Pontifícia Universidade Católica
do Rio de Janeiro

Erico de Souza Prado Lopes

Symbol-Level Transmit Processing for Multiuser MIMO Systems with PSK Modulation

Tese de Doutorado

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia Elétrica.

Advisor: Prof. Lukas T. N. Landau

Rio de Janeiro
January 2025

Pontifícia Universidade Católica
do Rio de Janeiro

Erico de Souza Prado Lopes

Symbol-Level Transmit Processing for Multiuser MIMO Systems with PSK Modulation

Dissertation presented to the Programa de Pós–graduação em Engenharia Elétrica of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia Elétrica. Approved by the Examination Committee.

Prof. Lukas T. N. Landau
Advisor
Departamento de Engenharia Elétrica (DEE) – PUC-Rio

Prof. Rodrigo C. de Lamare
Departamento de Engenharia Elétrica (DEE) – PUC-Rio

Dr. André Robert Flores Manrique
Departamento de Engenharia Elétrica (DEE) – PUC-Rio

Prof. Amine Mezghani
University of Manitoba – UM

Dr. Michael Joham
Technische Universität München – TUM

Prof. Tadeu N. Ferreira
Universidade Federal Fluminense – UFF

Rio de Janeiro, January the 8th, 2025

Erico de Souza Prado Lopes

Received the B.Sc. in electrical engineering with emphasis in telecommunications and electronics from the Pontifical Catholic University of Rio de Janeiro in 2018. He received the M.Sc. degree in electrical engineering in the area of communication systems in 2021 also by the Pontifical Catholic University of Rio de Janeiro. Currently, he works at Instituto Nacional de Propriedade Industrial (INPI) as a patent examiner, and is a Ph.D. student at the Center for Studies in Telecommunications, Pontifical Catholic University of Rio de Janeiro, Brazil. His research interests include communications, signal processing, and optimization.

Ficha Catalográfica

Lopes, Erico de Souza Prado

Symbol-Level Transmit Processing for Multiuser MIMO Systems with PSK Modulation / Erico de Souza Prado Lopes; advisor: Lukas T. N. Landau. – 2025.

144 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Departamento de Engenharia Elétrica (DEE).

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Pré-codificação a nível de símbolo. 3. Superfícies refletoras reconfiguráveis. 4. Sistemas MIMO multiusuário. 5. Sinalização de envelope constante. 6. Quantização de baixa resolução. I. Landau, Lukas T. N. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Departamento de Engenharia Elétrica (DEE). III. Título.

CDD: 621.3

This thesis is dedicated to my family
for their endless love, support, and encouragement.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to Prof. Lukas T. N. Landau for guiding me through the academic world. From my M.Sc. advising until the present work we had six years of fruitful conversations where I learned a lot. Today I can joyfully say that he is not only my advisor but also a great friend. This thesis would not have been possible without his guidance, enthusiasm, and unwavering support. I would also like to thank him for giving me his computer during the COVID-19 pandemic, as this thesis would not have been completed without it.

I would like to extend my thanks to Prof. Amine Mezghani, Dr. André Flores, Dr. Michael Joham, Prof. Rodrigo de Lamare, and Prof. Tadeu Ferreira for serving as referees on my defense committee.

I am deeply grateful to my mother, father, and wife for being my most fierce supporters. Their support has given me the strength to persevere through challenging times. I would like to thank my mother for always listening to me even when I was down and, in difficult times, doing her best to lift my spirit. She was and still is always there for me when I need the most, being not only lovely and caring but also hard when required. I would like to thank my father for encouraging me to pursue the PhD and carefully listening to my long (and probably boring) explanations about my research topics. He was the one who encouraged me the most to proceed through the hardest pieces of research, which were also the most fruitful ones. Without his love and excitement, I would not have reached this moment. I would like to thank my wife Marianne for being the most wonderful person I know and for her affection through good and bad times. Mari is always there to celebrate when I succeed and to cheer me up when I fail. She has always understood the resignations that come with pursuing a PhD and supported me throughout this time. This thesis would not be possible without all of you and is as much yours as it is mine. I also have the best mother and father-in-law. The lunches we've had together were not only fun but also had insightful ideas for life. My sincere thanks to Fátima and Sérgio for the best time and the wonderful teachings. I love you all.

I had the privilege of sharing a laboratory with some amazing people. I want to thank Ali and Diana for the fun we had studying together and for their friendship. Moreover, throughout the PhD, I started working at AMAZUL, which supported me in pursuing this work. I would like to thank my friends Ícaro, Alina, Juliana, Giovanna, Victor, Iara, Tiago, Daniel, Cayque, and Mariana for the great times we had together.

I would like to express my sincere thanks to all the professors, students, and staff at CETUC for providing a pleasant environment for me to pursue

my studies. Thanks to all of you, CETUC became my second home during my eight years of studying telecommunications.

Finally, I would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) for the financial support that made it possible to complete my PhD. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Abstract

Landau, Lukas T. N (Advisor). **Symbol-Level Transmit Processing for Multiuser MIMO Systems with PSK Modulation**. Rio de Janeiro, 2025. 144p. Tese de doutorado – Departamento de Departamento de Engenharia Elétrica (DEE), Pontifícia Universidade Católica do Rio de Janeiro.

This study proposes different symbol-level transmit processing methods for diverse multiuser MIMO setups. First, two symbol-level precoders are developed considering a strict per-antenna power constraint and PSK modulation for perfect and imperfect channel state information.
Then, a large-scale MIMO setup is considered where the energy consumption of the radio frequency front ends yields a bottleneck for realizing energy-efficient MIMO systems. With this, power reduction features such as constant envelope signaling and low-resolution quantization are applied to enable low-cost deployments, with low environmental impact, and better coverage. In this context, the minimum symbol-error probability formulation is considered as the design criterion for the case of QPSK data symbols, and, for arbitrary PSK modulation, the study proposes the novel minimum union-bound symbol-error probability formulation. Based on these criteria the study proposes different low-resolution symbol-level precoders based on the partial greedy search method and the proposed quality-of-service branch and bound algorithm.
Finally, a virtual multiuser MIMO system with PSK modulation realized via the reconfigurable intelligent surface-based passive transmitter setup is considered. Under this framework, this study considers both high-resolution and discrete phase shift reconfigurable intelligent surface models. With these frameworks, the study derives symbol-level power minimization problems under quality of service constraints. Both the symbol-error probability and union-bound symbol-error probability are considered for the quality of service formulation. The problems are solved by utilizing a bisection method and a branch-and-bound method for high and low-resolution reflecting elements, respectively.

Keywords

Symbol-Level Precoding, Reconfigurable Intelligent Surfaces, Multiuser MIMO Systems, Constant Envelope Signaling, Low-Resolition Quantization

Resumo

Lopes, Erico de Souza Prado; Landau, Lukas T. N. **Processamento de Sinais a Nível de Símbolo para Transmissão em Sistemas MIMO com modulação PSK**. 2025. 144p. Tese de Doutorado – Departamento de Departamento de Engenharia Elétrica (DEE), Pontifícia Universidade Católica do Rio de Janeiro.

Este estudo propõe diferentes métodos de processamento de transmissão a nível de símbolo para diversas configurações MIMO multiusuário. Primeiro, dois pré-codificadores a nível de símbolo são desenvolvidos considerando uma estrita restrição de potência por antena e modulação PSK para informações de estado de canal perfeito e imperfeito. Então, uma configuração MIMO em larga escala é considerada onde o consumo de energia dos front-ends de radiofrequência produz um gargalo para a realização de sistemas MIMO com eficiência energética. Com isso, recursos de redução de energia, como sinalização de envelope constante e quantização de baixa resolução, são aplicados para permitir implantações de baixo custo, com baixo impacto ambiental e melhor cobertura. Neste contexto, a formulação da mínima probabilidade de erro de símbolo é considerada como o critério de projeto para o caso de símbolos de dados QPSK e, para modulação PSK arbitrária, o estudo propõe a nova formulação de mínima probabilidade de erro de símbolo vinculada ao limitante da união. Com base nestes critérios, o estudo propõe diferentes pré-codificadores de baixa resolução a nível de símbolo, baseados no método de busca parcial gananciosa e no algoritmo proposto de branch-and-bound qualidade de serviço. Finalmente, é considerado um sistema MIMO multiusuário virtual com modulação PSK realizado através da utilização de um transmissor baseado em superfícies inteligentes reconfiguráveis. Com esta estrutura, este estudo considera modelos de superfície inteligentes reconfiguráveis de alta resolução e com mudança de fase discreta. Com essas estruturas, o estudo deriva problemas de minimização de potência a nível de símbolo sob restrições de qualidade de serviço. Tanto a probabilidade de erro de símbolo quanto a probabilidade de erro de símbolo vinculada ao limitante da união são consideradas para a formulação da qualidade do serviço. Os problemas são resolvidos utilizando um método de bissecção e um método branch-and-bound para elementos refletores de alta e baixa resolução, respectivamente.

Palavras-chave

Pré-codificação a nível de símbolo Superfícies refletoras reconfiguráveis Sistemas MIMO multiusuário Sinalização de envelope constante Quantização de baixa resolução

Table of contents

List of figures

List of tables

List of Abreviations

AMAF – Active Multi-Antenna Feeder
ANTP – Average Normalized Transmit Power
AWGN – Additive White Gaussian Noise
B&B – Branch-and-Bound
BER – Bit-Error Rate
BPSK – Binary Phase Shift Keying
BM – Bisection Method
BS – Base Station
CE – Constant Envelope
CI – Constructive Interference
CSI – Channel State Information
DPP – Discrete Programming Problem
EE – Energy Efficiency
ESDP – Euclidean Smallest Distance Projection
IPM – Interior Points Method
LHS – Left Hand Side
LMMSE – Linear Minimum Mean Squared Error
MDDT – Minimum Distance to the Decisions Threshold
MIMO – Multiple-Input Multiple-Output
MIP – Mixed-Integer Program
MMDDT – Maximum Minimum Distance to the Decisions Threshold
MMSE – Minimum Mean Squared Error
MRT – Maximum Ratio Transmission
MUBSEP – Minimum Union-Bound Symbol-Error Probability
MU-MIMO – Multiuser Multiple-Input Multiple-Output
MSEP – Minimum Symbol-Error Probability
MSM – Maximum Safety Margin
PA – Power Amplifier
PAPC – Per Antenna Power Constraint
PGS – Partial Greedy Search
PHMMDDT – Power Minimization for High-Resolution RIS Under MDDT Constraints
PHSEP – Power Minimization for High-Resolution RIS Under SEP Constraints
PHUBSEP – Power Minimization for High-Resolution RIS Under UBSEP Constraints

PSEP – Power Minimization Under SEP
PSK – Phase Shift Keying
PUBSEP – Power Minimization Under UBSEP Constraints
QAM – Quadrature Amplitude Modulation
QoS – Quality of Service
QP – Quadratic Program
QPSK – Quadrature Phase Shift Keying
RCG – Riemannian Conjugate Gradient
RF – Radio Frequency
RFFE – Radio Frequency Front End
RMMSE – Robust Minimum Mean Squared Error
RIS – Reconfigurable Intelligent Surfaces
UBCO – Upper Bound Complexity Order
UBSEP – Union-Bound Symbol-Error Probability
SEP – Symbol-Error Probability
SER – Symbol-Error Rate
SNR – Signal-to-Noise Ratio
SOCP – Second Order Cone Program
SPAPC – Strict Per Antenna Power Constraint
SLP – Symbol-Level Precoding
TPC – Total Power Constraint
TSBP – Tree Search Based Precoding
ZF – Zero Forcing

Published and Submitted Articles

The work presented in this thesis gave rise to the following conference and journal papers:

Conference Papers

1. E. S. P. Lopes, L. T. N. Landau and A. Mezghani, *Minimum Union Bound Symbol Error Probability Precoding for PSK Modulation and Phase Quantization* published in the proceedings of 2022 IEEE Globecom Workshops, Rio de Janeiro, Brazil.
2. E. S. P. Lopes and L. T. N. Landau, *Symbol-Error Probability Constrained Power Minimization for Reconfigurable Intelligent Surfaces-Based Passive Transmitter* published in the proceedings of 2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Los Sueños, Costa Rica.
3. E. S. P. Lopes, L. T. N. Landau and A. Mezghani, *Power Minimization Under QoS Requirements for Low-Resolution Multiuser MIMO Systems With Reconfigurable Intelligent Surfaces Based Transmitter* published in the proceedings of ICC 2024 - IEEE International Conference on Communications, Denver, CO, USA, 2024.
4. E. S. P. Lopes, L. T. N. Landau and A. Mezghani, *RIS-based Passive Transmitter Reflection Optimization with Performance Complexity Trade-Offs* presented in at the International Symposium on Wireless Communication Systems (ISWCS) 2024 (ISWCS'24), Rio de Janeiro, Brazil.

Journal Papers

1. E. S. P. Lopes and L. T. N. Landau, *Discrete MMSE Precoding for Multiuser MIMO Systems With PSK Modulation* published in IEEE Transactions on Wireless Communications in October 2022.
2. E. S. P. Lopes and L. T. N. Landau, *MMSE Symbol Level Precoding Under a Per Antenna Power Constraint for Multiuser MIMO Systems*

With PSK Modulation published in IEEE Wireless Communications Letters in November 2022.

3. E. S. P. Lopes, L. T. N. Landau and A. Mezghani, *Minimum Symbol Error Probability Discrete Symbol Level Precoding for MU-MIMO Systems with PSK Modulation* published in IEEE Transactions on Communications in October 2023.
4. E. S. P. Lopes, L. T. N. Landau and A. Mezghani, *Power Minimization under Quality of Service Constraints for RIS-based Passive Transmitter MIMO Systems* submitted to IEEE Transactions on Communications at August 2024.

1
Introduction

One central demand to enable the new key applications of the next generation of wireless communications is a higher data rate with high reliability and increased energy efficiency (EE) [1]. According to [2], the future wireless generation will require an improvement in the data rate of factor 100 for the uplink and factor 50 for the downlink while achieving 10 thousand times higher reliability when compared to 5G. Moreover, as stated in [3], 6G networks will require 10 to 100 times higher EE compared to 5G, to enable scalable low-cost deployments, with low environmental impact, and better coverage. Multiuser multiple-input multiple-output (MU-MIMO) systems are considered a promising physical-layer technique and are expected to be a key technology for attaining these requirements [2]. Yet, the design of EE MIMO systems with minimum error-rate compromise remains a challenge to overcome.

The conventional MU-MIMO implementation consists of equipping a base station (BS) with large antenna arrays to allow for large diversity gains. Yet, due to the high number of radio frequency front ends (RFFEs), the energy consumption of the radio frequency (RF) chains imposes a challenge for this kind of technology [4]. The EE requirement led to the development of different studies that analyzed the circuit of RFFEs to dissect the most consuming elements, e.g., [5, 6]. These works conclude that the power amplifiers (PAs) and data converters are two of the most consuming elements in the RFFE. With this, many recent studies consider adopting features to minimize the power consumption of these elements. In most cases, to increase the PA's efficiency the adoption of constant envelope (CE) signaling is considered, and, to decrease the power consumption of the data converters, low-resolution in amplitude is utilized.

Another method to realize low-cost EE MU-MIMO systems is the utilization of reconfigurable intelligent surfaces (RIS). RIS are two-dimensional surfaces with many reconfigurable reflecting elements that can independently adjust their reflection coefficient in a real-time programmable manner. As proposed in [7, 8, 9, 10, 11], one can construct RIS-based transmitter by illuminating a RIS with a carrier signal generated by a nearby RF signal generator and changing the parameters of the reflecting elements to modulate and transmit

information symbols. With this, RIS-based passive transmitters realize virtual MIMO systems with a small number of RF chains and cost-effective reflecting elements, which benefits the implementation of massive MIMO with reduced hardware complexity and increased EE.

1.1 Contributions

A fundamental problem for MU-MIMO systems is the design of low-complexity transmit processing algorithms that attain the high-reliability constraints of future wireless communications networks. This thesis proposes symbol-level transmit processing algorithms for different MIMO setups. It is divided into three parts where the contributions are delineated for the different setups.

1.1.1 Contributions under the Conventional MIMO Transmitter and Strict Per Antenna Power Constraint

The first part of the study considers a conventional MU-MIMO scenario and proposes different optimal symbol-level precoding (SLP) algorithms based on the minimum mean squared error (MMSE) and the robust MMSE (RMMSE) criteria. For the algorithms' design, a strict per antenna power constraint (SPAPC) is adopted as it is considered the most realistic model for this scenario [12].

1.1.2 Contributions under the Conventional MIMO Transmitter with Constant Envelope and Low Resolution Constraints

For large-scale MIMO where the energy consumption of the RFFE is significant to the EE of the system, power reduction features such as CE signaling and low-resolution quantization are necessary for low-cost deployments, with low environmental impact, and better coverage. In this scenario, the contributions of this study are listed in the following:

- Development of a novel precoding design criterion based on the minimization of the union-bound symbol error probability (MUBSEP).
- Development of practical partial greedy search (PGS) SLP algorithms based on the relaxation of the feasible sets to its convex hull utilizing the minimum symbol error probability (MSEP) and MUBSEP design criteria.

– Development of the quality of service (QoS) branch-and-bound (B&B) precoding algorithm which improves standard full B&B approaches in the sense of incorporating a symbol error probability (SEP) requirement in the method.
– Development of optimal RMMSE SLPs based on a B&B algorithm for imperfect channel state information (CSI) scenarios.
– Development of a suboptimal low complexity RMMSE SLP based on the relaxation of the discrete feasible set to its convex hull and subsequent Euclidean distance mapping, also for imperfect CSI scenarios.

1.1.3 Contributions under the RIS-Based Transmitter Framework

As mentioned RIS-based transmitters can be utilized to achieve low-cost implementations of MIMO systems. In this context, the MIMO scenarios that arise depend on the RIS' hardware implementation. When the RIS allows for a high number of phase shifts, the feasible set of the reflection coefficients can be well approximated by a Riemannian manifold. On the other hand, the general case yields a discrete feasible set.

For both scenarios, this study proposes power minimization problems under QoS constraints. While for the case of BPSK or QPSK users' data, the SEP is considered QoS requisite, for other PSK scenarios the union-bound SEP (UBSEP) functions are utilized as constraints. For the general case of discrete phase-shift RIS, the problem is constructed as a mixed-integer program (MIP) and solved via an improved version of the QoS B&B approach. In the high-resolution case, it becomes a multivariate problem that is solved via the combined utilization of a bisection method (BM) and the Riemannian Conjugate Gradient (RCG) algorithm.

1.2 Thesis Outline

The remainder of this thesis consists of six chapters which are structured as follows:

– Chapter 2 presents a brief review of the literature and dissects also some important baselines;
– Chapter 3 presents the developed SPAPC precoding techniques with the conventional MIMO transmitter;
– Chapter 4 proposes different SLPs under CE and low-resolution constraints with the conventional MIMO transmitter;

– Chapter 5 proposes the different symbol-level power minimization problems under a virtual MIMO system realized with the RIS-based passive transmitter setup;

– Chapter 6 presents the conclusion of the thesis;

– Chapter 7 discusses the possible extensions of the studies presented in this thesis.

Finally, the appendix presents the convexity analysis, the development of the minimum distance to the decisions threshold (MDDT) bound, and the signal-to-noise ratio (SNR) definitions.

1.3 Notation

Regarding the notation, bold lowercase and uppercase letters indicate vectors and matrices, respectively. Non-bold letters express scalars. The operators $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote complex conjugation, transposition and conjugate transposition, respectively. The i-th element of a given vector $\boldsymbol{a}$ is denoted by $[\boldsymbol{a}]_i$. Real and imaginary part operators, as well as the functions $\mathrm{erf}(z) = \int_0^z e^{-t^2} dt$, $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$, $\Phi(z) = (1/2)\,\mathrm{erfc}\,(z/\sqrt{2})$ and $\ln(\cdot)$ are also applied to vectors and matrices. The operator $R(\cdot)$ converts a complex-valued vector into a specific equivalent real-valued notation. For a given matrix $\boldsymbol{A} \in \mathbb{C}^{K \times M}$ the equivalent real-valued matrix $\boldsymbol{A}_\mathrm{r} = R(\boldsymbol{A})$ is given by

$$\boldsymbol{A}_\mathrm{r} = \begin{bmatrix} \mathrm{Re}\{a_{11}\} & -\mathrm{Im}\{a_{11}\} & \cdots & \mathrm{Re}\{a_{1M}\} & -\mathrm{Im}\{a_{1M}\} \\ \mathrm{Im}\{a_{11}\} & \mathrm{Re}\{a_{11}\} & \cdots & \mathrm{Im}\{a_{1M}\} & \mathrm{Re}\{a_{1M}\} \\ \vdots & & \ddots & & \vdots \\ \mathrm{Re}\{a_{K1}\} & -\mathrm{Im}\{a_{K1}\} & \cdots & \mathrm{Re}\{a_{KM}\} & -\mathrm{Im}\{a_{KM}\} \\ \mathrm{Im}\{a_{K1}\} & \mathrm{Re}\{a_{K1}\} & \cdots & \mathrm{Im}\{a_{KM}\} & \mathrm{Re}\{a_{KM}\} \end{bmatrix}. \tag{1-1}$$

For a given column vector $\boldsymbol{a} \in \mathbb{C}^M$ the equivalent real-valued vector $\boldsymbol{a}_\mathrm{r} = R(\boldsymbol{a})$ is given by

$$\boldsymbol{a}_\mathrm{r} = \begin{bmatrix} \mathrm{Re}\{[\boldsymbol{a}]_1\} & \mathrm{Im}\{[\boldsymbol{a}]_1\} & \cdots & \mathrm{Re}\{[\boldsymbol{a}]_M\} & \mathrm{Im}\{[\boldsymbol{a}]_M\} \end{bmatrix}^T. \tag{1-2}$$

The operator $C(\cdot)$ converts equivalent real-valued notation into complex-valued notation, meaning $C(\mathbf{A}_\mathrm{r}) = \mathbf{A}$. Finally, for the given vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, $\mathrm{P}(\boldsymbol{a} = \boldsymbol{b})$ denotes the probability of the event $\boldsymbol{a} = \boldsymbol{b}$.

2
Important Baselines

This chapter introduces some baselines considered in this study. The concepts presented in the following sections are necessary not only for the proper understanding of chapters 3 to 5 but also for localizing the contributions of the work.

2.1
The Minimum Distance to the Decisions Threshold

In the context of downlink transmission for multiuser MIMO systems, one of the most utilized concepts is the MDDT first introduced by [13]. It is generally used either for constructing a design criterion - by maximizing the MDDT leading to the maximum MDDT (MMDDT) concept [14, 15, 16, 17, 18] - or as a QoS constraint [19]. This section mathematically models the MDDTs of the k-th user as a function of the received signal considering a multiuser MIMO downlink system with K single-antenna users that expect to receive α_s-PSK data symbols. The derivation starts by denoting the data symbol for the k-th user as s_k and its noiseless received signal as y_k. With this, assuming hard detection, one can visualize the signal space for the k-th user in the left-hand side (LHS) of Fig. 2.1, where $\phi = \pi/\alpha_s$.

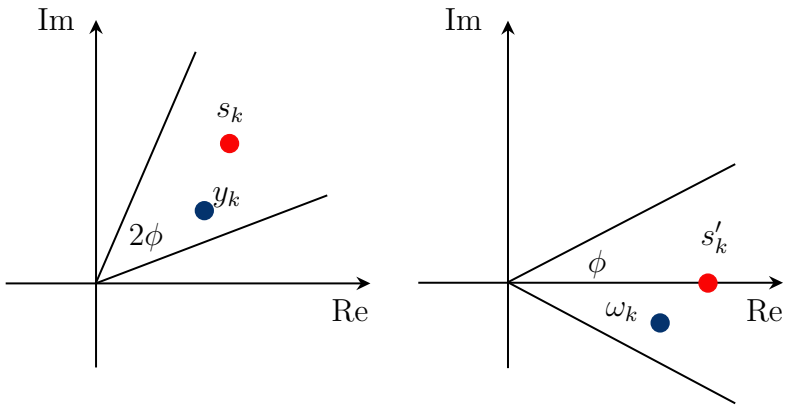

Figure 2.1: Signal space for the k-th user (right). Rotated coordinate system (left)

To compute the MDDT of the k-th user the first step is to consider a rotation by $\arg\{s_k^*\}$ of the coordinate system such that the symbol of interest

is placed on the real axis, as shown in the right-hand side (RHS) of Fig. 2.1. This is done by multiplying both the symbol of interest s_k and the noiseless received signal y_k by $e^{-j\arg\{s_k\}} = s_k^*$ which reads

$$s'_k = s_k s_k^* = 1, \quad \omega_k = y_k s_k^*. \tag{2-1}$$

Based on the rotated coordinate system two distances between the noiseless received signal and the decision's thresholds can be computed for each user, as shown in Fig. 2.2.

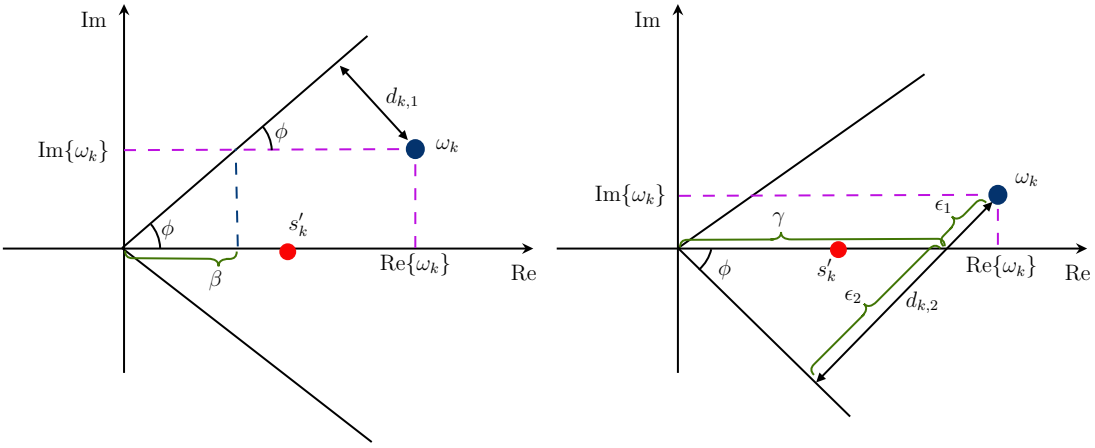

Figure 2.2: Distances to the Decisions' Threshold

The first distance, $d_{1,k}$, is computed based on the LHS of Fig. 2.2 as

$$\begin{aligned} d_{1,k} &= (\mathrm{Re}\{\omega_k\} - \beta)\sin(\phi) \\ &= \left(\mathrm{Re}\{\omega_k\} - \frac{\mathrm{Im}\{\omega_k\}}{\tan(\phi)}\right)\sin(\phi) \\ &= \mathrm{Re}\{\omega_k\}\sin(\phi) - \mathrm{Im}\{\omega_k\}\cos(\phi). \end{aligned}$$

The second distance of the $d_{2,k}$ is calculated based on the RHS of Fig. 2.2. It is easy to spot that

$$d_{2,k} = \epsilon_1 + \epsilon_2, \tag{2-2}$$

where $\epsilon_1 = \frac{\mathrm{Im}\{\omega_k\}}{\cos(\phi)}$ and $\epsilon_2 = \gamma\sin(\phi)$, with $\gamma = \mathrm{Re}\{\omega_k\} - \mathrm{Im}\{\omega_k\}\tan(\phi)$. Finally, the value of $d_{2,k}$ is given by

$$d_{2,k} = \mathrm{Re}\{\omega_k\}\sin(\phi) + \mathrm{Im}\{\omega_k\}\cos(\phi).$$

The smallest distance of the rotated symbol ω_k to the rotated decision

threshold is expressed as $d_k = \min_{\xi \in \{1,2\}} d_{\xi,k}$, which is summarized as

$$d_k = \mathrm{Re}\left\{s_k^* y_k\right\} \sin\phi - \left|\mathrm{Im}\left\{s_k^* y_k\right\}\right| \cos\phi. \tag{2-3}$$

Since the considered rotation also includes the decision thresholds, the distance expression in (2-3) also holds for y_k.

2.2 Full Branch-and-Bound Algorithm

The B&B, first created in 1960 by A. H. Land and A. G. Doig [20], is an established technique in the wireless communications area. It has important applications in this context, such as multiuser detection [21], discrete beamforming [22, 23] and, more recently, discrete precoding [17]. Although many different B&B methods exist in the literature [17, 18, 24, 25, 26, 27], this section presents the Full-B&B approach first proposed in [17] to optimally solve discrete programming problems (DPPs) with polynomial computational complexity. The Full-B&B method from [17] is a technique for solving problems in the form of

$$\begin{aligned} \boldsymbol{x}_{\mathrm{opt}} = \min_{\boldsymbol{x}} \ & g(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{x} \in \mathcal{X}^M, \end{aligned} \tag{2-4}$$

where $g : \mathbb{R}^M \to \mathbb{R}$ is a convex objective function, $\mathcal{X}$ is a discrete set with α_x elements and $\boldsymbol{x}_{\mathrm{opt}}$ is the optimal solution. The Full-B&B method consists of two stages, namely the initialization and the tree search-based stage.

2.2.1 Initialization Stage

The initialization step aims to compute a finite upper bound such that, in the tree search stage, many branches of the tree can be pruned early which is beneficial in terms of computational complexity. The initialization starts by constructing a lower bounding convex optimization problem on $g(\boldsymbol{x}_{\mathrm{opt}})$ by relaxing $\mathcal{X}^M$ to its convex hull $\mathcal{P}$ which yields

$$\begin{aligned} \boldsymbol{x}_{\mathrm{lb}} = \min_{\boldsymbol{x}} \ & g(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{x} \in \mathcal{P}. \end{aligned} \tag{2-5}$$

By solving (2-5) one obtains $\boldsymbol{x}_{\mathrm{lb}}$, if $\boldsymbol{x}_{\mathrm{lb}} \in \mathcal{X}^M$ then $\boldsymbol{x}_{\mathrm{opt}} = \boldsymbol{x}_{\mathrm{lb}}$ and the algorithm terminates with $\boldsymbol{x}_{\mathrm{lb}}$. Otherwise, an associated upper bound on

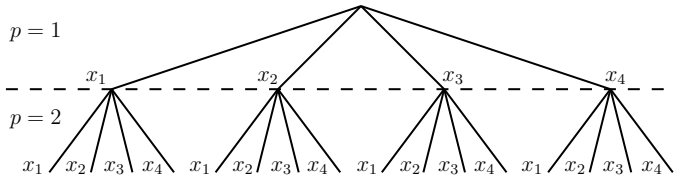

Figure 2.3: Tree representation of the set $\mathcal{X}^M$ for $M = 2$ and $\alpha_x = 4$

$g(\boldsymbol{x}_{\text{opt}})$ is obtained by projecting the solution of (2-5) to $\mathcal{X}^M$ and evaluating $g(\cdot)$ accordingly. In [17] the projection step is done via uniform quantization denoted by the operator $Q(\cdot)$. With this, the associated upper bound solution is given by $\boldsymbol{x}_{\text{ub}} = Q(\boldsymbol{x}_{\text{lb}})$ and smallest known upper bound, $\check{g}$, is initialized as $\check{g} = g(\boldsymbol{x}_{\text{ub}})$.

2.2.2 Tree Search-Based Stage

The Full-B&B method is a tree search-based algorithm. The tree represents the set of all possible solutions for the vector $\boldsymbol{x}$, i.e., represents the set $\mathcal{X}^M$. For structuring the tree M levels are considered and each node has one ingoing branch and α_x outgoing branches as shown in Fig. 2.3. The tree search-based stage starts at layer $p = 1$ by fixing p entries of $\boldsymbol{x}$. The vector $\boldsymbol{x}$ is then rewritten as $\boldsymbol{x} = [\boldsymbol{f}_i^T, \boldsymbol{v}^T]^T$, with $\boldsymbol{f}_i \in \mathcal{X}^p$. In the context of tree search the subvector $\boldsymbol{f}_i$ denotes the i-th branch of the layer p of the tree. Based on the fixed branch $\boldsymbol{f}_i$, a subproblem can be formulated as

$$\begin{aligned} \boldsymbol{v}_{\text{opt}|\boldsymbol{f}_i} = \min_{\boldsymbol{v}} \quad & g(\boldsymbol{v}, \boldsymbol{f}_i) \\ \text{s.t. } & \boldsymbol{v} \in \mathcal{X}^{M-p}. \end{aligned} \tag{2-6}$$

Relaxing the problem from (2-6) states

$$\begin{aligned} \boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i} = \min_{\boldsymbol{v}} \quad & g(\boldsymbol{v}, \boldsymbol{f}_i) \\ \text{s.t. } & \boldsymbol{v} \in \mathcal{J}, \end{aligned} \tag{2-7}$$

where $\mathcal{J}$ is the convex hull of $\mathcal{X}^{M-p}$. With $\boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i}$ a lower bound solution conditioned on $\boldsymbol{f}_i$ is obtained as $\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i} = [\boldsymbol{f}_i, \boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i}]$. An upper bound on $g(\boldsymbol{v}_{\text{opt}|\boldsymbol{f}_i}, \boldsymbol{f}_i)$ can be computed by projecting the vector $\boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i}$ to $\mathcal{X}^{M-p}$ resulting in $\boldsymbol{v}_{\text{ub}|\boldsymbol{f}_i} = Q(\boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i})$ and computing $g(\boldsymbol{x}_{\text{ub},i})$, with $\boldsymbol{x}_{\text{ub},i} = [\boldsymbol{f}_i, \boldsymbol{v}_{\text{ub}|\boldsymbol{f}_i}] \in \mathcal{X}^M$. The B&B algorithm benefits from having a small upper bound that allows for many branch exclusions. With this, for each $\boldsymbol{f}_i$ the algorithm updates $\check{g} = \min\left(g(\boldsymbol{x}_{\text{ub},i}), \check{g}\right)$. After the update, the algorithm proceeds by fixing the next branch $\boldsymbol{f}_{i+1}$.

Algorithm 1 Full B&B Algorithm [17]

initialization:
Solve problem (2-5) to get $\boldsymbol{x}_{\text{lb}}$
If $\boldsymbol{x}_{\text{lb}} \in \mathcal{X}^M \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{opt}} = \boldsymbol{x}_{\text{lb}}$
Compute $\boldsymbol{x}_{\text{ub}} = Q(\boldsymbol{x}_{\text{lb}})$ and initialize $\check{g} = g(\boldsymbol{x}_{\text{ub}})$
Define the first level ($p = 1$) of the tree by $\mathcal{G}_p := \mathcal{X}$
for $p = 1 : M - 1$ **do**
  Partition $\mathcal{G}_p$ in $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{|\mathcal{G}_p|}$
  for $i = 1 : |\mathcal{G}_p|$ **do**
    Conditioned on $\boldsymbol{f}_i$ solve (2-7) to get $\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i} = [\boldsymbol{f}_i, \boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i}]$
    Compute $\boldsymbol{v}_{\text{ub}|\boldsymbol{f}_i} = Q(\boldsymbol{v}_{\text{lb}|\boldsymbol{f}_i})$ and construct $\boldsymbol{x}_{\text{ub},i} = [\boldsymbol{f}_i, \boldsymbol{v}_{\text{ub}|\boldsymbol{f}_i}]$
    Update the smallest known upper bound with: $\check{g} = \min\left(\check{g}, g(\boldsymbol{x}_{\text{ub},i})\right)$
  end for
  Build the set $\mathcal{G}'_p := \left\{\boldsymbol{f}_i |\ g(\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i}) \leq \check{g}, i = 1, \ldots, |\mathcal{G}_p|\right\}$
  Define the set for the next level in the tree $\mathcal{G}_{d+1} := \mathcal{G}'_p \times \mathcal{X}$
end for
Partition $\mathcal{G}_M$ in $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{G}_M|}$
The global solution is $\boldsymbol{x}_{\text{opt}} = \min_{\boldsymbol{x}_i \in \mathcal{G}_M} g(\boldsymbol{x}_i)$

After all possible valid branches in a given layer are evaluated, i.e., all valid $\boldsymbol{f}_i$ were fixed and its conditioned upper and lower bounds computed the algorithm proceeds to the pruning step where the set of approved branches, $\mathcal{G}'_p$, in the current layer p is constructed. The pruning step proposed in [17] consists of excluding from the search set all $\boldsymbol{f}_i$ that cannot be a subvector from $\boldsymbol{x}_{\text{opt}}$. This is done by exploring the property that if $\boldsymbol{l}$ is lower bound solution on $\boldsymbol{x}_{\text{opt}}$ and $\boldsymbol{u}$ is an upper bound solution on $\boldsymbol{x}_{\text{opt}}$, then by definition $g(\boldsymbol{l}) \leq g(\boldsymbol{x}_{\text{opt}}) \leq g(\boldsymbol{u})$. With this, if $\boldsymbol{f}_i$ is a subvector of $\boldsymbol{x}_{\text{opt}}$ then $g(\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i}) \leq g(\boldsymbol{x}_{\text{opt}}) \leq \check{g}$. If, however, $\check{g} < g(\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i})$ then $\boldsymbol{f}_i$ cannot be a subvector of $\boldsymbol{x}_{\text{opt}}$ and it and all its evolutions are excluded from the search. Based on this, during the pruning step the approved set of branches in the p-th layer $\mathcal{G}'_p$ is constructed based on the following law $\mathcal{G}'_p = \{\boldsymbol{f}_i |\ g(\boldsymbol{x}_{\text{lb}|\boldsymbol{f}_i}) \leq \check{g}, \forall i\}$.

After pruning, the set of valid subvectors is updated and the algorithm repeats the process in the next layer. In the last layer, the global solution is computed as $\boldsymbol{x}_{\text{opt}} = \min_{\boldsymbol{x}_i \in \mathcal{G}_M} g(\boldsymbol{x}_i)$. The Full-B&B algorithm is detailed in Algorithm 1.

2.3 Linear MMSE Precoding

The Linear MMSE (LMMSE) precoder [28] is one of the most popular precoding approaches present in the literature due to its high performance with

low computational complexity. Although it is a relevant baseline of this work, the LMMSE setup implies some major differences compared to the techniques developed in this thesis. First, different from the SLP techniques proposed in this thesis, is a channel-level approach, which implies that the precoding matrix requires computation once per coherence time interval instead of a symbol-by-symbol computation. Moreover, the LMMSE approach considers an average total power constraint, which, according to [12], although mathematically tractable does not fully model the RF chain's hardware and the constraints imposed by it. Finally, the LMMSE technique can also be utilized with QAM signaling, when CSI is available at the receiver. In what follows the LMMSE precoding matrix is derived.

A MIMO system is considered which consists of a linear precoder with precoding matrix $\mathbf{P}$ at the transmitter and a linear equalizer represented by the matrix $\mathbf{G}$. The output signal of the detector is described by

$$\tilde{\mathbf{x}} = \mathbf{G}\left(\mathbf{HPx} + \mathbf{w}\right), \tag{2-8}$$

where $\mathbf{H}$ is the MIMO channel, $\mathbf{x}$ is the input signal and $\mathbf{w}$ is the additive noise. The input signal $\mathbf{x}$ has the covariance matrix $\mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{C}_\mathbf{x}$. The noise $\mathbf{w}$ has the covariance matrix $\mathrm{E}\left\{\mathbf{w}\mathbf{w}^H\right\} = \mathbf{C}_\mathbf{w}$. Because the transmit energy is constrained, it is considered that the received signal is scaled with factor f at the receiver, which is part of the optimization. The MMSE precoder problem under an average total power constraint reads as in the following

$$\begin{aligned} \mathbf{P}_{\mathrm{MMSE}} = \arg\min_{\mathbf{P},\mathrm{f}} \mathrm{E}\left\{\|\mathbf{x} - \mathrm{f}\tilde{\mathbf{x}}\|_2^2\right\} \\ \text{s.t.} \;\; \mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\} \leq E_{\mathrm{tx}}, \end{aligned} \tag{2-9}$$

where E_{tx} denotes the transmit energy. The Lagrangian function reads as

$$\begin{aligned} L\left(\mathbf{P}, \mathrm{f}, \lambda\right) =& \mathrm{E}\left\{\|\mathbf{x} - \mathrm{f}\tilde{\mathbf{x}}\|_2^2\right\} + \lambda\left(\mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\} - E_{\mathrm{tx}}\right) \\ =& \mathrm{tr}\left\{\mathbf{C}_\mathbf{x}\right\} - \mathrm{f}\ \mathrm{E}\left\{\mathbf{x}^H\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{x}\right\} - \mathrm{f}\ \mathrm{E}\left\{\mathbf{x}^H\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H\mathbf{x}\right\} \\ &+ \mathrm{f}^2\ \mathrm{E}\left\{\mathbf{x}^H\mathbf{P}^H\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{x}\right\} \\ &+ \mathrm{f}^2\ \mathrm{E}\left\{\mathbf{w}^H\mathbf{G}^H\mathbf{G}\mathbf{w}\right\} + \lambda\left(\mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\} - E_{\mathrm{tx}}\right), \end{aligned} \tag{2-10}$$

where λ denotes the Lagrangian multiplier. By making use of the trace operator

and its properties the Lagrangian function can be rewritten as follows

$$\begin{aligned} L\left(\mathbf{P}, \mathrm{f}, \lambda\right) =& \operatorname{tr}\left\{\mathbf{Cx}\right\} - \mathrm{f}\ \operatorname{tr}\left\{\mathbf{GHPCx}\right\} - \mathrm{f}\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C_x}\mathbf{P}^H\right\} \\ &+ \mathrm{f}^2\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{GHPCxP}^H\right\} + \mathrm{f}^2\ \operatorname{tr}\left\{\mathbf{GCwG}^H\right\} \\ &+ \lambda\left(\operatorname{tr}\left\{\mathbf{PC_xP}^H\right\} - E_{\mathrm{tx}}\right). \end{aligned} \tag{2-11}$$

Taking the derivative with respect to $\mathbf{P}^*$ yields

$$\frac{\partial L\left(\mathbf{P}, \mathrm{f}, \lambda\right)}{\partial \mathbf{P}^*} = -\ \mathrm{f}\ \mathbf{H}^H\mathbf{G}^H\mathbf{Cx} + \mathrm{f}^2\ \mathbf{H}^H\mathbf{G}^H\mathbf{GHPCx} + \lambda\ \mathbf{PC_x}. \tag{2-12}$$

Equating (2-12) to zero yields

$$\frac{1}{\mathrm{f}}\mathbf{H}^H\mathbf{G}^H = \left(\mathbf{H}^H\mathbf{G}^H\mathbf{GH} + \frac{\lambda}{\mathrm{f}^2}\mathbf{I}\right)\mathbf{P}, \tag{2-13}$$

such that $\mathbf{P}_{\mathrm{MMSE}}$ has the structure

$$\mathbf{P} = \frac{1}{\mathrm{f}}\left(\mathbf{H}^H\mathbf{G}^H\mathbf{GH} + \frac{\lambda}{\mathrm{f}^2}\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{G}^H. \tag{2-14}$$

Taking the derivative of (2-11) with respect to f yields

$$\begin{aligned} \frac{\partial L\left(\mathbf{P}, \mathrm{f}, \lambda\right)}{\partial \mathrm{f}} =& -\operatorname{tr}\left\{\mathbf{GHPC_x}\right\} - \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C_x}\mathbf{P}^H\right\} \\ &+ 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{GHPC_xP}^H\right\} + 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{GC_wG}^H\right\} \\ =& -2\operatorname{Re}\left\{\operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C_x}\mathbf{P}^H\right\}\right\} + 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{GHPC_xP}^H\right\} \\ &+ 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{GC_wG}^H\right\}. \end{aligned} \tag{2-15}$$

Equating (2-15) to zero gives

$$2\operatorname{Re}\left\{\operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C_x}\mathbf{P}^H\right\}\right\} = 2\mathrm{f}\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{GHPC_xP}^H\right\} + 2\mathrm{f}\ \operatorname{tr}\left\{\mathbf{GC_wG}^H\right\}. \tag{2-16}$$

Because of the structure of $\mathbf{P}_{\mathrm{MMSE}}$ in (2-14), (2-16) can be written without the real part operator, denoted by

$$2\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C_x}\mathbf{P}^H\right\} = 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{GHPC_xP}^H\right\} + 2\ \mathrm{f}\ \operatorname{tr}\left\{\mathbf{GC_wG}^H\right\}. \tag{2-17}$$

Multiplying from the right $2\ \mathrm{f}\ \mathbf{C_xP}^H$ in (2-13) and using the trace operator

yields

$$\begin{aligned} 2\ \mathrm{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\} =& 2\ \mathrm{f}\ \mathrm{tr}\left\{\left(\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}+\frac{\lambda}{\mathrm{f}^2}\mathbf{I}\right)\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\} \\ =& 2\ \mathrm{f}\ \mathrm{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}+2\ \mathrm{f}\frac{\lambda}{\mathrm{f}^2}\ \mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}. \end{aligned} \tag{2-18}$$

Putting together the RHS of (2-17) and the RHS of (2-18) yields

$$\begin{aligned} &2\ \mathrm{f}\ \mathrm{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}+2\ \mathrm{f}\ \mathrm{tr}\left\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\right\} \\ &=2\ \mathrm{f}\ \mathrm{tr}\left\{\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}+2\ \mathrm{f}\frac{\lambda}{\mathrm{f}^2}\ \mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}, \end{aligned} \tag{2-19}$$

which can be rearranged to

$$\frac{\lambda}{\mathrm{f}^2}=\frac{\mathrm{tr}\left\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\right\}}{\mathrm{tr}\left\{\mathbf{P}\mathbf{C}_\mathbf{x}\mathbf{P}^H\right\}}=\frac{\mathrm{tr}\left\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\right\}}{E_{\mathrm{tx}}}, \tag{2-20}$$

where it is considered that the transmit energy constraint holds with equality (in this problem it is obvious that more transmit energy helps to reduce the MSE). Substituting the diagonal loading in $\frac{\lambda}{\mathrm{f}^2}$ in (2-14) yields

$$\mathbf{P}_{\mathrm{MMSE}}=\frac{1}{\mathrm{f}}\left(\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}+\frac{\mathrm{tr}\left\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\right\}}{E_{\mathrm{tx}}}\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{G}^H. \tag{2-21}$$

Substituting (2-21) in the transmit energy constraint in (2-9) yields

$$E_{\mathrm{tx}}=\frac{1}{\mathrm{f}^2}\mathrm{tr}\left\{\left(\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}+\frac{\mathrm{tr}\left\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\right\}}{E_{\mathrm{tx}}}\mathbf{I}\right)^{-2}\mathbf{H}^H\mathbf{G}^H\mathbf{C}_\mathbf{x}\mathbf{G}\mathbf{H}\right\}, \tag{2-22}$$

which gives the MMSE scaling factor

$$\mathrm{f}=\sqrt{\frac{\mathrm{tr}\left\{\left(\mathbf{H}^H\mathbf{G}^H\mathbf{G}\mathbf{H}+\frac{\mathrm{tr}\{\mathbf{G}\mathbf{C}_\mathbf{w}\mathbf{G}^H\}}{E_{\mathrm{tx}}}\mathbf{I}\right)^{-2}\mathbf{H}^H\mathbf{G}^H\mathbf{C}_\mathbf{x}\mathbf{G}\mathbf{H}\right\}}{E_{\mathrm{tx}}}}. \tag{2-23}$$

Finally, the transmit signal is obtained by multiplying the input signal $\boldsymbol{x}$ by the precoding matrix $\mathbf{P}_{\mathrm{MMSE}}$.

3
Symbol-Level Precoding under a Strict Per Antenna Power Constraint with Conventional MIMO Transmitter

For MU-MIMO systems a fundamental problem is the design of low-complexity precoding algorithms that attain the high reliability constraints of future wireless communications networks. Linear techniques, such as zero-forcing (ZF) and matched filtering [29, 30], have been proposed in the literature. However, when considering linear precoding, an established assumption in the literature [28, 31] is that the transmit symbols are constrained by an average total power constraint (TPC). This yields a system that is easier to model, yet, according to [12], in a realistic scenario each BS antenna is connected to its own PA and thus has to meet its specific power constraints.

With this, several precoding techniques arose considering per antenna power constraints (PAPC). Linear channel-level precoding strategies considering an average PAPC are well studied in the literature [32, 33, 34, 35]. However, according to [36], the consideration of a SPAPC yields a more realistic scenario since the transmit power at each antenna is upper bounded by a threshold to avoid severe distortion at the PA due to clipping. With this, different linear precoding techniques have been developed considering SPAPCs [36, 37]. More recently, the SLP strategy has been receiving increasing attention since it allows for a higher degree of reliability. In what follows the system model considered for this chapter is exposed and a brief revision of the SLP contributions considering SPAPCs is provided.

3.1
System Model

The system model consists of a single-cell MU-MIMO scenario where the BS is equipped with M transmit antennas serving K single antenna users. A symbol level transmission is considered where s_k represents the data symbol to be delivered for the k-th user. Each symbol s_k is considered to belong to the set $\mathcal{S}$ that represents all possible symbols of a α_s-PSK modulation and is given by

$$\mathcal{S} = \left\{ s : s = e^{\frac{j\pi(2i+1)}{\alpha_s}}, \text{ for } i = 1, \ldots, \alpha_s \right\}. \tag{3-1}$$

The symbols of all users are described in a stacked vector notation as $\boldsymbol{s} = [s_1, \ldots, s_K]^T \in \mathcal{S}^K$. Based on $\boldsymbol{s}$ the precoder computes the transmit vector $\boldsymbol{x} = [x_1, \ldots, x_M]^T$. The entries of $\boldsymbol{x}$ are constrained by a SPAPC, meaning $|x_m|^2 \leq \mathrm{P_A}$ for $m \in \{1, \ldots, M\}$, where $\mathrm{P_A}$ represents the maximum per antenna transmit power. A frequency flat fading channel described by the matrix $\boldsymbol{H} \in \mathbb{C}^{K \times M}$ is considered. The BS is considered to receive the CSI coefficients from the users which correspond to the matrix $\tilde{\boldsymbol{H}} \in \mathbb{C}^{K \times M}$, which implies spatial correlation $\mathrm{E}\{\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}\} = K\boldsymbol{R}_\mathrm{s}$. It is considered that the spatial correlation matrix $\boldsymbol{R}_\mathrm{s}$ can be estimated and is known at the BS. It is considered that the spatial correlation is modeled by the Kronecker model [38], which implies that the entries of $\boldsymbol{R}_\mathrm{s}$ are in the form of $[r_\mathrm{s}]_{i,j} = \rho^{(i-j)^2}$ for $(i,j) \in \{1, \ldots, M\}^2$. The factor $\rho \in [0,1]$ is the correlation index of neighboring antennas. In this chapter, the channel model is described by

$$\boldsymbol{H} = \boldsymbol{N}\tilde{\boldsymbol{H}} + \sqrt{\boldsymbol{I} - \boldsymbol{N}^2}\boldsymbol{\Psi}\boldsymbol{R}_\mathrm{s}^{\frac{1}{2}}. \tag{3-2}$$

The matrix $\boldsymbol{N} = \mathrm{diag}\,(\boldsymbol{\eta})$, with $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_K] \in [0,1]^K$ describes the user-specific quality of the CSI which can also be interpreted as the temporal correlation factor. It is considered that $\boldsymbol{N}$ can be estimated and is known at the BS. The matrix $\boldsymbol{\Psi}$, with $\boldsymbol{\psi}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ being the k-th row of $\boldsymbol{\Psi}$ for $k \in \{1, \ldots K\}$, describes the random part of the channel model. The received signal for all users $\boldsymbol{z}$ can be described as

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \tag{3-3}$$

with its k-th entry z_k being the received signal from the k-th user. The vector $\boldsymbol{w} \sim \mathcal{CN}\,(0, \sigma_w^2 \boldsymbol{I})$ represents additive white Gaussian noise (AWGN). Each received symbol z_k is detected as $\hat{s}_k = D(z_k)$ where $\hat{s}_k$ denotes the detected symbol for the k-th user and $D(\cdot)$ the hard detection operation.

3.2 Literature Review

As mentioned the SLP strategy has become popular due to its high degree of reliability. In this section, we revise the works from [39] and [40]. In [39] SLP is considered with a SPAPC and two novel strategies based on the concept of *strict* and *non-strict rotation* for constructive interference (CI) based precoding are proposed. Moreover, [39] also proposes a ZF design for the considered system model. In [40] the MMDDT SLP under a SPAPC is written as a second-order cone program (SOCP) and solved via the primal-dual interior-points method (IPM). Both works considered in this literature review

rely on perfect CSI meaning $\boldsymbol{H} = \tilde{\boldsymbol{H}}$ and consider no spatial correlation, i.e., $\boldsymbol{R}_{\mathrm{s}} = \boldsymbol{I}$.

3.2.1 Zero-Forcing SPAPC Design

The ZF criterion is based on eliminating the interference considered harmful for detection. As proposed in [39], the SLP ZF under a SPAPC can be designed by imposing the ZF constraint and scaling it to satisfy the SPAPC. With this, the closed-form solution for the ZF-SPAPC precoding matrix is given as follows

$$\boldsymbol{P} = \frac{\sqrt{\mathrm{P_A}}\boldsymbol{H}^{\dagger}}{\max\limits_{m\in\{1,\ldots,M\}} \left|\left[\boldsymbol{H}^{\dagger}\boldsymbol{s}\right]\right|_m} \tag{3-4}$$

where $\boldsymbol{H}^{\dagger}$ is Moore Penrose pseudo-inverse of the matrix $\boldsymbol{H}$. After computing $\boldsymbol{P}$ the vector $\boldsymbol{x}$ is computed as $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{s}$.

3.2.2 Constructive Interference Designs

This section presents the state-of-the-art formulations based on constructive interference as well as the methods proposed for solving them. The section starts by introducing the formulations and then proceeds to the exposure of the techniques utilized for solving the related optimization problems.

3.2.2.1 Constructive Interference Strict Phase Rotation Formulation

The strict phase rotation design consists of exploiting the multiuser interference to increase the amplitude of the desired signal at the detector's side. To this end, the phases of the interfering signals are controlled and rotated such that they are strictly aligned to those of the data symbols of interest [41]. The SLP optimization problem of strict phase rotation under a SPAPC, proposed in [39], is cast as

$$\begin{aligned} \max_{\boldsymbol{x},\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & \boldsymbol{h}_k\boldsymbol{x} = \lambda_k s_k, \quad \text{for } k \in \{1,\ldots,K\} \\ & \lambda_k \geq \epsilon, \quad \text{for } k \in \{1,\ldots,K\} \\ & |x_m|^2 \leq \mathrm{P_A}, \quad \text{for } m \in \{1,\ldots,M\}, \end{aligned} \tag{3-5}$$

where $\boldsymbol{h}_k$ denotes the k-th row of the channel matrix $\boldsymbol{H}$, and ϵ is an optimization variable related to (3-5) being written in the epigraph form [42, Section 4.1.3]. Note that, the constraint $\boldsymbol{h}_k\boldsymbol{x} = \lambda_k s_k$ imposes the k-th user noiseless received signal to be a scaled version of the desired signal with scaling factor λ_k. Moreover, due to the constraints $\lambda_k \geq \epsilon$, for $k \in \{1, \ldots, K\}$, when maximizing ϵ one is maximizing the smallest λ_k which is then beneficial for detection since it leads to a larger noiseless received signal.

3.2.2.2 MMDDT Formulation

The MMDDT formulation is one of the most prominent design criteria in the SLP literature due to its asymptotically optimal SEP performance with the SNR increase. The MMDDT criterion has several different designations in the literature. The works from [39, 41] call it the Non-Strict Phase Rotation criterion. In [12, 25, 43, 44] it is called the constructive interference (CI) criterion. Finally, [14, 45], name it the maximum safety margin (MSM) criterion. All the mentioned designations express the same idea of maximizing the smallest MDDT of all users, denoted by ϵ in what follows.

To mathematically determine ϵ we start from the baseline from Section 2.1, where the MDDT from the k-th user was derived as

$$d_k = \mathrm{Re}\left\{s_k^* y_k\right\} \sin\phi - \left|\mathrm{Im}\left\{s_k^* y_k\right\}\right| \cos\phi, \tag{3-6}$$

with y_k denoting the noiseless received signal of the k-th user and $\phi = \pi/\alpha_s$. In the system model from this section $y_k = \boldsymbol{h}_k\boldsymbol{x}$, which implies that the MDDT for the k-th user is expressed as

$$d_k = \mathrm{Re}\left\{s_k^* \boldsymbol{h}_k\boldsymbol{x}\right\} \sin\phi - \left|\mathrm{Im}\left\{s_k^* \boldsymbol{h}_k\boldsymbol{x}\right\}\right| \cos\phi, \tag{3-7}$$

By definition, ϵ is the smallest MDDT between all users. With this, for the considered system model it is defined as

$$\begin{aligned} \epsilon &= \min_{k\in\{1,\ldots,K\}} d_k \\ &= \min_{k\in\{1,\ldots,K\}} \mathrm{Re}\left\{s_k^* \boldsymbol{h}_k\boldsymbol{x}\right\} \sin\phi - \left|\mathrm{Im}\left\{s_k^* \boldsymbol{h}_k\boldsymbol{x}\right\}\right| \cos\phi. \end{aligned} \tag{3-8}$$

Using (3-8) as the objective function of a SLP problem under an SPAPC yields

$$\begin{aligned} \left[\boldsymbol{x}_{\mathrm{opt}},\, \epsilon_{\,\mathrm{opt}}\right] = &\max_{\boldsymbol{x},\epsilon} \min_{k\in\{1,\ldots,K\}} \mathrm{Re}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\} \sin\phi - \left|\mathrm{Im}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}\right| \cos\phi \\ &\text{s.t.} \quad |x_m|^2 \leq \mathrm{P_A}, \quad m \in \{1, \ldots, M\}. \end{aligned} \tag{3-9}$$

3.2.3
Symbol-Level Precoding State-of-the-Art approaches under a SPAPC

This section presents the SLP contributions devised based on the constructive interference criteria (i.e., Strict Phase Rotation and MMDDT) under a SPAPC. The main contributions for CI precoding under a SPAPC are proposed in the studies from [39, 40], of which the results relevant to this thesis are detailed in what follows.

3.2.3.1
Projected Gradient for CI Precoding

The work from [39] proposes a projected gradient method for solving the Strict Phase Rotation and MMDDT problems with reduced time complexity compare with standard optimization tools, e.g., CVX. Following the path taken by the author, we start the exposure of the work from [39] by the Strict Phase Rotation problem. Departing from (3-5) the author starts by defining $\boldsymbol{f}_k = {}^{\boldsymbol{h}_k}/_{s_k^*}$, for $k \in \{1, \ldots, K\}$, hence the equality constraints in (3-5) can be expressed as $\boldsymbol{f}_k^H \boldsymbol{x} = \lambda_k$, $\forall k \in \{1, \ldots, K\}$. Since λ_k is real-valued, $\mathrm{Re}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} = \lambda_k$ and $\mathrm{Im}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} = 0$. The optimization problem in (3-5) is rewritten in epigraph form as

$$\begin{aligned} \max_{\boldsymbol{x},\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & \mathrm{Re}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} \geq \epsilon, \quad \text{for } k \in \{1, \ldots, K\} \\ & \mathrm{Im}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} = 0, \quad \text{for } k \in \{1, \ldots, K\} \\ & |x_m|^2 \leq \mathrm{P_A}, \quad \text{for } m \in \{1, \ldots, M\}. \end{aligned} \tag{3-10}$$

Such that the problem is written with real-valued variables the author defines $\boldsymbol{f}_{\mathrm{R},k} = \mathrm{Re}\{\boldsymbol{f}_k\}$, $\boldsymbol{f}_{\mathrm{I},k} = \mathrm{Im}\{\boldsymbol{f}_k\}$, $\tilde{\boldsymbol{x}} = [\mathrm{Re}\{\boldsymbol{x}\}^T, \mathrm{Im}\{\boldsymbol{x}\}^T]^T$, which yields the following optimization problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{x}}} \max_{k \in 1,\ldots,K} \quad & -\tilde{\boldsymbol{b}}_k^T \tilde{\boldsymbol{x}} \\ \text{s.t.} \quad & \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} = \boldsymbol{0}, \quad \tilde{x}_m^2 + \tilde{x}_{M+m}^2 \leq \mathrm{P_A}, \text{ for } m \in \{1, \ldots, M\}, \end{aligned} \tag{3-11}$$

where $\tilde{\boldsymbol{b}}_k = [\boldsymbol{f}_{\mathrm{R},k}^T, \boldsymbol{f}_{\mathrm{I},k}^T]^T$ and

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} -[\boldsymbol{f}_{\mathrm{I}}]_1 & \cdots & -[\boldsymbol{f}_{\mathrm{I}}]_K \\ [\boldsymbol{f}_{\mathrm{R}}]_1 & \cdots & [\boldsymbol{f}_{\mathrm{R}}]_K \end{bmatrix}^T. \tag{3-12}$$

To achieve a smooth objective the author considers the log-sum-exp approximation which yields the following problem

$$\begin{aligned} &\min_{\tilde{\boldsymbol{x}}} \ \gamma \ln\left(\sum_{k=1}^{K} e^{(-\tilde{\boldsymbol{b}}_k^T \tilde{\boldsymbol{x}}/\gamma)}\right) \\ &\text{s.t.} \ \ \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} = \boldsymbol{0}, \quad \tilde{x}_m^2 + \tilde{x}_{M+m}^2 \leq \mathrm{P_A}, \ \text{for } m \in \{1,\ldots,M\}. \end{aligned} \tag{3-13}$$

From this point, the author utilizes a standard projected gradient algorithm, which requires the gradient of the objective and the projection step. To the scope of this thesis, it is not necessary to do the complete exposure of the projected gradient method from [39] being sufficient to state that the developed algorithm optimally solves (3-13) with complexity order of $\mathcal{O}(M^3 + I_{\text{iter}} I_{\text{PG}} I_{\text{ADMM}} M^2)$, where I_{iter} is the number of iterations of the algorithm and $I_{\text{PG}} I_{\text{ADMM}}$ is the number of iterations required for projection.

For the MMDDT design, similar steps are done departing (3-9), which yields

$$\begin{aligned} \max_{\boldsymbol{x},\epsilon} \ \ & \epsilon \\ \text{s.t.} \ \ & \mathrm{Re}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} - \psi \mathrm{Im}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} \geq \epsilon, \ \ \text{for } k \in \{1,\ldots,K\} \\ & \mathrm{Re}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} + \psi \mathrm{Im}\left\{\boldsymbol{f}_k^H \boldsymbol{x}\right\} \geq \epsilon, \ \ \text{for } k \in \{1,\ldots,K\} \\ & |x_m|^2 \leq \mathrm{P_A}, \quad \text{for } m \in \{1,\ldots,M\}, \end{aligned} \tag{3-14}$$

where $\psi = \tan(\pi/\alpha_s)^{-1}$. Introducing $\tilde{\boldsymbol{\Theta}} = [\tilde{\boldsymbol{\theta}}_1,\ldots,\tilde{\boldsymbol{\theta}}_{2K}]$, where $\tilde{\boldsymbol{\theta}}_k = [\boldsymbol{f}_{\mathrm{R},k}^T + \psi \boldsymbol{f}_{\mathrm{I},k}^T, \boldsymbol{f}_{\mathrm{I},k}^T - \psi \boldsymbol{f}_{\mathrm{R},k}^T]^T$ for $k \in \{1,\ldots,K\}$ and $\tilde{\boldsymbol{\theta}}_k = [\boldsymbol{f}_{\mathrm{R},k-K}^T - \psi \boldsymbol{f}_{\mathrm{I},k-K}^T, \boldsymbol{f}_{\mathrm{I},k-K}^T + \psi \boldsymbol{f}_{\mathrm{R},k-K}^T]^T$ for $k \in \{K+1,\ldots,2K\}$, the problem can be rewritten as

$$\begin{aligned} &\min_{\tilde{\boldsymbol{x}}} \ \gamma \ln\left(\sum_{k=1}^{2K} e^{(-\tilde{\boldsymbol{\theta}}_k^T \tilde{\boldsymbol{x}}/\gamma)}\right) \\ &\text{s.t.} \ \ \tilde{x}_m^2 + \tilde{x}_{M+m}^2 \leq \mathrm{P_A}, \ \text{for } m \in \{1,\ldots,M\}. \end{aligned} \tag{3-15}$$

From this point, the author utilizes a standard projected gradient algorithm, which optimally solves the problem with complexity order of $\mathcal{O}(I_{\text{iter}} MK + I_{\text{iter}} I_{\text{PG}} M)$. The lack of proper evaluation of the growth of the parameters I_{iter}, I_{PG} and I_{ADMM} with system size hinders a proper comparison of the computational complexity of the method proposed in [39] with the other approaches.

3.2.3.2 MMDDT SPAPC problem in the SOCP Standard Form

The study from [40] formulates the optimization problem described in (3-9) as a SOCP in standard form. Note that this is beneficial since, although some tools (e.g., CVX) accept complex-valued functions, they generally apply some preprocessing such that the optimization problem is in SOCP standard form. Directly making the problem a SOCP in standard form reduces the preprocessing required and thus the computational complexity of solving the optimization problem. Introducing the matrix $\boldsymbol{H}_{s^*} = \mathrm{diag}(\boldsymbol{s}^*)\boldsymbol{H}$, one can write the (3-9) as a minimization problem as

$$\begin{aligned} \left[\boldsymbol{x}_{\mathrm{opt}},\, \epsilon_{\,\mathrm{opt}}\right] = \min_{\boldsymbol{x},\epsilon} \; & -\epsilon \\ \text{s.t.} \;\; & \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\phi - \left|\mathrm{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\right|\cos\phi \geq \epsilon\mathbf{1} \\ & |x_m|^2 \leq \mathrm{P_A}, \quad m \in \{1,\ldots,M\}\,. \end{aligned} \tag{3-16}$$

Applying standard optimization problem algorithms in general requires all involved functions to be convex and twice continuously differentiable. Unfortunately, the MDDT constraint, $\mathrm{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\phi - \left|\mathrm{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\right|\cos\phi \geq \epsilon\mathbf{1}$, does not meet the differentiability condition due to the absolute value. This, however, is easily solved by including a new set of constraints which reads as

$$\begin{aligned} \left[\boldsymbol{x}_{\mathrm{opt}},\, \epsilon_{\,\mathrm{opt}}\right] = \min_{\boldsymbol{x},\epsilon} \; & -\epsilon \\ \text{s.t.} \;\; & \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\phi - \mathrm{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\cos\phi \geq \epsilon\mathbf{1} \\ & \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\phi + \mathrm{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\cos\phi \geq \epsilon\mathbf{1} \\ & |x_m|^2 \leq \mathrm{P_A}, \quad m \in \{1,\ldots,M\}\,. \end{aligned} \tag{3-17}$$

The problem above can be reformulated with real-valued variables by introducing the new optimization variable $\boldsymbol{u} = \left[\epsilon, \boldsymbol{x}_{\mathrm{r}}^T\right]^T$, with $\boldsymbol{x}_{\mathrm{r}} = R(\boldsymbol{x})$, and reformulating the problem accordingly. With this, the equivalent real-valued optimization problem reads as

$$\begin{aligned} \boldsymbol{u}_{\mathrm{opt}} = & \arg\min_{\boldsymbol{u}} \; \boldsymbol{a}^T\boldsymbol{u} \\ \text{s.t.} \;\; & \boldsymbol{B}\boldsymbol{u} \leq \mathbf{0}, \\ & \|\boldsymbol{C}_{\mathrm{m}}\boldsymbol{u}\|_2 \leq \sqrt{\mathrm{P_A}}, \quad \text{for } m \in \{1,\ldots,M\}\,, \end{aligned} \tag{3-18}$$

where

$$\boldsymbol{a} = [-1, \mathbf{0}^T]^T, \quad \boldsymbol{B} = \left[\mathbf{1}, \boldsymbol{\Theta}_{\mathrm{r}}\right], \tag{3-19}$$

$$\boldsymbol{\Theta}_{\mathrm{r}} = \left[\boldsymbol{\gamma}_1^T, \boldsymbol{\lambda_1}^T, \cdots, \boldsymbol{\gamma}_K^T, \boldsymbol{\lambda}_K^T, \boldsymbol{\psi}_1^T, \boldsymbol{\delta}_1^T, \cdots, \boldsymbol{\psi}_K^T, \boldsymbol{\delta}_K^T\right], \tag{3-20}$$

$$\boldsymbol{C}_{\mathrm{m}} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0}^T & \boldsymbol{D}_{\mathrm{m}} \end{bmatrix}, \quad \boldsymbol{D}_{\mathrm{m}} = \mathrm{diag}\left(\boldsymbol{d}_{\mathrm{m}}\right), \tag{3-21}$$

where $\boldsymbol{d}_{\mathrm{m}} \in \mathbb{R}^{2M\times 1}$ being a vector of zeros with ones at entries $2m-1$ and $2m$, and $\boldsymbol{\gamma}_k$, $\boldsymbol{\lambda}_k$, $\boldsymbol{\psi}_k$ and $\boldsymbol{\delta}_k$ are the k-th row of the matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Delta}$, which are given by

$$\begin{aligned} \boldsymbol{\Gamma} = &\quad \mathrm{Im}\left\{\boldsymbol{H}_{s^*}\right\}\cos(\phi) - \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\right\}\sin(\phi) \\ \boldsymbol{\Lambda} = &\quad \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\right\}\cos(\phi) + \mathrm{Im}\left\{\boldsymbol{H}_{s^*}\right\}\sin(\phi) \\ \boldsymbol{\Psi} = &-\mathrm{Im}\left\{\boldsymbol{H}_{s^*}\right\}\cos(\phi) - \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\right\}\sin(\phi) \\ \boldsymbol{\Delta} = &\quad \mathrm{Im}\left\{\boldsymbol{H}_{s^*}\right\}\sin(\phi) - \mathrm{Re}\left\{\boldsymbol{H}_{s^*}\right\}\cos(\phi). \end{aligned} \tag{3-22}$$

The problem described in (3-18) is a SOCP cf.[42, Sec. 4.4.2] and can be readily solved with IPM. The optimal solution can be converted back to complex-valued notation by extracting $\boldsymbol{x}_{\mathrm{r}}$ of $\boldsymbol{u}_{\mathrm{opt}}$ and applying $\boldsymbol{x} = C(\boldsymbol{x}_{\mathrm{r}})$. The study from [40] utilizes the primal-dual IPM for solving the (3-18) with upper bound complexity order (UBCO) of $\mathcal{O}\left(M^{3.5}\log\left({}^{M}/_{\epsilon_{\mathrm{tol}}}\right)\right)$, where ϵ_{tol} denotes the optimality tolerance.

3.3 Contributions of this chapter

Section 3.2 shows some of the most prominent design concepts present in the literature and their related precoding algorithms. Besides the aforementioned concepts, one of the most prominent design criteria in the literature is the MMSE [46, 47]. The MMSE utilization ranges from the established channel-level linear precoding strategy presented in [28] to an SLP design considering coarse quantization [27].

In this context, by considering a SPAPC this chapter proposes two SLP techniques for PSK modulation. While the first method utilizes the MMSE criterion considering perfect CSI at the transmitter, the second approach allows for imperfect CSI scenarios by exploiting knowledge about second-order statistics of the CSI mismatch. The proposed approaches are formulated in the SOCPs form and are readily solved with polynomial complexity using the IPM. Numerical results indicate that the proposed MMSE methods are superior to the existing techniques in terms of bit-error-rate (BER) for the low

and medium SNR regimes. Moreover, regarding CSI imperfection the proposed RMMSE design outperforms the examined SPAPC state-of-the-art algorithms for all values of CSI mismatch.

3.4 Proposed MMSE Precoding Designs under a Strict Per Antenna Power Constraint

In this section, we propose SLP designs based on the MMSE objective under an SPAPC for two different scenarios. In the first scenario, perfect CSI is considered, meaning $\tilde{\boldsymbol{H}} = \boldsymbol{H}$. In the second scenario, it is considered that the BS has imperfect CSI and knowledge about the matrices $\tilde{\boldsymbol{H}}$, $\boldsymbol{N}$ and $\boldsymbol{R}_{\mathrm{s}}$. The MMSE objective, similar to as proposed in [28], can be utilized under a SPAPC with the following problem

$$\begin{aligned} &\min_{\boldsymbol{x},\beta} \mathrm{E}\left\{\|\beta\boldsymbol{z} - \boldsymbol{s}\|_2^2\right\} \\ &\text{s.t. } |x_m|^2 \leq \mathrm{P_A}, \text{ for } \; m \in \{1,\ldots,M\}, \; \beta \geq 0. \end{aligned} \tag{3-23}$$

Note that the real-valued factor β represents a theoretical automatic gain control which is part of the established MMSE objective as proposed in [28]. The factor β is computed by the precoder alongside the transmit vector $\boldsymbol{x}$. Yet, since in this study PSK modulation is considered, knowledge of β is not required for hard detection.

3.4.1 Proposed MMSE SPAPC Design

The MMSE optimization problem from (3-23) can be rewritten by substituting $\boldsymbol{z}$ in the objective which yields

$$\begin{aligned} &\min_{\boldsymbol{x},\beta} \mathrm{E}\{\|\beta\boldsymbol{H}\boldsymbol{x} + \beta\boldsymbol{w} - \boldsymbol{s}\|_2^2\} \\ &\text{s.t. } |x_m|^2 \leq \mathrm{P_A}, \text{ for } \; m \in \{1,\ldots,M\}, \; \beta \geq 0. \end{aligned} \tag{3-24}$$

Note that $\boldsymbol{x}$ is a complex-valued vector and β is a real-valued scaling factor. Since most established optimization algorithms consider real-valued variables, the problem from (3-24) is in the following rewritten in a real-valued notation which yields

$$\begin{aligned} &\min_{\boldsymbol{x}_{\mathrm{r}},\beta} \mathrm{E}\{\|\beta\boldsymbol{H}_{\mathrm{r}}\, \boldsymbol{x}_{\mathrm{r}} + \beta\boldsymbol{w}_{\mathrm{r}} - \boldsymbol{s}_{\mathrm{r}}\|_2^2\} \\ &\text{s.t. } \left\{[\boldsymbol{x}_{\mathrm{r}}]_{2m-1}^2 + [\boldsymbol{x}_{\mathrm{r}}]_{2m}^2\right\} \leq \mathrm{P_A}, \text{ for } m \in \{1,\ldots,M\}, \; \beta \geq 0, \end{aligned} \tag{3-25}$$

where $\boldsymbol{w}_\mathrm{r} = R(\boldsymbol{w})$, $\boldsymbol{s}_\mathrm{r} = R(\boldsymbol{s})$ and $\boldsymbol{H}_\mathrm{r} = R(\boldsymbol{H})$, with the operator $R(\cdot)$ introduced in (1-1) and (1-2). Considering that perfect CSI is available at the BS, i.e., $\boldsymbol{H} = \tilde{\boldsymbol{H}}$, the problem from (3-25) can be expressed as an equivalent problem with

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r},\beta} \beta^2 \boldsymbol{x}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r}\, \boldsymbol{x}_\mathrm{r} - 2\beta \boldsymbol{x}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r}^T \boldsymbol{s}_\mathrm{r} + \beta^2 K \sigma_w^2 \\ &\text{s.t. } \left\{ [\boldsymbol{x}_\mathrm{r}]_{2m-1}^2 + [\boldsymbol{x}_\mathrm{r}]_{2m}^2 \right\} \leq \mathrm{P_A}, \text{ for } m \in \{1, \ldots, M\}, \ \beta \geq 0. \end{aligned} \tag{3-26}$$

If $\beta \geq 0$ would be constant, the objective would be a convex quadratically constrained quadratic program, since $\boldsymbol{H}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r} \in S_+^{2M}$, [42, Sec. 4.4]. Yet, the objective is in general not jointly convex in β and $\boldsymbol{x}_\mathrm{r}$ [48, Appendix]. Nevertheless, it can be rewritten as an equivalent convex function by substituting the optimization variable $\boldsymbol{x}_\mathrm{r}$. In this context, we introduce a new optimization variable $\boldsymbol{x}_\mathrm{s} = \beta \boldsymbol{x}_\mathrm{r}$, similar as done in [27, 49]. With this, the optimization problem described in (3-26) can be rewritten as

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{s},\beta} \boldsymbol{x}_\mathrm{s}^T \boldsymbol{H}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r} \boldsymbol{x}_\mathrm{s} - 2\boldsymbol{x}_\mathrm{s}^T \boldsymbol{H}_\mathrm{r}^T \boldsymbol{s}_\mathrm{r} + \beta^2 K \sigma_w^2 \\ &\text{s.t. } [\boldsymbol{x}_\mathrm{s}]_{2m-1}^2 + [\boldsymbol{x}_\mathrm{s}]_{2m}^2 \leq \beta^2 \mathrm{P_A}, \text{ for } m \in \{1, \ldots, M\}, \ \beta \geq 0. \end{aligned} \tag{3-27}$$

The problem can be written in matrix form as

$$\begin{aligned} \min_{\boldsymbol{v}} \ & \boldsymbol{v}^T \boldsymbol{U} \boldsymbol{v} + \boldsymbol{p}^T \boldsymbol{v} \\ \text{s.t. } & \|\boldsymbol{E}_m \boldsymbol{v}\|_2 \leq \boldsymbol{g}^T \boldsymbol{v}, \quad \text{for } m \in \{1, \ldots, M\}, \\ & \boldsymbol{a}^T \boldsymbol{v} \leq 0 \end{aligned} \tag{3-28}$$

where $\boldsymbol{v} = \left[\beta, \boldsymbol{x}_\mathrm{s}^T\right]^T$, $\boldsymbol{a} = \left[-1, \boldsymbol{0}^T\right]^T$, $\boldsymbol{g} = \left[\sqrt{\mathrm{P_A}}, \boldsymbol{0}^T\right]^T$, $\boldsymbol{p} = \left[0, -2\boldsymbol{s}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r}\right]^T$,

$$\boldsymbol{U} = \begin{bmatrix} K\sigma_w^2 & \boldsymbol{0} \\ \boldsymbol{0}^T & \boldsymbol{H}_\mathrm{r}^T \boldsymbol{H}_\mathrm{r} \end{bmatrix}, \quad \boldsymbol{E}_m = \begin{bmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0}^T & \mathrm{diag}\,(\boldsymbol{d}_m) \end{bmatrix}, \tag{3-29}$$

with $\boldsymbol{d}_m \in \mathbb{R}^{2M \times 1}$ being a vector of zeros with ones at entries $2m-1$ and $2m$. Note that, the problem described in (3-28) is convex. In what follows it transformed into a SOCP in standard form, which significantly facilitates implementation. By introducing the additional variable t, cf. [42, Sec. 4.1.3],

the problem can be written with quadratic constraints as

$$\begin{aligned} \min_{t,\boldsymbol{v}} \ & \boldsymbol{p}^T\boldsymbol{v} + 2t + 1 \\ \text{s.t. } & \|\boldsymbol{E}_m\boldsymbol{v}\|_2 \leq \boldsymbol{g}^T\boldsymbol{v}, \quad \text{for } m \in \{1,\ldots,M\}, \\ & \boldsymbol{v}^T\boldsymbol{U}\boldsymbol{v} \leq 2t+1 \\ & \boldsymbol{a}^T\boldsymbol{v} \leq 0. \end{aligned} \tag{3-30}$$

Note that, since $\boldsymbol{U} \in S_+^{2M+1}$, it can be written as $\boldsymbol{U} = \boldsymbol{L}^T\boldsymbol{L}$, with $\boldsymbol{L} = \boldsymbol{U}^{\frac{1}{2}}$. By substituting $\boldsymbol{U} = \boldsymbol{L}^T\boldsymbol{L}$ and adding t^2 at both sides of the quadratic constraint the problem is rewritten as

$$\begin{aligned} \min_{t,\boldsymbol{v}} \ & \boldsymbol{p}^T\boldsymbol{v} + 2t \\ \text{s.t. } & \|\boldsymbol{E}_m\boldsymbol{v}\|_2 \leq \boldsymbol{g}^T\boldsymbol{v}, \quad \text{for } m \in \{1,\ldots,M\}, \\ & \boldsymbol{v}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{v} + t^2 \leq (t+1)^2 \\ & \boldsymbol{a}^T\boldsymbol{v} \leq 0. \end{aligned} \tag{3-31}$$

By using stacked vector notation in the form of the new optimization variable $\boldsymbol{u} = \left[\boldsymbol{v}^T, t\right]^T$ and taking the square root of the quadratic constraint the problem can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{u}} \ & \boldsymbol{r}^T\boldsymbol{u} \\ \text{s.t. } & \|\boldsymbol{F}_m\boldsymbol{u}\|_2 \leq \boldsymbol{l}^T\boldsymbol{u}, \quad \text{for } m \in \{1,\ldots,M\}, \\ & \|\boldsymbol{G}\boldsymbol{u}\|_2 \leq \boldsymbol{q}^T\boldsymbol{u} + 1 \\ & \boldsymbol{o}^T\boldsymbol{u} \leq 0, \end{aligned} \tag{3-32}$$

where $\boldsymbol{r} = \left[\boldsymbol{p}^T, 2\right]^T$, $\boldsymbol{l} = \left[\boldsymbol{g}^T, 0\right]^T$, $\boldsymbol{q} = \left[\boldsymbol{0}^T, 1\right]^T$, $\boldsymbol{o} = \left[\boldsymbol{a}^T, 0\right]^T$,

$$\boldsymbol{F}_m = \begin{bmatrix} \boldsymbol{E}_m & \boldsymbol{0} \\ \boldsymbol{0}^T & 0 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{bmatrix}, \tag{3-33}$$

The problem described in (3-32) is a SOCP, cf. [42, Sec. 4.4.2], and can be readily solved with IPM. The solution can be converted back to complex-valued notation by extracting $\boldsymbol{x}_\mathrm{s}$ and β from $\boldsymbol{u}_\mathrm{opt}$ and applying $\boldsymbol{x} = C\left(\frac{\boldsymbol{x}_\mathrm{s}}{\beta}\right)$.

3.4.2 Proposed Robust MMSE SPAPC Precoding Design

In this subsection, we propose an SLP design based on the MMSE objective under a SPAPC considering knowledge of $\tilde{\boldsymbol{H}}$, $\boldsymbol{N}$ and $\boldsymbol{R}_\mathrm{s}$. Such that

the MMSE objective under imperfect CSI is written in real-valued notation the matrices $\tilde{\boldsymbol{H}}_{\mathrm{r}} = R(\tilde{\boldsymbol{H}})$, $\boldsymbol{N}_{\mathrm{r}} = R(\boldsymbol{N})$, $\boldsymbol{\Psi}_{\mathrm{r}} = R(\boldsymbol{\Psi})$ and $\boldsymbol{R}_{\mathrm{s,r}} = R(\boldsymbol{R}_{\mathrm{s}})$ are defined. With this, the real-valued channel matrix can be written as $\boldsymbol{H}_{\mathrm{r}} = \boldsymbol{N}_{\mathrm{r}}\tilde{\boldsymbol{H}}_{\mathrm{r}} + \sqrt{\boldsymbol{I} - \boldsymbol{N}_{\mathrm{r}}^2}\boldsymbol{\Psi}_{\mathrm{r}}\boldsymbol{R}_{\mathrm{s,r}}^{\frac{1}{2}}$. By substituting $\boldsymbol{H}_{\mathrm{r}}$ in (3-23) and considering $\mathrm{E}\left\{\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{\Psi}_{\mathrm{r}}\right\} = \boldsymbol{0}$ the RMMSE problem reads as

$$\begin{aligned}&\min_{\boldsymbol{x}_{\mathrm{r}},\beta} \beta^2\boldsymbol{x}_{\mathrm{r}}^T\left(\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}^2\tilde{\boldsymbol{H}}_{\mathrm{r}} + \gamma\boldsymbol{R}_{\mathrm{s,r}}\right)\boldsymbol{x}_{\mathrm{r}} - 2\beta\boldsymbol{x}_{\mathrm{r}}^T\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}\boldsymbol{s}_{\mathrm{r}} + \beta^2K\sigma_w^2\\ &\text{s.t. } \left\{[\boldsymbol{x}_{\mathrm{r}}]_{2m-1}^2 + [\boldsymbol{x}_{\mathrm{r}}]_{2m}^2\right\} \le \mathrm{P_A}, \text{ for } m \in \{1,\dots,M\},\ \beta \ge 0,\end{aligned}$$

where $\gamma = \mathrm{trace}(\boldsymbol{I} - \boldsymbol{N}_{\mathrm{r}}^2)$. As before, this proposed objective is not jointly convex in $\boldsymbol{x}_{\mathrm{r}}$ and β. Yet, an equivalent convex problem can be cast by substituting $\boldsymbol{x}_{\mathrm{s}} = \beta\boldsymbol{x}_{\mathrm{r}}$, which yields

$$\begin{aligned}&\min_{\boldsymbol{x}_{\mathrm{s}},\beta} \boldsymbol{x}_{\mathrm{s}}^T\left(\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}^2\tilde{\boldsymbol{H}}_{\mathrm{r}} + \gamma\boldsymbol{R}_{\mathrm{s,r}}\right)\boldsymbol{x}_{\mathrm{s}} - 2\boldsymbol{x}_{\mathrm{s}}^T\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}\boldsymbol{s}_{\mathrm{r}} + \beta^2K\sigma_w^2\\ &\text{s.t. } [\boldsymbol{x}_{\mathrm{s}}]_{2m-1}^2 + [\boldsymbol{x}_{\mathrm{s}}]_{2m}^2 \le \beta^2\mathrm{P_A}, \text{ for } m \in \{1,\dots,M\},\ \beta \ge 0.\end{aligned}$$

The problem can be written in matrix form as

$$\begin{aligned}&\min_{\boldsymbol{v}} \ \boldsymbol{v}^T\tilde{\boldsymbol{U}}\boldsymbol{v} + \tilde{\boldsymbol{p}}^T\boldsymbol{v}\\ &\text{s.t. } \|\boldsymbol{E}_m\boldsymbol{v}\|_2 \le \boldsymbol{g}^T\boldsymbol{v}, \ \text{ for } m \in \{1,\dots,M\},\\ &\quad \boldsymbol{a}^T\boldsymbol{v} \le 0\end{aligned} \tag{3-34}$$

where

$$\tilde{\boldsymbol{U}} = \begin{bmatrix} K\sigma_w^2 & \boldsymbol{0} \\ \boldsymbol{0}^T & \tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}^2\tilde{\boldsymbol{H}}_{\mathrm{r}} + \gamma\boldsymbol{R}_{\mathrm{s,r}}\end{bmatrix}, \quad \tilde{\boldsymbol{p}} = \begin{bmatrix} 0 \\ -2\tilde{\boldsymbol{H}}_{\mathrm{r}}^T\boldsymbol{N}_{\mathrm{r}}\boldsymbol{s}_{\mathrm{r}}\end{bmatrix},$$

and the other quantities are defined in (3-29). Note that, since $\tilde{\boldsymbol{U}} \in S_+^{2M+1}$ the problem is convex. By following the same steps utilized in the section 3.4.1 one can write the problem described in (3-34) as the following SOCP

$$\begin{aligned}&\min_{\boldsymbol{u}} \ \tilde{\boldsymbol{r}}^T\boldsymbol{u}\\ &\text{s.t. } \|\boldsymbol{F}_m\boldsymbol{u}\|_2 \le \boldsymbol{l}^T\boldsymbol{u}, \ \text{ for } m \in \{1,\dots,M\},\\ &\quad \left\|\tilde{\boldsymbol{G}}\boldsymbol{u}\right\|_2 \le \boldsymbol{q}^T\boldsymbol{u} + 1\\ &\quad \boldsymbol{o}^T\boldsymbol{u} \le 0,\end{aligned} \tag{3-35}$$

where

$$\tilde{\boldsymbol{r}} = \begin{bmatrix} \tilde{\boldsymbol{p}} \\ 2 \end{bmatrix}, \quad \tilde{\boldsymbol{G}} = \begin{bmatrix} \tilde{\boldsymbol{L}} & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{bmatrix}, \quad \tilde{\boldsymbol{L}} = \tilde{\boldsymbol{U}}^{\frac{1}{2}} \tag{3-36}$$

and the other quantities are defined in (3-33). As before the problem described in (3-35) is a SOCP, cf. [42, Sec. 4.4.2], and can be readily solved with IPM. The solution can be converted back to complex-valued notation by extracting $\boldsymbol{x}_\mathrm{s}$ and β from $\boldsymbol{u}_\mathrm{opt}$ and applying $\boldsymbol{x} = C\left(\frac{\boldsymbol{x}_\mathrm{s}}{\beta}\right)$. Note that, the MSE associated to the solution of (3-34) is lower bounded by $\breve{\mathrm{MSE}}(\boldsymbol{s}_\mathrm{r}) = K - \boldsymbol{s}_\mathrm{r}^T \boldsymbol{N}_\mathrm{r}^T \tilde{\boldsymbol{H}}_\mathrm{r} \left(\tilde{\boldsymbol{H}}_\mathrm{r}^T \boldsymbol{N}_\mathrm{r}^2 \tilde{\boldsymbol{H}}_\mathrm{r} + \gamma \boldsymbol{R}_\mathrm{s,r}\right)^{-1} \tilde{\boldsymbol{H}}_\mathrm{r}^T \boldsymbol{N}_\mathrm{r} \boldsymbol{s}_\mathrm{r}$. This MSE bound, which is greater than zero due to the CSI imperfection, is computed by considering the unconstrained version of (3-34).

3.4.3 About the Complexity of the Proposed Designs

As mentioned the MMSE and the RMMSE optimization problems are SOCPs and thus can be solved via IPM. By solving them with the Barrier Method one can achieve a UBCO of $\mathcal{O}(M^{3.5})$. Another IPM approach that can be utilized is the primal-dual IPM. According to [50], the number of iterations of the primal-dual IPM can be upper bounded by $\sqrt{n}\log(n/\epsilon_\mathrm{tol})$ where n is the number of variables and ϵ_tol is the predefined optimality tolerance. Note that, the complexity of the iterations is dominated by solving a linear system needed to compute the primal-dual search direction. With this, considering that the linear systems can be solved with complexity $\mathcal{O}(n^3)$ via Gauss-Jordan elimination, the total complexity of the proposed approaches can be upper bounded by $\mathcal{O}(M^{3.5}\log(M/\epsilon_\mathrm{tol}))$.

3.5 Numerical Results

In this section, the proposed precoders are evaluated in terms of BER and computational complexity and compared with other state-of-the-art designs. To this end, the SNR is defined as $\mathrm{SNR} = (M \cdot \mathrm{P_A})/\sigma_w^2$, as derived in section C.1 of the appendix. The proposed methods are evaluated against the following state-of-the-art approaches:

1- The ZF SPAPC precoder [39];

2- The CVX-CIO precoder [12] designed for constant envelope;

3- The Strict CI SPAPC precoder [39];

4- The Non-Strict CI SPAPC precoder [39];

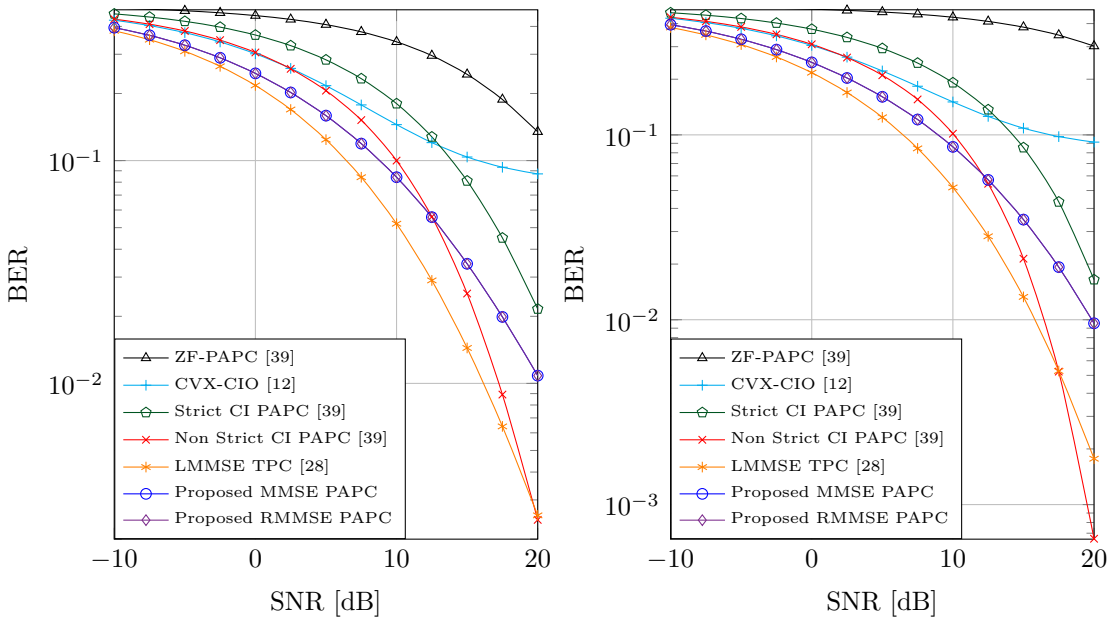

Figure 3.1: BER × SNR, for $\alpha_s = 4$ PSK users' data, CSI quality $\boldsymbol{\eta} = \mathbf{1}$, spatial correlation factor $\rho = 0$. $K = 15$ users, $M = 15$ antennas (left). $K = 60$ users, $M = 60$ antennas (right).

5- The LMMSE precoder [28] (average TPC).

3.5.1 BER evaluation under perfect CSI

In this subsection, a BER × SNR evaluation is considered assuming no spatial correlation, i.e., $\rho = 0$ and perfect CSI. In this context, the first experiment, shown on the LHS of Fig. 3.1, considers a MIMO scenario with a BS with $M = 15$ antennas serving $K = 15$ users with QPSK user symbols, meaning that $\alpha_s = 4$. As seen in the LHS of Fig. 3.1, the proposed methods outperform the existing approaches in terms of BER for the low and intermediate SNR regimes. For high-SNR, the proposed MMSE precoders outperform all investigated approaches except for the Non-Strict CI-based precoder [39]. This is expected since it is known that CI is nearly optimal for high SNR [51] and the MMSE criterion is favorable for low and medium SNR [27]. The second experiment, shown in the RHS of Fig. 3.1, evaluates the BER against the SNR in the larger MIMO context of a BS with $M = 60$ antennas serving $K = 60$ users also with QPSK user symbols. The RHS of Fig. 3.1, reaffirms the conclusions present in the LHS, underlining that the proposed MMSE-based SLPs are favorable for the low and intermediate SNR regimes.

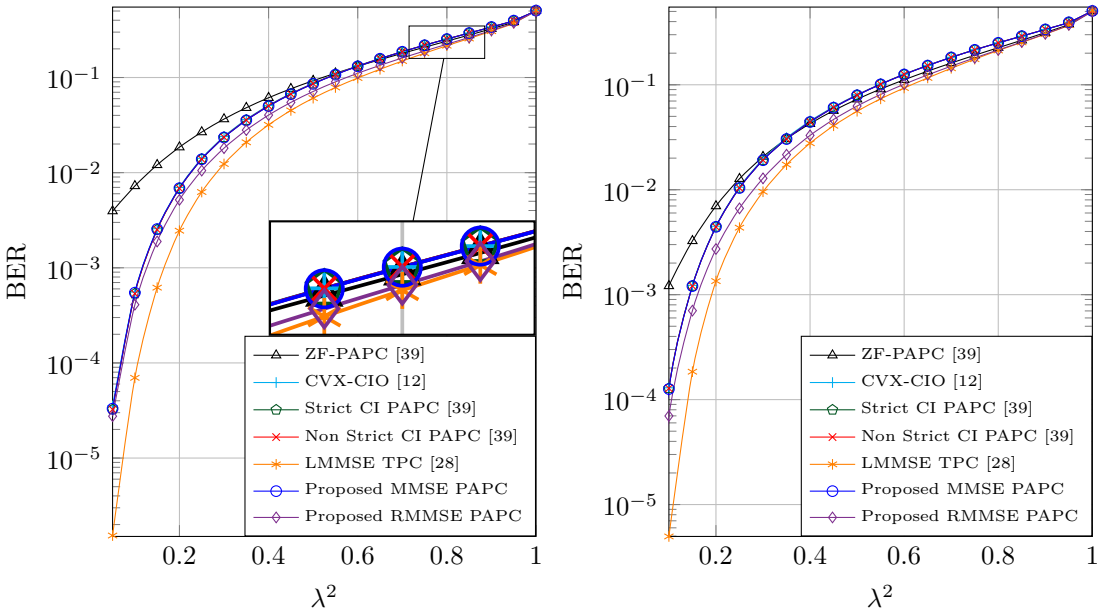

Figure 3.2: BER × CSI imperfection factor λ^2, for $K = 5$ users, $M = 50$ antennas, $\alpha_s = 8$ PSK users' data, spatial correlation factor $\rho = 0$. SNR= 12 dB (left). SNR= 15 dB (right)

3.5.2 BER evaluation under imperfect CSI

In this subsection, the proposed approaches are evaluated in terms of BER with CSI imperfection. The evaluated MIMO scenario consists of a BS with $M = 50$ antennas which serve $K = 5$ users with $\alpha_s = 8$. To facilitate the analysis during this subsection it is considered $\boldsymbol{\eta} = \xi\mathbf{1}$, meaning that all channels have the same CSI quality. The CSI imperfection is then expressed in terms of $\lambda^2 = \sqrt{1-\xi^2}$.

The first experiment, shown in Fig. 3.2, consists of a BER performance evaluation for different levels of CSI imperfection under SNR of 12 dB for the LHS, and SNR of 15 dB for the RHS. For this experiment no spatial correlation is considered, meaning $\rho = 0$. As can be seen in the LHS of Fig. 3.2 the proposed RMMSE SPAPC design outperforms in terms of BER all other examined SPAPC state-of-the-art approaches for $\lambda^2 > 0$. Moreover, the proposed RMMSE approach yields similar performance in terms of BER as the LMMSE [28] (average TPC) design for very low CSI quality ($\lambda^2 > 0.8$). The RHS of Fig. 3.2 reaffirms the conclusions discussed. Yet, the comparison of both plots reveals that the performance benefits of the proposed RMMSE approach are more pronounced in the RHS. This outcome is expected, as the RMMSE technique accounts for CSI imperfection, and as the SNR increases, performance becomes increasingly dominated by CSI mismatch rather than

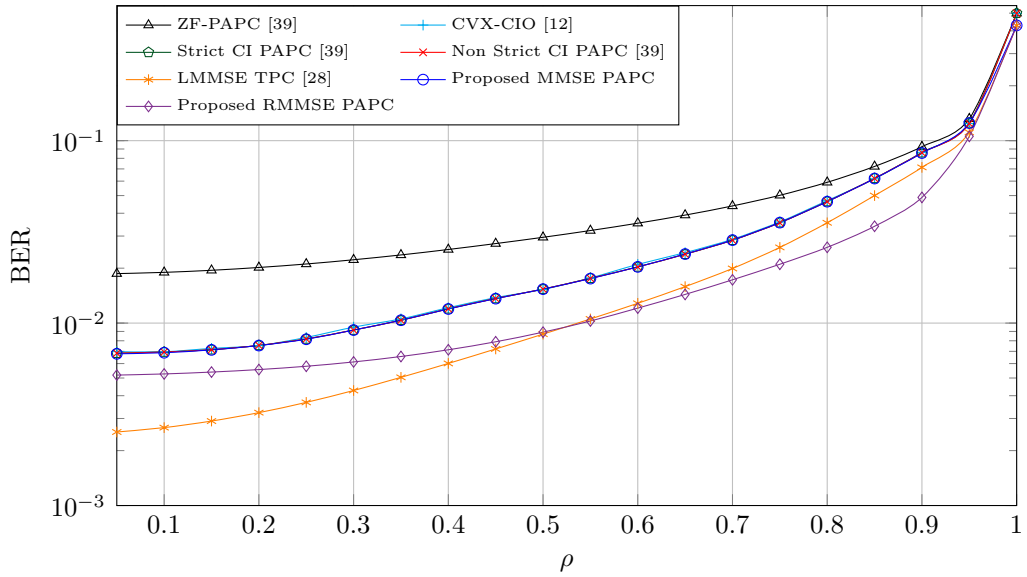

Figure 3.3: BER × Spatial correlation factor ρ, for $K = 5$ users, $M = 50$ antennas, $\alpha_s = 8$ PSK users' data, CSI imperfection factor $\lambda^2 = 0.2$ and SNR= 12 dB

noise.

The second experiment consists of a BER performance evaluation against the spatial correlation factor ρ for SNR = 12 dB and $\lambda^2 = 0.2$. As shown in Fig. 3.3, the proposed RMMSE approach outperforms in terms of BER all examined SPAPC designs for all examined ρ. Moreover, it also outperforms the LMMSE design for $\rho > 0.5$.

Finally, the third experiment consists of a BER × SNR evaluation considering both imperfect CSI and spatial correlation with the parameters $\lambda^2 = 0.2$ and $\rho = 0.15$. As can be seen in Fig. 3.4, the proposed RMMSE precoder outperforms all other SPAPC approaches in terms of BER. Note that, both proposed MMSE and RMMSE approaches yield similar performance for low SNR. Starting from medium SNR, as the SNR grows, the proposed RMMSE approach deviates in performance from the proposed MMSE counterpart. Finally, for very high SNR (SNR > 27.5 dB) the proposed RMMSE approach shows a significant advantage and outperforms also the LMMSE method (average TPC).

3.5.3 Complexity Analysis

As discussed in section 3.4.3 the complexity of the proposed methods is upper bounded by $\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$. Table 3.1 summarizes the complexity of the considered approaches.

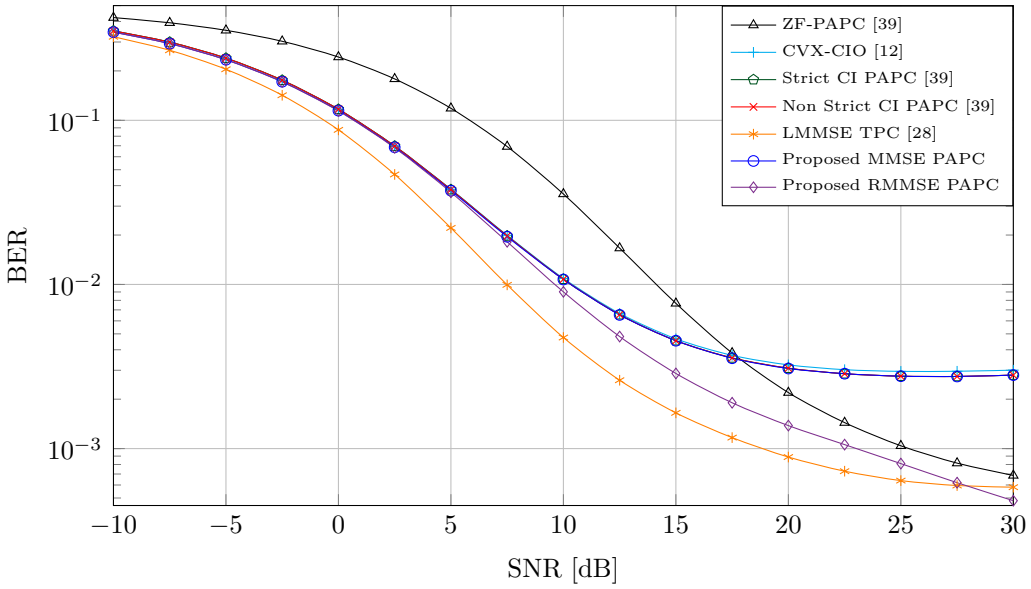

Figure 3.4: BER × SNR, for $K = 5$ users, $M = 50$ antennas, $\alpha_s = 8$ PSK users' data, spatial correlation factor $\rho = 0.15$ and CSI imperfection factor $\lambda^2 = 0.2$

Table 3.1: Computational Complexity of the Precoding Algorithms

Algorithm	Complexity
ZF SPAPC [39]	$\mathcal{O}\left(K^2M\right)$
CVX-CIO [12]	$\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$
Strict CI SPAPC [39]	$\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$
Non-Strict CI SPAPC [39]	$\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$
Linear MMSE TPC [28]	$\mathcal{O}\left(K^3\right)$
Proposed MMSE SPAPC	$\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$
Proposed RMMSE SPAPC	$\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$

Note that, the optimization-based state-of-the-art algorithms (namely CVX-CIO [12], Strict CI SPAPC [39] and Non-Strict CI SPAPC [39]) can be transformed in standard form SOCPs which can be solved via the primal-dual IPM with complexity $\mathcal{O}\left(M^{3.5}\log\left(M/\epsilon_{\text{tol}}\right)\right)$. With this, it can be concluded that these approaches yield similar complexity as the proposed methods.

4
Symbol-Level Precoding under Constant Envelope and Low-Resolution Constraints with the Conventional MIMO Transmitter

For large-scale MIMO systems where the energy consumption of the RFFE is significant to the EE of the system, power reduction features such as CE signaling and low-resolution quantization are necessary for low-cost deployments, with low environmental impact, and better coverage. To mitigate the error-rate performance degradation that these features yield CE low-resolution precoding has become prominent in the literature [52, 53]. In what follows the system model considered for this chapter is exposed and a brief revision of the CE low-resolution SLP literature is provided.

4.1
System Model

The system model consists of a single-cell MU-MIMO scenario where the BS is equipped with M transmit antennas that serve K single-antenna users. A symbol-level transmission is considered where s_k represents the data symbol of the k-th user. Each symbol s_k is considered to belong to the set $\mathcal{S}$ that represents all possible symbols of a α_s-PSK modulation and reads as $\mathcal{S} = \left\{ s : s = e^{\frac{j\pi(2i+1)}{\alpha_s}}, \text{for } i = 1, \ldots, \alpha_s \right\}$. The symbols of all users are described in a stacked vector notation as $\boldsymbol{s} = [s_1, \ldots, s_K]^T \in \mathcal{S}^K$. It is considered that different users' symbols are independent and that $\mathrm{P}\left(s_k = s_i\right) = {}^1\!/_{\alpha_s}, \forall i \in \{1, \ldots, \alpha_s\}$. Based on $\boldsymbol{s}$ the precoder computes the transmit vector $\boldsymbol{x} = [x_1, \ldots, x_M]^T$ with entries constrained to the set $\mathcal{X}$ which is given by $\mathcal{X} = \left\{ x : x = \sqrt{\mathrm{P_A}}\, e^{\frac{j\pi(2i+1)}{\alpha_x}}, \text{for } i = 1, \ldots, \alpha_x \right\}$ with $\mathrm{P_A}$ being the per antenna transmit power. The vector $\boldsymbol{x}$ is transmitted over a frequency flat fading channel described by the matrix $\boldsymbol{H} \in \mathbb{C}^{K \times M}$. The received signal corresponding to the k-th user reads as $z_k = y_k + w_k = \boldsymbol{h}_k\, \boldsymbol{x} + w_k$, where y_k is the noiseless received signal at the k-th user, $\boldsymbol{h}_k$ is the k-th row of the channel matrix $\boldsymbol{H}$ and the complex random variable $w_k \sim \mathcal{CN}(0, \sigma_w^2)$ represents additive white Gaussian noise. Each z_k is detected based on the decision region it belongs. The decision region of s_i, termed $\mathcal{S}_i$, is the set of points closer to s_i than all other valid candidates for detection. This implies

that z_k is detected as s_i if $z_k \in \mathcal{S}_i$. For PSK the decision regions are circle sectors with infinite radius and angle of 2θ, where $\theta = \pi/\alpha_s$. The detected symbol vector is written as $\hat{\boldsymbol{s}} = [\hat{s}_1, \ldots, \hat{s}_K]$.

4.2 Literature Review

This section revises some of the prominent discrete SLP methods in the literature. The techniques that are exposed are

1- The MSM precoder [14];

2- The MMDDT B&B precoder [18];

3- The MMSE Mapped precoder [48];

4- The MMSE B&B precoder [48].

All approaches considered in this section consider perfect CSI at the transmitter.

4.2.1 MMDDT-based Low-resolution Precoders

As mentioned in section 3.2.2.2 the MMDDT is one of the most prominent criteria in the literature. Different works utilize it as the design objective for low-resolution SLP. In this section, we expose the MMDDT problem formulation under low-resolution constraints and expose the methods from [12, 14, 18]. Considering the MMDDT objective, described in section 3.2.2.2, with low-resolution constraints yields the following optimization problem

$$\begin{aligned}\left[\boldsymbol{x}_{\text{opt}},\, \epsilon_{\,\text{opt}}\right] = \underset{\boldsymbol{x}\in\mathcal{X}^M,\epsilon}{\operatorname{argmin}} \; & -\epsilon \\ \text{s.t.} \;\; & \operatorname{Re}\{\boldsymbol{H}_{s^*}\boldsymbol{x}\}\sin\theta - |\operatorname{Im}\{\boldsymbol{H}_{s^*}\boldsymbol{x}\}|\cos\theta \geq \epsilon \mathbf{1}_{2K},\end{aligned} \tag{4-1}$$

where $\boldsymbol{H}_{s^*} = \operatorname{diag}(\boldsymbol{s}^*)\boldsymbol{H}$. The set $\mathcal{X}^M$ contains α^M elements, since the number of elements is finite, problem (4-1) is solvable by applying exhaustive search, which implies exponential complexity with the number of BS antennas. This yields prohibitive computational complexity even for small-scale MIMO systems. With this, different methods were devised to achieve more reasonable complexity performance trade-offs.

4.2.1.1 The MSM precoder

To build a reduced complexity technique, [14] considers a relaxation of the feasible set $\mathcal{X}^M$ to its convex hull $\mathcal{P}$. With this, problem (4-1) is relaxed as

$$\begin{aligned}\left[\boldsymbol{x}_{\text{opt}},\, \epsilon_{\,\text{opt}}\right] = \arg\min_{\boldsymbol{x}\in\mathcal{P},\epsilon} & -\epsilon \\ \text{s.t.}\;\; & \text{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\theta - \left|\text{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\right|\cos\theta \geq \epsilon\mathbf{1}_{2K}.\end{aligned} \tag{4-2}$$

The problem is equivalently written with real-valued variables as

$$\begin{aligned}\left[\boldsymbol{x}_{\text{lb}},\, \epsilon\right] = \arg\min_{\boldsymbol{x},\epsilon} & -\epsilon \\ \text{s.t.}\;\; & \text{Re}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\sin\theta - \left|\text{Im}\left\{\boldsymbol{H}_{s^*}\boldsymbol{x}\right\}\right|\cos\theta \geq \epsilon\mathbf{1}_{2K} \\ & \text{Re}\left\{x_m e^{j\phi_i}\right\} \leq \frac{\cos\left(\frac{\pi}{\alpha_x}\right)}{\sqrt{M}},\;\text{for } m = 1,\ldots,M \\ & \phi_i = \frac{2\pi i}{\alpha_x},\;\text{for } i = 1,\ldots\alpha_x.\end{aligned} \tag{4-3}$$

With the relaxation of the feasible set $\boldsymbol{x}_{\text{lb}}$ does not necessarily attain the low-resolution constraints, i.e., $\boldsymbol{x}_{\text{lb}}$ does not necessarily belong to $\mathcal{X}^M$. With this, to achieve a feasible solution, uniform quantization is considered, which yields $\boldsymbol{x} = Q(\boldsymbol{x}_{\text{lb}})$. The vector $\boldsymbol{x}$ adheres to the low-resolution constraints and is utilized for transmission.

4.2.1.2 MMDDT B&B Precoder

The MMDDT B&B Precoder, first considered in [17] for 1-bit quantizers, is generalized in [18] for quantization with arbitrary resolution. The idea is to utilize the MMDDT criterion (described in section 3.2.2.2) with the Full-B&B method (described in section 2.2) to achieve optimal MMDDT performance. As described in section 2.2 the Full-B&B method considers an initialization step for complexity reduction reasons. In the MMDDT B&B Precoder case, this initialization step consists of the computation of the MSM solution. The subsequent tree-search-based part of the algorithm consists of applying the MMDDT objective to the subproblems which yields convex optimization problems solvable with IPM.

As stated in [48], the MMDDT criterion asymptotically approaches the optimal SEP performance with the increase in SNR. With this, the MMDDT B&B technique can be considered as near MSEP precoder for the high-SNR

regime. This statement is confirmed through numerical simulations in section 4.6.

4.2.2 MMSE-based Low-resolution Precoders

Another established precoding design criterion is the MMSE [28]. In this section, we revise some of the low-resolution SLP techniques. Considering the MMSE objective with low-resolution constraints yields the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r},f} f^2\boldsymbol{x}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}\,\boldsymbol{x}_\mathrm{r} - 2f\boldsymbol{x}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}^T\boldsymbol{s}_\mathrm{r} + f^2\mathrm{E}\{\boldsymbol{w}_\mathrm{r}^T\boldsymbol{w}_\mathrm{r}\} \\ &\text{subject to: } \left\{[\boldsymbol{x}_\mathrm{r}]_{2m-1} + j\,[\boldsymbol{x}_\mathrm{r}]_{2m}\right\} \in \mathcal{X}, \quad \text{for } m \in \{1,\ldots,M\}, \\ &\qquad\qquad f \geq 0, \end{aligned} \tag{4-4}$$

where $\boldsymbol{x}_\mathrm{r} = R(\boldsymbol{x})$. Note that, similarly as in section 3.4.1 problem (4-4) is not jointly convex in $\boldsymbol{x}_\mathrm{r}$ and f. Yet, the utilization of a similar variable substitution leads to a joint convex objective in terms of $\boldsymbol{x}_\mathrm{r,f} = f\boldsymbol{x}_\mathrm{r}$ and $\boldsymbol{x}_\mathrm{r}$.

4.2.2.1 The MMSE Mapped Precoder

Similarly as in the MSM case, to build a reduced complexity technique, [48] considers a relaxation of the feasible set $\mathcal{X}^M$ to its convex hull $\mathcal{P}$, which yields the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r},f} f^2\boldsymbol{x}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}\,\boldsymbol{x}_\mathrm{r} - 2f\boldsymbol{x}_\mathrm{r}^T\boldsymbol{H}_\mathrm{r}^T\boldsymbol{s}_\mathrm{r} + f^2\mathrm{E}\{\boldsymbol{w}_\mathrm{r}^T\boldsymbol{w}_\mathrm{r}\} \\ &\text{subject to: } \boldsymbol{A}\boldsymbol{x}_\mathrm{r} \leq \boldsymbol{b}, \quad f \geq 0, \end{aligned} \tag{4-5}$$

where

$$\begin{aligned} \boldsymbol{A} &= \left[(\boldsymbol{I}_M \otimes \boldsymbol{\beta}_1)^T \quad (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_2)^T \quad \ldots \quad (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_{\alpha_x})^T\right]^T, \\ \boldsymbol{\beta}_i &= \left[\cos\left(\tfrac{2\pi i}{\alpha_x}\right) \quad -\sin\left(\tfrac{2\pi i}{\alpha_x}\right)\right], \quad i \in \{1,\ldots,\alpha_x\}, \\ \boldsymbol{b} &= \frac{\cos(\frac{\pi}{\alpha_x})}{\sqrt{M}}\mathbf{1}_{M\alpha_x}, \end{aligned} \tag{4-6}$$

with $\mathbf{1}_{M\alpha_x}$ being a column vector with length $M\alpha_x$. Writing (4-5) in terms of $\boldsymbol{x}_{\mathrm{r,f}}$ yields

$$\begin{aligned} &\min_{\boldsymbol{x}_{\mathrm{r,f}},f} \boldsymbol{x}_{\mathrm{r,f}}^T \boldsymbol{H}_{\mathrm{r}}^T \boldsymbol{H}_{\mathrm{r}} \boldsymbol{x}_{\mathrm{r,f}} - 2\boldsymbol{x}_{\mathrm{r,f}}^T \boldsymbol{H}_{\mathrm{r}}^T \boldsymbol{s}_{\mathrm{r}} + f^2 \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{w}_{\mathrm{r}}\} \\ &\text{subject to: } \boldsymbol{R} \begin{bmatrix} \boldsymbol{x}_{\mathrm{r,f}} \\ f \end{bmatrix} \leq \mathbf{0}, \qquad f \geq 0, \end{aligned} \tag{4-7}$$

where $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{A}, & -\boldsymbol{b} \end{bmatrix}$. Finally, the MMSE mapped problem consists of writing (4-7) as a standard quadratic program (QP) with

$$\begin{aligned} &\min_{\boldsymbol{v}} \frac{1}{2} \boldsymbol{v}^T \boldsymbol{U} \boldsymbol{v} + \boldsymbol{p}^T \boldsymbol{v} \\ &\text{subject to: } \boldsymbol{R}_{\mathrm{ext}} \boldsymbol{v} \leq \mathbf{0}, \end{aligned} \tag{4-8}$$

where

$$\begin{aligned} \boldsymbol{v} &= \begin{bmatrix} \boldsymbol{x}_{\mathrm{r,f}} \\ f \end{bmatrix}, \quad \boldsymbol{U} = 2 \begin{bmatrix} \boldsymbol{H}_{\mathrm{r}}^T \boldsymbol{H}_{\mathrm{r}} & \mathbf{0} \\ \mathbf{0}^T & \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{w}_{\mathrm{r}}\} \end{bmatrix}, \\ \boldsymbol{p} &= \begin{bmatrix} -2\boldsymbol{H}_{\mathrm{r}}^T \boldsymbol{s}_{\mathrm{r}} \\ 0 \end{bmatrix}, \quad \boldsymbol{R}_{\mathrm{ext}} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{\xi}^T \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix}. \end{aligned} \tag{4-9}$$

Note that, since $\boldsymbol{U} \in S_+^{2M+1}$, the objective is convex and can be solved utilizing standard optimization tools. After solving (4-8) the MMSE mapped one extracts $\boldsymbol{x}_{\mathrm{r,f}}$ from the optimal solution $\boldsymbol{v}$ and computes $\boldsymbol{x}_{\mathrm{lb}} = (\boldsymbol{x}_{\mathrm{r,f}}/f)$. Note that the solution $\boldsymbol{x}_{\mathrm{lb}}$ does not necessarily belong to $\mathcal{X}^M$. To arrive at a feasible solution [48] considers uniform quantization of $\boldsymbol{x}_{\mathrm{lb}}$ which yields $\hat{\boldsymbol{x}}_{\mathrm{ub}} = Q(\boldsymbol{x}_{\mathrm{lb}})$. Since the mapping step does not preserve the value of f, it is recomputed based on the mapped precoding vector $\hat{\boldsymbol{x}}_{\mathrm{ub}}$, with

$$f_{\mathrm{ub}} = \frac{\boldsymbol{s}_{\mathrm{r}}^T \boldsymbol{H}_{\mathrm{r}} \hat{\boldsymbol{x}}_{\mathrm{r,ub}}}{||\boldsymbol{H}_{\mathrm{r}} \hat{\boldsymbol{x}}_{\mathrm{r,ub}}||_2^2 + \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{w}_{\mathrm{r}}\}}, \tag{4-10}$$

where $\hat{\boldsymbol{x}}_{\mathrm{r,ub}} = R(\hat{\boldsymbol{x}}_{\mathrm{ub}})$. The scaling factor f_{ub} associated with the mapped solution can be negative corresponding to an unfeasible solution of (4-4). In this scenario, a feasible solution with equivalent MSE is computed by flipping the sign of $\hat{\boldsymbol{x}}_{\mathrm{ub}}$, leading to the transmit vector being computed as $\boldsymbol{x} = \mathrm{sign}(f_{\mathrm{ub}})\hat{\boldsymbol{x}}_{\mathrm{ub}}$.

4.2.2.2 The MMSE B&B Precoder

The MMSE B&B Precoder utilizes the MMSE criterion with the Full-B&B method (described in section 2.2) to achieve optimal MMSE performance.

For the initialization step, the MMSE-mapped solution is considered. The subsequent tree-search-based part of the algorithm consists of directly applying the MMSE objective to the subproblems which yields convex optimization problems solvable with IPM.

As stated in [48], the MMSE criterion is favorable in terms of BER for the low-SNR regime. As confirmed through numerical simulations in section 4.6, the MMSE B&B technique can be considered as near MSEP precoder for the low-SNR regime.

4.3 Contributions of this chapter

In this chapter different CE low-resolution SLPs are proposed for PSK modulation. Different criteria are considered, first, SEP-related criteria, namely MSEP and MUBSEP, are used, and novel precoding algorithms are proposed based on them. Then the RMMSE criterion is utilized considering imperfect CSI, and various precoding methods are proposed.

4.4 Discrete Precoding with SEP-related Criteria

This section proposes the utilization of SEP-based precoding criteria as a direct approach to optimize the QoS of the system. The probability of detecting the data vector $\boldsymbol{s}$ conditioned on the transmit vector $\boldsymbol{x}$ can be computed based on the probabilities of detection of the individual users as

$$\mathrm{P}(\hat{\boldsymbol{s}} = \boldsymbol{s}|\boldsymbol{x}) = \prod_{k=1}^{K} \mathrm{P}(\hat{s}_k = s_k|\boldsymbol{x}) \ . \tag{4-11}$$

To simplify the notation we denote $\mathrm{P}(\hat{\boldsymbol{s}} = \boldsymbol{s}|\boldsymbol{x})$ as $\mathrm{P}(\hat{\boldsymbol{s}}|\boldsymbol{x})$ and $\mathrm{P}(\hat{s}_k = s_k|\boldsymbol{x})$ as $\mathrm{P}(\hat{s}_k|\boldsymbol{x})$. With this, (4-11) is rewritten as $\mathrm{P}(\hat{\boldsymbol{s}}|\boldsymbol{x}) = \prod_{k=1}^{K} \mathrm{P}(\hat{s}_k|\boldsymbol{x})$. As stated before, the detector decides for s_k when the received symbol z_k belongs to $\mathcal{S}_k$. Thus, the individual user probabilities are given by

$$\mathrm{P}\left(\hat{s}_k|\boldsymbol{x}\right) = \mathrm{P}\left(z_k \in \mathcal{S}_k|\boldsymbol{x}\right) = \frac{1}{\pi\sigma_w^2} \int_{\mathcal{S}_k} \mathrm{e}^{-\frac{|t-y_k|^2}{\sigma_w^2}} dt. \tag{4-12}$$

The integral from (4-12) has tabled solutions for $\alpha_s \in \{2, 4\}$, allowing the BS to compute it with relatively low computational effort. Yet, for $\alpha_s \notin \{2, 4\}$, the exact computation of (4-12) requires the utilization of Monte Carlo methods. In this study, the exact computation of (4-12) is considered for $\alpha_s \in \{2, 4\}$. For $\alpha_s \notin \{2, 4\}$, we propose the union-bound probability as a lower bound

on $\mathrm{P}\left(\hat{s}_k|\boldsymbol{x}\right)$ allowing for a general formulation with a closed-form objective function.

4.4.1 MSEP Criterion

For $\alpha_s \in \{2, 4\}$, the real and imaginary parts of the data symbols are considered independent. This allows the decision region $\mathcal{S}_k$ to be written as $\mathcal{R}_k \cap \mathcal{I}_k$, where $\mathcal{R}_k$ and $\mathcal{I}_k$ are the decision regions the real and imaginary parts of s_k. The probability of the detector deciding for s_k, can be written as $\mathrm{P}\left(\hat{s}_k|\boldsymbol{x}\right) = \mathrm{P}\left(\mathrm{Re}\left\{z_k\right\} \in \mathcal{R}_k|\boldsymbol{x}\right)\mathrm{P}\left(\mathrm{Im}\left\{z_k\right\} \in \mathcal{I}_k|\boldsymbol{x}\right)$, where the probabilities of correct detection of the real and imaginary parts of s_k are given by

$$\begin{aligned}\mathrm{P}\left(z_k \in \mathcal{R}_k|\boldsymbol{x}\right) &= \int_0^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{\left(t-\mathrm{sign}\left(\mathrm{Re}\left\{s_k\right\}\right)\mathrm{Re}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}\right)^2}{\sigma_w^2}} dt \\ &= \Phi\left(\frac{\sqrt{2}\,\mathrm{sign}\left(\mathrm{Re}\left\{s_k\right\}\right)\mathrm{Re}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}}{\sigma_w}\right), \end{aligned} \tag{4-13}$$

$$\begin{aligned}\mathrm{P}\left(z_k \in \mathcal{I}_k|\boldsymbol{x}\right) &= \int_0^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{\left(t-\mathrm{sign}\left(\mathrm{Im}\left\{s_k\right\}\right)\mathrm{Im}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}\right)^2}{\sigma_w^2}} dt \\ &= \Phi\left(\frac{\sqrt{2}\,\mathrm{sign}\left(\mathrm{Im}\left\{s_k\right\}\right)\mathrm{Im}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}}{\sigma_w}\right). \end{aligned} \tag{4-14}$$

With this, the probability of correct detection is computed considering (4-11) which reads as

$$\mathrm{P}\left(\hat{\boldsymbol{s}}|\boldsymbol{x}\right) = \prod_{k=1}^{K} \Phi\left(\frac{\sqrt{2}\,\mathrm{sign}\left(\mathrm{Re}\left\{s_k\right\}\right)\mathrm{Re}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}}{\sigma_w}\right)\Phi\left(\frac{\sqrt{2}\,\mathrm{sign}\left(\mathrm{Im}\left\{s_k\right\}\right)\mathrm{Im}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}}{\sigma_w}\right). \tag{4-15}$$

The minimum SEP (MSEP) problem, equivalent as defined in [54], is written as the minimization of $-\ln\left(\mathrm{P}\left(\hat{\boldsymbol{s}}|\boldsymbol{x}\right)\right)$, which reads as

$$\min_{\boldsymbol{x}\in\mathcal{X}^M} -\sum_{k=1}^{K}\left(\ln\left(\Phi\left(\boldsymbol{u}_{\mathrm{r},k}\left(\boldsymbol{x}\right)\right)\right) + \ln\left(\Phi\left(\boldsymbol{u}_{\mathrm{i},k}\left(\boldsymbol{x}\right)\right)\right)\right), \tag{4-16}$$

where $\boldsymbol{u}_{\mathrm{r},k}\left(\boldsymbol{x}\right) = \left(\sqrt{2}/\sigma_w\right)\left(\mathrm{sign}\left(\mathrm{Re}\left\{s_k\right\}\right)\mathrm{Re}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}\right)$ and $\boldsymbol{u}_{\mathrm{i},k}\left(\boldsymbol{x}\right) = \left(\sqrt{2}/\sigma_w\right)\left(\mathrm{sign}\left(\mathrm{Im}\left\{s_k\right\}\right)\mathrm{Im}\left\{\boldsymbol{h}_k\boldsymbol{x}\right\}\right)$. An alternative real-valued formulation

can be cast as

$$\begin{aligned} \min_{\boldsymbol{x}_\mathrm{r}} \; & -\sum_{k=1}^{K}\left(\ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)+\ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)\right) \\ \text{s.t.} \quad & [\boldsymbol{x}_\mathrm{r}]_{2m-1}+j\,[\boldsymbol{x}_\mathrm{r}]_{2m}\in\mathcal{X} \quad \text{for } m=1,\ldots,M. \end{aligned} \tag{4-17}$$

where M denotes the number of BS antennas, $\boldsymbol{x}_\mathrm{r} = R(\boldsymbol{x})$ and $\boldsymbol{h}_{\mathrm{R},k}^{T}$ and $\boldsymbol{h}_{\mathrm{I},k}^{T}$ are the k-th rows of matrices $\boldsymbol{H}_\mathrm{R}$ and $\boldsymbol{H}_\mathrm{I}$, respectively. The matrices $\boldsymbol{H}_\mathrm{R}$ and $\boldsymbol{H}_\mathrm{I}$ are defined as $\boldsymbol{H}_\mathrm{R} = (\sqrt{2}/\sigma_w)\,\mathrm{diag}(\mathrm{sign}(\mathrm{Re}\{\boldsymbol{s}\}))\boldsymbol{H}_\mathrm{R}^{Q}$ and $\boldsymbol{H}_\mathrm{I} = (\sqrt{2}/\sigma_w)\,\mathrm{diag}(\mathrm{sign}(\mathrm{Im}\{\boldsymbol{s}\}))\boldsymbol{H}_\mathrm{I}^{Q}$, with

$$\boldsymbol{H}_\mathrm{R}^{Q}=\begin{bmatrix} \mathrm{Re}\{h_{11}\} & -\mathrm{Im}\{h_{11}\} & \cdots & \mathrm{Re}\{h_{1M}\} & -\mathrm{Im}\{h_{1M}\} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{Re}\{h_{K1}\} & -\mathrm{Im}\{h_{K1}\} & \cdots & \mathrm{Re}\{h_{KM}\} & -\mathrm{Im}\{h_{KM}\} \end{bmatrix}, \tag{4-18}$$

$$\boldsymbol{H}_\mathrm{I}^{Q}=\begin{bmatrix} \mathrm{Im}\{h_{11}\} & \mathrm{Re}\{h_{11}\} & \cdots & \mathrm{Im}\{h_{1M}\} & \mathrm{Re}\{h_{1M}\} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{Im}\{h_{K1}\} & \mathrm{Re}\{h_{K1}\} & \cdots & \mathrm{Re}\{h_{KM}\} & \mathrm{Im}\{h_{KM}\} \end{bmatrix}. \tag{4-19}$$

The MSEP objective in real-valued notation can be cast as

$$f_0(\boldsymbol{x}_\mathrm{r}) = -\sum_{k=1}^{K}\left(\ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)+\ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)\right). \tag{4-20}$$

The gradient and Hessian of $f_0(\boldsymbol{x}_\mathrm{r})$ are given by

$$\nabla f_0(\boldsymbol{x}_\mathrm{r}) = -\sum_{k=1}^{K}\frac{\boldsymbol{m}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)}+\frac{\boldsymbol{m}_{\mathrm{I},k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)}, \tag{4-21}$$

$$\begin{aligned} &\nabla^2 f_0(\boldsymbol{x}_\mathrm{r}) = \\ &\sum_{k=1}^{K}\frac{\boldsymbol{m}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{R,k}^{T}\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)^2}+\frac{\boldsymbol{m}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{I,k}^{T}\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)\right)^2}, \end{aligned} \tag{4-22}$$

where

$$\boldsymbol{m}_{R,k} = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{R},k}, \tag{4-23}$$

$$\boldsymbol{m}_{I,k} = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{I},k}, \tag{4-24}$$

$$\boldsymbol{\Psi}_{R,k} = \frac{\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)}{\sqrt{2\pi}}\,e^{-\frac{\left(\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^{T}\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{R},k}^{T}, \tag{4-25}$$

$$\boldsymbol{\Psi}_{I,k} = \frac{\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)}{\sqrt{2\pi}}\,e^{-\frac{\left(\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{I},k}\boldsymbol{h}_{\mathrm{I},k}^{T}\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{I},k}^{T}. \tag{4-26}$$

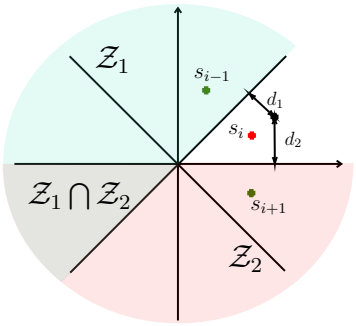

Figure 4.1: Representation of the union bound

The MSEP objective is convex in $\boldsymbol{x}_\mathrm{r}$ with the convexity proof given in Appendix A.1. As previously stated the MSEP formulation, first proposed in [54], is limited to $\alpha_s \in \{2, 4\}$.

4.4.2 Proposed MUBSEP Criterion

The union bound states that for any finite set of events, $\mathrm{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathrm{P}(A_i)$ with A_i being an event. With this, the error probability of the k-th user, $\mathrm{P_e}\left(\hat{s}_k|\boldsymbol{x}\right) = \mathrm{P}\left(z_k \in \mathcal{Z}_1 \cup \mathcal{Z}_2|\boldsymbol{x}\right)$, is upper bounded by $\mathrm{P_{ub}}(\hat{s}_k|\boldsymbol{x}) = \mathrm{P}\left(z_k \in \mathcal{Z}_1|\boldsymbol{x}\right) + \mathrm{P}\left(z_k \in \mathcal{Z}_2|\boldsymbol{x}\right)$, where $\mathcal{Z}_1$ and $\mathcal{Z}_2$, depicted in Fig. 4.1, are given by $\mathcal{Z}_1 = \{z_k : \|z_k - s_{i_-}\|_2 \leq \|z_k - s_i\|_2\}$ and $\mathcal{Z}_2 = \{s : \|z_k - s_{i_+}\|_2 \leq \|z_k - s_i\|_2\}$, with s_i being the i-th element of $\mathcal{S}$, index $i_- = \mathrm{mod}\left(i + \alpha_s - 2, \alpha_s\right) + 1$ and index $i_+ = \mathrm{mod}\left(i, \alpha_s\right) + 1$. The individual probabilities are computed based on the MDDTs, $d_{1,k}\left(\boldsymbol{x}\right)$ and $d_{2,k}\left(\boldsymbol{x}\right)$, as

$$\mathrm{P}\left(z_k \in \mathcal{Z}_1|\boldsymbol{x}\right) = \int_{d_{1,k}(\boldsymbol{x})}^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{t^2}{\sigma_w^2}} dt = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right), \tag{4-27}$$

$$\mathrm{P}\left(z_k \in \mathcal{Z}_2|\boldsymbol{x}\right) = \int_{d_{2,k}(\boldsymbol{x})}^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{t^2}{\sigma_w^2}} dt = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right). \tag{4-28}$$

The MDDTs are computed, similarly to in [14] and [18], by applying a rotation of $\arg\{s_k^*\} = -\phi_{s_k}$ to the coordinate system such that the symbol of interest is placed on the real axis. This is done by multiplying both s_k and y_k by $e^{-j\phi_{s_k}}$ which results in $e^{-j\phi_{s_k}} s_k = 1$ and $\chi_k = e^{-j\phi_{s_k}} y_k$. Based on the rotated coordinate system the MDDTs are computed as

$$d_{1,k}\left(\boldsymbol{x}\right) = \mathrm{Re}\left\{s_k^* \boldsymbol{h}_k \boldsymbol{x}\right\} \sin\theta - \mathrm{Im}\left\{s_k^* \boldsymbol{h}_k \boldsymbol{x}\right\} \cos\theta, \tag{4-29}$$

$$d_{2,k}\left(\boldsymbol{x}\right) = \mathrm{Re}\left\{s_k^* \boldsymbol{h}_k \boldsymbol{x}\right\} \sin\theta + \mathrm{Im}\left\{s_k^* \boldsymbol{h}_k \boldsymbol{x}\right\} \cos\theta. \tag{4-30}$$

With this, one can construct a bound on the probability of correct detection of the k-th user as

$$\begin{aligned}\mathrm{P}\left(\hat{s}_k|\boldsymbol{x}\right) &= 1-\mathrm{P_e}\left(\hat{s}_k|\boldsymbol{x}\right)\\ &\geq 1-\mathrm{P_{ub}}(\hat{s}_k|\boldsymbol{x})\\ &= \frac{1}{2}\mathrm{erf}\left(\frac{d_{1,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right)+\frac{1}{2}\mathrm{erf}\left(\frac{d_{2,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right),\end{aligned} \tag{4-31}$$

where $\mathrm{P_{ub}}(\hat{s}_k|\boldsymbol{x})$ is called the union-bound SEP of the k-th user. By combining the individual bounds, the probability of correct detection for all users, $\mathrm{P}(\hat{\boldsymbol{s}}|\boldsymbol{x})$, is lower bounded by

$$\mathrm{P_{ub}}(\hat{\boldsymbol{s}}|\boldsymbol{x})=\prod_{k=1}^{K}\mathrm{P_{ub}}(\hat{s}_k|\boldsymbol{x})=\left(\frac{1}{2}\right)^K\prod_{k=1}^{K}\left(\mathrm{erf}\left(\frac{d_{1,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right)+\mathrm{erf}\left(\frac{d_{2,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right)\right).$$

The minimum union-bound SEP (MUBSEP) optimization problem is written as the minimization of $-\ln\left(\mathrm{P_{ub}}(\hat{\boldsymbol{s}}|\boldsymbol{x})\right)$ as

$$\min_{\boldsymbol{x}\in\mathcal{X}^M}-\sum_{k=1}^{K}\ln\left(\mathrm{erf}\left(\frac{d_{1,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right)+\mathrm{erf}\left(\frac{d_{2,k}\left(\boldsymbol{x}\right)}{\sigma_w}\right)\right). \tag{4-32}$$

An equivalent real-valued formulation of (4-32) can be cast as

$$\begin{aligned}\min_{\boldsymbol{x}_\mathrm{r}}\ &-\sum_{k=1}^{K}\ln\left(\mathrm{erf}\left(\boldsymbol{u}_{1,k}^T\boldsymbol{x}_\mathrm{r}\right)+\mathrm{erf}\left(\boldsymbol{u}_{2,k}^T\boldsymbol{x}_\mathrm{r}\right)\right) \qquad (4\text{-}33)\\ \text{s.t.}\quad &\left[\boldsymbol{x}_\mathrm{r}\right]_{2m-1}+j\left[\boldsymbol{x}_\mathrm{r}\right]_{2m}\in\mathcal{X}\quad \text{for } m=1,\ldots,M.\end{aligned}$$

where $\boldsymbol{x}_\mathrm{r}=R(\boldsymbol{x})$, $\boldsymbol{u}_{1,k}=\left(\boldsymbol{h}_{\mathrm{R},\theta,k}^{s^*}-\boldsymbol{h}_{\mathrm{I},\theta,k}^{s^*}\right)^T$ and $\boldsymbol{u}_{2,k}=\left(\boldsymbol{h}_{\mathrm{R},\theta,k}^{s^*}+\boldsymbol{h}_{\mathrm{I},\theta,k}^{s^*}\right)^T$ with $\boldsymbol{h}_{\mathrm{R},\theta,k}^{s^*}$ and $\boldsymbol{h}_{\mathrm{I},\theta,k}^{s^*}$ being the k-th rows of matrices $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*}$. The matrices $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*}$ are given by $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*}=\sin(\theta)/\sigma_w\boldsymbol{H}_{\mathrm{R}}^{s^*}$, and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*}=\cos(\theta)/\sigma_w\boldsymbol{H}_{\mathrm{I}}^{s^*}$, with

$$\boldsymbol{H}_{\mathrm{R}}^{s^*}=\begin{bmatrix}\mathrm{Re}\{h_{11}^{s^*}\} & -\mathrm{Im}\{h_{11}^{s^*}\} & \cdots & \mathrm{Re}\{h_{1M}^{s^*}\} & -\mathrm{Im}\{h_{1M}^{s^*}\}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \mathrm{Re}\{h_{K1}^{s^*}\} & -\mathrm{Im}\{h_{K1}^{s^*}\} & \cdots & \mathrm{Re}\{h_{KM}^{s^*}\} & -\mathrm{Im}\{h_{KM}^{s^*}\}\end{bmatrix}$$

$$\boldsymbol{H}_{\mathrm{I}}^{s^*}=\begin{bmatrix}\mathrm{Im}\{h_{11}^{s^*}\} & \mathrm{Re}\{h_{11}^{s^*}\} & \cdots & \mathrm{Im}\{h_{1M}^{s^*}\} & \mathrm{Re}\{h_{1M}^{s^*}\}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \mathrm{Im}\{h_{K1}^{s^*}\} & \mathrm{Re}\{h_{K1}^{s^*}\} & \cdots & \mathrm{Re}\{h_{KM}^{s^*}\} & \mathrm{Im}\{h_{KM}^{s^*}\}\end{bmatrix},$$

where $h_{ij}^{s^*}$ is the element of the i-th row and j-th column of the matrix $\boldsymbol{H}^{s^*}=\mathrm{diag}\left\{s^*\right\}\boldsymbol{H}$.

Note that the equivalency between (4-32) and (4-33) implies ${}^{d_{1,k}(\boldsymbol{x})}/_{\sigma_w} = \boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}$ and ${}^{d_{2,k}(\boldsymbol{x})}/_{\sigma_w} = \boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}$. As discussed in section A.2 of the appendix, the proposed MUBSEP objective is convex for $d_{1,k}(\boldsymbol{x}) \geq 0$ and $d_{2,k}(\boldsymbol{x}) \geq 0, \forall k \in \{1,\ldots,K\}$, which can be reformulated in the form $\boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \boldsymbol{0}$ with $\boldsymbol{C} = \left[\left(\boldsymbol{H}_{\mathrm{R},\theta}^{s^*} - \boldsymbol{H}_{\mathrm{I},\theta}^{s^*}\right)^T, \left(\boldsymbol{H}_{\mathrm{R},\theta}^{s^*} + \boldsymbol{H}_{\mathrm{I},\theta}^{s^*}\right)^T\right]^T$. With this, a convex MUBSEP problem is formulated by including $\boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \boldsymbol{0}$ as an additional constraint which reads as

$$\begin{aligned} \min_{\boldsymbol{x}_\mathrm{r}} \; & -\sum_{k=1}^{K} \ln\left(\mathrm{erf}\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right) + \mathrm{erf}\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)\right) \\ \text{s.t.} \quad & \boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \boldsymbol{0}, \; [\boldsymbol{x}_\mathrm{r}]_{2m-1} + j\,[\boldsymbol{x}_\mathrm{r}]_{2m} \in \mathcal{X} \;\; \text{for } m = 1,\ldots,M. \end{aligned} \tag{4-34}$$

Due to the additional constraint, the optimal solution from (4-34) is, in general, a suboptimal solution of (4-33). However, different solutions mean that the solution of (4-33) violates $\boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \boldsymbol{0}$. This in turn implies that for at least one user i either $d_{1,i}(\boldsymbol{x}) < 0$ or $d_{2,i}(\boldsymbol{x}) < 0$. It is concluded by analyzing Fig. 4.1 that having $d_{1,i}(\boldsymbol{x}) < 0$ or $d_{2,i}(\boldsymbol{x}) < 0$ yields a noiseless received symbol y_i in the incorrect decision region leading to $\mathrm{P_e}(\hat{s}_i|\boldsymbol{x}) > 0.5$. With this, the optimal solution from (4-34) is only different from the one from (4-33) in high SEP settings. These are not relevant cases since future systems will be designed to provide high reliability and avoid these scenarios.

4.4.3 Precoding Algorithm Design

In this section, the previously presented formulations are utilized for the development of different low-resolution precoding algorithms. Since $\mathcal{X}$ is discrete, the optimization problems proposed in the previous section consist of the minimization of convex objectives over a discrete feasible set, which characterizes DPPs. In the section, solving several convex optimization problems is necessary. Although these problems are convex, using standard optimization tools runs into compatibility issues for MUBSEP. This is the case since, most standard optimization tools (e.g., CVX [42]) do not support the usage of the MUBSEP objective since its convexity is only guaranteed in the feasible set. Moreover, in practice, using standard optimization problem tools (e.g., fmincon) with the proposed objectives often runs into precision issues. With this, at the beginning of this section, an introduction to the algorithm implemented for solving the involved convex optimization problems is given.

Algorithm 2 Barrier Method

Inputs: Strictly feasible initial point $\boldsymbol{x}_0$, $t_0 > 0$, $\mu > 1$ and $\epsilon_{\text{tol}} > 0$ **Output**: $\boldsymbol{x}_{\text{opt}}$
Define $t = t_0$
Repeat
Compute $\boldsymbol{x}^*(t)$ by minimizing $f(\boldsymbol{x}) = tf_0(\boldsymbol{x}) + \phi(\boldsymbol{x})$ starting at $\boldsymbol{x}_0$ using Algorithm 3
Update $\boldsymbol{x}_0 = \boldsymbol{x}^*(t)$
Stopping criterion: **If** $\varphi/t \leq \epsilon_{\text{tol}} \rightarrow$ **Return** $\boldsymbol{x}_{\text{opt}} = \boldsymbol{x}_0$
Update $t = \mu t$

4.4.3.1 The Barrier Method

In this section, a particular case of the barrier method [42, Section 11.3] that solves convex optimization problems with only inequality constraints is presented. The optimization problems considered in this study have the following general form

$$\min_{\boldsymbol{x}} \ f_0(\boldsymbol{x}) \qquad \text{s.t.} \ \ f_i(\boldsymbol{x}) \leq 0 \quad \text{for } i = 1, \ldots, \varphi, \tag{4-35}$$

where the functions $f_i(\boldsymbol{x}) : \mathbb{R}^q \rightarrow \mathbb{R}$ for $i \in \{1, \ldots, \varphi\}$ are convex and twice continuously differentiable. The problem can be approximated as an unconstrained problem as

$$\min_{\boldsymbol{x}} \ tf_0(\boldsymbol{x}) + \phi(\boldsymbol{x}), \tag{4-36}$$

where $\phi(\boldsymbol{x}) = -\sum_{i=1}^{\varphi} \ln(-f_i(\boldsymbol{x}))$ is called the log-barrier function which serves as a penalty function to ensure attainability to the feasible set. As stated in [42, Section 11.2.2] solving (4-36) for a given value of t yields a solution $\boldsymbol{x}^*(t)$ that belongs to the φ/t-suboptimal set of the original problem (4-35). With this, the barrier method consists of sequentially computing $\boldsymbol{x}^*(t)$ for increasing values of t until $t \geq \varphi/\epsilon_{\text{tol}}$, where ϵ_{tol} is the given optimality tolerance. Regarding the initial value of t, termed t_0, the choice of any $t_0 > 0$ guarantees convergence, this study considers for implementation $t_0 = 1$. The steps of the method are summarized in Algorithm 2.

The utilization of Algorithm 2 implies a method for solving the unconstrained problem in (4-36). In this study, unconstrained problems are solved via the Newton method [42, Section 10.2.2] which requires the gradient and Hessian of the objective function. In the context of (4-36) the gradient and

Algorithm 3 Newton Method

Inputs: Starting point $\boldsymbol{x}_0 \in \mathbf{dom} f$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\epsilon_{\text{tol}} > 0$ **Output**: $\boldsymbol{x}_{\text{opt}}$
Repeat
Compute $\Delta\boldsymbol{x}_{\text{nt}} = -\nabla^2 f(\boldsymbol{x}_0)^{-1}\nabla f(\boldsymbol{x}_0)$ and $\lambda^2 = \nabla f(\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)^{-1}\nabla f(\boldsymbol{x}_0)$
Stopping criterion: **If** $\lambda^2/2 \leq \epsilon_{\text{tol}} \rightarrow$ **Return** $\boldsymbol{x}_{\text{opt}} = \boldsymbol{x}_0$
Choose a step size n via backtracking line search [42, Algorithm 9.2]
Update $\boldsymbol{x}_0 = \boldsymbol{x}_0 + n\Delta\boldsymbol{x}_{\text{nt}}$

Hessian of $f(\boldsymbol{x}) = tf_0(\boldsymbol{x}) + \phi(\boldsymbol{x})$ are defined as

$$\begin{aligned}\nabla f(\boldsymbol{x}) &= t\nabla f_0(\boldsymbol{x}) - \sum_{i=1}^{\varphi}\frac{\nabla f_i(\boldsymbol{x})}{f_i(\boldsymbol{x})},\\ \nabla^2 f(\boldsymbol{x}) &= t\nabla^2 f_0(\boldsymbol{x}) + \sum_{i=1}^{\varphi}\frac{\nabla f_i(\boldsymbol{x})\nabla f_i(\boldsymbol{x})^T}{f_i(\boldsymbol{x})^2} - \sum_{i=1}^{\varphi}\frac{\nabla^2 f_i(\boldsymbol{x})}{f_i(\boldsymbol{x})}.\end{aligned} \tag{4-37}$$

The implementation details are shown in Algorithm 3. Regarding the complexity of the barrier method, [42, Section 11.5.6] states that the best upper bound on the number of Newton steps required grows with $\sqrt{\varphi}$. Considering that the computation of a Newton step requires a matrix inversion, done with complexity $\mathcal{O}(q^3)$ via Gauss-Jordan elimination, an UBCO of the algorithm is given by $\mathcal{O}\left(\sqrt{\varphi}q^3\right)$.

4.4.3.2 Partial Greedy Search Precoding

Greedy search is a widely applied approach for discrete problem in wireless communications [55]. In this section, PGS precoding methods are proposed based on the optimization problems (4-17) and (4-34). The first step to assemble the proposed PGS algorithms is to relax the discrete feasible set $\mathcal{X}^M$ to its convex hull $\mathcal{P}$. As in [48, 27, 14], $\mathcal{P}$ is described with $\boldsymbol{A}\boldsymbol{x}_{\text{r}} - \boldsymbol{b} \preceq \boldsymbol{0}$, where

$$\boldsymbol{A} = \left[(\boldsymbol{I}_M \otimes \boldsymbol{\beta}_1)^T,\ \ (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_2)^T,\ \ \ldots,\ \ (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_{\alpha_x})^T\right]^T, \tag{4-38}$$

$$\boldsymbol{\beta}_i = \left[\cos\phi_i,\ \ -\sin\phi_i\right],\quad \phi_i = \frac{2\pi i}{\alpha_x},\quad i \in \{1, \ldots, \alpha_x\}, \tag{4-39}$$

$$\boldsymbol{b} = \sqrt{\mathrm{P_A}}\cos\left(\frac{\pi}{\alpha_x}\right)\boldsymbol{1}_{M\alpha_x}. \tag{4-40}$$

Replacing $\mathcal{X}^M$ by $\mathcal{P}$ yields real-valued convex optimization problems solvable utilizing the barrier method in the form of

$$\boldsymbol{x}_{\text{r,lb}} = \min_{\boldsymbol{x}_{\text{r}}} f_0(\boldsymbol{x}_{\text{r}}) \quad \text{s.t.}\ \ \boldsymbol{C}\boldsymbol{x}_{\text{r}} \preceq \boldsymbol{0},\quad \boldsymbol{A}\boldsymbol{x}_{\text{r}} - \boldsymbol{b} \preceq \boldsymbol{0}, \tag{4-41}$$

where f_0 is given by (4-20) for the MSEP case and given by (A-13) for the MUBSEP case, with the constraint $\boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \mathbf{0}$ only taken into account for the MUBSEP case. Note that, $\boldsymbol{x}_\mathrm{lb} = C(\boldsymbol{x}_\mathrm{r,lb}) \in \mathcal{P}$ can also belong to $\mathcal{X}^M$ as $\mathcal{P} \cap \mathcal{X}^M \neq \emptyset$. If this is the case, $\boldsymbol{x}_\mathrm{lb}$ is the optimal solution from the original DPP and can be utilized for transmission without requiring further processing. However, if $\boldsymbol{x}_\mathrm{lb} \notin \mathcal{X}^M$ the solution $\boldsymbol{x}_\mathrm{lb}$ must be projected to $\mathcal{X}^M$. The projection step considered consists of two stages. First, elementwise smallest distance projection (ESDP) is performed such that the projected vector reads as $\boldsymbol{x}_\mathrm{ub} = Q(\boldsymbol{x}_\mathrm{lb})$, where $Q(\cdot)$ represents the ESDP operation. By this method $[\boldsymbol{x}_\mathrm{ub}]_p$ is computed as $[\boldsymbol{x}_\mathrm{ub}]_p = \arg\min_{i\in\{1\dots\alpha_x\}} \left|[\boldsymbol{x}_\mathrm{lb}]_p - x_i\right|^2$, where x_i the i-th element of $\mathcal{X}$. Note that, the projected vector $\boldsymbol{x}_\mathrm{ub}$ attains the low-resolution constraints, meaning $\boldsymbol{x}_\mathrm{ub} \in \mathcal{X}^M$, and could be used for transmission. Having ESDP-based projection, although practical, causes a significant loss in performance. To mitigate this performance degradation the second step of the projection algorithm consists of utilizing PGS as a local optimization approach. The PGS projection method starts by determining $\mathcal{T} = \left\{p : [\boldsymbol{x}_\mathrm{lb}]_p \notin \mathcal{X}\right\}$ and computing $\boldsymbol{x}_\mathrm{ub} = Q(\boldsymbol{x}_\mathrm{lb})$ as the initial vector. Then, for each $p \in \mathcal{T}$ and for all $i \in \{1, \dots, \alpha_x\}$, the algorithm replaces $[\boldsymbol{x}_\mathrm{ub}]_p$ by $x_i \in \mathcal{X}$ and computes some objective $g(\cdot)$. By this, the algorithm aims to determine the value of $x_i \in \mathcal{X}$ that minimizes $g(\cdot)$ and update $[\boldsymbol{x}_\mathrm{ub}]_p = x_i$ for each $p \in \mathcal{T}$. The objective $g(\cdot)$ depends on the chosen criterion and is discussed in the following subsections. After all $p \in \mathcal{T}$ were considered the output vector $\boldsymbol{x}_\mathrm{pgs} = \boldsymbol{x}_\mathrm{ub}$ is utilized for transmission. This second step of the projection algorithm is denoted by the operator $P(\cdot)$, such that $\boldsymbol{x}_\mathrm{pgs} = P(Q(\boldsymbol{x}_\mathrm{lb}))$.

Proposed MSEP PGS algorithm The design of the proposed MSEP PGS algorithm starts by considering the relaxation of the discrete feasible set in (4-17) to its convex hull which yields

$$\begin{aligned} \min_{\boldsymbol{x}_\mathrm{r}} \; & -\sum_{k=1}^{K} \ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^T \boldsymbol{x}_\mathrm{r}\right)\right) + \ln\left(\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^T \boldsymbol{x}_\mathrm{r}\right)\right) \\ \text{s.t.} \; & \boldsymbol{A}\boldsymbol{x}_\mathrm{r} - \boldsymbol{b} \preceq \mathbf{0}. \end{aligned} \tag{4-42}$$

As stated in section 4.4.3.1, solving (4-42) with the barrier method requires the gradient and Hessian of the objective and constraint functions. For the MSEP objective, $\nabla f_0(\boldsymbol{x}_\mathrm{r})$ and $\nabla^2 f_0(\boldsymbol{x}_\mathrm{r})$ are given in (4-21) and (4-22), respectively. The constraint functions of problem (4-42) are described by $f_i(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{a}_i\boldsymbol{x}_\mathrm{r} - b_i$, for $i = 1, \dots, M\alpha_x$, with $\boldsymbol{a}_i$ being the i-th row of $\boldsymbol{A}$ and b_i being the i-th element of $\boldsymbol{b}$. With this, $\nabla f_i(\boldsymbol{x}_\mathrm{r})$ and $\nabla^2 f_i(\boldsymbol{x}_\mathrm{r})$ for the i-th constraint function

read as

$$\nabla f_i(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{a}_i, \quad \nabla^2 f_i(\boldsymbol{x}_\mathrm{r}) = \mathbf{0}. \tag{4-43}$$

After solving (4-42) the algorithm determines $\mathcal{T} = \left\{p : [\boldsymbol{x}_\mathrm{lb}]_p \notin \mathcal{X}\right\}$, computes $\boldsymbol{x}_\mathrm{ub} = Q\left(C\left(\boldsymbol{x}_\mathrm{r,lb}\right)\right)$ and proceeds to sequentially update $\boldsymbol{x}_\mathrm{ub}$ based on $g(\cdot)$. For MSEP the objective $g(\cdot)$ reads as

$$g\left(\boldsymbol{x}\right) = -\mathbf{1}_K^T \left(\ln\left(\Phi\left(\boldsymbol{S}_\mathrm{R}\ \mathrm{Re}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right)\right) + \ln\left(\Phi\left(\boldsymbol{S}_\mathrm{I}\ \mathrm{Im}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right)\right)\right), \tag{4-44}$$

where $S_\mathrm{R} = \sqrt{2}/\sigma_w\ \mathrm{diag}\left(\mathrm{sign}(\mathrm{Re}\left\{\boldsymbol{s}\right\})\right)$ and $S_\mathrm{I} = \sqrt{2}/\sigma_w\ \mathrm{diag}\left(\mathrm{sign}(\mathrm{Im}\left\{\boldsymbol{s}\right\})\right)$. The steps of the MSEP PGS algorithm are summarized in Algorithm 4. The algorithm's complexity is dominated by the complexity of the barrier method which, as stated in section 4.4.3.1, is upper bounded by $\mathcal{O}\left(\sqrt{\varphi}q^3\right)$, with φ being the number of inequality constraints and q being the number of optimization variables. Substituting $\varphi = M\alpha_x$ and $q = 2M$ yields UBCO of $\mathcal{O}\left(M^{3.5}\right)$.

Proposed MUBSEP PGS algorithm The design of the proposed MUBSEP PGS algorithm starts with the relaxation of the discrete feasible set in (4-34) to its convex hull which yields

$$\begin{aligned} \min_{\boldsymbol{x}_\mathrm{r}}\ & -\sum_{k=1}^{K} \ln\left(\mathrm{erf}\left(\boldsymbol{u}_{1,k}^T\boldsymbol{x}_\mathrm{r}\right) + \mathrm{erf}\left(\boldsymbol{u}_{2,k}^T\boldsymbol{x}_\mathrm{r}\right)\right) \\ \text{s.t.}\ & \boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \mathbf{0}, \quad \boldsymbol{A}\boldsymbol{x}_\mathrm{r} - \boldsymbol{b} \preceq \mathbf{0}. \end{aligned} \tag{4-45}$$

Solving (4-45) with the barrier method requires the gradient and Hessian of the objective and constraint functions. For MUBSEP, $f_0(\boldsymbol{x}_\mathrm{r})$ and $\nabla^2 f_0(\boldsymbol{x}_\mathrm{r})$ are given in (A-14) and (A-17), respectively. The constraint functions of (4-45), read as $f_i(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{a}_i\boldsymbol{x}_\mathrm{r} - b_i$, for $i = 1, \ldots, M\alpha_x$, and $f_i(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{c}_i\boldsymbol{x}_\mathrm{r}$ for $i = M\alpha_x + 1, \ldots, M\alpha_x + 2K$. With this, the gradient and Hessian of the first $M\alpha_x$ constraints are given in (4-43), and, for $i = M\alpha_x + 1, \ldots, M\alpha_x + 2K$, $\nabla f_i(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{c}_i$ and $\nabla^2 f_i(\boldsymbol{x}_\mathrm{r}) = \mathbf{0}$.

After solving (4-42) the algorithm determines $\mathcal{T} = \left\{p : [\boldsymbol{x}_\mathrm{lb}]_p \notin \mathcal{X}\right\}$, computes $\boldsymbol{x}_\mathrm{ub} = Q\left(C\left(\boldsymbol{x}_\mathrm{r,lb}\right)\right)$ and proceeds to sequentially update $\boldsymbol{x}_\mathrm{ub}$ based on $g(\cdot)$. For MUBSEP the objective $g(\cdot)$ reads as

$$g(\boldsymbol{x}) = -\mathbf{1}_K^T \ln\left(\mathrm{erf}\left(\boldsymbol{\rho}_1\right) + \mathrm{erf}\left(\boldsymbol{\rho}_2\right)\right), \tag{4-46}$$

Algorithm 4 MSEP/MUBSEP PGS Algorithm

Inputs: $\boldsymbol{H}$, $\boldsymbol{s}$, $\mathcal{X}$, θ, $\mathrm{P_A}$, Criterion and $g(\cdot)$ **Output**: $\boldsymbol{x}_{\mathrm{pgs}}, \mathcal{T}$
if Criterion = MSEP $\rightarrow$ Solve (4-42) to get $\boldsymbol{x}_{\mathrm{r,lb}}$ and compute $\boldsymbol{x}_{\mathrm{lb}} = C\left(\boldsymbol{x}_{\mathrm{r,lb}}\right)$
if Criterion = MUBSEP $\rightarrow$ Solve (4-45) to get $\boldsymbol{x}_{\mathrm{r,lb}}$ and compute $\boldsymbol{x}_{\mathrm{lb}} = C\left(\boldsymbol{x}_{\mathrm{r,lb}}\right)$
Construct the set $\mathcal{T} = \left\{p : [\boldsymbol{x}_{\mathrm{lb}}]_p \notin \mathcal{X}\right\}$ and compute $\boldsymbol{x}_{\mathrm{ub}} = Q\left(\boldsymbol{x}_{\mathrm{lb}}\right)$
for $p \in \mathcal{T}$ **do**
  for $i = 1 : \alpha_x$ **do**
    if Criterion = MSEP $\rightarrow$ Fix $[\boldsymbol{x}_{\mathrm{ub}}]_p$ as x_i and compute the $g_p^i = g(\boldsymbol{x}_{\mathrm{ub}})$ using (4-44)
    if Criterion = MUBSEP $\rightarrow$ Fix $[\boldsymbol{x}_{\mathrm{ub}}]_p$ as x_i and compute the $g_p^i = g(\boldsymbol{x}_{\mathrm{ub}})$ using (4-46)
  end for
  Update the p-th entry of $\boldsymbol{x}_{\mathrm{ub}}$ as $[\boldsymbol{x}_{\mathrm{ub}}]_p = x_i$ with $i = \underset{i=1,\ldots,\alpha_x}{\arg\min}\, g_p^i$
end for
The output vector is given by $\boldsymbol{x}_{\mathrm{pgs}} = \boldsymbol{x}_{\mathrm{ub}}$

where

$$\boldsymbol{\rho}_1 = \mathrm{Re}\left\{\boldsymbol{H}_{\mathrm{s}}\boldsymbol{x}\right\} - \mathrm{Im}\left\{\boldsymbol{H}_{\mathrm{c}}\boldsymbol{x}\right\}, \quad \boldsymbol{\rho}_2 = \mathrm{Re}\left\{\boldsymbol{H}_{\mathrm{s}}\boldsymbol{x}\right\} + \mathrm{Im}\left\{\boldsymbol{H}_{\mathrm{c}}\boldsymbol{x}\right\}, \tag{4-47}$$

where $\boldsymbol{H}_{\mathrm{s}} = {}^{\sin(\theta)}/_{\sigma_w}\left(\mathrm{diag}(\boldsymbol{s}^*)\boldsymbol{H}\right)$ and $\boldsymbol{H}_{\mathrm{c}} = {}^{\cos(\theta)}/_{\sigma_w}\left(\mathrm{diag}(\boldsymbol{s}^*)\boldsymbol{H}\right)$. The steps for PGS projection are detailed in Algorithm 4. The algorithm's complexity is dominated by the complexity of the barrier method, which, for the MUBSEP case, is calculated using $\varphi = M\alpha_x + 2K$ and $q = 2M$. With this, the algorithm yields an UBCO of $\mathcal{O}\left(M^3\sqrt{M+K}\right)$.

4.4.3.3 Precoding via QoS Branch-and-Bound

This section proposes a B&B method that accepts as a solution any vector $\boldsymbol{x}$ that attains the condition $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right) \preceq \boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is the QoS constraint vector and $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right)$ relates to each user's SEP. If attaining the condition is not possible for $\boldsymbol{x} \in \mathcal{X}^M$ the algorithm computes the optimal solution $\boldsymbol{x}_{\mathrm{opt}}$ of the corresponding DPP. In the MSEP case, $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right)$ is computed as

$$\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right) = \mathbf{1}_K - \Phi\left(\boldsymbol{S}_{\mathrm{R}}\,\mathrm{Re}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right) \circ \Phi\left(\boldsymbol{S}_{\mathrm{I}}\,\mathrm{Im}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right), \tag{4-48}$$

where $\circ$ denotes elementwise multiplication. For MSEP, $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right)$ denotes the exact SEP vector, meaning that $\forall k \in \{1,\ldots,K\}$, $\left[\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right)\right]_k = \mathrm{P_e}\left(\hat{s}_k|\boldsymbol{x}\right)$. For MUBSEP, $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right)$ is given by

$$\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}\right) = {}^1\!/_2\,\mathrm{erfc}\left(\boldsymbol{\rho}_1\right) + {}^1\!/_2\,\mathrm{erfc}\left(\boldsymbol{\rho}_2\right), \tag{4-49}$$

with $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ given in (4-47). In the MUBSEP case, $\mathbf{P}_{\mathrm{e}}(\boldsymbol{x})$ represents an upper bound on the SEP vector, meaning, that for all $k \in \{1,\ldots,K\}$, $\mathrm{P}_{\mathrm{e}}(\hat{s}_k|\boldsymbol{x}) \leq [\mathbf{P}_{\mathrm{e}}(\boldsymbol{x})]_k$. Note that, having a solution vector $\boldsymbol{x}$ that attains $\mathbf{P}_{\mathrm{e}}(\boldsymbol{x}) \preceq \boldsymbol{\lambda}$ implies that $\mathrm{P}_{\mathrm{e}}(\hat{s}_k|\boldsymbol{x}) \leq [\boldsymbol{\lambda}]_k$ for all $k \in \{1,\ldots,K\}$.

The proposed QoS B&B algorithm consists of two phases, namely the PGS stage and the Tree Search Based Precoding (TSBP) stage. For a given formulation, the first stage consists of executing the corresponding PGS method and evaluating the stopping criteria. If the stopping criteria are not met, the algorithm proceeds to the TSBP part where the practical PGS solution is continuously enhanced until either the SEP requirement is attained or $\boldsymbol{x}_{\mathrm{opt}}$ is computed.

QoS B&B PGS Stage For a given objective function $f_0(\boldsymbol{x}_{\mathrm{r}})$ (given in (4-20) for MSEP case or in (A-13) for MUBSEP) the DPPs proposed can be written in the following form

$$\begin{aligned}\boldsymbol{x}_{\mathrm{r,opt}} = &\arg\min_{\boldsymbol{x}_{\mathrm{r}}} f_0(\boldsymbol{x}_{\mathrm{r}}) \\ &\text{s.t.} \quad \boldsymbol{C}\boldsymbol{x}_{\mathrm{r}} \preceq \mathbf{0}, \quad [\boldsymbol{x}_{\mathrm{r}}]_{2m-1} + j\,[\boldsymbol{x}_{\mathrm{r}}]_{2m} \in \mathcal{X}, \text{ for } m \in \{1,\ldots,M\},\end{aligned} \tag{4-50}$$

where $\boldsymbol{x}_{\mathrm{r,opt}} = R(\boldsymbol{x}_{\mathrm{opt}})$ and $\boldsymbol{C}\boldsymbol{x}_{\mathrm{r}} \preceq \mathbf{0}$ is only taken into account for the MUBSEP case. A practical solution to (4-50) can be computed via executing Algorithm 4 with the criterion correspondent to the objective $f_0(\boldsymbol{x}_{\mathrm{r}})$, which yields the output solution $\boldsymbol{x}_{\mathrm{pgs}}$ and the set $\mathcal{T}$. During the QoS B&B PGS stage two stopping conditions are considered. The first evaluates the optimality of the PGS solution by checking if $\mathcal{T} = \varnothing$. If this condition holds it means that the solution of the relaxed problem $\boldsymbol{x}_{\mathrm{lb}}$ is already in the original feasible set $\mathcal{X}^M$. For this case, $\boldsymbol{x}_{\mathrm{pgs}} = \boldsymbol{x}_{\mathrm{lb}} = \boldsymbol{x}_{\mathrm{opt}}$ and the algorithm terminates with the transmit vector $\boldsymbol{x}_{\mathrm{pgs}}$. The second condition evaluates if the SEP requirement is attained, with this, the algorithm terminates if $\mathbf{P}_{\mathrm{e}}(\boldsymbol{x}_{\mathrm{pgs}}) \preceq \boldsymbol{\lambda}$ also returning $\boldsymbol{x}_{\mathrm{pgs}}$ as the transmit vector. If neither condition is satisfied the algorithm returns $\check{g} = g(\boldsymbol{x}_{\mathrm{pgs}})$ and $\check{\boldsymbol{x}} = \boldsymbol{x}_{\mathrm{pgs}}$ as the initial smallest known upper bound and its corresponding vector, respectively, with $g(\cdot)$ given by (4-44) for MSEP and (4-46) for MUBSEP. This is relevant for the pruning process of the QoS B&B TSBP stage and is explained in later sections. The steps of the QoS B&B PGS Stage are summarized in Algorithm 5.

QoS B&B Tree Search Based Precoding Stage If neither stopping criteria are met at the PGS step, the proposed QoS B&B algorithm proceeds to the

Algorithm 5 Proposed QoS B&B PGS Stage

Inputs: $\boldsymbol{H}$, $\boldsymbol{s}$, σ_w, Criterion, $\boldsymbol{\lambda}$ **Output**: $\boldsymbol{x}_{\text{out}}$, $\check{\boldsymbol{x}}$, $\check{g}$
if Criterion = MSEP
Execute Algorithm 4 with Criterion = MSEP to get $\boldsymbol{x}_{\text{pgs}}$ and $\mathcal{T}$ and compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{pgs}})$ using (4-48)
if $\mathcal{T} = \varnothing \wedge \mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{pgs}}) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_{\text{pgs}}$
Compute $g(\boldsymbol{x}_{\text{pgs}})$ with (4-44) and **return** $\check{\boldsymbol{x}} = \boldsymbol{x}_{\text{pgs}}$, $\check{g} = g(\boldsymbol{x}_{\text{pgs}})$, $\boldsymbol{x}_{\text{out}} = []$
else if Criterion = MUBSEP
Execute Algorithm 4 with Criterion = MUBSEP to get $\boldsymbol{x}_{\text{pgs}}$ and $\mathcal{T}$ and compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{pgs}})$ using (4-49)
if $\mathcal{T} = \varnothing \wedge \mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{pgs}}) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_{\text{pgs}}$
Compute $g(\boldsymbol{x}_{\text{pgs}})$ with (4-46) and **return** $\check{\boldsymbol{x}} = \boldsymbol{x}_{\text{pgs}}$, $\check{g} = g(\boldsymbol{x}_{\text{pgs}})$, $\boldsymbol{x}_{\text{out}} = []$
end if

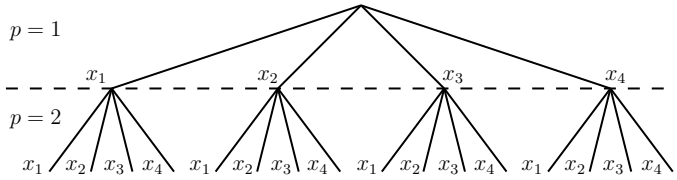

Figure 4.2: Tree representation of the set $\mathcal{X}^M$ for a system with $M = 2$ BS antennas and QPSK precoding modulation ($\alpha_x = 4$)

TSBP stage where the tree represents the set $\mathcal{X}^M$. The tree is constructed considering that the p-th BS antenna represents the p-th layer and each possible subvector $\boldsymbol{f} \in \mathcal{X}^p$ represents one branch. An example of a tree for a system with two transmit antennas and QPSK signaling is shown in Fig. 4.2. The QoS B&B TSBP stage performs breadth-first search in the feasible set $\mathcal{X}^M$ to minimize a given objective $f_0(\boldsymbol{x}_{\text{r}})$ which either represents MUBSEP or the MSEP criterion. While minimizing $f_0(\boldsymbol{x}_{\text{r}})$ if an intermediate solution $\boldsymbol{x}_{\text{int}}$ attains the condition $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{int}}) \preceq \boldsymbol{\lambda}$ the algorithm terminates with $\boldsymbol{x}_{\text{int}}$ as the transmit vector. The TSBP stage starts at the first layer $p = 1$ by fixing $2p$ entries of $\boldsymbol{x}_{\text{r}}$ such that the precoding vector becomes $\boldsymbol{x}_{\text{r}} = [\boldsymbol{f}_{\text{r},i}^T, \boldsymbol{v}_{\text{r}}^T]^T$, with $C(\boldsymbol{f}_{\text{r},i}) \in \mathcal{X}^p$. With this, a subproblem is constructed by rewriting (4-50) as

$$\begin{aligned} \boldsymbol{v}_{\text{r,opt}} &= \arg\min_{\boldsymbol{v}_{\text{r}}} f_0(\boldsymbol{v}_{\text{r}}, \boldsymbol{f}_{\text{r},i}) \qquad (4\text{-}51) \\ \text{s.t.} \;\; & \boldsymbol{C}[\boldsymbol{f}_{\text{r},i}^T, \boldsymbol{v}_{\text{r}}^T]^T \preceq \mathbf{0}, \quad [\boldsymbol{v}_{\text{r}}]_{2m-1} + j\,[\boldsymbol{v}_{\text{r}}]_{2m} \in \mathcal{X}, \;\text{for}\; \; m \in \{1, \ldots, M-p\}, \end{aligned}$$

where the constraint $\boldsymbol{C}[\boldsymbol{f}_{\text{r},i}^T, \boldsymbol{v}_{\text{r}}^T]^T \preceq \mathbf{0}$ is only taken into account for the MUBSEP case. A lower bounding problem on $f_0(\boldsymbol{v}_{\text{r,opt}})$ is obtained by relaxing feasible set $\mathcal{X}^{M-p}$ to its convex hull $\mathcal{J}$, which yields

$$\begin{aligned} \boldsymbol{v}_{\text{r,lb}} = &\arg\min_{\boldsymbol{v}_{\text{r}}} f_0(\boldsymbol{v}_{\text{r}}, \boldsymbol{f}_{\text{r},i}) \\ &\text{s.t.} \;\; \boldsymbol{C}[\boldsymbol{f}_{\text{r},i}^T, \boldsymbol{v}_{\text{r}}^T]^T \preceq \mathbf{0}, \;\; \text{for} \;\; C(\boldsymbol{v}_{\text{r}}) \in \mathcal{J}. \qquad (4\text{-}52) \end{aligned}$$

Solving (4-52) and evaluating $f_0(\boldsymbol{x}_{\mathrm{r,lb},i})$, with $\boldsymbol{x}_{\mathrm{r,lb},i} = \left[\boldsymbol{f}_{\mathrm{r},i}^T, \boldsymbol{v}_{\mathrm{r,lb}}^T\right]^T$, yields a lower bound on $f_0(\boldsymbol{v}_{\mathrm{r,opt}}, \boldsymbol{f}_{\mathrm{r},i})$, meaning that $f_0(\boldsymbol{x}_{\mathrm{r,lb},i}) \leq f_0(\boldsymbol{v}_{\mathrm{r,opt}}, \boldsymbol{f}_{\mathrm{r},i})$. Note that, if $\boldsymbol{f}_{\mathrm{r},i}$ is a subvector of $\boldsymbol{x}_{\mathrm{r,opt}}$ then $f_0(\boldsymbol{x}_{\mathrm{r,lb},i}) \leq f_0(\boldsymbol{v}_{\mathrm{r,opt}}, \boldsymbol{f}_{\mathrm{r},i}) = f_0(\boldsymbol{x}_{\mathrm{r,opt}})$. On the other hand, if $f_0(\boldsymbol{x}_{\mathrm{r,lb},i}) > f_0(\boldsymbol{x}_{\mathrm{r,opt}})$ then $\boldsymbol{f}_{\mathrm{r},i}$ cannot be a subvector of $\boldsymbol{x}_{\mathrm{r,opt}}$ and $f_0(\boldsymbol{v}_{\mathrm{r,opt}}, \boldsymbol{f}_{\mathrm{r},i}) > f_0(\boldsymbol{x}_{\mathrm{r,opt}})$. An upper bound on $f_0(\boldsymbol{v}_{\mathrm{r,opt}}, \boldsymbol{f}_{\mathrm{r},i})$ is computed by projecting the vector $\boldsymbol{v}_{\mathrm{lb}} = C(\boldsymbol{v}_{\mathrm{r,lb}})$ to $\mathcal{X}^{(M-p)}$ as $\boldsymbol{v}_{\mathrm{ub}} = P(Q(\boldsymbol{v}_{\mathrm{lb}}))$, with $P(\cdot)$ and $Q(\cdot)$ being the projection operators introduced in section 4.4.3.2, and computing $g(\boldsymbol{x}_{\mathrm{ub},i}, \boldsymbol{f}_i)$ with $\boldsymbol{x}_{\mathrm{ub},i} = \left[\boldsymbol{f}_i^T, \boldsymbol{v}_{\mathrm{ub}}^T\right]^T$. The solution $\boldsymbol{x}_{\mathrm{ub},i}$ attains the low-resolution constraints and is evaluated against the condition $\mathbf{P}_{\mathrm{e}}\left(\boldsymbol{x}_{\mathrm{ub},i}\right) \preceq \boldsymbol{\lambda}$. If the condition holds the algorithm terminates with $\boldsymbol{x}_{\mathrm{ub},i}$ as the transmit vector. Otherwise the algorithm proceeds by fixing the next subvector $\boldsymbol{f}_{\mathrm{r},i+1}$ and solving the corresponding subproblem, i.e. evaluating the next branch of the layer. After all branches in one layer were evaluated they are subjected to the pruning process where the search set is reduced by eliminating $\boldsymbol{f}_i$ that cannot be part of the optimal solution and the algorithm goes to the next layer, i.e., updates $p = p + 1$. In the following, the MSEP and MUBSEP lower bounding subproblems are derived and the pruning process is presented.

MSEP Lower Bounding Subproblem Formulation The MSEP subproblems are written considering the minimization of the objective described in (4-20) for $\boldsymbol{x}_{\mathrm{r}} = [\boldsymbol{f}_{\mathrm{r},i}^T, \boldsymbol{v}_{\mathrm{r}}^T]^T$. To this end, we split $\boldsymbol{H}_{\mathrm{R}} = \left[\boldsymbol{G}_{\mathrm{R}}\ ,\ \boldsymbol{T}_{\mathrm{R}}\right]$ and $\boldsymbol{H}_{\mathrm{I}} = \left[\boldsymbol{G}_{\mathrm{I}}\ ,\ \boldsymbol{T}_{\mathrm{I}}\right]$ where $\boldsymbol{G}_{\mathrm{R}}$ and $\boldsymbol{G}_{\mathrm{I}}$ consist of the first $2p$ columns of $\boldsymbol{H}_{\mathrm{R}}$ and $\boldsymbol{H}_{\mathrm{I}}$, respectively, and $\boldsymbol{T}_{\mathrm{R}}$ and $\boldsymbol{T}_{\mathrm{I}}$ consist of the subsequent $2(M-p)$ columns of $\boldsymbol{H}_{\mathrm{R}}$ and $\boldsymbol{H}_{\mathrm{I}}$, respectively. Considering that $\boldsymbol{A}'$ is obtained by selecting the last $2\left(M-p\right)$ columns of $\boldsymbol{A}$, the MSEP subproblem conditioned on $\boldsymbol{f}_{\mathrm{r},i}$ reads as

$$\begin{aligned} \min_{\boldsymbol{v}_{\mathrm{r}}}\ & -\sum_{k=1}^{K}\left(\ln\left(\Phi\left(\boldsymbol{g}_{\mathrm{R},k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T\boldsymbol{v}_{\mathrm{r}}\right)\right) + \ln\left(\Phi\left(\boldsymbol{g}_{\mathrm{I},k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T\boldsymbol{v}_{\mathrm{r}}\right)\right)\right) \\ \text{s.t.}\ & \boldsymbol{A}'\boldsymbol{v}_{\mathrm{r}} - \boldsymbol{b} \preceq \mathbf{0}, \end{aligned} \tag{4-53}$$

where $\boldsymbol{g}_{\mathrm{R},k}^T$, $\boldsymbol{g}_{\mathrm{I},k}^T$, $\boldsymbol{t}_{\mathrm{R},k}^T$, and $\boldsymbol{t}_{\mathrm{I},k}^T$ are the k-th rows of the matrices $\boldsymbol{G}_{\mathrm{R}}$, $\boldsymbol{G}_{\mathrm{I}}$, $\boldsymbol{T}_{\mathrm{R}}$, and $\boldsymbol{T}_{\mathrm{I}}$, respectively. The subproblems are solvable using the barrier method presented in section 4.4.3.1, which requires the gradient and Hessian of the objective and constraint functions. For the MSEP subproblems (4-53), $\nabla f_0(\boldsymbol{x})$

and $\nabla^2 f_0(\boldsymbol{x})$ are given by

$$\nabla f_0(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{\boldsymbol{n}_{\mathrm{R},k}(\boldsymbol{x}_\mathrm{r})}{\Phi\left(\boldsymbol{g}_{\mathrm{R},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T \boldsymbol{v}_\mathrm{r}\right)} + \frac{\boldsymbol{n}_{\mathrm{I},k}(\boldsymbol{x}_\mathrm{r})}{\Phi\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)}, \tag{4-54}$$

$$\nabla^2 f_0(\boldsymbol{x}) = \sum_{k=1}^{K} \frac{\boldsymbol{n}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right) \boldsymbol{n}_{R,k}^T\left(\boldsymbol{x}_\mathrm{r}\right) + \boldsymbol{\Upsilon}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)\right)^2} + \frac{\boldsymbol{n}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right) \boldsymbol{n}_{I,k}^T\left(\boldsymbol{x}_\mathrm{r}\right) + \boldsymbol{\Upsilon}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)\right)^2}, \tag{4-55}$$

where

$$\boldsymbol{n}_{R,k} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{g}_{\mathrm{R},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T \boldsymbol{v}_\mathrm{r}\right)^2}{2}} \boldsymbol{t}_{\mathrm{R},k}, \quad \boldsymbol{n}_{I,k} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)^2}{2}} \boldsymbol{t}_{\mathrm{I},k}, \tag{4-56}$$

$$\boldsymbol{\Upsilon}_{R,k} = \frac{\Phi\left(\boldsymbol{g}_{\mathrm{R},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T \boldsymbol{v}_\mathrm{r}\right)}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{g}_{\mathrm{R},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T \boldsymbol{v}_\mathrm{r}\right)^2}{2}} \boldsymbol{t}_{\mathrm{R},k}\left(\boldsymbol{g}_{\mathrm{R},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{R},k}^T \boldsymbol{v}_\mathrm{r}\right) \boldsymbol{t}_{\mathrm{R},k}^T, \tag{4-57}$$

$$\boldsymbol{\Upsilon}_{I,k} = \frac{\Phi\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right)^2}{2}} \boldsymbol{t}_{\mathrm{I},k}\left(\boldsymbol{g}_{\mathrm{I},k}^T \boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{\mathrm{I},k}^T \boldsymbol{v}_\mathrm{r}\right) \boldsymbol{t}_{\mathrm{I},k}^T. \tag{4-58}$$

The constraint functions, in the case of problem (4-53), are described by $f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{a}_i' \boldsymbol{v}_\mathrm{r} - b_i$, for $i = 1, \ldots, M\alpha_x$, with $\boldsymbol{a}_i'$ being the i-th row of $\boldsymbol{A}'$. With this, $\nabla f_i(\boldsymbol{x}_\mathrm{r})$ and $\nabla^2 f_i(\boldsymbol{x}_\mathrm{r})$ for the i-th constraint function read as

$$\nabla f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{a}_i', \quad \nabla^2 f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{0}. \tag{4-59}$$

As mentioned in section 4.4.3.1 the UBCO of the barrier method is $\mathcal{O}(\sqrt{\varphi} q^3)$. For the MSEP subproblems $\varphi = M\alpha_x$ and $q = 2(M-p)$, which yields $\mathcal{O}\left(M^{3.5}\right)$.

MUBSEP Lower Bounding Subproblem Formulation For the MUBSEP case, we split $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*} = \left[\boldsymbol{G}_{\mathrm{R},\theta}^{s^*},\ \boldsymbol{T}_{\mathrm{R},\theta}^{s^*}\right]$ and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*} = \left[\boldsymbol{G}_{\mathrm{I},\theta}^{s^*},\ \boldsymbol{T}_{\mathrm{I},\theta}^{s^*}\right]$ where $\boldsymbol{G}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{G}_{\mathrm{I},\theta}^{s^*}$ consist of the first $2p$ columns of $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*}$, respectively and $\boldsymbol{T}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{T}_{\mathrm{I},\theta}^{s^*}$ consist of the subsequent $2(M-p)$ columns of $\boldsymbol{H}_{\mathrm{R},\theta}^{s^*}$ and $\boldsymbol{H}_{\mathrm{I},\theta}^{s^*}$, respectively. To write the subproblems, we first define the vectors $\boldsymbol{g}_{1,k}^T$, $\boldsymbol{g}_{2,k}^T$ and $\boldsymbol{t}_k^T$ as the k-th rows of the matrices $\boldsymbol{G}_1 = \boldsymbol{G}_{\mathrm{R},\theta}^{s^*} - \boldsymbol{G}_{\mathrm{I},\theta}^{s^*}$, $\boldsymbol{G}_2 = \boldsymbol{G}_{\mathrm{R},\theta}^{s^*} + \boldsymbol{G}_{\mathrm{I},\theta}^{s^*}$ and $\boldsymbol{T}_1 = \boldsymbol{T}_{\mathrm{R},\theta}^{s^*} - \boldsymbol{T}_{\mathrm{I},\theta}^{s^*}$, $\boldsymbol{T}_2 = \boldsymbol{T}_{\mathrm{R},\theta}^{s^*} + \boldsymbol{T}_{\mathrm{I},\theta}^{s^*}$, respectively. With this, the MUBSEP subproblem is given by

$$\begin{aligned} \min_{\boldsymbol{v}_\mathrm{r}} & -\sum_{k=1}^{K} \ln\left(\operatorname{erf}\left(\boldsymbol{g}_{1,k}^T \boldsymbol{f}_\mathrm{r} + \boldsymbol{t}_{1,k}^T \boldsymbol{v}_\mathrm{r}\right) + \operatorname{erf}\left(\boldsymbol{g}_{2,k}^T \boldsymbol{f}_\mathrm{r} + \boldsymbol{t}_{2,k}^T \boldsymbol{v}_\mathrm{r}\right)\right) \\ \text{s.t.} & \quad \boldsymbol{R}\boldsymbol{v}_\mathrm{r} - \boldsymbol{l} \preceq \boldsymbol{0}, \quad \boldsymbol{A}'\boldsymbol{v}_\mathrm{r} - \boldsymbol{b} \preceq \boldsymbol{0}. \end{aligned} \tag{4-60}$$

where $\boldsymbol{R} = -\left[\boldsymbol{T}_1^T, \boldsymbol{T}_2^T\right]^T$ and $\boldsymbol{l} = \left[\boldsymbol{G}_1^T, \boldsymbol{G}_2^T\right]^T \boldsymbol{f}_{\mathrm{r},i}$. For MUBSEP subproblems (4-60), $\nabla f_0(\boldsymbol{x})$ and $\nabla^2 f_0(\boldsymbol{x})$ are given by

$$\nabla f_0(\boldsymbol{v}_\mathrm{r}) = -\sum_{k=1}^{K} \frac{\boldsymbol{\vartheta}_k\left(\boldsymbol{v}_\mathrm{r}\right)}{\varrho_k\left(\boldsymbol{v}_\mathrm{r}\right)}, \tag{4-61}$$

$$\nabla^2 f_0(\boldsymbol{v}_\mathrm{r}) = \sum_{k=1}^{K} \frac{\left(\boldsymbol{\Delta}_{1,k} + \boldsymbol{\Delta}_{2,k}\right)\varrho_k\left(\boldsymbol{v}_\mathrm{r}\right) + \boldsymbol{\vartheta}_k\left(\boldsymbol{v}_\mathrm{r}\right)\boldsymbol{\vartheta}_k^T\left(\boldsymbol{v}_\mathrm{r}\right)}{\left(\varrho_k\left(\boldsymbol{v}_\mathrm{r}\right)\right)^2}. \tag{4-62}$$

where

$$\boldsymbol{\vartheta}_k\left(\boldsymbol{v}_\mathrm{r}\right) = \frac{2}{\sqrt{\pi}}\left(e^{-\left(\boldsymbol{g}_{1,k}^T\boldsymbol{f}_{\mathrm{r},i}+\boldsymbol{t}_{1,k}^T\boldsymbol{v}_\mathrm{r}\right)^2}\boldsymbol{t}_{1,k} + e^{-\left(\boldsymbol{g}_{2,k}^T\boldsymbol{f}_{\mathrm{r},i}+\boldsymbol{t}_{2,k}^T\boldsymbol{v}_\mathrm{r}\right)^2}\boldsymbol{t}_{2,k}\right), \tag{4-63}$$

$$\varrho_k\left(\boldsymbol{v}_\mathrm{r}\right) = \mathrm{erf}\left(\boldsymbol{g}_{1,k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{1,k}^T\boldsymbol{v}_\mathrm{r}\right) + \mathrm{erf}\left(\boldsymbol{g}_{2,k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{2,k}^T\boldsymbol{v}_\mathrm{r}\right), \tag{4-64}$$

$$\boldsymbol{\Delta}_{1,k} = \frac{4}{\sqrt{\pi}}\left(e^{-\left(\boldsymbol{g}_{1,k}^T\boldsymbol{f}_{\mathrm{r},i}+\boldsymbol{t}_{1,k}^T\boldsymbol{v}_\mathrm{r}\right)^2}\boldsymbol{t}_{1,k}\left(\boldsymbol{g}_{1,k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{1,k}^T\boldsymbol{v}_\mathrm{r}\right)\boldsymbol{t}_{1,k}^T\right), \tag{4-65}$$

$$\boldsymbol{\Delta}_{2,k} = \frac{4}{\sqrt{\pi}}\left(e^{-\left(\boldsymbol{g}_{2,k}^T\boldsymbol{f}_{\mathrm{r},i}+\boldsymbol{t}_{2,k}^T\boldsymbol{v}_\mathrm{r}\right)^2}\boldsymbol{t}_{2,k}\left(\boldsymbol{g}_{2,k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{t}_{2,k}^T\boldsymbol{v}_\mathrm{r}\right)\boldsymbol{t}_{2,k}^T\right). \tag{4-66}$$

In (4-60) the constraint functions read as $f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{a}_i'\boldsymbol{v}_\mathrm{r} - b_i$, for $i = 1, \ldots, M\alpha_x$, and $f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{r}_i\boldsymbol{v}_\mathrm{r} - l_i$ for $i = M\alpha_x + 1, \ldots, M\alpha_x + 2K$, with $\boldsymbol{r}_i$ being the i-th row of $\boldsymbol{R}$ and l_i being the i-th element of $\boldsymbol{l}$. With this, the gradient and Hessian of the first $M\alpha_x$ constraints are given in (4-59), and, for $i = M\alpha_x + 1, \ldots, M\alpha_x + 2K$, $\nabla f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{r}_i$ and $\nabla^2 f_i(\boldsymbol{v}_\mathrm{r}) = \boldsymbol{0}$. For the MUBSEP subproblems $\varphi = M\alpha_x + 2K$ and $q = 2(M - p)$ which yields an UBCO of $\mathcal{O}\left(M^3\sqrt{M+K}\right)$.

Pruning Process The pruning process aims to exclude from the search subvectors that cannot be part of the optimal solution. Note that, $f_0(\boldsymbol{v}_{\mathrm{r,ub}}, \boldsymbol{\xi}_\mathrm{r})$ is an upper bound on $f_0(\boldsymbol{x}_{\mathrm{r,opt}})$ for any fixed subvector $\boldsymbol{\xi}_\mathrm{r}$. This implies that if $\boldsymbol{f}_{\mathrm{r},i}$ is a subvector of $\boldsymbol{x}_{\mathrm{r,opt}}$ then the relation $f_0(\boldsymbol{v}_{\mathrm{r,lb}}, \boldsymbol{f}_{\mathrm{r},i}) \leq f_0(\boldsymbol{x}_{\mathrm{r,opt}}) \leq f_0(\boldsymbol{v}_{\mathrm{r,ub}}, \boldsymbol{\xi}_\mathrm{r})$ is guaranteed to hold. Yet, if $f_0(\boldsymbol{v}_{\mathrm{r,lb}}, \boldsymbol{f}_{\mathrm{r},i}) > f_0(\boldsymbol{v}_{\mathrm{r,ub}}, \boldsymbol{\xi}_\mathrm{r})$ then $\boldsymbol{f}_{\mathrm{r},i}$ cannot be a subvector of $\boldsymbol{x}_{\mathrm{r,opt}}$ and it and all its evolutions can be excluded from the search. In this context, having a small upper-bound solution allows for a large number of exclusions which is beneficial for reducing the computational complexity of the algorithm. With this, the pruning process starts by updating the smallest known upper bound $\breve{g}$, for the given layer p, as $\breve{g} = \min\left(\breve{g}, f_0(\boldsymbol{x}_{\mathrm{r,ub},i})\right)$. Then, the lower-bound solutions $\boldsymbol{x}_{\mathrm{lb},i}$ are considered for the pruning criterion. Similar to the one from [27], the pruning criterion considered exploits the low-resolution constraints property that, for a sufficiently small ϵ, the ϵ-suboptimal set only contains the optimal solution. This implies that it is sufficient to find a solution in the ϵ-suboptimal set, which

Algorithm 6 Proposed QoS B&B Tree Search Based Precoding Stage

Inputs: $\boldsymbol{H}$, $\boldsymbol{s}$, σ_w, $\boldsymbol{\lambda}$, $\check{\boldsymbol{x}}$, $\check{g}$, Criterion **Output**: $\boldsymbol{x}_{\text{out}}$

Define the first level ($p = 1$) of the tree by $\mathcal{G}_p := \mathcal{X}$

for $p = 1 : M - 1$ **do**

Partition $\mathcal{G}_p$ in $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{|\mathcal{G}_p|}$

for $i = 1 : |\mathcal{G}_p|$ **do**

if Criterion = MSEP

Solve (4-53) conditioned on $\boldsymbol{f}_{\text{r},i} = R(\boldsymbol{f}_i)$ to get $\boldsymbol{v}_{\text{r,lb}|f_i}$

Construct $\boldsymbol{x}_{\text{lb},i} = \left[\boldsymbol{f}_i^T \ , \ C\left(\boldsymbol{v}_{\text{r,lb}|f_i}\right)^T\right]$ and compute $g_{\text{lb},i} = g(\boldsymbol{x}_{\text{lb},i})$ using (4-44)

Project $\boldsymbol{x}_{\text{ub},i} = P(Q(\boldsymbol{x}_{\text{lb},i}))$ and compute $g_{\text{ub},i} = g(\boldsymbol{x}_{\text{ub},i})$ using (4-44)

Compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{ub},i})$ with (4-48), **if** $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{ub},i}) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_{\text{ub},i}$

else if Criterion = MUBSEP

Solve (4-60) conditioned on $\boldsymbol{f}_{\text{r},i} = R(\boldsymbol{f}_i)$ to get $\boldsymbol{v}_{\text{r,lb}|f_i}$

Construct $\boldsymbol{x}_{\text{lb},i} = \left[\boldsymbol{f}_i^T \ , \ C\left(\boldsymbol{v}_{\text{r,lb}|f_i}\right)^T\right]$ and compute $g_{\text{lb},i} = g(\boldsymbol{x}_{\text{lb},i})$ using (4-46)

Project $\boldsymbol{x}_{\text{ub},i} = P(Q(\boldsymbol{x}_{\text{lb},i}))$ and compute $g_{\text{ub},i} = g(\boldsymbol{x}_{\text{ub},i})$ using (4-46)

Compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{ub},i})$ with (4-49), **if** $\mathbf{P}_{\text{e}}(\boldsymbol{x}_{\text{ub},i}) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_{\text{ub},i}$

end if

end for

Update the best upper bound with $\check{g} = \min(\check{g}, g_{\text{ub},i})$ and update $\check{\boldsymbol{x}}$ accordingly

Construct a reduced set as $\mathcal{G}'_p := \{\boldsymbol{x}_{\text{lb},i} \mid g_{\text{lb},i} < (1-\gamma)\ \check{g}, i = 1, \ldots, |\mathcal{G}_p|\}$

Define the set for the next level in the tree: $\mathcal{G}_{p+1} := \mathcal{G}'_p \times \mathcal{X}$

end for

Partition $\mathcal{G}_M$ in $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\mathcal{G}_M|}$

if Criterion = MSEP

for $i = 1 : |\mathcal{G}_M|$

Compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_i)$ with (4-48), **if** $\mathbf{P}_{\text{e}}(\boldsymbol{x}_i) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_i$

end for

The global solution is $\boldsymbol{x}_{\text{out}} = \underset{\boldsymbol{x} \in \{\mathcal{G}_M \cup \{\check{\boldsymbol{x}}\}\}}{\arg\min}\ g(\boldsymbol{x})$, with $g(\boldsymbol{x})$ computed with (4-44)

else if Criterion = MUBSEP

for $i = 1 : |\mathcal{G}_M|$

Compute $\mathbf{P}_{\text{e}}(\boldsymbol{x}_i)$ with (4-49), **if** $\mathbf{P}_{\text{e}}(\boldsymbol{x}_i) \preceq \boldsymbol{\lambda} \rightarrow$ **terminate with** $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}_i$

end for

The global solution is $\boldsymbol{x}_{\text{out}} = \underset{\boldsymbol{x} \in \{\mathcal{G}_M \cup \{\check{\boldsymbol{x}}\}\}}{\arg\min}\ g(\boldsymbol{x})$, with $g(\boldsymbol{x})$ computed with (4-46)

end if

can be done by adopting the pruning condition $f_0(\boldsymbol{x}_{\mathrm{r,lb},i}) < (1-\delta)\check{g}$, with a sufficiently small value for δ [27]. With this, a set is constructed only with $\boldsymbol{x}_{\mathrm{r,lb},i}$ that attain the pruning condition, and the algorithm goes to the next layer, i.e. updates $p = p+1$. If the algorithm reaches the last layer $p = M$, it is expected that the set of candidates $\mathcal{G}_M$ is relatively small. With this, it is evaluated if any $\boldsymbol{x}_i \in \mathcal{G}_M$ attains $\mathbf{P}_{\mathrm{e}}(\boldsymbol{x}_i) \preceq \boldsymbol{\lambda}$, if true the algorithm terminates with $\boldsymbol{x}_i$. Otherwise, the algorithm terminates with $\boldsymbol{x}_{\mathrm{out}} = C\left(\underset{\boldsymbol{x}\in\mathcal{G}_M\cup\{\check{\boldsymbol{x}}\}}{\operatorname{argmin}} g(\boldsymbol{x})\right)$ computed via brute force. The steps of the QoS B&B TSBP stage are further described in Algorithm 6.

Finally, the QoS B&B algorithm is summarized as the execution of Algorithm 5 representing the QoS B&B PGS stage, and if the process is not terminated with the transmit vector the execution of Algorithm 6 representing the QoS B&B TSBP stage.

4.5 Discrete Precoding with the RMMSE Criterion for Imperfect CSI Scenarios

The methods proposed in the previous section consider perfect knowledge about the CSI for their design. Although possible, the extension of the SEP-related methods for imperfect CSI scenarios is not straightforward. Robustness is a desired quality in transmit and receive processing schemes [56, 57, 58]. With this, this section presents the proposed precoding methods utilizing RMMSE as the design objective. The methods proposed are extensions of the approaches from [59] for imperfect CSI scenarios where the transmitter has knowledge about the second-order statistics of the CSI mismatch. Note that, when perfect CSI is available the proposed RMMSE technique becomes equivalent to the MMSE approaches described in section 4.2.2. The RMMSE problem considering quantization of the transmit signal can be cast

$$\begin{aligned} &\min_{\boldsymbol{x},f} \mathrm{E}\{\|f\boldsymbol{z}-\boldsymbol{s}\|_2^2\} \\ &\text{subject to: } \boldsymbol{x}\in\mathcal{X}^M, \quad f\geq 0, \end{aligned} \tag{4-67}$$

which is similar to the MMSE problem under low-resolution constraints from [59]. Yet, similarly as in chapter 3, considers that the CSI available at the transmitter side is imperfect and that the transmitter has knowledge about the second-order statistics of the CSI mismatch. With this, the channel is modeled as $\boldsymbol{H} = \sqrt{\lambda}\boldsymbol{H}_{\mathrm{est}} + \sqrt{1-\lambda}\boldsymbol{\Gamma}$, where $\lambda \in (0,1]$ is a positive real-valued scalar which is considered as known at the transmitter. The channel matrix can be written in real-valued form as $\boldsymbol{H}_{\mathrm{r}} = \sqrt{\lambda}\boldsymbol{H}_{\mathrm{r,est}} + \sqrt{1-\lambda}\boldsymbol{\Gamma}_{\mathrm{r}}$ where

$\boldsymbol{H}_{\mathrm{r,est}} = R(\boldsymbol{H}_{\mathrm{est}})$ and $\boldsymbol{\Gamma}_{\mathrm{r}} = R(\boldsymbol{\Gamma})$. The optimization problem from (4-67) can be written taking into consideration the CSI mismatch, which reads as

$$\begin{aligned} &\min_{\boldsymbol{x}_{\mathrm{r}},f} \mathrm{E}\{\|f\left(\sqrt{\lambda}\boldsymbol{H}_{\mathrm{r,est}} + \sqrt{1-\lambda}\boldsymbol{\Gamma}_{\mathrm{r}}\right)\boldsymbol{x}_{\mathrm{r}} + f\boldsymbol{w}_{\mathrm{r}} - \boldsymbol{s}_{\mathrm{r}}\|_2^2\} \\ &\text{subject to: } \left\{[\boldsymbol{x}_{\mathrm{r}}]_{2m-1} + j\,[\boldsymbol{x}_{\mathrm{r}}]_{2m}\right\} \in \mathcal{X}, \quad \text{for } m \in \{1,\ldots,M\}, \\ &\qquad f \geq 0. \end{aligned} \tag{4-68}$$

Considering $\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^T\boldsymbol{H}_{\mathrm{r,est}}\right\} = \boldsymbol{0}$, an equivalent problem can be cast as

$$\begin{aligned} &\min_{\boldsymbol{x}_{\mathrm{r}},f} f^2\boldsymbol{x}_{\mathrm{r}}^T\left(\lambda\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}\right)\boldsymbol{x}_{\mathrm{r}} - \\ &\qquad 2f\sqrt{\lambda}\,\boldsymbol{x}_{\mathrm{r}}^T\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{s}_{\mathrm{r}} + f^2\mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T\boldsymbol{w}_{\mathrm{r}}\} \\ &\text{subject to: } \left\{[\boldsymbol{x}_{\mathrm{r}}]_{2m-1} + j\,[\boldsymbol{x}_{\mathrm{r}}]_{2m}\right\} \in \mathcal{X}, \quad \text{for } m \in \{1,\ldots,M\}, \\ &\qquad f \geq 0, \end{aligned} \tag{4-69}$$

where $\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}} = \mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^T\boldsymbol{\Gamma}_{\mathrm{r}}\right\}$ is considered as known by the transmitter. The objective from (4-69) relates to the MSE as follows

$$\begin{aligned} \tilde{\mathrm{MSE}}\left(\boldsymbol{x}_{\mathrm{r}}, f\right) &= \mathrm{MSE}\left(\boldsymbol{x}_{\mathrm{r}}, f\right) - \mathrm{E}\left\{\boldsymbol{s}_{\mathrm{r}}^T\boldsymbol{s}_{\mathrm{r}}\right\} \\ &= f^2\boldsymbol{x}_{\mathrm{r}}^T\left(\lambda\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}\right)\boldsymbol{x}_{\mathrm{r}} - \\ &\quad 2f\sqrt{\lambda}\,\boldsymbol{x}_{\mathrm{r}}^T\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{s}_{\mathrm{r}} + f^2\mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T\boldsymbol{w}_{\mathrm{r}}\}. \end{aligned} \tag{4-70}$$

The objective described in (4-70) is not jointly convex in $\boldsymbol{x}_{\mathrm{r}}$ and f. Moreover, the feasible set of (4-69) is discrete and, thus, not convex.

4.5.1 Proposed RMMSE Mapped Precoder

In this subsection, we propose a practical approach for the problem described in (4-67). Since the feasible set of the optimization problem presented by (4-67) is non-convex we replace $\mathcal{X}^M$ by its convex hull $\mathcal{P}$, which is a polyhedron. With this, the problem reads as

$$\begin{aligned} &\min_{\boldsymbol{x},f\geq 0} \mathrm{E}\{\|f\boldsymbol{z} - \boldsymbol{s}\|_2^2\} \\ &\text{s.t } \mathrm{Re}\left\{x_m e^{j\frac{2\pi i}{\alpha_x}}\right\} \leq \sqrt{\mathrm{P_A}}\cos\left(\frac{\pi}{\alpha_x}\right),\ m \in \{1,\ldots,M\},\ i \in \{1,\ldots,\alpha_x\}. \end{aligned} \tag{4-71}$$

The equivalent problem in real-valued notation reads as

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r},f} \mathrm{E}\{\|f\boldsymbol{z}_\mathrm{r} - \boldsymbol{s}_\mathrm{r}\|_2^2\} \\ &\text{subject to: } \boldsymbol{A}\boldsymbol{x}_\mathrm{r} \leq \boldsymbol{b}, \quad f \geq 0, \end{aligned} \tag{4-72}$$

where $\boldsymbol{z}_\mathrm{r} = R(\boldsymbol{z})$. The inequality $\boldsymbol{A}\boldsymbol{x}_\mathrm{r} \leq \boldsymbol{b}$ restricts the elements of the precoding vector to be inside or on the border of the polyhedron. The polyhedron associated to uniformly phase quantized transmit symbols with α_x different phases can be expressed as proposed before in [18], which is similar to the description in [14]. The corresponding matrix notation reads as

$$\begin{aligned} &\boldsymbol{A} = \begin{bmatrix}(\boldsymbol{I}_M \otimes \boldsymbol{\beta}_1)^T & (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_2)^T & \dots & (\boldsymbol{I}_M \otimes \boldsymbol{\beta}_{\alpha_x})^T\end{bmatrix}^T, \\ &\boldsymbol{\beta}_i = \begin{bmatrix}\cos\left(\frac{2\pi i}{\alpha_x}\right) & -\sin\left(\frac{2\pi i}{\alpha_x}\right)\end{bmatrix}, \quad i \in \{1, \dots, \alpha_x\}, \quad \boldsymbol{b} = \sqrt{\mathrm{P_A}}\cos\left(\frac{\pi}{\alpha_x}\right)\boldsymbol{1}_{M\alpha_x}, \end{aligned} \tag{4-73}$$

with $\boldsymbol{1}_{M\alpha_x}$ being a column vector with length $M\alpha_x$. Expanding the objective from (4-72) leads to the following problem

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r},f} f^2\boldsymbol{x}_\mathrm{r}^T\left(\lambda\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}\right)\boldsymbol{x}_\mathrm{r} - 2f\sqrt{\lambda}\,\boldsymbol{x}_\mathrm{r}^T\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{s}_\mathrm{r} + f^2\mathrm{E}\{\boldsymbol{w}_\mathrm{r}^T\boldsymbol{w}_\mathrm{r}\} \\ &\text{subject to: } \boldsymbol{A}\boldsymbol{x}_\mathrm{r} \leq \boldsymbol{b}, \quad f \geq 0. \end{aligned} \tag{4-74}$$

An equivalent convex problem can be cast by substituting $\boldsymbol{x}_\mathrm{r,f} = f\boldsymbol{x}_\mathrm{r}$. With this, the problem from (4-74) is rewritten as

$$\begin{aligned} &\min_{\boldsymbol{x}_\mathrm{r,f},f} \boldsymbol{x}_\mathrm{r,f}^T\left(\lambda\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}\right)\boldsymbol{x}_\mathrm{r,f} - 2\sqrt{\lambda}\,\boldsymbol{x}_\mathrm{r,f}^T\boldsymbol{H}_{\mathrm{r,est}}^T\boldsymbol{s}_\mathrm{r} + f^2\mathrm{E}\{\boldsymbol{w}_\mathrm{r}^T\boldsymbol{w}_\mathrm{r}\} \\ &\text{subject to: } \boldsymbol{R}\begin{bmatrix}\boldsymbol{x}_\mathrm{r,f} \\ f\end{bmatrix} \leq \boldsymbol{0}, \qquad f \geq 0, \end{aligned}$$

where $\boldsymbol{R} = \begin{bmatrix}\boldsymbol{A}, & -\boldsymbol{b}\end{bmatrix}$. Finally, the problem can be written in the standard QP form as

$$\begin{aligned} &\min_{\boldsymbol{v}} \frac{1}{2}\boldsymbol{v}^T\tilde{\boldsymbol{U}}\boldsymbol{v} + \tilde{\boldsymbol{p}}^T\boldsymbol{v} \\ &\text{subject to: } \boldsymbol{R}_\mathrm{ext}\boldsymbol{v} \leq \boldsymbol{0}, \end{aligned} \tag{4-75}$$

where

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{x}_{\mathrm{r,f}} \\ f \end{bmatrix}, \quad \boldsymbol{R}_{\mathrm{ext}} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{\xi}^T \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix}, \quad \tilde{\boldsymbol{p}} = \begin{bmatrix} -2\sqrt{\lambda}\boldsymbol{H}_{\mathrm{r}}^T \boldsymbol{s}_{\mathrm{r}} \\ 0 \end{bmatrix} \tag{4-76}$$

$$\tilde{\boldsymbol{U}} = 2 \begin{bmatrix} \lambda \boldsymbol{H}_{\mathrm{r,est}}^T \boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}} & \mathbf{0} \\ \mathbf{0}^T & \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{w}_{\mathrm{r}}\} \end{bmatrix}. \tag{4-77}$$

Note that, since $\tilde{\boldsymbol{U}} \in S_+^{2M+1}$, the problem is a convex QP and can be solved by utilizing standard optimization tools.

By solving the relaxed problem (4-75) one obtains the solution $\boldsymbol{v} = \left[\boldsymbol{x}_{\mathrm{r,f}}^T,\ f\right]^T$ and readily extracts $\boldsymbol{x}_{\mathrm{r,lb}} = \frac{\boldsymbol{x}_{\mathrm{r,f}}}{f}$ and $\boldsymbol{x}_{\mathrm{lb}} = C(\boldsymbol{x}_{\mathrm{r,lb}})$. Yet, since the discrete feasible set is relaxed to its convex hull, the optimal solution of the relaxed problem is not necessarily in the feasible set of (4-67). Therefore, to find a feasible solution, mapping $\boldsymbol{x}_{\mathrm{lb}}$ to the closest Euclidean distance point in $\mathcal{X}^M$ is considered, which then yields $\hat{\boldsymbol{x}}_{\mathrm{ub}}$. Since the mapping step does not preserve the value of f, it must be recomputed as follows

$$f_{\mathrm{ub}} = \frac{\sqrt{\lambda}\ \boldsymbol{s}_{\mathrm{r}}^T \boldsymbol{H}_{\mathrm{r,est}} \hat{\boldsymbol{x}}_{\mathrm{r,ub}}}{\hat{\boldsymbol{x}}_{\mathrm{r,ub}}^T \left(\lambda \boldsymbol{H}_{\mathrm{r,est}}^T \boldsymbol{H}_{\mathrm{r,est}} + (1-\lambda)\,\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}\right) \hat{\boldsymbol{x}}_{\mathrm{r,ub}} + \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T \boldsymbol{w}_{\mathrm{r}}\}}, \tag{4-78}$$

where $\hat{\boldsymbol{x}}_{\mathrm{r,ub}} = R(\hat{\boldsymbol{x}}_{\mathrm{ub}})$. Note that, the scaling factor f_{ub} associated to the mapped solution can be negative which then corresponds to an unfeasible solution for (4-67). However, in this scenario, a feasible solution with equivalent MSE can be computed by flipping the sign of $\hat{\boldsymbol{x}}_{\mathrm{ub}}$, which then leads to the transmit vector being computed as $\boldsymbol{x}_{\mathrm{ub}} = \mathrm{sign}(f_{\mathrm{ub}})\hat{\boldsymbol{x}}_{\mathrm{ub}}$. Flipping the sign of $\hat{\boldsymbol{x}}_{\mathrm{ub}}$ leads to a feasible solution due to the symmetry of $\mathcal{X}$.

Note that, $\tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{lb}})$ is a lower bound on the optimal value of the original problem, meaning that $\tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{lb}}) \leq \tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{opt}})$, where $\boldsymbol{x}_{\mathrm{opt}}$ is the optimal solution of (4-67). Yet, as mentioned before, $\boldsymbol{x}_{\mathrm{lb}}$ is not necessarily in $\mathcal{X}^M$ and thus, is not necessarily a feasible solution of (4-67). After mapping, $\boldsymbol{x}_{\mathrm{ub}}$ is obtained. Since the mapping step does not preserve optimality, $\boldsymbol{x}_{\mathrm{ub}}$ is a feasible upper bound solution to (4-67). In terms of the objective function this means that $\tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{lb}}) \leq \tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{opt}}) \leq \tilde{\mathrm{MSE}}\,(\boldsymbol{x}_{\mathrm{ub}})$. In this sense, the proposed mapped precoder provides a practical suboptimal solution to (4-67).

4.5.2 Proposed Optimal Approach via Branch-and-Bound

As previously discussed, solving the relaxed problem provides a lower bound solution to (4-67), which, in general, does not attain the low-resolution constraints. A feasible solution can be practically computed by utilizing the

proposed mapped approach, yet, its MSE is an upper bound on MSE $(\boldsymbol{x}_{\text{opt}})$. In this subsection, we propose a branch-and-bound strategy that reliably computes the optimal solution for (4-67).

4.5.2.1 Introduction of the Branch-and-Bound Method

A branch-and-bound algorithm is a tree search based method. The tree represents the set of all possible solutions for the vector $\boldsymbol{x}$, i.e., it is a representation of the set $\mathcal{X}^M$. The construction of the tree is done similarly as in section 4.4.3.3 with an example of the tree shown in Fig. 4.2.

For constructing the precoding vector we consider the minimization of an objective function $g(\boldsymbol{x}, \boldsymbol{s})$ subject to the discrete feasible set, described by

$$\boldsymbol{x}_{\text{opt}} = \arg\min_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{s}) \quad \text{s.t. } \boldsymbol{x} \in \mathcal{X}^M. \tag{4-79}$$

In the context of solving (4-67), the objective function $g(\boldsymbol{x}, \boldsymbol{s})$ represents the MSE with imperfect CSI. A lower bound on $g(\boldsymbol{x}_{\text{opt}}, \boldsymbol{s})$ can be obtained by relaxing $\mathcal{X}^M$ to its convex hull. The relaxed problem is expressed as

$$\boldsymbol{x}_{\text{lb}} = \arg\min_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{s}) \quad \text{s.t. } \boldsymbol{x} \in \mathcal{P}. \tag{4-80}$$

An associated upper bound on $g(\boldsymbol{x}_{\text{opt}}, \boldsymbol{s})$ can be obtained by mapping the solution of (4-80) to the feasible set and evaluating $g(\cdot)$, as discussed previously in subsection 4.5.1. The upper bound value of (4-79) is termed $\check{g}$. Having an upper bound solution implies that $\check{g} \geq g(\boldsymbol{x}_{\text{opt}}) \geq g(\boldsymbol{x}_{\text{lb}})$, which means that the objective of the mapped vector is always greater or equal to the objective of the solution from the relaxed problem (4-80).

By fixing p entries of $\boldsymbol{x}$, the vector can be rewritten as $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$, with $\boldsymbol{x}_1 \in \mathcal{X}^p$. With this, a subproblem can be formulated as

$$\boldsymbol{x}_{2,\text{opt}|\boldsymbol{x}_1} = \arg\min_{\boldsymbol{x}_2} g(\boldsymbol{x}_2, \boldsymbol{x}_1, \boldsymbol{s}) \quad \text{s.t. } \boldsymbol{x}_2 \in \mathcal{X}^{M-p}. \tag{4-81}$$

Relaxing the problem from (4-81) we have

$$\boldsymbol{x}_{2,\text{lb}} = \arg\min_{\boldsymbol{x}_2} g(\boldsymbol{x}_2, \boldsymbol{x}_1, \boldsymbol{s}) \quad \text{s.t. } \boldsymbol{x}_2 \in \mathcal{J}, \tag{4-82}$$

where $\mathcal{J}$ is the convex hull of $\mathcal{X}^{M-p}$. If the optimal value of (4-82) is larger than a known upper bound $\check{g}$ on the solution of (4-79), then all members in the discrete set which include the previously fixed vector $\boldsymbol{x}_1$ can be excluded from the search process. By this strategy we intend to exclude most of the

candidates from the possible solution set such that the number of residual candidates is only a small fraction of its total number and, thus, they can be evaluated via exhaustive search.

4.5.2.2 Branch-and-Bound Initialization

The branch-and-bound algorithm converges faster when one computes, as early as possible, an upper bound that permits many exclusions. Therefore, it is recommended to have an initialization step where an upper bound $\check{g} < \infty$ is found before beginning with the search process. In this regard, for initialization, one solves the relaxed RMMSE problem (4-75). With this, $\boldsymbol{x}_{\mathrm{lb}}$ and $g(\boldsymbol{x}_{\mathrm{lb}}) = \tilde{\mathrm{MSE}}(\boldsymbol{x}_{\mathrm{lb}})$ are obtained. After mapping $\boldsymbol{x}_{\mathrm{lb}}$ to $\mathcal{X}^M$, $\hat{\boldsymbol{x}}_{\mathrm{ub}}$ and $\check{g} = \tilde{\mathrm{MSE}}(\hat{\boldsymbol{x}}_{\mathrm{ub}})$ are determined. Note that, if the solution of the relaxed problem is already in the feasible set (meaning if $\boldsymbol{x}_{\mathrm{lb}} \in \mathcal{X}^M$), upper and lower bound are equal which corresponds to

$$\boldsymbol{x}_{\mathrm{ub}} = \boldsymbol{x}_{\mathrm{lb}} = \boldsymbol{x}_{\mathrm{opt}} \rightarrow g(\boldsymbol{x}_{\mathrm{lb}}) = \check{g}. \tag{4-83}$$

This would imply that the optimal solution is found already by the approach from subsection 4.5.1 and the tree search process can be skipped.

4.5.2.3 Subproblems

When the condition from (4-83) is not met, the branch-and-bound tree search method is applied, which involves solving subproblems, as first mentioned on 4.5.2.1. To this end, the real-valued precoding vector $\boldsymbol{x}_{\mathrm{r}}$ is divided in a fixed vector of length $2p$ and a variable vector according to $\boldsymbol{x}_{\mathrm{r}} = \left[\boldsymbol{x}_{\mathrm{r,fixed}}^T \, , \, \boldsymbol{x}_{\mathrm{r}}'^T\right]^T$. The corresponding real-valued channel matrix can be written as

$$\boldsymbol{H}_{\mathrm{r}} = \left[\sqrt{\lambda}\boldsymbol{H}_{\mathrm{r,est,fixed}} + \sqrt{1-\lambda}\boldsymbol{\Gamma}_{\mathrm{r,fixed}} \, , \, \sqrt{\lambda}\boldsymbol{H}'_{\mathrm{r,est}} + \sqrt{1-\lambda}\boldsymbol{\Gamma}'_{\mathrm{r}}\right], \tag{4-84}$$

which then leads to the following robust equivalent MSE formulation

$$\begin{aligned}
&\tilde{\mathrm{MSE}} = f'^2(\boldsymbol{x}_{\mathrm{r,fixed}}^T(\lambda\boldsymbol{H}_{\mathrm{r,est,fixed}}^T\boldsymbol{H}_{\mathrm{r,est,fixed}} + \\
&(1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^T\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\})\boldsymbol{x}_{\mathrm{r,fixed}} + \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^T\boldsymbol{w}_{\mathrm{r}}\}) + \\
&f'\boldsymbol{x}_{\mathrm{r,fixed}}^T(2\lambda\boldsymbol{H}_{\mathrm{r,est,fixed}}^T\boldsymbol{H}'_{\mathrm{r,est}} + (1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^T\boldsymbol{\Gamma}'_{\mathrm{r}}\right\} + \\
&(1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}'^T\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\})\boldsymbol{x}'_{\mathrm{r,f}} + \boldsymbol{x}_{\mathrm{r,f}}'^T(\lambda\boldsymbol{H}_{\mathrm{r,est}}'^T\boldsymbol{H}'_{\mathrm{r,est}} + \\
&(1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}'^T\boldsymbol{\Gamma}'_{\mathrm{r}}\right\})\boldsymbol{x}'_{\mathrm{r,f}} - 2\sqrt{\lambda}(f'\boldsymbol{x}_{\mathrm{r,fixed}}^T\boldsymbol{H}_{\mathrm{r,est,fixed}}^T + \boldsymbol{x}_{\mathrm{r,f}}'^T\boldsymbol{H}_{\mathrm{r,est}}'^T)\boldsymbol{s}_{\mathrm{r}},
\end{aligned} \tag{4-85}$$

where the values of $\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime}\right\} \in \mathbb{R}^{2p\times 2p}$, $\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^{T}\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime}\right\} \in \mathbb{R}^{2p\times 2(M-p)}$, $\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\} \in \mathbb{R}^{2(M-p)\times 2p}$ and $\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^{T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\} \in \mathbb{R}^{2(M-p)\times 2(M-p)}$ can be taken from $\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}}$ with the following structure

$$\boldsymbol{R}_{\mathrm{r},\boldsymbol{\Gamma}} = \begin{bmatrix} \mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^{T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\} & \mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^{T}\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime}\right\} \\ \mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\} & \mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime}\right\} \end{bmatrix}. \tag{4-86}$$

With this, (4-85) can be rearranged with a stacked vector notation as

$$\tilde{\mathrm{MSE}} = \frac{1}{2}\boldsymbol{v}^{\prime T}\tilde{\boldsymbol{Q}}\boldsymbol{v}^{\prime} + \boldsymbol{l}^{T}\boldsymbol{v}^{\prime}, \tag{4-87}$$

where $\boldsymbol{v}^{\prime} = \left[\boldsymbol{x}_{\mathrm{r,f}}^{\prime T},\ f^{\prime}\right]^{T}$ and

$$\boldsymbol{l} = -2\begin{bmatrix} \boldsymbol{H}_{\mathrm{r}}^{\prime T} \\ \boldsymbol{x}_{\mathrm{r,\ fixed}}^{T}\boldsymbol{H}_{\mathrm{r,\ fixed}}^{T} \end{bmatrix}\boldsymbol{s}_{\mathrm{r}}, \quad \tilde{\boldsymbol{Q}} = 2\begin{bmatrix} \tilde{\boldsymbol{Q}}_{1} & \tilde{\boldsymbol{q}}_{2} \\ \tilde{\boldsymbol{q}}_{2}^{T} & \tilde{q}_{3} \end{bmatrix}, \tag{4-88}$$

with

$$\begin{aligned} \tilde{\boldsymbol{Q}}_{1} =& \lambda\boldsymbol{H}_{\mathrm{r,est}}^{\prime T}\boldsymbol{H}_{\mathrm{r,est}}^{\prime} + (1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime}\right\}, \\ \tilde{\boldsymbol{q}}_{2} =& \lambda\boldsymbol{H}_{\mathrm{r,est}}^{\prime T}\boldsymbol{H}_{\mathrm{r,est,fixed}}\boldsymbol{x}_{\mathrm{r,fixed}} + (1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r}}^{\prime T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\}\boldsymbol{x}_{\mathrm{r,fixed}}, \\ \tilde{q}_{3} =& \boldsymbol{x}_{\mathrm{r,fixed}}^{T}(\lambda\boldsymbol{H}_{\mathrm{r,est,fixed}}^{T}\boldsymbol{H}_{\mathrm{r,est,fixed}} + \\ & (1-\lambda)\mathrm{E}\left\{\boldsymbol{\Gamma}_{\mathrm{r,fixed}}^{T}\boldsymbol{\Gamma}_{\mathrm{r,fixed}}\right\})\boldsymbol{x}_{\mathrm{r,fixed}} + \mathrm{E}\{\boldsymbol{w}_{\mathrm{r}}^{T}\boldsymbol{w}_{\mathrm{r}}\}. \end{aligned} \tag{4-89}$$

Finally, with (4-87), the optimization problem that describes the RMMSE subproblems reads as

$$\min_{\boldsymbol{v}^{\prime}} \frac{1}{2}\boldsymbol{v}^{\prime T}\tilde{\boldsymbol{Q}}\boldsymbol{v}^{\prime} + \boldsymbol{l}^{T}\boldsymbol{v}^{\prime} \ \text{ s.t. } \boldsymbol{R}_{\mathrm{ext}}^{\prime}\boldsymbol{v}^{\prime} \leq \boldsymbol{0}. \tag{4-90}$$

where $\boldsymbol{R}_{\mathrm{ext}}^{\prime}$ is obtained by selecting the last $2\,(M-p)$ columns of $\boldsymbol{R}_{\mathrm{ext}}$.Since $\tilde{\boldsymbol{Q}} \in S_{+}^{2(M-p)+1}$ the problem described in (4-90) is a convex QP.

4.5.2.4 Pruning Step

By solving the subproblems one can compute a lower bound, and, after the mapping step, an upper bound on the optimal solution of (4-67). As mentioned before in subsection 4.5.2.1, these bounds are utilized to reduce the searching set such that the optimal solution can be found via exhaustive search. This subsection details the process of excluding candidate solutions from the searching set, which, in the context of tree search is named pruning the tree.

To formulate an efficient pruning criterion, in this subsection, we exploit a property of the low resolution constraints. As mentioned before, having low resolution data converters implies that the entries of the precoding vector to belong to a discrete set. Due to the discrete nature of the feasible set, by choosing a sufficiently small ϵ, for example $\epsilon = \delta\ \tilde{\mathrm{MSE}}_{\mathrm{opt}}$ with $0 \leq \delta \ll 1$, the ϵ-suboptimal set given by $\mathcal{X}_{\mathrm{opt},\epsilon} = \left\{\boldsymbol{x} :\ \tilde{\mathrm{MSE}}(\boldsymbol{x}) \leq \tilde{\mathrm{MSE}}_{\mathrm{opt}} + \epsilon\right\}$, cf. [42], contains only the global optimal solution. This implies for the tree search process that it is sufficient to find a solution in $\mathcal{X}_{\mathrm{opt},\epsilon}$. Implicitly the ϵ-suboptimal set can be addressed by the pruning condition $\tilde{\mathrm{MSE}}(\boldsymbol{x}_{\mathrm{lb}}) < (1-\delta)\breve{g}$, where $\breve{g}$ is the best known upper bound. By setting a sufficiently small value for δ, the optimal solution from (4-67) can be obtained with probability one, as is confirmed in [27].

With this, the major steps for constructing the proposed branch-and-bound algorithm are described and the branch-and-bound algorithm for solving (4-67) can be assembled. The steps of the method are detailed in Algorithm 7.

4.6 Numerical Results

For the numerical evaluation, the channel coefficients are modeled by independent Rayleigh fading [60], and for the B&B algorithms $\gamma = \delta = 10^{-7}$. As derived in section C.1 of the appendix, the SNR is defined as $\mathrm{SNR} = \frac{(M \cdot \mathrm{P_A})}{\sigma_w^2}$. The proposed methods are compared against the following state-of-the-art approaches:

1- The MMDDT B&B precoder [18];

2- The MMSE B&B precoder [48];

3- The CI 1-bit Partial B&B precoder [25];

4- The MMSE Mapped precoder [48];

5- The MSM precoder [14];

6- The quantized CVX-CIO precoder [12];

7- The unquantized ZF-P precoder [61];

8- The unquantized LMMSE precoder [28]

In sections 4.6.1 and 4.6.2 perfect CSI is considered for numerical evaluations. In this scenario, the proposed RMMSE approaches are equivalent to their MMSE counterparts and thus the proposed RMMSE techniques are not explicitly mentioned. In section 4.6.3, imperfect CSI is considered and the RMMSE methods, which in this scenario are not equivalent to MMSE, are compared with the mentioned state-of-the-art algorithms.

Algorithm 7 Proposed RMMSE B&B Precoding Algorithm

Solve (4-75) and get $\boldsymbol{x}_{\text{lb}} = C\left(\frac{\boldsymbol{x}_{\text{r,f}}}{f}\right)$
Map $\boldsymbol{x}_{\text{lb}}$ to $\mathcal{X}^M$ to get $\hat{\boldsymbol{x}}_{\text{ub}}$ and compute $f_{\text{ub}}(\hat{\boldsymbol{x}}_{\text{r,ub}})$ using (4-78), where $\hat{\boldsymbol{x}}_{\text{r,ub}} = R\left(\hat{\boldsymbol{x}}_{\text{ub}}\right)$
Compute $\boldsymbol{x}_{\text{ub}} = \text{sign}(f_{\text{ub}})\hat{\boldsymbol{x}}_{\text{ub}}$ and evaluate the optimality condition described in (4-83). If it holds return $\boldsymbol{x}_{\text{opt}} = \boldsymbol{x}_{\text{ub}}$
Otherwise, define $\breve{\boldsymbol{x}} = \boldsymbol{x}_{\text{ub}}$ and $\breve{g} = \frac{1}{2}\boldsymbol{v}^T\boldsymbol{U}\boldsymbol{v} + \boldsymbol{p}^T\boldsymbol{v}$, where $\boldsymbol{U}$ and $\boldsymbol{p}$ are given by (4-77) and $\boldsymbol{v} = \left[f_{\text{ub}}\boldsymbol{x}_{\text{r,ub}}^T, f_{\text{ub}}\right]^T$
for $d = 1 : M-1$ **do**
  Partition $\mathcal{G}_d$ in $\boldsymbol{x}_{\text{fixed},1}, \ldots, \boldsymbol{x}_{\text{fixed},|\mathcal{G}_d|}$
  for $i = 1 : |\mathcal{G}_d|$ **do**
    Based on $\boldsymbol{x}_{\text{r,fixed},i} = R(\boldsymbol{x}_{,\text{fixed},i})$ solve (4-90)
    to get $\boldsymbol{x}'_{\text{r,f}}$ and f'
    Compute $\tilde{\text{MSE}}_{\text{lb}}\left(\boldsymbol{x}'_{\text{r,f}}, f'\right)$ using (4-87)
    Extract $\boldsymbol{x}'_{\text{lb},i} = C\left(\frac{\boldsymbol{x}'_{\text{r,f}}}{f'}\right)$, map $\boldsymbol{x}'_{\text{lb},i}$ to get $\boldsymbol{x}'_{\text{ub}} \in \mathcal{X}^{M-d}$ and
    construct $\hat{\boldsymbol{x}}_{\text{r,ub}} = \left[\boldsymbol{x}_{\text{fixed},i}^T,\ R(\boldsymbol{x}'_{\text{ub}})^T\right]^T$
    Compute $f_{\text{ub}}(\hat{\boldsymbol{x}}_{\text{r,ub}})$ using (4-78) and $\tilde{\text{MSE}}_{\text{ub}}(\hat{\boldsymbol{x}}_{\text{r,ub}}, f_{\text{ub}})$
    using (4-87)
    Update the best upper bound with $\breve{g} = \min\left(\breve{g}, \tilde{\text{MSE}}_{\text{ub}}\right)$ and
    update $\breve{\boldsymbol{x}}$ accordingly
  end for
  Construct a reduced set by comparing conditioned lower bounds with the global upper bound $\breve{g}$
  $\mathcal{G}'_d := \left\{C(\boldsymbol{x}'_{\text{lb},i}) | \tilde{\text{MSE}}_{\text{lb}}(\boldsymbol{x}'_{\text{lb},i}) < (1-\delta)\breve{g}, i = 1, \ldots, |\mathcal{G}_d|\right\}$
  Define the set for the next level in the tree: $\mathcal{G}_{d+1} := \mathcal{G}'_d \times \mathcal{X}$
end for
Search method for the ultimate level $d = M$:
Define $\tilde{\mathcal{G}}_M = \mathcal{G}_M \cup \{\breve{\boldsymbol{x}}\}$ and partition $\tilde{\mathcal{G}}_M$ in $\boldsymbol{x}_{\text{fixed},1}, \ldots, \boldsymbol{x}_{\text{fixed},|\tilde{\mathcal{G}}_M|}$
Compute $f_{\text{ub},i}(R(\boldsymbol{x}_{\text{fixed},i}))$ using (4-78) for all $i \in \left\{1, \ldots, \left|\tilde{\mathcal{G}}_M\right|\right\}$
Based on all $f_{\text{ub},i}$ and $\boldsymbol{x}_{\text{r,fixed,i}}$, determine
$[\boldsymbol{x}_{\text{r,opt}}, f_{\text{ub,opt}}] = \underset{i \in \left\{1,\ldots,|\tilde{\mathcal{G}}_M|\right\}}{\text{argmin}} \tilde{\text{MSE}}(\boldsymbol{x}_{\text{r,fixed,i}}, f_{\text{ub},i})$ using (4-87)
The global solution is $\boldsymbol{x}_{\text{opt}} = \text{sign}(f_{\text{ub,opt}})C(\boldsymbol{x}_{\text{r,opt}})$

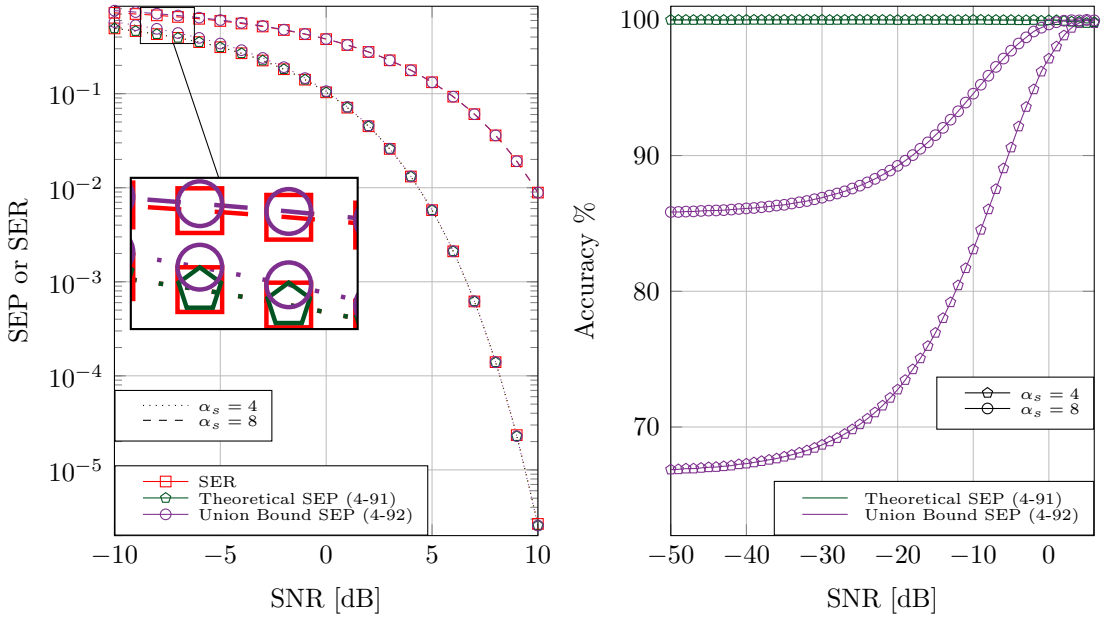

Figure 4.3: SEP or SER versus SNR (left) and Accuracy versus SNR (right), for $K = 30$ users and $M = 100$ BS antennas

4.6.1 Bound Evaluation

This subsection compares the theoretical methods for computing the SEP and union-bound SEP, discussed in sections 4.4.1 and 4.4.2, with the numerically computed SEP. To this end, Monte Carlo simulations are considered, and theoretical SEP (compatible with $\alpha_s \in \{2, 4\}$), union-bound SEP, and numerically computed SER are calculated. The theoretical SEP is given by

$$\mathrm{SEP}_{\mathrm{the}} = (1/K)\, \mathbf{1}_K^T \left(\mathbf{1}_K - \Phi\left(\boldsymbol{S}_{\mathrm{R}}\, \mathrm{Re}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right) \circ \Phi\left(\boldsymbol{S}_{\mathrm{I}}\, \mathrm{Im}\left\{\boldsymbol{H}\boldsymbol{x}\right\}\right)\right), \tag{4-91}$$

and union-bound SEP is computed as

$$\mathrm{SEP}_{\mathrm{ub}} = (1/2K)\, \mathbf{1}_K^T \left(\mathrm{erfc}\left(\boldsymbol{\rho}_1\right) + \mathrm{erfc}\left(\boldsymbol{\rho}_2\right)\right), \tag{4-92}$$

where $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are defined in (4-47). In this study, the accuracy of the SEP prediction methods is defined using the SER as a baseline which reads as $\mathrm{Accuracy}\% = \left(1 - \frac{|\mathrm{SER} - \mathrm{SEP}|}{\mathrm{SER}}\right) \times 100$. For the Monte Carlo simulations, the considered scenarios consist of a BS with $M = 100$ antennas serving $K = 30$ single antenna users with $\alpha_s = 4$ and $\alpha_s = 8$. For this simulation, a total of $3.2 \cdot 10^7$ transmit symbols were considered. Expressions (4-91) and (4-92) can be applied to any precoding approach, and, in this subsection, the users' symbols considered to be precoded with the full-resolution LMMSE method [28].

For the case of $\alpha_s = 4$, SER and theoretical SEP (4-91) are expected to be equivalent (100% accuracy). When using (4-92) it is expected, for the high-SNR limit, an accuracy of 100% and for the low-SNR limit, $\text{SEP}_{\text{ub}} = \left({}^{\alpha_s}/_{(\alpha_s-1)}\right)\text{SER}$ which results in an accuracy of $\left({}^{(\alpha_s-2)}/_{(\alpha_s-1)}\right) \times 100$, which corresponds to 66.6% and 85.7% accuracy, for $\alpha_s = 4$ and $\alpha_s = 8$, respectively. The LHS of Fig. 4.3 shows that the union-bound SEP yields a tight upper bound on the actual error probability and confirms the perfect prediction of the theoretical SEP formulation. The RHS of Fig. 4.3 confirms the expected accuracies of union-bound SEP and theoretical SEP predictions for the low and high SNR limits.

4.6.2 Performance Analysis with Constant Envelope Signals and Low-Resolution DACs

This section presents comparisons of the proposed precoding algorithms against the state-of-the-art approaches, and, of the MSEP and MUBSEP criteria against the MMSE and MMDDT formulations. These are made in terms of symbol-error rate (SER) and computational complexity considering a MIMO scenario of a BS with $M = 12$ antennas serving $K = 3$ single antenna users with $\alpha_s = \alpha_x = 4$. To facilitate the performance analysis of the proposed QoS B&B precoders the same SEP requirements are considered for all users such that $\boldsymbol{\lambda} = 10^{-\tau} \cdot \mathbf{1}_K$. The performance evaluation of the MSEP and MUBSEP formulations is based on full B&B methods that yield optimal precoding vectors in their corresponding design criterion, which can be realized with $\tau = \infty$. Since the works from [18] and [48] are also optimal in terms of the MMDDT and MMSE, respectively, comparing their performance is equivalent to comparing their design objectives.

4.6.2.1 Performance versus SNR evaluation

In this subsection, the SEP requirement parameter is set to $\tau = 3$, which corresponds to $\boldsymbol{\lambda} = 10^{-3} \cdot \mathbf{1}_K$. The first experiment, shown in the upper LHS of Fig. 4.4, consists of the evaluation of the SER. While the upper LHS of Fig. 4.4 shows similar SER for the MSEP and MUBSEP full B&B algorithms, it also shows that the full B&B methods using the MSEP and MUBSEP criteria outperform the MMSE and MMDDT B&B approaches for all examined SNR values in terms of SER. Since the full B&B methods are optimal in terms of their respective design criterion it can be concluded that utilizing the MSEP and MUBSEP formulations yields smaller SER than utilizing the MMSE and

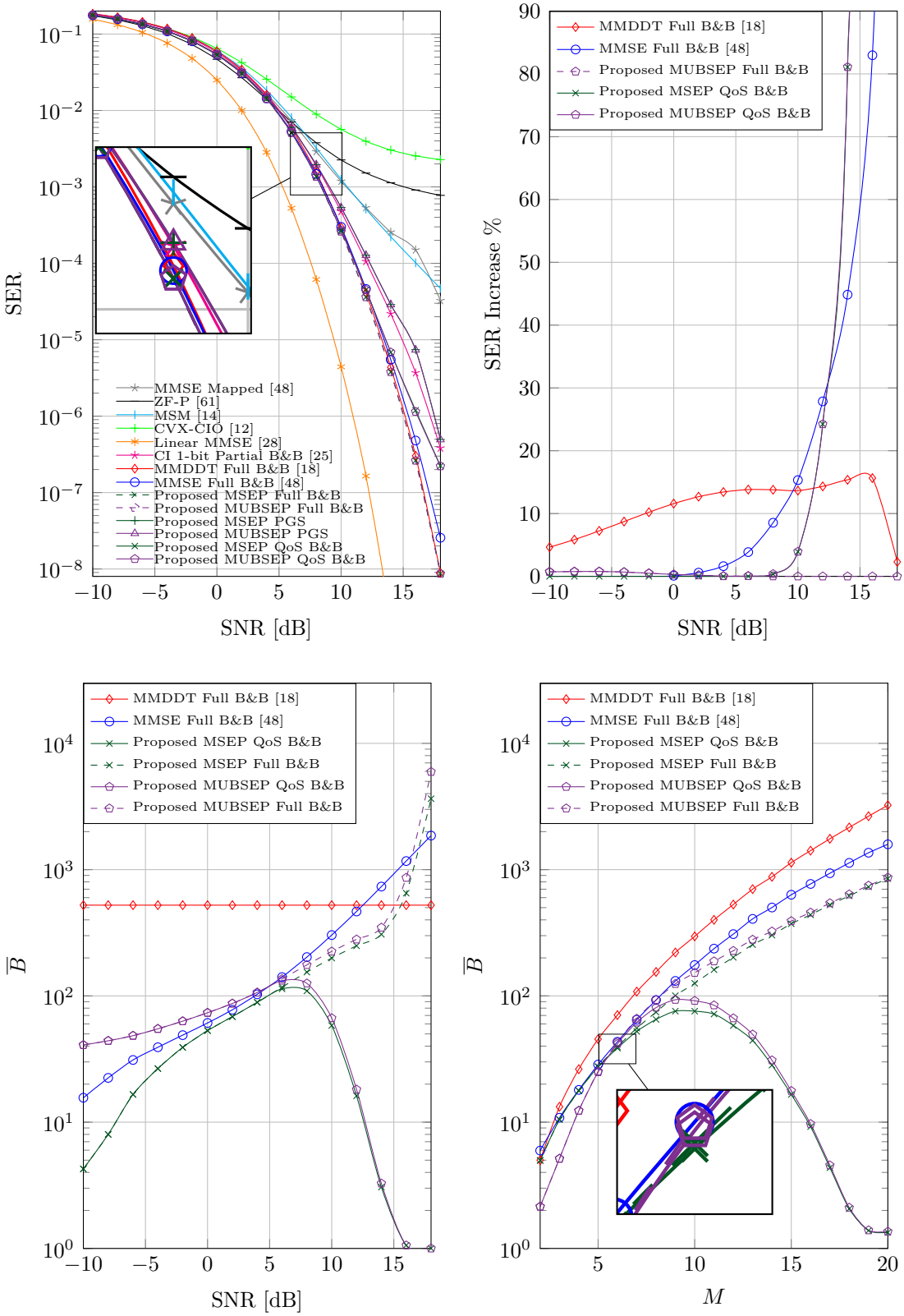

Figure 4.4: Considered scenario: $K = 3$ users, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols and SEP requirement parameter $\tau = 3$. SER $\times$ SNR for $M = 12$ antennas (Upper LHS). SER Increase % $\times$ SNR for $M = 12$ antennas (Upper RHS). Average number of convex optimization problems solved $\overline{B}$ $\times$ SNR for $M = 12$ antennas (Lower LHS). Average number of convex optimization problems solved $\overline{B} \times M$ for SNR = 10 dB (Lower RHS).

Table 4.1: Computational Complexity of the SLP Algorithms

Algorithm	Complexity	Problem Type
CVX-CIO [12]	$\mathcal{O}\left(M^{3.5}\right)$	Second Order Cone Program
MSM-Precoder [14]	$\mathcal{O}\left(M^{3.5}\right)$	Linear Program
MMSE Mapped [48]	$\mathcal{O}\left(M^{3.5}\right)$	Quadratic Program
CI-1 bit Partial B&B [25]	$\mathcal{O}\left(M^{3.5} + B\ K^{3.5}\right)$	Discrete Programming Problem
MMDDT B&B [18]	$\mathcal{O}\left(B\ M^{3.5}\right)$	Discrete Programming Problem
MMSE B&B [48]	$\mathcal{O}\left(B\ M^{3.5}\right)$	Discrete Programming Problem
Proposed MSEP PGS	$\mathcal{O}\left(M^{3.5}\right)$	Convex Optimization Problem
Proposed MSEP QoS B&B	$\mathcal{O}\left(B\ M^{3.5}\right)$	Discrete Programming Problem
Proposed MUBSEP PGS	$\mathcal{O}\left(M^{3}\sqrt{M+K}\right)$	Convex Optimization Problem
Proposed MUBSEP QoS B&B	$\mathcal{O}\left(B\ M^{3}\sqrt{M+K}\right)$	Discrete Programming Problem

MMDDT designs. Note that, the SER $\times$ SNR curve of the full MSEP B&B in Fig. 4.4 represents the minimum uncoded SER that any precoder that attains the system model can achieve. It is seen in the upper LHS of Fig. 4.4 that the proposed MSEP and MUBSEP QoS B&B approaches yield similar SER for SNR $<$ 10 dB, yet for SNR $\geq$ 10 dB the proposed QoS B&B approaches start deviating from full B&B methods. This is the case since, for this SNR region, the proposed QoS B&B methods attain the SEP requirement without the need for optimal vectors and return suboptimal solutions with reduced complexity. Furthermore, the upper LHS of Fig. 4.4 shows that the proposed MSEP PGS and MUBSEP PGS algorithms outperform all other suboptimal designs for all examined SNR values, except for the CI 1-bit Partial B&B precoder.

As seen in the upper RHS of Fig. 4.4 the MUBSEP full B&B algorithm yields approximately no increase in SER when compared with the MSEP full B&B baseline. On the other hand, the proposed QoS B&B approaches yield a significant increase in SER for SNR $\geq$ 10 dB. This happens since, in this SNR region, the SEP requirement is attained with lower complexity suboptimal vectors as will be discussed in the analysis of the lower plots of Fig. 4.4.

The computational complexity analysis is done by comparing the UBCO of each algorithm, achieved by considering that the optimization-based approaches are solved with the barrier method. While the complexity of the considered channel-level ZF-P and LMMSE techniques are $\mathcal{O}(K^2M)$ and $\mathcal{O}(K^3)$, respectively, the UBCO of the state-of-the-art SLPs is shown in Table 4.1, where B denotes a given number of optimization problems solved in the corresponding B&B algorithm. As shown in Table 4.1, the proposed MSEP and MUBSEP PGS algorithms have similar complexity as the other suboptimal designs and have reduced complexity order as compared with the CI 1-bit Partial B&B technique, which can yield unfavorable complexity for large K [41]. For comparing the complexity of the considered B&B algorithms an evaluation of B is required. This is done in the experiments presented in the lower

plots of Fig. 4.4 where the average value of B, termed $\overline{B}$, is evaluated against the SNR (on the LHS) and against the number of transmit antennas M (on the RHS). Since the CI 1-bit Partial B&B method's complexity scales with $\mathcal{O}\left(M^{3.5} + BK^{3.5}\right)$ instead of $\mathcal{O}\left(BM^{3.5}\right)$ from other B&B approaches, it is not considered for evaluation in the lower plots of Fig. 4.4.

The third experiment, present in the lower LHS of Fig. 4.4, shows that the $\overline{B}$ of the full MSEP B&B algorithm is smaller than the ones from MMDDT for SNR $<$ 16 dB and than MMSE for SNR $<$ 18 dB. Furthermore, the full MUBSEP B&B algorithm yields smaller values of $\overline{B}$ than MMDDT for SNR $<$ 16 dB and than MMSE for 4 dB $<$ SNR $<$ 18 dB. The proposed MSEP and MUBSEP QoS B&B designs yield similar $\overline{B}$ as the full B&B methods for SNR $\leq$ 6 dB, yet, as the SNR increases, the values of $\overline{B}$ of the proposed QoS B&B algorithms decrease until for SNR $=$ 18 dB both QoS B&B methods yield $\overline{B} \approx 1$. This happens since, due to the overall SER decreasing with the SNR, the proposed QoS B&B precoding algorithm requires, on average, fewer visited branches to find a vector that attains the SEP requirement. Note that, having $\overline{B} = 1$ means that the proposed QoS B&B methods have a similar UBCO as low-complexity state-of-the-art approaches. The fourth experiment, presented in the lower RHS of Fig. 4.4 for SNR $=$ 10 dB, shows that $\overline{B}$ of the full MSEP B&B method is smaller than all other state-of-the-art methods considered for all evaluated values of M. The full MUBSEP B&B algorithm yields smaller $\overline{B}$ than MMDDT for all evaluated values of M and than MMSE for $M \leq 6$ and $M \geq 9$. The number of optimization problems solved by the proposed full MUBSEP B&B algorithm approaches its MSEP counterpart as the number of BS antennas increases. Moreover, it can be seen that the value of $\overline{B}$ of the proposed QoS B&B algorithms initially grows with M, yet, for $M > 9$, $\overline{B}$ starts to decrease such that for $M = 20$ the proposed MSEP and MUBSEP QoS B&B methods yield $\overline{B} \approx 1.3$. This is explained as follows: since the SEP decreases with M, for fixed τ and SNR, including more antennas eventually makes the system attain the SEP requirement. After this, if M continues increasing, the benefits in SEP allow for the usage of suboptimal vectors, decreasing the number of optimization problems solved. To underline this effect consider the UBCO of $M = 10$ with $\overline{B}_{M=10} = 75.68$, which yields $\mathcal{O}(2.39{\cdot}10^5)$, and, $M = 20$ with $\overline{B}_{M=20} = 1.33$, which yields $\mathcal{O}(4.78{\cdot}10^4)$. With this, different than full B&B approaches, large-scale MIMO can be beneficial in terms of complexity for the proposed QoS B&B algorithms.

Fig. 4.4 also underlines that when operating at sufficiently high SNR the QoS B&B approach yields similar complexity as PGS as seen in the lower RHS. Yet, for an SNR decrease, the proposed QoS B&B algorithm adjusts by

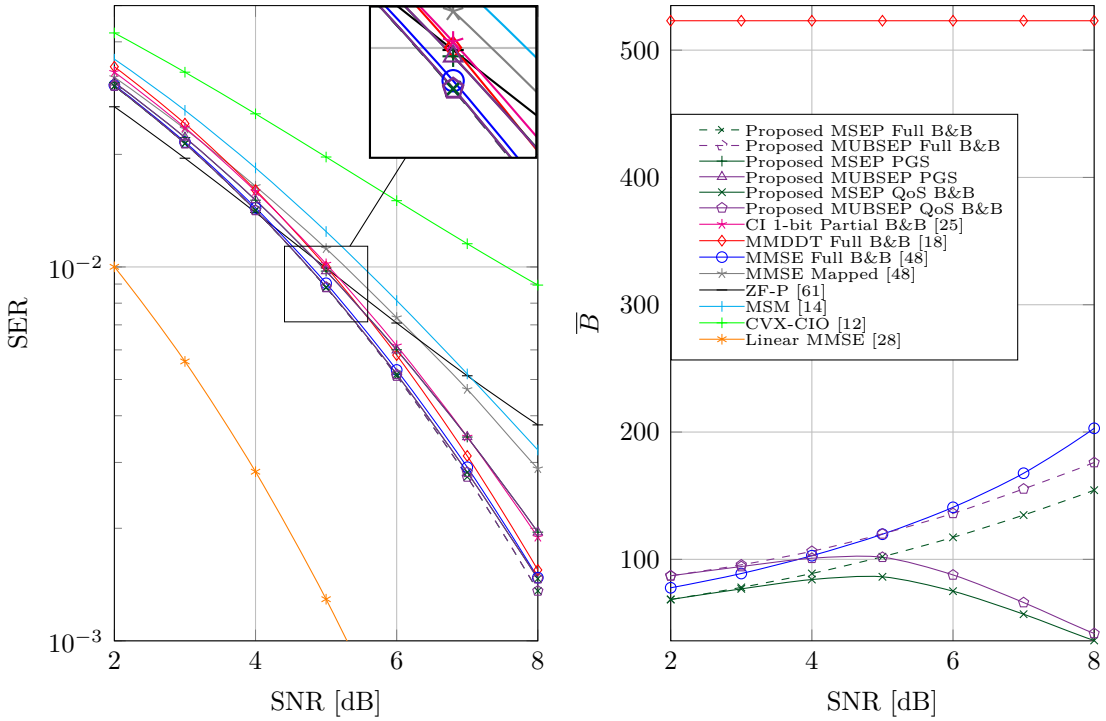

Figure 4.5: Considered scenario: $K = 3$ users, $M = 12$ BS antennas, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols and QoS constraint vector $\boldsymbol{\lambda} = 10^{-2} \cdot \mathbf{1}_K$. SER $\times$ SNR (left). Average number of convex optimization problems solved $\overline{B} \times$ SNR (right).

increasing the complexity such that the system's SEP is maintained smaller than the SEP requisite as seen in lower LHS and the upper RHS. For extreme scenarios where attaining the SEP requirements is not possible (or is possible only with the optimal solution) the proposed QoS B&B approach yields the optimal solution with full B&B complexity as seen in all plots of Fig. 4.4.

Performance Evaluation for a Reduced SNR Range Section 4.6.2 evaluates the performance of the considered approaches for a broad SNR range to illustrate the different aspects of the proposed methods that rise in the diverse SNR regimes. This section evaluates the performance of the proposed methods for a limited range allowing for a more adequate visualization of the benefits of the proposed methods in this SNR regime. In this section, $\tau = 2$ is considered, meaning $\boldsymbol{\lambda} = 10^{-2} \cdot \mathbf{1}_K$.

The LHS of Fig. 4.5 shows that the proposed MSEP and MUBSEP Full B&B and QoS B&B approaches outperform, in terms of SER, all state-of-the-art techniques for the considered SNR range. Moreover, as shown in the RHS of Fig. 4.5, the proposed MSEP and MUBSEP B&B approaches require less computational complexity when compared to the MMSE and

MMDDT Full B&B methods for all examined SNR and for SNR $>$ 5 dB, respectively. With this, one can state that the proposed B&B techniques yield higher SER performance with reduced computational complexity compared with the state-of-the-art B&B techniques in many scenarios. Furthermore, Fig. 4.5 also demonstrates that a significant reduction in computational complexity with a minor increase in SER can be achieved when utilizing the proposed MSEP and MUBSEP QoS B&B approaches instead of their full B&B counterparts. Finally, the LHS of Fig. 4.5 shows that the proposed PGS approaches outperform all other considered techniques with the same UBCO for all SNR, and outperform the CI 1-bit Partial B&B approach for SNR ≤ 7 dB.

Complexity Evaluation with Array-Gain The general SNR definition derived in section C.1 of the appendix, i.e., SNR $= {}^{(M \cdot \mathrm{P_A})}/_{\sigma_w^2}$, considers a generic array in which the transmit vector implies $\boldsymbol{C}_x = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\} = \mathrm{P_A}\boldsymbol{I}$. This consideration, in general, does not influence the analysis and information obtained from the experiments. Yet, for the special case of the lower RHS of Fig. 4.4, since the complexity of proposed QoS B&B decreases with the SNR, one could argue that the complexity benefits from the proposed QoS B&B methods come from the increase in receive power experienced by the users, that arises due to the actual receive SNR increasing with M, which is not taken into account in the SNR definition of section C.1 of the appendix.

To investigate this hypothesis, this section considers the alternative SNR definition derived in section C.2 of the appendix, i.e., SNR $= {}^{(M^2 \cdot \mathrm{P_A})}/_{\sigma_w^2}$. This definition considers an array that utilizes the maximum ratio transmission (MRT) beamformer which is known to maximize the received SNR [62, Section 7.3.1]. Since the considered precoders do not focus on received SNR maximization, the SNR definition, SNR $= {}^{(M^2 \cdot \mathrm{P_A})}/_{\sigma_w^2}$, can be considered as an upper bound on the average received SNR. With this, if an increase in M still leads to a QoS B&B complexity decrease, one can affirm that the complexity benefits of the QoS B&B design come from the increased number of BS antennas and not due to the growth in received SNR.

The experiment of this section evaluates the average number of optimization problems solved in the corresponding B&B algorithm, $\overline{B}$, versus the number of BS antennas, M, for a SEP requirement parameter of $\tau = 1$, meaning $\boldsymbol{\lambda} = 10^{-1} \cdot \mathbf{1}_K$. Fig. 4.6, shows that different from the Full B&B approaches, the value of $\overline{B}$ of the proposed MSEP and MUBSEP QoS B&B, after an initial increase, decreases until $\overline{B} \approx 1$ for $M = 20$. This implies that the decrease in computational complexity comes from the increase in M which corroborates

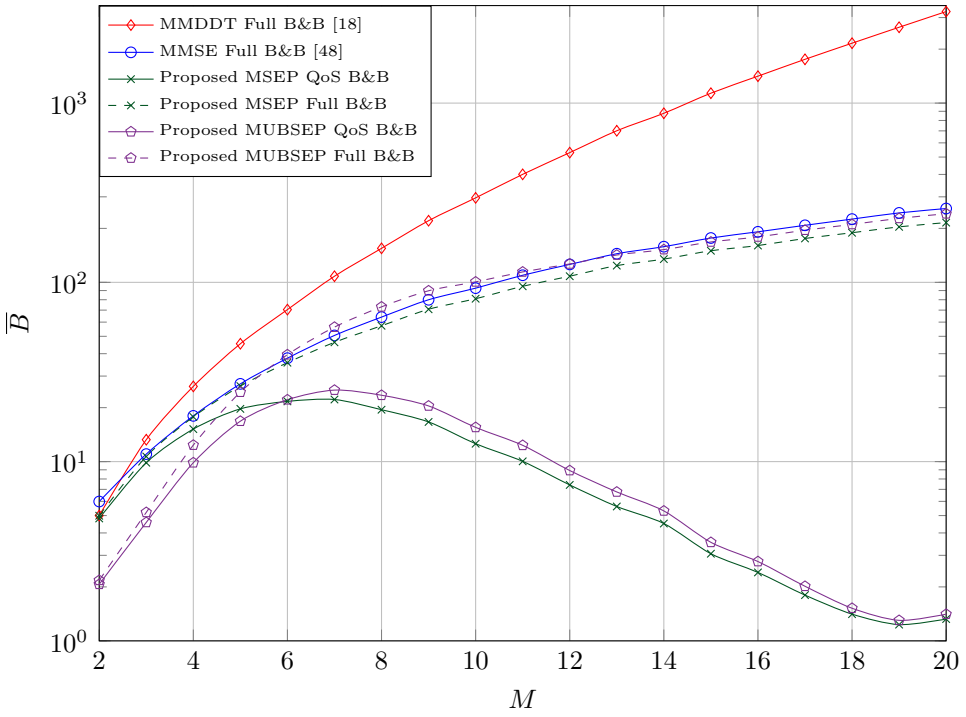

Figure 4.6: Average number of convex optimization problems solved versus number of antennas, $\overline{B} \times M$, for SNR $= {}^{(M^2 \cdot \mathrm{P_A})}/_{\sigma_w^2} = 16$ dB, $K = 3$ users, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols, SEP requirement parameter $\tau = 1$.

the idea that the large-scale MIMO can be beneficial in terms of computational complexity for the proposed QoS B&B algorithm.

4.6.2.2 Performance versus QoS parameter evaluation

This subsection evaluates the performance against the SEP parameter τ with SNR $= 12$ dB. The first experiment (shown in the upper plots of Fig. 4.7) consists of the evaluation of the SER $\times\, \tau$ and $\overline{B} \times\, \tau$. As seen in the upper LHS of Fig. 4.7, for small values of τ the proposed QoS B&B algorithms yield the same SER performance as the proposed PGS algorithms. Yet, as τ increases, meaning the system requires smaller SEP, the SER from proposed QoS B&B algorithms decreases until it reaches the optimal performance limit. As shown in the upper RHS of Fig. 4.7, $\overline{B} = 1$ for the proposed QoS B&B algorithms for small values of τ. This is expected since, for this τ region, the SER of PGS is sufficient for attaining the system's SEP requisite. As τ increases, the values of $\overline{B}$ also increase approaching $\overline{B}$ of full B&B techniques.

The second experiment (shown in the lower plots of Fig. 4.7) consists of the evaluation of the SER decrease and UBCO increase versus the value of τ, where we use the MSEP PGS as a baseline. With this, SER decrease % =

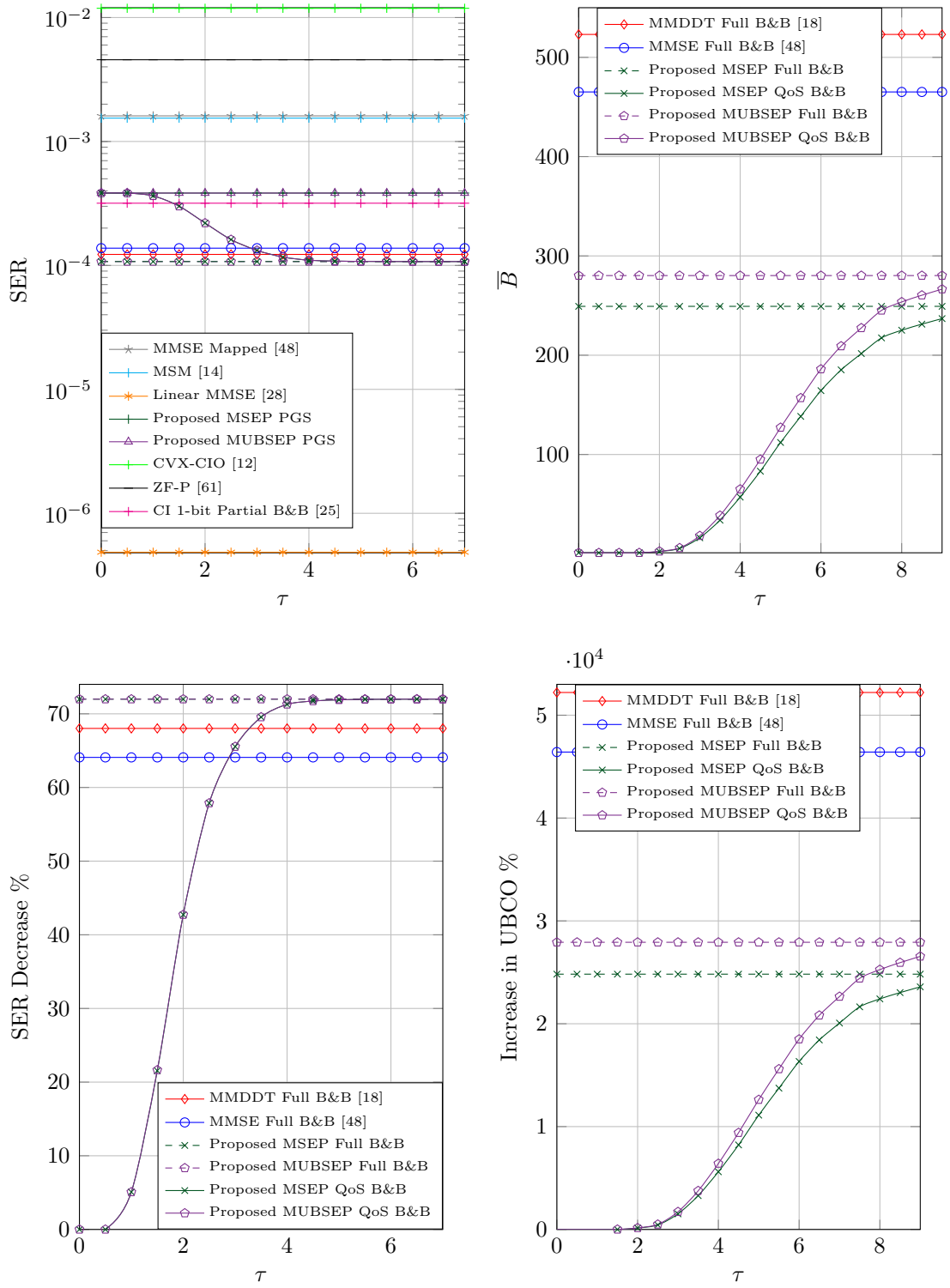

Figure 4.7: Considered scenario: $K = 3$ users, $M = 12$ antennas, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols, SNR = 12dB, QoS constraint vector $\boldsymbol{\lambda} = 10^{-\tau} \cdot \mathbf{1}_K$. SER $\times$ SEP requirement parameter τ (upper left). Average number of convex optimization problems solved $\overline{B} \times \tau$ (upper right). SER decrease % $\times \tau$ (lower left). Increase in $\overline{B}$ % $\times \tau$ (lower right).

$\frac{\text{SER}_{\text{MSEP PGS}}-\text{SER}}{\text{SER}_{\text{MSEP PGS}}} \times 100$, and, Increase in UBCO% $= \left(\overline{B}-1\right) \times 100$. As seen in the lower plots of Fig. 4.7 for $\tau \geq 4$ the proposed QoS B&B algorithms saturate approximately at 72% SER decrease outperforming the MMDDT and MMSE B&B algorithms while presenting a smaller increase in UBCO as compared to the full B&B algorithms.

Finally, a joint analysis of the plots of Fig. 4.7 shows that, even though for $\tau \geq 4$ the SER loss of utilizing the proposed QoS B&B approaches is negligible (approximately $2.7 \cdot 10^{-6}$ for $\tau = 4$), the complexity benefits can be substantial (complexity decrease factor of approximately 4.4 for $\tau = 4$). This underscores the effectiveness of the proposed QoS B&B method in achieving advantageous complexity-performance trade-offs, making it a robust choice for scenarios requiring high efficiency without significant performance sacrifice.

4.6.3 Performance under Imperfect CSI

In this section, the SER performance is evaluated for an imperfect CSI scenario. To this end, the channel is considered to be modeled as $\boldsymbol{H} = \sqrt{\xi}\hat{\boldsymbol{H}} + \sqrt{1-\xi}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a random matrix with i.i.d. zero-mean and unit-variance, $\xi \in (0,1]$ is a parameter that describes the level of the CSI mismatch and $\hat{\boldsymbol{H}}$ is considered as known at the transmitter. Note that, despite different notations, the channel model considered is equivalent to the one proposed in section 4.5. We highlight that, in the current scenario, the proposed RMMSE methods are not equivalent to their MMSE counterparts. For the simulations, a MIMO scenario with $K = 3$ users, $M = 8$ BS antennas, and $\alpha_s = \alpha_x = 4$ is considered. Moreover, the SEP requisite for the QoS B&B algorithms is set to $\tau = 3$ which implies $\boldsymbol{\lambda} = 10^{-3} \cdot \mathbf{1}_K$.

4.6.3.1 SER versus CSI Imperfection

This subsection evaluates the SER performance against the CSI imperfection with SNR = 12 dB, at the LHS of Fig. 4.8, and with SNR = 15 dB at the RHS of Fig. 4.8. To facilitate the analysis we define $\gamma = 1 - \xi$. As shown in the LHS Fig. 4.8, the MSEP and MUBSEP QoS B& B and full B&B approaches outperform all state-of-the-art quantized non-robust techniques for all values of γ. The proposed PGS approaches outperform all non-robust suboptimal methods except for the CI 1-bit Partial B&B which yields similar performance for $\gamma > 0.1$. The proposed PGS approaches also outperform the MMSE Full B&B method for $\gamma > 0.1$. Regarding the proposed robust approaches it can be seen that the proposed RMMSE B&B method outperforms

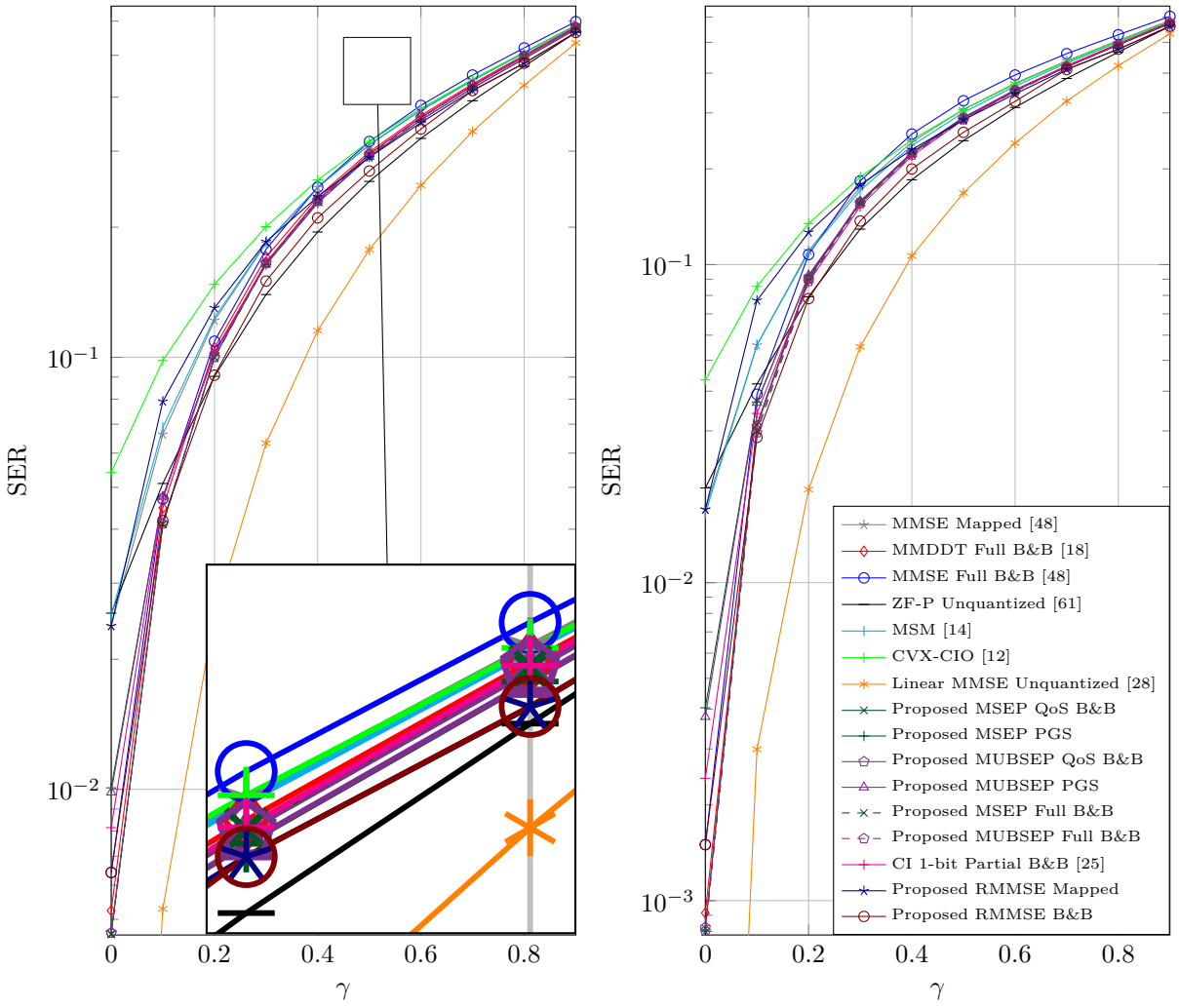

Figure 4.8: SER × CSI imperfection factor γ, for $K = 3$ users, $M = 8$ antennas, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols. SNR = 12 dB (left). SNR = 15 dB (right)

all other quantized approaches for all values of γ which then justifies its suitability for imperfect CSI scenarios. The proposed RMMSE Mapped approach outperforms all quantized approaches, except for the RMMSE B&B technique, for $\gamma \geq 0.5$ which justifies its usage for scenarios of high CSI imperfection. The RHS Fig. 4.8 shows similar results and reaffirms the conclusions discussed regarding the LHS Fig. 4.8. Yet, similarly as in section 3.5.2, comparing the plots reveals that the performance benefits of the proposed RMMSE approach are more pronounced in the RHS. This outcome is expected, as the RMMSE technique accounts for CSI imperfection, and as the SNR increases, performance becomes increasingly dominated by CSI mismatch rather than noise.

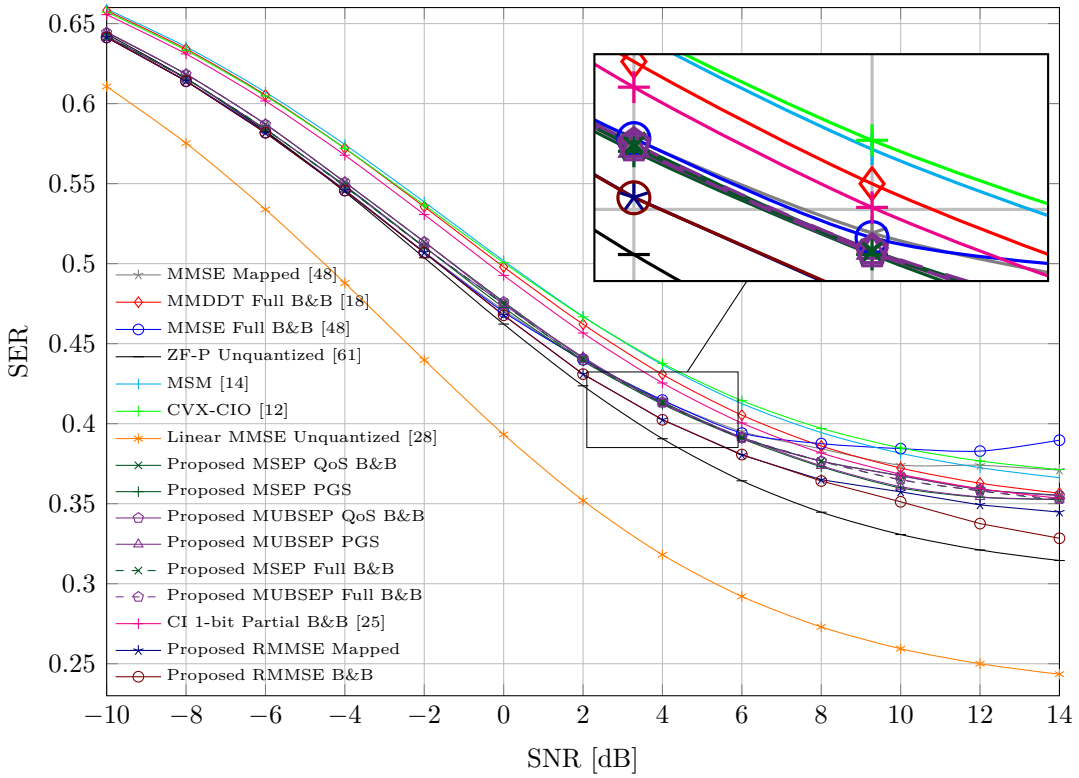

Figure 4.9: SER× SNR, for $K = 3$ users, $M = 8$ antennas, $\alpha_s = 4$ PSK users' data, $\alpha_x = 4$ PSK transmit symbols, CSI quality parameter $\xi = 0.4$.

4.6.3.2 SER versus SNR for a given CSI Imperfection

This subsection evaluates the SER performance against SNR for $\xi = 0.4$. As shown in Fig. 4.9, the MSEP and MUBSEP PGS approaches outperform all state-of-the-art quantized non-robust techniques for SNR $\geq$ 4 dB. The proposed QoS and Full B&B techniques yield a comparable performance as PGS for SNR $\leq$ 6 dB. For SNR $>$ 6 dB the QoS and full B&B approaches yield a decrease in performance when compared with their PGS counterpart. Regarding the proposed robust approaches it can be seen that the proposed RMMSE B&B method outperforms all other quantized approaches for all values of SNR. Finally, the proposed RMMSE Mapped approach outperforms all quantized approaches except for the RMMSE B&B method which yields similar performance for SNR $\leq$ 8 dB.

5
Power Minimization under Quality of Service Constraints for RIS-based Passive Transmitter MIMO Systems

As mentioned in chapter 1, to enable the foreseen applications of future generations of wireless communications, attainment of strict QoS requisites, e.g., low latency, high reliability, and, high data rate [63, 64], are necessary. While MU-MIMO systems are expected to be a key technology for attaining these requirements [2], as previously discussed, equipping a BS with large antenna arrays can lead to high hardware costs and increased energy consumption which yields bottlenecks for the practical implementation. In this context, RIS has emerged as a promising technology for beyond-5G/6G wireless communications as it can improve the EE of wireless systems while offering the benefits of low-cost and easy integration into the currently deployed wireless systems [64, 65]. In essence, a RIS comprises an array of reconfigurable passive reflective elements where the reflection coefficient of each element is real-time electrically controllable. By adjusting the RIS elements' reflection coefficients one can perform passive signal shaping without necessitating a power amplifier which yields an advantage in terms of EE when compared with conventional MIMO BSs [66, 67]. With this, RIS has become popular in the literature of mmWave and multi-antenna communications, most commonly for backscatter communications [68, 69], wireless propagation environment control [70, 71] and beamforming [72, 73]. Regarding beamforming for RIS, recent advances showed that RIS is also able to perform simultaneous passive beamforming and transmit physical information [74, 75], which led to the development of different studies for this context. The work from [76] compares RIS-based modulation with spatial multiplexing and discusses the former's benefits over the latter. In [77] the authors introduce the concept of modulating intelligent surfaces as RIS capable of performing passive beamforming for users served by a BS, embedding information through backscatter communication, or doing both simultaneously. The work from [8] demonstrates the advantage of using RIS for modulation over traditional beamforming, where RIS phase shifts are independent of transmitted information.

Following the works form [7, 8, 9, 10, 11] this study utilizes the RIS modulation capabilities with RIS-based passive transmitter setup which real-

izes low-cost energy-efficient massive MIMO. As proposed in [7, 8, 9, 10, 11], an efficient transmitter can be realized by illuminating a RIS with an unmodulated carrier signal generated by a nearby RF signal generator and changing the parameters of the reflecting elements to modulate and transmit information symbols. With this, RIS-based passive transmitters realize virtual MIMO systems with a single RF chain and cost-effective reflecting elements, which benefits the implementation of massive MIMO with reduced hardware complexity and increased EE. By employing the RIS-based passive transmission setup the optimization of the transmit signal can be done utilizing a similar mechanism as in symbol-level precoding [18, 27, 78, 26, 41, 79, 80], which achieves high-performance by varying the precoder for each symbol vector. In what follows the system model considered for this chapter is exposed and a brief revision of the optimization for RIS-based passive transmitters literature is provided.

5.1 System Model

The system model consists of an RF generator illuminating an RIS with N reflecting elements that serve K single antenna users as depicted in Fig. 5.1.

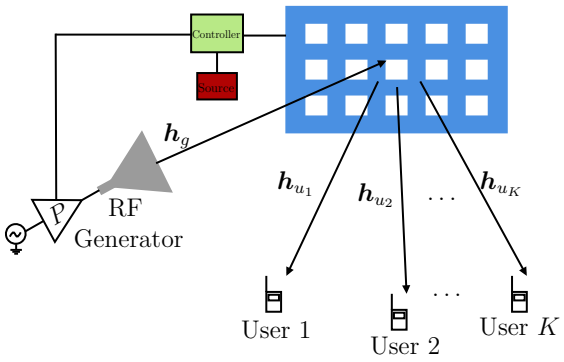

Figure 5.1: Multiuser MIMO downlink via passive RIS reflection

A symbol-level transmission is considered with the symbols generated by a memoryless source connected to the RIS controller. The data symbol of the k-th user is denoted as s_k, such that for all $k \in \mathcal{K} = \{1, \ldots, K\}$, $s_k \in \mathcal{S}$, where $\mathcal{S}$ represents all possible symbols of a α_s-PSK modulation. The symbols of all users are described in a stacked vector notation as $\boldsymbol{s} = [s_1, \ldots, s_K]^T$. Based on $\boldsymbol{s}$ the controller determines the phase shift vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T$, where θ_n is considered to belong to the set $\mathcal{T}$ which is given by $\mathcal{T} = \{\theta : \theta = e^{\frac{j\pi(2i+1)}{\alpha_\theta}}$, for $i = 1, \ldots, \alpha_\theta\}$. Given the availability of channel estimation approaches [81, 82, 83, 84], perfect channel state information at the RIS is

considered. The received signal of the k-th user z_k, for all $k \in \mathcal{K}$, reads as

$$z_k = \sqrt{P}\boldsymbol{h}_k^H \boldsymbol{\theta} + w_k, \tag{5-1}$$

with P being the transmit power of the RF generator, $w_k \sim \mathcal{CN}(0, \sigma_w^2)$ representing additive white Gaussian noise, and, $\boldsymbol{h}_k = \boldsymbol{h}_{u_k}^H \text{diag}\,(\boldsymbol{h}_g)$, being the effective channel, where $\boldsymbol{h}_g$ is the channel between the RF generator and the RIS and $\boldsymbol{h}_{u_k}$ is the channel between the RIS the k-th user.

Each z_k is hard detected based on the decision region it belongs. The decision region of s_i, termed $\mathcal{S}_i$, is the set of points closer to s_i than all other valid candidates for detection. This implies that z_k is detected as s_i if $z_k \in \mathcal{S}_i$. For PSK the decision regions are circle sectors with infinite radius and angle of 2ϕ, where $\phi = \pi/\alpha_s$. The detected symbol vector is written as $\hat{\boldsymbol{s}} = [\hat{s}_1, \ldots, \hat{s}_K]$. It is considered that the RIS controller knows the SEP requirement of all users, with the SEP requirement denoted as ρ_k.

5.2 Literature Review

Different works have arisen considering RIS-based passive transmission schemes. In [85] the authors jointly optimize the total power reflected from the RIS and the power allocation fraction assigned to each user. In [77] to maximize each user's spectral efficiency, the authors propose a joint non-convex optimization problem using the sum minimum mean-square error criterion.

This section exposes the work from [19] as it is the study most related to the methods proposed in the chapter. In [19] a power minimization problem is proposed under the condition that the MDDT of each user is greater or equal to a given requisite. Following the derivations presented in section 3.2.2.2, one can construct the MDDTs for the considered system model as

$$d_k = \sqrt{P}\left(\text{Re}\{s_k^* \boldsymbol{h}_k^H \boldsymbol{\theta}\}\sin\phi - \left|\text{Im}\{s_k^* \boldsymbol{h}_k^H \boldsymbol{\theta}\}\right|\cos\phi\right), \text{ for } k \in \mathcal{K}. \tag{5-2}$$

Although not directly explicated, the authors consider a sufficiently high-resolution RIS such that the feasible set of the reflection coefficients $\mathcal{T}$ can be well approximated by $\mathcal{T} = \{\theta : |\theta|^2 = 1\}$. With this, the power minimization problem for high-resolution RIS under MDDT constraints (PHMDDT) is

constructed as

$$\begin{aligned} \min_{\boldsymbol{\theta},P} \;& P \\ \text{s.t.} \;& \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}, \quad P \geq 0, \quad r_k = \sqrt{P}\boldsymbol{h}_k^H \boldsymbol{\theta} e^{-j\arg(s_k)}, \\ & \mathrm{Re}\{r_k\}\sin\phi - |\mathrm{Im}\{r_k\}|\cos\phi \geq \alpha_k, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-3}$$

where α_k for $k \in \mathcal{K}$ is the given MDDT requisite. To facilitate solving (5-3) the MDDT constraints are reformulated by diving both sides by $\alpha_k\sqrt{P}$, which yields

$$\frac{1}{\sqrt{P}} \leq \frac{1}{\alpha_k}\left(\mathrm{Re}\{\hat{r}_k\}\sin\phi - |\mathrm{Im}\{\hat{r}_k\}|\cos\phi\right), \tag{5-4}$$

where $\hat{r}_k = \boldsymbol{h}_k^H \boldsymbol{\theta} e^{-j\arg(s_k)}$. By introducing the variable $t = \frac{1}{\sqrt{P}}$ the power minimization problem described in (5-3) is reformulated as

$$\begin{aligned} \max_{\boldsymbol{\theta},t} \;& t \\ \text{s.t.} \;& \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}, \quad \hat{r}_k = \boldsymbol{h}_k^H \boldsymbol{\theta} e^{-j\arg(s_k)}, \\ & t \leq \frac{1}{\alpha_k}\left(\mathrm{Re}\{\hat{r}_k\}\sin\phi - |\mathrm{Im}\{\hat{r}_k\}|\cos\phi\right), \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-5}$$

which is equivalently written as

$$\begin{aligned} \max_{\boldsymbol{\theta}} \; & \min_{k\in\mathcal{K}} \; \frac{1}{\alpha_k}\left(\mathrm{Re}\{\hat{r}_k\}\sin\phi - |\mathrm{Im}\{\hat{r}_k\}|\cos\phi\right) \\ \text{s.t.} \; & \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}, \quad \hat{r}_k = \boldsymbol{h}_k^H \boldsymbol{\theta} e^{-j\arg(s_k)}. \end{aligned} \tag{5-6}$$

Defining $\boldsymbol{a}_k^H = \frac{1}{\alpha_k}\boldsymbol{h}_k^H e^{-j\arg(s_k)}$ for $k \in \mathcal{K}$ (5-6) is cast as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \; & \max_{k\in\mathcal{K}} \; \left|\mathrm{Im}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\right|\cos\phi - \mathrm{Re}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\sin\phi \\ \text{s.t.} \; & \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}. \end{aligned} \tag{5-7}$$

To avoid the absolute value and the max$(\cdot)$ and to get a twice differentiable objective the objective is written in terms of the log-sum-exp function which yields

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & \epsilon \ln \sum_{k=1}^{K} e^{\left(\frac{\mathrm{Im}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\cos\phi - \mathrm{Re}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\sin\phi}{\epsilon}\right)} + e^{\left(\frac{\mathrm{Im}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\cos\phi + \mathrm{Re}\{\boldsymbol{a}_k^H\boldsymbol{\theta}\}\sin\phi}{\epsilon}\right)} \\ \text{s.t.} \; & \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}. \end{aligned} \tag{5-8}$$

Problem (5-8) is the minimization of a convex function under the N-

dimensional complex circle manifold $\mathcal{M} = \left\{ \boldsymbol{\theta} \in \mathbb{C}^N |\ [\boldsymbol{\theta}]_n^* [\boldsymbol{\theta}]_n = 1,\ \text{for } n \in \mathcal{N} \right\}$. By defining

$$h(\boldsymbol{\theta}) = \epsilon \ln \sum_{k=1}^{K} e^{\left(\frac{\mathrm{Im}\{\boldsymbol{a}_k^H \boldsymbol{\theta}\} \cos\phi - \mathrm{Re}\{\boldsymbol{a}_k^H \boldsymbol{\theta}\} \sin\phi}{\epsilon} \right)} + e^{\left(\frac{\mathrm{Im}\{\boldsymbol{a}_k^H \boldsymbol{\theta}\} \cos\phi + \mathrm{Re}\{\boldsymbol{a}_k^H \boldsymbol{\theta}\} \sin\phi}{\epsilon} \right)}, \quad (5\text{-}9)$$

the problem is described as an unconstrained problem on $\mathcal{M}$ as

$$\min_{\boldsymbol{\theta} \in \mathcal{M}} h(\boldsymbol{\theta}), \quad (5\text{-}10)$$

where applying the Riemannian conjugate gradient (RCG) algorithm supports finding a local optimal solution. The authors of [19] propose a channel-level approach that involves solving (5-10) for each possibility $\boldsymbol{s} \in \mathcal{S}^K$. Although the methods proposed in the following sections of this thesis are also compatible with this idea, they are designed for symbol-level computation since, even for midsize systems, solving α_s^K optimization problems for each channel coherence time either leads to a high latency or requires a high number of parallel optimization problem hardware.

5.3 Contributions of this chapter

Following the path of [19], this chapter proposes a power minimization problem under QoS requisites. Yet, unlike in [19], it focuses on minimizing the power radiated by the RF generator to the RIS under the condition that the SEP of the users is below a given requisite. The discrete phase-shift RIS model is considered such that the reflecting elements' coefficients are restricted to a discrete set. For QPSK users' data, where the SEP can be expressed with tabled functions [26], the study proposes the power minimization under SEP constraints (PSEP) problem. For the general case of M-PSK users' data, utilizing the SEP would lead to constraint functions that require evaluation via Monte Carlo methods. With this, the PSEP problem is reformulated substituting the SEP by the UBSEP [26] in the constraint functions. Due to the UBSEP functions being an upper-bound on the SEP [86], the resulting problem of power minimization under UBSEP constraints (PUBSEP) is a restricted version of the PSEP problem. Based on the PSEP and PUBSEP problems we build on the QoS B&B algorithm by accepting not only solutions that attain the given target power budget of the system but also solutions that are sufficiently close to the optimal such that it is considered unnecessary to continue the search process. Numerical results underline that by utilizing the proposed B&B method significant complexity reduction can be achieved with

a minor increase in transmit power.

For the case where the number of discrete phase-shift RIS is large, i.e., for a high-resolution RIS, a reduced complexity method is proposed by approximating the discrete feasible set to its continuous counterpart. This leads to reformulating the proposed PSEP and PUBSEP problems as constrained optimization problems on an oblique manifold, which are solved via BMs. The proposed BM successively adjusts the transmit power while evaluating the feasibility of the QoS constraints by solving, via the RCG algorithm, an auxiliary problem dependent only on the coefficients of the RIS reflecting elements. Numerical results show that the proposed techniques yield low complexity with reduced transmit power compared to the state-of-the-art approach from [19].

5.4 Problem formulation

Under the considered system model the problem of power minimization under the condition that the SEP of each user k is below the given requisite, ρ_k, is cast as

$$\begin{aligned} \min_{\boldsymbol{\theta}, P \in \mathbb{R}_+} \quad & P \\ \text{s.t.} \quad & \mathrm{P}\left(\hat{s}_k \neq s_k | \boldsymbol{\theta}, P\right) \leq \rho_k, \text{ for } k \in \mathcal{K}, \\ & [\boldsymbol{\theta}]_n \in \mathcal{T}, \text{ for } n \in \mathcal{N}. \end{aligned} \tag{5-11}$$

As before, the detector decides for s_k when the received symbol z_k belongs to $\mathcal{S}_k$. With this, the SEP of the k-th user can be written as

$$\mathrm{P}\left(\hat{s}_k \neq s_k | \boldsymbol{\theta}, P\right) = 1 - \mathrm{P}\left(z_k \in \mathcal{S}_k | \boldsymbol{\theta}, P\right) = 1 - \frac{1}{\pi \sigma_w^2} \int_{\mathcal{S}_k} \mathrm{e}^{-\frac{\left|r - \sqrt{P} \boldsymbol{h}_k^H \boldsymbol{\theta}\right|^2}{\sigma_w^2}} dr. \tag{5-12}$$

The integral in (5-12) has tabled solutions for $\alpha_s \in \{2, 4\}$ and for $\alpha_s \notin \{2, 4\}$ requires solution via Monte-Carlo methods. With this, for the case of $\alpha_s \in \{2, 4\}$ the exact computation of the SEP is considered. For other PSK cases, this study considers substituting $\mathrm{P}\left(\hat{s}_k \neq s_k | \boldsymbol{\theta}, P\right)$ in (5-11) by the union-bound SEP [26].

5.4.1 Power Minimization under SEP constraints

For $\alpha_s \in \{2, 4\}$ the real and imaginary parts of the data symbols can be considered as independent. This allows the decision region $\mathcal{S}_k$ to be partitioned

as $\mathcal{R}_k \cap \mathcal{I}_k$, where $\mathcal{R}_k$ and $\mathcal{I}_k$ are the decision regions of the real and imaginary parts of s_k. With this, $\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right) = 1 - \mathrm{P}\left(z_k \in \mathcal{R}_k|\boldsymbol{\theta}, P\right)\mathrm{P}\left(z_k \in \mathcal{I}_k|\boldsymbol{\theta}, P\right)$, where

$$\begin{aligned}\mathrm{P}\left(z_k \in \mathcal{R}_k|\boldsymbol{\theta}, P\right) &= \int_0^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}} e^{-\frac{\left(t - \mathrm{sign}(\mathrm{Re}\{s_k\})\mathrm{Re}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}\right)^2}{\sigma_w^2}} dt \\ &= \Phi\left(\sqrt{\frac{2P}{\sigma_w^2}}\,\mathrm{sign}(\mathrm{Re}\left\{s_k\right\})\mathrm{Re}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}\right), \end{aligned} \tag{5-13}$$

$$\begin{aligned}\mathrm{P}\left(z_k \in \mathcal{I}_k|\boldsymbol{\theta}, P\right) &= \int_0^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{\left(t - \mathrm{sign}(\mathrm{Im}\{s_k\})\mathrm{Im}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}\right)^2}{\sigma_w^2}} dt \\ &= \Phi\left(\sqrt{\frac{2P}{\sigma_w^2}}\,\mathrm{sign}(\mathrm{Im}\left\{s_k\right\})\mathrm{Im}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}\right), \end{aligned} \tag{5-14}$$

where $\Phi(\cdot)$ denotes the cumulative Gaussian distribution function. As a consequence, the SEP of the k-th user can be cast as

$$\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right) = 1 - \Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{r},k}(\boldsymbol{\theta})\right)\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{i},k}(\boldsymbol{\theta})\right), \tag{5-15}$$

where $\boldsymbol{v}_{\mathrm{r},k}\left(\boldsymbol{\theta}\right) = \sqrt{2}\ \mathrm{sign}\left(\mathrm{Re}\left\{s_k\right\}\right)\mathrm{Re}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}$ and $\boldsymbol{v}_{\mathrm{i},k}(\boldsymbol{\theta}) = \sqrt{2}\ \mathrm{sign}\left(\mathrm{Im}\left\{s_k\right\}\right)\mathrm{Im}\{\boldsymbol{h}_k^H\boldsymbol{\theta}\}$. Constraining the SEP to be smaller or equal to ρ_k is equivalent to constraining the correct detection probability to be greater or equal to $1 - \rho_k$. With this, one can write (5-11) as

$$\begin{aligned}&\min_{\boldsymbol{\theta}, P \in \mathbb{R}_+} \quad P \\ &\text{s.t. } \left[\boldsymbol{\theta}\right]_n \in \mathcal{T}, \text{ for } n \in \mathcal{N}, \\ &\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{r},k}(\boldsymbol{\theta})\right)\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{i},k}(\boldsymbol{\theta})\right) \geq 1 - \rho_k, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-16}$$

Problem (5-16) is rewritten with real-valued variables as

$$\begin{aligned}&\min_{\boldsymbol{\theta}_{\mathrm{r}}, P \in \mathbb{R}_+} \quad P \\ &\text{s.t.} \quad \left[\boldsymbol{\theta}_{\mathrm{r}}\right]_{2n-1} + j\left[\boldsymbol{\theta}_{\mathrm{r}}\right]_{2n} \in \mathcal{T}, \text{for } n \in \mathcal{N}, \\ &\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{1,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right)\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{2,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right) \geq 1 - \rho_k, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-17}$$

where $\boldsymbol{\theta}_{\mathrm{r}} = R(\boldsymbol{\theta}) = [\mathrm{Re}\{[\boldsymbol{\theta}]_1\},\ \mathrm{Im}\{[\boldsymbol{\theta}]_1\},\ \cdots,\ \mathrm{Re}\{[\boldsymbol{\theta}]_N\},\ \mathrm{Im}\{[\boldsymbol{\theta}]_N\}]^T$,

$$\boldsymbol{h}_{1,k}^T = \left[[\boldsymbol{\gamma}_{R,k}]_1, -[\boldsymbol{\gamma}_{I,k}]_1, \ldots, [\boldsymbol{\gamma}_{R,k}]_N, -[\boldsymbol{\gamma}_{I,k}]_N\right], \tag{5-18}$$

$$\boldsymbol{h}_{2,k}^T = \left[[\boldsymbol{\gamma}_{I,k}]_1\}, [\boldsymbol{\gamma}_{R,k}]_1, \ldots, [\boldsymbol{\gamma}_{I,k}]_N\}, [\boldsymbol{\gamma}_{R,k}]_N\}\right], \tag{5-19}$$

with $\boldsymbol{\gamma}_{R,k} = \sqrt{2}\,\mathrm{sign}\left(\mathrm{Re}\{s_k\}\right)\mathrm{Re}\{\boldsymbol{h}_k^H\}$, and, $\boldsymbol{\gamma}_{I,k} = \sqrt{2}\,\mathrm{sign}\left(\mathrm{Im}\{s_k\}\right)\mathrm{Im}\{\boldsymbol{h}_k^H\}$. Finally, the PSEP problem is cast by taking the logarithm of the SEP constraints which reads as

$$\begin{aligned} &\min_{\boldsymbol{\theta}_{\mathrm{r}}, P\in\mathbb{R}_+} \quad P \\ &\text{s.t.} \quad [\boldsymbol{\theta}_{\mathrm{r}}]_{2n-1} + j\,[\boldsymbol{\theta}_{\mathrm{r}}]_{2n} \in \mathcal{T}, \text{for } n \in \mathcal{N}, \\ &-\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right)\right) - \beta_k \le 0, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-20}$$

with $\beta_k = -\ln(1-\rho_k)$. As mentioned before the PSEP formulation is restricted to the cases where the users' data is either BPSK or QPSK.

5.4.2 Power Minimization under Union-Bound SEP constraints

For the cases in which $\alpha_s \notin \{2, 4\}$ this study considers substituting $\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right)$ in (5-11) by the UBSEP. As the UBSEP, denoted by $\mathrm{P}_{\mathrm{ub}}\left(\hat{s}_k|\boldsymbol{\theta}, P\right)$, is an upper-bound on the SEP [86, 26], substituting the constraint $\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right) \le \rho_k$ by $\mathrm{P}_{\mathrm{ub}}\left(\hat{s}_k|\boldsymbol{\theta}, P\right) \le \rho_k$ yields a restriction of the feasible set of the original problem, implying that the restricted problem's optimal transmit power is larger or equal to the original one. The union-bound inequality states that for any finite set of events, $\mathrm{P}\left(\bigcup_i A_i\right) \le \sum_i \mathrm{P}(A_i)$, with A_i representing an event. With this, $\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right)$ is bounded by

$$\begin{aligned} \mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right) &= \mathrm{P}\left(z_k \in \mathcal{Z}_1 \cup \mathcal{Z}_2|\boldsymbol{\theta}, P\right) \\ &\le \mathrm{P}\left(z_k \in \mathcal{Z}_1|\boldsymbol{\theta}, P\right) + \mathrm{P}\left(z_k \in \mathcal{Z}_2|\boldsymbol{\theta}, P\right) = \mathrm{P}_{\mathrm{ub}}\left(\hat{s}_k|\boldsymbol{\theta}, P\right), \end{aligned} \tag{5-21}$$

with $\mathcal{Z}_1$ and $\mathcal{Z}_2$ depicted in Fig. 5.2. The individual probabilities are computed based on the MDDTs, $d_{1,k}$ and $d_{2,k}$, as

$$\mathrm{P}\left(z_k \in \mathcal{Z}_1|\boldsymbol{\theta}, P\right) = \int_{d_{1,k}}^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{t^2}{\sigma_w^2}} dt = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}}{\sigma_w}\right)$$

$$\mathrm{P}\left(z_k \in \mathcal{Z}_2|\boldsymbol{\theta}, P\right) = \int_{d_{2,k}}^{\infty} \frac{1}{\sqrt{\pi\sigma_w^2}}\, e^{-\frac{t^2}{\sigma_w^2}} dt = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}}{\sigma_w}\right).$$

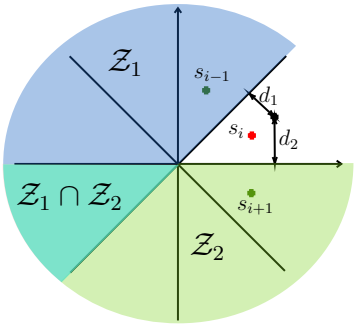

Figure 5.2: Representation of the union-bound

The MDDTs are computed by rotating the coordinate system such that the symbol of interest is placed on the real axis. This is done by multiplying both s_k and $\boldsymbol{h}_k^H\boldsymbol{\theta}$ by s_k^* which results in $s_k^* s_k = 1$ and $\omega_k = s_k^* \boldsymbol{h}_k^H \boldsymbol{\theta}$. With the rotated coordinate system the MDDTs are computed as

$$d_{1,k} = \sqrt{P}\left(\mathrm{Re}\{s_k^*\boldsymbol{h}_k^H\boldsymbol{\theta}\}\sin\phi - \mathrm{Im}\{s_k^*\boldsymbol{h}_k^H\boldsymbol{\theta}\}\cos\phi\right) \tag{5-22}$$

$$d_{2,k} = \sqrt{P}\left(\mathrm{Re}\{s_k^*\boldsymbol{h}_k^H\boldsymbol{\theta}\}\sin\phi + \mathrm{Im}\{s_k^*\boldsymbol{h}_k^H\boldsymbol{\theta}\}\cos\phi\right). \tag{5-23}$$

With this, the bound on $\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right)$ is given by

$$\begin{aligned}\mathrm{P}\left(\hat{s}_k \neq s_k|\boldsymbol{\theta}, P\right) &\leq \mathrm{P}_{\mathrm{ub}}\left(\hat{s}_k|\boldsymbol{\theta}, P\right) \\ &= \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}\left(\boldsymbol{\theta}, P\right)}{\sigma_w}\right) + \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}\left(\boldsymbol{\theta}, P\right)}{\sigma_w}\right).\end{aligned} \tag{5-24}$$

Substituting (5-24) in (5-11) yields the following problem

$$\begin{aligned}&\min_{\boldsymbol{\theta}, P\in\mathbb{R}_+} \quad P \\ &\text{s.t.} \quad [\boldsymbol{\theta}]_n \in \mathcal{T}, \ \text{for } n \in \mathcal{N}, \\ &\sum_{\xi=1}^{2}\frac{1}{2}\mathrm{erfc}\left(\frac{d_{\xi,k}\left(\boldsymbol{\theta}, P\right)}{\sigma_w}\right) \leq \rho_k, \ \text{for } k \in \mathcal{K}.\end{aligned} \tag{5-25}$$

Considering $\rho_k \leq 0.5$, for $k \in \mathcal{K}$, the UBSEP constraints are only achievable with all users having nonnegative MDDTs. With this, $d_{\xi,k}\left(\boldsymbol{\theta}, P\right) \geq 0$ for $\xi \in \{1, 2\}$ and $k \in \mathcal{K}$ is an implicit constraint of problem (5-25). The PUBSEP problem is finally cast writing (5-25) in real-valued formulation and explicitly

including the MDDTs restriction, which yields

$$\begin{aligned} &\min_{\boldsymbol{\theta}_{\mathrm{r}}, P\in\mathbb{R}_+} \quad P \\ &\text{s.t.} \quad [\boldsymbol{\theta}_{\mathrm{r}}]_{2n-1} + j\,[\boldsymbol{\theta}_{\mathrm{r}}]_{2n} \in \mathcal{T}, \text{ for } n \in \mathcal{N}, \quad \boldsymbol{U}\boldsymbol{\theta}_{\mathrm{r}} \preceq \mathbf{0}, \\ &\sum_{\xi=1}^{2} \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right) - \rho_k \leq 0, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-26}$$

where $\boldsymbol{u}_{1,k} = \left(\boldsymbol{\gamma}_{\mathrm{R},k}\sin(\phi) - \boldsymbol{\gamma}_{\mathrm{I},k}\cos(\phi)\right)^T$ and $\boldsymbol{u}_{2,k} = \left(\boldsymbol{\gamma}_{\mathrm{R},k}\sin(\phi) + \boldsymbol{\gamma}_{\mathrm{I},k}\cos(\phi)\right)^T$, with

$$\boldsymbol{\gamma}_{\mathrm{R},k} = [\mathrm{Re}\{[\boldsymbol{\zeta}_k]_1\}, -\mathrm{Im}\{[\boldsymbol{\zeta}_k]_1\}, \ldots, \mathrm{Re}\{[\boldsymbol{\zeta}_k]_N\}, -\mathrm{Im}\{[\boldsymbol{\zeta}_k]_N\}], \tag{5-27}$$

$$\boldsymbol{\gamma}_{\mathrm{I},k} = [\mathrm{Im}\{[\boldsymbol{\zeta}_k]_1\}, \mathrm{Re}\{[\boldsymbol{\zeta}_k]_1\}, \ldots, \mathrm{Im}\{[\boldsymbol{\zeta}_k]_N\}, \mathrm{Re}\{[\boldsymbol{\zeta}_k]_N\}], \tag{5-28}$$

$\boldsymbol{\zeta}_k = s_k^*\boldsymbol{h}_k^H$ and $\boldsymbol{U} = -[\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T]^T$, with $\boldsymbol{\beta}_1 = [\boldsymbol{u}_{1,1}, \ldots, \boldsymbol{u}_{1,K}]^T$ and $\boldsymbol{\beta}_2 = [\boldsymbol{u}_{2,1}, \ldots, \boldsymbol{u}_{2,K}]^T$. Note that, $d_{\xi,k} = \sqrt{P}\boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{r}}$, for $\xi \in \{1,2\}$ and $k \in \mathcal{K}$.

5.5 Proposed Branch-and-Bound Algorithm

This section proposes a B&B method that accepts as a solution any pair $(P_{\mathrm{out}}, \boldsymbol{\theta}_{\mathrm{out}}) \in \mathcal{H}$, where

$$\mathcal{H} = \left\{(P, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{T}^N \wedge (P \leq P_{\mathrm{b}} \vee 10\log_{10}(P/P_{\mathrm{opt}}) \leq \gamma)\right\}, \tag{5-29}$$

with P_{b} is the target power budget of the system, P_{opt} is the optimal transmit power and γ is the acceptable power increase factor. With this, the algorithm reduces the number of branches explored by allowing for suboptimal solutions that either attain the system's target power budget or are sufficiently close to the optimal solution such that further computation is considered unnecessary.

The first step of the proposed algorithm consists of the computation of an upper-bound solution and the evaluation of the stopping criteria. To this end, the original problems are relaxed substituting the discrete feasible set $\mathcal{T}^N$ by convex hull $\mathcal{P}$, described as $\boldsymbol{A}\boldsymbol{\theta}_{\mathrm{r}} \preceq \boldsymbol{b}$, where

$$\begin{aligned} &\boldsymbol{A} = \left[(\boldsymbol{I}_N \otimes \boldsymbol{\beta}_1)^T \quad (\boldsymbol{I}_N \otimes \boldsymbol{\beta}_2)^T \quad \ldots \quad (\boldsymbol{I}_N \otimes \boldsymbol{\beta}_{\alpha_\theta})^T\right]^T, \\ &\boldsymbol{\beta}_i = \left[\cos\left(\tfrac{2\pi i}{\alpha_\theta}\right) \quad -\sin\left(\tfrac{2\pi i}{\alpha_\theta}\right)\right], \quad i \in \{1, \ldots, \alpha_\theta\}, \quad \boldsymbol{b} = \cos\left(\frac{\pi}{\alpha_\theta}\right)\mathbf{1}_{N\alpha_\theta}, \end{aligned}$$

with $\mathbf{1}_{N\alpha_\theta}$ being the column vector with $N\alpha_\theta$ ones. By substituting $\mathcal{T}^N$ by its convex hull, the optimization problems described in (5-20) and (5-26) are cast

in the form of

$$\begin{aligned} \min_{\boldsymbol{\theta}_{\mathrm{r}}, P \in \mathbb{R}_+} \quad & P \\ \text{s.t.} \quad & f_k(\boldsymbol{\theta}_{\mathrm{r}}, P) \leq 0, \text{ for } k \in \mathcal{K}, \ \boldsymbol{G}\boldsymbol{\theta}_{\mathrm{r}} \preceq \boldsymbol{t}, \end{aligned} \tag{5-30}$$

where, for the PSEP case, $\boldsymbol{G} = \boldsymbol{A}$, $\boldsymbol{t} = \boldsymbol{b}$ and

$$f_k(\boldsymbol{\theta}_{\mathrm{r}}, P) = -\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right)\right) - \beta_k, \tag{5-31}$$

and for the PUBSEP case, $\boldsymbol{G} = [\boldsymbol{A}^T, \boldsymbol{U}^T]^T$ and $\boldsymbol{t} = [\boldsymbol{b}^T, \boldsymbol{0}^T]^T$, and

$$f_k(\boldsymbol{\theta}_{\mathrm{r}}, P) = \sum_{\xi=1}^{2} \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{r}}\right) - \rho_k. \tag{5-32}$$

By optimally solving (5-30) one gets $\boldsymbol{\theta}_{\mathrm{r,lb}}$, which implies $\boldsymbol{\theta}_{\mathrm{lb}} = C(\boldsymbol{\theta}_{\mathrm{r,lb}}) \in \mathcal{P}$, and its corresponding transmit power P_{lb}. Note that, $\boldsymbol{\theta}_{\mathrm{lb}} \in \mathcal{P}$ can also belong to $\mathcal{T}^N$ once $\mathcal{P} \cap \mathcal{T}^N \neq \emptyset$. If this is the case, $P_{\mathrm{opt}} = P_{\mathrm{lb}}$ and $\boldsymbol{\theta}_{\mathrm{opt}} = \boldsymbol{\theta}_{\mathrm{lb}}$, with P_{opt} and $\boldsymbol{\theta}_{\mathrm{opt}}$ being the optimal solutions of the original problem. Yet, if $\boldsymbol{\theta}_{\mathrm{lb}} \notin \mathcal{T}^N$ an upper bound solution on $\boldsymbol{\theta}_{\mathrm{opt}}$ is achieved by projecting to $\boldsymbol{\theta}_{\mathrm{lb}}$ to $\mathcal{T}^N$. The projection method considered is uniform quantization (UQ), denoted by the operator $Q(\cdot)$. By this approach, the p-th entry of $\boldsymbol{\theta}_{\mathrm{ub}} = Q(\boldsymbol{\theta}_{\mathrm{lb}})$, denoted as $\theta_{\mathrm{ub},p}$, is computed as $\theta_{\mathrm{ub},p} = \underset{i \in \{1,\ldots,\alpha_\theta\}}{\arg\min} |\theta_{\mathrm{lb},p} - \theta_i|$, where $\theta_{\mathrm{lb},p}$ denotes the p-th entry of $\boldsymbol{\theta}_{\mathrm{lb}}$ and θ_i is the i-th element of $\mathcal{T}$. Based on $\boldsymbol{\theta}_{\mathrm{r,ub}} = R(\boldsymbol{\theta}_{\mathrm{ub}})$, the corresponding transmit power is given by the solution of the following univariate upper bounding problem

$$\begin{aligned} P_{\mathrm{ub}} = \min_{P \in \mathbb{R}_+} \quad & P \\ \text{s.t.} \quad & f_k(\boldsymbol{\theta}_{\mathrm{r,ub}}, P) \leq 0, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-33}$$

If $P_{\mathrm{ub}} \leq P_{\mathrm{b}}$, the upper bound solution pair (P_{ub}, $\boldsymbol{\theta}_{\mathrm{ub}}$) attains the target power budget of the system. On the other hand, if $10\log_{10}\left({}^{P_{\mathrm{ub}}}/_{P_{\mathrm{lb}}}\right) \leq \gamma$ the solution pair is sufficiently close to the optimal solution such that further computation is considered unnecessary. With this, for both cases, the algorithm terminates with $\boldsymbol{\theta}_{\mathrm{out}} = \boldsymbol{\theta}_{\mathrm{ub}}$ and $P_{\mathrm{out}} = P_{\mathrm{ub}}$. In what follows, the PSEP and PUBSEP concepts are applied to (5-30) and (5-33). By substituting $\mathcal{T}^N$ by its convex

hull, the relaxed PSEP problem is cast as

$$\begin{aligned} &\min_{\boldsymbol{\theta}_\mathrm{r},P} \; P \\ &\text{s.t. } \boldsymbol{A}\boldsymbol{\theta}_\mathrm{r} \preceq \boldsymbol{b}, \; P \geq 0, \\ &-\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_\mathrm{r}\right)\right) \leq \beta_k, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-34}$$

An equivalent problem is formulated by defining the vector $\boldsymbol{x} = [\sqrt{P}\boldsymbol{\theta}_\mathrm{r}^T, \sqrt{P}]^T$ and applying the square root to the objective, which reads as

$$\begin{aligned} &\min_{\boldsymbol{x}} \; \boldsymbol{c}^T\boldsymbol{x} \\ &\text{s.t. } \boldsymbol{R}\boldsymbol{x} \preceq \mathbf{0}_{(N\alpha_\theta+1)}, \\ &-\sum_{\xi=1}^{2} \ln\left(\Phi\left(\boldsymbol{q}_{\xi,k}^T\boldsymbol{x}\right)\right) \leq \beta_k, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-35}$$

where $\boldsymbol{q}_{1,k} = (1/\sigma_w)[\boldsymbol{h}_{1,k}^T, 0]^T$, $\boldsymbol{q}_{2,k} = (1/\sigma_w)[\boldsymbol{h}_{2,k}^T, 0]^T$, $\boldsymbol{c} = [\mathbf{0}_{2N}^T, 1]^T$, $\boldsymbol{M} = [\boldsymbol{A}, \mathbf{0}_{N\alpha_\theta}]$, $\boldsymbol{R} = [(\boldsymbol{M} - \boldsymbol{b}\boldsymbol{c}^T)^T, -\boldsymbol{c}]^T$. As demonstrated in Appendix A.3 the SEP constraint functions from (5-35) are convex in $\boldsymbol{x}$. With this, (5-35) is a convex problem solvable with the barrier method [42, Section 11.3]. From the solution of (5-35), termed $\boldsymbol{x}_\mathrm{lb}$, one can extract P_lb and $\boldsymbol{\theta}_\mathrm{lb} = C(\boldsymbol{\theta}_\mathrm{r,lb})$. The upper bound solution $\boldsymbol{\theta}_\mathrm{ub} = Q(\boldsymbol{\theta}_\mathrm{lb})$ can then be converted to real-valued notation as $\boldsymbol{\theta}_\mathrm{r,ub} = R(\boldsymbol{\theta}_\mathrm{ub})$ and utilized for computing the upper bound transmit power P_ub with

$$\begin{aligned} P_\mathrm{ub} = \min_{P\in\mathbb{R}_+} \; & P \\ \text{s.t. } & -\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_\mathrm{r,ub}\right)\right) \leq \beta_k, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-36}$$

If either $\boldsymbol{h}_{1,k}^T\boldsymbol{\theta}_\mathrm{r,ub} \leq 0$ or $\boldsymbol{h}_{2,k}^T\boldsymbol{\theta}_\mathrm{r,ub} \leq 0$ for any $k \in \mathcal{K}$, problem (5-36) is infeasible for $\rho_k < 0.5$ and $P_\mathrm{ub} = \infty$. As in the PSEP case, the PUBSEP relaxed problem is formulated with $\boldsymbol{x} = [\sqrt{P}\boldsymbol{\theta}_\mathrm{r}^T, \sqrt{P}]^T$ as

$$\begin{aligned} &\min_{\boldsymbol{x}} \; \boldsymbol{c}^T\boldsymbol{x} \\ &\text{s.t. } \boldsymbol{D}\boldsymbol{x} \preceq \mathbf{0}_{(N\alpha_\theta+2K+1)}, \\ &\frac{1}{2}\sum_{\xi=1}^{2} \operatorname{erfc}\left(\boldsymbol{\nu}_{\xi,k}^T\boldsymbol{x}\right) \leq \rho_k, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-37}$$

where $\boldsymbol{\nu}_{1,k} = (1/\sigma_w)[\boldsymbol{u}_{1,k}^T, 0]^T$, $\boldsymbol{\nu}_{2,k} = (1/\sigma_w)[\boldsymbol{u}_{2,k}^T, 0]^T$, $\boldsymbol{D} = [\boldsymbol{R}^T, \boldsymbol{C}^T, -\boldsymbol{c}]^T$ and $\boldsymbol{C} = [\boldsymbol{U}, \mathbf{0}]$. As discussed in appendix A.4, the UBSEP constraints under the

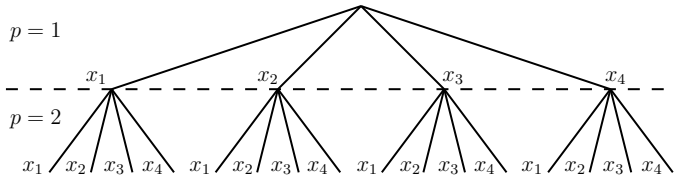

Figure 5.3: Tree representation of the set $\mathcal{T}^N$ for a system with $N = 2$ reflecting elements and QPSK precoding modulation ($\alpha_\theta = 4$)

condition $\boldsymbol{C}\boldsymbol{x} \preceq \boldsymbol{0}$ are convex, which implies that (5-37) is a convex problem solvable with the barrier method [42, Section 11.3]. From $\boldsymbol{x}_{\text{lb}}$ one can extract P_{lb} and $\boldsymbol{\theta}_{\text{lb}} = C(\boldsymbol{\theta}_{\text{r,lb}})$. The upper bound solution $\boldsymbol{\theta}_{\text{ub}} = Q(\boldsymbol{\theta}_{\text{lb}})$ can then be converted to real-valued notation as $\boldsymbol{\theta}_{\text{r,ub}} = R(\boldsymbol{\theta}_{\text{ub}})$ and utilized for computing the upper bound transmit power P_{ub} with

$$\begin{aligned} P_{\text{ub}} = \min_{P \in \mathbb{R}_+} \; & P \\ \text{s.t.} \; & \frac{1}{2}\sum_{\xi=1}^{2} \operatorname{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{u}_{\xi,k}^T \boldsymbol{\theta}_{\text{r,ub}}\right) \leq \rho_k, \; \text{for } k \in \mathcal{K}. \end{aligned} \tag{5-38}$$

As before, if $\boldsymbol{u}_{\xi,k}^T \boldsymbol{\theta}_{\text{r,ub}} \leq 0$, for any $\xi \in \{1, 2\}$ and $k \in \mathcal{K}$, problem (5-38) is infeasible for $\rho_k < 0.5$ and $P_{\text{ub}} = \infty$.

5.5.1 Branch-and-Bound Tree Search Stage

If neither stopping criteria are met at the initialization step, i.e., if $\boldsymbol{\theta}_{\text{lb}} \notin \mathcal{T}^N$, $P_{\text{ub}} > P_{\text{b}}$ and $10\log_{10}\left({}^{P_{\text{ub}}}/_{P_{\text{lb}}}\right) > \gamma$, the proposed B&B algorithm proceeds to the tree search stage where the tree represents the set $\mathcal{T}^N$. To this end, the smallest known upper bound is initialized as $\check{P} = P_{\text{ub}}$ and its corresponding reflection coefficient vector as $\check{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\text{ub}}$. The tree is constructed considering that the p-th reflection coefficient represents the p-th layer and each possible subvector $\boldsymbol{f} \in \mathcal{X}^p$ represents one branch. An example of a tree for a system with $N = 2$ reflecting elements and $\alpha_\theta = 4$, is shown in Fig. 5.3. The tree search process performs breadth-first search to find a vector $\boldsymbol{\theta}_{\text{out}} \in \mathcal{T}^N$ that attains the condition $10\log_{10}({}^{P(\boldsymbol{\theta}_{\text{out}})}/_{P_{\text{opt}}}) \leq \gamma$ with $P_{\text{opt}} = P(\boldsymbol{\theta}_{\text{opt}})$. During the search process, if an intermediate solution pair $(P_{\text{int}}, \boldsymbol{\theta}_{\text{int}})$ with $\boldsymbol{\theta}_{\text{int}} \in \mathcal{T}^N$ and $P_{\text{int}} \leq P_{\text{b}}$ is found, the process terminates with $\boldsymbol{\theta}_{\text{out}} = \boldsymbol{\theta}_{\text{int}}$ and $P_{\text{out}} = P_{\text{int}}$. The process starts at layer value $p = 1$ by fixing p entries of $\boldsymbol{\theta}$ such that the reflection coefficient vector becomes $\boldsymbol{\theta} = [\boldsymbol{f}_i^T, \boldsymbol{v}^T]^T$ with $\boldsymbol{f}_i \in \mathcal{T}^p$ and

$\boldsymbol{\theta}_{\mathrm{r}} = R(\boldsymbol{\theta}) = [\boldsymbol{f}_{\mathrm{r},i}^T, \boldsymbol{v}_{\mathrm{r}}^T]^T$. With this, a subproblem is assembled as

$$\begin{aligned} \{P_{\mathrm{opt}|\boldsymbol{f}_i}, \boldsymbol{v}_{\mathrm{r,opt}|\boldsymbol{f}_i}\} = \min_{\boldsymbol{v}_{\mathrm{r}}, P \in \mathbb{R}_+} \quad & P \\ \text{s.t.} \quad & [\boldsymbol{v}_{\mathrm{r}}]_{2n-1} + j\,[\boldsymbol{v}_{\mathrm{r}}]_{2n} \in \mathcal{T}, \text{ for } n \in \{1, \ldots, N-p\}, \\ & \boldsymbol{U}'\boldsymbol{v}_{\mathrm{r}} \preceq \boldsymbol{U}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r,i}}, \quad f_k(\boldsymbol{f}_{\mathrm{r},i}, \boldsymbol{v}_{\mathrm{r}}, P) \leq 0, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-39}$$

where $P_{\mathrm{opt}|\boldsymbol{f}_i}$ is the optimal transmit power for the fixed vector $\boldsymbol{f}_i$, $\boldsymbol{U}_{\mathrm{fixed}}$ and $\boldsymbol{U}'$ correspond to the first $2p$ and subsequent $2(N-p)$ columns of $\boldsymbol{U}$, and, the constraint $\boldsymbol{U}'\boldsymbol{v}_{\mathrm{r}} \preceq \boldsymbol{U}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r,i}}$ is only taken into account for the PUBSEP case. A lower bounding subproblem on $P_{\mathrm{opt}|\boldsymbol{f}_i}$ is obtained by relaxing $\mathcal{T}^{(N-p)}$ to its convex hull $\mathcal{J}$, which yields

$$\begin{aligned} \{P_{\mathrm{lb}|\boldsymbol{f}_i}, \boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i}\} = \min_{\boldsymbol{v}_{\mathrm{r}}, P \in \mathbb{R}_+} \quad & P \\ \text{s.t.} \quad & C(\boldsymbol{v}_{\mathrm{r}}) \in \mathcal{J}, \quad \boldsymbol{U}'\boldsymbol{v}_{\mathrm{r}} \preceq \boldsymbol{U}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r,i}}, \quad f_k(\boldsymbol{f}_{\mathrm{r},i}, \boldsymbol{v}_{\mathrm{r}}, P) \leq 0, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-40}$$

where the constraint $\boldsymbol{U}'\boldsymbol{v}_{\mathrm{r}} \preceq \boldsymbol{U}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r,i}}$ is only taken into account for the PUBSEP case. An upper bound on $P_{\mathrm{opt}|\boldsymbol{f}_i}$ can be computed by projecting the vector $\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i} = C(\boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i})$ to $\mathcal{T}^{(N-p)}$ resulting in $\boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i} = Q(\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i})$ and computing the corresponding transmit power $P_{\mathrm{ub}|\boldsymbol{f}_i}$, with $\boldsymbol{\theta}_{\mathrm{r,ub}|\boldsymbol{f}_i} = R(\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i}) = R([\boldsymbol{f}_i^T, \boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i}^T]^T)$ using (5-33). If the upper bound solution $P_{\mathrm{ub}|\boldsymbol{f}_i} \leq P_{\mathrm{b}}$, the solution pair, $(P_{\mathrm{ub}|\boldsymbol{f}_i}, \boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i})$, attains the low-resolution constraints and the target power budget of the system. With this, the algorithm terminates with $P_{\mathrm{out}} = P_{\mathrm{ub}|\boldsymbol{f}_i}$ and $\boldsymbol{\theta}_{\mathrm{out}} = \boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i}$. If, however, $P_{\mathrm{ub}|\boldsymbol{f}_i} > P_{\mathrm{b}}$, the algorithm proceeds by selecting the next branch $\boldsymbol{f}_{i+1}$ of the layer. After all possible valid branches in a given layer are evaluated, i.e., all valid $\boldsymbol{f}_i$ were fixed and its conditioned upper and lower bounds computed, the smallest known upper bound and its corresponding upper bound solution are updated as $\check{P} = \min_i(P_{\mathrm{ub}|\boldsymbol{f}_i}, \check{P})$ and $\check{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i}$. With $\check{P}$ the algorithm proceeds to the pruning step where the set of approved branches, $\mathcal{G}_p'$, in the current layer p is constructed. The proposed pruning step aims to exclude from the search set all reflection coefficient vectors $\boldsymbol{\theta}$ that belong to $\{\boldsymbol{\theta} : 10\log_{10}({}^{P(\boldsymbol{\theta})}/_{P_{\mathrm{opt}}}) > \gamma\}$. This can implicitly be done by approving branches $\boldsymbol{f}_i$ such that $P_{\mathrm{lb},\boldsymbol{f}_i} < (1-\delta)\check{P}$ with $\delta = 1 - 10^{-\frac{\gamma}{10}}$. With this, the set of approved branches for the given layer p is constructed as $\mathcal{G}_p' = \{\boldsymbol{f}_i | P_{\mathrm{lb},\boldsymbol{f}_i} < (1-\delta)\check{P}\}$. After pruning, the set of valid subvectors is updated and the algorithm repeats this process in the next layer. If the algorithm reaches the last layer, only a few valid candidate solutions are expected to remain. With this, they are all evaluated against $\check{P}$, and the optimal value is determined by the vector that yields the minimum value of P.

5.5.1.1 Subproblem Formulation

A general formulation of the lower bounding subproblems of the proposed B&B algorithm is given in (5-40). In what follows, the specific PSEP and PUBSEP subproblems are devised. Based on the PSEP problem from (5-34) one rewrites (5-40) as

$$\begin{aligned}\{P_{\mathrm{lb}|\boldsymbol{f}_i}, \boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i}\} = &\min_{\boldsymbol{v}_\mathrm{r}, P\in\mathbb{R}_+} \quad P \\ \text{s.t. } C(\boldsymbol{v}_\mathrm{r}) \in \mathcal{J}, \quad &-\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\left(\boldsymbol{\kappa}_{\xi,k}^T\boldsymbol{f}_{\mathrm{r},i} + \boldsymbol{\varrho}_{\xi,k}^T\boldsymbol{v}_\mathrm{r}\right)\right)\right) \le \beta_k, \quad \text{for } k\in\mathcal{K},\end{aligned} \tag{5-41}$$

where $\mathcal{J}$ is the convex hull of $\mathcal{T}^{(N-p)}$, $\boldsymbol{\kappa}_{\xi,k}$ and $\boldsymbol{\varrho}_{\xi,k}$ correspond to the first $2p$, and, of the subsequent $2(N-p)$ entries of $\boldsymbol{h}_{\xi,k}$, respectively. The lower bounding PSEP subproblem is cast by rewriting (5-41) using the auxiliary variable $\tilde{\boldsymbol{x}} = [\sqrt{P}\boldsymbol{v}_\mathrm{r}^T, \sqrt{P}]^T$, which yields

$$\begin{aligned}\tilde{\boldsymbol{x}}_{\boldsymbol{f}_i} = &\min_{\tilde{\boldsymbol{x}}} \ \boldsymbol{d}^T\tilde{\boldsymbol{x}} \\ \text{s.t. } &\left(\boldsymbol{R}' + \boldsymbol{R}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r},i}\boldsymbol{d}^T\right)\tilde{\boldsymbol{x}} \preceq \boldsymbol{0}, \\ &-\sum_{\xi=1}^{2}\ln\left(\Phi\left(\frac{\boldsymbol{\kappa}_{\xi,k}^T\boldsymbol{f}_{\mathrm{r},i}\boldsymbol{d}^T\tilde{\boldsymbol{x}} + \boldsymbol{\psi}_{\xi,k}^T\tilde{\boldsymbol{x}}}{\sigma_w}\right)\right) \le \beta_k, \ \text{for } k\in\mathcal{K},\end{aligned} \tag{5-42}$$

where $\boldsymbol{d} = [\boldsymbol{0}_{2(N-p)}^T, 1]^T$, $\boldsymbol{R}_{\mathrm{fixed}}$ is composed of the first $2p$ columns of $\boldsymbol{R}$, $\boldsymbol{R}'$ consists of the last $2(N-p)+1$ columns of $\boldsymbol{R}$ and $\boldsymbol{\psi}_{\xi,k} = [\boldsymbol{\varrho}_{\xi,k}^T, 0]^T$. Similar steps can be applied with the PUBSEP formulation such that the PUBSEP lower bounding subproblem is written as

$$\begin{aligned}\tilde{\boldsymbol{x}}_{\boldsymbol{f}_i} = &\min_{\tilde{\boldsymbol{x}}} \ \boldsymbol{d}^T\tilde{\boldsymbol{x}} \\ \text{s.t. } &\left(\boldsymbol{D}' + \boldsymbol{D}_{\mathrm{fixed}}\boldsymbol{f}_{\mathrm{r},i}\boldsymbol{d}^T\right)\tilde{\boldsymbol{x}} \preceq \boldsymbol{0}, \\ &\sum_{\xi=1}^{2}\frac{1}{2}\mathrm{erfc}\left(\frac{\boldsymbol{\eta}_{\xi,k}^T\boldsymbol{f}_{\mathrm{r},i}\boldsymbol{d}^T\tilde{\boldsymbol{x}} + \boldsymbol{\zeta}_{\xi,k}^T\tilde{\boldsymbol{x}}}{\sigma_w}\right) \le \rho_k, \ \text{for } k\in\mathcal{K}, \ \xi\in\{1,2\},\end{aligned} \tag{5-43}$$

where $\boldsymbol{\zeta}_{\xi,k} = [\boldsymbol{\lambda}_{\xi,k}^T, 0]^T$, $\boldsymbol{\eta}_{\xi,k}$ and $\boldsymbol{\lambda}_{\xi,k}$ correspond to the first $2p$ and of the subsequent $2(N-p)$ entries of $\boldsymbol{u}_{\xi,k}$, respectively, and, $\boldsymbol{D}_{\mathrm{fixed}}$ and $\boldsymbol{D}'$ are composed of the first $2p$ and of the last $2(N-p)+1$ columns of $\boldsymbol{D}$, respectively. Solving (5-42) and (5-43) yields $\tilde{\boldsymbol{x}}_{\boldsymbol{f}_i}$ from which $\boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i}$ and $P_{\mathrm{lb}|\boldsymbol{f}_i}$ are readily extracted. The steps of the proposed B&B method are detailed in Algorithm 8.

Algorithm 8 Proposed B&B Algorithm

Inputs: $\boldsymbol{h}_k$ for $k \in \mathcal{K}$, $\boldsymbol{s}$, $\mathcal{T}$, P_b, δ, *Criterion* **Output**: $\boldsymbol{\theta}_\mathrm{out}$, P_out

If *Criterion* = PSEP $\rightarrow$ Solve (5-35) to get $\boldsymbol{x}_\mathrm{lb} = [\sqrt{P_\mathrm{lb}}\boldsymbol{\theta}_\mathrm{r,lb}, \sqrt{P_\mathrm{lb}}]^T$
If *Criterion* = PUBSEP $\rightarrow$ Solve (5-37) to get $\boldsymbol{x}_\mathrm{lb} = [\sqrt{P_\mathrm{lb}}\boldsymbol{\theta}_\mathrm{r,lb}, \sqrt{P_\mathrm{lb}}]^T$
If $\boldsymbol{\theta}_\mathrm{lb} = C(\boldsymbol{\theta}_\mathrm{r,lb}) \in \mathcal{T}^N \rightarrow$ **terminate with** $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_\mathrm{lb}$ and $P_\mathrm{out} = P_\mathrm{lb}$
Compute $\boldsymbol{\theta}_\mathrm{ub} = Q(\boldsymbol{\theta}_\mathrm{lb})$ and get $\boldsymbol{\theta}_\mathrm{r,ub} = R(\boldsymbol{\theta}_\mathrm{ub})$
If *Criterion* = PSEP $\wedge\ \boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_\mathrm{r,ub} \geq 0\ \forall\ \xi \in \{1,2\}$ and $k \in \mathcal{K} \rightarrow$ Solve (5-36) to get P_ub
Else If *Criterion* = PUBSEP $\wedge\ \boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_\mathrm{r,ub} \geq 0\ \forall\ \xi \in \{1,2\}$ and $k \in \mathcal{K} \rightarrow$ Solve (5-38) to get P_ub
Else $\rightarrow$ Set $\boldsymbol{\theta}_\mathrm{ub} = \emptyset$ and $P_\mathrm{ub} = \infty$
If $P_\mathrm{ub} \leq P_\mathrm{b} \vee 10\log_{10}\left({}^{P_\mathrm{ub}}/_{P_\mathrm{lb}}\right) \leq \gamma \rightarrow$ **terminate with** $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_\mathrm{ub}$ and $P_\mathrm{out} = P_\mathrm{ub}$
Define $\check{\boldsymbol{\theta}} = \boldsymbol{\theta}_\mathrm{ub}$, $\check{P} = P_\mathrm{ub}$ and the first level ($p = 1$) of the tree by $\mathcal{G}_p := \mathcal{T}$
for $p = 1 : N-1$ **do**
  Partition $\mathcal{G}_p$ in $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{|\mathcal{G}_p|}$
  for $i = 1 : |\mathcal{G}_p|$ **do**
    If *Criterion* = PSEP
      Conditioned on $\boldsymbol{f}_{\mathrm{r},i} = R(\boldsymbol{f}_i)$ solve (5-42) to get $\tilde{\boldsymbol{x}}_{\boldsymbol{f}_i}$ and extract $P_{\mathrm{lb}|\boldsymbol{f}_i}$ and $\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i} = C(\boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i})$
      Map $\boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i} = Q(\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i})$ and get $\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i} = [\boldsymbol{f}_i^T, \boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i}^T]^T$
      If $\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i} \geq 0, \forall\xi \in \{1,2\}$ and $k \in \mathcal{K}$
        Based on $\boldsymbol{\theta}_{\mathrm{r,ub}|\boldsymbol{f}_i} = R(\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i})$ solve (5-36) to get $P_{\mathrm{ub}|\boldsymbol{f}_i}$
        If $P_{\mathrm{ub}|\boldsymbol{f}_i} \leq P_\mathrm{b} \rightarrow$ **terminate with** $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i}$ and $P_\mathrm{out} = P_{\mathrm{ub}|\boldsymbol{f}_i}$
    Else If *Criterion* = PUBSEP
      Conditioned on $\boldsymbol{f}_{\mathrm{r},i} = R(\boldsymbol{f}_i)$ solve (5-43) to get $\tilde{\boldsymbol{x}}_{\boldsymbol{f}_i}$ and extract $P_{\mathrm{lb}|\boldsymbol{f}_i}$ and $\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i} = C(\boldsymbol{v}_{\mathrm{r,lb}|\boldsymbol{f}_i})$
      Map $\boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i} = Q(\boldsymbol{v}_{\mathrm{lb}|\boldsymbol{f}_i})$ and get $\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i} = [\boldsymbol{f}_i^T, \boldsymbol{v}_{\mathrm{ub}|\boldsymbol{f}_i}^T]^T$
      If $\boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i} \geq 0, \forall\xi \in \{1,2\}$ and $k \in \mathcal{K}$
        Based on $\boldsymbol{\theta}_{\mathrm{r,ub}|\boldsymbol{f}_i} = R(\boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i})$ solve (5-38) to get $P_{\mathrm{ub}|\boldsymbol{f}_i}$
        If $P_{\mathrm{ub}|\boldsymbol{f}_i} \leq P_\mathrm{b} \rightarrow$ **terminate with** $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_{\mathrm{ub}|\boldsymbol{f}_i}$ and $P_\mathrm{out} = P_{\mathrm{ub}|\boldsymbol{f}_i}$
    end If
  end for
  Update $\check{P} = \min(\check{P}, P_{\mathrm{ub}|\boldsymbol{f}_i})$ and update $\check{\boldsymbol{\theta}}$ accordingly
  Based on $\check{P}$ and on the lower bounds build the set $\mathcal{G}'_p := \{\boldsymbol{\theta}_{\mathrm{lb}|\boldsymbol{f}_i} |\ P_{\mathrm{lb}|\boldsymbol{f}_i} < (1-\delta)\check{P},\ i = 1, \ldots, |\mathcal{G}_p|\}$
  Define the set for the next level in the tree: $\mathcal{G}_{p+1} := \mathcal{G}'_p \times \mathcal{T}$
end for
for $i = 1 : |\mathcal{G}_N|$ **do**
  If *Criterion* = PSEP $\wedge\ \boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_i \geq 0 \rightarrow$ Solve (5-36) with $\boldsymbol{\theta}_\mathrm{r,ub} = R(\boldsymbol{\theta}_i)$ and get P_i
  If *Criterion* = PUBSEP $\wedge\ \boldsymbol{u}_{\xi,k}^T\boldsymbol{\theta}_i \geq 0 \rightarrow$ Solve (5-38) with $\boldsymbol{\theta}_\mathrm{r,ub} = R(\boldsymbol{\theta}_i)$ and get P_i
  If $P_i \leq P_\mathrm{b} \rightarrow$ **terminate with** $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_i$ and $P_\mathrm{out} = P_i$
end for
The transmit power reads as $P_\mathrm{out} = \min_{i\in\{1,\ldots,|\mathcal{G}_N|\}}(\check{P}, P_i)$ and the reflection coefficients are given by $\boldsymbol{\theta}_\mathrm{out} = \boldsymbol{\theta}_i$

5.5.2 On the Computational Complexity of the Algorithm

The initialization problems, described in (5-35) and (5-37), the upper bounding problems from (5-36) and (5-38) and lower bounding subproblems, described in (5-42) and (5-43), are convex with twice continuously differentiable real-valued functions. With this, according to [42, Chapter 11] they are solvable with the barrier method. The UBCO of the barrier method can be summarized

as $\mathcal{O}(\sqrt{\varphi}q^3)$, [26],[42, Section 11.5.6], with φ being the number of inequality constraints and q being the number of optimization variables. By substituting the values of φ and q for the different problems one reaches the conclusion that solving (5-35) and (5-37) yields UBCOs of $\mathcal{O}(N^{3.5} + N^3\sqrt{K})$, (5-36) and (5-38) yields UBCOs of $\mathcal{O}(\sqrt{K})$, and (5-42) and (5-43) yields UBCOs of $\mathcal{O}(N^{3.5} + N^3\sqrt{K})$. For executing Algorithm 8 it is necessary to solve the initialization problem once and the upper bounding problems and lower bounding subproblems, J and B times, respectively. Since by the design of the algorithm $J \leq \alpha_\theta B$ the UBCO of the proposed B&B algorithms is given by $\mathcal{O}(B(N^{3.5} + N^3\sqrt{K}))$.

5.6 Problem Formulation for High-Resolution RIS

To derive a low-complexity approach for the case of high-resolution reflecting elements this section considers that, for a sufficiently large α_θ, the discrete set $\mathcal{T}$ can be well approximated by $\mathcal{C} = \{\theta : |\theta|^2 = 1\}$. With this, the PSEP and PUBSEP problems are reformulated based on the approximation of the original discrete set $\mathcal{T}$ by $\mathcal{C}$. Substituting $\mathcal{T}$ by $\mathcal{C}$ for the PSEP problem yields the following optimization problem

$$\begin{aligned} &\min_{\boldsymbol{\theta}, P\in\mathbb{R}_+} \quad P \\ &\text{s.t. } \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \text{ for } n \in \mathcal{N}, \\ &-\ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{r},k}(\boldsymbol{\theta})\right)\right) - \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{v}_{\mathrm{i},k}(\boldsymbol{\theta})\right)\right) \leq \beta_k, \text{ for } k \in \mathcal{K}. \end{aligned} \tag{5-44}$$

The optimization problem of power minimization for high-resolution RIS under SEP constraints (PHSEP) is cast by rewriting (5-44) with real-valued variables which yield

$$\begin{aligned} &\min_{\boldsymbol{\Theta}\in\mathcal{M}, P\in\mathbb{R}_+} \quad P \\ &-\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\ \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right)\right)\right) \leq \beta_k, \text{ for } k \in \mathcal{K}, \end{aligned} \tag{5-45}$$

where $\boldsymbol{b}_{\mathrm{R},k} = \sqrt{2}\ \mathrm{sign}(\mathrm{Re}\{s_k\})\boldsymbol{h}_k^H$, $\boldsymbol{b}_{\mathrm{I},k} = \sqrt{2}\ \mathrm{sign}(\mathrm{Im}\{s_k\})\boldsymbol{h}_k^H$,

$$\begin{aligned} &\mathcal{M} = \{\boldsymbol{\Theta} \in \mathbb{R}^{2\times N} : [\boldsymbol{\Theta}^T\boldsymbol{\Theta}]_{(n,n)} = 1, \text{ for } n \in \mathcal{N}\}, \\ &\boldsymbol{\Theta} = \begin{bmatrix} (\mathrm{Re}\{\boldsymbol{\theta}\})^T \\ (\mathrm{Im}\{\boldsymbol{\theta}\})^T \end{bmatrix}, \quad \boldsymbol{H}_{1,k} = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{b}_{\mathrm{R},k}^T\} \\ -\mathrm{Im}\{\boldsymbol{b}_{\mathrm{R},k}^T\} \end{bmatrix}^T, \quad \boldsymbol{H}_{2,k} = \begin{bmatrix} \mathrm{Im}\{\boldsymbol{b}_{\mathrm{I},k}^T\} \\ \mathrm{Re}\{\boldsymbol{b}_{\mathrm{I},k}^T\} \end{bmatrix}^T. \end{aligned} \tag{5-46}$$

Similarly, substituting $\mathcal{T}$ by $\mathcal{C}$ with PUBSEP formulation yields the following problem

$$\begin{aligned} &\min_{\boldsymbol{\theta},P} \; P \\ &\text{s.t.} \;\; |[\boldsymbol{\theta}]_n|^2 = 1, \; \text{for } n \in \mathcal{N}, \quad P \geq 0, \\ &\sum_{\xi=1}^{2} \frac{1}{2}\text{erfc}\left(\frac{d_{\xi,k}\left(\boldsymbol{\theta}, P\right)}{\sigma_w}\right) \leq p_k, \; \text{for } k \in \mathcal{K}. \end{aligned} \tag{5-47}$$

The problem for power minimization for high-resolution RIS under UBSEP constraints (PHUBSEP) is cast by rewriting (5-47) with real-valued variables which yields

$$\begin{aligned} &\min_{\boldsymbol{\Theta}\in\mathcal{M},P\in\mathbb{R}_+} \; P \\ &\sum_{\xi=1}^{2} \frac{1}{2}\text{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\, \text{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right)\right) \leq \rho_k, \; \text{for } k \in \mathcal{K}. \end{aligned} \tag{5-48}$$

where $\boldsymbol{a}_k = s_k^* \boldsymbol{h}_k^H$, and,

$$\begin{aligned} \boldsymbol{U}_{1,k} &= \begin{bmatrix} \text{Re}\{\boldsymbol{a}_k^T\}\sin(\phi) - \text{Im}\{\boldsymbol{a}_k^T\}\cos(\phi) \\ -\text{Re}\{\boldsymbol{a}_k^T\}\cos(\phi) - \text{Im}\{\boldsymbol{a}_k^T\}\sin(\phi) \end{bmatrix}^T, \\ \boldsymbol{U}_{2,k} &= \begin{bmatrix} \text{Re}\{\boldsymbol{a}_k^T\}\sin(\phi) + \text{Im}\{\boldsymbol{a}_k^T\}\cos(\phi) \\ \text{Re}\{\boldsymbol{a}_k^T\}\cos(\phi) - \text{Im}\{\boldsymbol{a}_k^T\}\sin(\phi) \end{bmatrix}^T. \end{aligned} \tag{5-49}$$

As demonstrated in Appendix A.5, the SEP constraint functions are convex and, although the UBSEP functions are not geodesically convex in $\mathcal{M}$, one can restrict the feasible set such that UBSEP functions are geodesically convex. Yet, due to the set $\mathcal{M}$ not being geodesically convex [87, section 2.3], the optimization problems from (5-45) and (5-48) are not geodesically convex, implying that the application of descent methods only guarantees local optimality.

5.7 Local Optimum via the Proposed Bisection Method

A locally optimal solution for the proposed high-resolution problems is computed via the proposed bisection method. The method is initialized with P_- as a lower bound on P_{lopt} and P_+ as an upper bound on P_{lopt}. The variable P is fixed as $P_0 = {}^{(P_+ + P_-)}/_2$ and the remaining problem's feasibility is evaluated. If feasible, P_+ is updated as P_0, otherwise, P_- is updated as P_0. This is done recursively until the power difference between two consecutive iterations is

Algorithm 9 Proposed Bisection Method

Inputs: $P_+ \geq P_{\text{opt}}$, $P_- \leq P_{\text{opt}}$, $f_a < 0$, $i_{\max} > 0$, $\epsilon_{\text{tol}} > 0$ **Output**: P_{opt}, $\boldsymbol{\theta}_{\text{opt}}$
Define $i = 0$, $P_a = P_0$
While $(P_a - P_0 \leq \epsilon_{\text{tol}} \ \vee \ i \leq i_{\max}) \ \wedge \ f_a < 0$
  Solve (5-56) with RCG [88] considering $P = P_0$ and get $\boldsymbol{\Theta}_{\text{opt}}$
  Compute $f_a = f(\boldsymbol{\Theta}_{\text{opt}})$ with (5-54)
  If $f_a \leq 0 \rightarrow$ Update $P_+ = P_0$
  Else $\rightarrow$ Update $P_- = P_0$
  Update $P_a = P_0$, $P_0 = {}^{(P_+ + P_-)}/_2$ and $i = i + 1$
Update $P_{\text{opt}} = P_0$ and $\boldsymbol{\theta}_{\text{opt}} = [\boldsymbol{\Theta}_{\text{opt}}]_{(1,:)} + j\,[\boldsymbol{\Theta}_{\text{opt}}]_{(2,:)}$

below an optimality tolerance ϵ_{tol}. For a given P, the general optimization problem is written as

$$\begin{aligned} &\text{find}_{\boldsymbol{\Theta}\in\mathcal{M}} \ \ \boldsymbol{\Theta} \\ &\text{s.t. } f_k(\boldsymbol{\Theta}) \leq 0, \ \text{for } k \in \mathcal{K}, \end{aligned} \tag{5-50}$$

where, for the PHSEP formulation

$$f_k(\boldsymbol{\Theta}) = -\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\ \text{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right)\right)\right) - \beta_k, \tag{5-51}$$

and for the PHUBSEP formulation

$$f_k(\boldsymbol{\Theta}) = \sum_{\xi=1}^{2} \frac{1}{2}\text{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\ \text{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right)\right) - \rho_k. \tag{5-52}$$

The strategy for solving (5-50) consists of minimizing the maximum constraint function and evaluating if the locally optimal solution attains it. This yields the following optimization problem

$$\boldsymbol{\Theta}_{\text{lopt}} = \min_{\boldsymbol{\Theta}\in\boldsymbol{\mathcal{M}}} \ \max_{k\in\mathcal{K}} f_k(\boldsymbol{\Theta}). \tag{5-53}$$

Based on $\boldsymbol{\Theta}_{\text{lopt}}$ the feasibility of (5-53) is evaluated by checking $f(\boldsymbol{\Theta}_{\text{lopt}}) \leq 0$, where

$$f(\boldsymbol{\Theta}) = \max_{k\in\mathcal{K}} \ f_k(\boldsymbol{\Theta}). \tag{5-54}$$

If the condition holds the problem is feasible and $\boldsymbol{\Theta}_{\text{lopt}}$ is a solution of (5-50). Otherwise, at least one constraint cannot be fulfilled with the given transmit power P, implying that (5-50) is infeasible. The steps of the BM are further detailed in algorithm 9.

5.7.1 Evaluating Feasibility via Riemannian Conjugate Gradient

The utilization of algorithm 9 implies a method for locally solving the unconstrained problem in (5-53). This is done with the RCG algorithm [89, Section 3.1], designed for solving unconstrained minimization problems in Riemannian manifolds. Since the RCG approach requires a twice continuously differentiable objective, $f(\boldsymbol{\Theta})$ is substituted by its softmax approximation computed with the log-sum-exp function $\mathrm{LSE}(\boldsymbol{x}) = \ln\left(\sum_i e^{x_i}\right)$, which yields

$$f_0(\boldsymbol{\Theta}) = \ln\left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})}\right). \tag{5-55}$$

The precision of the approximation obeys the following bound

$$\max_{\{i=1,\ldots,n\}} x_i \leq \mathrm{LSE}(x_1,\ldots,x_n) \leq \max_{\{i=1,\ldots,n\}} x_i + \log(n).$$

Note that, $\mathrm{LSE}(\boldsymbol{x})$ is a non-decreasing function, which implies that for the regions where $f(\boldsymbol{\Theta})$ is convex the convexity is preserved. With this, (5-53) is rewritten as

$$\min_{\boldsymbol{\Theta}\in\boldsymbol{\mathcal{M}}} \ \ln\left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})}\right). \tag{5-56}$$

A locally optimal solution to the optimization problem described in (5-56) is computed via the RCG algorithm. To this end, however, the Euclidean gradient of $f_0(\boldsymbol{\Theta})$ and a strictly feasible starting point $\boldsymbol{\Theta}_0$ are required. In what follows these values are computed for both design criteria.

5.7.1.1 PHSEP RCG

With the PHSEP formulation the Euclidean gradient $\nabla f_0(\boldsymbol{\Theta})$ is given by

$$\nabla f_0(\boldsymbol{\Theta}) = \left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})}\nabla f_k\right)\left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})}\right)^{-1}, \tag{5-57}$$

$$\nabla f_k(\boldsymbol{\Theta}) = -\sqrt{\frac{P}{2\pi\sigma_w^2}}\sum_{\xi=1}^{2}\frac{e^{-\frac{P}{2\sigma_w^2}\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k})^2}\boldsymbol{H}_{\xi,k}^T}{\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\ \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right)\right)}. \tag{5-58}$$

As demonstrated in appendix A.5, the SEP functions $f_k(\boldsymbol{\Theta})$ are matrix convex in $\mathbb{R}^{2\times N}$, which implies that $f_0(\boldsymbol{\Theta})$ is also matrix convex. With this, initializing the RCG algorithm with any $\boldsymbol{\Theta}_0 \in \mathcal{M}$ guarantees convergence to a locally optimal solution.

5.7.1.2 PHUBSEP RCG

For PHUBSEP the values of the Euclidean gradient read as

$$\nabla f_0(\boldsymbol{\Theta}) = \left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})} \nabla f_k\right) \left(\sum_{k=1}^{K} e^{f_k(\boldsymbol{\Theta})}\right)^{-1}, \tag{5-59}$$

$$\nabla f_k(\boldsymbol{\Theta}) = -\sqrt{\frac{P}{\pi\sigma_w^2}} \sum_{\xi=1}^{2} e^{-\frac{P}{\sigma_w^2}\cdot \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right)^2} \boldsymbol{U}_{\xi,k}^T. \tag{5-60}$$

As demonstrated in appendix A.5, the functions $f_k(\boldsymbol{\Theta})$ are matrix convex for $\boldsymbol{\Theta} \in \mathcal{Y}$ with $\mathcal{Y} = \{\boldsymbol{\Theta} : \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right) \geq 0, \text{ for } k \in \mathcal{K}, \ \xi \in \{1,2\}\}$, which implies convexity of $f_0(\boldsymbol{\Theta})$ for $\boldsymbol{\Theta} \in \mathcal{Y}$. Due to $f_0(\boldsymbol{\Theta})$ being convex for $\boldsymbol{\Theta} \in \mathcal{Y}$, if the optimal solution of (5-56), termed $\boldsymbol{\Theta}_{\mathrm{lopt}}$, belongs to $\mathcal{Y}$, initializing the RCG method with any value of $\boldsymbol{\Theta}_0 \in \mathcal{Y}$ supports finding a locally optimal solution. This is the case since $f_0(\boldsymbol{\Theta})$ grows for a decrease in $\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right)$ and thus by initializing the RCG algorithm with $\boldsymbol{\Theta}_0 \in \mathcal{Y}$ it will takes steps to stay on $\mathcal{Y}$. On the other hand, if $\boldsymbol{\Theta}_{\mathrm{lopt}} \notin \mathcal{Y}$, which corresponds to a noiseless received signal outside the correct decision region, the value of at least one constraint function $f_k(\boldsymbol{\Theta})$ is given by $f_k(\boldsymbol{\Theta}) \geq 0.5 - \rho_k$. Since $f(\boldsymbol{\Theta}) = \max_k f_k(\boldsymbol{\Theta})$, this implies that in this case (5-48) is infeasible for $\rho_k < 0.5 \ \forall k \in \mathcal{K}$. For this case, initializing the RCG algorithm for solving (5-56) with any starting point, including $\boldsymbol{\Theta}_0 \in \mathcal{Y}$, will yield an output $\boldsymbol{\Theta}_{\mathrm{out}}$ such that $f(\boldsymbol{\Theta}_{\mathrm{out}}) > 0$. With this, the optimization problem for computing the initial point $\boldsymbol{\Theta}_0 \in \mathcal{Y}$ can be cast as

$$\max_{\boldsymbol{\Theta}\in\boldsymbol{\mathcal{M}}} \min_{k\in\mathcal{K},\xi\in\{1,2\}} \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right). \tag{5-61}$$

To solve (5-61) via the RCG algorithm the log-sum-exp function is applied which yields

$$\boldsymbol{\Theta}_0 = \min_{\boldsymbol{\Theta}\in\boldsymbol{\mathcal{M}}} \ v_0(\boldsymbol{\Theta}), \tag{5-62}$$

with $v_0(\boldsymbol{\Theta}) = \ln\left(\sum_{k=1}^{K} e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{1,k}\right)} + e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{2,k}\right)}\right)$, and, $\nabla v_0(\boldsymbol{\Theta}) = -\frac{\sum_{k=1}^{K} e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{1,k}\right)} \boldsymbol{U}_{1,k}^T + e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{2,k}\right)} \boldsymbol{U}_{2,k}^T}{\sum_{k=1}^{K} e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{1,k}\right)} + e^{-\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{2,k}\right)}}$. Problem (5-62) is solved via the utilization of the RCG algorithm with any starting point $\boldsymbol{\Theta} \in \mathcal{M}$. The details of the RCG implementation are given in [88].

5.7.2 Final Considerations

The complexity of the RCG algorithm dominates the complexity of the proposed PHSEP and PHUBSEP methods. Considering that the computational cost of computing the Euclidean gradient is $\mathcal{O}(N^2)$ and that the number of iterations required for convergence scales with $\sqrt{N}$, one can summarize the UBCO of the RCG algorithm as $\mathcal{O}(N^{2.5})$. Since the number of times that RCG is required to run mainly depends on initialization of P_+ and P_- and does not grow with the size of the system the overall UBCO of the proposed algorithm is in the order of $\mathcal{O}(N^{2.5})$.

The approximation of $\mathcal{T}^N$ by $\mathcal{C}$ implies that the RIS has sufficiently high resolution such that formulating the problem with a continuous set is beneficial for achieving a reasonable solution. Yet, in practice, the reflection coefficients are constraints to a discrete set since although large the resolution is always finite. By applying algorithm 9, one computes $P_{\text{lopt}} \in \mathbb{R}_+$ and $\boldsymbol{\theta}_{\text{lopt}} \in \mathcal{C}^N$. Note that, since $\boldsymbol{\theta}_{\text{lopt}}$ does not necessarily belong to $\mathcal{T}^N$ a projection step is necessary. This is done, similarly as in section 5.5, via UQ such that $\boldsymbol{\theta} = Q(\boldsymbol{\theta}_{\text{lopt}})$. Accordingly, the transmit power P is computed by solving (5-33) with the corresponding formulation and $\boldsymbol{\theta}_{\text{r,ub}} = R(\boldsymbol{\theta})$. After the mentioned steps $\boldsymbol{\theta} \in \mathcal{T}^N$ and $P \in \mathbb{R}_+$ can be utilized for transmission.

5.8 Numerical Results

This section evaluates the proposed algorithms against the state-of-the-art approach from [19] in terms of UBCO and average normalized transmit power (ANTP) defined as $P_n = 10\log_{10}(P/\sigma_w^2)$. For solving the optimization problem from [19] the bisection method from section 5.7 is considered. For the simulations, the channel coefficients are modeled by independent Rayleigh fading, and the noise variance is considered to be $\sigma_w^2 = 1$. To simplify the analysis all users are considered to have the same SEP requirement such that $\rho_k = 10^{-\tau}$, for $k \in \mathcal{K}$. A normalized target power budget $P_{\text{B}} = 10\log_{10}(P_{\text{b}}/\sigma_w^2)$ [dB] is considered for the plots, with P_{b} being the target power budget in linear scale utilized in the proposed B&B algorithms.

5.8.1 Performance Analysis versus SEP requirement

This section considers a scenario with $K = 2$ users, $N = 15$ reflecting elements, and QPSK data and transmit symbols, i.e., $\alpha_s = \alpha_\theta = 4$. For the experiments of Fig. 5.4 the proposed B&B approaches utilize $P_{\text{B}} = 2$ dB and

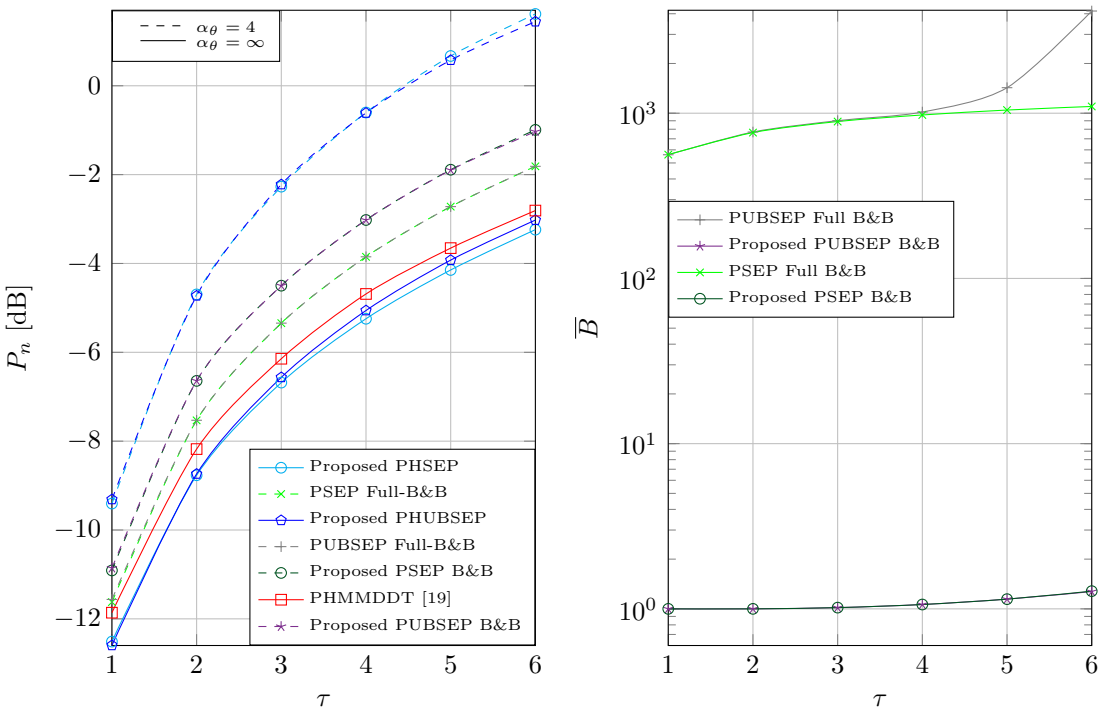

Figure 5.4: Considered scenario: $K = 2$ users, $N = 15$ reflecting elements, $\alpha_s = 4$ PSK users' data, $\alpha_\theta = 4$ available phase shifts, SEP requisites $\rho_k = 10^{-\tau}$ for $k \in \mathcal{K}$, target power budget $P_{\mathrm{B}} = 2$ dB and acceptable power increase factor $\gamma = 4$ dB. Average normalized transmit power $P_n \times \tau$ (left). Average number of optimization problems solved $\overline{B} \times \tau$ (right).

$\gamma = 4$ dB. The first experiment consists of the evaluation of the ANTP required for attaining the SEP requisites versus the SEP requirement parameter τ. The LHS of Fig. 5.4 shows that the unquantized PHSEP approach outperforms the unquantized PHUBSEP method in terms of ANTP for all values of τ. It also shows that the proposed unquantized PHSEP and PHUBSEP techniques require smaller ANTP for attaining the SEP requisites than the PMMDDT formulation from [19] for all values of τ. Note that, although not guaranteed due to the suboptimality of the BM utilized for solving the problems, this is expected since, as stated in section 5.4, the PUBSEP formulation is a restriction of the PSEP optimization problem, and, as shown in appendix B, the PHMMDDT formulation is a restriction of the PHUBSEP problem. Regarding the finite resolution methods, the LHS of Fig. 5.4 shows that the proposed PSEP and PUBSEP B&B methods and the PSEP and PUBSEP Full-B&B approaches yield a significant decrease in required ANTP compared to the high-resolution approaches after quantization. Moreover, it can be seen that the proposed PSEP and PUBSEP B&B and Full-B&B methods yield approximately 2 dB and 1.2 dB loss in relation to the infinite resolution approaches, respectively.

Table 5.1: UBCO of the Algorithms

Algorithm	UBCO
Proposed PHSEP	$\mathcal{O}(N^{2.5})$
Proposed PHUBSEP	$\mathcal{O}(N^{2.5})$
PHMMDDT [19]	$\mathcal{O}(N^{2.5})$
Proposed PSEP B&B	$\mathcal{O}\left(B(N^{3.5}+N^3\sqrt{K})\right)$
Proposed PUBSEP B&B	$\mathcal{O}\left(B(N^{3.5}+N^3\sqrt{K})\right)$
PSEP Full-B&B	$\mathcal{O}\left(B(N^{3.5}+N^3\sqrt{K})\right)$
PUBSEP Full-B&B	$\mathcal{O}\left(B(N^{3.5}+N^3\sqrt{K})\right)$

The UBCO analysis is done considering that the convex optimization-based approaches are solved with the barrier method. The UBCO of considered approaches is shown in Table 5.1, where B denotes the given number of sub-problems solved in the corresponding B&B algorithm. As shown in Table 5.1 the high-resolution approaches yield significantly smaller UBCO than the B&B approaches which justifies its utilization for scenarios where the resolution of the RIS elements is sufficiently high such that the decrease in ANTP performance is relatively small. These scenarios are explored in section 5.8.3. Since the UBCO of the B&B approaches depends on the value of B, for comparing their complexity an evaluation of B is necessary. This evaluation is done in the second experiment in terms of the average value of B, termed $\overline{B}$, and is shown in the RHS of Fig. 5.4. The RHS of Fig. 5.4 shows that the PSEP designs yield reduced $\overline{B}$ when compared with PUBSEP, with this it is concluded that PSEP is favorable in terms of UBCO when compared with PUBSEP. Moreover, the RHS of Fig. 5.4 shows a complexity reduction of at least factor 540 when utilizing the proposed B&B approaches when compared with its Full-B&B counterparts. Finally, a joint analysis of the plots in Fig. 5.4 summarizes the complexity-performance trade-off achieved when utilizing the $P_{\mathrm{B}} = 2$ dB and $\gamma = 4$ dB. Fig. 5.4 shows that utilizing the proposed B&B yields a power increase smaller than 1 dB and a UBCO decrease of at least factor 540 when compared with Full-B&B counterparts.

5.8.2 Performance-Complexity Trade-off Evaluation of the Proposed Branch-and-Bound Methods

The performance analysis of the proposed B&B approaches presented in Fig. 5.4 is computed considering the $P_{\mathrm{B}} = 2$ dB and $\gamma = 4$ dB, which corresponds to the specific complexity performance trade-off shown. Yet, by varying the values of P_{B} and γ different trade-offs are achievable. This section evaluates the performance of the proposed B&B approaches for the different values of the acceptable power increase γ and normalized target power budget P_{B}. For this section, the SEP requirements are set to $\rho_k = 10^{-4}$ for $k \in \mathcal{K}$.

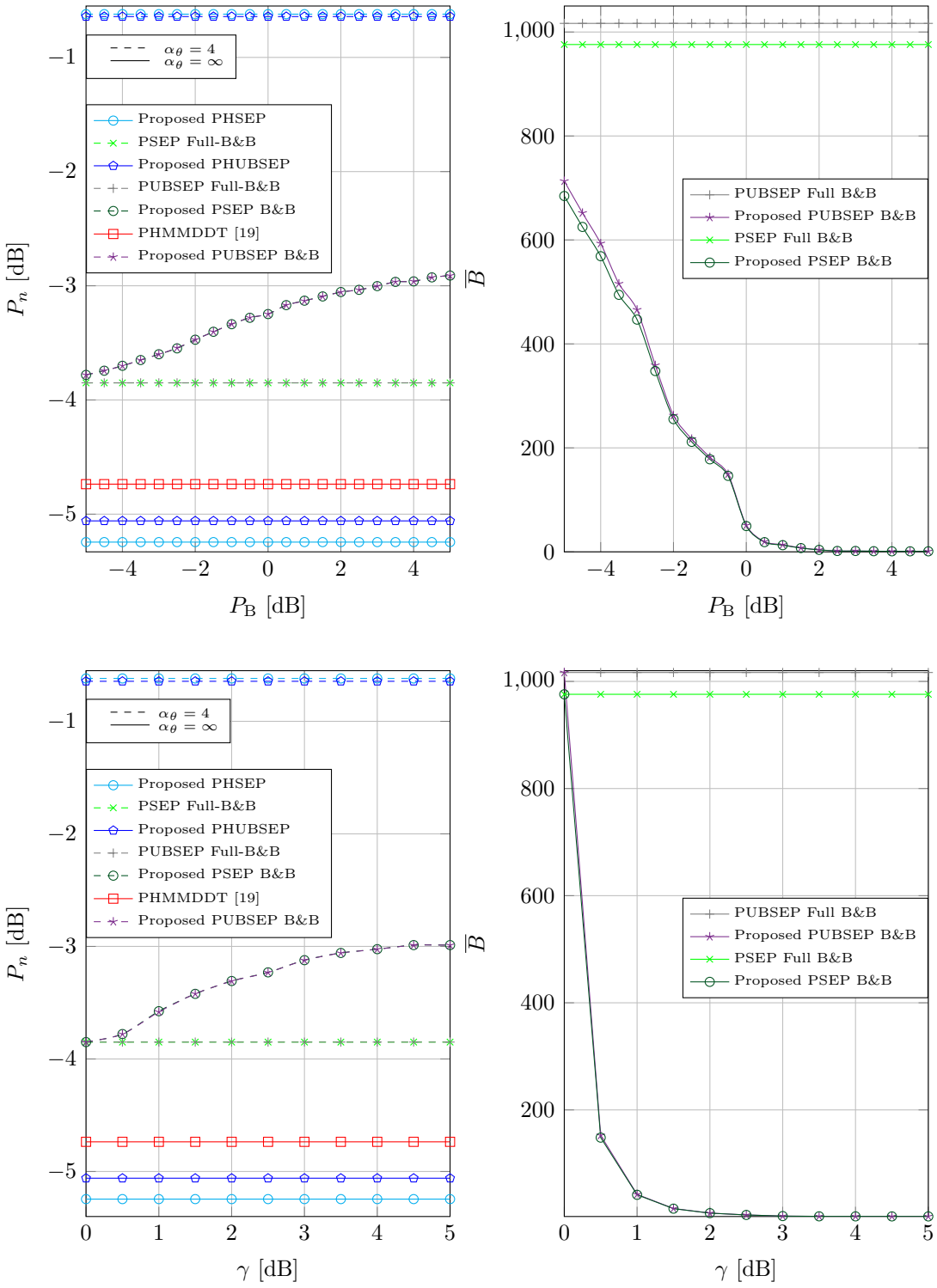

Figure 5.5: Considered scenario: $K = 2$ users, $N = 15$ reflecting elements, $\alpha_s = 4$ PSK users' data, $\alpha_\theta = 4$ available phase shifts, SEP requisites $\rho_k = 10^{-4}$ for $k \in \mathcal{K}$. Average normalized transmit power versus target power budget, P_n [dB] $\times$ P_{B} (upper left). Average number of optimization problems solved $\overline{B}$ $\times$ P_{B} (upper right), for acceptable power increase $\gamma = 0$ dB. P_n [dB] $\times$ γ (lower left), $\overline{B}$ $\times$ γ (lower right), for $P_{\mathrm{B}} = -\infty$ dB.

The first experiment, shown in the upper plots of Fig. 5.5, evaluates the impact of the target power budget P_{B} in the performance of the different approaches, for no acceptable power increase, meaning $\gamma = 0$ dB. As shown in the upper LHS of Fig. 5.5, for the low-end values of P_{B} the ANTP performance of the proposed PSEP and PUBSEP B&B methods is similar to its Full-B&B counterparts. This is the case since for extremely restrictive power budget scenarios where attaining $P_{\mathrm{out}} \leq P_{\mathrm{B}}$ is not possible (or is possible only with the optimal solution) the proposed B&B approach, when operating with $\gamma = 0$ dB, yields the optimal solution with Full-B&B complexity. Yet, as P_{B} increases, the target power budget of the system starts to be achieved with suboptimal reduced complexity reflection coefficients, and thus, the ANTP performance of the proposed B&B approaches starts to deviate from the performance of the Full-B&B methods. The upper RHS of Fig. 5.5 further highlights this behavior where it can be seen that, as the target power budget of the system increases, the average number of evaluated bounds explored by the proposed PSEP and PUBSEP B&B approaches decrease until, for $P_{\mathrm{B}} = 3$ dB, $\overline{B} \approx 1$ which implies an UBCO of $\mathcal{O}(N^{3.5} + N^3\sqrt{K})$.

In the second experiment, shown in the lower plots of Fig. 5.5, the performance of the proposed B&B is evaluated for different values of γ considering $P_{\mathrm{B}} = -\infty$ dB. The lower plots of Fig. 5.5 show, as expected, that for $\gamma = 0$ dB the proposed PSEP and PUBSEP B&B approaches yield the same ANTP and average number of subproblems solved as its Full-B&B counterparts. The lower LHS of Fig. 5.5 shows that as γ increases, i.e., as the system accepts a larger power increase, the ANTP of the proposed B&B approaches grows. Note, however, that the real increase in ANTP is always significantly smaller than the acceptable power increase being on average approximately 17% of the value of γ. The lower RHS of Fig. 5.5 on the other hand shows a rapid decrease in $\overline{B}$ with an increase in γ. This underlines the idea that, although computing the optimal solution requires exploration of a large number of branches when allowing for small ANTP compromises one can achieve significant reduction such that the resulting UBCO is $\mathcal{O}(N^{3.5}+N^3\sqrt{K})$. Finally, a joint analysis of the plots of Fig. 5.5 illustrates that a decrease in UBCO by a factor greater than 100 can be achieved with an increase of ANTP of less than 0.45 dB. With this, it can be stated that significant complexity reduction can be achieved with negligible ANTP compromise.

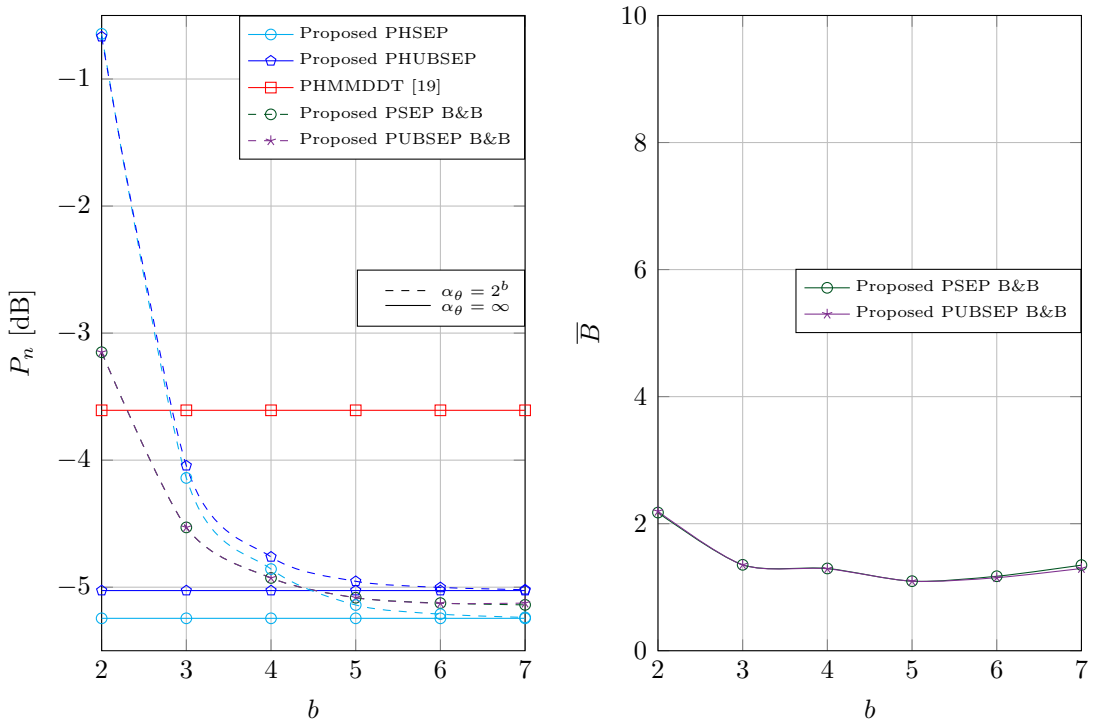

Figure 5.6: Considered scenario: $K = 2$ users, $N = 15$ reflecting elements, $\alpha_s = 4$ PSK users' data, SEP requites $\rho_k = 10^{-4}$ for $k \in \mathcal{K}$, target power budget $P_{\mathrm{B}} = 0$ dB and acceptable power increase factor $\gamma = 1$ dB. Average normalized transmit power versus number of bits $P_n \times b$ (left). Average number of optimization problems solved $\overline{B} \times b$ (right).

5.8.3 Performance Analysis versus Resolution

This section evaluates the effects of the resolution of the RIS elements, measured in bits as $b = \log_2(\alpha_\theta)$, on the performance of the proposed methods. The considered scenario consists of a system with $K = 2$ users, $N = 15$ reflecting elements, QPSK data symbols, $\rho_k = 10^{-4}$ for $k \in \mathcal{K}$, target power budget $P_{\mathrm{B}} = 0$ dB and $\gamma = 1$ dB. The experiment consists of an evaluation of P_n required for attaining the SEP constraints for different values of b. The LHS of Fig. 5.6 shows that, for $b = 3$, all proposed methods outperform the infinite resolution PHMMDDT baseline from [19]. Moreover, as expected, the LHS of Fig. 5.6 shows that as the number of resolution bits increases the proposed PHSEP and PHUBSEP quantized methods approach the ANTP performance of their infinite resolution counterparts. Considering that the proposed PHSEP and PHUBSEP quantized techniques yield UBCO of $\mathcal{O}(N^{2.5})$ one can understand that the proposed PHSEP and PHUBSEP methods yield reduced ANTP with low complexity which highlights the efficiency of the proposed methods. Regarding the proposed PSEP and PUBSEP B&B methods

the LHS of Fig. 5.6 shows that for $b \leq 4$ the proposed B&B techniques yield the smallest ANTP of the quantized approaches. Yet, for $b > 4$, they are outperformed by the PHSEP approach. Although the PHSEP approach yields a suboptimal solution even for $\alpha_\theta = \infty$, this is expected since $P_{\mathrm{B}} = 0$ dB and $\gamma = 1$ dB are considered, and, with this, complexity reduction is achieved at the expense of ANTP performance. The RHS of Fig. 5.6 addresses the UBCO of the proposed B&B approaches. As can be seen, when using $P_{\mathrm{B}} = 0$ dB and $\gamma = 1$ dB the proposed B&B approaches yield no significant increase in complexity with $1 < \overline{B} < 2.5$ for all values of b.

5.8.4 Transmit Power Analysis for Large-Scale Systems

This section evaluates the ANTP performance of the considered approaches for different values of the SEP requirement parameter τ, in a massive MIMO system. Unlike Full-B&B approaches, which can yield prohibitive complexity, the proposed B&B method is suitable for large-scale MIMO. For this section, the RIS elements are considered to have 2-bit resolution, and the B&B parameters are set to $P_{\mathrm{B}} = -8$ dB and $\gamma = 1$ dB. The first experiment, shown in the LHS of Fig. 5.7, considers a MIMO scenario of $K = 10$ users, $N = 100$ reflecting elements, and QPSK users' data. As shown in Fig. 5.7 the proposed PSEP and PUBSEP B&B techniques outperform the methods PHSEP and PHUBSEP methods after quantization, which highlights the suitability of the proposed B&B technique for large-scale MIMO. Regarding the infinite resolution approaches, the LHS of Fig. 5.7 shows that the proposed PHSEP and PHUBSEP techniques yield smaller ANTP when compared with the PHM-MDDT technique. The second experiment, present in the RHS of Fig. 5.7, considers a MIMO scenario of $K = 5$ users, $N = 120$ reflecting elements, and 8-PSK users' data. As shown in the RHS of Fig. 5.7, the proposed PUBSEP B&B approach outperforms the quantized PHUBSEP method for all values of τ. Yet, due to the small number of resolution bits of the RIS' elements, the PUBSEP-B&B approach yields approximately a 4 dB increase in ANTP compared to the unquantized PHUBSEP method. Finally, regarding the complexity of the proposed B&B approaches, the ANTP results from the LHS and RHS of Fig. 5.7 are achieved with the maximum values of $\overline{B} = 3.37$ and $\overline{B} = 44$ for $\tau = 6$, respectively.

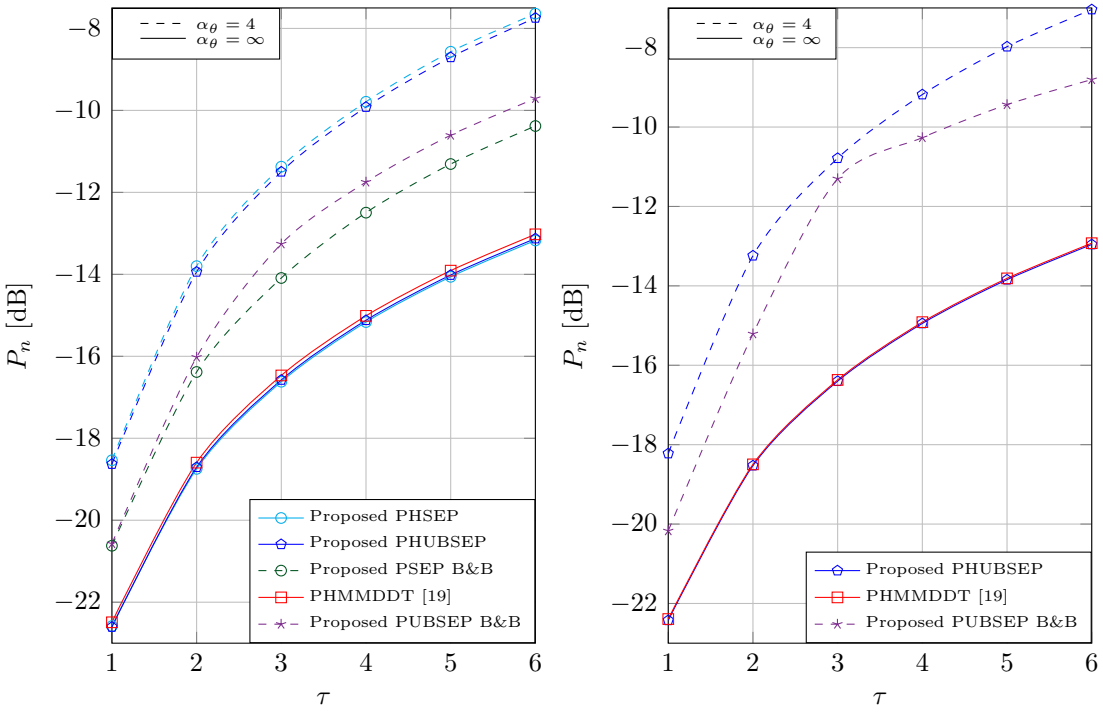

Figure 5.7: Evaluation: Average normalized transmit power versus SEP requisite parameter, $P_n \times \tau$, with SEP requisites $\rho_k = 10^{-\tau}$ for $k \in \mathcal{K}$, target power budget $P_{\mathrm{B}} = -8$ dB and acceptable power increase $\gamma = 1$ dB. First scenario: $K = 10$ users, $N = 100$ reflecting elements, $\alpha_s = 4$ PSK users' data, $\alpha_\theta = 4$ available phase shifts (left). Second scenario: $K = 5$ users, $N = 120$ reflecting elements, $\alpha_s = 8$ PSK users' data, $\alpha_\theta = 4$ available phase shifts (right).

6
Conclusions

This thesis proposes different symbol-level transmit processing methods for diverse MU-MIMO setups. First, this study proposes two SLP approaches considering a SPAPC and PSK modulation for perfect and imperfect CSI. The proposed precoding designs are formulated as SOCPs and are solved using the IPM in polynomial time. Numerical results confirm that for the perfect CSI scenario, the proposed designs are superior to the existing techniques in terms of BER for low and intermediate SNR. Moreover, when considering imperfect CSI numerical evaluations underline the superiority of the proposed RMMSE design.

For large-scale MIMO where the energy consumption of the RFFE is significant to the EE of the system, power reduction features such as CE signaling and low-resolution quantization are necessary for low-cost deployments, with low environmental impact, and better coverage. To mitigate the error-rate performance degradation that these features yield CE low-resolution precoding has become prominent in the literature. In this context, this thesis focuses on the development of SLP techniques for MU-MIMO downlink systems with arbitrary PSK modulation. While the MSEP formulation is considered as the design criterion for the case of QPSK data symbols, the study proposes the novel MUBSEP formulation for arbitrary PSK modulation. Based on these criteria the study proposes different low-resolution SLPs. First, PGS methods are devised as a low-complexity approach. Then the QoS B&B algorithm is proposed which differs from standard full B&B methods by stopping the tree search when a solution that attains the system's QoS constraint is found. Numerical results show that the proposed PGS approaches outperform the examined state-of-the-art methods either in terms of SER or computational complexity. Moreover, numerical results confirm that the proposed QoS B&B algorithm yields, in many scenarios, lower SER with lower computational complexity when compared with optimal state-of-the-art algorithms.

Finally, a virtual MU-MIMO system with PSK modulation realized via the RIS-based passive transmitter setup is considered. Under this framework, the work considers both high-resolution and discrete phase shift RIS models. For both cases, this study proposes power minimization problems under QoS

constraints. While for the case of BPSK or QPSK users' data, the SEP is considered QoS requisite, for the general M-PSK scenario the UBSEP functions are utilized as constraints. For the discrete phase-shift RIS case, the problems are formulated as MIPs and solved via an improved version of the QoS B&B approach. On the other hand, for high-resolution, they become multivariate problems that are solved via the combined utilization of the BM and the RCG algorithms. Numerical results show that the proposed power minimization approaches yield reduced transmit power compared to state-of-the-art techniques.

6.1 A Balance of the Achieved Results Regarding Branch-and-Bound Methods

A common conception in the literature is that the utilization of Full-B&B methods results in prohibitive computational complexity [90]. This perception has motivated the research community to move away from B&B techniques and, instead, develop a variety of projection-based algorithms for tackling discrete problems to achieve compromise solutions. While some of these solutions provide interesting trade-offs, they lack the performance guarantees that Full-B&B approaches can offer, which are critical for designing reliable communication networks.

In this thesis, we have revisited B&B methods to explore the balance between performance and complexity while maintaining performance guarantees. Our research demonstrates that although computing the optimal solution via a Full-B&B method can indeed be computationally expensive, allowing for small compromises can significantly reduce the computational effort required. This insight has enabled us to design B&B algorithms that achieve complexity levels comparable to state-of-the-art projection-based approaches while offering significantly improved performance and preserving the ability to guarantee performance. Based on these findings in chapter 4, we introduced the QoS B&B approach that ensures the satisfaction of SEP requirements for well-designed communication systems. Moreover, in chapter 5, we proposed a B&B design that either ensures the transmit power remains below or equal to the system's target power budget or provides a solution close to the optimal with a specified optimality factor. These contributions highlight the potential of B&B methods to deliver high-performance solutions with reduced complexity, challenging the prevailing notion of their impracticality.

In conclusion, the findings of this thesis have shown that a renewed focus on B&B algorithms within the research community can lead to fruitful results. By addressing the computational challenges associated with Full-B&B

methods, we understand that further advancements can make B&B techniques more efficient and practical for developing high-performance algorithms with manageable complexity.

6.2 A Balance of the Results with the Union-Bound SEP

Prior to the contributions presented in this thesis, most PSK modulation-based studies in the literature relied on indirect criteria for either minimizing the SEP or constraining it. Notably, methods based on MMSE and MMDDT have shown insightful results. However, the lack of a direct and quantifiable relationship between a user's MMSE and its SEP, coupled with the limitations of the MMDDT-based SEP bound from [14] — such as the absence of a closed-form expression or tabled solutions — has rendered the development of algorithms capable of real-time SEP control highly challenging.

To overcome these limitations, we proposed the union-bound SEP as an upper bound on the SEP that can be computed individually for each user. It was shown that the tightness of this bound increases with the modulation order and the SNR, achieving high accuracy in the SNR regions anticipated for future wireless communication networks. Since the proposed formulation essentially predicts the SEP for the expected cases of future wireless networks, it allows the design of algorithms that compute trade-offs while adhering to the system's QoS requirements. This enables adjustments in computational complexity, reductions in transmit power, and other optimizations without risking the delivery of the SEP necessary for the proper functioning of user applications.

The findings of this thesis have shown the benefits of incorporating SEP-related criteria into the design of transmit processing algorithms. This offers clear advantages by enabling the management of QoS requirements and supporting the real-time computation of favorable trade-offs for communication systems. Finally, this motivates a new research path in the field of real-time QoS management algorithms.

7
Future Work

In this final chapter, some possible extensions of the studies performed in this thesis are discussed. The discussion about the considered extensions is done in a chapter-by-chapter manner from chapters 3 to 5.

7.1
Future Work on Symbol-Level Precoding under Strict Per Antenna Power Constraints

The methods proposed in chapter 3 consider a narrow-band channel model such that, for the considered bandwidth, the channel $\boldsymbol{H}$ can be considered flat. An extension of the approaches from chapter 3 to broadband channels can be done via the consideration of a frequency selective channel as with $\boldsymbol{H}_l$ denoting the l-th channel tap for $l \in \{0, \ldots, L-1\}$. Different methods can be utilized for dealing with such channels. A formulation of the proposed approaches to deal with frequency-selectivity is possible. Moreover, an extension to the OFDM context would also be an interesting topic for investigation.

Another possible extension is the proposal of SPAPC precoders with the MSEP and MUBSEP concepts. With this, one could consider these either in the objective or as constraint functions for QoS guarantee. This would be especially useful for integrated sensing and communications where one could construct a Cramér-Rao lower bound minimization problem under either MSEP or MUBSEP constraints.

7.2
Future Work on Symbol-Level Precoding under Constant Envelope and Low-Resolution Constraints

The SEP-related methods proposed in chapter 4 consider perfect CSI. In practice, for the cases where the users are highly mobile, the proposed methods would require either channel tracking approaches or estimating the channel regularly to avoid the performance degradation that arises from the ever-changing channel from mobile systems. To make the proposed approaches more suitable for these imperfect CSI scenarios robust MSEP and MUBSEP formulations would be important.

7.3 Future Work on RIS-based Passive Transmitter MIMO Systems

The methods proposed in chapter 5 consider a single-antenna transmitter that radiates an unmodulated carrier signal. A possible extension is to consider the architecture from [11] which utilizes an active multi-antenna feeder (AMAF) present in the near-field from the RIS. The approach from [11] allows for a hybrid-beamforming scheme where the AMAF and the RIS perform the digital beamforming and analog beamforming tasks, respectively.

With this setup, symbol-level hybrid beamforming and channel-level hybrid beamforming are possible. When considering the symbol-level approach both the AMAF and the RIS are optimized in a symbol-by-symbol manner. In this scenario, the MSEP and MUBSEP concepts can be utilized to achieve high-performance systems. The channel-level hybrid beamforming schemes consider that both the AMAF and the RIS are optimized once per channel coherence time. In this scenario, one can consider the MMSE or the RMMSE as the objectives.

Bibliography

[1] VISWANATHAN, H.; MOGENSEN, P. E., **Communications in the 6G Era**, IEEE Access, vol. 8, pp. 57063–57074, 2020.

[2] KHAN, L. U.; YAQOOB, I.; IMRAN, M.; HAN, Z. ; HONG, C. S., **6G Wireless Systems: A Vision, Architectural Elements, and Future Directions**, IEEE Access, vol. 8, pp. 147029–147044, 2020.

[3] GIORDANI, M.; POLESE, M.; MEZZAVILLA, M.; RANGAN, S. ; ZORZI, M., **Toward 6G Networks: Use Cases and Technologies**, IEEE Commun. Mag., vol. 58, no. 3, pp. 55–61, 2020.

[4] RUSEK, F.; PERSSON, D.; LAU, B. K.; LARSSON, E. G.; MARZETTA, T. L.; EDFORS, O. ; TUFVESSON, F., **Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays**, IEEE Signal Process. Mag., vol. 30, no. 1, 2013.

[5] MEZGHANI, A.; NOSSEK, J. A., **Power Efficiency in Communication Systems from a Circuit Perspective**, In: 2011 IEEE INTERNATIONAL SYMPOSIUM OF CIRCUITS AND SYSTEMS (ISCAS), pp. 1896–1899, 2011.

[6] MEZGHANI, A.; NOSSEK, J. A., **Modeling and Minimization of Transceiver Power Consumption in Wireless Networks**, In: 2011 INTERNATIONAL ITG WORKSHOP ON SMART ANTENNAS, pp. 1–8, 2011.

[7] BASAR, E.; DI RENZO, M.; DE ROSNY, J.; DEBBAH, M.; ALOUINI, M.-S. ; ZHANG, R., **Wireless Communications Through Reconfigurable Intelligent Surfaces**, IEEE Access, vol. 7, pp. 116753–116773, 2019.

[8] KARASIK, R.; SIMEONE, O.; RENZO, M. D. ; SHAMAI SHITZ, S., **Adaptive Coding and Channel Shaping Through Reconfigurable Intelligent Surfaces: An Information-Theoretic Analysis**, IEEE Trans. Commun., 2021.

[9] BUZZI, S.; D'ANDREA, C. ; INTERDONATO, G., **Approaching Massive MIMO Performance with Reconfigurable Intelligent Surfaces:**

We Do Not Need Many Antennas, arXiv preprint arXiv:2203.07493, 2022.

[10] TIWARI, K. K.; CAIRE, G., **RIS-Based Steerable Beamforming Antenna with Near-Field Eigenmode Feeder**, In: ICC 2023 - IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, pp. 1293–1299, 2023.

[11] TIWARI, K. K.; CAIRE, G., **A New Old Idea: Beam-Steering Reflectarrays for Efficient Sub-THz Multiuser MIMO**, Authorea Preprints, 2023.

[12] AMADORI, P. V.; MASOUROS, C., **Constant Envelope Precoding by Interference Exploitation in Phase Shift Keying-Modulated Multiuser Transmission**, IEEE Trans. Commun., Jan 2017.

[13] LANDAU, L.; KRONE, S. ; FETTWEIS, G., **Intersymbol-Interference Design for Maximum Information Rates with 1-Bit Quantization and Oversampling at the Receiver**, In: SCC 2013; 9TH INTERNATIONAL ITG CONFERENCE ON SYSTEMS, COMMUNICATION AND CODING, Jan 2013.

[14] JEDDA, H.; MEZGHANI, A.; SWINDLEHURST, A. L. ; NOSSEK, J. A., **Quantized Constant Envelope Precoding With PSK and QAM Signaling**, IEEE Trans. Wireless Commun., vol. 17, no. 12, pp. 8022–8034, Dec 2018.

[15] ASKERBEYLI, F.; XU, W. ; NOSSEK, J. A., **1-Bit Precoding for Massive MIMO Downlink with Linear Programming and a Greedy Algorithm Extension**, In: 2021 IEEE 93RD VEHICULAR TECHNOLOGY CONFERENCE (VTC2021-SPRING), pp. 1–5, 2021.

[16] PARK, G.-J.; HONG, S.-N., **Construction of 1-Bit Transmit-Signal Vectors for Downlink MU-MISO Systems With PSK Signaling**, IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 8270–8274, 2019.

[17] LANDAU, L. T. N.; DE LAMARE, R. C., **Branch-and-Bound Precoding for Multiuser MIMO Systems With 1-Bit Quantization**, IEEE Wireless Commun. Lett., vol. 6, no. 6, pp. 770–773, Dec 2017.

[18] LOPES, E. S. P.; LANDAU, L. T. N., **Optimal Precoding for Multiuser MIMO Systems With Phase Quantization and PSK Modulation via Branch-and-Bound**, IEEE Wireless Commun. Lett., 2020.

[19] LIU, R.; LI, M.; LIU, Q.; SWINDLEHURST, A. L. ; WU, Q., **Intelligent Reflecting Surface Based Passive Information Transmission: A Symbol-Level Precoding Approach**, IEEE Transactions on Vehicular Technology, vol. 70, no. 7, pp. 6735–6749, 2021.

[20] A. H. LAND AND A. G. DOIG, **An Automatic Method of Solving Discrete Programming Problems**, Econometrica, vol. 28, no. 3, pp. 497–520, 1960.

[21] LUO, J.; PATTIPATI, K.; WILLETT, P. ; BRUNEL, L., **Branch-and-bound-Based Fast Optimal Algorithm for Multiuser Detection in Synchronous CDMA**, In: PROC. IEEE INT. CONF. COMMUN. (ICC), pp. 3336–3340 vol.5, 2003.

[22] ISRAEL, J.; FISCHER, A.; MARTINOVIC, J.; JORSWIECK, E. A. ; MESYAGUTOV, M., **Discrete Receive Beamforming**, IEEE Signal Process. Lett., vol. 22, no. 7, pp. 958–962, 2015.

[23] ISRAEL, J.; FISCHER, A. ; MARTINOVIC, J., **A Branch-and-Bound Algorithm for Discrete Receive Beamforming with Improved Bounds**, In: 2015 IEEE INTERNATIONAL CONFERENCE ON UBIQUITOUS WIRELESS BROADBAND (ICUWB), pp. 1–5, 2015.

[24] JEON, Y.; LEE, N.; HONG, S. ; HEATH, R. W., **One-Bit Sphere Decoding for Uplink Massive MIMO Systems With One-Bit ADCs**, IEEE Trans. Wireless Commun., vol. 17, no. 7, pp. 4509–4521, 2018.

[25] LI, A.; LIU, F.; MASOUROS, C.; LI, Y. ; VUCETIC, B., **Interference Exploitation 1-Bit Massive MIMO Precoding: A Partial Branch-and-Bound Solution With Near-Optimal Performance**, IEEE Trans. Wireless Commun., vol. 19, no. 5, pp. 3474–3489, 2020.

[26] LOPES, E. S. P.; LANDAU, L. T. N. ; MEZGHANI, A., **Minimum Symbol Error Probability Discrete Symbol Level Precoding for MU-MIMO Systems With PSK Modulation**, IEEE Transactions on Communications, 2023.

[27] LOPES, E. S. P.; LANDAU, L. T. N., **Discrete MMSE Precoding for Multiuser MIMO Systems with PSK Modulation**, IEEE Transactions on Wireless Communications, 2022.

[28] JOHAM, M.; UTSCHICK, W. ; NOSSEK, J. A., **Linear Transmit Processing in MIMO Communications Systems**, IEEE Trans. Signal Process., vol. 53, no. 8, pp. 2700–2712, Aug 2005.

[29] KAMMOUN, A.; MÜLLER, A.; BJÖRNSON, E. ; DEBBAH, M., **Linear Precoding Based on Polynomial Expansion: Large-Scale Multi-Cell MIMO Systems**, IEEE J. Sel. Areas Commun., 2014.

[30] HOYDIS, J.; TEN BRINK, S. ; DEBBAH, M., **Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?**, IEEE Journal on Selected Areas in Communications, 2013.

[31] PEEL, C.; HOCHWALD, B. ; SWINDLEHURST, A., **A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication-Part I: Channel Inversion and Regularization**, IEEE Trans. Commun., 2005.

[32] YU, W.; LAN, T., **Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints**, IEEE Trans. Signal Process., 2007.

[33] BOCCARDI, F.; HUANG, H., **Zero-Forcing Precoding for the MIMO Broadcast Channel under Per-Antenna Power Constraints**, In: 2006 IEEE 7TH WORKSHOP ON SIGNAL PROCESS. ADV. IN WIRELESS COMMUN. (SPAWC), 2006.

[34] KARAKAYALI, K.; YATES, R.; FOSCHINI, G. ; VALENZUELA, R., **Optimum Zero-forcing Beamforming with Per-antenna Power Constraints**, In: 2007 IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2007.

[35] FENG, C.; JING, Y., **Modified MRT and Outage Probability Analysis for Massive MIMO Downlink under Per-Antenna Power Constraint**, In: 2016 IEEE 17TH WORKSHOP ON SIGNAL PROCESS. ADV. IN WIRELESS COMMUN. (SPAWC), 2016.

[36] CHEN, C.-E., **MSE-Based Precoder Designs for Transmitter-Preprocessing-Aided Spatial Modulation Under Per-Antenna Power Constraints**, IEEE Transactions on Vehicular Technology, 2017.

[37] PI, Z., **Optimal Transmitter Beamforming with Per-Antenna Power Constraints**, In: PROC. IEEE INT. CONF. COMMUN. (ICC), 2012.

[38] DA-SHAN SHIU; FOSCHINI, G. J.; GANS, M. J. ; KAHN, J. M., **Fading Correlation and Its Effect on the Capacity of Multielement Antenna Systems**, IEEE Trans. Commun., March 2000.

[39] CHEN, C.-E., **Computationally Efficient Constructive Interference Precoding for PSK Modulations Under Per-Antenna Power Constraint**, IEEE Transactions on Vehicular Technology, 2020.

[40] LOPES, E. S. P.; LANDAU, L. T. N., **MMSE Symbol Level Precoding Under a Per Antenna Power Constraint for Multiuser MIMO Systems With PSK Modulation**, IEEE Wireless Communications Letters, vol. 11, no. 11, pp. 2440–2444, 2022.

[41] LI, A.; MASOUROS, C., **Interference Exploitation Precoding Made Practical: Optimal Closed-Form Solutions for PSK Modulations**, IEEE Trans. Wireless Commun., 2018.

[42] BOYD, S.; VANDENBERGHE, L., *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[43] LIU, F.; MASOUROS, C.; AMADORI, P. V. ; SUN, H., **An Efficient Manifold Algorithm for Constructive Interference Based Constant Envelope Precoding**, IEEE Signal Processing Letters, 2017.

[44] WEN, Y.; WANG, H.; LI, A.; LIAO, X. ; MASOUROS, C., **Low-Complexity Interference Exploitation MISO Precoding Under Per-Antenna Power Constraint**, IEEE Transactions on Wireless Communications, 2024.

[45] FESL, B.; JEDDA, H. ; NOSSEK, J. A., **Discrete One-Bit Precoding for Massive MIMO**, In: WSA 2019; 23RD INTERNATIONAL ITG WORKSHOP ON SMART ANTENNAS, pp. 1–6, April 2019.

[46] B. CUNHA, T. E.; DE LAMARE, R. C.; FERREIRA, T. N. ; HÄLSIG, T., **Joint Automatic Gain Control and MMSE Receiver Design for Quantized Multiuser MIMO Systems**, In: 2018 15TH INTERNATIONAL SYMPOSIUM ON WIRELESS COMMUNICATION SYSTEMS (ISWCS), pp. 1–5, 2018.

[47] DUARTE, F. L.; DE LAMARE, R. C., **Cloud-Driven Multi-Way Multiple-Antenna Relay Systems: Best-User-Link Selection and Joint Mmse Detection**, In: ICASSP 2020 - 2020 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), pp. 5160–5164, 2020.

[48] LOPES, E. S. P.; LANDAU, L. T. N., **Optimal and Suboptimal MMSE Precoding for Multiuser MIMO Systems Using Constant Envelope Signals with Phase Quantization at the Transmitter and PSK Modulation**, In: WSA 2020; 24TH INT. ITG WORKSHOP ON SMART ANTENNAS, 2020.

[49] JACOBSSON, S.; DURISI, G.; COLDREY, M.; GOLDSTEIN, T. ; STUDER, C., **Quantized Precoding for Massive MU-MIMO**, IEEE Trans. Commun., vol. 65, no. 11, pp. 4670–4684, 2017.

[50] PENG, J.; ROOS, C. ; TERLAKY, T., **New Complexity Analysis of the Primal—Dual Newton Method for Linear Optimization**, Annals of Operations Research, 2000.

[51] JEDDA, H.; NOSSEK, J. A. ; MEZGHANI, A., **Minimum BER precoding in 1-bit massive MIMO systems**, In: PROC. OF IEEE SENSOR ARRAY AND MULTICHANNEL SIGNAL PROCESSING WORKSHOP (SAM), July 2016.

[52] LI, A.; SPANO, D.; KRIVOCHIZA, J.; DOMOUCHTSIDIS, S.; TSINOS, C. G.; MASOUROS, C.; CHATZINOTAS, S.; LI, Y.; VUCETIC, B. ; OTTERSTEN, B., **A Tutorial on Interference Exploitation via Symbol-Level Precoding: Overview, State-of-the-Art and Future Directions**, IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 796–839, 2020.

[53] ALODEH, M.; SPANO, D.; KALANTARI, A.; TSINOS, C. G.; CHRISTOPOULOS, D.; CHATZINOTAS, S. ; OTTERSTEN, B., **Symbol-Level and Multicast Precoding for Multiuser Multiantenna Downlink: A State-of-the-Art, Classification, and Challenges**, IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1733–1757, 2018.

[54] MEZGHANI, A.; HEATH, R. W., **Massive MIMO Precoding and Spectral Shaping with Low Resolution Phase-only DACs and Active Constellation Extension**, IEEE Trans. Wireless Commun., pp. 1–1, 2022.

[55] MENDONÇA, M. O. K.; DINIZ, P. S. R.; FERREIRA, T. N. ; LOVISOLO, L., **Antenna Selection in Massive MIMO Based on Greedy Algorithms**, IEEE Transactions on Wireless Communications, vol. 19, no. 3, pp. 1868–1881, 2020.

[56] PINTO, E.; GALDINO, J. ; OTHERS, **Simple and Robust Analytically Derived Variable Step-Size Least Mean Squares Algorithm for Channel Estimation**, IET communications, vol. 3, no. 12, pp. 1832–1842, 2009.

[57] MOHAMMADZADEH, S.; NASCIMENTO, V. H.; LAMARE, R. C. D. ; KUKRER, O., **Covariance Matrix Reconstruction Based on Power Spectral Estimation and Uncertainty Region for Robust Adaptive Beamforming**, IEEE Trans. Aerosp. Electron. Syst., vol. 59, no. 4, pp. 3848–3858, 2023.

[58] PALHARES, V. M. T.; FLORES, A. R. ; DE LAMARE, R. C., **Robust MMSE Precoding and Power Allocation for Cell-Free Massive MIMO Systems**, IEEE Transactions on Vehicular Technology, vol. 70, no. 5, pp. 5115–5120, 2021.

[59] LOPES, E. D. S. P., *Discrete Precoding and Adjusted Detection for Multiuser MIMO Systems with PSK Modulation*, PhD thesis, PUC-Rio, 2021.

[60] YANG, H.; MARZETTA, T. L., **Performance of Conjugate and Zero-Forcing Beamforming in Large-Scale Antenna Systems**, IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 172–179, 2013.

[61] MOHAMMED, S. K.; LARSSON, E. G., **Per-Antenna Constant Envelope Precoding for Large Multi-User MIMO Systems**, IEEE Trans. Commun., vol. 61, no. 3, pp. 1059–1071, March 2013.

[62] GOLDSMITH, A., *Wireless communications*, Cambridge university press, 2005.

[63] PAN, C.; REN, H.; WANG, K.; KOLB, J. F.; ELKASHLAN, M.; CHEN, M.; DI RENZO, M.; HAO, Y.; WANG, J.; SWINDLEHURST, A. L.; YOU, X. ; HANZO, L., **Reconfigurable Intelligent Surfaces for 6G Systems: Principles, Applications, and Research Directions**, IEEE Communications Magazine, vol. 59, no. 6, pp. 14–20, 2021.

[64] RAPPAPORT, T. S.; XING, Y.; KANHERE, O.; JU, S.; MADANAYAKE, A.; MANDAL, S.; ALKHATEEB, A. ; TRICHOPOULOS, G. C., **Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond**, IEEE Access, vol. 7, pp. 78729–78757, 2019.

[65] WU, Q.; ZHANG, R., **Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network**, IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, 2020.

[66] LIU, Y.; LIU, X.; MU, X.; HOU, T.; XU, J.; DI RENZO, M. ; AL-DHAHIR, N., **Reconfigurable Intelligent Surfaces: Principles and Opportunities**, IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1546–1577, 2021.

[67] NGO, H. Q.; LARSSON, E. G. ; MARZETTA, T. L., **Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems**, IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436–1449, 2013.

[68] ZHAO, H.; SHUANG, Y.; WEI, M.; CUI, T. J.; HOUGNE, P. D. ; LI, L., **Metasurface-Assisted Massive Backscatter Wireless Communication with Commodity Wi-Fi Signals**, Nature communications, vol. 11, no. 1, pp. 3926, 2020.

[69] GALAPPATHTHIGE, D. L.; REZAEI, F.; TELLAMBURA, C. ; HERATH, S., **RIS-Empowered Ambient Backscatter Communication Systems**, IEEE Wireless Communications Letters, vol. 12, no. 1, pp. 173–177, 2022.

[70] ZAPPONE, A.; DI RENZO, M.; SHAMS, F.; QIAN, X. ; DEBBAH, M., **Overhead-Aware Design of Reconfigurable Intelligent Surfaces in Smart Radio Environments**, IEEE Transactions on Wireless Communications, vol. 20, no. 1, pp. 126–141, 2021.

[71] ALEXANDROPOULOS, G. C.; SHLEZINGER, N.; ALAMZADEH, I.; IMANI, M. F.; ZHANG, H. ; ELDAR, Y. C., **Hybrid Reconfigurable Intelligent Metasurfaces: Enabling Simultaneous Tunable Reflections and Sensing for 6G Wireless Communications**, IEEE Vehicular Technology Magazine, vol. 19, no. 1, pp. 75–84, 2024.

[72] JUNG, M.; SAAD, W.; DEBBAH, M. ; HONG, C. S., **On the Optimality of Reconfigurable Intelligent Surfaces (RISs): Passive Beamforming, Modulation, and Resource Allocation**, IEEE Trans. Wireless Commun., vol. 20, no. 7, pp. 4347–4363, 2021.

[73] LIN, S.; ZHENG, B.; ALEXANDROPOULOS, G. C.; WEN, M.; DI RENZO, M. ; CHEN, F., **Reconfigurable Intelligent Surfaces with Reflection Pattern Modulation: Beamforming Design and Performance Analysis**, IEEE Trans. Wireless Commun., 2020.

[74] YAN, W.; YUAN, X.; HE, Z.-Q. ; KUAI, X., **Passive Beamforming and Information Transfer Design for Reconfigurable Intelligent Surfaces Aided Multiuser MIMO Systems**, IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1793–1808, 2020.

[75] LIN, S.; ZHENG, B.; ALEXANDROPOULOS, G. C.; WEN, M.; DI RENZO, M. ; CHEN, F., **Joint Passive Beamforming and Information Transfer for RIS-Empowered Wireless Communications**, In: GLOBECOM 2020 - 2020 IEEE GLOBAL COMMUNICATIONS CONFERENCE, pp. 1–6, 2020.

[76] LI, Q.; WEN, M. ; DI RENZO, M., **Single-RF MIMO: From Spatial Modulation to Metasurface-Based Modulation**, IEEE Wireless Communications, vol. 28, no. 4, pp. 88–95, 2021.

[77] REHMAN, H. U.; BELLILI, F.; MEZGHANI, A. ; HOSSAIN, E., **Modulating Intelligent Surfaces for Multiuser MIMO Systems: Beamforming and Modulation Design**, IEEE Trans. Commun., 2022.

[78] LOPES, E. S. P.; LANDAU, L. T. N., **MMSE Symbol Level Precoding Under a Per Antenna Power Constraint for Multiuser MIMO Systems With PSK Modulation**, IEEE Wireless Communications Letters, vol. 11, no. 11, pp. 2440–2444, 2022.

[79] SHAO, M.; LI, Q.; MA, W.-K. ; SO, A. M.-C., **A Framework for One-Bit and Constant-Envelope Precoding Over Multiuser Massive MISO Channels**, IEEE Transactions on Signal Processing, vol. 67, no. 20, pp. 5309–5324, 2019.

[80] SHAO, M.; LI, Q. ; MA, W.-K., **Minimum Symbol-Error Probability Symbol-Level Precoding With Intelligent Reflecting Surface**, IEEE Wireless Commun. Lett., vol. 9, no. 10, pp. 1601–1605, 2020.

[81] WU, Q.; ZHANG, S.; ZHENG, B.; YOU, C. ; ZHANG, R., **Intelligent Reflecting Surface-Aided Wireless Communications: A Tutorial**, IEEE Transactions on Communications, vol. 69, no. 5, pp. 3313–3351, 2021.

[82] ZHENG, B.; YOU, C. ; ZHANG, R., **Intelligent Reflecting Surface Assisted Multi-User OFDMA: Channel Estimation and Training Design**, IEEE Transactions on Wireless Communications, vol. 19, no. 12, pp. 8315–8329, 2020.

[83] DE ARAÚJO, G. T.; DE ALMEIDA, A. L. F. ; BOYER, R., **Channel Estimation for Intelligent Reflecting Surface Assisted MIMO**

Systems: A Tensor Modeling Approach, IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 3, pp. 789–802, 2021.

[84] YOU, C.; ZHENG, B. ; ZHANG, R., **Channel Estimation and Passive Beamforming for Intelligent Reflecting Surface: Discrete Phase Shift and Progressive Refinement**, IEEE Journal on Selected Areas in Communications, vol. 38, no. 11, pp. 2604–2620, 2020.

[85] LI, Q.; EL-HAJJAR, M.; HEMADEH, I.; JAGYASI, D.; SHOJAEIFARD, A.; BASAR, E. ; HANZO, L., **The Reconfigurable Intelligent Surface-Aided Multi-Node IoT Downlink: Beamforming Design and Performance Analysis**, IEEE Internet of Things Journal, vol. 10, no. 7, pp. 6400–6414, 2022.

[86] LOPES, E. S. P.; LANDAU, L. T. N. ; MEZGHANI, A., **Minimum Union Bound Symbol Error Probability Precoding for PSK Modulation and Phase Quantization**, In: 2022 IEEE GLOBECOM WORKSHOPS (GC WKSHPS), pp. 1681–1686, 2022.

[87] BOUMAL, N., *An introduction to optimization on smooth manifolds*, Cambridge University Press, 2023.

[88] BOUMAL, N.; MISHRA, B.; ABSIL, P.-A. ; SEPULCHRE, R., **Manopt, a Matlab Toolbox for Optimization on Manifolds**, Journal of Machine Learning Research, vol. 15, no. 42, pp. 1455–1459, 2014.

[89] BOUMAL, N., **Optimization and Estimation on Manifolds.**, 2014.

[90] LI, A.; MASOUROS, C.; SWINDLEHURST, A. L. ; YU, W., **1-bit massive MIMO transmission: Embracing interference with symbol-level precoding**, IEEE Communications Magazine, vol. 59, no. 5, pp. 121–127, 2021.

A
Convexity Analysis

This section of the appendix provides proof of the convexity of the SEP-related functions and derives the conditions of which the UBSEP-related functions are convex.

A.1
Proof of Convexity of the MSEP objective

Convexity of the MSEP objective is established by proving that the Hessian of the objective is positive semi-definite (PSD) for all values of $\boldsymbol{x}_\mathrm{r}$. As written in (4-22), $\nabla^2 f_0(\boldsymbol{x}_\mathrm{r})$ is given by

$$\nabla^2 f_0(\boldsymbol{x}_\mathrm{r}) = \sum_{k=1}^{K} \frac{\boldsymbol{m}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{R,k}^T\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)\right)^2}+\frac{\boldsymbol{m}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{I,k}^T\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)}{\left(\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^T\boldsymbol{x}_\mathrm{r}\right)\right)^2}, \tag{A-1}$$

where

$$\boldsymbol{m}_{R,k} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{R},k}, \tag{A-2}$$

$$\boldsymbol{m}_{I,k} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{h}_{\mathrm{I},k}^T\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{I},k}, \tag{A-3}$$

$$\boldsymbol{\Psi}_{R,k} = \frac{\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{R},k}^T, \tag{A-4}$$

$$\boldsymbol{\Psi}_{I,k} = \frac{\Phi\left(\boldsymbol{h}_{\mathrm{I},k}^T\boldsymbol{x}_\mathrm{r}\right)}{\sqrt{2\pi}}\, e^{-\frac{\left(\boldsymbol{h}_{\mathrm{I},k}^T\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{I},k}\boldsymbol{h}_{\mathrm{I},k}^T\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{I},k}^T. \tag{A-5}$$

Note that, since for $\left(\Phi\left(\alpha\right)\right)^2 \geq 0$ for $\alpha \in \mathbb{R}$, a sufficient condition for convexity is proving that $\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{m}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{R,k}^T\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{R,k}\left(\boldsymbol{x}_\mathrm{r}\right)$ and $\boldsymbol{\Gamma}_{I,k}(\boldsymbol{x}_\mathrm{r}) = \boldsymbol{m}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{m}_{I,k}^T\left(\boldsymbol{x}_\mathrm{r}\right)+\boldsymbol{\Psi}_{I,k}\left(\boldsymbol{x}_\mathrm{r}\right)$ are PSD for $k \in \{1,\ldots,K\}$. Expanding $\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r})$

yields

$$\begin{aligned}\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r}) &= \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{h_{R,k}^T\boldsymbol{x}_\mathrm{r}}{2}}\right)^2 \boldsymbol{h}_{R,k}\boldsymbol{h}_{R,k}^T + \frac{\Phi\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)}{\sqrt{2\pi}}\, e^{-\frac{\left(h_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)^2}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{R},k}^T \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{h_{R,k}^T\boldsymbol{x}_\mathrm{r}}{2}}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{h_{R,k}^T\boldsymbol{x}_\mathrm{r}}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T + \Phi(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r})\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{R},k}^T\right).\end{aligned}$$

Since $\frac{1}{\sqrt{2\pi}}e^{-\frac{h_{R,k}^T\boldsymbol{x}_\mathrm{r}}{2}} \geq 0$, the matrix $\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r})$ is PSD if and only if $\boldsymbol{\Upsilon}_k(\boldsymbol{x}_\mathrm{r}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{h_{R,k}^T\boldsymbol{x}_\mathrm{r}}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T + \Phi\left(\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\right)\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}\boldsymbol{h}_{\mathrm{R},k}^T$ is PSD for $k \in \{1,\ldots,K\}$. Calling $a = \boldsymbol{h}_{\mathrm{R},k}^T\boldsymbol{x}_\mathrm{r}$ yields

$$\boldsymbol{\Upsilon}_k(a) = \frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}}\boldsymbol{h}_{\mathrm{R},k}\boldsymbol{h}_{\mathrm{R},k}^T + \boldsymbol{h}_{\mathrm{R},k}\left(a\Phi(a)\right)\boldsymbol{h}_{\mathrm{R},k}^T \tag{A-6}$$

$$= \boldsymbol{h}_{\mathrm{R},k}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} + a\Phi(a)\right)\boldsymbol{h}_{\mathrm{R},k}^T. \tag{A-7}$$

Note that, $\boldsymbol{v}\ \alpha\ \boldsymbol{v}^T \succeq \mathbf{0}$ for $\boldsymbol{v} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^+$. As a consequence, a sufficient condition for PSD is having $g(a) = \frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} + a\Phi(a) \geq 0$ for $a \in \mathbb{R}$. To prove that $g(a)$ is positive for $a \in \mathbb{R}$, $\frac{\partial g(a)}{\partial a}$ is computed in what follows,

$$\frac{\partial g(a)}{\partial a} = \frac{\partial}{\partial a}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} + a\Phi(a)\right) \tag{A-8}$$

$$= -\frac{a}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} + \Phi(a) + \frac{a}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} \tag{A-9}$$

$$= \Phi(a). \tag{A-10}$$

Note that, since $\frac{\partial g(a)}{\partial a} = \Phi(a)$ is always greater than zero, the function $g(a)$ is monotonically increasing. This implies that $g(a)$ approaches its minimum value as a tends to $-\infty$. With this, to prove that $g(a) \geq 0\ \forall\ a \in \mathbb{R}$ it is sufficient to prove that $\lim_{a\to-\infty} g(a) \geq 0$. Computing the limit,

$$\lim_{a\to-\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}} + a\Phi(a) = \lim_{a\to-\infty} a\Phi(a), \tag{A-11}$$

using the identity $\Phi(a) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)\right)$ yields

$$\lim_{a\to-\infty}\frac{a}{2}\left(1 + \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)\right) = 0. \tag{A-12}$$

With this, the minimum value of $g(a)$ is zero, which implies that for $\boldsymbol{x}_\mathrm{r} \in \mathbb{R}^{2M}$, $\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}$ for $k \in \{1,\ldots,K\}$. Similar steps can be taken to prove that for $\boldsymbol{x}_\mathrm{r} \in \mathbb{R}^{2M}$, $\boldsymbol{\Gamma}_{I,k}(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}$ for $k \in \{1,\ldots,K\}$. Finally, having $\boldsymbol{\Gamma}_{R,k}(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}$ for

$k \in \{1, \ldots, K\}$ and $\boldsymbol{\Gamma}_{I,k}(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}$ for $k \in \{1, \ldots, K\}$ yields that $\nabla^2 f_0(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}$ and, thus, $f_0(\boldsymbol{x}_\mathrm{r})$ is convex in $\boldsymbol{x}_\mathrm{r}$.

A.2 Conditions for Convexity of the MUBSEP Objective

Considering the real-valued formulation described in (4-33) the MUBSEP objective reads as

$$f_0(\boldsymbol{x}_\mathrm{r}) = -\sum_{k=1}^{K} \ln\left(\operatorname{erf}\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right) + \operatorname{erf}\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)\right). \tag{A-13}$$

Convexity can be proven by evaluating the conditions under which the Hessian is PSD [42]. Taking the derivative of $f_0(\boldsymbol{x}_\mathrm{r})$ with respect to $\boldsymbol{x}_\mathrm{r}$ yields

$$\nabla f_0(\boldsymbol{x}_\mathrm{r}) = -\sum_{k=1}^{K} \frac{\frac{2}{\sqrt{\pi}}\, e^{-\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{1,k} + \frac{2}{\sqrt{\pi}}\, e^{-\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{2,k}}{\operatorname{erf}\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right) + \operatorname{erf}\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)}. \tag{A-14}$$

The gradient can be written in the form $\nabla f_0(\boldsymbol{x}_\mathrm{r}) = -\sum_{k=1}^{K} \frac{\boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r})}{q_k(\boldsymbol{x}_\mathrm{r})}$ where $q_k(\boldsymbol{x}_\mathrm{r}) = \operatorname{erf}\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right) + \operatorname{erf}\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)$, and, $\boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r}) = \frac{2}{\sqrt{\pi}}\left(e^{-\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{1,k} + e^{-\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{2,k}\right)$. The Hessian, then, reads as

$$\nabla^2 f_0(\boldsymbol{x}_\mathrm{r}) = \frac{\partial^2 f_0(\boldsymbol{x}_\mathrm{r})}{\partial \boldsymbol{x}_\mathrm{r} \partial \boldsymbol{x}_\mathrm{r}^T} = -\sum_{k=1}^{K} \frac{\frac{\partial \boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r})}{\partial \boldsymbol{x}_\mathrm{r}^T} q_k(\boldsymbol{x}_\mathrm{r}) - \boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r}) \frac{\partial q_k(\boldsymbol{x}_\mathrm{r})}{\partial \boldsymbol{x}_\mathrm{r}^T}}{\left(q_k(\boldsymbol{x}_\mathrm{r})\right)^2}, \tag{A-15}$$

with $\frac{\partial \boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r})}{\partial \boldsymbol{x}_\mathrm{r}^T} = -\left(\boldsymbol{\Psi}_{1,k} + \boldsymbol{\Psi}_{2,k}\right)$ and $\frac{\partial q_k(\boldsymbol{x}_\mathrm{r})}{\partial \boldsymbol{x}_\mathrm{r}^T} = \boldsymbol{m}_k^T(\boldsymbol{x}_\mathrm{r})$, and, $\boldsymbol{\Psi}_{1,k}$ and $\boldsymbol{\Psi}_{2,k}$ given by

$$\boldsymbol{\Psi}_{1,k} = \frac{4}{\sqrt{\pi}} e^{-\left(\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{1,k} \boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r} \boldsymbol{u}_{1,k}^T, \quad \boldsymbol{\Psi}_{2,k} = \frac{4}{\sqrt{\pi}} e^{-\left(\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r}\right)^2} \boldsymbol{u}_{2,k} \boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r} \boldsymbol{u}_{2,k}^T. \tag{A-16}$$

The Hessian then reads as

$$\nabla^2 f_0(\boldsymbol{x}_\mathrm{r}) = \sum_{k=1}^{K} \frac{\left(\boldsymbol{\Psi}_{1,k} + \boldsymbol{\Psi}_{2,k}\right) q_k(\boldsymbol{x}_\mathrm{r}) + \boldsymbol{m}_k(\boldsymbol{x}_\mathrm{r}) \boldsymbol{m}_k^T(\boldsymbol{x}_\mathrm{r})}{\left(q_k(\boldsymbol{x}_\mathrm{r})\right)^2}. \tag{A-17}$$

A sufficient condition for PSD is $\left(\boldsymbol{\Psi}_{1,k} + \boldsymbol{\Psi}_{2,k}\right) q_k(\boldsymbol{x}_\mathrm{r}) \succeq \mathbf{0}\ \forall\ k \in \{1, \ldots, K\}$. With this, positive semi-definiteness is achieved for $\boldsymbol{u}_{1,k}^T \boldsymbol{x}_\mathrm{r} \geq 0$, $\boldsymbol{u}_{2,k}^T \boldsymbol{x}_\mathrm{r} \geq 0, \forall k \in \{1, \ldots, K\}$. Note that, this implies $d_{1,k}(\boldsymbol{x}) \geq 0$, $d_{2,k}(\boldsymbol{x}) \geq 0, \forall k \in \{1, \ldots, K\}$. Finally, the condition for convexity of the MUBSEP objective function can be cast in a stacked manner for all k as $\boldsymbol{C}\boldsymbol{x}_\mathrm{r} \preceq \mathbf{0}$, where $\boldsymbol{C} = \left[\left(\boldsymbol{H}_{\mathrm{R},\theta}^{s^*} - \boldsymbol{H}_{\mathrm{I},\theta}^{s^*}\right)^T, \left(\boldsymbol{H}_{\mathrm{R},\theta}^{s^*} + \boldsymbol{H}_{\mathrm{I},\theta}^{s^*}\right)^T\right]^T$.

A.3
Proof of Convexity of the SEP Functions

For proving convexity we depart from a function with known properties and apply a series of operations to arrive at $f_k(\boldsymbol{\theta}_\mathrm{r}) = -\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\boldsymbol{h}_{\xi,k}^T\boldsymbol{\theta}_\mathrm{r}\right)\right) - \beta_k$. To this end, consider the log-concave function [42, Example 3.39], $\Phi(\theta) = \int_{-\infty}^{\theta} e^{-\frac{u^2}{2}} du$. A function f is log-concave if $\log f$ is concave [42, Definition 3.5.1]. With this, $g(\theta) = \ln(\Phi(\theta))$ is a concave function. Note that, $h(\boldsymbol{\theta}_\mathrm{r}) = g(\boldsymbol{A}\boldsymbol{\theta}_\mathrm{r} + \boldsymbol{b})$ is concave if $g(\theta)$ is concave [42, Section 3.2.2]. With this, it follows that $h_1(\boldsymbol{\theta}_\mathrm{r}) = \ln(\Phi(\boldsymbol{h}_{1,k}\boldsymbol{\theta}_\mathrm{r}))$ and $h_2(\boldsymbol{\theta}_\mathrm{r}) = \ln(\Phi(\boldsymbol{h}_{2,k}\boldsymbol{\theta}_\mathrm{r}))$ are concave functions. As stated in [42, Section 3.2.1], if $w_i \geq 0$ and $f_i(\boldsymbol{\theta}_\mathrm{r})$ is concave for all i, then $u(\boldsymbol{\theta}_\mathrm{r}) = \sum_i w_i f_i(\boldsymbol{\theta}_\mathrm{r})$ is concave. By setting $w_i = 1$ for $i \in \{1, 2, 3\}$ and $f_1(\boldsymbol{\theta}_\mathrm{r}) = \ln(\Phi(\boldsymbol{h}_{1,k}\boldsymbol{\theta}_\mathrm{r}))$, $f_2(\boldsymbol{\theta}_\mathrm{r}) = \ln(\Phi(\boldsymbol{h}_{2,k}\boldsymbol{\theta}_\mathrm{r}))$ and $f_3(\boldsymbol{\theta}_\mathrm{r}) = \beta_k$ one can assemble the concave function $u(\boldsymbol{\theta}_\mathrm{r}) = \sum_{\xi=1}^{2} \ln(\Phi(\boldsymbol{h}_{\xi,k}\boldsymbol{\theta}_\mathrm{r})) + \beta_k$. Note that $f_k(\boldsymbol{\theta}_\mathrm{r})$ is convex since $f_k(\boldsymbol{\theta}_\mathrm{r}) = -u(\boldsymbol{\theta}_\mathrm{r})$ and $u(\boldsymbol{\theta}_\mathrm{r})$ is concave.

A.4
Condition for convexity of the Union-Bound SEP functions

This section derives the conditions in which the UBSEP functions $f_k(\boldsymbol{x}) = \frac{1}{2}\sum_{\xi=1}^{2} \mathrm{erfc}(\boldsymbol{\nu}_{\xi,k}^T\boldsymbol{x})$ are convex. As stated in [42, Section 3.1.4] convexity can be proven by evaluating the conditions under which the Hessian is PSD. Taking the derivative of $f_k(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ yields

$$\frac{\partial f_k(\boldsymbol{x})}{\partial \boldsymbol{x}} = -\frac{1}{\sqrt{\pi}} \sum_{\xi=1}^{2} e^{-\left(\boldsymbol{v}_{\xi,k}^T\boldsymbol{x}\right)^2} \boldsymbol{v}_{\xi,k}. \tag{A-18}$$

The Hessian is then computed by taking the derivative with respect to $\boldsymbol{x}^T$, which yields

$$\frac{\partial f_k^2(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^T} = \frac{2}{\sqrt{\pi}} \sum_{\xi=1}^{2} e^{-\left(\boldsymbol{v}_{\xi,k}^T\boldsymbol{x}\right)^2} \boldsymbol{v}_{\xi,k}\left(\boldsymbol{v}_{\xi,k}^T\boldsymbol{x}\right)\boldsymbol{v}_{\xi,k}^T. \tag{A-19}$$

Note that, a sufficient condition for $\nabla^2 f_k(\boldsymbol{x})$ to be PSD is $\boldsymbol{v}_{1,k}^T\boldsymbol{x} \geq 0$ and $\boldsymbol{v}_{2,k}^T\boldsymbol{x} \geq 0$.

A.5
Convexity Analysis for the High-Resolution Constraint Functions

This section examines the SEP and UBSEP constraint functions formulated for high-resolution cases, proves that the SEP functions are matrix convex, and demonstrates the conditions for matrix convexity UBSEP functions.

First, consider the SEP constraint function

$$f_k(\boldsymbol{\Theta}) = -\sum_{\xi=1}^{2} \ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\ \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right)\right)\right) - \beta_k. \tag{A-20}$$

As proven in appendix A.3, the function $g(x) = -\ln\left(\Phi\left(\sqrt{\frac{P}{\sigma_w^2}}\ x\right)\right) - \beta_k$ is convex. Note that, by definition $\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right) = \sum_{i=1}^{2}\sum_{j=1}^{N}\theta_{i,j}h_{i,j}^{\xi}$, where $\theta_{i,j}$ and $h_{i,j}^{\xi}$ denote the entry on the i-th row and j-th column of $\boldsymbol{\Theta}$ and $\boldsymbol{H}_{\xi,k}$, respectively. With this, $f_k(\boldsymbol{\Theta}) = g(\sum_{i=1}^{2}\sum_{j=1}^{N}\theta_{i,j}h_{i,j}^{\xi}) - \beta_k$ is a composition of the convex nondecreasing function g with a linear function $\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{H}_{\xi,k}\right)$, which yields a convex function [42, Section 3.2.4].

A similar path can be taken to derive the conditions of matrix convexity of the UBSEP functions

$$f_k(\boldsymbol{\Theta}) = \sum_{\xi=1}^{2}\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}\ \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right)\right) - \rho_k. \tag{A-21}$$

Note, however, that function $g(x) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{P}{\sigma_w^2}}x\right) - \rho_k$ is convex for only $x \geq 0$. With this, the composed function $f_k(\boldsymbol{\Theta}) = g(\sum_{i=1}^{2}\sum_{j=1}^{N}\theta_{i,j}u_{i,j}^{\xi}) - \rho_k$, with $u_{i,j}^{\xi}$ denoting the entry on the i-th row and j-th column of $\boldsymbol{U}_{\xi,k}$ is convex in the regions where $\sum_{i=1}^{2}\sum_{j=1}^{N}\theta_{i,j}u_{i,j}^{\xi} \geq 0$. This implies that $f_k(\boldsymbol{\Theta})$ is matrix convex for $\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{U}_{\xi,k}\right) \geq 0,\ \text{for } k \in \mathcal{K},\ \xi \in \{1,2\}$.

B
MDDT-Bound on the Symbol Error Probability

In this section, we construct an MDDT-based bound on the SEP and prove that this bound is also an upper bound on the UBSEP. Based on the MDDT bound the PHMMDDT problem is proven to be a restriction of the PHSEP and PHUBSEP problems.

B.1
MDDT-based Bound as an Upper bound on the Union-Bound SEP

As mentioned in chapters 4 and 5, for a given k-th user's noiseless received signal y_k, the UBSEP relates to the SEP as

$$\begin{aligned} \mathrm{P}\left(\hat{s}_k \neq s_k | y_k\right) &= \mathrm{P}\left(z_k \in \mathcal{Z}_1 \cup \mathcal{Z}_2 | y_k\right) \\ &\leq \mathrm{P}\left(z_k \in \mathcal{Z}_1 | y_k\right) + \mathrm{P}\left(z_k \in \mathcal{Z}_2 | y_k\right) = \mathrm{P}_{\mathrm{ub}}(\hat{s}_k | y_k), \end{aligned} \tag{B-1}$$

where $\mathcal{Z}_1$ and $\mathcal{Z}_2$, depicted in Fig. 5.2. For computing the UBSEP of the k-th user the MDDTs, $d_{1,k}\left(y_k\right)$ and $d_{2,k}\left(y_k\right)$, are considered, such that

$$\mathrm{P}\left(z_k \in \mathcal{Z}_1 | y_k\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}\left(y_k\right)}{\sigma_w}\right), \quad \mathrm{P}\left(z_k \in \mathcal{Z}_2 | y_k\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}\left(y_k\right)}{\sigma_w}\right).$$

With this, the UBSEP of the k-th user reads as

$$\mathrm{P}\left(\hat{s}_k \neq s_k | y_k\right) \leq \mathrm{P}_{\mathrm{ub}}(\hat{s}_k | y_k) = \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}\left(y_k\right)}{\sigma_w}\right) + \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}\left(y_k\right)}{\sigma_w}\right). \tag{B-2}$$

An upper bound on the UBSEP of the k-th user can be constructed by considering the minimum between both MDDTs as

$$\begin{aligned} \mathrm{P}_{\mathrm{ub}}(\hat{s}_k | y_k) &= \frac{1}{2}\mathrm{erfc}\left(\frac{d_{1,k}\left(y_k\right)}{\sigma_w}\right) + \frac{1}{2}\mathrm{erfc}\left(\frac{d_{2,k}\left(y_k\right)}{\sigma_w}\right) \\ &\leq \mathrm{erfc}\left(\frac{\min\left(d_{1,k}\left(y_k\right), d_{2,k}\left(y_k\right)\right)}{\sigma_w}\right). \end{aligned} \tag{B-3}$$

Note that, $\min\left(d_{1,k}\left(y_k\right), d_{2,k}\left(y_k\right)\right)$ is the MDDT of the k-th user which, as mentioned in section 2.1, can be written for PSK data as

$$d_k\left(y_k\right) = \min\left(d_{1,k}\left(y_k\right), d_{2,k}\left(y_k\right)\right) = \mathrm{Re}\{s_k^* y_k\}\sin\phi - |\mathrm{Im}\{s_k^* y_k\}|\cos\phi, \quad \text{(B-4)}$$

where $\phi = {}^{\phi}/_{\alpha_s}$ and α_s denotes the PSK modulation order. With this, one can construct a MDDT-based bound on the SEP as

$$\mathrm{P}_{\mathrm{mddt}}(\hat{s}_k|y_k) = \mathrm{erfc}\left(\frac{d_k\left(y_k\right)}{\sigma_w}\right). \quad \text{(B-5)}$$

The relation between SEP, UBSEP, and $\mathrm{P}_{\mathrm{mddt}}(\hat{s}_k|y_k)$ reads as

$$\mathrm{P}\left(\hat{s}_k \neq s_k|y_k\right) \leq \mathrm{P}_{\mathrm{ub}}(\hat{s}_k|y_k) \leq \mathrm{P}_{\mathrm{mddt}}(\hat{s}_k|y_k). \quad \text{(B-6)}$$

B.2
MMDDT Problem as a Restriction of and PHUBSEP Problem

Considering equation (B-6) and the system model exposed in section 5.1 one can construct the high-resolution RIS power minimization problem with the MDDT bound as

$$\begin{aligned} &\min_{\boldsymbol{\theta},P} \; P \\ &\text{s.t.} \;\; \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \; \text{for } n \in \mathcal{N}, \quad P \geq 0, \\ &\mathrm{erfc}\left(\frac{d_k\left(\boldsymbol{\theta}, P\right)}{\sigma_w}\right) \leq \rho_k, \; \text{for } k \in \mathcal{K}. \end{aligned} \quad \text{(B-7)}$$

Applying the inverse complementary error function to the inequality constraints, the previous problem is rewritten as

$$\begin{aligned} &\min_{\boldsymbol{\theta},P} \; P \\ &\text{s.t.} \;\; \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \; \text{for } n \in \mathcal{N}, \quad P \geq 0, \\ &\qquad d_k\left(\boldsymbol{\theta}, P\right) \geq \sigma_w\left(\mathrm{erfc}^{-1}\left(\rho_k\right)\right), \; \text{for } k \in \mathcal{K}. \end{aligned} \quad \text{(B-8)}$$

Finally, problem (B-8) is written as exposed in section 5.2 with

$$\begin{aligned} &\min_{\boldsymbol{\theta},P} \; P \\ &\text{s.t.} \;\; \left|[\boldsymbol{\theta}]_n\right|^2 = 1, \; \text{for } n \in \mathcal{N}, \quad P \geq 0, \quad r_k = \sqrt{P}\boldsymbol{h}_k^H \boldsymbol{\theta} \mathrm{e}^{-j\arg(s_k)}, \\ &\qquad \mathrm{Re}\{r_k\}\sin\phi - |\mathrm{Im}\{r_k\}|\cos\phi \geq \alpha_k, \; \text{for } k \in \mathcal{K}, \end{aligned} \quad \text{(B-9)}$$

where $\alpha_k = \sigma_w \left(\text{erfc}^{-1}\left(\rho_k\right)\right)$. Since the MDDT constraint functions are upper bounds on the UBSEP constraints, one can understand the problem from (B-9) as a restricted version of the proposed PHUBSEP formulation. This implies that, for attaining the same SEP requisite the optimal transmit power minimization under MDDT constraints is greater or equal to the optimal transmit power of the proposed approach.

C
SNR Definition

In this section, we derive the SNR definitions for different beamforming scenarios. By analyzing these approaches, we illustrate how each technique impacts SNR focusing the signal toward the intended receivers. The derived expressions serve as a foundation for the SNR definitions utilized in the numerical results sections of the thesis. The average received SNR is defined as

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{E}\left\{\mathrm{P}_{\mathrm{R},k}\right\}}{\mathrm{E}\left\{\mathrm{P}_{\mathrm{N},k}\right\}}, \tag{C-1}$$

where $\mathrm{P}_{\mathrm{R},k}$ is the received signal power of the k-th user and $\mathrm{P}_{\mathrm{N},k}$ is the k-th user's noise power. For the system models considered in this thesis, the received SNR can be rewritten as

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{E}\left\{\|\boldsymbol{h}_k\boldsymbol{x}\|_2^2\right\}}{\mathrm{E}\left\{\|w_k\|_2^2\right\}} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{E}\left\{\boldsymbol{x}^H\boldsymbol{h}_k^H\boldsymbol{h}_k\boldsymbol{x}\right\}}{\sigma_w^2}. \tag{C-2}$$

For given channels $\boldsymbol{h}_k$ for $k \in \{1,\ldots,K\}$, the SNR expression can be further simplified as

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{tr}\left(\boldsymbol{h}_k\mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}\boldsymbol{h}_k^H\right)}{\sigma_w^2} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{tr}\left(\boldsymbol{h}_k\boldsymbol{C}_x\boldsymbol{h}_k^H\right)}{\sigma_w^2}, \tag{C-3}$$

where $\boldsymbol{C}_x = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}$. The definition of the average SNR in (C-3) considers deterministic channels and should be extended for $\boldsymbol{h}_k$ for $k \in \{1,\ldots,K\}$ as random vectors. To this end, the average SNR is written as

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{E}\left\{\mathrm{tr}\left(\boldsymbol{h}_k\boldsymbol{C}_x\boldsymbol{h}_k^H\right)\right\}}{\sigma_w^2}. \tag{C-4}$$

As can be concluded from (C-4), different ways of defining the transmit vector $\boldsymbol{x}$ lead to different received SNRs. In what follows we expose the methods that imply the SNR definitions utilized in the numerical results of this study.

C.1
Average Receive SNR with a Generic Transmitter

Applying the SNR definition in (C-4) for a generic transmitter with $\boldsymbol{C}_x = \mathrm{P_A}\boldsymbol{I}$ yields

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\mathrm{P_A}\ \mathrm{E}\left\{\mathrm{tr}\left(\boldsymbol{h}_k\boldsymbol{I}\boldsymbol{h}_k^H\right)\right\}}{\sigma_w^2}.$$

Considering $\boldsymbol{h}_{k,m} \sim \mathcal{CN}(0,1)$, for $k \in \{1,\ldots,K\}$, and, $m \in \{1,\ldots,M\}$, yields

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{M\mathrm{P_A}}{\sigma_w^2} = \frac{M\mathrm{P_A}}{\sigma_w^2}. \tag{C-5}$$

C.2
Maximum Average Receive SNR

The received SNR can be maximized with the application of the MRT beamformer [62, Section 7.3.1]. With this, for a given channel $\boldsymbol{H}$, the transmit vector is read as $\boldsymbol{x} = \sqrt{\frac{M\mathrm{P_A}}{\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{H}^H\right)}}\boldsymbol{H}^H\boldsymbol{s}$, which implies $\boldsymbol{C}_x = M\mathrm{P_A}\left(\frac{\boldsymbol{H}^H\boldsymbol{C}_s\boldsymbol{H}}{\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{H}^H\right)}\right)$, with $\boldsymbol{C}_s = \mathrm{E}\left\{\boldsymbol{s}\boldsymbol{s}^H\right\} = \boldsymbol{I}$. Applying MRT to the definition in (C-4) yields

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{M\mathrm{P_A}\ \mathrm{E}\left\{\frac{\mathrm{tr}\left(\boldsymbol{h}_k\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{h}_k^H\right)}{\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{H}^H\right)}\right\}}{\sigma_w^2}.$$

Considering $\boldsymbol{h}_{k,m} \sim \mathcal{CN}(0,1)$, for $k \in \{1,\ldots,K\}$ and $m \in \{1,\ldots,M\}$ yields

$$\mathrm{SNR} = \frac{1}{K}\sum_{k=1}^{K}\frac{M^2\mathrm{P_A}}{\sigma_w^2} = \frac{M^2\mathrm{P_A}}{\sigma_w^2}. \tag{C-6}$$