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, Jul. 1996.
- [10] J. Li and P. Stoica. *Robust Adaptive Beamforming*. John Wiley & Sons Inc., 2006.
- [11] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, vol. 57 (no. 8): pp. 1408–1418, Aug. 1969.
- [12] J. C. Preisig. Robust maximum energy adaptive matched field processing. *IEEE Trans. Signal Process.*, vol. 42 (no. 7): pp. 1585–1593, Jul. 1994.
- [13] M. Rübsamen. *Advanced direction-of-arrival estimation and beamforming techniques for multiple antenna systems*. PhD thesis, Technische Universität Darmstadt, May 2011.
- [14] O. L. Forst. An algorithm for linearly constrained adaptive processing. *Proc. IEEE*, vol. 60 (no. 8): pp. 926–935, Aug. 1972.

- [15] L. Chang and C. C. Yeh. Performance of dmi and eigenspace-based beamformers. *IEEE Trans. Antennas Propagat.*, vol. 40 (no. 11): pp. 1336–1347, Nov. 1992.
- [16] B. D. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24 (no. 4): pp. 397–401, Jul. 1988.
- [17] K. L. Bell, Y. Ephraim, and H. L. Van Trees. A Bayesian approach to robust adaptive beamforming. *IEEE Trans. Signal Process.*, vol. 48 (no. 2): pp. 386–398, Feb. 2000.
- [18] S. A. Vorobyov, A. B. Gershman, and Zhi-Quan Luo. Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem. *IEEE Trans. Signal Process.*, vol. 51 (no. 2): pp. 313–324, Feb. 2003.
- [19] P. Stoica, Z. Wang, and J. Li. Robust Capon beamforming. *IEEE Signal Process. Lett.*, vol. 10 (no. 6): pp. 172–175, Jun. 2003.
- [20] J. Li, P. Stoica, and Z. Wang. On robust Capon beamforming and diagonal loading. *IEEE Trans. Signal Process.*, vol. 51 (no. 7): pp. 1702–1714, Jul. 2003.
- [21] 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 (no. 9): pp. 2257–2269, Sep. 2003.
- [22] 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 (no. 2): pp. 108–111, Feb. 2004.
- [23] J. Li, P. Stoica, and Z. Wang. Doubly constrained robust Capon beamformer. *IEEE Trans. Signal Process.*, vol. 52 (no. 9): pp. 2407–2423, Sep. 2004.
- [24] R. G. Lorenz and S. P. Boyd. Robust minimum variance beamforming. *IEEE Trans. Signal Process.*, vol. 53 (no. 5): pp. 1684–1696, May 2005.
- [25] A. Mutapcic, S.-J. Kim, and S. P. Boyd. Beamforming with uncertain weights. *IEEE Signal Process. Lett.*, vol. 14 (no. 5): pp. 348–351, May 2007.
- [26] C.-Y. Chen and P. P. Vaidyanathan. Quadratically constrained beamforming robust against direction-of-arrival mismatch. *IEEE Trans. Signal Process.*, vol. 55 (no. 8): pp. 4139–4150, Aug. 2007.
- [27] S.-J. Kim, A. Magnani, A. Mutapcic, S. P. Boyd, and Z.-Q. Luo. Robust beamforming via worst-case SINR maximization. *IEEE Trans. Signal Process.*, vol. 56 (no. 4): pp. 1539–1547, Apr. 2008.
- [28] Z. L. Yu, M. H. Er, and W. Ser. A novel adaptive beamformer based on semidefinite programming (SDP) with magnitude response constraints. *IEEE Trans. Antennas Propagat.*, vol. 56 (no. 5): pp. 1297–1307, May 2008.
- [29] 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 (no. 11): pp. 5719–5724, Nov. 2008.

- [30] A. Hassaniien, S. A. Vorobyov, and K. M. Wong. Robust adaptive beamforming using sequential quadratic programming: An iterative solution to the mismatch problem. *IEEE Signal Process. Lett.*, vol. 15: pp. 733–736, 2008.
- [31] Y. Zhang, J. P. Lie, B. P. Ng, and C. M. S. See. Robust minimum ℓ_1 -norm adaptive beamformer against intermittent sensor failure and steering vector error. *IEEE Trans. Antennas Propagat.*, vol. 58 (no. 5): pp. 1796–1801, May 2010.
- [32] M. RübSamen, A. El-Keyi, A. B. Gershman, and T. Kirubarajan. Robust broadband adaptive beamforming using convex optimization. In D. P. Palomar and Y. C. Eldar, editors, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, Cambridge, 2010.
- [33] M. RübSamen and A. B. Gershman. Robust adaptive beamforming using multi-dimensional covariance fitting. *IEEE Trans. Signal Process.*, vol. 60 (no. 2): pp. 740–753, Feb. 2012.
- [34] A. Khabbazibasmenj, S. A. Vorobyov, and A. Hassaniien. Robust adaptive-beamforming based on steering vector estimation with as little as possible prior information. *IEEE Trans. Signal Process.*, vol. 60 (no. 6): pp. 2974–2987, Jun. 2012.
- [35] M. RübSamen and M. Pesavento. Maximally robust Capon beamformer. *IEEE Trans. Signal Process.*, vol. 61 (no. 8): pp. 2030–2041, Apr. 2013.

- [36] A. Khabbazi-basmenj and S. A. Vorobyov. Robust adaptive beamforming for general-rank signal model with positive semi-definite constraint via POTDC. *IEEE Trans. Signal Process.*, vol. 61 (no. 23): pp. 6103–6117, Dec. 2013.
- [37] C. R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas. Blind equalization using the constant modulus criterion: A review. *Proc. IEEE*, vol. 86 (no. 10): pp. 1927–1950, Oct. 1998.
- [38] M. Shao and C. L. Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proc. IEEE*, vol. 81 (no. 7): pp. 986–1010, Jul. 1993.
- [39] K. P. Balanda and H. L. MacGillivray. Kurtosis: A critical review. *The American Statistician*, vol. 42 (no. 2): pp. 111–119, May 1988.
- [40] J. V. Stone. *Independent Component Analysis: A Tutorial Introduction*. MIT Press, 2004.
- [41] V. Zarzoso and A. K. Nandi. Exploiting non-Gaussianity in blind identification and equalization of MIMO FIR channels. *IEE Proc.-Vis. Image Signal Process.*, vol. 151 (no. 1): pp. 69–75, Feb. 2004.
- [42] S. Gazor and W. Zhang. Speech probability distribution. *IEEE Signal Process. Lett.*, vol. 10 (no. 7): pp. 204–207, Jul. 2003.
- [43] D. Middleton. Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models. *IEEE Trans. Inf. Theory*, vol. 45 (no. 4): pp. 1129–1149, May 1999.

- [44] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proc. F Radar Signal Process.*, vol. 140 (no. 6): pp. 362–370, Dec. 1993.
- [45] T.-J. Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33 (no. 4): pp. 806–811, Aug. 1985.
- [46] J. S. Reed, J. D. Mallett, and L. E. Brennan. Robust minimum variance beamforming. *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10 (no. 6): pp. 853–863, Nov. 1974.
- [47] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagat.*, vol. 34 (no. 3): pp. 276–280, Mar. 1986.
- [48] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33 (no. 3): pp. 387–392, Apr. 1985.
- [49] X. Jiang, W.-J. Zeng, A. Yasocharan, H. C. So, and T. Kirubarajan. Minimum dispersion beamforming for non-Gaussian signals. *IEEE Trans. Signal Process.*, vol. 62 (no. 7): pp. 1879–1893, Apr. 2014.
- [50] X. Jiang, W.-J. Zeng, A. Yasocharan, H. C. So, and T. Kirubarajan. Robust beamforming by linear programming. *IEEE Trans. Signal Process.*, vol. 62 (no. 7): pp. 1834–1849, Apr. 2014.
- [51] X. Jiang, W.-J. Zeng, A. Yasocharan, H. C. So, and T. Kirubarajan. Gradient projection for robust minimum dispersion beamforming. Submitted to *IEEE Trans. Signal Process.*

- [52] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Process. Mag.*, vol. 29 (no. 4): pp. 61–80, Jul. 2012.
- [53] P. Tsakalides and C. L. Nikias. Robust adaptive beamforming in alpha-stable noise environments. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Atlanta, Georgia, USA, May 1996.
- [54] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley & Sons Inc., 1995.
- [55] Y. Nesterov and A. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
- [56] R. H. Byrd and D. A. Pyne. Convergence of the iteratively reweighted least-squares algorithm for robust regression. Technical report, John Hopkins University, 1979.
- [57] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math*, vol. 63 (no. 1): pp. 1–38, 2010.
- [58] M. Novey, T. Adali, and A. Roy. A complex generalized Gaussian distribution–characterization, generation, and estimation. *IEEE Trans. Signal Process.*, vol. 58 (no. 3): pp. 1427–1433, Mar. 2010.
- [59] D. Ge, X. Jiang, and Y. Ye. A note on the complexity of l_p minimization. *Math. Program., Ser. B*, vol. 129 (no. 2): pp. 285–299, Oct. 2011.

- [60] H. F. Walker and P. Ni. Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.*, vol. 49 (no. 4): pp. 1715–1735, Jul. 2011.
- [61] W.-J. Zeng, H. C. So, and L. Huang. ℓ_p -music: Robust direction-of-arrival estimator for impulsive noise environments. *IEEE Trans. Signal Process.*, vol. 61 (no. 17): pp. 4296–4308, Sep. 2013.
- [62] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2006.
- [63] E. K. P. Chong and S. H. Zak. *An Introduction to Optimization*. John Wiley & Sons Inc., 2007.
- [64] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.
- [65] A. N. Iusem. On the convergence properties of the projected gradient method for convex optimization. *Computational and Applied Mathematics*, vol. 22 (no. 1): pp. 37–52, 2003.
- [66] Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: A general approach. *Annals of Operations Research*, vol. 46 (no. 1): pp. 157–178, 1993.
- [67] M. Abramowitz. Algebraic equations. In M. Abramowitz and I. A. Stegun, editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, 1972.
- [68] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1985.

- [69] D. N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Commun.*, vol. 28 (no. 11): pp. 1867–1875, Nov. 1980.
- [70] M. Grant and S. P. Boyd. CVX: Matlab software for disciplined convex programming. <http://cvxr.com/cvx/>, Mar. 2014.
- [71] F. Solomon. *Probability and Stochastic Processes*. Cambridge Univ. Press, 1985.
- [72] G. B. Dantzig. *Linear Programming and Extensions*. Princeton Univ. Press, 1998.
- [73] M. S. Lobo, L. Vandenberghe, S. P. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra Appl.*, vol. 284 (no. 1): pp. 193–228, Nov. 1998.
- [74] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program., Ser. B*, vol. 95 (no. 1): pp. 3–51, Jan. 2003.
- [75] X.-L. Li and X.-D. Zhang. A family of generalized constant modulus algorithms for blind equalization. *IEEE Trans. Commun.*, vol. 54 (no. 11): pp. 1913–1917, Nov. 2006.
- [76] W.-J. Zeng, X.-L. Li, and X.-D. Zhang. Adaptive newton algorithms for blind equalization using the generalized constant modulus criterion. In *Proceedings of the IEEE ICASSP*, Taipei, Taiwan, 2009.
- [77] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Univ. Press, 1988.

- [78] C. Forbes, M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. John Wiley & Sons Inc., 2010.
- [79] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1964.
- [80] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, 1996.
- [81] W. Liu and S. Weiss. *Wideband Beamforming: Concepts and Techniques*. Wiley, 2010.
- [82] M. Honig, U. Madhow, and S. Verdú. Blind adaptive multiuser detection. *IEEE Trans. Inform. Theory*, vol. 41 (no. 4): pp. 944–960, Jul. 1995.
- [83] H. V. Poor and X. Wang. Code-aided interference suppression for DS/CDMA spread-spectrum communications – part II: Parallel blind adaptive implementations. *IEEE Trans. Commun.*, vol. 45 (no. 9): pp. 1112–1122, Sep. 1997.
- [84] M. K. Tsatsanis and Z. Xu. Performance analysis of minimum variance CDMA receivers. *IEEE Trans. Signal Process.*, vol. 46 (no. 11): pp. 3014–3022, Nov. 1998.
- [85] H. V. Poor and X. Wang. Blind multiuser detection: A subspace approach. *IEEE Trans. Inform. Theory*, vol. 44 (no. 2): pp. 677–690, Mar. 1998.
- [86] Z. Xu, P. Liu, and X. Wang. Blind multiuser detection: From MOE to subspace methods. *IEEE Trans. Signal Process.*, vol. 52 (no. 2): pp. 510–524, Feb. 2004.

- [87] S. Shahbazpanahi and A. B. Gershman. Robust blind multiuser detection for synchronous CDMA systems using worst-case performance optimization. *IEEE Trans. Wireless Commun.*, vol. 3 (no. 6): pp. 2232–2245, Nov. 2004.
- [88] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Z.-Q. Luo. Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver. *IEEE Trans. Signal Process.*, vol. 53 (no. 1): pp. 295–305, Jan. 2005.
- [89] S. Cui, M. Kisiailiou, Z.-Q. Luo, and Z. Ding. Robust blind multiuser detection against signature waveform mismatch based on second-order cone programming. *IEEE Trans. Wireless Commun.*, vol. 4 (no. 4): pp. 1285–1291, Jul. 2005.
- [90] A. Elnashar, S. Elnoubi, and H. A. El-Mikati. Performance analysis of blind adaptive MOE multiuser receivers using inverse QRD-RLS algorithm. *IEEE Trans. Circuits Syst. I, Regular Papers*, vol. 55 (no. 1): pp. 398–411, Feb. 2008.
- [91] R. C. de Lamare and R. Sampaio-Neto. Reduced-rank space-time adaptive interference suppression with joint iterative least squares algorithms for spread-spectrum systems. *IEEE Trans. Veh. Technol.*, vol. 59 (no. 3): pp. 1217–1228, Mar. 2010.
- [92] S. Shahbazpanahi, M. Beheshti, A. B. Gershman, M. Gharavi-Alkhansari, and K. M. Wong. Minimum variance linear receivers for multiaccess MIMO wireless systems with space-time block coding. *IEEE Trans. Signal Process.*, vol. 52 (no. 12): pp. 3306–3313, Dec. 2004.

- [93] Y. Rong, S. Shahbazpanahi, and A. B. Gershman. Robust linear receivers for space-time block coded multi-access MIMO systems with imperfect channel state information. *IEEE Trans. Signal Process.*, vol. 53 (no. 8): pp. 3081–3090, Aug. 2005.
- [94] Y. Rong, S. A. Vorobyov, and A. B. Gershman. Robust linear receivers for multi-access space-time block coded MIMO systems: A probabilistically constrained approach. *IEEE J. Sel. Areas Commun.*, vol. 24 (no. 8): pp. 1560–1570, Aug. 2006.
- [95] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE J. Sel. Areas Commun.*, vol. 16 (no. 8): pp. 1437–1450, Oct. 1998.
- [96] C. Farsakh and J. A. Nossek. Spatial covariance-based downlink beamforming in an SDMA mobile radio system. *IEEE Trans. Commun.*, vol. 46 (no. 11): pp. 1497–1506, Nov. 1998.
- [97] B. K. Chalise, S. Shahbazpanahi, A. Czylik, and A. B. Gershman. Robust downlink beamforming based on outage probability specifications. *IEEE Trans. Wireless Commun.*, vol. 6 (no. 10): pp. 3498–3503, Oct. 2007.
- [98] N. Vucic and H. Boche. Robust QoS-constrained optimization of downlink multiuser MISO systems. *IEEE Trans. Signal Process.*, vol. 57 (no. 2): pp. 714–725, Feb. 2009.

- [99] N. Vucic and H. Boche. A tractable method for chance-constrained power control in downlink multiuser MISO systems with channel uncertainties. *IEEE Signal Process. Lett.*, vol. 16 (no. 5): pp. 346–349, May 2009.
- [100] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo. Distributed beamforming for relay networks based on second-order statistics of the channel state information. *IEEE Trans. Signal Process.*, vol. 56 (no. 9): pp. 4306–4316, Sep. 2008.
- [101] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor. Multiple peer-to-peer communications using a network of relays. *IEEE Trans. Signal Process.*, vol. 57 (no. 8): pp. 3053–3062, Aug. 2009.
- [102] H. Chen, A. B. Gershman, and S. Shahbazpanahi. Filter-and-forward distributed beamforming in relay networks with frequency selective fading. *IEEE Trans. Signal Process.*, vol. 53 (no. 3): pp. 1251–1262, Mar. 2010.
- [103] M. F. A. Ahmed and S. A. Vorobyov. Sidelobe control in collaborative beamforming via node selection. *IEEE Trans. Signal Process.*, vol. 58 (no. 12): pp. 6168–6180, Dec. 2010.
- [104] C.-I Kuo, S.-H. Wu, and C.-K. Tseng. Robust linear beamformer designs for coordinated multi-point AF relaying in downlink multi-cell networks. *IEEE Trans. Wireless Commun.*, vol. 11 (no. 9): pp. 3272–3283, Sep. 2012.
- [105] D. Ponukumati, F. Gao, and C. Xing. Robust peer-to-peer relay beamforming: A probabilistic approach. *IEEE Commun. Lett.*, vol. 17 (no. 2): pp. 305–308, Feb. 2013.

-
- [106] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inform. Theory*, vol. 50 (no. 12): pp. 3062–3080, Dec. 2004.
- [107] J. Mattingley and S. P. Boyd. Real-time convex optimization in signal processing. *IEEE Signal Process. Mag.*, vol. 27 (no. 3): pp. 50–61, Mar. 2010.