Fixed Point Algorithms and Superiorization in Communication Systems

vorgelegt von
M. Sc.
Jochen Fink

an der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Giuseppe Caire, PhD
Gutachter: Prof. Dr.-Ing. Sławomir Stańczak
Gutachter: Prof. Isao Yamada, PhD
Gutachter: Prof. Nikolaos Sidiropoulos, PhD

Tag der wissenschaftlichen Aussprache: 22. Juli 2022

Berlin 2022

Abstract

This work studies fixed point algorithms and superiorization in wireless communication systems. Modern wireless systems perform a great variety of signal processing tasks, including channel estimation, precoding, combining, signal detection, and peak-to-average power ratio (PAPR) reduction. The growing demand for mobile data traffic calls for systems with higher bandwidths, larger antenna arrays, and the capability to serve an increasing number of devices. As a result, the dimensions of various optimization problems arising in wireless networks are growing continuously, which increases the computational cost of algorithmic solutions. Hence, scalable algorithms with low complexity are vital to meet both real-time requirements and power budgets. For various types of problems, finding optimal solutions can be too computationally demanding in practice. In this case, it is desirable to strike a balance between performance and complexity. Ranging between feasibility seeking and constrained optimization, the superiorization methodology is a promising means of achieving this trade-off. It relies on the concept of bounded perturbation resilience of an iterative algorithm. A feasibility-seeking fixed point algorithm is said to be bounded perturbation resilient if its convergence to a fixed point can be guaranteed even if certain perturbations are added to the iterate in each step. In this case, the superiorization methodology can be used to define a sequence of perturbations leading to a reduced (not necessarily minimal) value of a given objective function. Compared to exact constrained minimization, superiorization often results in a lower computational cost.

In this thesis, we investigate the bounded perturbation resilience of several algorithmic frameworks, including the well-known projections onto convex sets (POCS) algorithm, the adaptive projected subgradient method (APSM), and certain extrapolated projection methods. By doing so, we enable their use as *basic algorithms* for superiorization. We propose an algorithm for the nonconvex multi-group multicast beamforming problem based on a perturbed POCS algorithm. The proposed perturbations simultaneously reduce two objective functions, one of which is nonconvex. Then we harness the bounded perturbation resilience of the APSM by proposing an algorithm for detection in multiple-input multiple-output (MIMO) systems based on a superiorized APSM. We also devise a deep unfolded version of this algorithm, in which the design parameters are learned using a stochastic

gradient descent method. Moreover, we propose online algorithms for estimating and tracking time-varying channels with hybrid-beamforming architectures based on an APSM. The proposed channel estimation algorithms can compensate for random delay and phase variations in wideband channels. Furthermore, we devise a data-driven analog combining policy. Finally, we propose extrapolated projection methods for PAPR reduction. We devise perturbations that aim at incorporating a nonconvex constraint set, which allows the simultaneous use of certain subcarriers for channel estimation and peak cancellation. Simulations at the end of each chapter show that the proposed methods can outperform state-of-the-art techniques, while often resulting in a reduced computational cost.

Zusammenfassung

Diese Arbeit untersucht Fixpunktalgorithmen und *superiorization* in drahtlosen Kommunikationssystemen. Moderne Funksysteme bewältigen eine Vielzahl von Signalverarbeitungsaufgaben, wie beispielsweise Kanalschätzung, Precoding, Combining, Signaldetektion oder der Reduzierung des Verhältnisses von Spitzen- zu Durchschnittsleistung (peak-to-average power ratio, PAPR) von Sendesignalen. Die wachsende Nachfrage nach mobilem Datenverkehr macht Systeme mit höheren Bandbreiten und größeren Antennenarrays erforderlich, die darüber hinaus in der Lage sind, eine wachsende Anzahl von Geräten zu versorgen. Infolgedessen nehmen die Dimensionen vieler Optimierungsprobleme in drahtlosen Netzen ständig zu, was den Rechenaufwand für ihre algorithmische Lösung erhöht. Daher sind skalierbare Algorithmen mit geringer Komplexität von entscheidender Bedeutung, um sowohl Echtzeitanforderungen zu erfüllen als auch Energiebudgets einzuhalten. Die Suche nach einer optimalen Lösung kann in der Praxis für bestimmte Arten von Optimierungsproblemen zu rechenintensiv sein. In diesem Fall ist es von Interesse, einen Kompromiss zwischen Approximationsgüte und Rechenaufwand zu finden. Die *superiorization methodology* stellt ein vielversprechendes Mittel dar, um diesen Trade-off zu erreichen. Hierbei handelt es sich um ein algorithmisches Framework, welches einen Mittelweg zwischen der Lösung von Zulässigkeitsproblemen und restringierter Optimierung darstellt. Sie stützt sich auf das Konzept der *bounded perturbation resilience* iterativer Algorithmen. Ein Fixpunktalgorithmus zur Lösung eines Zulässigkeitsproblems gilt als *bounded perturbation resilient*, wenn dessen Konvergenz zu einem Fixpunkt auch dann garantiert werden kann, wenn in jeder Iteration bestimmte Störungen addiert werden. In diesem Fall kann die *superiorization methodology* verwendet werden, um eine Folge von Störungen zu definieren, welche zu einem reduzierten (nicht notwendigerweise minimalen) Wert einer gegebenen Zielfunktion führen. Auf diese Weise geht *superiorization* oft in einem deutlich geringeren Rechenaufwand einher als Verfahren zur exakten Lösung von restringierten Optimierungsproblemen.

In dieser Arbeit untersuchen wir die *bounded perturbation resilience* verschiedener Algorithmen, darunter der bekannte *projections onto convex sets (POCS)*-Algorithmus, die *adaptive projected subgradient method (APSM)* sowie extrapolierte Projektionsmethoden. Auf diese Weise ermöglichen wir deren Verwendung als sogenannte *basic algorithms* für die

superiorization. Wir entwickeln einen Algorithmus für das nichtkonvexe *multi-group multicast beamforming*-Problem, der auf einem gestörten POCS-Algorithmus basiert. Die vorgeschlagenen Störungen reduzieren gleichzeitig zwei Zielfunktionen, von denen eine nichtkonvex ist. Anschließend nutzen wir die *bounded perturbation resilience* der APSM, indem wir einen Algorithmus zur Detektion in *multiple-input multiple-output (MIMO)*-Systemen vorschlagen, welcher auf einer superiorisierten APSM beruht. Zusätzlich entwickeln wir mittels *deep unfolding* eine Version dieses Algorithmus, deren Designparameter mit Hilfe eines stochastischen Gradientenabstiegsverfahrens gelernt werden. Darüber hinaus schlagen wir Online-Algorithmen zur Schätzung und Verfolgung (Tracking) von zeitvarianten Kanälen mit *hybrid beamforming*-Architekturen vor, die auf einer APSM basieren. Die vorgeschlagenen Kanalschätzverfahren können zufällige Laufzeit- und Phasenschwankungen in Breitbandkanälen kompensieren. Darüber hinaus entwickeln wir eine datengesteuerte *analog combining policy*. Schlussendlich verwenden wir extrapolierte Projektionsmethoden zur PAPR-Reduktion. Wir schlagen Störungen vor, die darauf abzielen, eine nichtkonvexe Nebenbedingung einzubeziehen, welche die gleichzeitige Verwendung bestimmter Unterträger für die Kanalschätzung und die Auslöschung von Signalspitzen ermöglicht. Simulationen am Ende der jeweiligen Kapitel zeigen, dass die vorgeschlagenen Methoden den Stand der Technik übertreffen können, wobei sie oft mit geringeren Rechenkosten verbunden sind.

Acknowledgments

First and foremost, I would like to thank my advisor Prof. Dr.-Ing. Sławomir Stańczak for his guidance throughout the entire PhD period, and for allowing me to focus on the topics that interested me the most. I would like to express special thanks to Dr. Renato Cavalcante for teaching me the foundations this thesis is based on, and for his endless patience to discuss detailed questions that emerged during my work. Also, I would like to thank the coauthors of my publications for our collaboration.

I am very grateful to Prof. Isao Yamada and Prof. Nikolaos Sidiropoulos for serving as referees of this dissertation. Moreover, I would like to thank Prof. Giuseppe Caire for chairing the dissertation committee.

My friends and colleagues at Technische Universität Berlin and Fraunhofer Heinrich-Hertz-Institute deserve thanks for the fruitful discussions on research-related and -unrelated topics. Finally, I would like thank all my friends and family – especially those involved on a daily basis – for their constant encouragement and support during my work on this thesis.

Contents

1. Introduction **1**

1.1. Motivation 1

1.2. Contributions and Outline 2

1.3. Notation and Preliminaries 5

2. Bounded Perturbation Resilience of Iterative Algorithms **15**

2.1. Projections Onto Convex Sets 18

2.2. Adaptive Projected Subgradient Method 21

2.3. Extrapolated Alternating Projection Methods 32

3. Multicast Beamforming **41**

3.1. Introduction 41

3.2. Problem Statement 43

3.2.1. System Model and Original Problem 43

3.2.2. Problem Formulation in a Real Hilbert Space 45

3.3. Algorithmic Solutions 49

3.3.1. Feasibility-Seeking Basic Algorithm 50

3.3.2. Proposed Perturbations 52

3.3.3. Convergence of the Proposed Algorithm 57

3.3.4. Relation to the Superiorization Methodology 58

3.3.5. Summary of the Proposed Algorithm 59

3.4. Numerical Results 60

3.4.1. Performance vs. Computation Time 63

3.4.2. Varying number of antennas 65

3.4.3. Varying number of users 66

3.4.4. Varying Target SINR 66

3.5. Conclusion and Final Remarks 67

4. MIMO Detection **71**

4.1. Introduction 71

4.2. Problem Statement 72

4.3. Algorithmic Solutions . . . 73
4.3.1. An Adaptive Projected Subgradient Method for MIMO Detection . 74
4.3.2. Superiorization . . . 75
4.3.3. Convergence of the Proposed Algorithms . . . 77
4.3.4. Summary of the Proposed Algorithms . . . 79
4.3.5. Deep-Unfolded APSM for MIMO Detection . . . 79
4.4. Numerical Results . . . 80
4.4.1. Performance of the Untrained Detectors . . . 81
4.4.2. Performance of a Deep-Unfolded APSM Detector . . . 82
4.5. Conclusion . . . 84

5. Online Channel Estimation for Hybrid Beamforming Architectures 85
5.1. Introduction . . . 85
5.2. Problem Statement . . . 87
5.3. Algorithmic Solutions . . . 88
5.3.1. Channel Estimation using Existing Batch Methods . . . 88
5.3.2. Online Algorithm for Channel Estimation and Tracking . . . 89
5.3.3. Weighted ℓ_1-Norm Regularization via Superiorization . . . 94
5.3.4. A Sparse Representation of the Channel . . . 96
5.3.5. Delay and Phase Compensation . . . 97
5.3.6. Data-driven Analog Combining Scheme . . . 102
5.3.7. Summary of the Proposed Channel Estimation Scheme . . . 104
5.4. Numerical Results . . . 104
5.4.1. Narrowband Channels . . . 106
5.4.2. Performance of the Proposed Delay and Phase Estimators . . . 107
5.4.3. Wideband Channels . . . 109
5.5. Conclusion . . . 110

6. PAPR Reduction 111
6.1. Introduction . . . 111
6.2. System Model and Basic Definitions . . . 112
6.3. PAPR Reduction by Set Theoretic Estimation . . . 113
6.3.1. Problem Statement . . . 114
6.3.2. Some Relevant Constraint Sets . . . 115
6.4. Algorithmic Solutions . . . 120
6.4.1. Previous Methods . . . 121
6.4.2. Extrapolated Projection Method for Tone Reservation . . . 121

6.4.3. PAPR Reduction with Arbitrary Constraints 122
6.4.4. Phaseless Pilot Reuse by Superiorization 123
6.4.5. Practical Aspects: Feasibility, Computational Efficiency 125
6.5. Numerical Results 126
6.6. Conclusion 127

7. Discussion and Outlook **129**

Appendices **131**

A. Multicast Beamforming **133**
A.1. Proof of Remark 3.1 133
A.2. Proof of Proposition 3.1 133
A.3. Technical Lemmas Required to Prove Proposition 3.2(a) 134

B. MIMO Detection **137**

C. Online Channel Estimation **141**
C.1. Proof of Proposition 5.4 141
C.2. Proof of Proposition 5.7 143

Publication List **147**

Bibliography **149**

Copyright Information

Abbreviations

3GPP	3rd Generation Partnership Project
4G	fourth generation
ACE	active constellation extension
ADMM	alternating direction method of multipliers
AMP	approximate message passing
APSM	adaptive projected subgradient method
BER	bit error ratio
BS	base station
CCDF	complementary cumulative distribution function
CCP	convex-concave procedure
CPU	central processing unit
CRLB	Cramér-Rao lower bound
CSI	channel state information
DFT	discrete Fourier transform
EAPM	extrapolated alternating projection method
EM	electromagnetic
EPPM	extrapolated parallel projection method
EVM	error vector magnitude
FFT	fast Fourier transform
GPR	Gurin-Polyak-Raik
GPU	graphics processing unit
i.i.d.	independent and identically distributed
ICF	iterative clipping and filtering
IO-LAMA	individually-optimal large-MIMO AMP
IoT	Internet of Things
LASSO	least absolute shrinkage and selection operator
LMMSE	linear minimum mean square error
MIMO	multiple-input multiple-output
ML	maximum likelihood

MMF	max-min fair
MMPC	multigroup multicast power control
NMSE	normalized mean squared error
OAMP	orthogonal approximate message passing
OFDM	orthogonal frequency division multiplexing
PAPR	peak-to-average power ratio
POCS	projections onto convex sets
PSD	positive semidefinite
QAM	quadrature amplitude modulation
QCQP	quadratically constrained quadratic program
QoS	quality of service
QPSK	quadrature phase shift keying
RF	radio frequency
SCA	successive convex approximation
SCF	simplified clipping and filtering
SDP	semidefinite program
SER	symbol error ratio
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SVD	singular value decomposition
TR	tone reservation
UE	user equipment
ULA	uniform linear array
UPA	uniform planar array

1. Introduction

1.1. Motivation

Digital signal processing is ubiquitous in all parts of modern wireless communication systems. While linear processing was prevalent in the early days of digital communications, the rising demand for mobile communication has driven the development of increasingly complex system architectures, calling for more sophisticated nonlinear processing techniques. Over the past three decades, growing computational power has facilitated the application of convex optimization techniques to approach a vast variety of problems in signal processing and communications. Although convex optimization problems typically admit polynomial-time solutions, strict latency requirements combined with ever-increasing system dimensions of wireless networks can hinder the use of general-purpose convex solvers in practice. Moreover, novel communication schemes often involve nonconvex optimization problems, many of which are known to be NP-hard and hence do not admit polynomial-time solutions (unless NP=P).

When finding an optimal solution is not possible within the available time budget, one typically has to resort to suboptimal approximation techniques. In this case, fixed point algorithms are of particular interest. Given some initial point, a fixed point algorithm produces a sequence of estimates by iteratively applying certain mappings. In some cases, this sequence provably converges to an optimal solution. Apart from that, it may exhibit weaker guarantees including monotone approximation of a solution or convergence to a feasible point. As fixed point methods often have constant computational cost throughout all iterations, they naturally allow to strike a balance between performance and complexity by terminating the algorithm after a certain number of iterations.

The *superiorization methodology* [CDH10] is an algorithmic framework that ranges between constrained minimization and feasibility seeking [Cen15]. By interleaving fixed point iterations (with respect to the feasible set) with perturbations that aim at reducing the objective value, superiorization can achieve considerable complexity reduction compared to iterative techniques for constrained minimization [CDH$^+$14]. Whereas the goal of constrained minimization is to find a feasible point (i.e., a point satisfying all constraints) for

which the objective value is minimal, superiorization[1] typically builds upon a simple fixed point algorithm that produces a sequence of points provably converging to a feasible point. This fixed point algorithm serves as the so-called *basic algorithm*. For a given objective function, the superiorization methodology generates a *superiorized version of the basic algorithm* by adding small perturbations in each iteration with the intent to find a feasible point with reduced (not necessarily minimal) objective value. A central concept in the superiorization methodology is the *bounded perturbation resilience* of iterative algorithms. Feasibility seeking algorithms having this property are guaranteed to converge to a feasible point even if summable perturbations are added to their iterates in each iteration. Once bounded perturbation resilience of an algorithm has been established, the convergence of superiorized versions of this algorithm is guaranteed automatically.

1.2. Contributions and Outline

This thesis aims at studying the potential of iterative algorithms to tackle a range of problems arising in wireless communication systems. One particular goal is to develop efficient algorithms with guaranteed convergence that can be used to approximate solutions to (NP-hard) nonconvex optimization problems. Some of the proposed techniques build upon the concepts of bounded perturbation resilience and superiorization. Therefore, a second goal of this thesis is to show the bounded perturbation resilience of algorithmic frameworks, from which a variety of practical algorithms can be derived. The main contributions of this thesis are organized as follows.

In **Chapter 2**, we investigate the bounded perturbation resilience of iterative algorithms that will be used in the subsequent chapters. We derive convergence criteria for perturbed versions of the projections onto convex sets (POCS) algorithm (see, e.g., [SY98]), the adaptive projected subgradient method (APSM) [YO05], and the extrapolated alternating projection methods proposed in [GPR67] and [BCK06], respectively. In particular, we derive conditions for weak and strong convergence in (possibly infinite dimensional) real Hilbert spaces.
Some of the results on the bounded perturbation resilience of POCS in Section 2.1 have been presented in the conference publication [3] and the journal paper [7]. The results on the bounded perturbation resilience of the APSM in Section 2.2 have been partially presented in the conference paper [10], which is to appear, and they are covered in full by the journal paper [11], which has been submitted for publication.

In **Chapter 3**, we propose a superiorized version of POCS to approximate solutions to

[1] In this thesis, we only consider *weak superiorization*, in which the underlying feasibility problem is assumed to be consistent. See [Cen15] for a detailed overview of weak and strong superiorization.

the nonconvex multi-group multicast beamforming problem with quality of service (QoS) constraints and per-antenna power constraints. We apply POCS to a convex relaxation of the original problem, so that we can guarantee its convergence. Then we introduce perturbations that aim at simultaneously reducing two superiorization objectives, one of which is nonconvex. By proving that the proposed perturbations are bounded, we can guarantee the convergence of the superiorized algorithm.
The work in Chapter 3 has been partially presented in the conference paper [3] and to its full extent in the journal publication [7].

In **Chapter 4**, we propose an iterative algorithm with low computational complexity for multiple-input multiple-output (MIMO) detection. We devise a variant of the APSM to minimize asymptotically a sequence of convex cost functions over the convex hull of the constellation constraint. Then we introduce perturbations that steer the iterates of the APSM towards the discrete constellation constraint set. Unlike existing iterative MIMO detectors based on approximate message passing (AMP), this set-theoretic approach does not impose any assumptions on the channel matrix. Thus we can guarantee its convergence even for realistic channel models. Moreover, we show that the proposed method gives rise to a new type of deep unfolding-based MIMO detector.
The major part of the work in Chapter 4 has been presented in the conference paper [10] and the journal paper [11].

In **Chapter 5**, we consider the APSM in an online setting. We propose an algorithm for online channel estimation and tracking with hybrid beamforming architectures. Additionally, inspired by the superiorization methodology, we propose a heuristic that implements a regularization to promote sparsity of the channel estimates in delay- and angular domain. In contrast to many existing approaches, our proposed technique does not assume block fading, so it can be used in realistic settings in which the channel coefficients change continuously over time. The online structure of the proposed algorithms allows us to combine them with mechanisms that compensate for random delay and phase variations, and that adapt the analog beamforming policy based on previous channel observations.
Some parts of the work in Chapter 5 have been published in the conference paper [6].

In **Chapter 6**, we address the peak-to-average power ratio (PAPR) problem in orthogonal frequency division multiplexing (OFDM) systems, which is caused by the superposition of a large number of subcarriers. We investigate the potential of extrapolated alternating projection methods to improve the performance and to reduce the complexity compared to the POCS algorithm, which has been used in several previous works. Furthermore, we introduce a nonconvex constraint set that allows to reduce the PAPR by modifying the phase of subcarriers that are used for channel estimation with phaseless pilots [WBJ15a, WBJ15b]. Exploiting the bounded perturbation resilience of the proposed

algorithms, we introduce perturbations that incorporate this nonconvex constraint.
The work in Chapter 6 is an extended version of the results published in the conference paper [1].

Historical Notes

Since the early work by von Neumann [vN33], the POCS algorithm considered in Section 2.1 has been rediscovered and refined many times: The alternating projection method in [vN33] finds a point in the intersection of two subspaces. As pointed out in [NS06], this result was rediscovered independently in [Aro50], [Nak53], and [Wie55]. Kaczmarcz [Kac37] used cyclic projections onto hyperplanes to solve a system of linear equations. Agmon, Motzkin and Schoenberg [Agm54, MS54] generalized this cyclic projection method by introducing relaxed projections, and by extending the convergence proof to half-spaces. Eremin [Ere65] further generalized these results to closed convex sets in finite dimensional spaces, Bregman [Bre65] showed the weak convergence of the sequence generated by this method in the general case, and Gurin, Polyak and Raik [GPR67] derived sufficient conditions for strong convergence.[2]

By now, the POCS algorithm has become ubiquitous in engineering. Exemplary applications include medical imaging [YW82, SS84, SKPJ04, GMZ$^+$16, TBL$^+$18], seismic imaging [GCL$^+$10, YGC12, GWC$^+$15, ZZZ$^+$20], road design [BK15], and distributed learning [AMB17]. Applications in wireless communications include localization in wireless sensor networks [BH06, RSS06], PAPR reduction [GP97, Arm02, KJ03], multiuser detection [VLH13], and radar waveform design [EBL$^+$15], among many others. The APSM covered in Section 2.2 has been used for multiaccess interference suppression [CY08], acoustic feedback cancellation [YY06, WZQZ10], robust beamforming [STY09], robust subspace tracking [CKT14], online radio-map reconstruction [KCV$^+$15], kernel-based online classification [STY08], PAPR reduction [CY09], distributed learning in diffusion networks [CYM09, CST11, SYCD18], and adaptive symbol detection [ACYS18, ACYS20]. The extrapolated alternating projection methods addressed in Section 2.3 can be used to solve feasibility problems with two closed convex sets. Hence they are applicable to many of the applications of POCS listed above. Applications of extrapolated alternating projection methods in the literature include image reconstruction [CCC$^+$12] and uplink-downlink conversion of spatial covariance matrices [MCS18, MCS21]. Owing to the widespread use of the algorithmic frameworks under consideration, the theoretical results in Chapter 2 can be valuable for practical applications beyond those presented in this thesis.

[2] Further details regarding the history of projection methods can be found in [Com93, BB96, CC15].

1.3. Notation and Preliminaries

Unless otherwise specified, lowercase letters denote scalars, lowercase letters in bold typeface denote vectors, uppercase letters in bold typeface denote matrices. The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$, $\mathbb{R}$ and $\mathbb{C}$ denote the nonnegative integers, integers, nonnegative real numbers, positive real numbers, real numbers and complex numbers, respectively. The set of summable sequences in $\mathbb{R}_+$ is denoted by $\ell_+^1(\mathbb{N})$. The imaginary unit, i.e., the solution in $\mathbb{C}$ to $j^2 = -1$, is denoted by j. The real part, imaginary part, magnitude, angle, and complex conjugate of a complex number $z \in \mathbb{C}$ are denoted by $\mathrm{Re}\{z\}$, $\mathrm{Im}\{z\}$, $|z|$, $\angle z$, and z^*, respectively. To complex vectors and matrices, these functions are applied entry-wise. The sign and the nonnegative part of a real number $x \in \mathbb{R}$ are denoted by $\mathrm{sgn}(x)$ and $(x)_+ := \max\{x, 0\}$, respectively. We write $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$ for the floor function, which is given by $\lfloor x \rfloor = \max\{ m \in \mathbb{Z} \mid m \leq x\}$. The cardinality of a discrete set $\mathcal{I}$ is denoted by $|\mathcal{I}|$. Given two sets $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \subset \mathcal{B}$ or $\mathcal{B} \supset \mathcal{A}$ if $(\forall \mathbf{x} \in \mathcal{A})\ \mathbf{x} \in \mathcal{B}$. Given two statements A and B, we write $A \implies B$ if A is sufficient for B, and $A \iff B$ if A is necessary and sufficient for B.

The transpose, Hermitian transpose, inverse, and Moore-Penrose inverse of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$, respectively. An orthogonal matrix is a matrix $\mathbf{Q} \in \mathbb{R}^{N \times N}$ satisfying $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and a unitary matrix is a matrix $\mathbf{U} \in \mathbb{C}^{N \times N}$ satisfying $\mathbf{U}^{-1} = \mathbf{U}^H$. The Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ and their direct sum is denoted by $\mathbf{A} \oplus \mathbf{B}$. We write I for the identity operator and $\mathbf{I}_N$ for the $N \times N$-identity matrix. If the dimension is clear from the context, we sometimes omit the subscript N. The all-zero vector or matrix is denoted by $\mathbf{0}$ and the ith Cartesian unit vector is denoted by $\mathbf{e}_i$, where the dimension of the space will be clear from the context. The singular values of a matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ are denoted by $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_N(\mathbf{A})$. For square matrices $\mathbf{A}$ we define $\mathrm{diag}(\mathbf{A})$ to be the column vector composed of the diagonal entries of $\mathbf{A}$, and for row or column vectors $\mathbf{a}$ we define $\mathrm{diag}(\mathbf{a})$ to be a square diagonal matrix having $\mathbf{a}$ as its diagonal. For a Hermitian or symmetric matrix $\mathbf{A}$, write $\mathbf{A} \succcurlyeq \mathbf{0}$ if $\mathbf{A}$ is positive semidefinite (PSD). The principal square root of a PSD matrix $\mathbf{A} \succcurlyeq \mathbf{0}$ is denoted by $\mathbf{A}^{\frac{1}{2}}$. The submatrix consisting of the entries in rows i_1 through i_2 and columns k_1 through k_2 of a matrix $\mathbf{A}$ is written as $\mathbf{A}_{i_1:i_2,k_1:k_2}$. The vectorization operator that stacks the columns of a matrix into a column vector is denoted by $\mathrm{vec}(\cdot)$. For $p \geq 1$, we write $\|\mathbf{x}\|_p$ for the p-norm of a column vector $\mathbf{x}$, where $\|\cdot\|_2$ is the Euclidean norm. In compliance with established literature, we sometimes refer to $\|\cdot\|_1$ as ℓ_1-norm, even if the underlying space is finite-dimensional. The Frobenius norm, spectral norm, and nuclear norm of a matrix $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_2$, and $\|\mathbf{A}\|_*$. We denote by $\mathcal{U}(a, b)$ the uniform distribution over an interval $[a, b] \subset \mathbb{R}$. The multivariate normal distribution with

mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance $\boldsymbol{\Gamma} \succcurlyeq \mathbf{0} \in \mathbb{R}^{N \times N}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ and the circularly-symmetric multivariate complex normal distribution with mean $\boldsymbol{\mu} \in \mathbb{C}^N$ and covariance $\boldsymbol{\Gamma} \succcurlyeq \mathbf{0} \in \mathbb{C}^{N \times N}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$. The expected value of a random variable x is denoted by $\mathbb{E}[x]$. We write arg min and arg max for the sets of points with minimal and maximal objective value, respectively. These sets are singletons if the minimum (or maximum) of the respective function is unique.

Throughout this thesis, we denote by $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ a real Hilbert space with induced norm $(\forall \mathbf{x} \in \mathcal{H})\ \|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Given a function $f : \mathcal{H} \to \mathbb{R}$ and a constant $c \in \mathbb{R}$, we denote by $\operatorname{lev}_{\leq c} := \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) \leq c\}$ a sublevel set of f. We say that a function f is closed if all of its sublevel sets are closed. Moreover, we say that a function $f : \mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\}$ is coercive if $f(\mathbf{x}) \to +\infty$ whenever $\|\mathbf{x}\| \to +\infty$. The distance between two points $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ in a real Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, where $\| \cdot \|$ is the norm induced by the inner product $\langle \cdot, \cdot \rangle$. The distance between a point $\mathbf{x} \in \mathcal{H}$ and a nonempty set $\mathcal{C} \subset \mathcal{H}$ is defined as $d(\mathbf{x}, \mathcal{C}) = \inf_{\mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|$. Following [BCL02], we define the projection of a point $\mathbf{x} \in \mathcal{H}$ onto a nonempty subset $\mathcal{C} \subset \mathcal{H}$ as the set

$$\Pi_{\mathcal{C}}(\mathbf{x}) = \{\mathbf{y} \in \mathcal{C} \mid d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathcal{C})\},$$

and we denote by $P_{\mathcal{C}} : \mathcal{H} \to \mathcal{C}$ an arbitrary but fixed selection of $\Pi_{\mathcal{C}}$, i.e., $(\forall \mathbf{x} \in \mathcal{H})$ $P_{\mathcal{C}}(\mathbf{x}) \in \Pi_{\mathcal{C}}(\mathbf{x})$. If $\mathcal{C}$ is nonempty, closed, and convex, the set $\Pi_{\mathcal{C}}(\mathbf{x})$ is a singleton for all $\mathbf{x} \in \mathcal{H}$, so $\Pi_{\mathcal{C}}$ has a unique selection $P_{\mathcal{C}}$, which itself is called a projector. For nonempty closed nonconvex sets $\mathcal{C}$ in finite-dimensional Hilbert spaces, $\Pi_{\mathcal{C}}(\mathbf{x})$ is nonempty for all $\mathbf{x} \in \mathcal{H}$, although it is not in general a singleton. Nevertheless, we will refer to the selection $P_{\mathcal{C}}$ as the projector, as the distinction from the set-valued mapping $\Pi_{\mathcal{C}}$ will always be clear.

Fact 1.1. *(Projection Theorem) [BC11, Theorem 3.16] Let $\mathcal{C} \subset \mathcal{H}$ be a nonempty closed convex set. Then $(\forall \mathbf{x} \in \mathcal{H})$ $\Pi_{\mathcal{C}}(\mathbf{x})$ is a singleton, and $(\forall \mathbf{p} \in \mathcal{H})$ its unique element $P_{\mathcal{C}}(\mathbf{x})$ satisfies*

$$\mathbf{p} = P_{\mathcal{C}}(\mathbf{x}) \iff [\mathbf{p} \in \mathcal{C} \text{ and } (\forall \mathbf{y} \in \mathcal{C})\ \langle \mathbf{y} - \mathbf{p}, \mathbf{x} - \mathbf{p} \rangle \leq 0].$$

The following property will be used in Chapters 3 and 6 to derive closed-form expressions for the projection onto the intersection of certain closed convex subsets.

Proposition 1.1. *Let $\mathcal{I} = \{1, \ldots, K\}$ be a finite set and let $(\forall i \in \mathcal{I})$ $\mathcal{C}_i \subset \mathcal{H}$ be a nonempty closed convex set. If $\mathcal{C} := \bigcap_{i \in \mathcal{I}} \mathcal{C}_i \neq \emptyset$ and $(\forall \mathbf{x} \in \mathcal{H})(\forall i \in \mathcal{I})(\forall k \in \mathcal{I} \setminus \{i\})$*

$\langle P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_k}(\mathbf{x}) - \mathbf{x}\rangle = 0$, *the projection of a point* $\mathbf{x} \in \mathcal{H}$ *onto the intersection* $\mathcal{C}$ *satisfies*

$$P_{\mathcal{C}}(\mathbf{x}) \overset{(a)}{=} \mathbf{x} + \sum_{i\in\mathcal{I}}(P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}) \overset{(b)}{=} P_{\mathcal{C}_{i_1}} \cdots P_{\mathcal{C}_{i_K}}(\mathbf{x}),$$

where $i_1, \ldots, i_K$ *is an arbitrary permutation of the indices in* $\mathcal{I}$.

Proof.

(a) Define $\mathbf{p} := \mathbf{x} + \sum_{i\in\mathcal{I}}(P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x})$. *Then* $(\forall \mathbf{y} \in \mathcal{C})$ *it holds that*

$$\begin{aligned}
\langle \mathbf{y} - \mathbf{p}, \mathbf{x} - \mathbf{p}\rangle &= \langle \mathbf{y} - \mathbf{x} - \sum_{k\in\mathcal{I}}(P_{\mathcal{C}_k}(\mathbf{x}) - \mathbf{x}), -\sum_{i\in\mathcal{I}}(P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x})\rangle \\
&\overset{(i)}{=} -\sum_{i\in\mathcal{I}}\left(\langle \mathbf{y} - \mathbf{x}, P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}\rangle - \sum_{k\in\mathcal{I}}\langle P_{\mathcal{C}_k}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}\rangle\right) \\
&\overset{(ii)}{=} -\sum_{i\in\mathcal{I}}\left(\langle \mathbf{y} - \mathbf{x}, P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}\rangle - \langle P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}\rangle\right) \\
&\overset{(iii)}{=} -\sum_{i\in\mathcal{I}}\underbrace{\langle \mathbf{y} - P_{\mathcal{C}_i}(\mathbf{x}), P_{\mathcal{C}_i}(\mathbf{x}) - \mathbf{x}\rangle}_{\geq 0} \overset{(iv)}{\leq} 0
\end{aligned}$$

where (i) and (iii) follow from additivity of inner products, (ii) holds by assumption, and (iv) follows from Fact 1.1. Consequently, the equality (a) follows immediately from Fact 1.1.

(b) Let $\mathcal{A} \subset \mathcal{H}$ *and* $\mathcal{B} \subset \mathcal{H}$ *be closed convex subsets such that* $\mathcal{A}\cap\mathcal{B} \neq \emptyset$ *and* $(\forall \mathbf{x} \in \mathcal{H})$ $\langle P_{\mathcal{A}}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\rangle = 0$. *According to (a), it holds that* $(\forall \mathbf{x} \in \mathcal{H})$ $P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}) = P_{\mathcal{B}}(\mathbf{x}) + (P_{\mathcal{A}}(\mathbf{x}) - \mathbf{x})$.

Now, we fix $\mathbf{x} \in \mathcal{H}$ *and define*

$$\begin{aligned}
\mathcal{M} &:= \{\mathbf{y} \in \mathcal{H} \mid \langle \mathbf{y} - P_{\mathcal{A}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle \leq 0\} \\
&= \{\mathbf{y} \in \mathcal{H} \mid \langle \mathbf{y}, \mathbf{a}\rangle \leq b\},
\end{aligned}$$

where $\mathbf{a} := \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})$ *and* $b := \langle P_{\mathcal{A}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle$. *Note that for* $\mathbf{x} \notin \mathcal{A}$, $\mathcal{M} \subset \mathcal{H}$ *is a half-space, whereas for* $\mathbf{x} \in \mathcal{A}$, $\mathcal{M} = \mathcal{H}$. *Thus according to [SY98, Eq. (4.2-12)], the projection of a point* $\mathbf{y} \in \mathcal{H}$ *onto* $\mathcal{M}$ *is given by*

$$P_{\mathcal{M}}(\mathbf{y}) = \begin{cases} \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{a}\rangle - b}{\|\mathbf{a}\|^2}\mathbf{a} & \text{if } \mathbf{x} \notin \mathcal{A} \text{ and } \mathbf{y} \notin \mathcal{M} \\ \mathbf{y} & \text{otherwise.} \end{cases}$$

If $\mathbf{x} \notin \mathcal{A}$*, we have that* $\|\mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\| > 0$*, so*

$$\langle P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{A}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle = \underbrace{\langle P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}, \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle}_{=0} + \|\mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\|^2 > 0.$$

Consequently, since $P_{\mathcal{A}}(\mathbf{x}) \in \mathcal{M}$*, Fact 1.1 implies that* $P_{\mathcal{B}}(\mathbf{x}) \notin \mathcal{M}$ *whenever* $\mathbf{x} \notin \mathcal{A}$*. Hence, the projection of* $\mathbf{y} = P_{\mathcal{B}}(\mathbf{x})$ *onto* $\mathcal{M}$ *can be written as*

$$\begin{aligned} P_{\mathcal{M}} P_{\mathcal{B}}(\mathbf{x}) &= \begin{cases} P_{\mathcal{B}}(\mathbf{x}) - \frac{\langle P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x} + \mathbf{x} - P_{\mathcal{A}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle}{\|\mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\|^2}(\mathbf{x} - P_{\mathcal{A}}(\mathbf{x})) & \text{if } \mathbf{x} \notin \mathcal{A} \\ P_{\mathcal{B}}(\mathbf{x}) & \text{otherwise.} \end{cases} \\ &= P_{\mathcal{B}}(\mathbf{x}) - (\mathbf{x} - P_{\mathcal{A}}(\mathbf{x})) = P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}), \end{aligned}$$

where the last equality follows from (a). According to the projection theorem in Fact 1.1, $(\forall \mathbf{z} \in \mathcal{A})\ \langle \mathbf{z} - P_{\mathcal{A}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{A}}(\mathbf{x})\rangle \leq 0$*, so it is clear that* $\mathcal{A}\cap\mathcal{B} \subset \mathcal{A} \subset \mathcal{M}$*. Thus we have*

$$\begin{aligned} d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A}\cap\mathcal{B}) \geq d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A}) \geq d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{M}) &= \|P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{M}} P_{\mathcal{B}}(\mathbf{x})\| \\ &= \|P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x})\| \geq d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A}\cap\mathcal{B}), \end{aligned}$$

whereby $d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A}) = \|P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x})\|$*. Since the projection of* $P_{\mathcal{B}}(\mathbf{x})$ *onto the closed convex set* $\mathcal{A}$*, i.e., the point in* $\mathcal{A}$ *with distance* $d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A})$ *from* $P_{\mathcal{B}}(\mathbf{x})$*, is unique,* $d(P_{\mathcal{B}}(\mathbf{x}), \mathcal{A}) = \|P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x})\|$ *implies that* $P_{\mathcal{A}} P_{\mathcal{B}}(\mathbf{x}) = P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x})$*.*

Now, for an arbitrary permutation $i_1, \ldots, i_K$ *of the indices in* $\mathcal{I}$*, we define* $(\forall n \in \{1, \ldots, K\})\ \mathcal{A}_n := \bigcap_{k=1}^{n} \mathcal{C}_{i_k}$*. According to (a), we have* $(\forall n \in \{1, \ldots, K-1\})$

$$\langle P_{\mathcal{A}_n}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_{i_{n+1}}}(\mathbf{x}) - \mathbf{x}\rangle = \sum_{k=1}^{n} \langle P_{\mathcal{C}_{i_k}}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_{i_{n+1}}}(\mathbf{x}) - \mathbf{x}\rangle = 0,$$

whereby $P_{\mathcal{A}_{n+1}} = P_{\mathcal{A}_n \cap \mathcal{C}_{i_{n+1}}} = P_{\mathcal{A}_n} P_{\mathcal{C}_{i_{n+1}}}$*. Consequently, it holds that*

$$P_{\mathcal{C}} = P_{\mathcal{A}_K} = P_{\mathcal{A}_{K-1}} P_{\mathcal{C}_{i_K}} = P_{\mathcal{A}_{K-2}} P_{\mathcal{C}_{i_{K-1}}} P_{\mathcal{C}_{i_K}} = \cdots = P_{\mathcal{C}_{i_1}} \cdots P_{\mathcal{C}_{i_K}},$$

which is the desired result. □

In this work, we extend the notion of proximal mappings to proper closed (possibly nonconvex) functions.

Definition 1.1 (Proximal Mapping)**.** Let $f : \mathcal{H} \to (-\infty, +\infty]$ be a proper, closed

function. The proximal mapping $\text{prox}_f : \mathcal{H} \to \mathcal{H}$ associated with f satisfies $(\forall \mathbf{x} \in \mathcal{H})$

$$\text{prox}_f(\mathbf{x}) \in \mathcal{P}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathcal{H}} \left(f(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 \right),$$

where we assume that $(\forall \mathbf{x} \in \mathcal{H})$ $\mathcal{P}(\mathbf{x}) \neq \emptyset$.

If f is proper, lower-semicontinuous, and convex, the set $\mathcal{P}(\mathbf{x})$ is a singleton for all $\mathbf{x} \in \mathcal{H}$, and its unique element is referred to as $\text{prox}_f(\mathbf{x})$. In this case, the assumption in Definition 1.1 is always satisfied. If f is nonconvex, we denote by $\text{prox}_f(\mathbf{x})$ a unique point selected deterministically from the set $\mathcal{P}(\mathbf{x})$. Note that $\mathcal{P}(\mathbf{x})$ is nonempty for all $\mathbf{x} \in \mathcal{H}$ if $(\mathcal{H} = \mathbb{R}^N, \langle\cdot,\cdot\rangle)$ is a finite dimensional real Hilbert space and the function $\mathbf{y} \mapsto f(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2$ is coercive for all $\mathbf{x} \in \mathcal{H}$ [Bec17, Theorem 6.4]. The following fact is a special case of [CP11, Table 10.1.x]. It states a property of proximal mappings that will be useful in the subsequent chapters.

Fact 1.2. *Let $(\mathcal{H} = \mathbb{R}^N, \langle\cdot,\cdot\rangle)$ be a finite dimensional real Hilbert space. Given a proper closed function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ and an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{N\times N}$, the proximal mapping associated with $f(\mathbf{U}\cdot)$ can be expressed in terms of the proximal mapping associated with $f(\cdot)$, where*

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \text{prox}_{f(\mathbf{U}\cdot)}(\mathbf{x}) = \mathbf{U}^T \text{prox}_{f(\cdot)}(\mathbf{U}\mathbf{x}).$$

Given a subset $\mathcal{X} \subset \mathcal{H}$, a fixed point of a mapping $T : \mathcal{X} \to \mathcal{H}$ is a point $\mathbf{x} \in \mathcal{X}$ satisfying $T(\mathbf{x}) = \mathbf{x}$. The set $\text{Fix}(T) = \{\mathbf{x} \in \mathcal{H} \mid T(\mathbf{x}) = \mathbf{x}\}$ is called the fixed point set of T [YO05]. The domain of a mapping T is denoted by $\text{dom}(T)$. Given two mappings $T_1 : \mathcal{H} \supset \mathcal{D}_1 \to \mathcal{R}_1 \subset \mathcal{H}$ and $T_2 : \mathcal{H} \supset \mathcal{D}_2 \to \mathcal{R}_2 \subset \mathcal{H}$ such that $\mathcal{R}_1 \subset \mathcal{D}_2$, we use the shorthand $T_2T_1 := T_2 \circ T_1$ to denote their concatenation, which is defined by the composition $(\forall \mathbf{x} \in \mathcal{H})$ $T_2T_1(\mathbf{x}) := (T_2 \circ T_1)(\mathbf{x}) = T_2(T_1(\mathbf{x}))$. We use the shorthand $T = T_m \cdots T_1$ to denote the compositions of m mappings T_i for $i \in \{1, \ldots, m\}$. Moreover, $(\forall t \in \mathbb{N})$, we write T^t to denote the t-times composition of T with itself, where we use the convention that $T^0 = I$.

Definition 1.2 (Selected Properties of Mappings)**.** [BC11, Definition 4.1,4.33, Proposition 4.2], [YO05], [HX17], [Ceg12, Definition 2.2.1] Let $\mathcal{X} \subset \mathcal{H}$ be a nonempty subset of $\mathcal{H}$. A mapping $T : \mathcal{X} \to \mathcal{H}$ is called

- *nonexpansive* if

$$(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathcal{X}) \quad \|T(\mathbf{x}) - T(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

- *averaged nonexpansive* or *α-averaged nonexpansive* if there exist $\alpha \in (0, 1)$ and a nonexpansive mapping $R : \mathcal{X} \to \mathcal{H}$ such that $T = (1 - \alpha)I + \alpha R$.

- *firmly nonexpansive* (1/2-averaged nonexpansive) if

$$(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathcal{X}) \quad \|T(\mathbf{x}) - T(\mathbf{y})\|^2 \leq \langle T(\mathbf{x}) - T(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle,$$

  or equivalently, if $2T - I$ is nonexpansive.

- *quasi-nonexpansive* if $\mathrm{Fix}(T) \neq \emptyset$ and

$$(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathrm{Fix}(T)) \quad \|T(\mathbf{x}) - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

- *averaged quasi-nonexpansive* or *α-averaged quasi-nonexpansive* if there exist $\alpha \in (0, 1)$ and a quasi-nonexpansive mapping $R : \mathcal{X} \to \mathcal{H}$ such that $T = (1 - \alpha)I + \alpha R$.

- *firmly quasi-nonexpansive* (1/2-averaged quasi-nonexpansive) if $\mathrm{Fix}(T) \neq \emptyset$ and

$$(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathrm{Fix}(T)) \quad \|T(\mathbf{x}) - \mathbf{y}\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2 - \|T(\mathbf{x}) - \mathbf{x}\|^2,$$

  or equivalently, if $2T - I$ is quasi-nonexpansive.

- *κ-attracting quasi-nonexpansive* if $\mathrm{Fix}(T) \neq \emptyset$ and

$$(\exists \kappa > 0)(\forall \mathbf{x} \in \mathcal{X})(\forall \mathbf{y} \in \mathrm{Fix}(T)) \quad \|T(\mathbf{x}) - \mathbf{y}\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2 - \kappa \|\mathbf{x} - T(\mathbf{x})\|^2.$$

To derive the theoretical results in Chapter 2, we use the following additional facts about nonexpansive and quasi-nonexpansive mappings.

Fact 1.3. *[Ceg12, Theorem 2.1.13] Let $\mathcal{X} \subset \mathcal{H}$, let $T_i : \mathcal{X} \to \mathcal{X}, i \in \mathcal{I} = \{1, \ldots, m\}$ be nonexpansive mappings, and define $T := T_m \cdots T_1$. If $T_j(\mathcal{X}) = \{\mathbf{y} \in \mathcal{H} \mid (\exists \mathbf{x} \in \mathcal{X})\ \mathbf{y} = T_j(\mathbf{x})\}$ is bounded for at least one $j \in \mathcal{I}$, then $\mathrm{Fix}(T) \neq \emptyset$.*

Fact 1.4. *[Ceg12, Theorem 2.1.14] Let $\mathcal{X} \subset \mathcal{H}$, let $T_i : \mathcal{X} \to \mathcal{H}, i \in \mathcal{I} = \{1, \ldots, m\}$ be nonexpansive mappings with a common fixed point, and define $T := \sum_{i\in\mathcal{I}} w_i T_i$, where $(\forall i \in \mathcal{I})$ $w_i > 0$ and $\sum_{i\in\mathcal{I}} w_i = 1$. Then $\operatorname{Fix}(T) = \bigcap_{i\in\mathcal{I}} \operatorname{Fix}(T_i)$.*

Fact 1.5. *[YO05, Fact 1], [HX17, Proposition 2.3] Let $T_1, \ldots, T_m : \mathcal{H} \to \mathcal{H}$ be (averaged) nonexpansive mappings. Then the composition $T_m \cdots T_1$ is also (averaged) nonexpansive. Moreover, $\operatorname{Fix}(T_m \cdots T_1) = \bigcap_{i\in\{1,\ldots,m\}} \operatorname{Fix}(T_i)$ whenever the mappings have a common fixed point, i.e., when $\bigcap_{i\in\{1,\ldots,m\}} \operatorname{Fix}(T_i) \neq \emptyset$.*

Fact 1.6. *Let $\mathcal{X} \subset \mathcal{H}$ and let the mapping $T : \mathcal{X} \to \mathcal{H}$ be (α-averaged) nonexpansive with $\operatorname{Fix}(T) \neq \emptyset$. Then T is also (α-averaged) quasi-nonexpansive.*

Proof. *By nonexpansivity of T, $(\forall \mathbf{x} \in \mathcal{X})$ $(\forall \mathbf{y} \in \operatorname{Fix}(T))$*

$$\|T(\mathbf{x}) - T(\mathbf{y})\| = \|T(\mathbf{x}) - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|,$$

so T is also quasi-nonexpansive. Moreover, if T is α-averaged nonexpansive, there exists a nonexpansive mapping $R : \mathcal{X} \to \mathcal{H}$ such that $T = (1 - \alpha)I + \alpha R$. Since R and I are nonexpansive and $\alpha \in (0, 1)$, Fact 1.4 implies that $\operatorname{Fix}(T) = \operatorname{Fix}(R)$, so $\operatorname{Fix}(R) \neq \emptyset$. Hence R is also quasi-nonexpansive, whereby T is α-averaged quasi-nonexpansive. □

Fact 1.7. *[YO05, Proposition 1(b)] Let $\alpha > 0$. Then $T : \mathcal{H} \to \mathcal{H}$ is α-averaged quasi-nonexpansive if and only if $T : \mathcal{H} \to \mathcal{H}$ is $\frac{1-\alpha}{\alpha}$-attracting quasi-nonexpansive with $\operatorname{Fix}(R) = \operatorname{Fix}(T)$*

The following fact is a slight generalization of the result in [YO05, Proposition 1(d)] and [Ceg12, Corollary 2.1.47] to the case in which the two mappings do not necessarily have the same domain.

Fact 1.8. *[YO05, Proposition 1(d)], [Ceg12, Theorem 2.1.46 & Corollary 2.1.47] Let $\mathcal{X}_1, \mathcal{X}_2 \subset \mathcal{H}$, let $T_1 : \mathcal{X}_1 \to \mathcal{X}_2$ be η_1-attracting quasi-nonexpansive and let $T_2 : \mathcal{X}_2 \to \mathcal{X}_1$ be η_2-attracting quasi-nonexpansive with $\operatorname{Fix}(T_1) \cap \operatorname{Fix}(T_2) \neq \emptyset$. Then $T_2T_1 : \mathcal{X}_1 \to \mathcal{X}_1$ is $\frac{\eta_1\eta_2}{\eta_1+\eta_2}$-attracting quasi-nonexpansive with $\operatorname{Fix}(T_2T_1) = \operatorname{Fix}(T_1) \cap \operatorname{Fix}(T_2)$.*

The proof of Fact 1.8 is identical to the proof of [Ceg12, Theorem 2.1.46].

Fact 1.9. *[BC11, Section 16.1] Let $f : \mathcal{H} \to \mathbb{R}$ be a continuous convex function*[a] *and denote by*

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \partial f(\mathbf{x}) := \{\mathbf{g} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H})\ \langle \mathbf{y} - \mathbf{x}, \mathbf{g} \rangle + f(\mathbf{x}) \leq f(\mathbf{y})\} \tag{1.1}$$

the subdifferential *(i.e., the set of all* subgradients*) of f at $\mathbf{x}$. Then $(\forall \mathbf{x} \in \mathcal{H})$ $\partial f(\mathbf{x}) \neq \emptyset$.*

[a] Note that convex functions $f : \mathcal{H} \to \mathbb{R}$ are not in general continuous if $\mathcal{H}$ is infinite dimensional.

Fact 1.10. *[BC01, Proposition 2.3] Let $f : \mathcal{H} \to \mathbb{R}$ be a continuous convex function such that $\operatorname{lev}_{\leq 0} f := \{\mathbf{x} \in \mathcal{H} \mid f(x) \leq 0\} \neq \emptyset$ and let $\mathbf{g}(\mathbf{x}) \in \partial f(\mathbf{x})$ be a subgradient of f at $\mathbf{x} \in \mathcal{H}$. Then the* subgradient projector

$$T : \mathcal{H} \to \mathcal{H} : \ \mathbf{x} \mapsto \begin{cases} \mathbf{x} - \frac{f(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|^2} \mathbf{g}(\mathbf{x}) & \text{if } f(\mathbf{x}) > 0 \\ \mathbf{x} & \text{if } f(\mathbf{x}) \leq 0 \end{cases} \tag{1.2}$$

is firmly quasi-nonexpansive*, i.e., the mapping $2T - I$ is quasi-nonexpansive. Thus, by Definition 1.2, the mapping $(1 - \lambda)I + \lambda T$ is quasi-nonexpansive for all $\lambda \in [0, 2]$.*

Fact 1.11. *[YO05, Proposition 2] Let $\mathcal{K} \subset \mathcal{H}$ be a nonempty closed convex set, and let T be the subgradient projector in* (1.2) *relative to be a continuous convex function f with $\operatorname{lev}_{\leq 0} f \neq \emptyset$. Then for any $\lambda \in (0, 2)$, the mapping*

$$\hat{T}_\lambda := P_{\mathcal{K}} \left((1 - \lambda)I + \lambda T \right)$$

is $\left(1 - \frac{\lambda}{2}\right)$-attracting quasi-nonexpansive with fixed point set $\operatorname{Fix}(\hat{T}) = \mathcal{K} \cap \operatorname{lev}_{\leq 0} f$.

Definition 1.3 (Bounded Perturbations)**.** A sequence $(\beta_n \mathbf{y}_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$ is called a *sequence of bounded perturbations* if $(\beta_n)_{n \in \mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(\exists r \in \mathbb{R})(\forall n \in \mathbb{N})\ \|\mathbf{y}_n\| \leq r$.

Definition 1.4 (Weak and Strong Convergence)**.** [Ceg12, Section 1.1.1.3] A sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$ is said to *converge weakly* to $\mathbf{x} \in \mathcal{H}$ if for any $\mathbf{y} \in \mathcal{H}$ the sequence $(\langle \mathbf{y}, \mathbf{x}_k \rangle)_{n \in \mathbb{N}}$ converges to $\langle \mathbf{y}, \mathbf{x} \rangle$. If a subsequence $(\mathbf{x}_{k_n})_{n \in \mathbb{N}}$ of $(\mathbf{x}_n)_{n \in \mathbb{N}}$ converges weakly to $\mathbf{x}$, then $\mathbf{x}$ is called a *weak cluster point* of $(\mathbf{x}_n)_{n \in \mathbb{N}}$. A sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ in

$\mathcal{H}$ is said to *converge strongly* to $\mathbf{x} \in \mathcal{H}$ if $\lim_{n\to\infty} \|\mathbf{x}_n - \mathbf{x}\| = 0$. In this case, we also write $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$.

The following important properties of weakly convergent sequences will be used to prove the results in Chapter 2.

Fact 1.12. *[Ceg12, Section 1.1.1.3] Weakly convergent sequences have the following properties:*

(a) A bounded sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ includes a weakly convergent subsequence.

(b) Strong convergence of a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ to $\mathbf{x} \in \mathcal{H}$ implies weak convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ to $\mathbf{x}$.

(c) If $\mathcal{H}$ is finite dimensional, weak convergence of a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ to $\mathbf{x} \in \mathcal{H}$ implies strong convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ to $\mathbf{x}$.

Convergence proofs for various algorithms rely on the notion of Fejér monotonicity. A sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ is said to be Fejér monotone relative to a set $\mathcal{S} \subset \mathcal{H}$, if $(\forall \mathbf{z} \in \mathcal{S})(\forall n \in \mathbb{N})\ \|\mathbf{x}_{n+1} - \mathbf{z}\| \leq \|\mathbf{x}_n - \mathbf{z}\|$. To prove the convergence of perturbed algorithms, we make use of the following types of quasi-Fejér monotonicity.

Definition 1.5 (Quasi-Fejér Monotonicity)**.** [Com01, Definition 1.1] Let $\mathcal{S}$ be a nonempty subset of $\mathcal{H}$ and let $(\mathbf{x}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is

- *quasi-Fejér (monotone) of Type-I* relative to $\mathcal{S}$ if

$$(\exists (\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall \mathbf{z} \in \mathcal{S})(\forall n \in \mathbb{N}) \quad \|\mathbf{x}_{n+1} - \mathbf{z}\| \leq \|\mathbf{x}_n - \mathbf{z}\| + \varepsilon_n.$$

- *quasi-Fejér (monotone) of Type-II* relative to $\mathcal{S}$ if

$$(\exists (\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall \mathbf{z} \in \mathcal{S})(\forall n \in \mathbb{N}) \quad \|\mathbf{x}_{n+1} - \mathbf{z}\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 + \varepsilon_n.$$

- *quasi-Fejér (monotone) of Type-III* relative to $\mathcal{S}$ if

$$(\forall \mathbf{z} \in \mathcal{S})(\exists (\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall n \in \mathbb{N}) \quad \|\mathbf{x}_{n+1} - \mathbf{z}\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 + \varepsilon_n.$$

Fact 1.13. *[Com01, Proposition 3.2]. The different types of quasi-Fejér sequences relative to a set $\mathcal{S} \subset \mathcal{H}$ are related as follows.*

- *Type-I $\implies$ Type-III*
- *Type-II $\implies$ Type-III*
- *if $\mathcal{S}$ is bounded, then Type-I $\implies$ Type-II.*

Moreover, will use the following known results to prove the convergence of perturbed algorithms in Chapter 2.

Fact 1.14. *[BC11, Lemma 5.31] Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_+$, and let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ be sequences in $\ell^1_+(\mathbb{N})$ such that*

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} \leq (1+\gamma_n)\,\alpha_n - \beta_n + \delta_n.$$

Then the sequence $(\alpha_n)_{n\in\mathbb{N}}$ converges and $\sum_{n\in\mathbb{N}} \beta_n$ converges.

Fact 1.15. *[Com01, Proposition 3.2-3.3]. Let $(\mathbf{x}_n)_{n\in\mathbb{N}}$ be a quasi-Fejér sequence (of Type-I, Type-II, or Type-III) relative to a nonempty set $\mathcal{S} \subset \mathcal{H}$. Then $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is bounded and $(\forall \mathbf{z} \in \mathcal{S})$ $(\|\mathbf{x}_n - \mathbf{z}\|)_{n\in\mathbb{N}}$ converges.*

Fact 1.16. *[Opi67, Lemma 2], [Ceg12, Lemma 3.2.5] Let $\mathcal{X} \subset \mathcal{H}$, $T : \mathcal{X} \to \mathcal{H}$ be nonexpansive, and $\mathbf{y} \in \mathcal{X}$ be a weak cluster point of a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$. If $\lim_{n\to\infty} \|T(\mathbf{x}_n) - \mathbf{x}_n\| = 0$, then $\mathbf{y} \in \operatorname{Fix}(T)$.*

Fact 1.17. *[Com01, Theorem 3.8] Let $(\mathbf{x}_n)_{n\in\mathbb{N}}$ be quasi-Fejér of Type-III relative to a nonempty set $\mathcal{S} \subset \mathcal{H}$. Then $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathcal{S}$ if and only if the set of weak cluster points of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is a subset of $\mathcal{S}$. Note that quasi-Fejér monotonicity implies that $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is bounded (see Fact 1.15), so the set of weak cluster points is nonempty.*

Additional notation is introduced when needed. Variables are defined independently within each chapter.

2. Bounded Perturbation Resilience of Iterative Algorithms

In this section, we investigate the bounded perturbation resilience of algorithmic frameworks for fixed point problems in Hilbert spaces. Throughout this thesis, we denote by $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ a real Hilbert space equipped with inner product $\langle\cdot,\cdot\rangle$ inducing a norm by

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}.$$

The following propositions will be useful for proving the convergence of algorithms considered in the subsequent sections.

Proposition 2.1. *Let $(T_n : \mathcal{H} \to \mathcal{H})_{n\in\mathbb{N}}$ be a sequence of quasi-nonexpansive mappings such that $\mathcal{C} := \bigcap_{n\in\mathbb{N}} \mathrm{Fix}(T_n) \neq \emptyset$, and let $(\beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ be a sequence of bounded perturbations in $\mathcal{H}$. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by*

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_n\left(\mathbf{x}_n + \beta_n \mathbf{y}_n\right), \quad \mathbf{x}_0 \in \mathcal{H},$$

is quasi-Fejér of Type-I relative to $\mathcal{C}$.

Proof. *By quasi-nonexpansivity of T_n, it holds that* $(\forall \mathbf{z} \in \mathcal{C})(\forall n \in \mathbb{N})$

$$\begin{aligned}
\|\mathbf{x}_{n+1} - \mathbf{z}\|^2 &= \|T_n(\mathbf{x}_n + \beta_n \mathbf{y}_n) - \mathbf{z}\|^2 \\
&\overset{(i)}{\leq} \|\mathbf{x}_n + \beta_n \mathbf{y}_n - \mathbf{z}\|^2 \\
&= \|\mathbf{x}_n - \mathbf{z}\|^2 + 2\beta_n \langle \mathbf{x}_n - \mathbf{z}, \mathbf{y}_n\rangle + \beta_n^2 \|\mathbf{y}_n\|^2 \\
&\overset{(ii)}{\leq} \|\mathbf{x}_n - \mathbf{z}\|^2 + 2\beta_n \|\mathbf{x}_n - \mathbf{z}\| \cdot \|\mathbf{y}_n\| + \beta_n^2 \|\mathbf{y}_n\|^2 \\
&= \left(\|\mathbf{x}_n - \mathbf{z}\| + \beta_n \|\mathbf{y}_n\|\right)^2,
\end{aligned}$$

where (i) follows from quasi-nonexpansivity of T_n and (ii) is an application of the Cauchy-Schwartz inequality. Since $(\beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ is a sequence of bounded perturbations, there exists $r > 0$ such that $(\forall n \in \mathbb{N})$ $\|\mathbf{y}_n\| \leq r$, and $(\gamma_n)_{n\in\mathbb{N}} := (r\beta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$.

Consequently, $\left(\exists(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})\right)(\forall \mathbf{z}\in\mathcal{C})(\forall n\in\mathbb{N})$

$$\|\mathbf{x}_{n+1}-\mathbf{z}\| \leq \|\mathbf{x}_n-\mathbf{z}\|+\gamma_n,$$

which is the desired result. □

Proposition 2.2. *Let* $\kappa>0$ *and let* $(T_n:\mathcal{H}\to\mathcal{H})_{n\in\mathbb{N}}$ *be a sequence of* κ*-attracting quasi-nonexpansive mappings such that* $\mathcal{C}:=\bigcap_{n\in\mathbb{N}}\operatorname{Fix}(T_n)\neq\emptyset$*, and let* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$ *be a sequence of bounded perturbations in* $\mathcal{H}$*. Then for any bounded subset* $\mathcal{U}\subset\mathcal{C}$ *the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *generated by*

$$(\forall n\in\mathbb{N})\quad \mathbf{x}_{n+1}=T_n\left(\mathbf{x}_n+\beta_n\mathbf{y}_n\right),\quad \mathbf{x}_0\in\mathcal{H}$$

satisfies the following: $\left(\exists(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})\right)(\forall \mathbf{z}\in\mathcal{U})(\forall n\in\mathbb{N})$

$$\|\mathbf{x}_{n+1}-\mathbf{z}\|^2 \leq \|\mathbf{x}_n-\mathbf{z}\|^2-\kappa\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2+\gamma_n.$$

Proof. *Since* $(\forall n\in\mathbb{N})$ T_n *is* κ*-attracting quasi-nonexpansive with* $\operatorname{Fix}(T_n)\supset\mathcal{C}\supset\mathcal{U}$*, it holds that* $(\forall \mathbf{z}\in\mathcal{U})(\forall n\in\mathbb{N})$

$$\begin{aligned}
\|\mathbf{x}_{n+1}-\mathbf{z}\|^2 &\leq \|\mathbf{x}_n+\beta_n\mathbf{y}_n-\mathbf{z}\|^2-\kappa\|\mathbf{x}_{n+1}-(\mathbf{x}_n+\beta_n\mathbf{y}_n)\|^2\\
&= \|\mathbf{x}_n-\mathbf{z}\|^2+2\beta_n\langle\mathbf{x}_n-\mathbf{z},\mathbf{y}_n\rangle+\beta_n^2\|\mathbf{y}_n\|^2\\
&\quad -\kappa\left(\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2+2\beta_n\langle\mathbf{y}_n,\mathbf{x}_n-\mathbf{x}_{n+1}\rangle+\beta_n^2\|\mathbf{y}_n\|^2\right)\\
&\overset{(i)}{\leq} \|\mathbf{x}_n-\mathbf{z}\|^2+2\beta_n\langle\mathbf{x}_n-\mathbf{z},\mathbf{y}_n\rangle+\beta_n^2\|\mathbf{y}_n\|^2\\
&\quad -\kappa\left(\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2+2\beta_n\langle\mathbf{y}_n,\mathbf{x}_n-\mathbf{x}_{n+1}\rangle\right)\\
&\overset{(ii)}{\leq} \|\mathbf{x}_n-\mathbf{z}\|^2+2\beta_n\|\mathbf{x}_n-\mathbf{z}\|\|\mathbf{y}_n\|+\beta_n^2\|\mathbf{y}_n\|^2\\
&\quad -\kappa\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2+2\kappa\beta_n\|\mathbf{y}_n\|\|\mathbf{x}_n-\mathbf{x}_{n+1}\|\\
&= \|\mathbf{x}_n-\mathbf{z}\|^2-\kappa\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2\\
&\quad +2\beta_n\|\mathbf{y}_n\|\left(\|\mathbf{x}_n-\mathbf{z}\|+\kappa\|\mathbf{x}_n-\mathbf{x}_{n+1}\|\right)+\beta_n^2\|\mathbf{y}_n\|^2,
\end{aligned}$$

where (i) follows from nonnegativity of $\beta_n\|\mathbf{y}_n\|$ *and (ii) is a two-fold application of the Cauchy-Schwartz inequality. By Proposition 2.1,* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér of Type-I relative to* $\mathcal{C}$*, so Fact 1.15 ensures that* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is bounded. Boundedness of* $(\mathbf{x}_n)_{n\in\mathbb{N}}$*,* $(\mathbf{y}_n)_{n\in\mathbb{N}}$ *and* $\mathcal{U}$*, guarantee the existence of some* $r>0$ *such that* $(\forall n\in\mathbb{N})(\forall \mathbf{z}\in\mathcal{U})$

$\|\mathbf{x}_n - \mathbf{z}\| + \kappa\|\mathbf{x}_n - \mathbf{x}_{n+1}\| \leq r$ *and* $(\forall n \in \mathbb{N})\ \|\mathbf{y}_n\| \leq r$. *Consequently we can write*

$$\begin{aligned}\|\mathbf{x}_{n+1} - \mathbf{z}\|^2 &\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - \kappa\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 + \beta_n r^2 (2 + \beta_n) \\ &\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - \kappa\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 + \beta_n r^2 (2 + b),\end{aligned}$$

where we defined $b := \sum_{n\in\mathbb{N}} \beta_n$. *Therefore, defining* $c := r^2(2 + b)$ *and* $(\gamma_n)_{n\in\mathbb{N}} := (c\beta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ *yields the desired result.* □

Proposition 2.3. *Let* $\mathcal{S} \subset \mathcal{H}$ *be a nonempty subset, let* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *be quasi-Fejér of Type-I relative to* $\mathcal{S}$, *and let* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$ *be a sequence of bounded perturbations. Then the sequence* $(\mathbf{z}_n)_{n\in\mathbb{N}}$ *given by* $(\forall n \in \mathbb{N})\ \mathbf{z}_n = \mathbf{x}_n + \beta_n\mathbf{y}_n$ *is also quasi-Fejér of Type-I relative to* $\mathcal{S}$.

Proof. *By (Type-I) quasi-Fejér monotonicity of* $(\mathbf{x}_n)_{n\in\mathbb{N}}$, $(\exists(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall \mathbf{z} \in \mathcal{S})(\forall n \in \mathbb{N})\ \|\mathbf{x}_{n+1} - \mathbf{z}\| \leq \|\mathbf{x}_n - \mathbf{z}\| + \gamma_n$. *Thus it holds that* $(\exists(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N}))(\forall \mathbf{z} \in \mathcal{S})(\forall n \in \mathbb{N})$

$$\begin{aligned}\|\mathbf{z}_{n+1} - \mathbf{z}\| &= \|\mathbf{x}_{n+1} + \beta_{n+1}\mathbf{y}_{n+1} - \mathbf{z}\| \\ &\overset{(i)}{\leq} \|\mathbf{x}_{n+1} - \mathbf{z}\| + \beta_{n+1}\|\mathbf{y}_{n+1}\| \\ &\overset{(ii)}{\leq} \|\mathbf{x}_n - \mathbf{z}\| + \gamma_n + \beta_{n+1}\|\mathbf{y}_{n+1}\| \\ &= \|\mathbf{z}_n - \beta_n\mathbf{y}_n - \mathbf{z}\| + \gamma_n + \beta_n\|\mathbf{y}_{n+1}\| \\ &\overset{(iii)}{\leq} \|\mathbf{z}_n - \mathbf{z}\| + \underbrace{\gamma_n + \beta_n\|\mathbf{y}_n\| + \beta_{n+1}\|\mathbf{y}_{n+1}\|}_{=:\varepsilon_n},\end{aligned}$$

where we used the triangle inequality in (i) and (iii), and Type-I quasi-Fejér monotonicity of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *in (ii). Moreover,* $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ *because* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$ *is a sequence of bounded perturbations. Consequently,* $(\mathbf{z}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér of Type-I relative to* $\mathcal{S}$. □

Proposition 2.4. *Let* $T : \mathcal{H} \to \mathcal{H}$ *be quasi-nonexpansive with* $\mathcal{C} := \mathrm{Fix}(T) \neq \emptyset$ *and let* $\mathcal{B} \subset \mathcal{H}$ *be a bounded set. Then the image* $T(\mathcal{B}) = \{\mathbf{y} \in \mathcal{H} \mid (\exists \mathbf{x} \in \mathcal{B})\ \mathbf{y} = T(\mathbf{x})\}$ *of* $\mathcal{B}$ *under* T *is also bounded.*

Proof. *Note that* $(\forall \mathbf{x} \in \mathcal{H})(\forall \mathbf{z} \in \mathcal{C})$

$$\|T(\mathbf{x})\| \overset{(i)}{\leq} \|T(\mathbf{x}) - \mathbf{z}\| + \|\mathbf{z}\|$$

$$\overset{(ii)}{\leq} \|\mathbf{x}-\mathbf{z}\| + \|\mathbf{z}\|,$$

where (i) is an application of the triangle inequality and (ii) follows from quasi-nonexpansivity of T. Since $\mathcal{B}$ is bounded, there exists a bounded set $\mathcal{D} \supset \mathcal{B}$ such that $\mathcal{D} \cap \mathcal{C} \neq \emptyset$. Moreover, boundedness of $\mathcal{D}$ implies that $(\exists c \in \mathbb{R}_+)(\forall \mathbf{x} \in \mathcal{D})(\forall \mathbf{z} \in \mathcal{D} \cap \mathcal{C})$ $\|T(\mathbf{x})\| \leq \|\mathbf{x}-\mathbf{z}\| + \|\mathbf{z}\| \leq c$, which proves that $T(\mathcal{B}) \subset T(\mathcal{D})$ is bounded. □

In the remainder of this chapter, we investigate the bounded perturbation resilience of the POCS algorithm, of the APSM, and of two types of extrapolated alternating projection methods.

2.1. Projections Onto Convex Sets

A vast variety of engineering problems can be posed in terms of convex feasibility problems, which aim at finding a point in the intersection of a collection of closed convex sets. Each of the sets represents a certain property of the sought point (estimandum), and any point in the intersection of all sets possesses all of the desired properties. More formally, let $\mathcal{I} := \{1, \ldots, K\}$, let $(\forall k \in \mathcal{I})$ $\mathcal{C}_k \subset \mathcal{H}$ be a nonempty closed convex set, and consider the convex feasibility problem

$$\text{find } \mathbf{x} \in \mathcal{C}_\star := \bigcap_{k \in \mathcal{I}} \mathcal{C}_k, \tag{2.1}$$

where typically $\mathcal{C}_\star$ is assumed to be nonempty. A well-known technique to obtain solutions to this problem is the POCS algorithm, which is described below. Denote by $(\forall \lambda \in [0,2])$

$$T_{\mathcal{C}}^{(\lambda)} : \mathcal{H} \to \mathcal{H} : \ \mathbf{x} \mapsto \mathbf{x} + \lambda(P_{\mathcal{C}}(\mathbf{x}) - \mathbf{x})$$

the relaxed projection of $\mathbf{x} \in \mathcal{H}$ onto a closed convex set $\mathcal{C} \subset \mathcal{H}$. Now, we can define a POCS mapping $T : \mathcal{H} \to \mathcal{H}$ by

$$T := T_{\mathcal{C}_K}^{(\lambda_K)} \cdots T_{\mathcal{C}_1}^{(\lambda_1)}, \tag{2.2}$$

where $(\forall k \in \mathcal{I})$ $\lambda_k \in (0,2)$. Starting from any point $\mathbf{x}_0 \in \mathcal{H}$, the POCS algorithm produces a sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$ via

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T(\mathbf{x}_n). \tag{2.3}$$

The POCS algorithm above is a special case of the string-averaging projection method proposed in [CEH01]. In [CZ13] and [BRZ18], the authors investigate the convergence of perturbed variants of dynamic string averaging projection methods (see the formal

definition in [CZ13]). However, the results in these works are restricted to unrelaxed projections, i.e., to the case where $(\forall k \in \mathcal{I})\ \lambda_k = 1$. In [7], we showed that the mapping in (2.2) is α-averaged nonexpansive. We include this result below.

Remark 2.1. [7, Remark 5] The mapping T in (2.2) is averaged nonexpansive. Moreover, $\operatorname{Fix}(T) = \mathcal{C}_\star$ whenever the solution set $\mathcal{C}_\star$ in (2.1) is nonempty.

Proof. Note that, for every nonempty closed convex subset $\mathcal{C} \subset \mathcal{H}$, the reflector $R_\mathcal{C} = I + 2(P_\mathcal{C} - I)$ is nonexpansive [BC11, Corollary 4.18]. Therefore, according to Definition 1.2, $(\forall \lambda \in (0,2))$ the relaxed projector

$$T_\mathcal{C}^{(\lambda)} = I + \lambda(P_\mathcal{C} - I) = I + \frac{\lambda}{2}(R_\mathcal{C} - I)$$

is $\lambda/2$-averaged. Further (see Fact 1.5), the composite of finitely many averaged mappings is α-averaged for some $\alpha \in (0,1)$.[a] Moreover, as both the identity mapping and the the projection onto a closed convex set are nonexpansive, Fact 1.4 implies that $(\forall \mathcal{C} \subset \mathcal{H})(\forall \lambda \in (0,2))\ \operatorname{Fix}\left(T_\mathcal{C}^{(\lambda)}\right) = \operatorname{Fix}(I) \cap \operatorname{Fix}(P_\mathcal{C}) = \mathcal{C}$. Thus by Fact 1.5, $\mathcal{C}_\star \neq \emptyset$ implies that $\operatorname{Fix}(T) = \bigcap_{k\in\mathcal{I}} \operatorname{Fix}\left(T_{\mathcal{C}_k}^{(\lambda_k)}\right) = \mathcal{C}_\star$. □

[a] See [OY02, CY15] for an exact evaluation of the averagedness constant α of the composition of finitely many averaged mappings.

The perturbation resilience of α-averaged nonexpansive mappings (which by Remark 2.1 include the POCS mapping in (2.2)) in finite dimensional real Hilbert spaces $\mathcal{H} = \mathbb{R}^J$ has been proven in [HX17]. Lemma 2.1 generalizes this result to arbitrary real Hilbert spaces.

Lemma 2.1. *Let $T : \mathcal{H} \to \mathcal{H}$ be averaged nonexpansive with $\operatorname{Fix}(T) \neq \emptyset$ and let $(\beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ be a sequence of bounded perturbations. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ produced by the recurrence*

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T(\mathbf{x}_n + \beta_n \mathbf{y}_n), \quad \mathbf{x}_0 \in \mathcal{H}$$

converges weakly to a point in $\operatorname{Fix}(T)$.

Proof. *Since $\operatorname{Fix}(T) \neq \emptyset$ by assumption, T is also averaged quasi-nonexpansive, according to Fact 1.6. Hence by Fact 1.7, T is κ-attracting quasi-nonexpansive for some $\kappa > 0$. Therefore, Proposition 2.2 implies that there exist $\mathbf{z} \in \operatorname{Fix}(T)$ and $(\gamma_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ such that $(\mathbf{x}_n)_{n\in\mathbb{N}}$ satisfies*

$$(\forall n \in \mathbb{N}) \quad \|\mathbf{x}_{n+1} - \mathbf{z}\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 - \kappa\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 + \gamma_n.$$

Thus, according to Fact 1.14, we have that $\sum_{n\in\mathbb{N}} \kappa\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2 < \infty$, *which implies* $\lim_{n\to\infty}\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2 = 0$. *Now, define*

$$(\forall n\in\mathbb{N})\quad \bar{\mathbf{y}}_n := \begin{cases} \beta_n^{-1}\left(T(\mathbf{x}_n+\beta_n\mathbf{y}_n)-T(\mathbf{x}_n)\right) & \text{if } \beta_n\neq 0\\ \mathbf{0} & \text{otherwise}\end{cases}$$

and note that

$$\beta_n\|\bar{\mathbf{y}}_n\| = \|T(\mathbf{x}_n+\beta_n\mathbf{y}_n)-T(\mathbf{x}_n)\| \leq \beta_n\|\mathbf{y}_n\|,$$

because T *is nonexpansive. Therefore,* $(\exists r>0)\ (\forall n\in\mathbb{N})\ \|\bar{\mathbf{y}}_n\|\leq r$ *since* $(\mathbf{y}_n)_{n\in\mathbb{N}}$ *is bounded by assumption. Moreover, note that* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is bounded due to quasi-nonexpansivity of* T, *Proposition 2.1, and Fact 1.15, so owing to nonexpansivity (i.e., 1-Lipschitz continutity) of* T, $(\|T(\mathbf{x}_n)\|)_{n\in\mathbb{N}}$ *is also bounded. Hence the triangle inequality shows that* $(\|T(\mathbf{x}_n)-\mathbf{x}_n\|)_{n\in\mathbb{N}}$ *is bounded and we obtain*

$$\begin{aligned}
0 &= \lim_{n\to\infty}\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2\\
&= \lim_{n\to\infty}\|T(\mathbf{x}_n+\beta_n\mathbf{y}_n)-\mathbf{x}_n\|^2\\
&= \lim_{n\to\infty}\|T(\mathbf{x}_n)+\beta_n\bar{\mathbf{y}}_n-\mathbf{x}_n\|^2\\
&= \lim_{n\to\infty}\Big(\|T(\mathbf{x}_n)-\mathbf{x}_n\|^2+2\langle T(\mathbf{x}_n)-\mathbf{x}_n,\beta_n\bar{\mathbf{y}}_n\rangle+\beta_n^2\|\bar{\mathbf{y}}_n\|^2\Big)\\
&\geq \lim_{n\to\infty}\Big(\|T(\mathbf{x}_n)-\mathbf{x}_n\|^2-\underbrace{2\beta_n\|T(\mathbf{x}_n)-\mathbf{x}_n\|\|\bar{\mathbf{y}}_n\|}_{\to 0}+\underbrace{\beta_n^2\|\bar{\mathbf{y}}_n\|^2}_{\to 0}\Big)\\
&= \lim_{n\to\infty}\|T(\mathbf{x}_n)-\mathbf{x}_n\|^2 \geq 0.
\end{aligned}$$

Consequently, Fact 1.16 and the nonexpansivity of T *imply that any weak cluster point of* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is a fixed point of* T. *Finally, since* T *is also quasi-nonexpansive, Proposition 2.1 proves that* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér of Type-I (and hence by Fact 1.13 of Type-III). Therefore Fact 1.17 implies the weak convergence of* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *to a point in* $\mathrm{Fix}(T)$. □

Using Lemma 2.1, we can now derive conditions for the convergence of perturbed versions of the POCS algorithm in (2.3).

Theorem 2.1. *Let* $T:\mathcal{H}\to\mathcal{H}$ *be the mapping in* (2.2) *with* $(\forall k\in\mathcal{I})\ \lambda_k\in(0,2)$ *and let* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$ *be a sequence of bounded perturbations. Then the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *produced by the perturbed POCS algorithm*

$$(\forall n\in\mathbb{N})\quad \mathbf{x}_{n+1}=T(\mathbf{x}_n+\beta_n\mathbf{y}_n),\quad \mathbf{x}_0\in\mathcal{H}$$

satisfies the following:

(a) If T has a fixed point, $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathrm{Fix}(T)$.

(b) If $(\exists k \in \mathcal{I})$ such that the set $\mathcal{C}_k$ in (2.1) *is bounded and $\lambda_k = 1$, then T has a fixed point and $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in* $\mathrm{Fix}(T)$.

(c) If the solution set $\mathcal{C}_\star$ in (2.1) *is nonempty, $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\mathbf{x}_\star \in \mathcal{C}_\star$.*

(d) If $\mathcal{H}$ is finite dimensional, convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in (a)-(c) is strong.

Proof.

(a) As shown in Remark 2.1, the mapping T is averaged nonexpansive, so this result is immediate from Lemma 2.1.

(b) By assumption, there exists $k \in \mathcal{I}$ such that $\lambda_k = 1$, i.e., $T_{\mathcal{C}_k}^{(\lambda_k)} = P_{\mathcal{C}_k}$. Hence it holds that $T_{\mathcal{C}_k}^{(\lambda_k)}(\mathcal{H}) = P_{\mathcal{C}_k}(\mathcal{H}) = \{\mathbf{y} \in \mathcal{H} \mid (\exists \mathbf{x} \in \mathcal{H})\ \mathbf{y} = P_{\mathcal{C}_k}(\mathbf{x})\} = \mathcal{C}_k$, which is bounded by assumption. Therefore, Fact 1.3 implies that the mapping T in (2.2) *has a fixed point. Consequently, weak convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ to a point in* $\mathrm{Fix}(T)$ *follows from (a).*

(c) According to Remark 2.1, $\mathcal{C}_\star \neq \emptyset \implies \mathrm{Fix}(T) = \mathcal{C}_\star$, so (c) follows from (a).

(d) This result is immediate from Fact 1.12(c). □

2.2. Adaptive Projected Subgradient Method

The POCS algorithm in the previous section can only be applied to convex feasibility problems with a finite number of sets. In many practical applications, information about a sought point arrives sequentially, e.g., based on periodic measurements. A means of approaching estimation problems in this online setting is the APSM [YO05]. This algorithmic framework generalizes Polyak's subgradient method [Pol69] to the case of time-varying cost functions, i.e., it can be used to minimize asymptotically a sequence of convex functions over a closed convex set. One particular application of the APSM, which we will use in Chapters 4 and 5, is as follows. Given a closed convex set $\mathcal{K} \subset \mathcal{H}$, and an infinite collection $(\mathcal{C}_n)_{n\in\mathbb{N}}$ of closed convex sets with a nonempty intersection, we can use the APSM to find a point in the intersection of all but finitely many sets of the sequence $(\mathcal{C}_n)_{n\in\mathbb{N}}$. More

precisely, we can use the APSM to solve the problem

$$\text{find } \mathbf{x} \in \mathcal{K} \quad \text{s.t.} \quad \mathbf{x} \in \left(\bigcap_{n \geq n_0} \mathcal{C}_n \right)$$

for some $n_0 \in \mathbb{N}$. For instance, the set $\mathcal{C}_n$ (and the corresponding cost function Θ_n) can be used to encode new information obtained at time instant n, and the convex set $\mathcal{K}$ can incorporate prior knowledge. In this section, we show that the APSM is bounded perturbation resilient. By doing so, we justify the use of superiorized heuristics based on the APSM. The proofs follow closely the structure in [YO05], while extending the results in [YO05] to perturbed versions of the APSM. As in [YO05], Lemma 2.2 and Theorem 2.2 are technical results, which we use to prove the main result in Theorem 2.3. The following lemma is a generalization of [YO05, Lemma 1] to quasi-Fejér monotone sequences. It follows line by line the proof in [YO05].

Lemma 2.2. *Suppose that a sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ is quasi-Fejér monotone of Type-I relative to a closed convex set $\mathcal{C} \subset \mathcal{H}$. In addition, suppose that $\mathcal{C}$ has a nonempty relative interior with respect to a linear variety $\mathcal{V} \subset \mathcal{H}$, i.e., there exist $\mathbf{x}_0 \in \mathcal{C} \cap \mathcal{V}$ and $\varepsilon > 0$ satisfying $\mathcal{U} := \{\mathbf{x} \in \mathcal{V} \mid \|\mathbf{x} - \mathbf{x}_0\| \leq \varepsilon\} \subset \mathcal{C}$. Then $(P_{\mathcal{V}}(\mathbf{x}_n))_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{V}$.*

Proof. *It is sufficient to show that $(P_{\mathcal{V}}(\mathbf{x}_n))_{n\in\mathbb{N}}$ is a Cauchy sequence. To do so, we first show that there exists $(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ such that*

$$(\forall n \in \mathbb{N}) \quad 2\varepsilon \|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\| \leq \|\mathbf{u}_n - \mathbf{x}_0\|^2 - \|\mathbf{u}_{n+1} - \mathbf{x}_0\|^2 + \gamma_n. \tag{2.4}$$

If $P_{\mathcal{V}}(\mathbf{u}_n) = P_{\mathcal{V}}(\mathbf{u}_{n+1})$ for some $n \in \mathbb{N}$, quasi-Fejér monotonicity of $(\mathbf{u}_n)_{n\in\mathbb{N}}$ ensures that (2.4) holds for this n. Therefore it is sufficient to consider $n \in \mathbb{N}$ such that $P_{\mathcal{V}}(\mathbf{u}_n) \neq P_{\mathcal{V}}(\mathbf{u}_{n+1})$. In this case, we have $\mathbf{x}_0 + \varepsilon \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} \in \mathcal{C} \cap \mathcal{V}$, thus by Type-I quasi-Fejér monotonicity of $(\mathbf{u}_n)_{n\in\mathbb{N}}$ there exists $(\delta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ such that

$$\left\| \mathbf{x}_0 + \varepsilon \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} - \mathbf{u}_{n+1} \right\| \leq \left\| \mathbf{x}_0 + \varepsilon \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} - \mathbf{u}_n \right\| + \delta_n.$$

Squaring and expanding the above inequality yields

$$\begin{aligned} &\|\mathbf{x}_0 - \mathbf{u}_{n+1}\|^2 + 2\varepsilon \left\langle \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|}, \mathbf{x}_0 - \mathbf{u}_{n+1} \right\rangle + \varepsilon^2 \\ &\leq \|\mathbf{x}_0 - \mathbf{u}_n\|^2 + 2\varepsilon \left\langle \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|}, \mathbf{x}_0 - \mathbf{u}_n \right\rangle + \varepsilon^2 \end{aligned}$$

$$+ 2\delta_n \left\| \mathbf{x}_0 + \varepsilon \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} - \mathbf{u}_n \right\| + \delta_n^2.$$

By rearranging and applying the triangle inequality, we obtain

$$\begin{aligned}
&2\varepsilon \frac{\langle P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1}), \mathbf{u}_n - \mathbf{u}_{n+1} \rangle}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} \\
&\leq \|\mathbf{x}_0 - \mathbf{u}_n\|^2 - \|\mathbf{x}_0 - \mathbf{u}_{n+1}\|^2 + \delta_n \left(2 \left\| \mathbf{x}_0 + \varepsilon \frac{P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} - \mathbf{u}_n \right\| + \delta_n \right) \\
&\leq \|\mathbf{x}_0 - \mathbf{u}_n\|^2 - \|\mathbf{x}_0 - \mathbf{u}_{n+1}\|^2 + \delta_n \left(2 \|\mathbf{x}_0 - \mathbf{u}_n\| + 2\varepsilon + \delta_n \right).
\end{aligned}$$

Since $(\mathbf{u}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér monotone with respect to* $\mathcal{C}$ *and* $\mathbf{x}_0 \in \mathcal{C}$*, the sequence* $(\|\mathbf{x}_0 - \mathbf{u}_n\|)_{n\in\mathbb{N}}$ *converges (see Fact 1.15). Therefore* $(\exists r > 0)(\forall n \in \mathbb{N})\ \|\mathbf{x}_0 - \mathbf{u}_n\| < r$*. By defining* $a := (2r + 2\varepsilon + \sum_{n\in\mathbb{N}} \delta_n)$ *we obtain a sequence* $(\gamma_n)_{n\in\mathbb{N}} = (a\delta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ *such that*

$$2\varepsilon \frac{\langle P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1}), \mathbf{u}_n - \mathbf{u}_{n+1} \rangle}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|} \leq \|\mathbf{x}_0 - \mathbf{u}_n\|^2 - \|\mathbf{x}_0 - \mathbf{u}_{n+1}\|^2 + \gamma_n.$$

From firm nonexpansivity of $P_{\mathcal{V}}$ *(see Definition 1.2) we have*

$$0 \leq \|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\| \leq \frac{\langle P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1}), \mathbf{u}_n - \mathbf{u}_{n+1} \rangle}{\|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+1})\|},$$

which proves (2.4)*. Since* $(\forall n \in \mathbb{N})\ \gamma_n \geq 0$*, the inequality in* (2.4) *implies that* $(\forall n \in \mathbb{N})(\forall k \in \mathbb{N})$

$$2\varepsilon \|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+k})\| \leq \|\mathbf{u}_n - \mathbf{x}_0\|^2 - \|\mathbf{u}_{n+k} - \mathbf{x}_0\|^2 + \sum_{i=n}^{\infty} \gamma_i.$$

Moreover, since $(\|\mathbf{u}_n - \mathbf{x}_0\|)_{n\in\mathbb{N}}$ *converges and* $(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$*,* $(\forall \delta > 0)(\exists N \in \mathbb{N})(\forall n \geq N)(\forall k \in \mathbb{N})$

$$2\varepsilon \|P_{\mathcal{V}}(\mathbf{u}_n) - P_{\mathcal{V}}(\mathbf{u}_{n+k})\| \leq \|\mathbf{u}_n - \mathbf{x}_0\|^2 - \|\mathbf{u}_{n+k} - \mathbf{x}_0\|^2 + \sum_{i=n}^{\infty} \gamma_i < \delta,$$

which shows that $(P_{\mathcal{V}}(\mathbf{u}_n))_{n\in\mathbb{N}}$ *is a Cauchy sequence.* □

In Theorem 2.2 below, we generalize [YO05, Theorem 1] to quasi-Fejér monotone sequences. The proof is a slightly altered version of the proof in [YO05]. It comprises a somewhat involved geometric argument, which we summarize in the following: By harnessing Lemma 2.2 and quasi-Fejér monotonicity of the sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}$ relative to a closed convex set $\mathcal{C}$, we construct a set $\mathcal{S}(\varepsilon) \subset \mathcal{H}$ containing every element of the subsequence

$(\mathbf{u}_n)_{n\geq N_1}$ for some $N_1 \in \mathbb{N}$. Using the same geometric arguments as in [YO05], we show that the set $\mathcal{S}(\varepsilon)$ is a subset of the union of two induced norm balls with radius $\varepsilon > 0$, which we denote by $\mathcal{B}(\varepsilon)$. As the premise of Theorem 2.2 implies that $\lim_{n\to\infty} \|\mathbf{u}_{n+1} - \mathbf{u}_n\| = 0$, we conclude that for sufficiently small $\varepsilon > 0$ there exists $N_2 \in \mathbb{N}$ such that all elements of the subsequence $(\mathbf{u}_n)_{n\geq N_2}$ are contained in a single one of the two induced norm balls. Thus we deduce that the sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}$ converges to the center of this ball. To provide some intuition, Figure 2.1 illustrates the geometric properties used in the proof. Note that the arguments in Theorem 2.2 remain valid in arbitrary (possibly infinite dimensional) real Hilbert spaces.

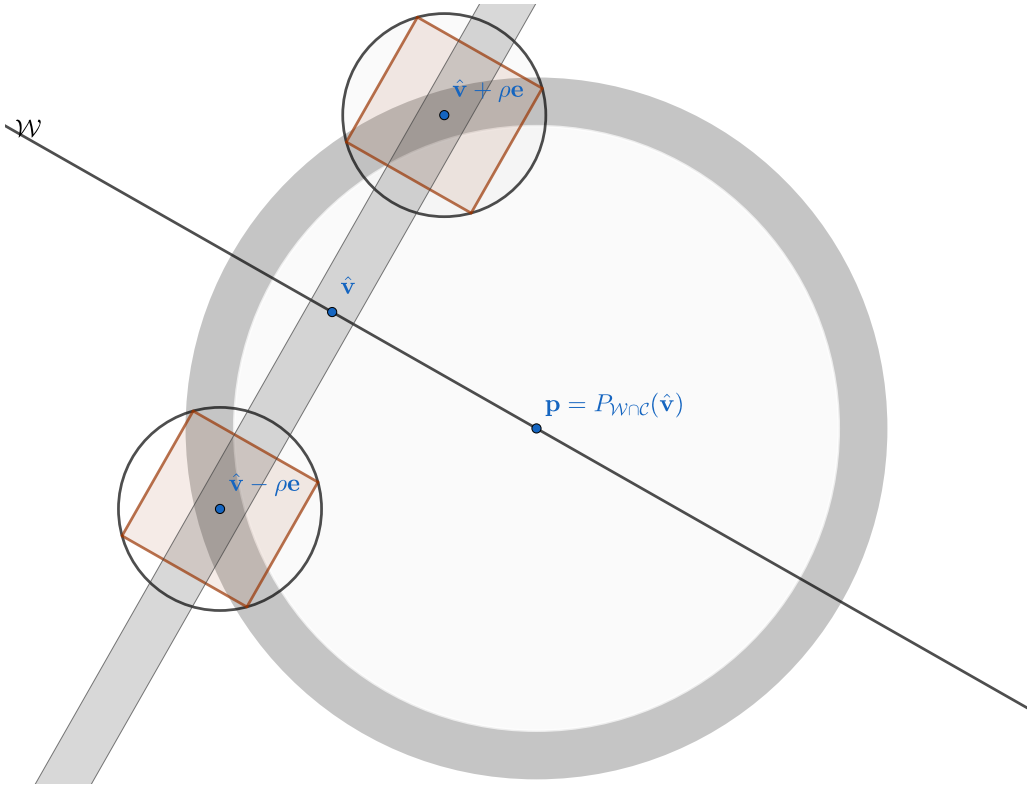


Figure 2.1.: Illustration of the arguments in Theorem 2.2. The intersection of the grey regions corresponds to the set $\mathcal{S}(\varepsilon)$. The red squares have a sidelength of $\sqrt{2}\varepsilon$, so each of them is enclosed in a ball of radius ε. The union of these two balls represents the set $\mathcal{B}(\varepsilon)$.

Theorem 2.2. *Let $(\mathbf{u}_n)_{n\in\mathbb{N}}$ be a quasi-Fejér sequence of Type-I relative to a closed*

convex set $\mathcal{C} \subset \mathcal{H}$ *and suppose that there exist* $\kappa > 0$ *and* $(\gamma_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ *such that*

$$(\forall \mathbf{z} \in \mathcal{C})(\forall n \in \mathbb{N}) \quad \kappa\|\mathbf{u}_n - \mathbf{u}_{n+1}\|^2 \leq \|\mathbf{u}_n - \mathbf{z}\|^2 - \|\mathbf{u}_{n+1} - \mathbf{z}\|^2 + \gamma_n. \tag{2.5}$$

Then $(\mathbf{u}_n)_{n\in\mathbb{N}}$ *converges strongly to a point in* $\mathcal{H}$ *if* $\mathcal{C}$ *has a nonempty relative interior with respect to a hyperplane* $\mathcal{W} \subset \mathcal{H}$.

Proof. *According to Lemma 2.2, the sequence* $(P_{\mathcal{W}}(\mathbf{u}_n))_{n\in\mathbb{N}}$ *converges strongly a point in* $\mathcal{W}$. *Hence we can define*

$$\hat{\mathbf{v}} := \lim_{n\to\infty} P_{\mathcal{W}}(\mathbf{u}_n), \tag{2.6}$$

$\mathbf{p} := P_{\mathcal{C}\cap\mathcal{W}}(\hat{\mathbf{v}})$, *and* $\mathbf{e} \in \mathcal{H}$ *satisfying* $\mathcal{W} = \{\mathbf{x} \in \mathcal{H} \mid \langle \mathbf{e}, \mathbf{x} - \hat{\mathbf{v}}\rangle = 0\}$ *and* $\|\mathbf{e}\| = 1$. *Moreover, according to Fact 1.15, the sequence* $(\|\mathbf{u}_n - \mathbf{z}\|)_{n\in\mathbb{N}}$ *converges for all* $\mathbf{z} \in \mathcal{C}$, *so we can define*

$$a := \lim_{n\to\infty} \|\mathbf{u}_n - \mathbf{p}\|, \tag{2.7}$$

$$\tau := \lim_{n\to\infty} \|P_{\mathcal{W}}(\mathbf{u}_n) - \mathbf{p}\| = \|\hat{\mathbf{v}} - \mathbf{p}\|, \tag{2.8}$$

$$\rho := \sqrt{a^2 - \tau^2}. \tag{2.9}$$

Note that by nonexpansivity of $P_{\mathcal{W}}$ *we have* $a \geq \lim_{n\to\infty} \|P_{\mathcal{W}}(\mathbf{u}_n) - P_{\mathcal{W}}(\mathbf{p})\| = \|\hat{\mathbf{v}} - \mathbf{p}\| = \tau$, *so* ρ *is well-defined.*

If $\rho = 0$, *we have* $a = \tau$, *whereby*

$$\begin{aligned}\lim_{n\to\infty} \|\mathbf{u}_n - \hat{\mathbf{v}}\|^2 &\overset{(i)}{=} \lim_{n\to\infty} \left(\|\mathbf{u}_n - P_{\mathcal{W}}(\mathbf{u}_n)\|^2 + \|P_{\mathcal{W}}(\mathbf{u}_n) - \hat{\mathbf{v}}\|^2\right) \\ &\overset{(ii)}{=} \lim_{n\to\infty} \left(\|\mathbf{u}_n - \mathbf{p}\|^2 - \|\mathbf{p} - P_{\mathcal{W}}(\mathbf{u}_n)\|^2 + \|P_{\mathcal{W}}(\mathbf{u}_n) - \hat{\mathbf{v}}\|^2\right) \\ &\overset{(iii)}{=} a^2 - \tau^2 + 0 = 0.\end{aligned}$$

Here, (i) and (ii) follow the Pythagorean theorem and the properties of a projection onto a hyperplane and (iii) follows from (2.7), (2.8) *and* (2.6). *Thus for* $\rho = 0$ *it holds that* $\lim_{n\to\infty} \mathbf{u}_n = \hat{\mathbf{v}}$.

For the case $\rho > 0$, *we define* $\left(\forall \varepsilon \in \left[0, \sqrt{2}\rho\right]\right)$

$$\mathcal{S}(\varepsilon) := \{\mathbf{x} \in \mathcal{H} \mid \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\| \leq \delta_\varepsilon, |\|\mathbf{x} - \mathbf{p}\| - a| \leq \delta_\varepsilon\}, \tag{2.10}$$

where

$$\delta_\varepsilon := \frac{\rho\frac{\varepsilon}{\sqrt{2}} - \frac{\varepsilon^2}{4}}{a+\tau}. \tag{2.11}$$

Note that $\rho > 0 \implies a + \tau > 0$, *so* δ_ε *in* (2.11) *is well-defined. Moreover, observe that*

$$\begin{aligned}(\forall \mathbf{x} \in \mathcal{S}(\varepsilon)) \quad \tau - \delta_\varepsilon &\overset{(i)}{\leq} \|\mathbf{p} - \hat{\mathbf{v}}\| - \|\hat{\mathbf{v}} - P_{\mathcal{W}}(\mathbf{x})\| \overset{(ii)}{\leq} \|P_{\mathcal{W}}(\mathbf{x}) - \mathbf{p}\| \\ &\overset{(iii)}{\leq} \|\mathbf{p} - \hat{\mathbf{v}}\| + \|\hat{\mathbf{v}} - P_{\mathcal{W}}(\mathbf{x})\| \overset{(iv)}{\leq} \tau + \delta_\varepsilon,\end{aligned} \tag{2.12}$$

where (i) and (iv) follow from (2.8) *and* (2.10) *and (ii) and (iii) are applications of the triangle inequality. By the Pythagorean theorem,* $(\forall \mathbf{x} \in \mathcal{H})$ $\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\|^2 = \|\mathbf{x} - \mathbf{p}\|^2 - \|P_{\mathcal{W}}(\mathbf{x}) - \mathbf{p}\|^2$, *so according to* (2.10) *and* (2.12) *we can write*

$$(\forall \mathbf{x} \in \mathcal{S}(\varepsilon)) \quad \underbrace{(a-\delta_\varepsilon)^2 - (\tau+\delta_\varepsilon)^2}_{\substack{=a^2-\tau^2-2(a+\tau)\delta_\varepsilon \\ =\rho^2-2(a+\tau)\delta_\varepsilon}} \leq \|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\|^2 \leq \underbrace{(a+\delta_\varepsilon)^2 - (\tau-\delta_\varepsilon)^2}_{\substack{=a^2-\tau^2+2(a+\tau)\delta_\varepsilon \\ =\rho^2+2(a+\tau)\delta_\varepsilon}}.$$

This implies that[a] $(\forall \mathbf{x} \in \mathcal{S}(\varepsilon))$

$$\rho - \|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| \leq \rho - \sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon} = (1 - \sqrt{1-\gamma})\rho \tag{2.13}$$

$$\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| - \rho \leq \sqrt{\rho^2 + 2(a+\tau)\delta_\varepsilon} - \rho = (\sqrt{1+\gamma} - 1)\rho, \tag{2.14}$$

where we substituted $\gamma := 2(a+\tau)\delta_\varepsilon\rho^{-2}$. *Note that* $\left(\forall \varepsilon \in \left[0, \sqrt{2}\rho\right]\right)$ $\gamma \in [0,1]$, *and that* $(\forall \gamma \in [0,1])$ $1 - \sqrt{1-\gamma} \geq \sqrt{1+\gamma} - 1$, *so*

$$\left(\forall \varepsilon \in \left[0, \sqrt{2}\rho\right]\right) \quad \rho - \sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon} \geq \sqrt{\rho^2 + 2(a+\tau)\delta_\varepsilon} - \rho.$$

Thus we can summarize (2.13) *and* (2.14) *as*

$$(\forall \mathbf{x} \in \mathcal{S}(\varepsilon)) \quad |\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| - \rho| \leq \rho - \sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon}. \tag{2.15}$$

For (2.11) *to hold,* ε *must satisfy* $-\frac{\varepsilon^2}{4} + \rho\frac{\varepsilon}{\sqrt{2}} - \delta_\varepsilon(a+\tau) = 0$, *i.e.,*

$$\varepsilon \in \left\{\sqrt{2}\rho \mp \sqrt{2}\sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon}\right\}.$$

This implies that $\varepsilon \geq \sqrt{2}\left(\rho - \sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon}\right)$, *so by* (2.15), *we have*

$$(\forall \mathbf{x} \in \mathcal{S}(\varepsilon)) \quad |\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| - \rho| \leq \rho - \sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon} \leq \frac{\varepsilon}{\sqrt{2}}. \tag{2.16}$$

By (2.9), *it holds that* $\rho \leq a \leq a + \tau$, *so according to* (2.10) *and* (2.11), *it holds that*

$$(\forall \mathbf{x} \in \mathcal{S}(\varepsilon)) \quad \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\| \leq \delta_\varepsilon \leq \frac{\varepsilon}{\sqrt{2}} \cdot \frac{\rho}{a+\tau} \leq \frac{\varepsilon}{\sqrt{2}}. \tag{2.17}$$

By definition of $\mathcal{W}$, *there exists* $\nu \in \mathbb{R}$ *such that* $(\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})) = \nu\mathbf{e}$. *Hence, according to* (2.16) *and* (2.17), *we can write*

if $\nu \geq 0$:

$$\begin{aligned}\|\mathbf{x} - (\hat{\mathbf{v}} + \rho\mathbf{e})\|^2 &\overset{(i)}{=} \|(\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})) - \rho\mathbf{e}\|^2 + \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\|^2 \\ &= |\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| - \rho|^2 + \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\|^2 \leq \varepsilon^2,\end{aligned}$$

if $\nu \leq 0$:

$$\begin{aligned}\|\mathbf{x} - (\hat{\mathbf{v}} - \rho\mathbf{e})\|^2 &\overset{(i)}{=} \|(\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})) + \rho\mathbf{e}\|^2 + \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\|^2 \\ &= |\|\mathbf{x} - P_{\mathcal{W}}(\mathbf{x})\| - \rho|^2 + \|P_{\mathcal{W}}(\mathbf{x}) - \hat{\mathbf{v}}\|^2 \leq \varepsilon^2,\end{aligned}$$

where (i) follows from the Pythagorean theorem. Consequently, $\left(\forall \varepsilon \in \left[0, \sqrt{2}\rho\right]\right) \mathcal{S}(\varepsilon) \subset \mathcal{B}(\varepsilon) := \mathcal{B}_1(\varepsilon) \cup \mathcal{B}_2(\varepsilon)$, *where*

$$\begin{aligned}\mathcal{B}_1(\varepsilon) &:= \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x} - (\hat{\mathbf{v}} + \rho\mathbf{e})\| \leq \varepsilon\}, \\ \mathcal{B}_2(\varepsilon) &:= \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x} - (\hat{\mathbf{v}} - \rho\mathbf{e})\| \leq \varepsilon\}.\end{aligned}$$

Now, we fix $\varepsilon \in (0, \rho/2) \subset \left[0, \sqrt{2}\rho\right]$. *Since* $\lim_{n\to\infty} \|P_{\mathcal{W}}(\mathbf{u}_n) - \hat{\mathbf{v}}\| = 0$ *and* $\lim_{n\to\infty} \|\mathbf{u}_n - P_{\mathcal{C}\cap\mathcal{W}}(\hat{\mathbf{v}})\| = a$, *there exists* $N_1 \in \mathbb{N}$ *such that*

$$(\forall n \geq N_1) \quad \mathbf{u}_n \in \mathcal{S}(\varepsilon) \subset \mathcal{B}(\varepsilon).$$

Since $(\forall \mathbf{z} \in \mathcal{C})$ $(\|\mathbf{u}_n - \mathbf{z}\|)_{n\in\mathbb{N}}$ *converges and* $(\gamma_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, (2.5) *implies that* $\lim_{n\to\infty} \|\mathbf{u}_n - \mathbf{u}_{n+1}\| = 0$, *so there exists* $N_2 \in \mathbb{N}$ *such that*

$$(\forall n \geq N_2) \quad \|\mathbf{u}_n - \mathbf{u}_{n+1}\| < 2\varepsilon < \rho,$$

which ensures the unique existence of $i \in \{1,2\}$ *satisfying*

$$(\forall n \geq N_2) \quad \mathbf{u}_n \in \mathcal{B}_i(\varepsilon).$$

This implies the strong convergence of $(\mathbf{u}_n)_{n\in\mathbb{N}}$ *to either* $\hat{\mathbf{v}} + \rho\mathbf{e}$ *or* $\hat{\mathbf{v}} - \rho\mathbf{e}$. □

[a] Note that $\sqrt{\rho^2 - 2(a+\tau)\delta_\varepsilon}$ is well-defined because the definition in (2.11) ensures that $(\forall \varepsilon \in \left[0, \sqrt{2}\rho\right])\ 2(a+\tau)\delta_\varepsilon \leq \rho^2$.

Finally, Theorem 2.3, which is based on [YO05, Theorem 2], states the main result of this section. It shows that perturbed versions of the APSM essentially enjoy the same convergence guarantees as their unperturbed counterpart in [YO05], except for monotone approximation.

Theorem 2.3. *Let* $(\Theta_n : \mathcal{H} \to \mathbb{R}_+)_{n\in\mathbb{N}}$ *be a sequence of continuous convex functions, let* $\mathcal{K} \subset \mathcal{H}$ *be a nonempty closed convex set, and denote the APSM update for the* n*th iteration by*[a] $(\forall n \in \mathbb{N})\ T_n : \mathcal{H} \to \mathcal{H}$

$$T_n(\mathbf{x}) = \begin{cases} P_{\mathcal{K}}\left(\mathbf{x} - \lambda_n \frac{\Theta_n(\mathbf{x})}{\|\Theta_n'(\mathbf{x})\|^2}\Theta_n'(\mathbf{x})\right) & \text{if } \Theta_n'(\mathbf{x}) \neq \mathbf{0}, \\ P_{\mathcal{K}}(\mathbf{x}) & \text{otherwise,} \end{cases} \tag{2.18}$$

where $\Theta_n'(\mathbf{x}_n) \in \partial\Theta_n(\mathbf{x}_n)$ *and* $\lambda_n \in [0,2]$. *Moreover, let* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ *be a sequence of bounded perturbations, define*

$$(\forall n \in \mathbb{N}) \quad \Omega_n := \left\{\mathbf{x} \in \mathcal{K} \;\middle|\; \Theta_n(\mathbf{x}) = \Theta_n^\star := \inf_{\mathbf{x}\in\mathcal{K}} \Theta_n(\mathbf{x})\right\},$$

and suppose that

$$(\forall n \in \mathbb{N}) \quad \Theta_n^\star = 0 \quad \text{and} \quad \Omega := \bigcap\nolimits_{n\in\mathbb{N}} \Omega_n \neq \emptyset. \tag{2.19}$$

Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}} \subset \mathcal{K}$ *generated by the perturbed APSM*[b]

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_n\left(\mathbf{x}_n + \beta_n\mathbf{y}_n\right), \quad \mathbf{x}_0 \in \mathcal{K} \tag{2.20}$$

satisfies the following:

(a) The sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér monotone of Type-I relative to* Ω, *so* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is bounded.*

(b) Moreover, if in addition to (2.19), $(\exists(\varepsilon_1,\varepsilon_2) \in \mathbb{R}_+^2)\ (\forall n \in \mathbb{N})\ \lambda_n \in [\varepsilon_1, 2-\varepsilon_2] \subset$

$(0,2)$ *and* $(\Theta_n'(\mathbf{x}_n+\beta_n\mathbf{y}_n))_{n\in\mathbb{N}}$ *is bounded, then* $\lim_{n\to\infty}\Theta_n(\mathbf{x}_n+\beta_n\mathbf{y}_n)=0$.

(c) *Assume* (2.19) *and suppose that* Ω *has some relative interior w.r.t. a hyperplane* $\mathcal{W}\subset\mathcal{H}$, *i.e.,* $(\exists\tilde{\mathbf{u}}\in\Omega\cap\mathcal{W})$ *and* $(\exists\varepsilon>0)$ *satisfying* $\mathcal{U}:=\{\mathbf{u}\in\mathcal{W}\mid\|\mathbf{u}-\tilde{\mathbf{u}}\|\leq\varepsilon\}\subset\Omega$. *Then by using* $(\forall n\in\mathbb{N})\ \lambda_n\in[\varepsilon_1,2-\varepsilon_2]\subset(0,2)$, *the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *in* (2.20) *converges strongly to a point* $\hat{\mathbf{u}}\in\mathcal{K}$, *i.e.,* $\lim_{n\to\infty}\|\mathbf{x}_n-\hat{\mathbf{u}}\|=0$. *Moreover,* $\lim_{n\to\infty}\Theta_n(\hat{\mathbf{u}})=0$ *provided that (i)* $(\Theta_n'(\mathbf{x}_n+\beta_n\mathbf{y}_n))_{n\in\mathbb{N}}$ *is bounded and (ii) there exists bounded* $(\Theta_n'(\hat{\mathbf{u}}))_{n\in\mathbb{N}}$, *where* $(\forall n\in\mathbb{N})\ \Theta_n'(\hat{\mathbf{u}})\in\partial\Theta_n(\hat{\mathbf{u}})$.

(d) *In addition to* (2.19) *and the conditions (i) and (ii) in (c), assume that* Ω *has an interior point* $\tilde{\mathbf{u}}$, *i.e.,* $(\exists\rho>0)$ *satisfying* $\{\mathbf{v}\in\mathcal{H}\mid\|\mathbf{v}-\tilde{\mathbf{u}}\|\leq\rho\}\subset\Omega$. *Define* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *by using* $(\forall n\in\mathbb{N})\ \lambda_n\in[\varepsilon_1,2-\varepsilon_2]\subset(0,2)$, *and let* $\hat{\mathbf{u}}:=\lim_{n\to\infty}\mathbf{x}_n\subset\mathcal{K}$ *(the existence of* $\hat{\mathbf{u}}$ *is guaranteed by (c)). In this case, if*

$$(\forall\varepsilon>0)(\forall r>0)(\exists\delta>0)\quad\inf_{\substack{d(\mathbf{x}_n,\operatorname{lev}_{\leq 0}\Theta_n)\geq\varepsilon\\ \|\tilde{\mathbf{u}}-\mathbf{x}_n\|\leq r}}\Theta_n(\mathbf{x}_n)\geq\delta,$$

the limit $\hat{\mathbf{u}}$ *satisfies* $\hat{\mathbf{u}}\in\overline{\liminf_{n\to\infty}\Omega_n}$, *where* $\liminf_{n\to\infty}\Omega_n:=\bigcup_{n=0}^{\infty}\bigcap_{k\geq n}\Omega_k$ *and the overbar denotes the closure of a set.*

Proof.

(a) *Note that* $(\forall n\in\mathbb{N})\ \operatorname{lev}_{\leq 0}\Theta_n\neq\emptyset$ *by assumption. Hence by Fact 1.11,* $(\forall n\in\mathbb{N})$ *the mapping* T_n *is quasi-nonexpansive. According to Proposition 2.1, the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér monotone of Type-I relative to* Ω. *This in turn implies that* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is bounded (see Fact 1.15).*

(b) *Introducing the shorthand* $(\forall n\in\mathbb{N})\ \mathbf{z}_n:=\mathbf{x}_n+\beta_n\mathbf{y}_n$ *and*

$$\boldsymbol{\Phi}_n=\begin{cases}\lambda_n\frac{\Theta_n(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2}\Theta_n'(\mathbf{z}_n) & \text{if } \Theta_n'(\mathbf{z}_n)\neq\mathbf{0},\\ \mathbf{0} & \text{otherwise}\end{cases}\tag{2.21}$$

we can write $(\forall n\in\mathbb{N})\ (\forall\mathbf{x}_n\in\mathcal{K})\ (\forall\mathbf{z}\in\Omega)$

$$\begin{aligned}\|\mathbf{x}_{n+1}-\mathbf{z}\|^2&=\|P_{\mathcal{K}}(\mathbf{z}_n-\boldsymbol{\Phi}_n)-P_{\mathcal{K}}(\mathbf{z})\|^2\\&\overset{(i)}{\leq}\|\mathbf{x}_n+\beta_n\mathbf{y}_n-\boldsymbol{\Phi}_n-\mathbf{z}\|^2\\&=\|\mathbf{x}_n-\mathbf{z}\|^2+2\langle\beta_n\mathbf{y}_n-\boldsymbol{\Phi}_n,\mathbf{x}_n-\mathbf{z}\rangle\\&\quad+\|\beta_n\mathbf{y}_n-\boldsymbol{\Phi}_n\|^2\end{aligned}$$

$$
\begin{aligned}
&= \|\mathbf{x}_n - \mathbf{z}\|^2 + 2\langle \beta_n \mathbf{y}_n - \mathbf{\Phi}_n, \mathbf{x}_n - \mathbf{z}\rangle \\
&\quad + \beta_n^2 \|\mathbf{y}_n\|^2 - 2\langle \mathbf{\Phi}_n, \beta_n \mathbf{y}_n\rangle + \|\mathbf{\Phi}_n\|^2 \\
&= \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\langle \mathbf{\Phi}_n, \mathbf{x}_n + \beta_n \mathbf{y}_n - \mathbf{z}\rangle \\
&\quad + 2\beta_n \langle \mathbf{y}_n, \mathbf{x}_n - \mathbf{z}\rangle + \beta_n^2 \|\mathbf{y}_n\|^2 + \|\mathbf{\Phi}_n\|^2 \\
&\overset{(ii)}{\leq} \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\langle \mathbf{\Phi}_n, \mathbf{z}_n - \mathbf{z}\rangle + \|\mathbf{\Phi}_n\|^2 \\
&\quad + 2\beta_n \|\mathbf{y}_n\| \|\mathbf{x}_n - \mathbf{z}\| + \beta_n^2 \|\mathbf{y}_n\|^2
\end{aligned}
$$

where (i) follows from nonexpansivity of $P_{\mathcal{K}}$, and (ii) is an application of the Cauchy-Schwarz inequality.

Since $(\mathbf{x}_n)_{n\in\mathbb{N}}$ and $(\mathbf{y}_n)_{n\in\mathbb{N}}$ are bounded, for any bounded subset $\mathcal{U} \subset \Omega$ there exists $r > 0$ such that $(\forall \mathbf{z} \in \mathcal{U})$ $(\forall n \in \mathbb{N})$ $\|\mathbf{x}_n - \mathbf{z}\| \leq r$ and $\|\mathbf{y}_n\| \leq r$. Hence by defining $c := (2r^2 + r^2 \sum_{n\in\mathbb{N}} \beta_n)$ and $(\gamma_n)_{n\in\mathbb{N}} := (c\beta_n)_{n\in\mathbb{N}}$ we have $(\forall \mathbf{z} \in \mathcal{U})$

$$
\begin{aligned}
\|\mathbf{x}_{n+1} - \mathbf{z}\|^2 &\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\langle \mathbf{\Phi}_n, \mathbf{z}_n - \mathbf{z}\rangle + \|\mathbf{\Phi}_n\|^2 + 2\beta_n r^2 + \beta_n^2 r^2 \\
&\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\langle \mathbf{\Phi}_n, \mathbf{z}_n - \mathbf{z}\rangle + \|\mathbf{\Phi}_n\|^2 + \gamma_n.
\end{aligned}
$$

If $\Theta_n(\mathbf{z}_n) = 0$ or $\Theta_n'(\mathbf{z}_n) = \mathbf{0}$, (2.21) yields $\mathbf{\Phi}_n = \mathbf{0}$, whereby

$$
\|\mathbf{x}_{n+1} - \mathbf{z}\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 + \gamma_n. \tag{2.22}
$$

Otherwise, i.e., if $\Theta_n(\mathbf{z}_n) \neq 0$ and $\Theta_n'(\mathbf{z}_n) \neq \mathbf{0}$, it follows from (1.1) that

$$
\begin{aligned}
\|\mathbf{x}_{n+1} - \mathbf{z}\|^2 &\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\lambda_n \frac{\Theta_n(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} \langle \Theta_n'(\mathbf{z}_n), \mathbf{z}_n - \mathbf{z}\rangle \\
&\quad + \lambda_n^2 \frac{\Theta_n(\mathbf{z}_n)^2}{\|\Theta_n'(\mathbf{z}_n)\|^2} + \gamma_n \\
&\leq \|\mathbf{x}_n - \mathbf{z}\|^2 - 2\lambda_n \frac{\Theta_n(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} (\Theta_n(\mathbf{z}_n) - \Theta_n(\mathbf{z})) \\
&\quad + \lambda_n^2 \frac{\Theta_n(\mathbf{z}_n)^2}{\|\Theta_n'(\mathbf{z}_n)\|^2} + \gamma_n \\
&= \|\mathbf{x}_n - \mathbf{z}\|^2 \\
&\quad - \lambda_n \left(2\left(1 - \frac{\Theta_n(\mathbf{z})}{\Theta_n(\mathbf{z}_n)}\right) - \lambda_n\right) \frac{\Theta_n^2(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} + \gamma_n.
\end{aligned} \tag{2.23}
$$

Since $(\forall n \in \mathbb{N})\ \lambda_n \in [\varepsilon_1, 2-\varepsilon_2]$ *and* $(\forall n \in \mathbb{N})\ (\forall \mathbf{z} \in \mathcal{U} \subset \Omega)\ \Theta_n(\mathbf{z}) = 0$, *we have*

$$\lambda_n \left(2\left(1 - \frac{\Theta_n(\mathbf{z})}{\Theta_n(\mathbf{z}_n)}\right) - \lambda_n\right) \geq \varepsilon_1\varepsilon_2.$$

Hence, by defining a sequence

$$(\forall n \in \mathbb{N}) \quad c_n := \begin{cases} 0 & \text{if } \Theta_n'(\mathbf{z}_n) = \mathbf{0} \\ \varepsilon_1\varepsilon_2 \frac{\Theta_n^2(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} & \text{otherwise,} \end{cases}$$

we can summarize (2.22) *and* (2.23) *as*

$$(\forall n \in \mathbb{N}) \quad \|\mathbf{x}_{n+1} - \mathbf{z}\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 - c_n + \gamma_n.$$

Because (1.1) *ensures that* $\Theta_n'(\mathbf{x}) = \mathbf{0} \implies \Theta_n(\mathbf{x}) = \Theta_n^\star = 0$, *it is sufficient to consider the case* $\Theta_n'(\mathbf{z}_n) \neq \mathbf{0}$. *Moreover, if this case occurs finitely many times, there exists* N_0 *such that* $(\forall n \geq N_0)\ \Theta_n^\star = 0$. *Thus it remains to show that* $\lim_{k\to\infty} \Theta_{n_k}(\mathbf{z}_{n_k}) = 0$, *where* $(n_k)_{k\in\mathbb{N}}$ *is the subsequence of* $(n)_{n\in\mathbb{N}}$ *comprised of all elements of the infinite set* $\mathcal{J} := \{n \in \mathbb{N} \mid \Theta_n'(\mathbf{x}) \neq \mathbf{0}\}$. *According to Fact 1.14,* $(\|\mathbf{x}_n - \mathbf{z}\|^2)_{n\in\mathbb{N}}$ *converges and* $(c_n)_{n\in\mathbb{N}}$ *is summable, whereby* $\sum_{n\in\mathcal{J}} \varepsilon_1\varepsilon_2 \frac{\Theta_n^2(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} < \infty$. *Moreover, since* $(\forall n \in \mathcal{J})\ \varepsilon_1\varepsilon_2 \frac{\Theta_n^2(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2} \geq 0$, *it follows that*

$$\lim_{k\to\infty} \varepsilon_1\varepsilon_2 \frac{\Theta_{n_k}^2(\mathbf{z}_{n_k})}{\|\Theta_{n_k}'(\mathbf{z}_{n_k})\|^2} = 0.$$

Therefore, boundedness of $(\Theta_n'(\mathbf{x}_n + \beta_n\mathbf{y}_n))_{n\in\mathbb{N}}$ *ensures that* $\lim_{n\to\infty} \Theta_n(\mathbf{x}_n + \beta_n\mathbf{y}_n) = 0$.

(c) It holds by assumption in (2.19) *that* $(\forall n \in \mathbb{N})\ \mathcal{K} \cap \operatorname{lev}_{\leq 0}\Theta_n \neq \emptyset$. *Therefore, according to Fact 1.11, the mapping* T_n *in* (2.18) *is* $\left(1 - \frac{\lambda_n}{2}\right)$*-attracting quasi-nonexpansive. Since the set* $\mathcal{U} \subset \Omega$ *is bounded and* $(\forall n \in \mathbb{N})\ \lambda_n \leq 2 - \varepsilon_2$, *Proposition 2.2 implies that* $(\exists (\gamma_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall \mathbf{z} \in \mathcal{U})$

$$\frac{\varepsilon_2}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 \leq \|\mathbf{x}_n - \mathbf{z}\|^2 - \|\mathbf{x}_{n+1} - \mathbf{z}\|^2 + \gamma_n.$$

Consequently, since $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is quasi-Fejér of Type-I relative to* Ω *(see (a)), Theorem 2.2 guarantees that the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *converges strongly to a point* $\hat{\mathbf{u}} \in \mathcal{H}$. *More precisely, it holds that* $\hat{\mathbf{u}} \in \mathcal{K}$, *because* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is a sequence in the closed set* $\mathcal{K}$. *Since* $(\beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$ *are bounded perturbations, the sequence* $(\mathbf{z}_n := \mathbf{x}_n + \beta_n\mathbf{y}_n)_{n\in\mathbb{N}}$

satisfies $\lim_{n\to\infty} \mathbf{z}_n = \lim_{n\to\infty} \mathbf{x}_n = \hat{\mathbf{u}}$. *By assumption (ii) there exists* $R > 0$ *such that* $(\forall n \in \mathbb{N})\ \|\Theta_n'(\hat{\mathbf{u}})\| \leq R$. *Thus we can use Fact 1.9 and the Cauchy-Schwartz inequality to obtain*

$$\begin{aligned} 0 \leq \Theta_n(\hat{\mathbf{u}}) &\leq \Theta_n(\mathbf{z}_n) - \langle \mathbf{z}_n - \hat{\mathbf{u}}, \Theta_n'(\hat{\mathbf{u}}) \rangle \\ &\leq \Theta_n(\mathbf{z}_n) + R\|\mathbf{z}_n - \hat{\mathbf{u}}\| \to 0. \end{aligned}$$

(d) The proof is identical to the proof of [YO05, Theorem 2(d)]. □

[a] The projection onto $\mathcal{K}$ for $\Theta_n'(\mathbf{x}) = \mathbf{0}$ ensures that the perturbed APSM generates a sequence in $\mathcal{K}$ regardless of the perturbations. It is not part of the definition in [YO05], where the absence of perturbations guarantees that $(\forall n \in \mathbb{N})\ \mathbf{u}_n \in \mathcal{K}$.

[b] We can assume without loss of generality that $\mathbf{x}_0 \in \mathcal{K}$, since $(\forall \mathbf{x} \in \mathcal{H})\ T_0(\mathbf{x}) \in \mathcal{K}$.

Remark 2.2. We note that in [YO05], the condition in (2.19) does not concern the initial n_0 iterations, allowing for a finite number of cost functions that lead to an empty intersection of zero level sets. Nevertheless, Theorem. 2.3 still covers this case if we let $\mathbf{x}_0 := \tilde{\mathbf{x}}_{n_0}$, where $\tilde{\mathbf{x}}_{n_0} \in \mathcal{H}$ denotes the estimate after the first n_0 iterations.

2.3. Extrapolated Alternating Projection Methods

Many problems arising in practice can be posed in terms of the two-set feasibility problem[1]

$$\text{find } \mathbf{x} \in \mathcal{A} \cap \mathcal{B}, \tag{2.24}$$

where $\mathcal{A} \subset \mathcal{H}$ and $\mathcal{B} \subset \mathcal{H}$ are closed convex sets such that $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Clearly, Problem 2.24 can be solved using alternating projection methods such as POCS. However, it is well-known that the convergence of sequences generated by this approach can be very slow [GPR67], [CCC+12]. To mitigate this problem, the authors of [GPR67] have proposed the following extrapolated alternating projection method. Starting from an arbitrary point $\mathbf{x}_0 \in \mathcal{H}$, this algorithm produces a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ via

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T^{\mathrm{GPR}} P_{\mathcal{A}}(\mathbf{x}_n),$$

where the mapping $T^{\mathrm{GPR}} : \mathcal{A} \to \mathcal{H}$ is given by

$$T^{\mathrm{GPR}}(\mathbf{x}) = \mathbf{x} + \sigma(\mathbf{x})(P_{\mathcal{A}} P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}),$$

[1] In Remark 2.3 at the end of this section, we show that the multi-set convex feasibility problem in (2.1) can be expressed in terms of a two-set convex feasibility problem in a product Hilbert space.

and the function $\sigma : \mathcal{A} \to [1, \infty)$ defines an extrapolation factor by

$$\sigma(\mathbf{x}) = \begin{cases} \frac{\|P_{\mathcal{B}}(\mathbf{x})-\mathbf{x}\|^2}{\langle P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x})-\mathbf{x}, P_{\mathcal{B}}(\mathbf{x})-\mathbf{x}\rangle} & \text{if } \mathbf{x} \notin \mathcal{B} \\ 1 & \text{otherwise.} \end{cases} \tag{2.25}$$

In [Ceg12, Section 5.2.1.1], a more general version of this algorithm is introduced, which uses a relaxed mapping $T_\lambda^{\mathrm{GPR}} : \mathcal{A} \to \mathcal{H}$ given by

$$T_\lambda^{\mathrm{GPR}} := I + \lambda(T^{\mathrm{GPR}} - I), \tag{2.26}$$

where $\lambda \in (0, 2)$ is a relaxation parameter. The algorithm proposed in [Ceg12, Section 5.2.1.1] generates a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ via the recursion

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathcal{A}} T_{\lambda_n}^{\mathrm{GPR}}(\mathbf{x}_n), \quad \mathbf{x}_0 \in \mathcal{A}, \tag{2.27}$$

where $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2)$.

In the following, we investigate the bounded perturbation resilience of this algorithm. Since adding perturbations to the iterate $\mathbf{x}_n$ may result in points $\mathbf{x}_n + \beta_n \mathbf{y}_n \notin \mathcal{A} = \mathrm{dom}(T^{\mathrm{GPR}})$, we will use the mapping $T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}$ instead of $P_{\mathcal{A}} T_{\lambda_n}^{\mathrm{GPR}}$ to define a sequence of estimates.

Fact 2.1. *[Ceg12, Corollary 4.3.14, Remark 2.4.2(d)] Let $\lambda \in (0, 2)$. Then the mapping T_λ^{GPR} defined in (2.26) is $\frac{2-\lambda}{\lambda}$-attracting quasi-nonexpansive with fixed point set $\mathrm{Fix}(T_\lambda^{\mathrm{GPR}}) = \mathcal{A} \cap \mathcal{B}$.*

Theorem 2.4. *Consider two closed convex sets $\mathcal{A} \subset \mathcal{H}$ and $\mathcal{B} \subset \mathcal{H}$, let $(\beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ be a sequence of bounded perturbations in $\mathcal{H}$, and let $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2)$. The sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by the perturbed extrapolated alternating projection method*

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}(\mathbf{x}_n + \beta_n \mathbf{y}_n), \quad \mathbf{x}_0 \in \mathcal{H} \tag{2.28}$$

satisfies the following:

(a) Assume that $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathcal{A} \cap \mathcal{B}$.

(b) Assume that $\mathcal{H}$ is finite dimensional and that $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{A} \cap \mathcal{B}$.

(c) Assume that $\mathcal{A}\cap\mathcal{B}$ has a nonempty relative interior w.r.t. a linear variety $\mathcal{V}\subset\mathcal{H}$, i.e., $(\exists\tilde{\mathbf{u}}\in\mathcal{A}\cap\mathcal{B}\cap\mathcal{V})$ and $(\exists\varepsilon>0)$ satisfying $\mathcal{U}:=\{\mathbf{u}\in\mathcal{V}\mid\|\mathbf{u}-\tilde{\mathbf{u}}\|\leq\varepsilon\}\subset\Omega$. Then the sequence $(P_{\mathcal{V}}(\mathbf{x}_n))_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{V}$.

(d) Assume that $\mathcal{A}\cap\mathcal{B}$ has a nonempty relative interior w.r.t. a hyperplane $\mathcal{W}\subset\mathcal{H}$, i.e., $(\exists\tilde{\mathbf{u}}\in\mathcal{A}\cap\mathcal{B}\cap\mathcal{W})$ and $(\exists\varepsilon>0)$ satisfying $\mathcal{U}:=\{\mathbf{u}\in\mathcal{W}\mid\|\mathbf{u}-\tilde{\mathbf{u}}\|\leq\varepsilon\}\subset\Omega$. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in (2.28) converges strongly to a point in $\mathcal{A}\cap\mathcal{B}$.

Proof.

(a) By Fact 1.8, Fact 2.1, and firm (quasi-)nonexpansivity of $P_{\mathcal{A}}$, the mapping $T_\lambda^{\mathrm{GPR}}P_{\mathcal{A}}$ is $\left(1-\frac{\lambda}{2}\right)$-attracting quasi-nonexpansive with $\mathrm{Fix}(T_\lambda^{\mathrm{GPR}}P_{\mathcal{A}})=\mathcal{A}\cap\mathcal{B}$. Since $(\forall n\in\mathbb{N})\ \lambda_n\leq 2-\varepsilon_2$, $T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}$ is $\frac{\varepsilon_2}{2}$-attracting quasi-nonexpansive, which implies the following:

- *According to Proposition 2.1, $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is quasi-Fejér of Type-I relative to $\mathcal{A}\cap\mathcal{B}$. Therefore, Proposition 2.3 shows that the sequence $(\mathbf{z}_n)_{n\in\mathbb{N}}$ defined by $(\forall n\in\mathbb{N})\ \mathbf{z}_n:=\mathbf{x}_n+\beta_n\mathbf{y}_n$ is also quasi-Fejér of Type-I relative to $\mathcal{A}\cap\mathcal{B}$, so Fact 1.15 ensures that $(\mathbf{z}_n)_{n\in\mathbb{N}}$ is bounded.*
- *By Proposition 2.2 $(\exists\mathbf{z}\in\mathcal{A}\cap\mathcal{B})$ $(\exists(\gamma_n)_{n\in\mathbb{N}}\in\ell_+^1(\mathbb{N}))$ such that*

$$\|\mathbf{x}_{n+1}-\mathbf{z}\|^2\leq\|\mathbf{x}_n-\mathbf{z}\|^2-\kappa\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2+\gamma_n,$$

so Fact 1.14 ensures that $\lim_{n\to\infty}\|\mathbf{x}_{n+1}-\mathbf{x}_n\|=0$.

Consequently, we have that

$$\begin{aligned}
0&=\lim_{n\to\infty}\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2\\
&=\lim_{n\to\infty}\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n+\beta_n\mathbf{y}_n\|^2\\
&=\lim_{n\to\infty}\Big(\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|^2+\beta_n^2\|\mathbf{y}_n\|^2+2\beta_n\langle T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n,\mathbf{y}_n\rangle\Big)\\
&\geq\lim_{n\to\infty}\Big(\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|^2+\underbrace{\beta_n^2\|\mathbf{y}_n\|^2}_{\overset{(i)}{\to}0}-\underbrace{2\beta_n\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|\cdot\|\mathbf{y}_n\|}_{\overset{(ii)}{\to}0}\Big)\\
&=\lim_{n\to\infty}\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|^2\geq 0.
\end{aligned}$$

Here (i) follows from boundedness of $(\mathbf{y}_n)_{n\in\mathbb{N}}$ and (ii) follows from boundedness of $(\mathbf{y}_n)_{n\in\mathbb{N}}$ and $(\mathbf{z}_n)_{n\in\mathbb{N}}$, quasi-nonexpansivity of $T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}$, and Proposition 2.4.

Next, we show that $\|P_{\mathcal{A}}P_{\mathcal{B}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|\to 0$ whenever $\|T_{\lambda_n}^{\mathrm{GPR}}P_{\mathcal{A}}(\mathbf{z}_n)-\mathbf{z}_n\|\to 0$.

To see this, we substitute

$$\begin{aligned}
&\|T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\|^2 \\
&= \|P_{\mathcal{A}}(\mathbf{z}_n) + \lambda_n \sigma(P_{\mathcal{A}}(\mathbf{z}_n)) \left(P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n)\right) - \mathbf{z}_n\|^2 \\
&= \|P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\|^2 + \|\lambda_n \sigma(P_{\mathcal{A}}(\mathbf{z}_n)) \left(P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n)\right)\|^2 \\
&\quad + \underbrace{2\lambda_n \sigma(P_{\mathcal{A}}(\mathbf{z}_n))}_{\geq 0} \underbrace{\langle P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n), P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n \rangle}_{\geq 0},
\end{aligned}$$

where nonnegativity of the last term follows from Fact 1.1. Hence we can write

$$\|T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\|^2 \geq \|P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\|^2 + \underbrace{\lambda_n \sigma(P_{\mathcal{A}}(\mathbf{z}_n))}_{\geq \varepsilon_1} \|P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n)\|^2,$$

where $\sigma(P_{\mathcal{A}}(\mathbf{z}_n)) \geq 1$ and $\lambda_n \geq \varepsilon_1 > 0$ is bounded away from zero. As a result, nonnegativity of the norm and $\|T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\| \to 0$ imply that $\|P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\| \to 0$ and $\|P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n)\| \to 0$. Hence the reverse triangle inequality yields

$$\begin{aligned}
0 &= \lim_{n\to\infty} \|P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - P_{\mathcal{A}}(\mathbf{z}_n)\| \\
&\geq \lim_{n\to\infty} \Big(\|P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\| - \underbrace{\|\mathbf{z}_n - P_{\mathcal{A}}(\mathbf{z}_n)\|}_{\to 0} \Big) \\
&= \lim_{n\to\infty} \|P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}(\mathbf{z}_n) - \mathbf{z}_n\| \geq 0.
\end{aligned}$$

Because $P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}$ is nonexpansive with $\mathrm{Fix}(P_{\mathcal{A}} P_{\mathcal{B}} P_{\mathcal{A}}) = \mathcal{A} \cap \mathcal{B}$, Fact 1.16 guarantees that all weak cluster points of the bounded sequence $(\mathbf{z}_n)_{n\in\mathbb{N}}$ are contained in $\mathcal{A} \cap \mathcal{B}$. Consequently, as $(\mathbf{z}_n)_{n\in\mathbb{N}}$ is quasi-Fejér of Type-I (and hence by Fact 1.13 of Type-III), Fact 1.17 guarantees the weak convergence of $(\mathbf{z}_n)_{n\in\mathbb{N}}$ to a point $\hat{\mathbf{u}} \in \mathcal{A} \cap \mathcal{B}$. Finally, $\lim_{n\to\infty} \beta_n \mathbf{y}_n = \mathbf{0}$ implies that $(\mathbf{x}_n = \mathbf{z}_n - \beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ converges weakly to $\hat{\mathbf{u}}$.

(b) This is immediate from (a) and Fact 1.12(c).

(c) This follows from Type-I quasi-Fejér monotonicity of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ and Lemma 2.2

(d) Since $(\forall n \in \mathbb{N})$, $T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}$ is $\frac{\varepsilon_2}{2}$-attracting quasi-nonexpansive $\mathrm{Fix}(T_{\lambda_n}^{\mathrm{GPR}} P_{\mathcal{A}}) = \mathcal{A} \cap \mathcal{B}$ has a nonempty relative interior w.r.t. a hyperplane, Proposition 2.2 and Theorem 2.2 imply strong convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ to a point $\hat{\mathbf{u}} \in \mathcal{H}$. As we showed in (a), the set of weak cluster points (which by Fact 1.12(b) includes the strong

limit point $\hat{\mathbf{u}}$*) is a subset of* $\mathcal{A} \cap \mathcal{B}$*, so* $\hat{\mathbf{u}} \in \mathcal{A} \cap \mathcal{B}$. □

A special case of Problem 2.24, where one of the sets (say, $\mathcal{A}$) is an affine subspace, is considered in [BCK06]. The authors propose an extrapolated alternating projection method for affine-convex feasibility problems. Starting from a point $\mathbf{x}_0 \in \mathcal{A}$, this algorithm produces a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ by applying the update rule

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_\lambda^{\mathrm{EAPM}}(\mathbf{x}_n), \tag{2.29}$$

where $\lambda \in (0,2)$ and $(\forall \mathbf{x} \in \mathcal{A})(\forall \lambda \in (0,2)$, the mapping $T_\lambda^{\mathrm{EAPM}} : \mathcal{A} \to \mathcal{A}$ is given by

$$T_\lambda^{\mathrm{EAPM}}(\mathbf{x}) = \mathbf{x} + \lambda K(\mathbf{x}) \left(P_{\mathcal{A}} P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\right). \tag{2.30}$$

The extrapolation factor determined by the function $K : \mathcal{A} \to [1, \infty)$

$$(\forall \mathbf{x} \in \mathcal{A}) \quad K(\mathbf{x}) = \begin{cases} \frac{\|P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\|^2}{\|P_{\mathcal{A}} P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\|^2} & \text{if } \mathbf{x} \notin \mathcal{B} \\ 1 & \text{otherwise} \end{cases} \tag{2.31}$$

was originally introduced in [Pie84]. It was observed in [CCC+12] that the extrapolated alternating projection method in (2.29) can achieve considerable convergence acceleration compared to projection methods without extrapolation, such as POCS.

Theorem 2.5 assesses the bounded perturbation resilience of the algorithm in (2.29). We consider a slightly more general definition of this algorithm, in which the relaxation parameter λ may vary throughout the iterations. Since adding perturbations $\beta_n \mathbf{y}_n$ to $\mathbf{x}_n$ may result in points $\mathbf{x}_n + \beta_n \mathbf{y}_n \notin \mathcal{A} = \mathrm{dom}(T_\lambda^{\mathrm{EAPM}})$, we project the perturbed estimate onto $\mathcal{A}$ before applying the mapping $T_{\lambda_n}^{\mathrm{EAPM}}$ at each iteration n.[2] Note that this modification does not change the unperturbed algorithm in (2.29), because $(\forall \mathbf{x} \in \mathcal{A})\ P_{\mathcal{A}}(\mathbf{x}) = \mathbf{x}$.

Theorem 2.5. *Consider a closed affine subspace* $\mathcal{A} \subset \mathcal{H}$ *and a closed convex set* $\mathcal{B} \subset \mathcal{H}$*, let* $(\beta_n \mathbf{y}_n)_{n\in\mathbb{N}}$ *be a sequence of bounded perturbations in* $\mathcal{H}$*, and let* $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon_1, 2 - \varepsilon_2]$ *for small* $\varepsilon_1, \varepsilon_2 > 0$*. The sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *in* $\mathcal{A}$ *generated by the perturbed extrapolated alternating projection method*

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_{\lambda_n}^{\mathrm{EAPM}} P_{\mathcal{A}}(\mathbf{x}_n + \beta_n \mathbf{y}_n), \quad \mathbf{x}_0 \in \mathcal{A} \tag{2.32}$$

satisfies the following:

[2] Alternatively, we could allow only such perturbations $\mathbf{y}_n$ that lie direction of the affine subspace $\mathcal{A}$, i.e., that satisfy $\mathbf{y}_n + P_{\mathcal{A}}(\mathbf{0}) \in \mathcal{A}$.

(a) If $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathcal{A} \cap \mathcal{B}$.

(b) If $\mathcal{H}$ is finite dimensional and $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{A} \cap \mathcal{B}$.

(c) If $\mathcal{B}$ has a nonempty relative interior w.r.t. the affine subspace $\mathcal{A}$, i.e., $(\exists \tilde{\mathbf{u}} \in \mathcal{A} \cap \mathcal{B})$ and $(\exists \varepsilon > 0)$ satisfying $\mathcal{U} := \{\mathbf{u} \in \mathcal{A} \mid \|\mathbf{u} - \tilde{\mathbf{u}}\| \leq \varepsilon\} \subset \mathcal{B}$, the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{A} \cap \mathcal{B}$.

Proof.

(a) $(\forall \mathbf{x} \in \mathcal{A})$ $P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x} \in \mathcal{A}_0$, where $\mathcal{A}_0 := \mathcal{A} - P_{\mathcal{A}}(\mathbf{0})$ is the linear subspace associated with the affine subspace $\mathcal{A}$. Moreover, $P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{B}}(\mathbf{x})$ is orthogonal to $\mathcal{A}_0$. Therefore $(\forall \mathbf{x} \in \mathcal{A})$

$$\begin{aligned}\|P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\|^2 = \langle P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\rangle \\ + \underbrace{\langle P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{B}}(\mathbf{x})\rangle}_{=0},\end{aligned}$$

i.e., the extrapolation factors in (2.25) and (2.31) satisfy $(\forall \mathbf{x} \in \mathcal{A})$ $K(\mathbf{x}) = \sigma(\mathbf{x})$. Consequently, the algorithm in (2.32) is a particular case of the algorithm in (2.28) for affine sets $\mathcal{A}$, whereby weak convergence of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ to a point in $\mathcal{A} \cap \mathcal{B}$ follows directly from Theorem 2.4(a).

(b) This follows directly from Theorem 2.4(b).

(c) According to Theorem. 2.4(b). the sequence $(P_{\mathcal{A}}(\mathbf{x}_n))_{n\in\mathbb{N}}$ converges strongly to a point $\hat{\mathbf{u}} \in \mathcal{A}$. Moreover, $(\mathbf{x}_n)_{n\in\mathbb{N}} = (P_{\mathcal{A}}(\mathbf{x}_n))_{n\in\mathbb{N}}$ because $(\forall n \in \mathbb{N})$ $\mathbf{x}_n \in \mathcal{A}$. By Fact 1.12(b), the strong limit point $\hat{\mathbf{u}}$ is also a weak cluster point, so it follows from (a) that $\hat{\mathbf{u}} \in \mathcal{A} \cap \mathcal{B}$. □

Although both algorithms in Theorem 2.4 and Theorem 2.5 are extrapolated alternating projection methods, we adopt the naming from [CCC+12], where the algorithm in Theorem 2.5 is referred to as extrapolated alternating projection method (EAPM). Thus we refer to the algorithm in Theorem 2.4 as Gurin-Polyak-Raik (GPR) algorithm, throughout this thesis. Since many practical problems involve more than two constraint sets, it is worth mentioning that the algorithms in this section can be used to solve the more general convex feasibility problem in (2.1) by using Pierra's product space formalism [Pie75, Pie84] (see also [Com97], [SY98, Section 2.9]). We sketch this idea in the following remark.

Remark 2.3. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real Hilbert space with induced norm $\|\cdot\|$ and let $\mathcal{C}_1, \ldots, \mathcal{C}_K$ be a family of closed convex subsets of $\mathcal{H}$ such that $\mathcal{C}_\star = \bigcap_{k=1}^K \mathcal{C}_k \neq \emptyset$. A point $\mathbf{x}_\star \in \mathcal{C}_\star$ can be found by applying the EAPM in (2.29) to Problem (2.24) in a product Hilbert space $(\mathcal{H}^K, \langle\!\langle\cdot,\cdot\rangle\!\rangle)$ with inner product

$$\left(\forall \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K\right)\left(\forall \mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K\right) \quad \langle\!\langle \mathbf{X}, \mathbf{Y}\rangle\!\rangle := \sum_{k=1}^K w_k \langle \mathbf{x}_k, \mathbf{y}_k\rangle$$

and induced norm

$$\left(\forall \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K\right) \quad |\!|\!| \mathbf{X} |\!|\!| = \sqrt{\sum_{k=1}^K w_k \|\mathbf{x}_k\|^2},$$

where $w_1, \ldots, w_k$ are positive weights such that $\sum_{k=1}^K w_k = 1$. To represent the multi-set feasibility problem in (2.1) in terms of a two-set feasibility problem in the product space $\left(\mathcal{H}^K, \langle\!\langle\cdot,\cdot\rangle\!\rangle\right)$, we define a closed subspace $\mathcal{A} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K \mid \mathbf{x}_1 = \cdots = \mathbf{x}_K\}$ and a closed convex set $\mathcal{B} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K \mid (\forall k \in \{1, \ldots, K\})\ \mathbf{x}_k \in \mathcal{C}_k\}$. It is easy to verify that the EAPM applied to the two-set problem in the product Hilbert space $(\mathcal{H}^K, \langle\!\langle\cdot,\cdot\rangle\!\rangle)$ is equivalent to the extrapolated parallel projection method (EPPM)[a] [Pie84]

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \lambda_n L\left(\mathbf{x}^{(n)}\right)\left(\sum_{k=1}^K w_k P_{\mathcal{C}_k}\left(\mathbf{x}^{(n)}\right) - \mathbf{x}^{(n)}\right), \quad \mathbf{x}^{(0)} \in \mathcal{H}$$

where $(\forall n \in \mathbb{N})\ \lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2)$ and

$$(\forall \mathbf{x} \in \mathcal{H}) \quad L(\mathbf{x}) = \begin{cases} \frac{\sum_{k=1}^K w_k \|P_{\mathcal{C}_k}(\mathbf{x}) - \mathbf{x}\|^2}{\left\|\sum_{k=1}^K w_k P_{\mathcal{C}_k}(\mathbf{x}) - \mathbf{x}\right\|^2} & \text{if } \mathbf{x} \notin \mathcal{C}_\star \\ 1 & \text{otherwise.} \end{cases}$$

To see this, note that the projections of a point $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_K) \in \mathcal{H}^K$ onto $\mathcal{A}$ and $\mathcal{B}$ are given by [SY98, Eq. (2.9-11)]

$$P_{\mathcal{A}}(\mathbf{X}) = \left(\sum_{k=1}^K w_k \mathbf{x}_k, \ldots, \sum_{k=1}^K w_k \mathbf{x}_k\right) \tag{2.33}$$

and [SY98, Eq. (2.9-8)]

$$P_{\mathcal{B}}(\mathbf{X}) = \left(P_{\mathcal{C}_1}(\mathbf{x}_1), \ldots, P_{\mathcal{C}_K}(\mathbf{x}_K)\right), \tag{2.34}$$

respectively. Substituting (2.33) and (2.34) in (2.29), (2.30), and (2.31), we can show that the EAPM generates a sequence $\left(\mathbf{X}^{(n)} = (\mathbf{x}^{(n)}, \ldots, \mathbf{x}^{(n)})\right)_{n\in\mathbb{N}}$ in $\mathcal{A}$ by applying the recursion $\mathbf{X}^{(0)} \in \mathcal{A}$, $(\forall n \in \mathbb{N})$

$$\mathbf{X}^{(n+1)} = \mathbf{X}^{(n)} + \lambda_n K\left(\mathbf{X}^{(n)}\right)\left(\left(\sum_{k=1}^{K} w_k P_{\mathcal{C}_k}\left(\mathbf{x}^{(n)}\right), \ldots, \sum_{k=1}^{K} w_k P_{\mathcal{C}_k}\left(\mathbf{x}^{(n)}\right)\right) - \mathbf{X}^{(n)}\right),$$

where $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$ and $(\forall \mathbf{X} = (\mathbf{x}, \ldots, \mathbf{x}) \in \mathcal{A})$

$$\begin{aligned}
K(\mathbf{X}) &= \begin{cases} \frac{\|P_{\mathcal{B}}(\mathbf{X})-\mathbf{X}\|^2}{\|P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{X})-\mathbf{X}\|^2} & \text{if } \mathbf{X} \notin \mathcal{B} \\ 1 & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{\sum_{k=1}^{K} w_k \|P_{\mathcal{C}_k}(\mathbf{x})-\mathbf{x}\|^2}{\sum_{i=1}^{K} w_i \|\sum_{k=1}^{K} w_k P_{\mathcal{C}_k}(\mathbf{x})-\mathbf{x}\|^2} & \text{if } \mathbf{X} \notin \mathcal{B} \\ 1 & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{\sum_{k=1}^{K} w_k \|P_{\mathcal{C}_k}(\mathbf{x})-\mathbf{x}\|^2}{\|\sum_{k=1}^{K} w_k P_{\mathcal{C}_k}(\mathbf{x})-\mathbf{x}\|^2} & \text{if } \mathbf{X} \notin \mathcal{B} \\ 1 & \text{otherwise} \end{cases} \\
&= L(\mathbf{x}).
\end{aligned}$$

Consequently, the sequence $\left(\mathbf{x}^{(n)}\right)_{n\in\mathbb{N}}$ in $\mathcal{H}$ produced by the EPPM converges to a point in $\mathcal{C}_\star$, if and only if the sequence $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ in $\mathcal{H}^K$ produced by the EAPM converges to a point in $\mathcal{A}\cap\mathcal{B}$. Thus Theorem 2.5 also ensures the bounded perturbation resilience of the EPPM.

[a] The original definition of the EPPM in [Pie84], which is also used in [Com97] and [SY98], restricts the relaxation parameters to $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon, 1] \subset (0, 1]$. In this remark, we derive a more general variant of the EPPM, which allows for step sizes that are twice as large. This algorithm is a special case of the EPPM2 in [Com97]. However, the variant considered here is sometimes also referred to as EPPM (see, e.g., [CCC+12]).

3. Multicast Beamforming

In this chapter, we use Theorem 2.1 to devise an algorithm with convergence guarantees to address the nonconvex multi-group multicast beamforming problem with QoS constraints and per-antenna power constraints. We formulate a convex relaxation of the problem as a semidefinite program (SDP) in a real Hilbert space, which allows us to approximate a point in the feasible set by iteratively applying a bounded perturbation resilient fixed point mapping. Inspired by the superiorization methodology, we use this mapping as a basic algorithm, and we add in each iteration a small perturbation with the intent to reduce the objective value and the distance to nonconvex rank constraint sets. We prove that the sequence of perturbations is bounded, so the algorithm is guaranteed to converge to a feasible point of the relaxed SDP. Simulations show that the proposed approach outperforms existing algorithms in terms of both computation time and approximation gap in many cases. The content of this chapter is borrowed from [7] with minimal changes.

3.1. Introduction

Many applications in wireless networks involve multicast communication, which can be defined as the transmission of identical information to multiple receivers. One example is connected driving, where applications such as platooning can benefit from transmitting the same status or control information to a group of vehicles [ZLW+15]. Another example is the transmission of audio signals for live events, where each spectator can select from a variety of audio streams. Both use cases can benefit considerably from physical layer precoders that ensure a given QoS level for the requested stream at each receiver while reusing the same time and frequency resources for all receivers.

Physical layer multicasting schemes have been extensively investigated in the last two decades. The authors of [SDL06] show that the performance of multicast transmission can be greatly improved by exploiting channel state information (CSI) at the transmitter. They consider two beamforming problems for single-group multicast beamforming, the max-min fair (MMF) multicast beamforming problem and the QoS constrained multicast beamforming problem. While the MMF formulation aims at maximizing the lowest signal-to-noise ratio (SNR) among a group of users subject to a unit power constraint on the

beamforming vector, the objective of the QoS constrained formulation is to minimize the transmit power subject to SNR constraints for the individual users. Moreover, the authors of [SDL06] show that the solutions to both problems are equivalent up to a scaling factor.

The more general case with multiple cochannel multicast groups is considered in [KSL08]. Unlike the single-group case, the QoS constrained and MMF versions of the multi-group multicast beamforming problem are different in the sense that a solution to one version cannot generally be obtained by scaling a solution to the other. However, algorithms for the QoS constrained formulation can be straightforwardly extended to approximate the MMF version, by performing a bisection search over the target signal-to-interference-plus-noise ratio (SINR) values. In the presence of per-antenna power constraints, the MMF version and other variations such as the sum-rate maximization [CCO14a] and weighted max-min fair [CCO14b] formulations can be approximated by performing a bisection over instances of a per-antenna power minimization problem [CCO14b]. These formulations are outside the scope of this work. In this chapter, we will therefore restrict our attention to the QoS constained problem.

The QoS-constrained multi-group multicast beamforming problem is a well-studied non-convex quadratically constrained quadratic program (QCQP), for which various algorithmic approximations have been proposed. Existing approaches such as semidefinite relaxation with Gaussian randomization [SDL06, KSL08] and successive convex approximation (SCA) algorithms — also known as convex-concave procedures (CCPs) — involve solving a sequence of convex subproblems. Solutions to these subproblems can be approximated either using off-the-shelf interior-point methods [MHG$^+$14, CCO15] or using first order algorithms such as the alternating direction method of multipliers (ADMM) [HS16, CT17]. While the use of interior-point methods typically results in a high computational complexity, the ADMM can require a large number of iterations to achieve a certain accuracy. Regardless of the algorithm used to approximate each subproblem, the CCP results in nested approximation loops. Terminating the inner iteration after a finite number of steps can hinder the feasibiltiy of estimates, which is required to ensure that the CCP converges. By contrast, if we assume the singular value decomposition (SVD) of a matrix to be computable,[1] the algorithm proposed in [7], which is presented in this chapter, is free of nested optimization loops.

In this chapter, we consider the QoS-constrained multi-group multicast beamforming problem in [KSL08] with optional per-antenna power constraints as introduced in [CT17]. We propose an algorithmic approximation based on superiorization of a bounded perturbation resilient fixed point mapping. To do so, we formulate the problem in a product Hilbert space composed of subspaces of Hermitian matrices. This allows us to approximate

[1]The convergence of algorithms for computing the SVD is well-studied (see, e.g., [VLG83]).

a feasible point of the relaxed problem with the POCS algorithm in (2.3). The bounded perturbation resilience of POCS, which is guaranteed by Theorem 2.1, allows us to add small perturbations in each iteration with the intent to reduce the objective value and the distance to the nonconvex rank-one constraints. In this way, the proposed algorithm follows the set theoretic philosophy of the POCS algorithm in combined with the general framework of superiorization, applied to a particular problem in communications. Simulations show that, compared to existing methods, the proposed approach can provide better approximations at a lower computational cost in many cases.

3.2. Problem Statement

In Section 3.2.1, we define the system model and state the multi-group multicast beamforming problem with QoS- and per-antenna-power-constraints, and we reformulate it in terms of a nonconvex SDP. A well-known approach to approximate solutions to such problems resorts to solving a convex relaxation: First, the original problem is relaxed and solved using, e.g., interior point methods. Subsequently, randomization techniques are applied to obtain candidate solutions to the original problem [KSL08], [LMS$^+$10]. However, in real-time applications, the complexity of interior point solvers becomes prohibitive as it grows very fast with the system size (i.e., the number of users and the number of antennas). Therefore, in Section 3.2.2, we formulate the problem in a real product Hilbert space composed of complex (Hermitian) matrices. This formulation makes the problem accessible by a variety of first order algorithms with low complexity and provable convergence properties.

3.2.1. System Model and Original Problem

Following the system model in [KSL08], we consider the downlink in a network with a transmitter equipped with N antenna elements, each of them represented by an element of the set $\mathcal{N} := \{1, \dots, N\}$. Each user $k \in \mathcal{K} := \{1, \dots, K\}$ is equipped with a single receive antenna. The users are grouped into M disjoint multicast groups $\mathcal{G}_m \subset \mathcal{K}$ indexed by $m \in \mathcal{M} := \{1, \dots, M\}$, such that $\bigcup_{m=1}^{M} \mathcal{G}_m = \mathcal{K}$. Each member of a multicast group $\mathcal{G}_m$ is intended to receive the same information-bearing symbol $x_m \in \mathbb{C}$. The receive signal for the kth user can be written as $y_k = \sum_{m=1}^{M} \mathbf{w}_m^H \mathbf{h}_k x_m + n_k$, where $\mathbf{w}_m \in \mathbb{C}^N$ is the beamforming vector for the mth multicast group, $\mathbf{h}_k \in \mathbb{C}^N$ is the instantaneous channel to user k, and $n_k \in \mathbb{C}$ — drawn independently from the distribution $\mathcal{CN}(0, \sigma_k^2)$ — is the noise sample at the receiver. Consequently, the transmit power for group $\mathcal{G}_m$ is proportional to $\|\mathbf{w}_m\|_2^2$.

In this chapter, we consider the multi-group multicast beamforming problem with QoS-constraints [KSL08], which has the objective to minimize the total transmit power subject to constraints on the QoS expressed in terms of SINR requirements. We use the following problem formulation from [CT17], with an individual power-constraint for each transmit antenna:

$$\underset{\{\mathbf{w}_m \in \mathbb{C}^N\}_{m=1}^M}{\text{minimize}} \quad \sum_{m=1}^{M} \|\mathbf{w}_m\|_2^2 \tag{3.1a}$$

$$\text{s.t.} \quad (\forall m \in \mathcal{M})(\forall k \in \mathcal{G}_m) \quad \frac{|\mathbf{w}_m^H \mathbf{h}_k|^2}{\sum_{l \neq m} |\mathbf{w}_l^H \mathbf{h}_k|^2 + \sigma_k^2} \geq \gamma_k \tag{3.1b}$$

$$(\forall i \in \mathcal{N}) \quad \sum_{m=1}^{M} \mathbf{w}_m^H \mathbf{e}_i \mathbf{e}_i^T \mathbf{w}_m \leq p_i \tag{3.1c}$$

The objective function in (3.1a) corresponds to the total transmit power. The inequalities in (3.1b) constitute the SINR-constraints, where γ_k is the SINR required by user k. The inequalities in (3.1c) correspond to the per-antenna power constraints, where $\mathbf{e}_i \in \mathbb{R}^N$ is the ith Cartesian unit vector.

The problem in (3.1) is a nonconvex QCQP, which is known to be NP-hard [SDL06]. A well-known strategy for approximating solutions to such problems is the semidefinite relaxation technique [KSL08], [LMS+10]. By this technique, we obtain a convex relaxation of the original problem by reformulating it as a nonconvex SDP and by dropping the nonconvex rank constraints. More precisely, using the trace identity $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ for matrices $\mathbf{A}, \mathbf{B}$ of compatible dimensions, we can write $\|\mathbf{w}_m\|_2^2 = \mathbf{w}_m^H \mathbf{w}_m = \text{tr}(\mathbf{w}_m^H \mathbf{w}_m) = \text{tr}(\mathbf{w}_m \mathbf{w}_m^H)$ and $|\mathbf{w}_m^H \mathbf{h}_k|^2 = \mathbf{w}_m^H \mathbf{h}_k (\mathbf{w}_m^H \mathbf{h}_k)^* = \text{tr}(\mathbf{w}_m^H \mathbf{h}_k \mathbf{h}_k^H \mathbf{w}_m) = \text{tr}(\mathbf{w}_m \mathbf{w}_m^H \mathbf{h}_k \mathbf{h}_k^H)$. By defining $(\forall k \in \mathcal{K})$ $\mathbf{Q}_k = \mathbf{h}_k \mathbf{h}_k^H$, and replacing the expression $\mathbf{w}_m \mathbf{w}_m^H$ by a positive semidefinite rank-one matrix $\mathbf{X}_m \in \mathbb{C}^{N \times N}$ for all $m \in \mathcal{M}$, we obtain the nonconvex SDP

$$\underset{\{\mathbf{X}_m \in \mathbb{C}^{N \times N}\}_{m=1}^M}{\text{minimize}} \quad \sum_{m=1}^{M} \text{tr}(\mathbf{X}_m) \tag{3.2a}$$

$$\text{s.t.} \quad (\forall m \in \mathcal{M})(\forall k \in \mathcal{G}_m) \quad \text{tr}(\mathbf{Q}_k \mathbf{X}_m) \geq \gamma_k \sum_{l \neq m} \text{tr}(\mathbf{Q}_k \mathbf{X}_l) + \gamma_k \sigma_k^2 \tag{3.2b}$$

$$(\forall i \in \mathcal{N}) \quad \sum_{m=1}^{M} \text{tr}(\mathbf{e}_i \mathbf{e}_i^T \mathbf{X}_m) \leq p_i \tag{3.2c}$$

$$(\forall m \in \mathcal{M}) \quad \mathbf{X}_m \succcurlyeq \mathbf{0} \tag{3.2d}$$

$$(\forall m \in \mathcal{M}) \quad \text{rank}(\mathbf{X}_m) \leq 1, \tag{3.2e}$$

The formulation in (3.2) is equivalent to (3.1) in the sense that $\{\mathbf{X}_m = \mathbf{w}_m \mathbf{w}_m^H\}_{m=1}^M$ solves

(3.2) if and only if $\{\mathbf{w}_m\}_{m=1}^M$ solves (3.1). A convex relaxation of Problem (3.2) can be obtained by simply dropping the rank constraints in (3.2e). The approach in [SDL06], [KSL08] solves this relaxed problem and, subsequently, generates candidate approximations for Problem (3.2) (and hence (3.1)) using randomization techniques. A solution to the relaxed problem is typically found using general-purpose interior point solvers, which results in high computational cost for large-scale problems. In the multi-group setting [KSL08], each randomization step involves solving an additional power control problem, which further increases the computational burden.

3.2.2. Problem Formulation in a Real Hilbert Space

The objective of this section is to show that Problem (3.2) can be formulated in a real Hilbert space, which enables us to approach the problem by means of efficient projection-based methods. To this end, we consider the *real* vector space $\mathcal{V} := \mathbb{C}^{N\times N}$ of complex $N\times N$-matrices. More precisely, we define vector addition in the usual way, and we restrict scalar multiplication to real scalars $a \in \mathbb{R}$, where each coefficient of a vector $\mathbf{X} \in \mathcal{V}$ is multiplied by a to obtain the vector $a\mathbf{X} \in \mathcal{V}$. In this way, $\mathcal{V}$ is a real vector space, i.e., a vector space over the field $\mathbb{R}$.

If we equip the space $\mathcal{V}$ with a real inner product[2]

$$(\forall \mathbf{X}, \mathbf{Y} \in \mathcal{V}) \quad \langle \mathbf{X}, \mathbf{Y} \rangle := \mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{X}^H\mathbf{Y}\right)\right\}, \tag{3.3}$$

which induces the standard Frobenius norm

$$||\mathbf{X}|| = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle} = \sqrt{\mathrm{tr}\,(\mathbf{X}^H\mathbf{X})},$$

we obtain a *real* Hilbert space $(\mathcal{V}, \langle\cdot,\cdot\rangle)$.

Remark 3.1. The function $\langle\cdot,\cdot\rangle$ defined in (3.3) is a real inner product.
Proof: See Appendix A.1

In the remainder of this chapter, we restrict our attention to the subspace $\mathcal{H} := \{\mathbf{X} \in \mathcal{V} \mid \mathbf{X} = \mathbf{X}^H\}$ of Hermitian matrices. Following the notation in [SY98], we define a product space $\mathcal{H}^M$ as the M-fold Cartesian product

$$\mathcal{H}^M := \underbrace{\mathcal{H} \times \cdots \times \mathcal{H}}_{M \text{ times}}$$

[2] See Remark 3.1.

of $\mathcal{H}$. In this vector space, the sum of two vectors $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_M) \in \mathcal{H}^M$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_M) \in \mathcal{H}^M$ is given by $\mathbf{X}+\mathbf{Y} := (\mathbf{X}_1 + \mathbf{Y}_1, \ldots, \mathbf{X}_M + \mathbf{Y}_M)$ and scalar multiplication is restricted to real scalars $a \in \mathbb{R}$, where $a\,(\mathbf{X}_1, \ldots, \mathbf{X}_M) := (a\mathbf{X}_1, \ldots, a\mathbf{X}_M)$. We equip the space $\mathcal{H}^M$ with the inner product

$$\langle\!\langle \mathbf{X}, \mathbf{Y} \rangle\!\rangle := \sum_{m=1}^{M} \langle \mathbf{X}_m, \mathbf{Y}_m \rangle, \tag{3.4}$$

which induces the norm

$$|||\mathbf{X}|||^2 = \langle\!\langle \mathbf{X}, \mathbf{X} \rangle\!\rangle = \sum_{m=1}^{M} \|\mathbf{X}_m\|^2,$$

where $(\forall m \in \mathcal{M})\ \mathbf{X}_m \in \mathcal{H}$ and $\mathbf{Y}_m \in \mathcal{H}$. Consequently, $\left(\mathcal{H}^M, \langle\!\langle \cdot, \cdot \rangle\!\rangle\right)$ is also a real Hilbert space.

In order to pose Problem (3.2) in this Hilbert space, we express the objective function in (3.2a) and the constraints in (3.2b)–(3.2e) in terms of a convex function and closed sets in $\left(\mathcal{H}^M, \langle\!\langle \cdot, \cdot \rangle\!\rangle\right)$ as shown below:

1) The objective function in (3.2a) can be written as the following inner product:

$$\sum_{m=1}^{M} \operatorname{tr}(\mathbf{X}_m) = \langle\!\langle \mathbf{J}, \mathbf{X} \rangle\!\rangle, \tag{3.5}$$

where $\mathbf{J} = (\mathbf{I}_N, \ldots, \mathbf{I}_N)$. This follows from (3.3), (3.4), and the fact that $(\forall \mathbf{W} \in \mathcal{H})$ $\operatorname{Im}\{\operatorname{tr}(\mathbf{W})\} = 0$.

2) The SINR constraint for user $k \in \mathcal{K}$ in (3.2b) corresponds to the closed half-space

$$\mathcal{Q}_k = \left\{ \mathbf{X} \in \mathcal{H}^M \,\middle|\, \langle\!\langle \mathbf{X}, \mathbf{Z}^k \rangle\!\rangle \geq \sigma_k^2 \right\}, \tag{3.6}$$

where $(\forall k \in \mathcal{K})\ \mathbf{Z}^k \in \mathcal{H}^M$ is given by

$$\mathbf{Z}^k = \Big(\underbrace{-\mathbf{Q}_k, \cdots, -\mathbf{Q}_k}_{1,\cdots,g_k-1}, \underbrace{\gamma_k^{-1}\mathbf{Q}_k}_{g_k}, \underbrace{-\mathbf{Q}_k, \cdots, -\mathbf{Q}_k}_{g_k+1,\cdots,M} \Big).$$

Here, we introduced indices $\{g_k\}_{k\in\mathcal{K}}$ that assign to each receiver $k \in \mathcal{K}$ the multicast group $\mathcal{G}_m$ to which it belongs (i.e., $g_k = m$, if $k \in \mathcal{G}_m$). In order to verify that the set

$\mathcal{Q}_k$ in (3.6) indeed represents the SINR constraint for user k in (3.2b), we rearrange[3]

$$\langle\!\langle \mathbf{X}, \mathbf{Z}^k \rangle\!\rangle = \frac{1}{\gamma_k} \langle \mathbf{X}_{g_k}, \mathbf{Q}_k \rangle - \sum_{\substack{l \in \mathcal{M} \\ l \neq g_k}} \langle \mathbf{X}_l, \mathbf{Q}_k \rangle.$$

Using the definition of the inner product in (3.3), and the fact that $(\forall \mathbf{W} \in \mathcal{H}^M)$ $\mathbf{W}^H = \mathbf{W}$ and $\mathrm{Im}\{\mathrm{tr}(\mathbf{W})\} = 0$, we can rewrite the constraint $\mathcal{Q}_k$ as

$$\mathrm{tr}(\mathbf{X}_{g_k} \mathbf{Q}_k) - \gamma_k \sum_{\substack{l \in \mathcal{M} \\ l \neq g_k}} \mathrm{tr}(\mathbf{X}_l \mathbf{Q}_k) \geq \gamma_k \sigma_k^2,$$

which corresponds to the kth SINR constraint in (3.2b).

3) The per-antenna power constraints in (3.2c) are expressed by the closed convex set

$$\mathcal{P} = \left\{ \mathbf{X} \in \mathcal{H}^M \;\middle|\; (\forall i \in \mathcal{N})\ \langle\!\langle \mathbf{D}^i, \mathbf{X} \rangle\!\rangle \leq p_i \right\},$$

where

$$(\forall i \in \mathcal{N}) \quad \mathbf{D}^i := (\mathbf{e}_i \mathbf{e}_i^T, \ldots, \mathbf{e}_i \mathbf{e}_i^T) \in \mathcal{H}^M. \tag{3.7}$$

This follows immediately from (3.3) and (3.4).

4) The PSD constraints in (3.2d) correspond to the closed convex cone $\mathcal{C}_+$ given by

$$\mathcal{C}_+ = \left\{ (\mathbf{X}_1, \ldots, \mathbf{X}_M) \in \mathcal{H}^M \;\middle|\; (\forall m \in \mathcal{M})\ \mathbf{X}_m \succcurlyeq \mathbf{0} \right\}. \tag{3.8}$$

5) The rank constraints in (3.2e) can be represented by the nonconvex set

$$\mathcal{R} = \left\{ \mathbf{X} \in \mathcal{H}^M \mid (\forall m \in \mathcal{M})\ \mathrm{rank}(\mathbf{X}_m) \leq 1 \right\}. \tag{3.9}$$

Consequently, we can pose Problem (3.2) as

$$\begin{aligned} \underset{\mathbf{X} \in \mathcal{H}^M}{\text{minimize}}\ & \langle\!\langle \mathbf{J}, \mathbf{X} \rangle\!\rangle \\ \text{s.t.}\ & (\forall k \in \mathcal{K})\ \mathbf{X} \in \mathcal{Q}_k \\ & \mathbf{X} \in \mathcal{P}, \quad \mathbf{X} \in \mathcal{C}_+, \quad \mathbf{X} \in \mathcal{R}. \end{aligned} \tag{3.10}$$

The problems in (3.2) and (3.10) are equivalent in the sense that $\{\mathbf{X}_m \in \mathcal{V}\}_{m \in \mathcal{M}}$ solves

[3] In the remainder of this chapter, we use the convention that $\mathbf{X}_m \in \mathcal{H}$ denotes the mth component matrix of an M-tuple $\mathbf{X} \in \mathcal{H}^M$.

Problem (3.2) if and only if $(\mathbf{X}_1, \ldots, \mathbf{X}_M) \in \mathcal{H}^M$ solves Problem (3.10). The advantage of the formulation in (3.10) is that it enables us to (i) streamline notation, (ii) express the updates of the algorithm proposed later in Section 3.3 in terms of well-known projections, and (iii) apply the convergence results for real Hilbert spaces, which were presented in Chapter 2.

It is worth noting that all constraint sets described above are closed, so a projection onto each of the sets exists for any point $\mathbf{X} \in \mathcal{H}^M$. This property is crucial to derive projection-based algorithms, such as the proposed algorithm. In particular, note that we cannot replace the inequality in (3.2e) with an equality, as commonly done in the literature. The reason is that, with an equality, the corresponding set is not closed, as shown in Remarks 3.2 and 3.3, and the practical implication is that the projection may not exist everywhere. Specifically, this happens whenever $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_M)$ satisfies $\mathbf{X}_m = \mathbf{0}$ for some $m \in \mathcal{M}$, which would leave the update rule at such points undefined in projection-based methods. This is illustrated for the case $\mathbf{X} = \mathbf{0} \in \mathcal{H}^M$ in Example 3.1 below.

Remark 3.2. The rank constraint set $\mathcal{R}$ in (3.9) is closed.

Proof. Let $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ be a sequence of points in $\mathcal{R}$ converging to a point $\mathbf{X}^\star = (\mathbf{X}_1^\star, \ldots, \mathbf{X}_M^\star) \in \mathcal{H}^M$ and denote by $(\forall m \in \mathcal{M})(\forall n \in \mathbb{N})\ \mathbf{X}_m^{(n)} = \mathbf{U}_m^{(n)}\mathbf{S}_m^{(n)}(\mathbf{V}_m^{(n)})^H$ the SVD of the mth component matrix of $\mathbf{X}^{(n)}$. It follows from $\mathbf{X}^{(n)} \in \mathcal{R}$ that $(\forall m \in \mathcal{M})\ \mathbf{S}_m^{(n)} = \operatorname{diag}([s_m^{(n)}, 0, \ldots, 0])$. Since a sequence of zeros can only converge to zero, the SVD $\mathbf{X}_m^\star = \mathbf{U}_m^\star\mathbf{S}_m^\star(\mathbf{V}_m^\star)^H$ of the mth component matrix of $\mathbf{X}^\star$ satisfies $\mathbf{S}_m^\star = \operatorname{diag}([s_m^\star, 0, \ldots, 0])$ for some $s_m^\star \in \mathbb{R}_+$. Therefore $(\forall m \in \mathcal{M})\ \operatorname{rank}(\mathbf{X}_m^\star) \leq 1$, so $\mathbf{X}^\star \in \mathcal{R}$. The above shows that $\mathcal{R}$ contains all its limit points, so it is closed. □

Remark 3.3. By contrast,

$$\mathcal{R}' = \left\{\mathbf{X} \in \mathcal{H}^M \mid (\forall m \in \mathcal{M})\ \operatorname{rank}(\mathbf{X}_m) = 1\right\}$$

is not a closed set, since for all $\mathbf{X} \in \mathcal{R}'$ and $\alpha \in (0, 1)$, the sequence $(\alpha^n\mathbf{X})_{n\in\mathbb{N}}$ in $\mathcal{R}'$ converges to $\mathbf{0} \notin \mathcal{R}'$.

Example 3.1. The set-valued projection of $\mathbf{0} \in \mathcal{H}^M$ onto the set $\mathcal{R}'$ in Remark 3.3 is empty.

Proof. Suppose that $\Pi_{\mathcal{R}'}(\mathbf{0}) \neq \emptyset$ and let $\mathbf{Z} \in \Pi_{\mathcal{R}'}(\mathbf{0})$, i.e., $\mathbf{Z}$ is any of the closest points of the set $\mathcal{R}'$ to the zero vector $\mathbf{0}$. Since $\Pi_{\mathcal{R}'}(\mathbf{0}) \subset \mathcal{R}'$, $(\forall m \in \mathcal{M})\ \operatorname{rank}(\mathbf{Z}_m) = 1$, i.e., $(\forall m \in \mathcal{M})\ \sigma_1(\mathbf{Z}_m) > 0$. Therefore, for any $\alpha \in (0, 1)$, $\alpha\mathbf{Z} \in \mathcal{R}'$ and $d(\mathbf{0}, \alpha\mathbf{Z}) < d(\mathbf{0}, \mathbf{Z})$, i.e., $\alpha\mathbf{Z} \in \mathcal{R}'$ is closer to the zero vector than $\mathbf{Z} \in \mathcal{R}'$, thus contradicting our assumption that $\mathbf{Z}$ is one of the closest points in $\mathcal{R}'$ to the vector $\mathbf{0}$. □

3.3. Algorithmic Solutions

The main difficulty in solving (3.10) is the presence of the nonconvex rank constraint. A well-known technique for approximating rank-constrained SDPs using convex optimization methods is the semidefinite relaxation approach [SDL06], [KSL08], [LMS+10]. This approach first solves (3.10) without the rank constraint, and then it applies heuristics to obtain rank-one approximations based on the solution to this relaxed problem. Similarly, we can obtain a convex relaxation

$$\begin{aligned} \underset{\mathbf{X}\in\mathcal{H}^M}{\text{minimize}} \; & \langle\langle \mathbf{J}, \mathbf{X} \rangle\rangle \\ \text{s.t.} \quad & (\forall k \in \mathcal{K}) \; \mathbf{X} \in \mathcal{Q}_k \\ & \mathbf{X} \in \mathcal{P}, \quad \mathbf{X} \in \mathcal{C}_+, \end{aligned} \tag{3.11}$$

of Problem (3.10) by dropping the nonconvex constraint set $\mathcal{R}$. In principle, we could solve this relaxed problem using first order techniques for constrained convex minimization. For instance, we could apply a projected (sub-)gradient method (see, e.g., [Nes18, Section 3.2.3]), which interleaves (sub-)gradient steps for the objective function with projections onto the feasible set of Problem (3.11). However, computing the projection onto the intersection of all constraint sets in Problem (3.11) typically requires an inner optimization loop because no simple expression for this projection is known. As it was shown in [CDH+14], superiorization can significantly reduce the computation time compared to the projected gradient method in some applications if the projection onto the feasible set is difficult to compute.

The superiorization methodology typically relies on an iterative process that solves a convex feasibility problem (i.e., that produces a sequence of points converging to a point within the intersection of all constraint sets) by repeatedly applying a computationally simple mapping. This iterative algorithm is called the *basic algorithm*. Based on this basic algorithm, the superiorization methodology automatically produces a *superiorized version of the basic algorithm*, by adding bounded perturbations (see Definition 1.3) to the iterates of the basic algorithm in every iteration. The perturbations are typically generated based on subgradient steps for a given objective function, in a way that ensures the sequence of perturbations to be bounded. By showing that the basic algorithm is bounded perturbation resilient (i.e., that the resulting sequence is guaranteed to converge to a feasible point, even when bounded perturbations are added in each iteration), one can ensure that the sequence produced by the superiorized version of the basic algorithm also converges to a feasible point. In contrast to constrained minimization, superiorization does not guarantee that the objective value of the resulting approximation is minimal. However,

the limit point of the superiorized algorithm typically has a lower objective value than the limit point of the unperturbed basic algorithm [Cen15].

To apply the superiorization methodology to Problem (3.10), we proceed as follows. In Section 3.3.1, we propose a bounded perturbation resilient basic algorithm by defining a mapping $T_\star : \mathcal{H}^M \to \mathcal{H}^M$. Given any point $\mathbf{X}^{(0)} \in \mathcal{H}^M$, this mapping generates a sequence of points converging to a feasible point of Problem (3.11) by

$$(\forall n \in \mathbb{N}) \quad \mathbf{X}^{(n+1)} = T_\star\left(\mathbf{X}^{(n)}\right).$$

In Section 3.3.2, we define a sequence $\left(\beta^{(n)}\mathbf{Y}^{(n)}\right)_{n\in\mathbb{N}}$ of bounded perturbations, with the intent to reduce slightly (i) the objective value of Problem (3.10) and (ii) the distance to the nonconvex rank constraint $\mathcal{R}$ in every iteration. As we show in Proposition 3.2 below, the proposed perturbations can achieve both goals simultaneously. The sequence of perturbations yields a superiorized version of the basic algorithm in (3.3) given by

$$(\forall n \in \mathbb{N}) \quad \mathbf{X}^{(n+1)} = T_\star\left(\mathbf{X}^{(n)} + \beta^{(n)}\mathbf{Y}^{(n)}\right), \quad \mathbf{X}^{(0)} \in \mathcal{H}^M. \tag{3.12}$$

In Section 3.3.3, we prove that the algorithm in (3.12) converges to a feasible point of Problem (3.11) by showing that $\left(\beta^{(n)}\mathbf{Y}^{(n)}\right)_{n\in\mathbb{N}}$ is a sequence of bounded perturbations. The relation between the proposed method and the superiorization methodology is discussed in detail in Section 3.3.4. Finally, the proposed algorithm is summarized in Section 3.3.5.

3.3.1. Feasibility-Seeking Basic Algorithm

A feasible point for the relaxed SDP in (3.11) can be found by solving the convex feasibility problem

$$\text{find } \mathbf{X} \in \mathcal{H}^M \quad \text{s.t.} \quad \mathbf{X} \in \mathcal{C}_\star := \bigcap_{k=1}^{K} \mathcal{Q}_k \cap \mathcal{P} \cap \mathcal{C}_+. \tag{3.13}$$

A well-known technique for solving problems of this kind is the POCS algorithm (see Section 2.1), which generates a sequence of estimates by $\mathbf{x}_0 \in \mathcal{H}^M$

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = T_\star(\mathbf{x}_n).$$

Here, the mapping $T_\star : \mathcal{H}^M \to \mathcal{H}^M$ is given by

$$T_\star := T_{\mathcal{C}_+}^{\mu_{K+2}} T_{\mathcal{P}}^{\mu_{K+1}} T_{\mathcal{Q}_K}^{\mu_K} \dots T_{\mathcal{Q}_1}^{\mu_1}, \tag{3.14}$$

where for a nonempty closed convex set $\mathcal{C} \in \mathcal{H}^M$,

$$T_{\mathcal{C}}^{\mu} = I + \mu(P_{\mathcal{C}} - I)$$

denotes the relaxed projector onto $\mathcal{C}$ with relaxation parameter $\mu \in (0, 2)$. The formal expressions for the projections of $\mathbf{X} \in \mathcal{H}^M$ onto each of the sets in (3.13) are given below.

1) The SINR constraint sets $\mathcal{Q}_k \in \mathcal{H}^M$ are half-spaces, the projections onto which are given by [BC11, Example 29.20] $(\forall k \in \mathcal{K})(\forall \mathbf{X} \in \mathcal{H}^M)$

$$P_{\mathcal{Q}_k}(\mathbf{X}) = \begin{cases} \mathbf{X}, & \text{if } \mathbf{X} \in \mathcal{Q}_k \\ \mathbf{X} + \frac{\sigma_k^2 - \langle\langle \mathbf{X}, \mathbf{Z}^k \rangle\rangle}{\left\| \left\| \mathbf{Z}^k \right\| \right\|^2} \mathbf{Z}^k, & \text{otherwise.} \end{cases}$$

2) The per-antenna power constraint set $\mathcal{P}$ is an intersection of the N half-spaces defined by the normal vectors $\mathbf{D}^i$ in (3.7) for $i \in \mathcal{N}$. Since these vectors are mutually orthogonal, i.e., $(\forall i \in \mathcal{N})(\forall j \in \mathcal{N} \setminus \{i\})\ \langle\langle \mathbf{D}^i, \mathbf{D}^j \rangle\rangle = 0$, the projection onto $\mathcal{P}$ can be written in closed form as

$$P_{\mathcal{P}}(\mathbf{X}) = \mathbf{X} + \sum_{i: p_i < \langle\langle \mathbf{X}, \mathbf{D}^i \rangle\rangle} \frac{p_i - \langle\langle \mathbf{X}, \mathbf{D}^i \rangle\rangle}{\left\| \left\| \mathbf{D}^i \right\| \right\|^2} \mathbf{D}^i,$$

according to Proposition 1.1. Alternatively, we could prove this result by applying [SY98, Thm 4.3-1] and Halperin's Theorem (see [Hal62], [Gin18, Theorem 4.2]).

3) The set $\mathcal{C}_+$ is the intersection of PSD cones in orthogonal subspaces of $\mathcal{H}^M$. The projection of $\mathbf{X} \in \mathcal{H}^M$ onto $\mathcal{C}_+$ is therefore given component-wise by

$$P_{\mathcal{C}_+}(\mathbf{X}) = \left(P_{\mathcal{H}_+}(\mathbf{X}_1), \ldots, P_{\mathcal{H}_+}(\mathbf{X}_M) \right),$$

where, $\mathcal{H}_+ = \{\mathbf{X} \in \mathcal{H} \mid \mathbf{X} \succcurlyeq \mathbf{0}\}$ is the cone of PSD matrices in $\mathcal{H}$. We use the eigendecomposition $\mathbf{X}_m = \mathbf{V}_m \mathbf{\Lambda}_m \mathbf{V}_m^H$ with (real) eigenvalues

$$\mathbf{\Lambda}_m = \operatorname{diag}\left(\lambda_1(\mathbf{X}_m), \ldots, \lambda_N(\mathbf{X}_m) \right) \tag{3.15}$$

to define the projection of $\mathbf{X}_m \in \mathcal{H}$ onto $\mathcal{H}_+$ as[4]

$$P_{\mathcal{H}_+}(\mathbf{X}_m) = \mathbf{V}_m \mathbf{\Lambda}_m^+ \mathbf{V}_m^H, \tag{3.16}$$

where $\mathbf{\Lambda}_m^+ := \operatorname{diag}\left((\lambda_1(\mathbf{X}_m))_+, \ldots, (\lambda_N(\mathbf{X}_m))_+ \right)$.

[4] For the case of real symmetric matrices, see, e.g., [GNR20, Lemma 2.1]. The result in [GNR20] is based on [HJ13, Corollary 7.4.9.3], which assumes complex Hermitian matrices. The generalization of [GNR20, Lemma 2.1] to complex Hermitian matrices is straightforward.

According to Theorem 2.1, the sequence $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ of vectors $\mathbf{X}^{(n)} \in \mathcal{H}^M$ produced by the update rule in (3.14) is guaranteed to converge to a solution of the feasibility problem in (3.13) for any $\mathbf{X}^{(0)} \in \mathcal{H}^M$, if a solution exists (i.e., if $\mathcal{C}_\star \neq \emptyset$). Note that this is the case if the relaxed SDP in (3.11) is feasible. Moreover, Theorem 2.1 also guarantees the convergence of perturbed variants of POCS.

3.3.2. Proposed Perturbations

In the following, we devise perturbations that steer the iterates of the fixed point algorithm in (3.12) towards a solution to the nonconvex problem in (3.2) and (3.10). To do so, we introduce a mapping that reduces the objective value and a mapping that reduces the distance to rank constraint sets. Then we define the proposed perturbations based on the composition of these two mappings As proven in Proposition 3.2 below, the resulting perturbations can achieve both goals simultaneously.

Power Reduction by Bounded Perturbations

In the literature on superiorization, the perturbations are typically defined based on subgradient steps of the objective function (see, e.g., [Cen15]). For the linear objective function in (3.11), this would result in perturbations of the form $-\alpha\,(\mathbf{I}_N, \ldots, \mathbf{I}_N)$ for some $\alpha > 0$. These perturbations are problematic for the problem considered here because we are interested in solutions comprised of positive semidefinite rank-one matrices, and adding these perturbations to an iterate $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_M)$ may result in indefinite full-rank component matrices $\mathbf{X}_m - \alpha\mathbf{I}_N$. To avoid this problem, we introduce the function $f_1 : \mathcal{H}^M \to \mathbb{R}_+$ given by

$$f_1(\mathbf{X}) := \sum_{m=1}^{M} \|\mathbf{X}_m\|_*, \tag{3.17}$$

where $\|\cdot\|_*$ is the nuclear norm. Since $\mathcal{C}_\star \subset \mathcal{C}_+$ by (3.13), we have $(\forall \mathbf{X} \in \mathcal{C}_\star)(\forall m \in \mathcal{M})(\forall i \in \mathcal{N})\ \sigma_i(\mathbf{X}_m) = \lambda_i(\mathbf{X}_m)$, where $\lambda_i(\mathbf{X}_m)$ and $\sigma_i(\mathbf{X}_m)$ denote the ith eigenvalue and singular value of the mth component matrix of $\mathbf{X}$, respectively. Hence we can write

$$\begin{aligned} f_1(\mathbf{X}) &= \sum_{m=1}^{M}\sum_{i=1}^{N} \sigma_i(\mathbf{X}_m) \\ &= \sum_{m=1}^{M}\sum_{i=1}^{N} \lambda_i(\mathbf{X}_m) = \sum_{m=1}^{M} \operatorname{tr}(\mathbf{X}_m). \end{aligned} \tag{3.18}$$

Therefore, by (3.5), minimizing f_1 over $\mathcal{C}_\star$ is equivalent to minimizing the linear objective function in (3.10) (or (3.11)) over $\mathcal{C}_\star$, in the sense that the solution sets to both formulations

are the same. As we will show below, this surrogate objective function gives rise to power-reducing perturbations, which are guaranteed not to increase the rank of their arguments' component matrices (see Remark 3.4).

The power-reducing perturbations are designed according to two criteria. Firstly, they should decrease the value of the surrogate function f_1. Secondly, they should not be too large in order to avoid slowing down convergence of the basic algorithm. For a given point $\mathbf{X} \in \mathcal{H}^M$ we derive a perturbation $\mathbf{Y}_\tau^\star$ satisfying these two criteria by solving the problem

$$\mathbf{Y}_\tau^\star := \mathbf{Y}_\tau^\star(\mathbf{X}) \in \underset{\mathbf{Y}\in\mathcal{H}^M}{\arg\min} \left(\tau f_1(\mathbf{X}+\mathbf{Y}) + \frac{1}{2}|||\mathbf{Y}|||^2 \right). \tag{3.19}$$

Here, $|||\mathbf{Y}|||^2$ acts as a regularization on the perturbations' magnitude, and the parameter $\tau \geq 0$ balances the two design criteria. The next proposition shows that $\mathbf{Y}_\tau^\star$ can be easily computed.

Proposition 3.1. *The unique solution to* (3.19) *is given by*

$$(\forall m \in \mathcal{M}) \quad \mathbf{Y}_\tau^\star|_m = \chi_\tau(\mathbf{X}_m) - \mathbf{X}_m, \tag{3.20}$$

where $\chi_\tau : \mathcal{H} \to \mathcal{H}$ *is the singular value shrinkage operator [CCS10]*

$$\begin{aligned} \chi_\tau(\mathbf{X}_m) &:= \mathbf{U}_m \chi_\tau(\mathbf{\Sigma}_m) \mathbf{V}_m^H, \\ \chi_\tau(\mathbf{\Sigma}_m) &= \operatorname{diag}\left(\left\{ (\sigma_i(\mathbf{X}_m) - \tau)_+ \right\}_{i\in\mathcal{N}} \right), \end{aligned} \tag{3.21}$$

and $(\forall m \in \mathcal{M})$ $\mathbf{X}_m = \mathbf{U}_m \mathbf{\Sigma}_m \mathbf{V}_m$ *is the singular value decomposition of* $\mathbf{X}_m$ *such that* $\mathbf{\Sigma}_m = \operatorname{diag}\left(\{\sigma_i(\mathbf{X}_m)\}_{i\in\mathcal{N}}\right)$.

Proof: *See Appendix A.2.*

By defining

$$(\forall \mathbf{X} \in \mathcal{H}^M) \quad \sigma_{\max}(\mathbf{X}) := \max_{\substack{m\in\mathcal{M} \\ i\in\mathcal{N}}} \sigma_i(\mathbf{X}_m)$$

we can express the power-reducing perturbation for a point $\mathbf{X} \in \mathcal{H}^M$ as $\mathbf{Y} = T_{\mathrm{P}}^\alpha(\mathbf{X}) - \mathbf{X}$, where the mapping $T_{\mathrm{P}}^\alpha := \operatorname{prox}_{\alpha\sigma_{\max}(\mathbf{X})f_1}$ is given component-wise by $(\forall m \in \mathcal{M})$

$$T_{\mathrm{P}}^\alpha(\mathbf{X})|_m = \chi_\tau(\mathbf{X}_m) \quad \text{with} \quad \tau = \alpha\sigma_{\max}(\mathbf{X}). \tag{3.22}$$

Note that $T_{\mathrm{P}}^0(\mathbf{X}) = \mathbf{X}$, and $(\forall \alpha \geq 1)$ $T_{\mathrm{P}}^\alpha(\mathbf{X}) = \mathbf{0}$. Therefore, the magnitude of the power-reducing perturbations can be controlled by choosing the parameter $\alpha \in [0, 1]$.

Moreover, in contrast to performing subgradient steps for the original cost function in (3.10), applying the perturbations in (3.22) cannot increase the rank:

Remark 3.4. For all $\alpha \geq 0$, T_{P}^{α} maps any point $\mathbf{X} = (\mathbf{X}_m)_{m\in\mathcal{M}} \in \mathcal{C}_+$ to a point $\mathbf{Z} = (\mathbf{Z}_m)_{m\in\mathcal{M}} = T_{\mathrm{P}}^{\alpha}(\mathbf{X}) \in \mathcal{C}_+$ satisfying $(\forall m \in \mathcal{M})$ $\operatorname{rank}(\mathbf{Z}_m) \leq \operatorname{rank}(\mathbf{X}_m)$. This follows immediately from (3.21).

Incorporating the Rank Constraints by Bounded Perturbations

Next, we define perturbations that steer the iterate towards the rank constraint set $\mathcal{R}$ in (3.9). While objective functions used for superiorization are usually convex, the function $f_2 : \mathcal{H}^M \to \mathbb{R}_+$

$$f_2(\mathbf{X}) := d(\mathbf{X}, \mathcal{R}), \tag{3.23}$$

i.e., the distance to the set $\mathcal{R}$, constitutes a nonconvex superiorization objective, so our approach does not follow exactly the superiorization methodology in [Cen15] (but we can still prove convergence).

As the perturbations may steer the iterates away from the feasible set, their magnitude should not be unnecessarily large. Therefore, we choose the rank-reducing perturbations as $P_{\mathcal{R}}(\mathbf{X}) - \mathbf{X}$, where $P_{\mathcal{R}}(\mathbf{X}) \in \Pi_{\mathcal{R}}(\mathbf{X})$ denotes a (generalized) projection of a given point $\mathbf{X} \in \mathcal{H}^M$ onto the closed nonconvex set $\mathcal{R}$. Since $\mathcal{R}$ is a closed set, the set-valued projection $\Pi_{\mathcal{R}}(\mathbf{X})$ is nonempty for all $\mathbf{X} \in \mathcal{H}^M$. A projection onto $\mathcal{R}$ can be computed by truncating all but the largest singular value of each component matrix to zero. We formally state this fact below.

Fact 3.1. *Let $\mathbf{X}_m = \mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^H \in \mathcal{H}$ be the SVD of the mth component matrix of $\mathbf{X}$ with $\mathbf{\Sigma}_m = \operatorname{diag}(\sigma_1(\mathbf{X}_m), \ldots, \sigma_N(\mathbf{X}_m))$. Then, $(\forall \mathbf{X} \in \mathcal{H}^M)$ the mth component matrix of a point $P_{\mathcal{R}}(\mathbf{X}) \in \Pi_{\mathcal{R}}(\mathbf{X})$ is given by [Luk13, Lemma 3.2]*

$$P_{\mathcal{R}}(\mathbf{X})|_m = \mathbf{U}_m \operatorname{diag}\left(\sigma_1(\mathbf{X}_m), 0, \ldots, 0\right) \mathbf{V}_m^H. \tag{3.24}$$

Combining Power- and Rank Perturbations

Since both T_{P}^{α} in (3.22) and $P_{\mathcal{R}}$ in (3.24) operate on the singular values of the component matrices, their composition is given by $(\forall m \in \mathcal{M})$

$$P_{\mathcal{R}}T_{\mathrm{P}}^{\alpha}(\mathbf{X})|_m = \left(\sigma_1(\mathbf{X}_m) - \alpha\sigma_{\max}(\mathbf{X})\right)_+ \mathbf{u}_{m1}\mathbf{v}_{m1}^H \in \mathcal{H},$$

where, $(\forall m \in \mathcal{M})$ $\mathbf{U}_m = [\mathbf{u}_{m1}, \ldots, \mathbf{u}_{mN}]$ and $\mathbf{V}_m = [\mathbf{v}_{m1}, \ldots, \mathbf{v}_{mN}]$. Moreover, it is easy to verify that $(\forall \mathbf{X} \in \mathcal{H}^M)(\forall \alpha \geq 0)$, $T_{\mathrm{P}}^{\alpha} P_{\mathcal{R}}(\mathbf{X}) = P_{\mathcal{R}} T_{\mathrm{P}}^{\alpha}(\mathbf{X})$. We will now use the composition of T_{P}^{α} and $P_{\mathcal{R}}$ to define a mapping $\Upsilon_\alpha : \mathcal{H}^M \to \mathcal{H}^M$ by $\Upsilon_\alpha := P_{\mathcal{R}} T_{\mathrm{P}}^{\alpha} - I$, i.e., $(\forall \mathbf{X} = (\mathbf{X}_m)_{m \in \mathcal{M}} \in \mathcal{H}^M)(\forall m \in \mathcal{M})$

$$\Upsilon_\alpha(\mathbf{X})|_m = (\sigma_1(\mathbf{X}_m) - \alpha \sigma_{\max}(\mathbf{X}))_+ \mathbf{u}_{m1} \mathbf{v}_{m1}^H - \mathbf{X}_m. \tag{3.25}$$

Finally, we define the sequence $\left(\beta^{(n)} \mathbf{Y}^{(n)}\right)_{n \in \mathbb{N}}$ of perturbations in (3.12) by

$$(\forall n \in \mathbb{N}) \quad \mathbf{Y}^{(n)} := \Upsilon_{\alpha^{(n)}} \left(\mathbf{X}^{(n)}\right), \tag{3.26}$$

where $\left(\alpha^{(n)}\right)_{n \in \mathbb{N}}$ is a sequence in $[0, 1]$ and $\left(\beta^{(n)}\right)_{n \in \mathbb{N}}$ is a summable sequence in $[0, 1]$. The following proposition shows that the perturbations in (3.26) can simultaneously reduce the objective value and the distance to the rank constraint set.

Proposition 3.2. *Let $\alpha \in \mathbb{R}_+$ and $\lambda \in [0, 1]$. Then each of the following holds for $\Upsilon_\alpha : \mathcal{H}^M \to \mathcal{H}^M$ in (3.25).*

(a) The perturbations cannot increase the distance to the set $\mathcal{C}_+$, i.e., $(\forall \mathbf{X} \in \mathcal{H}^M)$ $d(\mathbf{X} + \lambda \Upsilon_\alpha(\mathbf{X}), \mathcal{C}_+) \leq d(\mathbf{X}, \mathcal{C}_+)$. In particular, $(\forall \mathbf{X} \in \mathcal{H}^M)$ $\mathbf{X} \in \mathcal{C}_+ \implies \mathbf{X} + \lambda \Upsilon_\alpha(\mathbf{X}) \in \mathcal{C}_+$.

(b) If $\alpha > 0$, the perturbations decrease the value of the function f_1 in (3.18): $\left(\forall \mathbf{X} \in \mathcal{H}^M\right)$ $f_1(\mathbf{X} + \lambda \Upsilon_\alpha(\mathbf{X})) < f_1(\mathbf{X})$ whenever $f_1(\mathbf{X}) > 0$.

(c) If $\alpha > 0$ and $\mathbf{X} \in \mathcal{C}_+$, then the perturbations decrease the objective value of Problem (3.10), i.e., $\langle\langle \mathbf{J}, \mathbf{X} + \lambda \Upsilon_\alpha(\mathbf{X}) \rangle\rangle < \langle\langle \mathbf{J}, \mathbf{X} \rangle\rangle$ whenever $\langle\langle \mathbf{J}, \mathbf{X} \rangle\rangle > 0$.

(d) If $\lambda > 0$, the perturbations decrease the distance to the rank constraint set $\mathcal{R}$. More precisely, $\left(\forall \mathbf{X} \in \mathcal{H}^M\right)$ $f_2(\mathbf{X} + \lambda \Upsilon_\alpha(\mathbf{X})) < f_2(\mathbf{X})$ whenever $f_2(\mathbf{X}) > 0$.

Proof.

(a) Denote by $(\forall m \in \mathcal{M})$ $\mathbf{X}_m = \mathbf{U}_m \mathbf{\Sigma}_m \mathbf{V}_m^H$ the SVD with singular values $\mathbf{\Sigma}_m = \mathrm{diag}(\sigma_1(\mathbf{X}_m), \ldots, \sigma_N(\mathbf{X}_m))$, which is used to define (3.25). As we show in Lemma A.2 in the Appendix, there exists an eigendecomposition $\mathbf{X}_m = \mathbf{V}_m \mathbf{V}_m \mathbf{V}_m^H$, were $\mathbf{\Lambda}_m = \mathrm{diag}(\lambda_1(\mathbf{X}_m), \ldots, \lambda_N(\mathbf{X}_m))$. Now, define $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_M) \in \mathcal{H}^M$ as

$$(\forall m \in \mathcal{M}) \quad \mathbf{Z}_m := (\sigma_1(\mathbf{X}_m) - \alpha \sigma_{\max}(\mathbf{X}))_+ \mathbf{u}_{m1} \mathbf{v}_{m1}^H.$$

Again, by Lemma A.2, there exists an eigendecomposition $\mathbf{Z}_m = \lambda_1(\mathbf{Z}_m)\mathbf{v}_{m1}\mathbf{v}_{m1}^H$. *Moreover, by Lemma A.3, we can write*

$$\lambda_1(\mathbf{Z}_m) = \mathbf{v}_{m1}^H\mathbf{u}_{m1}\sigma_1(\mathbf{Z}_m) = \mathbf{v}_{m1}^H\mathbf{u}_{m1}\left(\sigma_1(\mathbf{X}_m) - \alpha\sigma_{\max}(\mathbf{X})\right)_+ .$$

As $\mathbf{X}_m$ *and* $\mathbf{Z}_m$ *have the same first left and right singular vectors (*$\mathbf{u}_{m1}$ *and* $\mathbf{v}_{m1}$*), it holds that* $\lambda_1(\mathbf{X}_m) \geq 0 \implies \lambda_1(\mathbf{Z}_m) \geq 0$. *Furthermore,* $|\lambda_1(\mathbf{Z}_m)| = \sigma_1(\mathbf{Z}_m) \leq \sigma_1(\mathbf{X}_m) = |\lambda_1(\mathbf{X}_m)|$, *so it is immediate from the definition of* $d(\cdot, \mathcal{C}_+)$ *in Lemma A.1 that* $d(\mathbf{Z}, \mathcal{C}_+) \leq d(\mathbf{X}, \mathcal{C}_+)$. *Finally, because the distance to a convex set is convex, we have that* $d(\mathbf{X} + \lambda\Upsilon_\alpha(\mathbf{X}), \mathcal{C}_+) = d((1-\lambda)\mathbf{X} + \lambda\mathbf{Z}, \mathcal{C}_+) \leq (1-\lambda)d(\mathbf{X}, \mathcal{C}_+) + \lambda d(\mathbf{Z}, \mathcal{C}_+) \leq d(\mathbf{X}, \mathcal{C}_+)$.

(b) It follows from (3.21) *and* (3.22) *that* $(\forall \mathbf{X} \in \mathcal{H}^M)(\forall \alpha > 0)\ f_1(\mathbf{X}) > 0 \implies f_1(T_{\mathrm{P}}^\alpha(\mathbf{X})) < f_1(\mathbf{X})$. *Moreover, by* (3.24) *we have that* $(\forall \lambda \in [0,1])\ f_1((1-\lambda)\mathbf{X} + \lambda P_{\mathcal{R}}(\mathbf{X})) \leq f_1(\mathbf{X})$. *This implies* $f_1(\mathbf{X}+\lambda\Upsilon_\alpha(\mathbf{X})) = f_1\left((1-\lambda)\mathbf{X} + \lambda P_{\mathcal{R}}T_{\mathrm{P}}^\alpha(\mathbf{X})\right) \leq f_1\left(T_{\mathrm{P}}^\alpha(\mathbf{X})\right) < f_1(\mathbf{X})$ *whenever* $f_1(\mathbf{X}) > 0$.

(c) This result follows from (a) and (b), since $(\forall \mathbf{X} \in \mathcal{C}_+)\ \langle\!\langle \mathbf{J}, \mathbf{X} \rangle\!\rangle = f_1(\mathbf{X})$ *according to* (3.18).

(d) Since $\mathcal{R}$ *is closed, we can write*

$$f_2(\mathbf{X}) = d(\mathbf{X}, \mathcal{R}) = |||\mathbf{X} - P_{\mathcal{R}}(\mathbf{X})||| = \sqrt{\sum_{m\in\mathcal{M}}\sum_{i=2}^{N}\sigma_i^2(\mathbf{X}_m)}.$$

Therefore, it follows from (3.21) *that* $(\forall \mathbf{X} \in \mathcal{H}^M)(\forall \alpha \in \mathbb{R}_+)\ f_2(T_{\mathrm{P}}^\alpha(\mathbf{X})) \leq f_2(\mathbf{X})$. *Moreover, by* (3.24), $(\forall \lambda \in (0,1])\ f_2(\mathbf{X}) > 0$ *implies that* $f_2((1-\lambda)\mathbf{X}+\lambda P_{\mathcal{R}}(\mathbf{X})) < f_2(\mathbf{X})$. *This in turn implies* $f_2(\mathbf{X} + \lambda\Upsilon_\alpha(\mathbf{X})) = f_2\left((1-\lambda)\mathbf{X} + \lambda P_{\mathcal{R}}T_{\mathrm{P}}^\alpha(\mathbf{X})\right) < f_2\left(T_{\mathrm{P}}^\alpha(\mathbf{X})\right) \leq f_2(\mathbf{X})$ *whenever* $f_2(\mathbf{X}) > 0$. □

With the perturbations defined in (3.26), the iteration in (3.12) yields the update rule

$$(\forall n \in \mathbb{N}) \quad \mathbf{X}^{(n+1)} = T_\star\left(\mathbf{X}^{(n)} + \beta^{(n)}\Upsilon_{\alpha^{(n)}}\left(\mathbf{X}^{(n)}\right)\right) \tag{3.27}$$

of the proposed algorithm, where $\mathbf{X}^{(0)} \in \mathcal{H}^M$ is arbitrary, $\left(\alpha^{(n)}\right)_{n\in\mathbb{N}}$ is a sequence in $[0,1]$, and $\left(\beta^{(n)}\right)_{n\in\mathbb{N}}$ is a summable sequence in $[0,1]$.

3.3.3. Convergence of the Proposed Algorithm

We will now examine the convergence of the proposed algorithm in (3.27). For this purpose, let $\left(\beta^{(n)}\right)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, let $\left(\alpha^{(n)}\right)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+$, and denote by $\left(\mathbf{Y}^{(n)}\right)_{n\in\mathbb{N}}$ the sequence of perturbations according to (3.26). Then the sequence $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ produced by the algorithm in (3.27) converges to a feasible point of Problem (3.11) for all $\mathbf{X}^{(0)} \in \mathcal{H}^M$.

Theorem 2.1 proves the weak convergence of the sequence $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ produced by the superiorized POCS algorithm in (3.27), given that the set $\mathcal{C}_\star$ is nonempty and that $\left(\beta^{(n)}\mathbf{Y}^{(n)}\right)_{n\in\mathbb{N}}$ is a sequence of bounded perturbations. Moreover, since $\mathcal{H}^M$ is finite dimensional, the convergence of $\left(\mathbf{X}^{(n)}\right)_{n\in\mathbb{N}}$ to a point in $\mathcal{C}_\star$ is strong. Therefore, it remains to show that the sequence $\left(\mathbf{Y}^{(n)}\right)_{n\in\mathbb{N}}$ is bounded for all sequences $\left(\alpha^{(n)}\right)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ and $\left(\beta^{(n)}\right)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, regardless of the choice of $\mathbf{X}^{(0)} \in \mathcal{H}^M$. To this end, we note that $(\forall n \in \mathbb{N})\ \left|\!\left|\!\left|\mathbf{Y}^{(n)}\right|\!\right|\!\right| \leq \left|\!\left|\!\left|\mathbf{X}^{(n)}\right|\!\right|\!\right|$ for any sequence $\left(\alpha^{(n)}\right)_{n\in\mathbb{N}}$ of nonnegative numbers:

Lemma 3.1. *The mapping* Υ_α *in* (3.25) *satisfies*

$$\left(\forall \mathbf{X} \in \mathcal{H}^M\right)(\forall \alpha \in \mathbb{R}_+) \quad |\!|\!|\Upsilon_\alpha(\mathbf{X})|\!|\!|^2 \leq |\!|\!|\mathbf{X}|\!|\!|^2.$$

Proof. *Let* $(\forall m \in \mathcal{M})\ \mathbf{X}_m = \mathbf{U}_m \mathrm{diag}(\{\sigma_i(\mathbf{X}_m)\}_{i\in\mathcal{N}})\mathbf{V}_m^H$ *denote the singular value decomposition of the* m*th component matrix of* $\mathbf{X}$*. According to* (3.25)*, the* m*th component matrix of* $\Upsilon_\alpha(\mathbf{X})$ *is given by* $\Upsilon_\alpha(\mathbf{X})|_m = -\mathbf{U}_m\mathbf{S}_m\mathbf{V}_m^H$*, where* $(\forall m \in \mathcal{M})$

$$\mathbf{S}_m = \mathrm{diag}\left(\min(\sigma_1(\mathbf{X}_m), \tau), \sigma_2(\mathbf{X}_m), \ldots, \sigma_N(\mathbf{X}_m)\right)$$

with $\tau = \alpha\sigma_{\max}(\mathbf{X})$*. Since* $(\forall \mathbf{W} \in \mathcal{H}^M)\ \|\mathbf{W}\|^2 = \sum_{i\in\mathcal{N}} \sigma_i^2(\mathbf{W})$*, we can write*

$$|\!|\!|\Upsilon_\alpha(\mathbf{X})|\!|\!|^2 = \sum_{m=1}^{M} \|\mathbf{S}_m\|^2 \leq \sum_{m=1}^{M} \|\mathbf{X}_m\|^2 = |\!|\!|\mathbf{X}|\!|\!|^2,$$

which concludes the proof. □

Using Lemma 3.1, we can show that the sequence of perturbations in (3.26) is bounded:

Lemma 3.2. *Let* $\left(\beta^{(n)}\right)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ *and suppose that* $(\forall n \in \mathbb{N})\ \alpha^{(n)} \geq 0$*. Then the sequence of perturbations* $\left(\beta^{(n)}\mathbf{Y}^{(n)}\right)$ *with* $\mathbf{Y}^{(n)}$ *defined by* (3.26) *is bounded.*

Proof. *We need to show that* $(\exists R \in \mathbb{R})(\forall n \in \mathbb{N})\ \left|\!\left|\!\left|\mathbf{Y}^{(n)}\right|\!\right|\!\right| \leq R$. *To this end, observe that* $\left(\forall \mathbf{X}^{(n)} \in \mathcal{H}^M\right)(\forall \mathbf{Z} \in \operatorname{Fix}(T_\star))$ *it holds that*

$$\begin{aligned}\left|\!\left|\!\left|\mathbf{X}^{(n+1)} - \mathbf{Z}\right|\!\right|\!\right| &= \left|\!\left|\!\left|T_\star\left(\mathbf{X}^{(n)} + \beta^{(n)}\mathbf{Y}^{(n)}\right) - \mathbf{Z}\right|\!\right|\!\right| \\ &\overset{(i)}{\leq} \left|\!\left|\!\left|\mathbf{X}^{(n)} + \beta^{(n)}\mathbf{Y}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| \\ &\overset{(ii)}{\leq} \left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| + \beta^{(n)}\left|\!\left|\!\left|\mathbf{Y}^{(n)}\right|\!\right|\!\right|,\end{aligned}$$

where (i) follows from the nonexpansivity of $T_\star$*, and (ii) is a consequence of the triangle inequality. By Lemma 3.1, the perturbations defined in* (3.26) *satisfy* $(\forall n \in \mathbb{N})$ $\left|\!\left|\!\left|\mathbf{Y}^{(n)}\right|\!\right|\!\right| \leq \left|\!\left|\!\left|\mathbf{X}^{(n)}\right|\!\right|\!\right|$. *Consequently, applying the triangle inequality again yields*

$$\left|\!\left|\!\left|\mathbf{X}^{(n+1)} - \mathbf{Z}\right|\!\right|\!\right| \leq \left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| + \beta^{(n)}\left(\left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| + |\!|\!|\mathbf{Z}|\!|\!|\right).$$

By defining $(\forall n \in \mathbb{N})$ $a^{(n)} = \left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right|$ *and* $\gamma^{(n)} = \beta^{(n)}|\!|\!|\mathbf{Z}|\!|\!|$*, we can deduce from Fact 1.14 that the sequence* $\left(a^{(n)}\right)_{n\in\mathbb{N}}$ *converges. This implies that there exists* $r \in \mathbb{R}$ *such that* $(\forall n \in \mathbb{N})$ $\left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| \leq r$. *Consequently, we have* $(\forall n \in \mathbb{N})$

$$\left|\!\left|\!\left|\mathbf{Y}^{(n)}\right|\!\right|\!\right| \overset{(i)}{\leq} \left|\!\left|\!\left|\mathbf{X}^{(n)}\right|\!\right|\!\right| \overset{(ii)}{\leq} \left|\!\left|\!\left|\mathbf{X}^{(n)} - \mathbf{Z}\right|\!\right|\!\right| + |\!|\!|\mathbf{Z}|\!|\!| \overset{(iii)}{\leq} r + |\!|\!|\mathbf{Z}|\!|\!| =: R$$

where (i) follows from Lemma 3.1, (ii) follows from the triangle inequality, and (iii) follows from Fact 1.14. □

By combining Fact 1.14 and Lemma 3.2 we can show that the proposed algorithm converges to a feasible point of the relaxed SDP in (3.11). This is summarized in the following proposition.

Proposition 3.3. *Assume that Problem 3.10 is feasible. Then the sequence produced by the algorithm in* (3.12) *with perturbations given by* (3.26) *is guaranteed to converge to a feasible point of Problem* (3.11) *for all* $\mathbf{X}^{(0)} \in \mathcal{H}^M$ *if* $\left(\alpha^{(n)}\right)_{n\in\mathbb{N}}$ *is a sequence in* $\mathbb{R}_+$ *and* $\left(\beta^{(n)}\right)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$.

Proof. *Since* $\mathcal{H}^M$ *is finite dimensional, this follows directly from Theorem 2.1 and Lemma 3.2.*

3.3.4. Relation to the Superiorization Methodology

The author of [Cen15] defines superiorization as follows:

> *'The superiorization methodology works by taking an iterative algorithm, investigating its perturbation resilience, and then, using proactively such permitted perturbations, forcing the perturbed algorithm to do something useful in addition to what it is originally designed to do.'*

Although our proposed algorithm matches this informal definition, there are some slight differences to the formal definition in [Cen15], where the perturbations $\mathbf{Y}^{(n)}$ are required to be nonascending vectors for a convex superiorization objective function.

Definition 3.1 (Nonascending Vectors [Cen15])**.** Given a function $\phi : \mathbb{R}^J \to \mathbb{R}$ and a point $\mathbf{y} \in \mathbb{R}^J$, a vector $\mathbf{d} \in \mathbb{R}^J$ is said to be nonascending for ϕ at $\mathbf{y}$ iff $\|\mathbf{d}\| \leq 1$ and there is a $\delta > 0$ such that for all $\lambda \in [0, \delta]$ we have $\phi(\mathbf{y} + \lambda\mathbf{d}) \leq \phi(\mathbf{y})$.

In our case, the goal of superiorization is two-fold, in the sense that it is expressed by two separate functions $f_1 : \mathcal{H}^M \to \mathbb{R}$ and $f_2 : \mathcal{H}^M \to \mathbb{R}$. While the function f_1 in (3.17) is convex, the function f_2 in (3.23) (i.e., the distance to nonconvex rank constraint set $\mathcal{R}$ in (3.9)) is a nonconvex function. Moreover, we use perturbations that are not restricted to a unit ball, and therefore they are not necessarily nonascending vectors. However, since the sequence $(\mathbf{Y}^{(n)})_{n\in\mathbb{N}}$ is bounded, $(\exists r > 0)(\forall n \in \mathbb{N})$ $\left|\left|\left|\mathbf{Y}^{(n)}\right|\right|\right| \leq r$, so we can define $(\forall n \in \mathbb{N})$ $\bar{\beta}^{(n)} = r\beta^{(n)}$ and $\bar{\mathbf{Y}}^{(n)} := \frac{1}{r}\mathbf{Y}^{(n)}$. Now $\bar{\mathbf{Y}}^{(n)}$ is nonascending for both superiorization objectives f_1 and f_2 for all $n \in \mathbb{N}$, and $\left(\bar{\beta}^{(n)}\right)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$. Keeping these slight distinctions in mind, we will refer to the proposed algorithm in (3.12) as *Superiorized Projections onto Convex Sets*.

3.3.5. Summary of the Proposed Algorithm

The proposed multi-group multicast beamforming algorithm is summarized in Algorithm 1. It is defined by the relaxation parameters $\mu_1, \ldots, \mu_{K+2}$ of the mapping $T_\star$ in (3.14), a scalar $a \in (0, 1)$ controlling the decay of the power-reducing perturbations, a scalar $b \in (0, 1)$ controlling the decay of the sequence of perturbation scaling factors, i.e., $(\forall n \in \mathbb{N})$ $\alpha^{(n)} = a^n$ and $\beta^{(n)} = b^n$. The stopping criterion is based on a tolerance value $\epsilon > 0$, and a maximum number $n_{\max}$ of iterations.

The arguments of the algorithm are the indices $g_1, \ldots, g_K$ assigning a multicast group to each user, the channel vectors $\mathbf{h}_1, \ldots, \mathbf{h}_K \in \mathbb{C}^N$, SINR requirements $\gamma_1, \ldots, \gamma_K$, and noise powers $\sigma_1, \ldots, \sigma_K$ of all users as well as the per-antenna power constraints $p_1, \ldots, p_N$. At each step, the algorithm computes a perturbation according to (3.25) and applies the feasibility seeking operator $T_\star$ in (3.14). It terminates when the relative variation of the estimate falls within the tolerance ϵ, or when the maximum number $n_{\max}$ of iterations is

reached. Finally, the beamforming vectors $\mathbf{w} = \{\mathbf{w}_m\}_{m\in\mathcal{M}}$ are computed by extracting the strongest principal component

$$(\forall m \in \mathcal{M}) \quad \mathbf{w}_m = \psi(\mathbf{X}_m) := \sqrt{\sigma_1(\mathbf{X}_m)}\mathbf{u}_{m1}, \tag{3.28}$$

where $(\forall m \in \mathcal{M})$ $\mathbf{X}_m = \mathbf{U}_m\mathbf{\Sigma}_m\mathbf{V}_m^H$, $\mathbf{\Sigma}_m = \operatorname{diag}(\sigma_1(\mathbf{X}_m), \ldots, \sigma_N(\mathbf{X}_m))$, and $\mathbf{U}_m = [\mathbf{u}_{m1}, \cdots, \mathbf{u}_{mN}]$.

Algorithm 1 Superiorized Projections onto Convex Sets

Parameters: $\{\mu_k\}_{k=1}^{K+2}$, $a, b \in (0,1)$, $\epsilon > 0$, $n_{\max} \in \mathbb{N}$
Input: $\{g_k\}_{k\in\mathcal{K}}$, $\{\mathbf{h}_k\}_{k\in\mathcal{K}}$, $\{\gamma_k\}_{k\in\mathcal{K}}$, $\{\sigma_k\}_{k\in\mathcal{K}}$, $\{p_i\}_{i\in\mathcal{N}}$
Output: $\{\mathbf{w}_m \in \mathbb{C}^N\}_{m\in\mathcal{M}}$
Initialization: Choose arbitrary $\mathbf{X}^{(0)} \in \mathcal{H}^M$
for $n = 0, \ldots, n_{\max} - 1$ **do**
  $\mathbf{Y}^{(n)} \leftarrow \Upsilon_{a^n}\left(\mathbf{X}^{(n)}\right)$ ▷ Eq. (3.25)
  $\mathbf{X}^{(n+1)} \leftarrow T_\star\left(\mathbf{X}^{(n)} + b^n\mathbf{Y}^{(n)}\right)$ ▷ Eq. (3.14)
  if $\left|\left|\left|\mathbf{X}^{(n+1)} - \mathbf{X}^{(n)}\right|\right|\right| < \epsilon\left|\left|\left|\mathbf{X}^{(n+1)}\right|\right|\right|$ **then**
    break
  end if
end for
return $\mathbf{w} = \left\{\psi\left(\mathbf{X}_m^{(n+1)}\right)\right\}_{m\in\mathcal{M}}$ ▷ Eq. (3.28)

3.4. Numerical Results

In this section, we compare Algorithm 1 (`S-POCS`) to several other methods from the literature. We choose identical noise levels and target SINRs for all users, i.e., $(\forall k \in \mathcal{K})$ $\sigma_k = \sigma$ and $\gamma_k = \gamma$. For each problem instance, we generate K independent and identically distributed (i.i.d.) Gaussian channel vectors $(\forall k \in \mathcal{K})$ $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$. In the first simulation, we drop the per-antenna power constraints, i.e., we set $(\forall i \in \mathcal{N})$ $p_i = \infty$, and we consider the following algorithms:

- The proposed method summarized in Algorithm 1 (`S-POCS`)
- Semidefinite relaxation with Gaussian randomization [KSL08] (`SDR-GauRan`)
- The successive convex approximation algorithm from [MHG+14, CCO15] (`FPP-SCA`)
- The ADMM-based convex-concave procedure from [CT17] (`CCP-ADMM`)

The `S-POCS` algorithm is as described in Algorithm 1, with parameters $a = 0.95$, $b = 0.999$, $\epsilon = 10^{-6}$, $n_{\max} = 10^5$. For the QoS-constraint sets, we use relaxation parameters $(\forall k \in \mathcal{K})$

$\mu_k = 1.9$, and for the per-antenna power constraint set $\mathcal{P}$ and the PSD constraint $\mathcal{C}_+$, we use unrelaxed projections, i.e., $\mu_{K+2} = \mu_{K+1} = 1$. We initialize the `S-POCS` algorithm with $\mathbf{X}^{(0)} = \mathbf{0}$. The convex optimization problems in the `SDR-GauRan` and `FPP-SCA` algorithms are solved with the interior point solver SDPT3 [TTT99]. Calling this solver from MATLAB results in a small computational overhead. While SDPT3 is written partly in MATLAB and partly in C, the remaining methods are implemented in pure MATLAB. The parameters of the `CCP-ADMM` algorithm are as specified in [CT17]. `FPP-SCA` and `CCP-ADMM` are initialized with coefficients drawn i.i.d. from $\mathcal{CN}(0,1)$. If $[\mathbf{h}_1, \dots \mathbf{h}_K] \in \mathbb{C}^{N\times K}$ has full column rank, we initialize `CCP-ADMM` with the closed-form starting point proposed in [CT17].

The `SDR-GauRan` algorithm begins by solving the relaxed problem in (3.11), and, subsequently, generates random candidate beamforming vectors using the `RandA` method [SDL06], [KSL08]. In the multi-group setting, where $M > 1$, an additional convex optimization problem (multigroup multicast power control (MMPC), [KSL08]) needs to be solved for each candidate vector. If no feasible MMPC problem is found during the `RandA` procedure, we define the output of the `SDR-GauRan` algorithm to be $\{\psi(\mathbf{X}_m^\star)\}_{m\in\mathcal{M}}$, where $\mathbf{X}^\star \in \mathcal{H}^M$ is a solution to the relaxed SDP in (3.11).

The `FPP-SCA` algorithm from [MHG$^+$14] works by solving a sequence of convex subproblems. By introducing slack variables, the feasibility of each subproblem is ensured. This obviates the need for a feasible initialization point, which is typically required to ensure convergence of CCP/SCA algorithms. The `CCP-ADMM` algorithm uses an ADMM algorithm to find a feasible starting point for the CCP. Subsequently, a similar ADMM algorithm is used to approximate each subproblem of the CCP. Because the ADMM is a first order method, the performance of `CCP-ADMM` is heavily dependent on the stopping criteria of the inner ADMM algorithm. Unlike `CCP-ADMM`, the `S-POCS` algorithm does not require an initialization phase, and it works by iteratively applying a sequence of mappings, which can be computed in a fixed number of steps.

Owing to the considerable differences in the functioning of these methods, achieving a fair comparison of their complexity is difficult. In the worst case, obtaining a solution with accuracy $\epsilon > 0$ for a convex SDP with k constraints over the n-dimensional (real-valued) PSD cone can take $\mathcal{O}\left(\max\{n,k\}^4 n^{\frac{1}{2}} \log(1/\epsilon)\right)$ operations asymptotically [LMS$^+$10]. CCP/SCA algorithms approximate nonconvex problems, so convergence to an optimal point can not be guaranteed in general. As a result, their total complexity cannot be expressed as a function of the accuracy ϵ. The complexity of each `FPP-SCA` iteration is $\mathcal{O}([MN+K]^{3.5})$ [MHG$^+$14]. However, general purpose solvers often apply heuristics that achieve considerable complexity reduction. Therefore, the aforementioned asymptotic results can be rather pessimistic in practice. The `CCP-ADMM` involves inverting a matrix, the cost of which is $\mathcal{O}(N^3)$ [CT17]. However, this matrix inverse needs to be computed only

once for each problem, and the complexity of the subsequent iterative steps is significantly lower. By contrast, the proposed `S-POCS` algorithm requires M eigen-decompositions in every iteration, resulting in a per-iteration complexity of $\mathcal{O}(MN^3)$. Unfortunately, without theoretical guarantees on the convergence rate, the use of per-iteration complexities is limited to assess the overall computational cost of an algorithm.

Therefore, as in [CT17], [CCC+12], we use the computation time of a specific implementation to assess the computational cost of each method. Although we exclude the time required for evaluating the performance, we note that the computation time required by each of the methods severely depends on the particular implementation. All algorithms are run on identical nodes of a high-performance computing central processing unit (CPU) cluster. This implementation does not exploit the potential of parallel computation using graphics processing units (GPUs), from which we expect significant acceleration especially for `CCP-ADMM`. Nevertheless, we emphasize that the proposed algorithm also achieves a smaller approximation gap in many scenarios.

The authors of [CT17] assess the performance of the considered algorithms by comparing the transmit power achieved by the resulting beamformers. However, none of the methods considered here can guarantee feasibility of the beamforming vectors, when the algorithms are terminated after a finite number of iterations. Furthermore, in the multi-group case, it may not be possible to scale an arbitrary candidate beamformer $\mathbf{w} = \{\mathbf{w}_m \in \mathbb{C}^N\}_{m\in\mathcal{M}}$ such that it satisfies all constraints in Problem (3.1). In principle, we could evaluate the performance by observing both the objective value (i.e., the transmit power of the beamformers) and a measure of constraints violation such as the normalized proximity function used in [CCC+12]. However, defining this measure of constraints violation is not straightforward, as the considered methods approach the problem in different spaces. Moreover, we are interested in expressing the quality of a beamforming vector by a single value to simplify the presentation. Therefore, we will compare the performance based on the minimal SINR achieved by the beamformer $\sqrt{\rho(\mathbf{w})}\cdot\mathbf{w}$ with

$$\rho(\mathbf{w}) = \min\left(\frac{P^\star_{\mathrm{SDR}}}{\sum_{m=1}^M \mathbf{w}_m^H\mathbf{w}_m}, \min_{i\in\mathcal{N}}\left(\frac{p_i}{\sum_{m=1}^M |w_{im}|^2}\right)\right).$$

The scaled vector $\sqrt{\rho(\mathbf{w})}\cdot\mathbf{w}$ satisfies all power constraints, and its total power is bounded by the optimal objective value $P^\star_{\mathrm{SDR}}$ of the relaxed SDP in (3.11). More compactly, given a candidate beamformer $\mathbf{w} = \{\mathbf{w}_m \in \mathbb{C}^N\}_{m\in\mathcal{M}}$ for Problem (3.1), we assess its performance

based on the function[5]

$$\mathrm{SINR}_{\rho}^{\min}(\mathbf{w}) = \min_{k\in\mathcal{K}} \frac{|\mathbf{w}_m^H \mathbf{h}_k|^2}{\sum_{l\neq m} |\mathbf{w}_l^H \mathbf{h}_k|^2 + \frac{\sigma_k^2}{\rho(\mathbf{w})}}. \tag{3.29}$$

Since $P_{\mathrm{SDR}}^{\star}$ is a lower bound on the objective value of the original problem in (3.1), it holds $(\forall\{\mathbf{w}_m \in \mathbb{C}^N\}_{m\in\mathcal{M}})$ that $\mathrm{SINR}_{\rho}^{\min}(\mathbf{w}) \leq \gamma$, where equality can only be achieved, if the relaxed problem in (3.11) has a solution composed of rank-one matrices.

3.4.1. Performance vs. Computation Time

We will now examine how the figure of merit in (3.29) evolves over time for beamforming vectors produced by the respective algorithms. Figure 3.1 shows the performance comparison for an exemplary scenario with $N = 20$ antennas, and $K = 20$ users split evenly into $M = 2$ groups, where $\sigma = 1$, $\gamma = 1$, and $(\forall i \in \mathcal{N})$ $p_i = \infty$. It can be seen that the

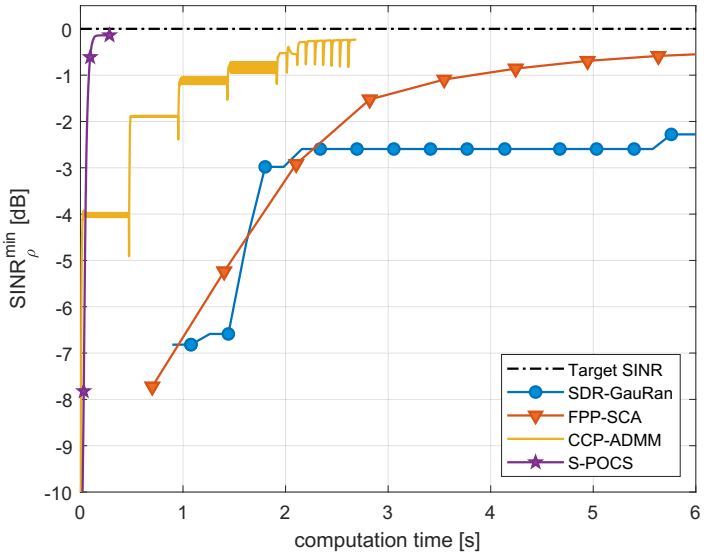


Figure 3.1.: $\mathrm{SINR}_{\rho}^{\min}(\mathbf{w}^{(t)})$ over time in a system with $N = 20$ antennas and $K = 20$ users users split evenly into $M = 2$ multicast groups.

`S-POCS` algorithm quickly converges to a point achieving an SINR close to the specified target value γ. The discontinuities in the SINR curve for the `CCP-ADMM` algorithm are due to the inner- and outer optimization loops. For the `SDR-GauRan` algorithm, the SINR

[5]For the sake of simplicity, we will refer to the figure of merit $\mathrm{SINR}_{\rho}^{\min}$ in (3.29), as *SINR* in the following. Note that $\mathrm{SINR}_{\rho}^{\min} < \gamma$ does not imply that $\mathbf{w}$ is infeasible, because $\mathrm{SINR}_{\rho}^{\min}$ is the minimal SINR achieved by the *scaled* beamformer $\sqrt{\rho(\mathbf{w})} \cdot \mathbf{w}$.

increases whenever the randomization produces a beamformer with better performance than the previous one. The SINR of the `FPP-SCA` algorithm improves continuously, albeit more slowly than the `S-POCS` and `CCP-ADMM` algorithms.

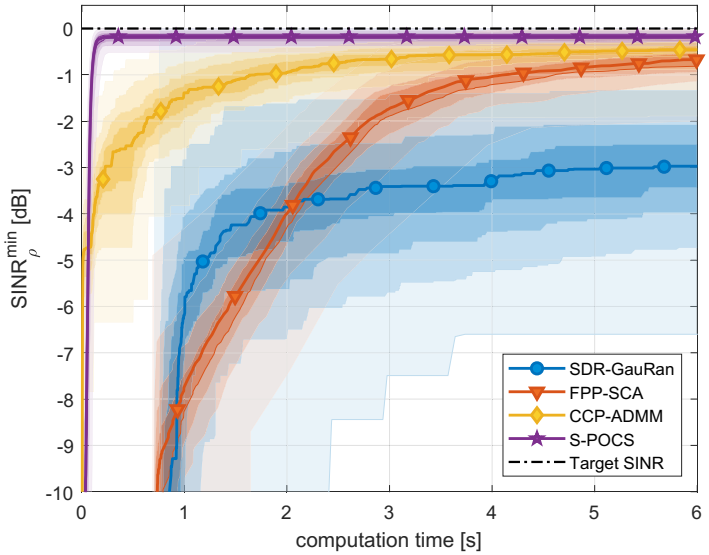


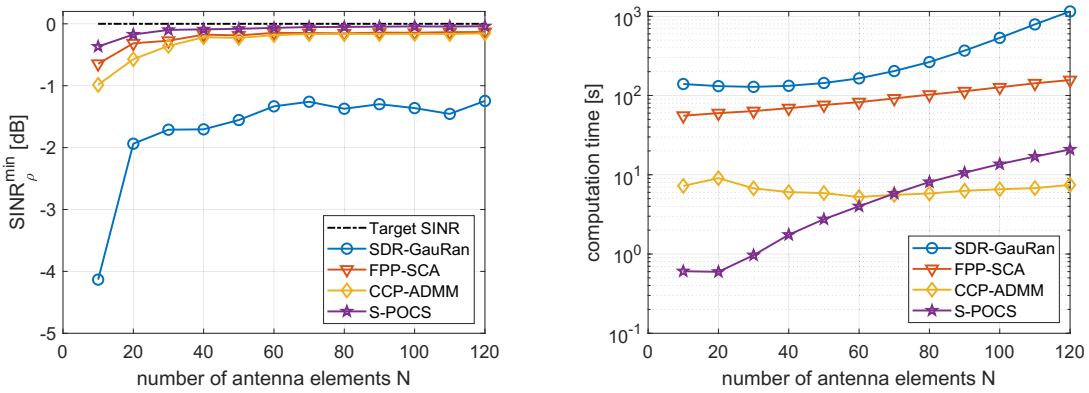
Figure 3.2.: $\mathrm{SINR}_\rho^{\min}(\mathbf{w}^{(t)})$ over time in a system with $N = 20$ antennas and $K = 20$ users split evenly into $M = 2$ multicast groups. The shaded regions include the outcomes for 100%, 75%, 50%, and 25% out of 100 problem instances, respectively, and the bold line represents the median.

Next, we evaluate the performance over 100 randomly generated problems. Since the SINR does not increase monotonically for all of the methods considered, we assume that each algorithm can keep track of the best beamformer produced so far. In this way, the oscillations in the SINR metric for the `CCP-ADMM` algorithm do not have a negative impact on its average performance. Figure 3.2 shows the performance of the beamforming vectors computed with the respective algorithms over time for a system with $N = 20$ transmit antennas, and $K = 20$ users split evenly into $M = 2$ multicast groups. The shaded regions correspond to the 100%, 75%, 50%, and 25% quantiles over all randomly generated problems. More precisely, the margins of the shaded regions correspond to the 1st, 13th, 26th, 38th, 63rd, 75th, 88th, and 100th out of 100 sorted y-axis values. For each algorithm, the median is represented by a bold line. The `S-POCS` algorithm achieves the highest median SINR, while requiring the lowest computation time among all methods considered. Moreover, it can be seen that the variation around this median value is less severe compared to the remaining approaches. Put differently, the time required for reaching a certain SINR varies much less severely for the `S-POCS` algorithm than for

the remaining methods. This can be of particular interest in delay sensitive applications, where a beamforming vector for a given channel realization must be computed within a fixed time period.

3.4.2. Varying number of antennas

In this subsection, we investigate the impact of the transmit antenna array size N on the performance of the respective beamforming algorithms. To do so, we generate 100 random problem instances for each array size N with $K = 20$ users split evenly in to $M = 2$ multicast groups. We choose unit target SINR and unit noise power for all users, and unit per-antenna power constraints, i.e., $\gamma = 1$, $\sigma = 1$ and $(\forall i \in \mathbb{N})\ p_i = 1$. For the



(a) (b)

Figure 3.3.: (a) $\mathrm{SINR}_\rho^{\min}(\mathbf{w})$ and (b) computation time for $K = 20$ users split evenly into $M = 2$ groups for varying antenna array sizes N.

`SDR-GauRan` algorithm, we generate 200 candidate beamforming vectors for each problem instance. We use the `CCP-ADMM` algorithm with parameters as specified in [CT17]. Since the inner ADMM iteration converges slowly for some problem instances, we set the maximal number of steps of the ADMM to $j_{\max} = 300$. For the outer CCP loop, we use the stopping criteria specified in [CT17], i.e., we stop the algorithm once the relative decrease of the objective value is below 10^{-3} or $t_{\max} = 30$ outer iterations are exceeded. For the `FPP-SCA` algorithm, we use a fixed number of 30 successive convex approximation steps.

Figure 3.3(a) shows the figure of merit in (3.29) for different numbers N of transmit antennas, averaged over 100 random problem instances each. For all N, `S-POCS` achieves highest value for $\mathrm{SINR}_\rho^{\min}(\cdot)$, followed by the `FPP-SCA`, `CCP-ADMM`, and `SDR-GauRan` algorithms. For $N \geq 80$, the `S-POCS` algorithm achieves an SINR of $\mathrm{SINR}_\rho^{\min}(\mathbf{w}_{\texttt{S-POCS}}) \geq -0.05\,\mathrm{dB}$. By contrast, the remaining methods do not exceed $\mathrm{SINR}_\rho^{\min}(\mathbf{w}_{\texttt{FPP-SCA}}) =$

$-0.12\,\mathrm{dB}$, $\mathrm{SINR}_\rho^{\min}(\mathbf{w}_{\texttt{CCP-ADMM}}) \geq -0.15\,\mathrm{dB}$, $\mathrm{SINR}_\rho^{\min}(\mathbf{w}_{\texttt{SDR-GauRan}}) \geq -1.18\,\mathrm{dB}$, respectively. The corresponding average computation times are shown in Figure 3.3(b). The `S-POCS` algorithm requires 0.26 %–2.38 % of the computation time required by `SDR-GauRan`, 0.95 %–11.64 % of the computation time required by `FPP-SCA`, and 6.49 %–233.6 % of the computation time required by `CCP-ADMM`. For $N \geq 80$, the computation time of `S-POCS` exceeds that of `CCP-ADMM`.

3.4.3. Varying number of users

In the following simulation, we fix an array size of $N = 50$ antenna elements, and we evaluate the performance of each method for varying numbers K of users split evenly into $M = 4$ multicast groups. Figure 3.4(a) shows the figure of merit in (3.29) averaged over 100 random problem instances for each K. As before, we choose $\gamma = 1$, $\sigma = 1$, and $(\forall i \in \mathcal{N})\ p_i = 1$. The second y-axis shows the average rank[6] of the M component matrices of the solution $\mathbf{X}^\star$ to the relaxed SDP in (3.11) computed by `SDR-GauRan`. For $K \leq 8$, the components of $\mathbf{X}^\star$ are rank-one in all cases. With increasing K, the average rank grows monotonically, reaching 2.88 for $K = 128$.

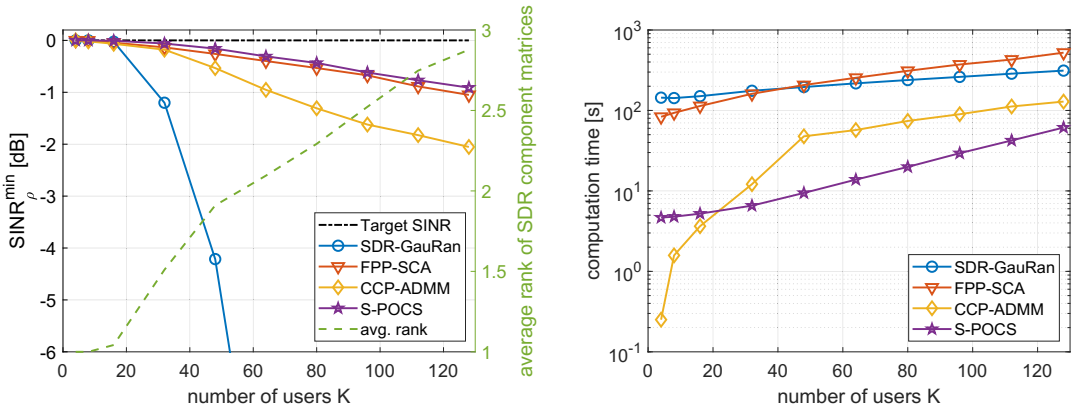
As expected, the optimality gap (represented by the difference between the target SINR the scaled SINR figure of merit $\mathrm{SINR}_\rho^{\min}$ in (3.29)) increases with the average rank of solutions to the relaxed SDP. While all algorithms achieve close to optimal performance for small numbers of users, the SINR in (3.29) decreases considerably faster for `SDR-GauRan` than for the remaining methods. For all values of K, `S-POCS` achieves the highest value for $\mathrm{SINR}_\rho^{\min}(\cdot)$ among all methods.

The corresponding average computation times are shown in Figure 3.4(b). `S-POCS` requires 3.23 %–19.59 % of the computation time required by `SDR-GauRan`, 4.07 %–11.72 % of the computation time required by `FPP-SCA`, and 19.78 %–1858 % of the computation time required by `CCP-ADMM`. While the `CCP-ADMM` takes only a fraction of the time required by `S-POCS` for small K, it slows down considerably as K increases. For moderate and large numbers of users, `S-POCS` outperforms the remaining methods in terms of both approximation gap and computation time.

3.4.4. Varying Target SINR

In the following simulation, we evaluate the impact of the target SINR on the respective algorithms in a system with $N = 30$ antenna elements, $K = 20$ users split evenly into $M = 2$ multicast groups, and unit noise power $\sigma = 1$. Since the target SINR has a strong

[6] More precisely, we determine the rank of a matrix by counting the number of singular values with magnitude greater than 10^{-6}.



(a) (b)

Figure 3.4.: (a) $\mathrm{SINR}_\rho^{\min}(\mathbf{w})$ for a system with $N = 50$ transmit antennas and a varying number of users split evenly into $M = 4$ multicast groups. The right y-axis shows the average rank of solutions to the relaxed SDP in (3.11). (b) The corresponding computation times.

impact on the transmit power, we set $(\forall i \in \mathcal{N})\ p_i = \infty$, to avoid generating infeasible instances of Problem (3.1). Figure 3.5(a) shows the figure of merit in (3.29) achieved by each method for the respective target SINR. Except for the `SDR-GauRan` algorithm, which exhibits a gap of about 2 dB to the target SINR, all methods achieve close to optimal performance for each target SINR. Figure 3.5(b) shows the computation time required by each algorithm for varying target SINR γ. The average computation time of `SDR-GauRan` and `FPP-SCA` is almost constant. For `CCP-ADMM`, the computation decreases slightly with an increasing target SINR. While the proposed `S-POCS` algorithm converges quickly for low target SINR levels, its computation time exceeds that of the `CCP-ADMM` for target SINRs above 8 dB. This indicates that the best choice of first order algorithms for multicast beamforming depends on the regime in which the system is operated.

3.5. Conclusion and Final Remarks

In this chapter, we presented an algorithm for multi-group multicast beamforming with per-antenna power constraints. We showed that the sequence produced by this algorithm is guaranteed to converge to a feasible point of the relaxed SDP, while the perturbations added in each iteration reduce the objective value and the distance to the nonconvex rank constraints. Numerical comparisons show that the proposed method outperforms state-of-the-art algorithms in terms of both approximation gap and computation time in many cases. While for very small problem dimensions, existing algorithms are similarly

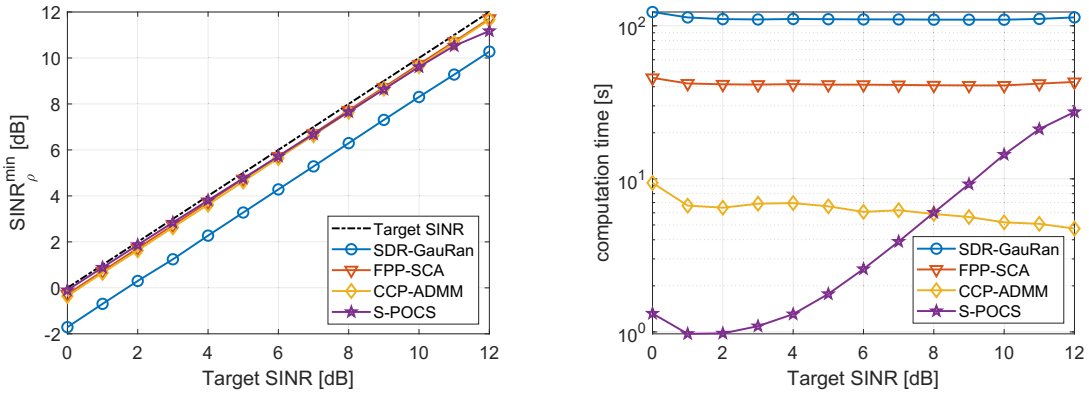


Figure 3.5.: (a) $\mathrm{SINR}_\rho^{\min}(\mathbf{w})$ for a system with $N = 30$ transmit antennas and $K = 20$ users split evenly into $M = 2$ multicast groups. (b) The corresponding computation times.

or even more efficient in terms of performance versus complexity, the advantage of the proposed method is particularly pronounced in the low target SINR regime as well as for large numbers of receivers. This makes the proposed method particularly relevant for low-energy or massive access applications. In comparison to other techniques, the computation time of the proposed method varies less severely across different problem instances of the same dimension. In communication systems, which are typically subject to strict latency constraints, the iteration can be terminated after a fixed number of steps without suffering severe performance loss. The applicability of the proposed algorithm is not restricted to the multicast beamforming problem considered here. A slight modification of the rank constraint naturally leads to algorithms for rank-two Alamouti multicast beamforming (see [WSM12], [WMS13]) or general rank multicast beamforming (see [LWP13], [TP20]). Several advantages of the proposed method are summarized below:

- The proposed method is guaranteed to converge to a feasible point of the relaxed SDP.
- The perturbations added in each iteration provably reduce both the objective value and the distance to the rank constraint.
- The proposed method is free from nested optimization loops, so its implementation does not require trading off inner against outer optimization steps.
- The bounded perturbation resilience of POCS guarantees the convergence of the proposed algorithm under mild conditions, even if projections onto the PSD-cone

are only computed approximately.

- The simple structure of the proposed method allows for a straightforward implementation in real-world systems.
- The proposed algorithm typically achieves most progress towards a solution during the initial iterations. This allows a simple trade-off between performance and complexity by choosing the maximal number of iterations.
- Unlike CCP/SCA algorithms, the proposed method can be extended straightforwardly to higher-rank multicast beamforming settings.

4. MIMO Detection

In this chapter, we use Theorem 2.3 to devise a superiorized APSM with provable convergence properties to tackle the nonconvex MIMO detection problem. Although various low-complexity MIMO detection algorithms achieve excellent performance on i.i.d. Gaussian channels, they typically incur high performance losses when more realistic channel models are considered. In this chapter, we propose a set-theoretic framework for MIMO detection. Compared to existing low-complexity iterative detectors such as AMP, the proposed algorithms can achieve considerably lower symbol error ratios over correlated channels. At the same time, the proposed methods do not require matrix inverses, such that their complexity is similar to AMP. This chapter is an extended version of [10]. Most results in this chapter, together with the results in Section 2.2, are included in the journal paper [11], which has been submitted for publication. A major part of the formulations are identical to the version in [11].

4.1. Introduction

Throughout the past decades, data rates in mobile networks have been growing exponentially [FZ08]. On a global scale, mobile data consumption is increasing rapidly [Cis14, Cis19]. In the near future, the growing number of mobile subscriptions, improvements in the performance of mobile devices, and new service types such as Internet of Things (IoT) communication are expected to continue driving the growth of global mobile data consumption [Eri21]. This continuing trend calls for an increasingly efficient use of spectral resources in order to meet the growing demand for mobile data. It is well-known that multi-antenna systems can increase the achievable rate without broadening the bandwidth [Tel99]. Therefore, such MIMO communication systems have gained much attention during the past two decades. While commercial fourth generation (4G) base stations typically only use two or four antennas for transmission and reception [KCKP16], future communication systems are envisioned to rely on base stations with large antenna arrays, which simultaneously serve many users [LVdP17].

Signals transmitted over MIMO channels typically suffer from co-channel interference, which complicates signal detection. In fact, the problem of optimal detection in MIMO

systems is known to be NP-hard [vEB81, Ver89, Mic01]. The MIMO detection problem has been studied for decades. Yet the growing interest in large-scale multi-antenna systems still drives the need for low-complexity approximation techniques. A comprehensive overview of MIMO detection algorithms can be found in [YH15] and [AJS19]. The authors of [JGMS15] propose a low-complextity MIMO detector based on AMP. They show that this individually-optimal large-MIMO AMP (IO-LAMA) algorithm is optimal for MIMO detection over i.i.d. Gaussian channels in the large-system limit under some additional conditions. In [MP17], the authors relax the assumption of i.i.d. Gaussian channels by proposing an orthogonal approximate message passing (OAMP) algorithm for MIMO detection over the more general class of unitarily invariant channel matrices. In contrast to the AMP detector proposed in [JGMS15], each iteration of OAMP involves a matrix inversion in order to compute the linear minimum mean square error (LMMSE) estimate, making OAMP more computationally complex than IO-LAMA. Many recent publications on MIMO detection [SDW17, SDW19, GDN18, USMC19, NL20, KAHF20, HWJL18b] propose deep-unfolded versions of iterative detectors. Despite their celebrated success, some of these techniques have been found to suffer considerable performance losses on realistic channels. The authors of [KAHF20] mitigate this problem by proposing an online training scheme, which in turn increases the computational cost compared to deep-unfolded algorithms that are trained offline.

In this chapter, we approach MIMO detection from a set-theoretic perspective. By posing the problem in a real Hilbert space, we can apply Theorem 2.3 to devise iterative MIMO detectors with provable convergence properties based on a superiorized APSM. The proposed detectors have a per-iteration complexity similar to IO-LAMA. At the same time, unlike IO-LAMA, the proposed methods are guaranteed to converge for arbitrary channel matrices. Owing to their iterative structure, the proposed algorithms can be readily used as a basis for deep-unfolded detection algorithms. Simulations show that, despite their low complexity, the proposed methods can outperform the more complex OAMP detector on realistic channels. Moreover, deep-unfolded versions of the proposed methods can achieve competitive performance within considerably fewer iterations.

4.2. Problem Statement

We consider a MIMO system with N_{T} transmit- and N_{R} receive antennas. For square constellations such as quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM), which are commonly used in practice, we can describe the system

using the real-valued signal model [Tel99]

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w},$$

where $\mathbf{y} \in \mathbb{R}^{2N_{\mathrm{R}}}$ is the received signal, $\mathbf{H} \in \mathbb{R}^{2N_{\mathrm{R}} \times 2N_{\mathrm{T}}})$ is the channel matrix, $\mathbf{s} \in \mathbb{R}^{2N_{\mathrm{T}}}$ is the transmit signal with coefficients $(\forall i \in \mathcal{I}_{\mathrm{T}} := \{1, \ldots, 2N_{\mathrm{T}}\})$ $s_i \in \mathcal{A} \subset \mathbb{R}$ drawn independently from a uniform distribution over the set $\mathcal{A}$ of real-valued constellation points, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I}_{2N_{\mathrm{R}}})$ is a vector of i.i.d. Gaussian noise samples.

The goal of MIMO detection is to estimate the transmit signal vector $\mathbf{x}$ based on knowledge of the channel $\mathbf{H}$ and the received signal vector $\mathbf{y}$. Since the entries of $\mathbf{s}$ are distributed uniformly over the constellation alphabet and $\mathbf{w}$ is a vector of Gaussian noise, the optimal detector uses the maximum likelihood (ML) criterion given by

$$\mathbf{s}^\star \in \arg\max_{\mathbf{x} \in \mathcal{S}} p\left(\mathbf{y}|\mathbf{x}\right) = \arg\min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2, \tag{4.1}$$

where $\mathcal{S} := \mathcal{A}^{2N_{\mathrm{T}}} \subset \mathbb{R}^{2N_{\mathrm{T}}}$ is the discrete set of feasible transmit signal vectors. The ML problem is known to be NP-hard [Mic01] (and in fact, NP-complete [DPDM17]). Therefore, various suboptimal approximations have been proposed.

4.3. Algorithmic Solutions

In this chapter, we approach the problem in (4.1) from a set-theoretic perspective, which allows us to devise low-complexity approximation techniques with provable convergence properties without imposing any additional assumptions. To apply the results from Chapter 2, we formulate Problem (4.1) in a real Hilbert space $\left(\mathcal{H} := \mathbb{R}^{2N_{\mathrm{T}}}, \langle \cdot, \cdot \rangle\right)$ equipped with standard Euclidean inner product

$$(\forall \mathbf{x}, \mathbf{y} \in \mathcal{H}) \quad \langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^T \mathbf{x}$$

inducing the Euclidean norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. In this Hilbert space, we can express the ML problem in (4.1) as

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimize}} \ \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \iota_{\mathcal{S}}(\mathbf{x}), \tag{4.2}$$

where and $\iota_{\mathcal{S}} : \mathcal{H} \to \mathbb{R}_+ \cap \{+\infty\}$ is the indicator function of $\mathcal{S}$ given by

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \iota_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in \mathcal{S} \\ +\infty & \text{otherwise.} \end{cases}$$

In this work, we propose a low-complexity approximation technique for the problem in (4.2) by following similar steps as in Section 3.3. In Section 4.3.1, we replace the finite set $\mathcal{S}$ in Problem (4.2) with its convex hull and we propose an APSM to approximate a solution to the relaxed problem. Subsequently, in Section 4.3.2, we propose superiorized version of this algorithm by adding bounded perturbations in each iteration with the intent to steer the iterate towards a solution to the nonconvex ML problem. Similarly to AMP, which alternates between gradient steps and (Gaussian) denoising steps, the proposed algorithm interleaves subgradient projections onto sublevel sets of increasing cost with denoising steps defined by hard slicing or soft thresholding. A convergence proof for the proposed method is provided in Section 4.3.3.

4.3.1. An Adaptive Projected Subgradient Method for MIMO Detection

In principle, a solution to the ML problem can be approximated by iterative techniques that interleave gradient steps for the cost function with projections onto the nonconvex constraint set in (4.1). Such algorithms, based on projected gradient methods or the ADMM, have been discussed in [LYSM17, SDW19, USMC19]. However, owing to the projection onto the nonconvex constellation alphabet, convergence of these algorithms can usually not be guaranteed, or the convergence proof imposes stringent assumptions on the channel matrix [LYSM17]. Instead of directly approaching the nonconvex ML problem in (4.1), some authors [TRL01, TAXH16, WDCS16] have applied iterative algorithms to a relaxed version of Problem (4.1), in which the discrete set $\mathcal{S}$ is replaced with its convex hull

$$\mathcal{B} := \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x}\|_\infty \leq a_{\max}\},$$

where $a_{\max} = \max_{a \in \mathcal{A}} |a|$. In the following, we devise basic algorithm based on an APSM that aims at minimizing the objective of (4.1) over the closed convex set $\mathcal{B}$. According to Theorem 2.3 and 2.2, convergence of the APSM can only be guaranteed if all but finitely many of its cost functions attain the value zero. Hence we cannot directly use the objective function in (4.1) as a cost function for all iterations of the APSM. If the optimal objective value $\rho^\star = \|\mathbf{H}\mathbf{s}^\star - \mathbf{y}\|_2^2$ of Problem (4.1) were known, we could use the APSM to solve the problem

$$\underset{\mathbf{x} \in \mathcal{B}}{\text{minimize}} \ \left(\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 - \rho^\star\right)_+ .$$

This convex minimization problem is equivalent to the convex feasibility problem

$$\text{find } \mathbf{x} \in \mathcal{C}_{\rho^\star} \cap \mathcal{B},$$

where $(\forall \rho \geq 0)\ \mathcal{C}_\rho := \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 \leq \rho\}$ is a sublevel set of the objective function in (4.1), also known as stochastic property set [YSY02]. In the following, we build upon a technique shown in [CSSK18], where the objective is to solve the problem

$$\text{find } \mathbf{x} \in \left(\bigcap_{n \geq n_0} \mathcal{C}_{\rho_n} \right) \cap \mathcal{B},$$

for some $n_0 \in \mathbb{N}$, given a sequence $(\mathcal{C}_{\rho_n})_{n \in \mathbb{N}}$ of stochastic property sets. As in [CSSK18], we define a sequence of continuous convex functions $\Theta_n : \mathcal{H} \to \mathbb{R}_+$ by

$$(\forall n \in \mathbb{N})(\forall \mathbf{x} \in \mathcal{H}) \quad \Theta_n(\mathbf{x}) := \Big(\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 - \rho_n \Big)_+$$

and we use the APSM to minimize asymptotically this sequence of functions over the set $\mathcal{B}$ by iteratively applying the recursion

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} := T_n(\mathbf{x}_n), \quad \mathbf{x}_0 \in \mathcal{H} \tag{4.3}$$

where

$$T_n(\mathbf{x}) := \begin{cases} P_{\mathcal{B}} \left(\mathbf{x} - \mu_n \frac{\Theta_n(\mathbf{x})}{\|\Theta_n'(\mathbf{x})\|^2} \Theta_n'(\mathbf{x}) \right) & \text{if } \Theta_n(\mathbf{x}) > 0 \\ P_{\mathcal{B}}(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{4.4}$$

Here, $(\forall n \in \mathbb{N})\ \Theta_n' : \mathcal{H} \to \mathcal{H}$ defines a subgradient

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \partial\Theta_n(\mathbf{x}) \ni \Theta_n'(\mathbf{x}) = \begin{cases} 2\mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y}) & \text{if } \Theta_n(\mathbf{x}) > 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

of Θ_n at $\mathbf{x}$, and $\mu_n \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$ is a relaxation parameter. If we choose the elements of $(\rho_n)_{n \in \mathbb{N}}$ to increase monotonically in such a way that $(\exists n_0 \in \mathbb{N})\ \rho_n > \rho^\star$, the recursion in (4.3) is guaranteed to converge (see Section 4.3.3). Moreover, if ρ_0 is sufficiently small and $(\rho_n)_{n \in \mathbb{N}}$ increases sufficiently slowly, the final objective value $\lim_{n \to \infty} \|\mathbf{H}\mathbf{x}_n - \mathbf{y}\|^2$ will be close to optimal.

4.3.2. Superiorization

Replacing the discrete constellation alphabet $\mathcal{S}$ with its convex hull $\mathcal{B} \supset \mathcal{S}$ can potentially limit the performance of the algorithm in (4.3), as it ignores available information on the prior distribution of $\mathbf{x}$. Therefore, we use the APSM in (4.3) as a basic algorithm, and we devise a superiorized version by adding small perturbations to its iterates with the intent to reduce slightly the value of some superiorization objective. According to Theorem 2.3,

convergence of the sequence generated by the recursion

$$(\forall n \in \mathbb{N})\ \mathbf{x}_{n+1} := T_n(\mathbf{x}_n + \beta_n \mathbf{v}_n), \quad \mathbf{x}_0 \in \mathcal{H} \tag{4.5}$$

can still be guaranteed if $(\beta_n \mathbf{v}_n)_{n\in\mathbb{N}}$ are bounded perturbations according to Definition 1.3.

Objective functions for superiorization are typically convex. Nevertheless, as in Chapter 3, we will consider nonconvex objective functions in the following. Moreover, in slight deviation from [CDH10] and [Cen15], we use proximal mappings instead of subgradients of the superiorization objective to define the perturbations. This allows for a simple trade-off between the perturbations' magnitude and their contribution to reducing the objective value. In order to incorporate our prior information about the transmit signal, we are interested in superiorization objective functions $f : \mathcal{H} \to \mathbb{R}_+ \cup \{+\infty\}$ that satisfy $f(\mathbf{x}) = 0 \iff \mathbf{x} \in \mathcal{S}$. One example of such a function is the indicator function $f_{\ell_2} := \iota_{\mathcal{S}}$. The proximal mapping associated with f_{ℓ_2} is given by $\operatorname{prox}_{f_{\ell_2}}(\mathbf{x}) = P_{\mathcal{S}}(\mathbf{x})$. Here, $P_{\mathcal{S}}$ denotes a projection onto the set $\mathcal{S}$. Since $\mathcal{S}$ is not convex, this point is not unique for all $\mathbf{x} \in \mathcal{H}$. However, a projection onto $\mathcal{S}$ always exists because the set is closed. Therefore, we can devise perturbations of the form

$$(\forall n \in \mathbb{N}) \quad \mathbf{v}_n^{(\ell_2)} := P_{\mathcal{S}}(\mathbf{x}_n) - \mathbf{x}_n. \tag{4.6}$$

As the primary objective of MIMO detection is to reduce the symbol error ratio (SER), one could instead use a superiorization objective that penalizes the number of coefficients of the estimate $\hat{\mathbf{x}}$ which lie outside the the set of valid constellation points, i.e.,

$$\sum_{k:\hat{x}_k \notin \mathcal{A}} 1 = \|\hat{\mathbf{x}} - P_{\mathcal{S}}(\hat{\mathbf{x}})\|_0, \tag{4.7}$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm. Borrowing a well-known result from compressed sensing [Don06], we replace the ℓ_0 pseudo-norm in (4.7) with the ℓ_1-norm to define an alternative superiorization objective $(\forall \mathbf{x} \in \mathcal{H})\ f_{\ell_1}(\mathbf{x}) := \|\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\|_1$. Note that f_{ℓ_1} is still nonconvex due to the projection onto the nonconvex set $\mathcal{S}$. Nevertheless, we can derive a proximal mapping associated with τf_{ℓ_1} as follows:

Proposition 4.1. *Let $\mathcal{A} \subset \mathbb{R}$ be a discrete set of equidistant real-valued constellation points and let $\mathcal{S} = \mathcal{A}^{2N_{\mathrm{T}}} \subset \mathcal{H}$. Then a proximal mapping associated with τf_{ℓ_1} is given by*

$$\operatorname{prox}_{\tau f_{\ell_1}}(\mathbf{x}) = \phi_\tau\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right) + P_{\mathcal{S}}(\mathbf{x}),$$

where $(\forall \tau \geq 0)$ $\phi_\tau : \mathcal{H} \to \mathcal{H}$ *is the soft-thresholding operator*

$$(\forall \mathbf{x} \in \mathcal{H})(\forall i \in \mathcal{I}_{\mathrm{T}}) \quad \phi_\tau(\mathbf{x})|_i := \operatorname{sgn}(x_i)(|x_i| - \tau)_+.$$

Proof: *See Appendix B.*

Using Proposition 4.1, we can define perturbations of the form

$$(\forall n \in \mathbb{N}) \quad \mathbf{v}_n^{(\ell_1)} := \operatorname{prox}_{\tau_n f_{\ell_1}}(\mathbf{x}_n) - \mathbf{x}_n. \tag{4.8}$$

4.3.3. Convergence of the Proposed Algorithms

In the following, we investigate the convergence of the proposed algorithms. According to Theorem 2.3 and Remark 2.2, the sequence produced by superiorized APSM in (4.5) converges (strongly) to a point $\mathbf{x}^\star \in \mathcal{B}$, given that the perturbations are bounded and that

(C1) $(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)$ $\Theta_n^\star = 0$, i.e., $\Omega_n := \{\mathbf{x} \in \mathcal{B} \mid \Theta_n(\mathbf{x}) = \Theta_n^\star\} = \mathcal{B} \cap \mathcal{C}_{\rho_n}$, and $\Omega := \bigcap_{n \geq n_0} \Omega_n \neq \emptyset$

(C2) $(\exists \mathbf{z} \in \Omega)(\exists \eta > 0)$ $\{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x} - \mathbf{z}\| \leq \eta\} \subset \Omega$, i.e., the set Ω has an interior point.

Moreover, the point $\mathbf{x}^\star$ minimizes all but finitely many functions of the sequence $(\Theta_n)_{n\in\mathbb{N}}$ if

(C3) $(\Theta_n'(\mathbf{x}_n + \beta_n \mathbf{y}_n))_{n\in\mathbb{N}}$ is bounded

(C4) there exists a bounded sequence $(\Theta_n'(\mathbf{x}^\star))_{n\in\mathbb{N}}$, where $(\forall n \in \mathbb{N})$ $\Theta_n'(\mathbf{x}^\star) \in \partial\Theta_n(\mathbf{x}^\star)$.

The objective of the remainder of this subsection is to show that these conditions are satisfied. We begin by showing that the proposed perturbations are bounded.

Proposition 4.2. *The proposed perturbations in* (4.6) *and* (4.8) *are bounded.*

Proof. *Since* $\mathcal{B}$ *is compact, we can define* $c := \max_{\mathbf{x}\in\mathcal{B}} \|\mathbf{x}\|$. *By* (4.4) *and the definition of a projection,* $(\forall n \in \mathbb{N})$ $\mathbf{x}_n \in \mathcal{B}$ *and* $(\forall \mathbf{x} \in \mathcal{H}) P_{\mathcal{S}}(\mathbf{x}) \in \mathcal{S} \subset \mathcal{B}$. *Consequently, we have*

$$\|\mathbf{v}_n^{(\ell_2)}\| = \|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \leq \|P_{\mathcal{S}}(\mathbf{x})\| + \|\mathbf{x}_n\| \leq 2c$$

and

$$\begin{aligned} \|\mathbf{v}_n^{(\ell_1)}\| &= \|\phi_\tau\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right) + P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \\ &\leq \|\phi_\tau\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right)\| + \|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \\ &\leq 2\|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \leq 4c, \end{aligned}$$

which concludes the proof. □

Finally, we apply Theorem 2.3 and Proposition 4.2 to prove the convergence of the proposed method.

Proposition 4.3. *Let* $(\rho_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}_+$ *satisfying* $(\exists n_0 \in \mathbb{N})$ $(\exists \eta > 0)$ $(\forall n \geq n_0)$ $\rho_n \geq \rho^\star + \eta$. *Then the proposed algorithm in* (4.5) *with perturbations according to* (4.6) *or* (4.8) *and* $(\beta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ *is guaranteed to converge to a point* $\mathbf{x}^\star \in \mathcal{B}$ *minimizing all but finitely many functions of the sequence* $(\Theta_n)_{n\in\mathbb{N}}$.

Proof. *In light of Theorem 2.3, it remains to show that the conditions (C1)-(C4) above are satisfied. Let* $\mathbf{s}^\star$ *denote a solution to Problem* (4.1).

(C1) By assumption, $(\forall n \geq n_0)$ $\rho_n \geq \rho^\star = \|\mathbf{H}\mathbf{s}^\star - \mathbf{y}\|^2$, *whereby* $0 \leq \Theta_n^\star \leq \left(\|\mathbf{H}\mathbf{s}^\star - \mathbf{y}\|^2 - \rho_n\right)_+ \leq 0$. *Moreover,* $(\forall n \geq n_0)$ $\mathbf{s}^\star \in \mathcal{C}_{\rho_n} \cap \mathcal{B} = \Omega_n$, *i.e.,* $\Omega \neq \emptyset$.

(C2) Define $\mathcal{E} := \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{s}^\star - \mathbf{x}\| \leq \varepsilon\}$ *with some positive* $\varepsilon \leq \frac{\sqrt{\rho^\star+\eta}-\sqrt{\rho^\star}}{\|\mathbf{H}\|_2}$. *All* $\mathbf{u} \in \mathcal{H}$ *with* $\|\mathbf{u}\| \leq 1$ *satisfy*

$$\begin{aligned}\|\mathbf{H}(\mathbf{s}^\star + \varepsilon\mathbf{u}) + \mathbf{y}\|^2 &= \rho^\star + 2\varepsilon\langle \mathbf{H}\mathbf{s}^\star - \mathbf{y}, \mathbf{H}\mathbf{u}\rangle + \varepsilon^2\|\mathbf{H}\mathbf{u}\|^2 \\ &\leq \rho^\star + 2\varepsilon\sqrt{\rho^\star}\|\mathbf{H}\|_2 + \varepsilon^2\|\mathbf{H}\|_2^2 \\ &\leq \rho^\star + \eta.\end{aligned}$$

Therefore, by the premise of this proposition, $(\forall n \geq n_0)$ $(\forall \mathbf{x} \in \mathcal{E})$ $\Theta_n(\mathbf{x}) = 0$, *i.e.,* $\mathcal{E} \subset \mathcal{C}_{\rho_n}$. *Now, we define a set with nonempty interior by* $\mathcal{Q} := \{\mathbf{x} \in \mathcal{H} \mid (\forall i \in \mathcal{I}_\mathrm{T})\ s_l \leq x_i \leq s_u\}$, *where* $(\forall i \in \mathcal{I}_\mathrm{T})$

$$\tilde{s}_i := \operatorname{sgn}(s_i^\star) \cdot \left(|s_i^\star| - \frac{\varepsilon}{\sqrt{2N_\mathrm{T}}}\right),$$

$s_l := \min(\tilde{s}_i, s_i)$, *and* $s_u := \max(\tilde{s}_i, s_i)$. *Note that* $(\forall n \geq n_0)$ $\mathcal{Q} \subset \mathcal{E} \subset \mathcal{C}_{\rho_n}$. *Moreover,* $\mathcal{Q} \subset \mathcal{B}$ *for sufficiently small* $\varepsilon > 0$, *so it holds that* $\mathcal{Q} \subset \Omega$.

(C3) Let $(\forall n \in \mathbb{N})$ $\mathbf{z}_n := \mathbf{x}_n + \beta_n\mathbf{y}_n$. *Since* $(\forall n \in \mathbb{N})$ $\Theta_n(\mathbf{x}) = 0 \implies \Theta_n'(\mathbf{x}) = \mathbf{0}$, *it is sufficient to consider the case* $\Theta_n(\mathbf{x}) > 0$. *In this case, we have that* $\Theta_n'(\mathbf{x}) = 2\mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y})$, *so* $(\forall \mathbf{x}_n \in \mathcal{B})$

$$\begin{aligned}\|\Theta_n'(\mathbf{z}_n)\| &\overset{(i)}{\leq} 2\|\mathbf{H}^T\mathbf{H}\mathbf{z}_n\| + 2\|\mathbf{H}^T\mathbf{y}\| \\ &\overset{(ii)}{\leq} 2\|\mathbf{H}^T\mathbf{H}\|_2 \cdot \|\mathbf{z}_n\| + 2\|\mathbf{H}^T\mathbf{y}\|\end{aligned}$$

$$\overset{(iii)}{\leq} 2\|\mathbf{H}^T\mathbf{H}\|_2 \cdot (\|\mathbf{x}_n\| + \beta_n\|\mathbf{y}_n\|) + 2\|\mathbf{H}^T\mathbf{y}\|$$
$$\overset{(iv)}{\leq} 2\left(c + r\max_{n\in\mathbb{N}}\beta_n\right)\|\mathbf{H}^T\mathbf{H}\|_2 + 2\|\mathbf{H}^T\mathbf{y}\|.$$

Here, (i) and (iii) follow from the triangle inequality, (ii) follows from the definition of an operator norm, and (iv) follows from the definition of the constant c in Proposition 4.2 and the fact that $(\exists r \in \mathbb{R})\ (\forall n \in \mathbb{N})\ \|\mathbf{y}_n\| \leq r$. Consequently, the sequence of subgradients $(\Theta_n'(\mathbf{x}_n + \beta_n\mathbf{y}_n))_{n\in\mathbb{N}}$ is bounded.

(C4) Since $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is a convergent sequence in the compact set $\mathcal{B}$, its limit $\mathbf{x}^\star$ also belongs to $\mathcal{B}$. Therefore, we can apply the same argument as above. □

4.3.4. Summary of the Proposed Algorithms

The proposed iterative MIMO dectector based on superiorized APSM is summarized in Algorithm 2 below.

Algorithm 2 Superiorized APSM for MIMO Detection

Parameters: $(\rho_n \geq 0)_{n\in\mathbb{N}}$, $(\mu_n \in (0,2))_{n\in\mathbb{N}}$, $(\beta_n \geq 0)_{n\in\mathbb{N}}$, $(\tau_n \geq 0)_{n\in\mathbb{N}}$
Input: $\mathbf{H} \in \mathbb{R}^{2N_\mathrm{R}\times 2N_\mathrm{T}}, \mathbf{y} \in \mathbb{R}^{2N_\mathrm{R}}$
Output: $\hat{\mathbf{x}} \in \mathbb{R}^{2N_\mathrm{T}}$
Initialization: Choose arbitrary $\mathbf{x}_0 \in \mathcal{H}$
for $n = 0, \ldots, n_{\max} - 1$ **do**
  $\mathbf{v}_n = \mathrm{prox}_{\tau_n f_{\ell_j}}(\mathbf{x}_n) - \mathbf{x}_n$, $j \in \{1,2\}$ ▷ perturbation according to (4.6) or (4.8)
  $\mathbf{z}_n = \mathbf{x}_n + \beta_n\mathbf{v}_n$
  $\Theta_n(\mathbf{z}_n) = (\|\mathbf{H}\mathbf{z}_n - \mathbf{y}\|^2 - \rho_n)_+$
  $\Theta_n'(\mathbf{z}_n) = 2\mathbf{H}^T(\mathbf{H}\mathbf{z}_n - \mathbf{y})$
  $\mathbf{x}_{n+1} = \begin{cases} P_{\mathcal{B}}\left(\mathbf{z}_n - \mu_n \frac{\Theta_n(\mathbf{z}_n)}{\|\Theta_n'(\mathbf{z}_n)\|^2}\Theta_n'(\mathbf{z}_n)\right) & \text{if } \Theta_n(\mathbf{z}_n) > 0 \\ P_{\mathcal{B}}(\mathbf{z}_n) & \text{otherwise} \end{cases}$
end for
return $\hat{\mathbf{x}} = \mathbf{x}_{n+1}$

4.3.5. Deep-Unfolded APSM for MIMO Detection

The superiorized APSM summarized in Algorithm 2 comprises four sequences of design parameters, which have a considerable impact on its performance. Since no theoretical criteria on the optimal choice of these parameters are available, we apply deep unfolding to learn suitable parameters based on simulated data. To do so, we create independent training and test datasets using the same system parameters but different random user

equipment (UE) placements. As the weights are trained in an offline manner, they need to generalize across arbitrary channel realizations. In order to achieve this, we create batches of size b of (single-subcarrier) detection problem tuples $(\mathbf{H}_i, \mathbf{x}_i, \mathbf{y}_i)$ corresponding to different random UE locations, where $i = 1, \ldots, b$. Once the model has been trained, inference is carried out using the dataset from [KAHF20], where one problem realization corresponds to a unique UE arrangement with an arbitrary number of subcarriers.

Owing to the vanishing gradient problem, training the unfolded algorithm in a single shot is difficult. Many existing works (e.g., [KAHF20]) mitigate the vanishing gradient problem by averaging the loss (e.g., the mean squared error) over all layers

$$\Theta(\{\hat{\mathbf{x}}_l\}, \mathbf{x}) = \frac{1}{2N_{\mathrm{T}}L} \sum_{l=1}^{L} \|\hat{\mathbf{x}}_l - \mathbf{x}\|_2^2,$$

where L denotes the number of layers in the network. However, this approach can result in suboptimal choices of the weights, as the performance of the trained network only depends on the loss in the output layer. Therefore, instead of averaging the loss over all layers, we adopt the incremental training technique proposed in [ITW19]. This training method consists of multiple *generations*. In the lth generation, an unfolded model with l layers is trained based on the loss function

$$\Theta(\{\hat{\mathbf{x}}_l\}, \mathbf{x}) = \frac{1}{2N_{\mathrm{T}}} \|\hat{\mathbf{x}}_l - \mathbf{x}\|_2^2.$$

At the beginning of the subsequent generation $l + 1$, the next layer is added to the model and training is performed based on the loss in the $(l + 1)$th layer. To ensure that the learned parameters are nonnegative, we learn the quantities $\log(\rho_l)$, $\log(\mu_l)$, $\log(\beta_l)$, and $\log(\tau_l)$, instead of the actual parameters ρ_l, μ_l, β_l, and τ_l for $l \in \{1, \ldots, L\}$. The actual parameters are computed by applying the exponential function to the learned weights in the forward pass. Moreover, to reduce the number of iterations required by APSM (and hence the number of layers required for its deep-unfolded version), we initialize the algorithm with the LMMSE estimate $\hat{\mathbf{x}}_0 = (\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}$.

4.4. Numerical Results

In this section, we compare the performance of the following algorithms:

- The APSM basic algorithm in (4.3) (`APSM`)
- The superiorized APSM in (4.5) with perturbations according to (4.6) (`APSM-L2`)
- The superiorized APSM in (4.5) with perturbations according to (4.8) (`APSM-L1`)

- The AMP-based MIMO detector (IO-LAMA) proposed in [JGMS15] (`AMP`)
- The detector based on OAMP [MP17] (`OAMP`)
- The LMMSE estimate given by $\mathbf{x}_{\texttt{LMMSE}} = (\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}$ (`LMMSE`).

We consider a system with $N_{\mathrm{T}} = 16$ single antenna transmitters and $N_{\mathrm{R}} = 64$ receive antennas and 16-QAM constellation. As in [KAHF20], we assume perfect power allocation, i.e., we normalize the columns of the channel matrix $\mathbf{H}$ to unit Euclidean norm. For the APSM algorithms, we set $(\forall n \in \mathbb{N})$ $\rho_n = 5 \cdot 10^{-5} \cdot 1.06^n$ and $\mu_n = 0.7$. The perturbations of `APSM-L2` are scaled using the sequence $(\beta_n = b^n)_{n\in\mathbb{N}}$ with $b = 0.9$. For `APSM-L1`, we set $(\forall n \in \mathbb{N})$ $\tau_n = 0.005$ and $\beta_n = 0.9999$.

4.4.1. Performance of the Untrained Detectors

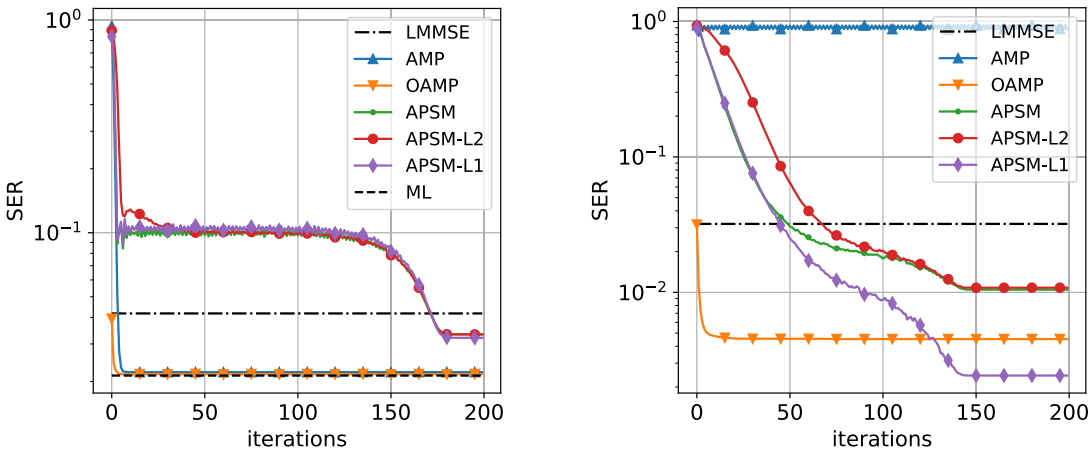


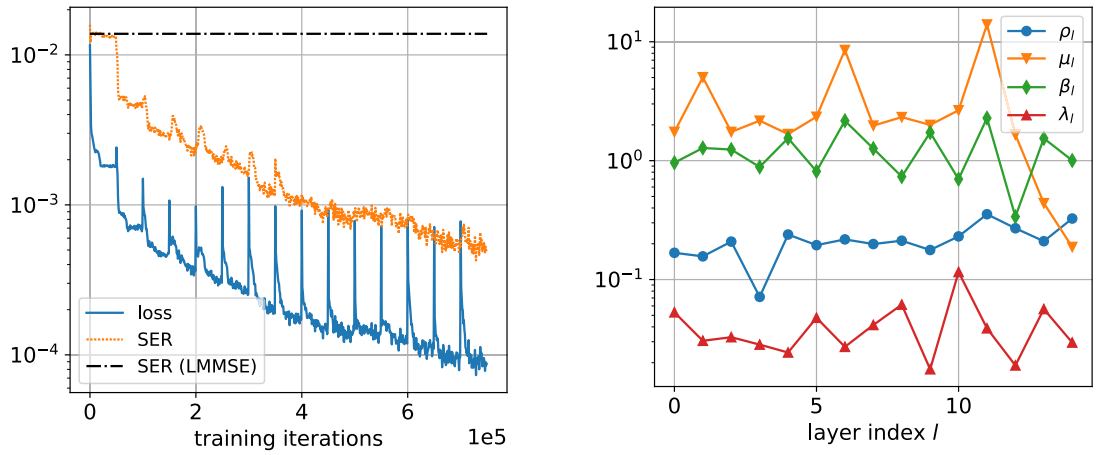
Figure 4.1.: SER as a function of the number of iterations averaged over (a) 10000 realizations of i.i.d. Gaussian channels with 9 dB SNR and (b) 10000 3GPP channels with 18 dB SNR.

Figure 4.1(a) shows the SER throughout the iterations, averaged over 10000 i.i.d. Gaussian channel matrices with 9 dB SNR. It can be seen that both `AMP` and `OAMP` achieve maximum likelihood performance within about 10 iterations. The proposed methods do not achieve ML performance. However, they still outperform `LMMSE`. Figure 4.1(b) shows the he SER as a function of the number of iterations averaged over 10000 single-subcarrier 3rd Generation Partnership Project (3GPP) MIMO channels with 18 dB SNR. The channels are generated with the Quasi Deterministic Radio Channel Generator (QuaDRiGa) [Fra19]

using the code provided with [KAHF20]. The simulation adopts the `3GPP_3D_UMa_NLOS` scenario for a base station (BS) equipped with a 4×8 dual polarized antenna array ($N_{\mathrm{R}} = 64$) according to [3GP17] and $N_{\mathrm{T}} = 16$ UEs equipped with omnidirectional antennas.

While all APSM-type algorithms achieve a SER below `LMMSE`, `AMP` fails to reduce the SER throughout the iterations. Superiorization based on the indicator function $f_{\ell_2} = \iota_{\mathcal{S}}$ (`APSM-L2`) does not improve the performance compared to the unperturbed basic algorithm (`APSM`). By contrast, the SER achieved by `APSM-L1` is about an order of magnitude below SER achieved by the unperturbed basic algorithm `APSM`, even outperforming the more complex `OAMP` detector.

4.4.2. Performance of a Deep-Unfolded APSM Detector



(a) (b)

Figure 4.2.: (a) Loss and SER of the `APSMNet` model trained at 18 dB SNR as a function of the training iterations. (b) Weights of the resulting trained `APSM-L1` model with $L = 15$ layers.

In the following, we train a deep-unfolded instance of `APSM-L1` with $L = 15$ layers using a batch size of 50 and incremental training with 50000 training iterations per generation. As described in Subsection 4.3.5, we use the LMMSE estimate as the input to the first layer. In the style of the unfolded MIMO detectors DetNet [SDW17, SDW19], ADMM-Net [USMC19], MMNet [KAHF20], or FS-Net [NL20], we refer to this method as `APSMNet` in the following. Training is performed using Adam optimizer [KB14] with a learning rate of 0.0005. Figure 4.2(a) shows the loss value and SER during training as a function of the

training iterations for an exemplary training run (the respective values are averaged over periods of 1000 successive training iterations to smooth the curve). It can be seen that the loss increases whenever a new layer is added to the model. This is due to the fact that the weights in the newly added layer have not been optimized yet. During the course of a generation of incremental training, the loss decreases and typically settles at a lower level than in the previous generation. For some generations, the the loss is still decreasing when the next layer is added. This indicates that the weights have not yet converged to their final values. Consequently, we expect that the performance of the trained model could be further improved by increasing the number of training iterations.

Figure 4.2(b) shows the weights learned by the same exemplary incremental training procedure. Interestingly, the learned parameters exhibit an oscillating behavior throughout the layers of the network. This effect has also been observed in [ITW19]. A theoretical explanation for this oscillating behavior, which relates the learned step size parameters to Chebyshev steps, was recently provided in [TW20a], [TW20b].

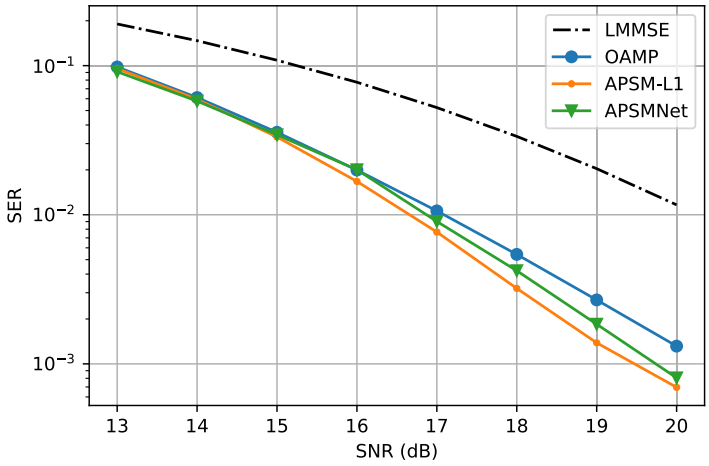


Figure 4.3.: SER for different methods averaged over 10000 3GPP channels as a function of the SNR.

Figure 4.3 shows a performance comparison between untrained APSM with $L = 200$ iterations (`APSM-L1`), trained APSM with $L = 15$ layers (`APSMNet`), OAMP [MP17] with $L = 30$ iterations (`OAMP`), and LMMSE. The `APSMNet` instances have been trained using an independent training dataset. For low SNRs, `APSM-L1` and `APSMNet` achieve the same SER as `OAMP`. In the high SNR regime, both `APSM-L1` and `APSMNet` outperform `OAMP`. This is remarkable, as `APSMNet` uses only one matrix inverse to initialize $\hat{\mathbf{x}}_0$, whereas OAMP needs to compute matrix inverses in each iteration, resulting in a considerably higher

computational cost.

4.5. Conclusion

In this chapter, we proposed iterative MIMO detectors with convergence guarantees based on a superiorized adaptive projected subgradient method. Unlike IO-LAMA, the proposed methods are not restricted to i.i.d. Gaussian channels. Simulations show that the proposed methods can outperform OAMP on realistic channels. Furthermore, we showed that even lower symbol error ratios can be achieved with deep-unfolded variants of the proposed algorithms. In contrast to OAMP, the APSM-based detectors do not require matrix inverses, so they have a per-iteration complexity similar to IO-LAMA. The proposed deep-unfolded detectors require only one matrix inverse, and they achieve a lower SER than OAMP within only 15 iterations.

5. Online Channel Estimation for Hybrid Beamforming Architectures

In this chapter, we propose online channel estimation algorithms based on the APSM in Theorem 2.3. The results are based on our initial work in [6], where we proposed algorithms for online channel estimation and tracking for hybrid beamforming architectures in a narrowband setting. In the following, we extend the approach in [6] to a wideband setting, and we additionally take into account random variations in the delay and phase shift of subsequent channel realizations. Some passages in this chapter are borrowed from [6].

5.1. Introduction

Massive MIMO systems are envisioned to increase greatly the rates experienced by users in future mobile networks. By exploiting channel state information, multi-antenna transceivers can precode and combine the transmit and receive signals, in order to improve the performance of communication systems. In fully digital beamforming architectures, every antenna is connected to a dedicated radio frequency (RF) chain. Consequently, both power consumption and hardware costs of these fully digital architectures grow rapidly as the number of antenna elements increases, potentially prohibiting their deployment especially in systems with large-scale antenna arrays. Hybrid analog-/digital beamforming architectures use an analog network of digitally controlled phase shifters or switches to map the transmit and receive signals for all antenna elements to a lower number of RF chains. These architectures are a promising means of reducing power consumption and hardware costs in large multi-antenna transceivers. However, compared to fully digital beamforming architectures, hybrid architectures render channel estimation more difficult, because only a subspace of the channel can be observed at any time instant. For this reason, multiple measurements (pilot signals in subsequent time slots) are required to obtain an accurate estimate of the channel. Wireless channels can vary quickly in high mobility scenarios, so the time required to collect sufficiently many samples can exceed the coherence time of the channel even if compressive sensing techniques are applied. This calls for channel estimation schemes that exploit temporal correlations between channel samples.

A major part of the literature on channel estimation for hybrid beamforming considers the estimation of static channels. The schemes proposed in [AEALH14, BLW15, MRRA+15, LGL16, VAHP17] estimate the channel based on multiple training symbols that are transmitted sequentially. In these papers, the authors (implicitly or explicitly) impose the block fading model, which assumes the channel to remain constant for a certain number of observations [MH99]. The authors of [YWJG18, HWJL18a, CYX+19] propose channel estimation schemes for hybrid architectures with electromagnetic (EM) lens antenna arrays. The methods proposed in these works yield high quality channel estimates based on a single training symbol. However, their complexity may be prohibitive in mobile scenarios with short channel coherence time. To the best of our knowledge, computationally simple online methods for estimating and tracking fast time-varying channels with hybrid architectures have not been proposed so far.

In [6], we proposed online channel estimation algorithms for hybrid analog-digital beamforming architechtures that estimate and track the channel by exploiting side-information such as sparsity in the angular domain. In contrast to previous works, the proposed methods do not rely on the conventional block fading assumption, but rather they allow for small variations in subsequent channel realizations. Moreover, we proposed an adaptive data-driven analog beamforming scheme that reduces the channel estimation error compared to a random analog beamforming policy.

This chapter extends the results in [6], which are restricted to narrowband channels having only a single subcarrier, to wideband channels with multiple subcarriers. In addition to that, we augment the system model by introducing random delay and phase variations to subsequent channel observations, which model the effects of hardware inaccuracies and switching processes at the UE. To compensate for these variations, we propose a mechanism in Subsection 5.3.5, which estimates the delay and phase differences between subsequent channel realizations. We derive theoretical lower bounds on the phase and delay estimation error, and we assess the the performance of the proposed phase and delay estimation scheme using an idealized signal model. The proposed phase and delay compensation scheme is integrated with two online channel estimation algorithms. The first algorithm, described in Subsection 5.3.2, uses a variant of the APSM in (2.18) without adding perturbations. The second algorithm, which is introduced in Subsection 5.3.3, is a heuristic that is motivated by a superiorized variant of the APSM in (2.18), in which we heuristically set $(\forall n \in \mathbb{N})\ \beta_n = 1$.

Simulations with a spatially consistent quasi-deterministic channel model are provided for a single moving UE. We show that proposed online channel estimation scheme can outperform batch optimization techniques in terms of both accuracy and computational cost, while permitting an adaptation of the analog combiners at runtime. Moreover, in

contrast to the baseline, the proposed method can be used to estimate wideband channels with random delay and phase variations. The simulations reveal that a significant performance improvement can be achieved by jointly estimating the channels on all subcarriers, compared to a parallel implementation of narrowband channel estimators.

5.2. Problem Statement

In this chapter, we consider the uplink in a single-cell wireless system OFDM with N_{F} subcarriers indexed by $k \in \mathcal{I}_{\mathrm{sc}} := \{1, \ldots, N_{\mathrm{F}}\}$. The network comprises a BS equippend with N_{R} antenna elements, which are connected to $L_{\mathrm{R}} < N_{\mathrm{R}}$ RF chains, and a single UE equipped with N_{T} antenna elements, each of them having a separate RF chain. Transmission takes place in regular time intervals of duration τ, which are indexed by $n \in \mathbb{N}$. The signal transmitted on the kth subcarrier during the nth time interval is denoted by $\mathbf{x}_{n,k} \in \mathbb{C}^{N_{\mathrm{T}}}$, so that the received signal after analog-to-digital conversion is given by

$$(\forall n \in \mathbb{N})(\forall k \in \mathcal{I}_{\mathrm{sc}}) \quad \mathbf{y}_{n,k} = \mathbf{B}_n \mathbf{H}_{n,k} \mathbf{x}_{n,k} + \mathbf{n}_{n,k} = (\mathbf{x}_{n,k}^T \otimes \mathbf{B}_n)\mathbf{h}_{n,k} + \mathbf{n}_{n,k}. \tag{5.1}$$

Here, $\mathbf{B}_n \in \mathbb{C}^{L_{\mathrm{R}} \times N_{\mathrm{R}}}$ represents the analog combiner, $\mathbf{h}_{n,k} := \mathrm{vec}(\mathbf{H}_{n,k})$ is the vectorized channel matrix, and $\mathbf{n}_{n,k} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{L_{\mathrm{R}}})$ is a vector of i.i.d. complex Gaussian noise samples. The entire wideband channel at time interval n can be represented by the matrix $\mathbf{H}_n = [\mathbf{h}_{n,1}, \ldots, \mathbf{h}_{n,N_{\mathrm{F}}}] \in \mathbb{C}^{N_{\mathrm{R}} N_{\mathrm{T}} \times N_{\mathrm{F}}}$. To simplify the notation, we define $(\forall n \in \mathbb{N})$ $\mathbf{X}_n = [\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,N_{\mathrm{F}}}] \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{F}}}$ and $\mathbf{Y}_n = [\mathbf{y}_{n,1}, \ldots, \mathbf{y}_{n,N_{\mathrm{F}}}] \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{F}}}$. The objective of the methods proposed in this chapter is to estimate and track the sequence $(\mathbf{H}_n)_{n \in \mathbb{N}}$ as the UE moves along a specified trajectory. The channel is assumed to be spatially (and hence temporally) consistent, i.e., the variations $\|\mathbf{H}_{n+1} - \mathbf{H}_n\|_F$ between subsequent channel realizations are assumed to be small if the time interval duration τ is sufficiently short. Therefore we do not use the conventional block fading assumption.

In practical systems, hardware imperfections can introduce varying phase shifts to signals transmitted in subsequent time intervals. Moreover, technical processes such as temporarily switching off the transceiver of the UE can add random offsets to the signal delay. As the digital baseband signal is restricted to a subspace of the channel by the analog combining hardware, multiple subsequent channel observations are required to reconstruct the channel coefficients. Therefore, phase shifts and delays between two subsequent channel observations can cause severe reconstruction errors. These effects can be taken into account by introducing a channel model in which delay and phase of the receive signal vary randomly between subsequent channel observations. To this end, consider the (spatially/temporally consistent) sequence $(\mathbf{H}_n)_{n \in \mathbb{N}}$ of wideband channels introduced

above. Given random delay[1] $\tau_n \sim \mathcal{U}(0,1)$ and phase shift $\varphi_n \sim \mathcal{U}(0,2\pi)$, the channel $\tilde{\mathbf{H}}_n = [\tilde{\mathbf{h}}_{n,1}, \ldots, \tilde{\mathbf{h}}_{n,N_\mathrm{F}}]$ at time interval $n \in \mathbb{N}$, including switching processes and hardware imperfections at the UE, can be modeled as

$$(\forall k \in \mathcal{I}_\mathrm{sc}) \quad \tilde{\mathbf{h}}_{n,k} = e^{j\varphi_n} e^{-j2\pi(k-1)\tau_n} \mathbf{h}_{n,k}, \tag{5.2}$$

which results in the receive signal

$$(\forall k \in \mathcal{I}_\mathrm{sc}) \quad \tilde{\mathbf{y}}_{n,k} = e^{j\varphi_n} e^{-j2\pi(k-1)\tau_n} \mathbf{B}_n \mathbf{H}_{n,k} \mathbf{x}_{n,k} + \mathbf{n}_{n,k}. \tag{5.3}$$

5.3. Algorithmic Solutions

The goal of the methods presented in the following is to estimate and track the sequence $(\mathbf{H}_n)_{n\in\mathbb{N}}$ of channel matrices based on the known pilots $(\mathbf{X}_n)_{n\in\mathbb{N}}$, analog combiners $(\mathbf{B}_n)_{n\in\mathbb{N}}$, and received signals $(\mathbf{Y}_n)_{n\in\mathbb{N}}$.

5.3.1. Channel Estimation using Existing Batch Methods

In this subsection, we restrict our attention to the narrowband setting, in which the system model is described by (5.1) with $N_\mathrm{F} = 1$. Dropping the subcarrier index k for notational convenience, we obtain

$$(\forall n \in \mathbb{N}) \quad \mathbf{y}_n = (\mathbf{x}_n^T \otimes \mathbf{B}_n)\mathbf{h}_n + \mathbf{n}_n, \tag{5.4}$$

where $\mathbf{y}_n \in \mathbb{C}^{N_\mathrm{R}}$, $\mathbf{n}_n \in \mathbb{C}^{N_\mathrm{R}}$, $\mathbf{x}_n \in \mathbb{C}^{N_\mathrm{T}}$, and $\mathbf{h}_n \in \mathbb{C}^{N_\mathrm{R}N_\mathrm{T}}$. Many techniques in the literature (e.g., [AEALH14, BLW15, MRRA+15, LGL16, VAHP17]) work by collecting batches of the signals $(\mathbf{B}_n, \mathbf{x}_n, \mathbf{y}_n)$ and by xploiting sparsity of the angular domain representation of the channel. For example, suppose that the vectorized channel $\mathbf{h}_n$ has a sparse representation $\boldsymbol{\theta}_n$ such that $\mathbf{h}_n = \boldsymbol{\Psi}\boldsymbol{\theta}_n$, where $\boldsymbol{\Psi} \in \mathbb{C}^{N_\mathrm{T}N_\mathrm{R} \times N_\mathrm{T}N_\mathrm{R}}$ is nonsingular. Under the block fading model, a good estimate of $\mathbf{h}_n$ can be obtained by solving the least absolute shrinkage and selection operator (LASSO) [Tib96] problem

$$\hat{\mathbf{h}}_n \in \underset{\mathbf{h}\in\mathbb{C}^{N_\mathrm{T}N_\mathrm{R}}}{\arg\min} \frac{1}{N_\mathrm{b}} \sum_{i=N_\mathrm{b}\left\lfloor \frac{n}{N_\mathrm{b}} \right\rfloor}^{(N_\mathrm{b}+1)\left\lfloor \frac{n}{N_\mathrm{b}} \right\rfloor - 1} \left\| \left(\mathbf{x}_i^T \otimes \mathbf{B}_i\right) \mathbf{h} - \mathbf{y}_i \right\|_2^2 + \lambda_\mathrm{b} \left\| \boldsymbol{\Psi}^{-1}\mathbf{h} \right\|_1, \tag{5.5}$$

[1] Without loss of generality, we set the OFDM symbol duration (without cyclic prefix) to 1 for notational convenience. Note that this does not limit the choice of τ in our simulations.

where λ_b is a nonnegative regularization parameter, and N_b is the batch size. The resulting channel estimate is used to perform detection for the entire duration $N_\mathrm{b}\Delta t$ of a batch. Aside from the computational complexity of solving Problem (5.5), which can be avoided by matching pursuit algorithms (see, e.g., [MRRA+15]), this formulation has the additional drawback that the channel estimate is not available before the last time interval in the respective batch. As a result, the analog combiner cannot be adapted immediately after a new channel realization is observed. Furthermore, if the channel estimate is used for precoding in the downlink, it might be outdated as soon as it becomes available.

In principle, this method can be extended to the wideband setting in (5.1) by estimating each of the N_F subchannels independently. However, this approach does not exploit correlations in the frequency domain (or more precisely, sparsity of the channel impulse response). Although formulating Problem (5.5) jointly over all subcarriers is straightforward, the resulting number of $N_\mathrm{R}N_\mathrm{T}N_\mathrm{F}$ complex variables likely prohibits the use of general purpose solvers, even for offline simulations [KKL+07]. Moreover, batch methods are not suited for the generalized system model in (5.3), as the randomly varying delay and phase shift violate the underlying block fading assumption.

5.3.2. Online Algorithm for Channel Estimation and Tracking

To overcome the limitations of batch methods, we propose an online channel estimation and tracking technique with low complexity in the following. Generalizing the narrowband formulation in [6] to the wideband setting in (5.1), we pose the channel estimation problem in a *real* Hilbert space $\left(\mathcal{H} = \mathbb{C}^{N_\mathrm{R}N_\mathrm{T}N_\mathrm{F}}, \langle\cdot,\cdot\rangle\right)$ with inner product

$$(\forall \mathbf{x}, \mathbf{y} \in \mathcal{H}) \quad \langle \mathbf{x}, \mathbf{y} \rangle := \mathrm{Re}\{\mathbf{y}^H \mathbf{x}\}, \tag{5.6}$$

inducing the Euclidean norm $(\forall \mathbf{x} \in \mathcal{H})\ \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \|\mathbf{x}\|_2$. The following proposition shows that $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is in fact a real Hilbert space.

Proposition 5.1. *The Hilbert space $\mathcal{H} = (\mathbb{C}^N, \langle\cdot,\cdot\rangle)$ equipped with the inner product $(\forall \mathbf{x}, \mathbf{y} \in \mathcal{H})\ \langle \mathbf{x}, \mathbf{y} \rangle := \mathrm{Re}\{\mathbf{x}^H \mathbf{y}\}$ is isometrically isomorphic to a real Hilbert space with standard Euclidean inner product.*

Proof. *Consider the bijective linear mapping $\xi : \mathbb{C}^N \to \mathbb{R}^{2N}$ defined by $(\forall \mathbf{x} \in \mathbb{C}^N)$*

$$\xi(\mathbf{x}) := \left[\mathrm{Re}\{\mathbf{x}\}^T, \mathrm{Im}\{\mathbf{x}\}^T\right]^T. \tag{5.7}$$

Then $(\forall \mathbf{x} \in \mathbb{C}^N)$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathrm{Re}\{\mathbf{x}^H \mathbf{y}\} = \mathrm{Re}\{\mathbf{x}\}^T \mathrm{Re}\{\mathbf{y}\} + \mathrm{Im}\{\mathbf{x}\}^T \mathrm{Im}\{\mathbf{y}\} = \xi(\mathbf{x})^T \xi(\mathbf{y}),$$

which completes the proof. □

For the remainder of this Chapter we define $N := N_{\mathrm{R}} N_{\mathrm{T}} N_{\mathrm{F}}$. Moreover, we use the shorthand $(\forall k \in \mathcal{I}_{\mathrm{sc}})$ $\mathbf{h}[k] = \mathbf{h}_{(k-1)N_{\mathrm{F}}+1:kN_{\mathrm{F}}} \in \mathbb{C}^{N_{\mathrm{R}} N_{\mathrm{T}}}$ to denote the kth component vector of a vector $\mathbf{h} = \left[\mathbf{h}[1]^T, \ldots, \mathbf{h}[N_{\mathrm{F}}]^T\right]^T \in \mathcal{H}$.

In order to devise an APSM channel estimation algorithm we define *1)* a sequence of closed convex sets $(\mathcal{S}_n \subset \mathcal{H})_{n \in \mathbb{N}}$ comprising all points that are consistent with the respective channel observation according to (5.1), *2)* $(\forall n \in \mathbb{N})$ a convex cost function $\Theta_n : \mathcal{H} \to \mathbb{R}_+$ that takes the value zero for arguments $\mathbf{x} \in \bigcap_{k=1}^{n} \mathcal{S}_n$, and *3)* a closed convex set $\mathcal{K}$ representing prior knowledge on the true channel. Then the update rule follows from (2.18). We detail these three steps below.

1) Denote by $\mathcal{P}_n \subset \mathcal{I}_{\mathrm{sc}}$ the set of subcarrier indices, on which pilot signals are transmitted at time interval n. In the absence of measurement noise, the set of vectorized channels that satisfy (5.1) is the affine subspace

$$\mathcal{S}_n = \left\{ \mathbf{h} \in \mathcal{H} \;\middle|\; (\forall k \in \mathcal{P}_n)\ (\mathbf{x}_{n,k}^T \otimes \mathbf{B}_n)\mathbf{h}[k] = \mathbf{y}_{n,k} \right\}. \tag{5.8}$$

Extending the set $\mathcal{S}_n$ by a tolerance margin that allows for a positive noise power is straightforward. However, we restrict our attention to the formulation in (5.8) for the simplicity of presentation.

2) We consider the sequence of cost functions given by

$$(\forall n \in \mathbb{N}) \quad \Theta_n : \mathcal{H} \to \mathbb{R}_+, \quad \Theta_n(\mathbf{h}) := \frac{1}{c_n} \sum_{i \in \mathcal{T}_n} w_i^{(n)} d(\hat{\mathbf{h}}_n, \mathcal{S}_i) d(\mathbf{h}, \mathcal{S}_i), \tag{5.9}$$

where $c_n = \sum_{i \in \mathcal{T}_n} w_i^{(n)} d\left(\mathbf{x}_n, \mathcal{S}_i\right)$ is a positive scaling factor, $\hat{\mathbf{h}}_n$ is the channel estimate in the nth iteration, and $\mathcal{T}_n = \{n - L, \ldots, n\}$ is a discrete set indexing the $L + 1$ most recent time intervals at time n. A convenient choice for the weights is

$$(\forall n \in \mathbb{N})(\forall i \in \mathcal{T}_n) \quad w_i^{(n)} := \frac{\eta^{n-i}}{\sum_{l \in \mathcal{T}_n} \eta^{n-l}}, \tag{5.10}$$

where $\eta \in (0, 1]$. In this way, for $\eta < 1$, the weight given to past measurements decays exponentially, whereas $\eta = 1$ results in a uniform weighting. Nonsmooth cost functions

of the type in (5.9) have been widely used in the literature (see, e.g., [YO05, YCY05, CY08, TSY10]).

Given a nonempty closed convex set $\mathcal{K} \subset \mathcal{H}$, we can substitute the cost functions in (5.9) in the definition of the APSM in (2.18) to obtain a sequence $(T_n^{\mathcal{K}})_{n\in\mathbb{N}}$ of mappings $(\forall n \in \mathbb{N})\ T_n^{\mathcal{K}} : \mathcal{H} \to \mathcal{H}$ given by

$$T_n^{\mathcal{K}}(\mathbf{h}) := P_{\mathcal{K}}\left(\mathbf{h} + \mu_n M_n \left(\sum_{i\in\mathcal{T}_n} w_i^{(n)} P_{\mathcal{S}_i}(\mathbf{h}) - \mathbf{h}\right)\right), \tag{5.11}$$

where $(\forall n \in \mathbb{N})\ \mu_n \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$, and

$$M_n = \frac{\sum_{i\in\mathcal{T}_n} w_i^{(n)} \left\| P_{\mathcal{S}_i}\left(\hat{\mathbf{h}}_n\right) - \hat{\mathbf{h}}_n \right\|^2}{\left\| \sum_{i\in\mathcal{T}_n} w_i^{(n)} P_{\mathcal{S}_i}\left(\hat{\mathbf{h}}_n\right) - \hat{\mathbf{h}}_n \right\|^2}. \tag{5.12}$$

3) If the channel has a sparse representation of the form $\mathbf{h}_n = \boldsymbol{\Psi}\boldsymbol{\theta}_n$, this side-information can be encoded in the constraint set $\mathcal{K}$ of the APSM in (5.11). A common choice for encouraging sparsity in a given domain $\boldsymbol{\Psi}$ is the ℓ_1-norm constraint

$$\mathcal{A} := \left\{ \mathbf{h} \in \mathcal{H} \,\middle|\, \|\boldsymbol{\Psi}^{-1}\mathbf{h}\|_1 \le z \right\}, \tag{5.13}$$

for some $z > 0$. If $\boldsymbol{\Psi}$ is a unitary matrix, the projection onto $\mathcal{A}$ can be computed in a fixed number of steps, as stated in Proposition 5.3 below. Using the ℓ_1-norm constraint in (5.13), we can now define a sequence of channel estimates by

$$(\forall n \in \mathbb{N}) \quad \hat{\mathbf{h}}_{n+1} = T_n^{\mathcal{A}}(\hat{\mathbf{h}}_n), \quad \hat{\mathbf{h}}_0 \in \mathcal{H}. \tag{5.14}$$

Although the convergence of the APSM can only be guaranteed if the intersection of all sets in (5.8) contains the true time-invariant channel vector, many studies have successfully applied it to solve online learning and adaptive filtering problems (see, e.g., [TSY10, YSY02, CYM09]). To implement the mapping $T_n^{\mathcal{A}}$, we need to compute the projections onto the sets $\mathcal{S}_n$ and $\mathcal{A}$. The projection onto the set $\mathcal{S}_n$ is given in Proposition 5.2 below.

Proposition 5.2. *For all $n \in \mathbb{N}$ the projection $\mathbf{p}_n = P_{\mathcal{S}_n}(\mathbf{h})$ of a point $\mathbf{h} \in \mathcal{H}$ onto*

$\mathcal{S}_n$ is given component-wise by

$$(\forall k \in \mathcal{I}_{\mathrm{sc}}) \quad \mathbf{p}_n[k] = \begin{cases} \mathbf{h}[k] - \mathbf{A}_{n,k}^{\dagger} \left(\mathbf{A}_{n,k} \mathbf{h}[k] - \mathbf{y}_{n,k} \right) & \text{if } k \in \mathcal{P}_n \\ \mathbf{h}[k] & \text{otherwise,} \end{cases} \tag{5.15}$$

where $(\forall n \in \mathbb{N})(\forall k \in \mathcal{P}_n)$ $\mathbf{A}_{n,k} = \mathbf{x}_{n,k}^T \otimes \mathbf{B}_n$.

Proof. *By defining $(\forall k \in \mathcal{I}_{\mathrm{sc}} \setminus \mathcal{P}_n)$ $\mathbf{A}_{n,k} := \mathbf{0} \in \mathbb{C}^{L_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}$, $\mathbf{A}_n := \bigoplus_{k=1}^{N_{\mathrm{F}}} \mathbf{A}_{n,k}$, and $\mathbf{b}_n = [\mathbf{b}_{n,1}^T, \ldots, \mathbf{b}_{n,N_{\mathrm{F}}}^T]^T$, where*

$$(\forall k \in \mathcal{I}_{\mathrm{sc}}) \quad \mathbf{b}_{n,k} := \begin{cases} \mathbf{y}_{n,k} & \text{if } k \in \mathcal{P}_n \\ \mathbf{0} \in \mathbb{C}^{L_{\mathrm{R}}} & \text{otherwise,} \end{cases}$$

we can rewrite (5.8) as

$$\mathcal{S}_n = \{ \mathbf{h} \in \mathcal{H} \mid \mathbf{A}_n \mathbf{h} = \mathbf{b}_n \}.$$

In the real Euclidean Hilbert space space $\mathbb{R}^{2N}$, which is isometrically isomorphic to $\mathcal{H}$ (see Proposition 5.1), the complex vector-matrix multiplication $\mathbf{A}_n \mathbf{h}$ can be expressed by $\hat{\mathbf{A}}_n \xi_N(\mathbf{h}) = \xi_{L_{\mathrm{R}} N_{\mathrm{F}}}(\mathbf{b}_n)$, where

$$\hat{\mathbf{A}}_n = \begin{bmatrix} \mathrm{Re}\{\mathbf{A}_n\} & -\mathrm{Im}\{\mathbf{A}_n\} \\ \mathrm{Im}\{\mathbf{A}_n\} & \mathrm{Re}\{\mathbf{A}_n\} \end{bmatrix}.$$

Consequently, $\mathcal{S}_n$ is an affine subspace, the projection onto which is given by [SY98, Section 4.5]

$$P_{\mathcal{S}_n}(\mathbf{h}) = \mathbf{h} - \mathbf{A}_n^{\dagger} \left(\mathbf{A}_n \mathbf{h} - \mathbf{b}_n \right).$$

The Moore-Penrose inverse of the direct sum satisfies $\left(\bigoplus_{m=1}^{M} \mathbf{A}_m \right)^{\dagger} = \bigoplus_{m=1}^{M} \mathbf{A}_m^{\dagger}$. Hence the projection can be computed for each subcarrier independently, which yields the result in (5.15). □

Lemma 5.1. *Let $(\mathcal{H} = \mathbb{C}^N, \langle \cdot, \cdot \rangle)$ be a real Hilbert space with inner product defined in (5.6) and let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper closed function and let $\mathbf{U} \in \mathbb{C}^{N \times N}$ be a unitary matrix. Then the proximal mapping associated with $f(\mathbf{U}\cdot)$ can be expressed in terms of the proximal mapping associated with $f(\cdot)$, where*

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \mathrm{prox}_{f(\mathbf{U}\cdot)}(\mathbf{x}) = \mathbf{U}^H \mathrm{prox}_{f(\cdot)}(\mathbf{U}\mathbf{x}).$$

Proof. *According to Proposition 5.1, we can represent $\mathbf{x} \in \mathcal{H}$ in terms of the*

real Euclidean vector $\hat{\mathbf{x}} = \xi(\mathbf{x}) \in \mathbb{R}^{2N}$*, by applying the bijective linear mapping* ξ*. Moreover, by [Tel99, Eq. (4a)], the matrix*

$$\hat{\mathbf{U}} = \begin{bmatrix} \mathrm{Re}\{\mathbf{U}\} & -\mathrm{Im}\{\mathbf{U}\} \\ \mathrm{Im}\{\mathbf{U}\} & \mathrm{Re}\{\mathbf{U}\} \end{bmatrix} \in \mathbb{R}^{2N \times 2N}$$

is orthogonal whenever $\mathbf{U}$ *is unitary. Hence by defining* $\hat{f} : \mathbb{R}^{2N} \to \mathbb{R} \cup \{+\infty\}$ *such that* $(\forall \mathbf{x} \in \mathcal{H})\ \hat{f}(\xi(\mathbf{x})) = f(\mathbf{x})$ *and applying Fact 1.2 we obtain*

$$\mathrm{prox}_{f(\mathbf{U}\cdot)}(\mathbf{x}) = \xi^{-1}\left(\hat{\mathbf{U}}^T \mathrm{prox}_{\hat{f}(\cdot)}(\hat{\mathbf{U}}\hat{\mathbf{x}})\right) = \mathbf{U}^H \mathrm{prox}_{f(\cdot)}(\mathbf{U}\mathbf{x}),$$

which is the desired result. □

The projection onto the set $\mathcal{A}$ in (5.13) is given in Proposition 5.3, which requires Algorithm 3.

Algorithm 3 Algorithm for projection onto the simplex [DSSSC08]

1: **Input:** $\mathbf{v} \in \mathbb{R}^N$ and a simplex $\mathcal{D} = \{\mathbf{v} \in \mathbb{R}_+^N |\ \sum_{i=1}^N v_i \leq z\}$
2: Sort $\mathbf{v}$ into $\mathbf{u}$: $u_1 \geq u_2 \geq \cdots \geq u_N$
3: Find $\rho = \max\left\{i \in \{1, \ldots, N\} \,\middle|\, u_i - \frac{1}{i}\left(\sum_{r=1}^i u_r - z\right) > 0\right\}$
4: Define $\theta = \frac{1}{\rho}\left(\sum_{i=1}^\rho u_i - z\right)$
5: **Output:** $P_{\mathcal{D}}(\mathbf{v}) = \mathbf{w}$, where $w_i = \max\{v_i - \theta, 0\}$

Proposition 5.3. *Let* $\mathcal{A}$ *be the closed convex set in* (5.13) *defined by some unitary matrix* $\mathbf{\Psi}$ *and* $z > 0$*. The projection of* $\mathbf{x} \in \mathbb{C}^N$ *onto* $\mathcal{A}$ *is given by* $P_{\mathcal{A}}(\mathbf{x}) = \mathbf{\Psi} P_{\mathcal{B}}(\mathbf{\Psi}^H \mathbf{x})$ *where* $\mathcal{B} = \left\{\mathbf{x} \in \mathbb{C}^N |\ \|\mathbf{x}\|_1 \leq z\right\}$*. The projection onto* $\mathcal{B}$ *is given by* $(\forall \mathbf{x} \in \mathcal{H})$ $P_{\mathcal{B}}(\mathbf{x}) = \mathbf{Q}_{\mathbf{x}} P_{\mathcal{D}}(\mathbf{Q}_{\mathbf{x}}^H \mathbf{x})$ *where* $\mathbf{Q}_{\mathbf{x}} := \mathrm{diag}(\exp(j\angle x_1), \ldots, \exp(j\angle x_N))$ *and* $P_{\mathcal{D}}$ *is defined in Algorithm 3.*

Proof. *Note that* $(\forall \mathbf{x} \in \mathcal{H})$ $\mathbf{Q}_{\mathbf{x}}^H \mathbf{x} \in \mathbb{R}^N$ *and* $\mathbf{Q}_{\mathbf{x}}^H \mathbf{x} \in \mathcal{D} \iff \mathbf{x} \in \mathcal{B}$*. Thus* $(\forall \mathbf{x} \in \mathcal{H})$ $\iota_{\mathcal{B}}(\mathbf{x}) = \iota_{\mathcal{D}}(\mathbf{Q}^H \mathbf{x})$*. Hence by Lemma 5.1 it holds that* $(\forall \mathbf{x} \in \mathcal{H})$

$$P_{\mathcal{B}}(\mathbf{x}) = \mathrm{prox}_{\iota_{\mathcal{B}}(\cdot)}(\mathbf{x}) = \mathrm{prox}_{\iota_{\mathcal{D}}(\mathbf{Q}_{\mathbf{x}}^H \cdot)}(\mathbf{x}) = \mathbf{Q}_{\mathbf{x}} \mathrm{prox}_{\iota_{\mathcal{D}}(\cdot)}(\mathbf{Q}_{\mathbf{x}}^H \mathbf{x}) = \mathbf{Q}_{\mathbf{x}} P_{\mathcal{D}}(\mathbf{Q}_{\mathbf{x}}^H \mathbf{x}).$$

Applying Lemma 5.1 again yields $P_{\mathcal{A}}(\mathbf{x}) = \mathbf{\Psi} P_{\mathcal{B}}(\mathbf{\Psi}^H \mathbf{x})$*, which concludes the proof.* □

A drawback of the algorithm in (5.11) is that it requires knowledge of a suitable upper-bound z on the ℓ_1-norm to define the set $\mathcal{A}$, and the optimal choice of Z is highly dependent on the problem dimensions. To mitigate this problem, we proposed a heuristic in [6], in which the projection onto $\mathcal{A}$ is replaced by the proximal mapping associated with a

weighted ℓ_1-norm. In Subsection 5.3.3, we provide some theoretical justification for this heuristic by applying Theorem 2.3.

5.3.3. Weighted ℓ_1-Norm Regularization via Superiorization

According to Theorem 2.3, perturbed versions of the APSM exhibit the same desirable properties as their unperturbed counterparts, except for monotone approximation. Therefore, we devise a superiorized variant of the algorithm in (5.11) by dropping the ℓ_1-norm constraint $\mathcal{A}$ (i.e., by defining $\mathcal{K} = \mathcal{H}$), by introducing a sparsity encouraging superiorization objective, and by devising bounded perturbations $(\beta_n \mathbf{v}_n)_{n\in\mathbb{N}}$ that aim at reducing the value of this objective function. The superiorized version of (5.14) generates a sequence by

$$(\forall n \in \mathbb{N}) \quad \hat{\mathbf{h}}_{n+1} = T_n^{\mathcal{H}} \left(\hat{\mathbf{h}}_n + \beta_n \mathbf{v}_n \right), \quad \hat{\mathbf{h}}_0 \in \mathcal{H}.$$

It has been observed in [CWB08] that the performance of sparse reconstruction algorithms can be improved, compared to plain ℓ_1 minimization, by solving a sequence of reweighted ℓ_1 minimization problems. Hence we devise a sequence $(f_n)_{n\in\mathbb{N}}$ of superiorization objective functions, where $(\forall n \in \mathbb{N})$ $f_n : \mathcal{H} \to \mathbb{R}_+$ is given by

$$(\forall \mathbf{x} \in \mathcal{H}) \quad f_n(\mathbf{x}) := \lambda \|\boldsymbol{\Psi}\mathbf{x}\|_1^{\boldsymbol{\omega}^{(n)}}. \tag{5.16}$$

Here, $\boldsymbol{\Psi} \in \mathbb{C}^N$ is a unitary matrix, $\lambda \in \mathbb{R}_+$ is a regularization constant, and $(\forall \boldsymbol{\omega} \in \mathbb{R}_{++}^N)$ $\|\cdot\|_1^{\boldsymbol{\omega}} : \mathcal{H} \to \mathbb{R}_+$ is a weighted ℓ_1-norm given by

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \|\mathbf{x}\|_1^{\boldsymbol{\omega}} = \sum_{i=1}^{N} \omega_i |x_i| = \sum_{i=1}^{N} \omega_i \sqrt{(\mathrm{Re}\{x_i\})^2 + (\mathrm{Im}\{x_i\})^2}. \tag{5.17}$$

To define the sequence of weight vectors $\left(\boldsymbol{\omega}^{(n)}\right)_{n\in\mathbb{N}}$, we use the iterative reweighting scheme proposed in [CWB08], where $(\forall n \in \mathbb{N})$ $(\forall i \in \{1, \ldots, N\})$

$$\omega_i^{(0)} = 1; \quad (\forall n > 0) \quad \omega_i^{(n)} = \frac{1}{\left|\theta_i^{(n)}\right| + \nu}; \quad \boldsymbol{\theta}^{(n)} = \boldsymbol{\Psi}^H \hat{\mathbf{h}}_n, \tag{5.18}$$

Similarly to Chapters 3 and 4 (where the same superiorization objective was used throughout all iterations), we use the proximal mapping associated with f_n to define perturbation $\mathbf{v}_n$. A formal definition of this proximal mapping is provided in the following propositions.

Proposition 5.4. *The proximal mapping associated with the weighted ℓ_1-norm in* (5.17) *is given by*[a] $\operatorname{prox}_{\lambda\|\cdot\|_1^{\boldsymbol{\omega}}} : \mathcal{H} \to \mathcal{H}$

$$\operatorname{prox}_{\lambda\|\cdot\|_1^{\boldsymbol{\omega}}}(\mathbf{x}) = \sum_{i=1}^{N} x_i \left(1 - \frac{\lambda\omega_i}{|x_i|}\right)_+ \mathbf{e}_i.$$

Proof: *See Appendix C.1.*

[a]Note that the weighted ℓ_1-norm is defined for complex vectors $\mathbf{x} \in \mathcal{H} = \mathbb{C}^N$ whereas the proximal mapping is derived with respect to their real representation $\xi(\mathbf{x}) \in \mathbb{R}^{2N}$.

Proposition 5.5. *For all $n \in \mathbb{N}$ the proximal mapping associated with the function f_n in* (5.16) *is given by*

$$\operatorname{prox}_{f_n}(\mathbf{x}) = \boldsymbol{\Psi}\operatorname{prox}_{\lambda\|\cdot\|_1^{\boldsymbol{\omega}^{(n)}}}(\boldsymbol{\Psi}^H\mathbf{x}),$$

where $\operatorname{prox}_{\lambda\|\cdot\|_1^{\boldsymbol{\omega}}}$ *is defined in Proposition 5.4.*

Proof. *Since $\boldsymbol{\Psi}$ is a unitary matrix and $(\forall n \in \mathbb{N})$ $f_n = \lambda\|\cdot\|_1^{\boldsymbol{\omega}^{(n)}}$ is a proper closed function, this proposition follows immediately from Lemma 5.1.* □

Applying the proximal mapping in Proposition 5.5 to the channel estimate $\hat{\mathbf{h}}_n$ reduces the weighted ℓ_1-norm of its sparse representation $\boldsymbol{\Psi}^H\hat{\mathbf{h}}_n$. Therefore, we define perturbations by

$$(\forall n \in \mathbb{N}) \quad \mathbf{v}_n := \operatorname{prox}_{f_n}\left(\hat{\mathbf{h}}_n\right) - \hat{\mathbf{h}}_n.$$

We note that, for $(\beta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, boundedness of this type of perturbation can be guaranteed using arguments analogous to those in Lemmas 3.1 and 3.2. However, we omit the formal proof, and we heuristically choose $(\forall n \in \mathbb{N})$ $\beta_n = 1$, which results in $(\beta_n)_{n\in\mathbb{N}} \neq \ell_+^1(\mathbb{N})$. This choice of the sequence $(\beta_n)_{n\in\mathbb{N}}$ yields $(\forall n \in \mathbb{N})$

$$\begin{aligned}\breve{\mathbf{h}}_n &= \operatorname{prox}_{f_n}\left(\hat{\mathbf{h}}_n\right), \quad \hat{\mathbf{h}}_0 \in \mathcal{H}\\ \breve{\mathbf{h}}_{n+1} &= \operatorname{prox}_{f_{n+1}}\left(T_n^{\mathcal{H}}\left(\breve{\mathbf{h}}_n\right)\right).\end{aligned}$$

As the choice of the initial point is arbitrary, this algorithm is equivalent to the heuristic proposed in [6], which is given by

$$(\forall n \in \mathbb{N}) \quad \hat{\mathbf{h}}_{n+1} = \overline{T}_n(\hat{\mathbf{h}}_n), \quad \hat{\mathbf{h}}_0 \in \mathcal{H}.$$

Here, $(\forall n \in \mathbb{N})$ the mapping $\overline{T}_n : \mathcal{H} \to \mathcal{H}$ is defined by $(\forall \mathbf{h} \in \mathcal{H})$

$$\overline{T}_n(\mathbf{h}) = \operatorname{prox}_{f_{n+1}} \left(\mathbf{h} + \mu_n M_n \left(\sum_{k \in \mathcal{T}_n} w_k^{(n)} P_{\mathcal{S}_k}(\mathbf{h}) - \mathbf{h} \right) \right), \tag{5.19}$$

where $w_k^{(n)}$ and M_n are as defined in (5.10) and (5.12), respectively, and $\mu_n \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$. To accelerate the convergence of the proposed algorithms, the update rules in (5.11) and (5.19) can be applied t times in each iteration n, resulting in the recursions $\hat{\mathbf{h}}_{n+1} = (T_n^{\mathcal{A}})^t(\hat{\mathbf{h}}_n)$ and $\hat{\mathbf{h}}_{n+1} = \overline{T}_n^t(\hat{\mathbf{h}}_n)$, respectively.

5.3.4. A Sparse Representation of the Channel

The algorithms presented above exploit sparsity of the channel in a certain domain. When both transmitter and receiver are equipped with a uniform linear array (ULA), a sparse representation of the channel $\mathbf{H}_{n,k} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$ can be obtained by a two-dimensional discrete Fourier transform (DFT) $\mathbf{F}_{N_\mathrm{R}}^H \mathbf{H}_{n,k} \mathbf{F}_{N_\mathrm{T}}$ (see, e.g., [HGPR+16]), where $\mathbf{F}_{N_\mathrm{R}} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{R}}$ and $\mathbf{F}_{N_\mathrm{T}} \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{T}}$ are unitary DFT matrices. Applying the identity $\operatorname{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \operatorname{vec}(\mathbf{B})$, we can express this sparse representation by $\bar{\boldsymbol{\theta}}_{n,k} = \bar{\boldsymbol{\Psi}}^H \mathbf{h}_{n,k}$, where $\bar{\boldsymbol{\Psi}}^H := (\mathbf{F}_{N_\mathrm{T}}^T \otimes \mathbf{F}_{N_\mathrm{R}}^H)$ and $\mathbf{h}_{n,k} = \operatorname{vec}(\mathbf{H}_{n,k})$. Note that the Kronecker product of two unitary matrices is unitary [RS89, Property 2.5], so $\bar{\boldsymbol{\Psi}}$ is a unitary matrix. For uniform planar arrays (UPAs), the matrix $\bar{\boldsymbol{\Psi}}$ can be constructed in a similar way from lower-dimensional DFT matrices, which implement DFTs on fibers of a tensor that is obtained by reshaping the channel matrix $\mathbf{H}_{n,k}$ into vertical and horizontal antenna elements at the transmitter and the receiver.

In addition to sparsity in the angular domain, we can exploit sparsity of the channel impulse response, which is commonly assumed for wideband channels [Mol05]. To do so, we define $\bar{\boldsymbol{\Theta}}_n = [\bar{\boldsymbol{\theta}}_{n,1}, \ldots, \bar{\boldsymbol{\theta}}_{N_\mathrm{F}}]$ and $\boldsymbol{\Theta}_n = \bar{\boldsymbol{\Theta}}_n \mathbf{F}_{N_\mathrm{F}}^H$, where $\mathbf{F}_{N_\mathrm{F}} \in \mathbb{C}^{N_\mathrm{F} \times N_\mathrm{F}}$ is a unitary DFT matrix. Applying once more the identity $\operatorname{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \operatorname{vec}(\mathbf{B})$, we can express the vectorized channel in the time- and angular domain by

$$\boldsymbol{\theta}_n = \operatorname{vec}(\boldsymbol{\Theta}_n) = \boldsymbol{\Psi}^H \mathbf{h}_n,$$

where $\boldsymbol{\Psi}^H = \mathbf{F}_{N_\mathrm{F}}^* \otimes \bar{\boldsymbol{\Psi}}^H = \mathbf{F}_{N_\mathrm{F}}^* \otimes \mathbf{F}_{N_\mathrm{T}}^T \otimes \mathbf{F}_{N_\mathrm{R}}^H$ is unitary. Note that multiplication by $\boldsymbol{\Psi}$ or $\boldsymbol{\Psi}^H$ can be implemented very efficiently via low-dimensional fast Fourier transforms (FFTs) on fibers of a tensor that is obtained by reshaping the channel vector $\mathbf{h}_n$.

5.3.5. Delay and Phase Compensation

The proposed online algorithms in Sections 5.3.2 and 5.3.3 do not assume that the channel remains constant. However, large variations of the channel coefficients in subsequent time intervals can render tracking the channel impossible. One reason for this is that the cost function in (5.9) depends on measurements obtained in multiple time intervals, so the true current channel might be far outside the set of minimizers of this cost function if the difference between subsequent channel realizations is large. Note that this is the case for the generalized channel model in (5.3), in which the absolute phase shift and delay vary randomly between subsequent channel observations. Therefore, applying the proposed online algorithms to the generalized channel model in (5.2) calls for a mechanism that compensates for the random delay and phase variations. To simplify the following presentation, we assume that known pilot signals are transmitted on all subcarriers in each time interval n, i.e., $(\forall n \in \mathbb{N})\ \mathcal{P}_n = \mathcal{I}_{\mathrm{sc}}$. The application of the proposed delay and phase estimation scheme in the general case $\mathcal{P}_n \subset \mathcal{I}_{\mathrm{sc}}$ is described in Remark 5.1.

According to (5.2), the sequence of channels $(\tilde{\mathbf{H}}_n)_{n\in\mathbb{N}}$ is composed of a sequence $(\mathbf{H}_n)_{n\in\mathbb{N}}$ of spatially consistent (and hence slowly varying) channels, where each channel realization experiences a random absolute delay and phase shift. More formally, by introducing the shorthand $(\forall \varphi \in [0, 2\pi])(\forall \tau \in [0, 1])$

$$\mathbf{Q}(\varphi, \tau) := e^{j\varphi}\mathrm{diag}\left(1, e^{-j2\pi\tau}, \ldots, e^{-j2\pi(N_{\mathrm{F}}-1)\tau}\right), \tag{5.20}$$

we can express (5.2) as $(\forall n \in \mathbb{N})\ \tilde{\mathbf{H}}_n = \mathbf{H}_n\mathbf{Q}(\varphi_n, \tau_n)$. Although the decomposition of $\tilde{\mathbf{H}}_n$ into τ_n, φ_n, and $\mathbf{H}_n$ is ambiguous, it is always possible to decompose $(\tilde{\mathbf{H}}_n)_{n\in\mathbb{N}}$ into a sequence $(\check{\mathbf{H}}_n)_{n\in\mathbb{N}}$ of slowly varying channels (i.e., $(\forall n \in \mathbb{N})\ \left\|\check{\mathbf{H}}_{n+1} - \check{\mathbf{H}}_n\right\|_F$ is small) and the corresponding delays $(\check{\tau}_n)_{n\in\mathbb{N}}$ and phase shifts $(\check{\varphi}_n)_{n\in\mathbb{N}}$ such that

$$(\forall n \in \mathbb{N}) \quad \tilde{\mathbf{H}}_n = \check{\mathbf{H}}_n\mathbf{Q}(\check{\varphi}_n, \check{\tau}_n).$$

Therefore, we can use the proposed algorithms to track the sequence $(\check{\mathbf{H}}_n)_{n\in\mathbb{N}}$ of slowly varying channels. Given an estimate $\hat{\mathbf{H}}_n = [\hat{\mathbf{h}}_{n,1}, \ldots, \hat{\mathbf{h}}_{n,N_{\mathrm{F}}}] \approx \check{\mathbf{H}}_n$, a prediction $\hat{\mathbf{Y}}_n = [\hat{\mathbf{y}}_{n,1}, \ldots, \hat{\mathbf{y}}_{n,N_{\mathrm{F}}}]$ of the (noiseless) receive signal at time interval $n \in \mathbb{N}$ can be computed based on the known pilot signals and analog combining matrix as

$$(\forall k \in \mathcal{I}_{\mathrm{sc}}) \quad \hat{\mathbf{y}}_{n,k} = \left(\mathbf{x}_{n,k}^T \otimes \mathbf{B}_n\right)\hat{\mathbf{h}}_{n,k}. \tag{5.21}$$

Upon reception of the true receive signal $\tilde{\mathbf{Y}}_n = [\tilde{\mathbf{y}}_{n,1}, \ldots \mathbf{y}_{n,N_{\mathrm{F}}}]$, we can estimate the

delay and phase offset as

$$(\hat{\varphi}_n, \hat{\tau}_n) \in \arg\min_{\varphi,\tau} \left\| \tilde{\mathbf{Y}}_n - \hat{\mathbf{Y}}_n \mathbf{Q}(\varphi, \tau) \right\|_F^2 . \tag{5.22}$$

Then we can construct the constraint set $\mathcal{S}_n$ in (5.8) using the delay- and phase compensated versions of the receive signal $\tilde{\mathbf{Y}}^{(t)}$ given by

$$\mathbf{Y}_n := \tilde{\mathbf{Y}}_n \left(\mathbf{Q}(\hat{\varphi}_n, \hat{\tau}_n) \right)^H ,$$

and compute the next estimate $\hat{\mathbf{h}}_{n+1} = \mathrm{vec}(\hat{\mathbf{H}}_{n+1}) = T_n^t(\hat{\mathbf{h}}_n)$ of the delay- and phase compensated channel by t-fold application of either of the mappings $T_n = T_n^{\mathcal{A}}$ in (5.11) or $T_n = \overline{T}_n$ in (5.19). Finally, the sequence $(\tilde{\mathbf{H}}_n)_{n\in\mathbb{N}}$ of channels including random delay and phase variations can be approximated by $\tilde{\mathbf{H}}_n \approx \hat{\mathbf{H}}_{n+1} \mathbf{Q}(\hat{\varphi}_n, \hat{\tau}_n)$.

Proposed Phase and Delay Estimators

In the following, we propose an algorithm for the delay and phase estimation problem in (5.22). To illustrate this approach, we consider an idealized noiseless setting in which $\tilde{\mathbf{Y}}_n = \hat{\mathbf{Y}}_n \mathbf{Q}(\varphi_n, \tau_n)$. Then in Proposition 5.6, we show that the proposed estimators in fact approximate a solution to Problem 5.22. A very common technique for estimating delay and phase shift is based on the cross-correlation. To illustrate this idea, consider two continuous (complex valued) time-domain signals $s_1(t)$ and $s_2(t)$, where s_2 is a delayed and phase-shifted version of s_1 given by

$$s_2(t) = e^{j\varphi} s_1(t - \tau_0).$$

Under this model, the cross-correlation

$$\begin{aligned} R_{s_1 s_2}(\tau) &= \int_{-\infty}^{\infty} s_1^*(t) s_2(t+\tau) d\tau \\ &= \int_{-\infty}^{\infty} s_1^*(t) e^{j\varphi} s_1(t - \tau_0 + \tau) d\tau \\ &= e^{j\varphi} R_{s_1 s_1}(\tau - \tau_0) \end{aligned}$$

is a translated and phase-shifted version of the auto-correlation of $s_1(t)$. It follows from the Cauchy-Schwartz inequality, that the auto-correlation achieves its maximum at zero. Thus the cross-correlation function $R_{s_1 s_2}(\tau)$ has a maximum at $\tau = \tau_0$. Moreover, the auto-correlation is real-valued at the origin in general, i.e., $\mathrm{Im}\{R_{s_1,s_1}(0)\} = 0$. Therefore,

the phase shift φ can be obtained from the cross-correlation function, since

$$\angle R_{s_1 s_2}(\tau_0) = \angle e^{j\varphi} R_{s_1 s_1}(0) = \varphi.$$

Consequently, under mild assumptions on the wave-form of $s_1(t)$, both delay and phase shift can in principle be recovered with arbitrary precision in the absence of noise. For discrete signals, however, the resolution is restricted by the sampling rate. To alleviate this problem, we can compute the discrete cross-correlation of an upsampled time-domain version of the discrete frequency-domain signals considered in this work. To this end, we note that the delay is assumed to be restricted to one OFDM symbol duration, which allows us to estimate delay and phase shift based on the cyclic cross-correlation.

As the delay and phase estimation problem in (5.22) is solved independently for each time interval n, we drop the index n in the following derivation for notational convenience. Denoting by $(\forall l \in \{1, \dots, L_{\mathrm{R}}\})$ $\hat{\mathbf{y}}_l = (\hat{y}_{l,1}, \dots, \hat{y}_{l,N_f})^T \in \mathbb{C}^{N_{\mathrm{F}}}$ and $\tilde{\mathbf{y}}_l = (\tilde{y}_{l,1}, \dots, \tilde{y}_{l,N_f})^T \in \mathbb{C}^{N_{\mathrm{F}}}$ the predicted and received signals according to (5.28), we obtain one upsampled cross-correlation for each RF chain, which we denote by $\mathbf{r}_l$. To combine the observations from all RF chains, we sum over all cyclic cross-correlation vectors $\mathbf{r}_l$, as we expect all L_{R} vectors $\mathbf{r}_l$ to be in phase at the position of their largest peak. The resulting sum of all cross-correlations is given by

$$\mathbf{r} = \sum_{l=1}^{L_{\mathrm{R}}} \mathbf{r}_l = \mathbf{F}^H \sum_{l=1}^{L_{\mathrm{R}}} \hat{\mathbf{y}}_l^* \odot \tilde{\mathbf{y}}_l. \tag{5.23}$$

Here $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{F}} \times M N_{\mathrm{F}}}$ is a DFT matrix with upsampling factor M, which is given by

$$(\forall k \in \mathcal{I}_{\mathrm{sc}})(\forall i \in \{1, \dots, M N_{\mathrm{F}}\}) \quad F_{i,k} = \frac{1}{\sqrt{M N_{\mathrm{F}}}} e^{-j2\pi k \frac{i-1}{M N_{\mathrm{F}}}}. \tag{5.24}$$

In practice, the cross-correlation in (5.23) can be computed very efficiently using the FFT-algorithm. Now we compute

$$i^\star \in \underset{i \in \{1, \dots, M N_{\mathrm{F}}\}}{\arg\max} |r_i|,$$

and estimate the delay and phase shift according to

$$\hat{\tau} = \frac{i^\star - 1}{M N_{\mathrm{F}}} \qquad \text{and} \qquad \hat{\varphi} = \angle c_{k_\star}. \tag{5.25}$$

The following proposition shows that the estimators in (5.25) are in fact suited to approximate solutions to Problem (5.22).

Proposition 5.6. *The estimates $\hat{\varphi}$ and $\hat{\tau}$ in* (5.25) *approximate a solution to Problem* (5.22) *as $M \to \infty$.*

Proof. *Dropping the time index n, we can write the objective function in* (5.22) *as*

$$\begin{aligned}\left\|\tilde{\mathbf{Y}} - \hat{\mathbf{Y}}\mathbf{Q}(\varphi,\tau)\right\|_F^2 &= \|\tilde{\mathbf{Y}}\|_F^2 - 2\mathrm{Re}\left\{\mathrm{tr}\left(\left(\hat{\mathbf{Y}}\mathbf{Q}(\varphi,\tau)\right)^H \tilde{\mathbf{Y}}\right)\right\} + \|\hat{\mathbf{Y}}\mathbf{Q}(\varphi,\tau)\|_F^2 \\ &= \|\tilde{\mathbf{Y}}\|_F^2 - 2\mathrm{Re}\left\{\mathrm{tr}\left(\left(\mathbf{Q}(\varphi,\tau)\right)^H \hat{\mathbf{Y}}^H \tilde{\mathbf{Y}}\right)\right\} + \|\hat{\mathbf{Y}}\|_F^2,\end{aligned}$$

where the second equality follows from the fact that $\mathbf{Q}(\varphi,\tau)$ is a unitary matrix and the that Frobenius norm is invariant under uniform transformation. Since both $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$ are independent from the optimization variables φ and τ, we have that

$$\arg\min_{\varphi,\tau} \left\|\tilde{\mathbf{Y}} - \hat{\mathbf{Y}}\mathbf{Q}(\varphi,\tau)\right\|_F^2 = \arg\max_{\varphi,\tau} \mathrm{Re}\left\{\mathrm{tr}\left(\left(\mathbf{Q}(\varphi,\tau)\right)^H \hat{\mathbf{Y}}^H \tilde{\mathbf{Y}}\right)\right\}.$$

By definition in (5.20), *$\mathbf{Q}(\varphi,\tau)$ is a diagonal matrix. Moreover, for complex matrices $\mathbf{A}$ and $\mathbf{B}$ of identical size, $\mathrm{tr}(\mathbf{A}^H\mathbf{B}) = \mathrm{vec}(\mathbf{A})^H\mathrm{vec}(\mathbf{B})$, so it holds that*

$$\begin{aligned}\mathrm{Re}\left\{\mathrm{tr}\left(\left(\mathbf{Q}(\varphi,\tau)\right)^H \hat{\mathbf{Y}}^H \tilde{\mathbf{Y}}\right)\right\} &= \mathrm{Re}\left\{\left(\mathrm{diag}\left(\mathbf{Q}(\varphi,\tau)\right)\right)^H \mathrm{diag}\left(\hat{\mathbf{Y}}^H \tilde{\mathbf{Y}}\right)\right\} \\ &= \mathrm{Re}\left\{\sum_{k=1}^{N_\mathrm{F}}\sum_{l=1}^{L_\mathrm{R}} \hat{Y}_{l,k}^* \tilde{Y}_{l,k} e^{-j\varphi} e^{j2\pi k\tau}\right\}.\end{aligned}$$

Hence by defining a vector $\mathbf{c} \in \mathbb{C}^{N_\mathrm{F}}$ entry-wise as

$$(\forall n \in \mathcal{I}_\mathrm{sc}) \quad c_k := \sum_{l=1}^{L_\mathrm{R}} \hat{Y}_{l,k}^* \tilde{Y}_{l,k} \tag{5.26}$$

we obtain a reformulation of Problem (5.22) *given by*

$$(\varphi^\star, \tau^\star) \in \arg\max_{\varphi,\tau} \mathrm{Re}\left\{e^{-j\varphi}\sum_{k=1}^{N_\mathrm{F}} c_k \mathbf{e}^{j2\pi k\tau}\right\}. \tag{5.27}$$

A solution to this problem can be approximated by discretizing the continuous variable τ. This can be done by via the DFT matrix $\mathbf{F}$ defined in (5.24), *resulting in $\hat{\tau} = \frac{i^\star - 1}{MN_\mathrm{F}}$, where*

$$(\hat{\varphi}, i^\star) \in \arg\max_{\varphi,i} \mathrm{Re}\left\{e^{-j\varphi} r_i\right\},$$

and $\mathbf{r} = \mathbf{F}^H\mathbf{c} \in \mathbb{C}^{MN_\mathrm{F}}$. Clearly, the real part of $\mathbf{e}^{-j\varphi} r_i$ is maximized when $|r_i|$ is

maximal and $\varphi = \angle r_i$, which corresponds to the estimators proposed in (5.25)*. As the upsampling factor M increases, the grid points τ_i approximate arbitrarily closely the continuous value τ in* (5.27)*. Moreover, the objective function in* (5.27) *is infinitely differentialble w.r.t. τ, so it is smooth. Therefore, $|r_{i^\star}|$ approaches the optimal value of* (5.27)*, as M increases, which proves the proposition.* □

Remark 5.1. We note that estimates of φ and τ can also be obtained for $\mathcal{P}_n \neq \mathcal{I}_{\mathrm{sc}}$, i.e., if not all subcarriers carry pilot signals. Because receive signal can only be predicted for pilot subcarries, we simply set $(\forall k \in \mathcal{I}_{\mathrm{sc}} \setminus \mathcal{P}_n)\ c_k = 0$ in (5.26).

Lower Bounds on the Delay and Phase Estimation Errors

In order to assess the performance of the proposed delay and phase estimators, we now derive bounds on the achievable error variance for the phase and delay estimation problem. To this end, note that the error variance of any unbiased estimator is lower-bounded by the Cramér-Rao lower bound (CRLB). In the following, we consider the ideal setting in which the channel is static and known perfectly (except for the random delay and phase shift), i.e., $(\forall k \in \mathcal{I}_{\mathrm{sc}})$, $\hat{\mathbf{H}}_{n,k} = \mathbf{H}_{n,k}$. In this case, $(\forall n \in \mathbb{N})$ $(\forall k \in \mathcal{I}_{\mathrm{sc}})$

$$\hat{\mathbf{y}}_{n,k} = \mathbf{B}_n \mathbf{H}_{n,k} \mathbf{x}_{n,k} \quad \text{and} \quad \tilde{\mathbf{y}}_{n,k} = e^{j\varphi_n} e^{-j2\pi(k-1)\tau_n} \mathbf{B}_n \mathbf{H}_{n,k} \mathbf{x}_{n,k} + \mathbf{n}_{n,k},$$

where $\mathbf{n}_{n,k} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_{\mathrm{R}}})$. Consequently, the measured receive signal $\tilde{\mathbf{Y}}_n$ is a noisy, delayed, and phase-shifted version of the predicted receive signal $\hat{\mathbf{Y}}_n$. To simplify the proof, we suppose that $\mathbf{B}_n \mathbf{H}_{n,k} \mathbf{x}_{n,k} \sim \mathcal{CN}(\mathbf{0}, \frac{\gamma}{\sigma^2} \mathbf{I}_{N_{\mathrm{R}}})$, where γ denotes the SNR. Then the CRLBs for delay and phase estimation are given by Proposition 5.7.

Proposition 5.7. *Let $\mathcal{P} \subset \mathcal{I}_{\mathrm{sc}}$, let $(\forall k \in \mathcal{P})$ $\hat{\mathbf{y}}_k \sim \mathcal{CN}\left(\mathbf{0}, \frac{\gamma}{\sigma^2} \mathbf{I}_{N_{\mathrm{R}}}\right)$, and define*

$$(\forall k \in \mathcal{P}) \quad \tilde{\mathbf{y}}_k = e^{j\varphi} e^{-j2\pi(k-1)\tau} \hat{\mathbf{y}}_k + \mathbf{n}_k, \tag{5.28}$$

where $\tau \sim \mathcal{U}(0,1)$, $\varphi \sim \mathcal{U}(0, 2\pi)$, and $(\forall k \in \mathcal{P})$ $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_{\mathrm{R}}})$.

Then the CRLBs for unbiased estimators of phase φ and delay τ are given by

$$\mathrm{var}\{\varphi\} \geq \frac{1}{2\gamma L_{\mathrm{R}} \left(|\mathcal{P}| - \frac{C_1^2}{C_2}\right)} \qquad \textit{and} \qquad \mathrm{var}\{\tau\} \geq \frac{1}{8\pi^2 \gamma L_{\mathrm{R}} \left(C_2 - \frac{C_1^2}{|\mathcal{P}|}\right)},$$

where $C_1 := \sum_{k \in \mathcal{P}} (k-1)$ and $C_2 := \sum_{k \in \mathcal{P}} (k-1)^2$. For the special case $\mathcal{P} = \mathcal{I}_{\mathrm{sc}}$ these

expressions are simplify to

$$\mathrm{var}\{\varphi\} \geq \frac{2N_{\mathrm{F}} - 1}{\gamma L_{\mathrm{R}} N_{\mathrm{F}}(N_{\mathrm{F}} + 1)} \qquad \textit{and} \qquad \mathrm{var}\{\tau\} \geq \frac{3}{2\pi^2 \gamma L_{\mathrm{R}} N_{\mathrm{F}}(N_{\mathrm{F}}^2 - 1)}.$$

Proof: *See Appendix C.2*

For $\mathcal{P} = \mathcal{I}_{\mathrm{sc}}$, Proposition 5.7 shows the following important facts about minimum variance unbiased delay and phase estimators. The minimal variance of an unbiased delay estimator is heavily dependent on the number of subcarriers. Asymptotically, the CRLB for delay estimation scales with the third power of the reciprocal of N_{F}, and linearly with the reciprocal of L_{R}. Consequently, increasing the number of subcarriers improves the precision of a minimum variance unbiased delay estimator more drastically than increasing the number of RF chains. In case of the phase estimator, the minimal variance asymptotically decreases proportional to the reciprocals of both N_{F} and L_{R}.

5.3.6. Data-driven Analog Combining Scheme

The algorithms proposed in Subsections 5.3.2 and 5.3.3 can be applied independently of the analog combining policy $(\mathbf{B}_n)_{n\in\mathbb{N}}$. A common approach in the literature (see, e.g., [ALH15]) is to draw the phases of the constant magnitude entries of $\mathbf{B}_n$ in each time interval i.i.d. from a discrete set of quantized angles. In the following, we propose a data-driven combining scheme that exploits channel estimates from previous time intervals to increase the signal power at the RF chains. For a fully connected analog combining network consisting of digitally controlled phase shifters with q bits, the set of potential analog combining matrices is given by

$$\mathcal{L}_q := \left\{ \mathbf{B} \in \mathbb{C}^{L_{\mathrm{R}} \times N_{\mathrm{R}}} \;\middle|\; (\forall (i,k) \in \mathcal{I}_{L_{\mathrm{R}}} \times \mathcal{I}_{N_{\mathrm{R}}}) \;\; |B_{ik}| = 1, \; \angle B_{ik} \bmod 2^{1-q}\pi = 0 \right\},$$

where $\mathcal{I}_{N_{\mathrm{R}}} = \{1, \ldots, N_{\mathrm{R}}\}$ and $\mathcal{I}_{L_{\mathrm{R}}} = \{1, \ldots, L_{\mathrm{R}}\}$. In this architecture, the analog combiner that maximizes the SNR at the RF chains in time interval $n \in \mathbb{N}$ is given by

$$\mathbf{B}_n^{\star} \in \operatorname*{arg\,max}_{\mathbf{B}\in\mathcal{L}_q} \frac{1}{\sigma^2 N_{\mathrm{R}} N_{\mathrm{F}}} \mathbb{E}\left[\sum_{k=1}^{N_{\mathrm{F}}} \|\mathbf{B}\mathbf{H}_{n,k}\mathbf{x}_{n,k}\|_2^2 \right]. \tag{5.29}$$

If $\mathbb{E}\left[\mathbf{x}_{n,k}\mathbf{x}_{n,k}^H\right] = P\mathbf{I}$ for transmit power $P > 0$, we can rewrite (5.29) as

$$\mathbf{B}_n^{\star} \in \operatorname*{arg\,max}_{\mathbf{B}\in\mathcal{L}_q} \mathrm{tr}\left(\mathbb{E}\left[\sum_{k=1}^{N_{\mathrm{F}}} \mathbf{H}_{n,k}\mathbf{H}_{n,k}^H \right] \mathbf{B}^H\mathbf{B} \right), \tag{5.30}$$

where the covariance matrix can be approximated by $\mathbf{R}_n \approx \mathbb{E}\left[\sum_{k=1}^{N_\mathrm{F}} \mathbf{H}_{n,k}\mathbf{H}_{n,k}^H\right]$, based on the channel estimates in the previous time steps

$$(\forall n \in \mathbb{N}) \quad \mathbf{R}_{n+1} := \mathbf{R}_n + \alpha\left(\sum_{k=1}^{N_\mathrm{F}} \hat{\mathbf{H}}_{n,k}\hat{\mathbf{H}}_{n,k}^H - \mathbf{R}_n\right), \quad \mathbf{R}_0 := \mathbf{I}_{N_\mathrm{R}}. \tag{5.31}$$

Here, the constant $\alpha \in (0,1]$ determines how fast $\mathbf{R}_n$ adapts to new estimates of the channel. Solving Problem (5.30) is difficult because of its combinatorial nature. Therefore, we devise a low-complexity heuristic by replacing the constraint set $\mathcal{L}_q$ in (5.30) with

$$\mathcal{C} := \left\{\mathbf{B} \in \mathbb{C}^{L_\mathrm{R}\times N_\mathrm{R}} |\ \|\mathbf{B}^H\mathbf{B}\|_F \leq r\right\},$$

for some $r > 0$.[2] Approximating the covariance matrix according to (5.31) and replacing the nonconvex set $\mathcal{L}_q$ by by the convex set $\mathcal{C}$ yields the relaxed problem

$$\mathbf{W}_n^\star \in \operatorname*{arg\,max}_{\mathbf{B}\in\mathcal{C}} \operatorname{tr}\left(\mathbf{B}^H\mathbf{B}\mathbf{R}_n\right). \tag{5.32}$$

Now, we solve this relaxed problem and we project the result onto the nonconvex set $\mathcal{L}_q$. Using the Eckart-Young-Mirsky theorem [EY36], [Mir60, Theorem 2], we can deduce as a solution to (5.32) the matrix $\mathbf{W}_n^\star = a(\mathbf{S}_{1:L_\mathrm{R},1:L_\mathrm{R}}^{(n)})^{\frac{1}{2}}(\mathbf{U}_{1:N_\mathrm{R},1:L_\mathrm{R}}^{(n)})^H$ for $a > 0$ such that the constraint is satisfied with equality. Here, $\mathbf{R}_n = \mathbf{U}^{(n)}\mathbf{S}^{(n)}(\mathbf{U}^{(n)})^H \succcurlyeq \mathbf{0}$, in which $\mathbf{S}^{(n)}$ is a diagonal matrix of singular values of $\mathbf{R}_n$ arranged in nonascending order. Because the (generalized) projection

$$P_{\mathcal{L}_q}(\mathbf{W}) = \exp\left(j2^{1-q}\pi\left\lfloor\frac{\phi_{ik}^{(n)}}{2\pi}2^q + \frac{1}{2}\right\rfloor\right), \qquad \phi_{ik}^{(n)} = \angle W_{ik} \tag{5.33}$$

of $\mathbf{W} \in \mathbb{C}^{L_\mathrm{R}\times N_\mathrm{R}}$ onto $\mathcal{L}_q$ is independent of $|W_{ik}|$, a projection of $\mathbf{W}_n^\star$ onto $\mathcal{L}_q$ is given by

$$\mathbf{B}_n = P_{\mathcal{L}_q}\left(\left(\mathbf{U}_{1:N_\mathrm{R},1:L_\mathrm{R}}^{(n)}\right)^H\right). \tag{5.34}$$

By setting the phase shifters according to (5.34), we can expect to improve the SNR at the RF chains given that a good estimate of the channel covariance matrix is available; i.e., given that $\left\|\mathbf{R}_n - \mathbb{E}\left[\sum_{k=1}^{N_\mathrm{F}} \mathbf{H}_{n,k}\mathbf{H}_{n,k}^H\right]\right\|_F$ is sufficiently small. To achieve this, we implement a simple trade-off between exploration and exploitation. At each time instant n, we either draw a random analog combiner from a uniform distribution over $\mathcal{L}_q$ (with

[2] Since the projection onto the nonconvex set $\mathcal{L}_q$ in (5.33) only depends on the phase of its argument, the final result in (5.34) is independent of the scaling factor r. Thus we can set $r = 1$ without loss of generality.

probability $p_{\text{explore}} \in [0,1]$)), or we use the analog combiner in (5.34) (with probability $1 - p_{\text{explore}}$).

5.3.7. Summary of the Proposed Channel Estimation Scheme

The propopsed online channel estimation and tracking scheme with data-driven analog combining and phase and delay compensation is summarized in Algorithm 4 below. Here,

Algorithm 4 Online Channel Estimation and Tracking for Hybrid Beamforming Architectures

Parameters: p_{explore}, α, $(\mu_n)_{n\in\mathbb{N}}$, $(\mathcal{T}_n)_{n\in\mathbb{N}}$, $\eta \in (0,1]$, $\{(T_n)_{n\in\mathbb{N}} = (T_n^{\mathcal{A}})_{n\in\mathbb{N}}, z\}$ or $\{(T_n)_{n\in\mathbb{N}} = (\overline{T}_n)_{n\in\mathbb{N}}, \lambda, \nu\}$, t
Initialization: $\mathbf{R}_0 = \mathbf{I}_{N_{\text{R}}}$, $\hat{\mathbf{h}}_0 \in \mathbb{C}^{N_{\text{R}}N_{\text{T}}N_{\text{F}}}$
for $n \in \mathbb{N}$ **do**
  Draw $b_n \sim \mathcal{U}(0,1)$
  if $b_n \leq p_{\text{explore}}$ **then**
    Draw $\mathbf{B}_n$ from a uniform distribution over $\mathcal{L}_q$ ▷ exploration
  else
    Compute $\mathbf{B}_n$ according to (5.34) ▷ exploitation
  end if
  Predict $\hat{\mathbf{Y}}_n$ according to (5.21)
  Measure $\tilde{\mathbf{Y}}_n$
  Estimate $(\hat{\varphi}_n, \hat{\tau}_n)$ according to (5.25)
  Define $\mathcal{S}_n$ using $\mathbf{Y}_n = \tilde{\mathbf{Y}}_n \left(\mathbf{Q}(\hat{\varphi}_n, \hat{\tau}_n)\right)^H$
  Compute $\hat{\mathbf{h}}_{n+1} = T_n^t(\hat{\mathbf{h}}_n)$ with $T_n = T_n^{\mathcal{A}}$ (5.11) or $T_n = \overline{T}_n$ (5.19)
  Compute $\mathbf{R}_{n+1}$ according to (5.31)
  Output channel estimate $\hat{\mathbf{H}}_{n+1}\mathbf{Q}(\varphi_n, \tau_n) \approx \tilde{\mathbf{H}}_n$
end for

$p_{\text{explore}} \in [0,1]$ is the exploration probability, $\alpha \in [0,1]$ is the learning rate in (5.31), $(\mu_n)_{n\in\mathbb{N}} \subset [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$ is the sequence of relaxation parameters in (5.11) or (5.19), the sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ indicates the sets time intervals used in the cost function at time interval n, and $\eta \in (0,1]$ determines the weighting in (5.10). For the original APSM $T_n^{\mathcal{A}}$ in (5.11), $z > 0$ denotes the size of the ℓ_1-norm constraint in (5.13). For the heuristic $\overline{T}_n$ in (5.19), $\lambda > 0$ and $\nu > 0$ denote the scaling factor of the superiorization objective in (5.16) and the stabilization constant in (5.18), respectively. The positive integer t determines the number of algorithmic updates computed in each time interval.

5.4. Numerical Results

In this section, we evaluate the performance of the proposed methods using simulated channel matrices obtained with the Quasi Deterministic Radio Channel Generator (QuaDRiGa)

[Fra19]. This MATLAB-based simulator generates spatially consistent channels that reflect the spatio-temporal correlations typical of wireless channels. The simulation results are averaged over 200 scenarios according to Table 5.4, divided evenly into line-of-sight and non-line-of-sight scenarios. In each scenario, the UE orientation and the direction of the track are drawn uniformly at random. The starting positions of the track are chosen such that the angles of arrival are uniformly distributed over the interval $[-60°, 60°]$, and the distances to the BS are uniformly distributed over the interval $[50\,\mathrm{m}, 500\,\mathrm{m}]$. For each scenario, a tensor of 100 channel matrices is generated and scaled such that the average power of all elements is equal to one.

Table 5.1.: Simulation Parameters

BS antenna height	25 m
UE antenna height	1.5 m
UE speed	30 km/h
Scenario	`BERLIN_UMa_LOS`; `BERLIN_UMa_NLOS`
Carrier frequency	2.53 GHz
Interval duration	250 μs
BS antenna	$N_\mathrm{R} = 64$, 4×8 dual polarized [3GP17]
UE antenna	$N_\mathrm{T} = 4$, ULA
Analog combiner	$q = 3$ bit phase shifters, fully-connected
RF chains	$L_\mathrm{R} = 16$

In Subsection 5.4.1, we compare the performance of the proposed algorithms with the baseline in (5.5). To achieve a fair comparison, we consider the narrowband signal model in (5.4) which is free from random delay and phase variations. Subsection 5.4.2 compares the variance of the phase and delay estimators proposed in Subsection 5.3.5 to the lower bounds in 5.7. In Subsection 5.4.3 we investigate the performance of the proposed algorithms when applied to the wideband signal model in (5.3) with randomly varying delay and phase shift.

For the simulations in Subsections 5.4.1 and 5.4.3 we define

$$\mathrm{SNR}_\mathrm{dB} := 10 \log_{10} \left(\frac{\mathbb{E}\left[\| \mathbf{H}_{n,k} \mathbf{x}_{n,k} \|_2^2 \right]}{N_\mathrm{R} \sigma^2} \right),$$

where the expectation is with respect to transmit signals and subcarriers. This definition of the SNR ensures that the noise power is independent of the analog combiner $\mathbf{B}_n$. The performance of algorithms is assessed via the normalized mean squared error (NMSE)

$$\mathrm{NMSE}\left(\hat{\mathbf{h}}_n \right) = \frac{\| \hat{\mathbf{h}}_n - \mathbf{h}_n \|^2}{\| \mathbf{h}_n \|^2}$$

In each time interval, the UE transmits pilot signals $(\forall k \in \mathcal{I}_{\mathrm{sc}})$ $\mathbf{x}_{n,k} \in \mathbb{C}^{N_{\mathrm{T}}}$ with unit magnitude entries, the phase of which is drawn i.i.d. from $\mathcal{U}(0, 2\pi)$. The receive signal $\mathbf{y}_{n,k}$ is determined by the respective signal model.

5.4.1. Narrowband Channels

Although the signal model in (5.1) does not satisfy the block fading assumption, it is free from random delay and phase variations, i.e., the sequence of channels is temporally consistent. In the following simulation, we use this signal model to compare the batch method in (5.5) with the proposed online channel estimation algorithms. As the baseline in (5.5) assumes narrowband channels, we set $N_{\mathrm{F}} = 1$, resulting in the signal model in (5.4). Figure 5.1 shows the NMSE for $\mathrm{SNR}_{\mathrm{dB}} = 20\,\mathrm{dB}$ averaged over all 200 scenarios.

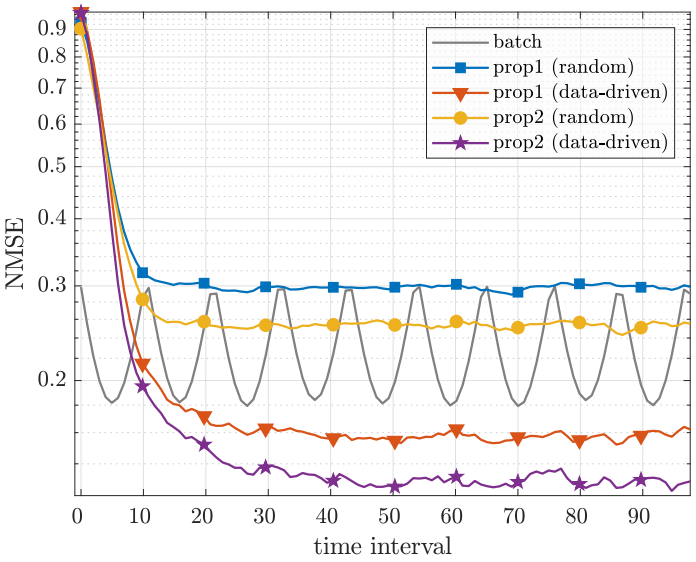


Figure 5.1.: NMSE for different methods averaged over 200 scenarios

The curve labeled `batch` refers to the solution to (5.5) obtained with a standard interior point solver. We use a batch size of $N_{\mathrm{b}} = 11$ and a regularization parameter $\lambda_{\mathrm{b}} = 1.5$. These parameters have been found to achieve the lowest average NMSE using grid search. The update rules defined by the mappings $(T_n^{\mathcal{A}})^t$ and $\overline{T}_n^t$ in (5.11) and (5.19) are denoted by `prop1` and `prop2`, respectively. For these methods, the relaxation parameter is chosen as $(\forall n \in \mathbb{N})$ $\mu_n := 1$ and the buffer size in (5.9) is $|\mathcal{T}_n| = \min(n, 5)$, i.e., the cost function in (5.9) takes into account the 5 most recent measurements. The weighting in (5.10) is defined by $\eta = 0.8$ and for each time int $t = 5$, i.e., each of the respective mappings are applied $t = 5$ times in each time interval n. For `prop1`, the size of the ℓ_1-norm ball in (5.13) is set to

$z = 30$, which was found to achieve the best average NMSE. For `prop2`, the regularization parameter of the weighted ℓ_1-norm is set to $\lambda = 0.003$, and the stabilization constant in (5.18) is $\nu = 0.05$. For the proposed online methods, two analog combining schemes are considered. Curves labeled `random` correspond to the case in which the entries of $\mathbf{B}_n$ are drawn i.i.d. from a uniform distribution over $\mathcal{L}_q$, whereas curves labeled `data-driven` correspond to the combining scheme proposed in Subsection 5.3.6, where $p_{\text{explore}} = 0.15$ and $\alpha = 0.9$. In both cases, the resolution of the digitally controlled phase shifters is set to $q = 3\,\text{bit}$.

Because the `batch` method produces a single channel estimate for the entire batch, the NMSE increases towards beginning and end of each batch as the true channel varies continuously. For the random analog combining policy, `prop1` and `prop2` incur a slightly higher NMSE than the `batch` method. The proposed data-driven analog combining policy considerably reduces the NMSE to the random policy for both `prop1` and `prop2`. Moreover, independently of the analog combining policy, the heuristic `prop2` using the weighted ℓ_1-norm achieves a lower NMSE than the original APSM `prop1`. The proposed methods with data-driven analog combining outperform the baseline. However, none of the algorithms achieves an NMSE below 0.16. On average, proposed online algorithms required less than 4 % of the computation time required by the batch method in our simulations. This is despite the fact that the batch method needs to solve only one optimization problem every N_b time intervals, and that the proposed online algorithms were not implemented on GPUs. We note that the estimation error of all methods can be reduced by decreasing the speed of the UE or by reducing the time interval duration τ. As will be shown in Subsection 5.4.3, the performance of the proposed algorithms can also be improved by jointly processing multiple subcarriers.

5.4.2. Performance of the Proposed Delay and Phase Estimators

In the following, we compare the variance of the estimators proposed in Section 5.3.5 to the respective CRLBs. The simulations are performed in the simplified system model in Proposition 5.6, where the actual receive signal $\tilde{\mathbf{Y}}_n$ is a noisy, delayed, and phase-shifted version of the predicted receive signal $\hat{\mathbf{Y}}_n$, and the entries of $\hat{\mathbf{Y}}_n$ are drawn i.i.d. from a complex Gaussian distribution. Since the phase is periodic with period 2π, we define phase error as

$$\varepsilon_\varphi(\varphi, \hat{\varphi}) = \min_{i \in \mathbb{Z}} |\hat{\varphi} - \varphi - 2\pi i|^2.$$

This definition avoids the dependency of the estimation error on the absolute value of φ, as for φ close to $-\pi$, small deviations can result in an extremely large the absolute error $|\varphi - \hat{\varphi}|$. Moreover, we note that the proposed method cannot resolve delays above one

symbol duration (i.e. $\tau > 1$). Since we assume that the random delays are restricted to the interval $[0, 1)$, we can define the delay estimation error in the same way.

$$\varepsilon_\tau(\tau, \hat{\tau}) = \min_{i \in \mathbb{Z}} |\hat{\tau} - \tau - i|^2.$$

Note that the restriction of φ and τ to the intervals $[0, 2\pi)$ and $[0, 1)$, respectively, do not limit the applicability of the proposed method for channel tracking, as adding integer multiples of the respective period does not change the reconstructed channel coefficients. Figure 5.2 shows the unbiased sample variance

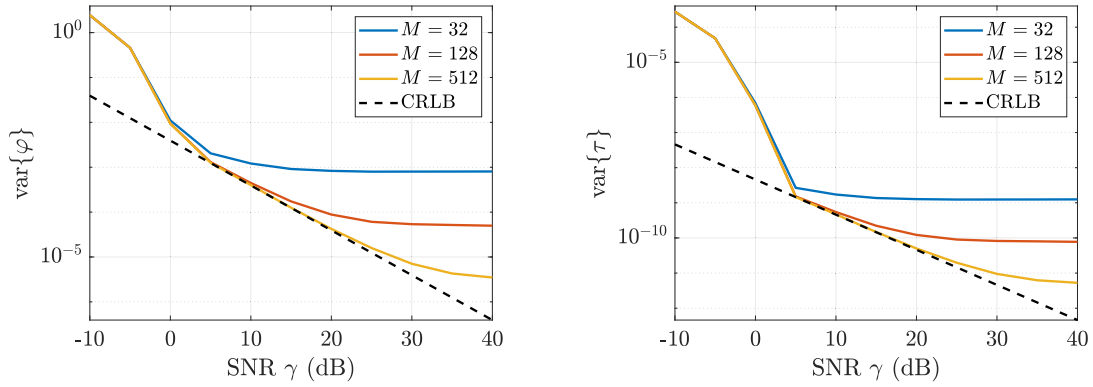


Figure 5.2.: Performance of the proposed (a) phase and (b) delay estimators in a system with $N_\mathrm{F} = 256$ subcarriers, $|\mathcal{P}| = 32$ pilot subcarriers, $L_\mathrm{R} = 16$ RF chains, and upsampling factors $M \in \{32, 128, 512\}$.

$$\mathrm{var}\{\varphi\} = \frac{1}{K-1} \sum_{n=1}^{K} \varepsilon_\varphi(\varphi_n, \hat{\varphi}_n) \qquad \text{and} \qquad \mathrm{var}\{\tau\} = \frac{1}{K-1} \sum_{n=1}^{K} \varepsilon_\tau(\tau_n, \hat{\tau}_n)$$

of the phase and delay estimators for $K = 10000$ random realizations according to the signal model in (5.6) for a system with $N_\mathrm{F} = 256$ subcarriers, $|\mathcal{P}| = 32$ pilot subcarriers and $L_\mathrm{R} = 16$ RF chains as a function of the SNR. The set of pilot subcarriers used in the simulation comprises eight equidistant blocks of four adjacent subcarrier indices each, i.e., $\mathcal{P} = \{15, 16, 17, 18, 47, 48, \ldots, 210, 239, 240, 241, 242\} \subset \mathcal{I}_\mathrm{sc}$. In the high SNR regime, both estimators achieve a variance close to the CRLB for sufficiently large upsampling factors. With an upsampling factor of $M = 512$ both estimators nearly achieve the CRLB for SNRs below 20 dB. Clearly, precision can be traded for computational complexity in both cases.

Although the proposed scheme can be used with any set $\mathcal{P} \subset \mathcal{I}_{\mathrm{sc}}$ of pilot subcarrier indices, we note that the choice of $\mathcal{P}$ has a significant impact on the performance of the proposed estimators. In particular, an equidistant pilot subcarrier spacing results in a periodically repeating cross-correlation vector $\mathbf{r}$, which renders delay estimation ambiguous.

5.4.3. Wideband Channels

The goal of this subsection is to evaluate the performance of the proposed scheme for the wideband signal model in (5.3), in which the delay and phase shift vary randomly between subsequent time intervals. Owing to this random variation, the batch method in (5.5) can not be applied in this case so it is excluded from the comparison.

We compare the NMSE of the proposed methods in the same simulation setting considered in Subsection 5.4.1, with the only differences being that we increase the number of subcarriers to $N_{\mathrm{F}} = 256$, and that we introduce random phase and delay variations to the channel in subsequent time intervals. The proposed algorithms can be readily applied to this setting, using the phase and delay compensation scheme in Subsection 5.3.5. Here, we use an upsampling factor of $M = 512$, which achieved close to optimal performance in the simulation in Subsection 5.4.2. As the increased number of subcarriers results in an increased ℓ_1-norm of the true channel, we change the ℓ_1-norm constraint of `prop1` to $z = 2000$, which again was found to achieve good average performance for $N_{\mathrm{F}} = 256$ via grid search. For `prop2`, we use the exact same parameters as in Subsection 5.4.1.

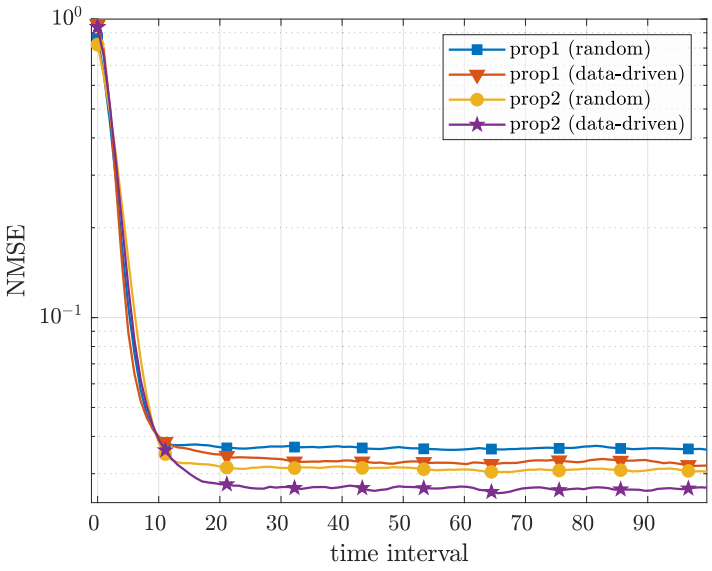


Figure 5.3.: NMSE for different methods averaged over 200 scenarios

As can be seen in Figure 5.3, all methods achieve a considerably lower NMSE than in the narrowband setting. As before, the heuristic in (5.19) performs slightly better than the original APSM in (5.11), and for both methods, the data-driven analog combining policy outperforms the random policy. It is interesting to note that no parameter changes were necessary for the heuristic in (5.19) even though the number of subcarriers was increased by a factor of 256. Whereas none of the methods reached an NMSE below 0.16 in the narrowband setting, all online algorithms settle at an NMSE below 0.04 within only 10 time intervals for $N_{\mathrm{F}} = 256$. This shows that the estimation error can be reduced by jointly performing channel estimation over all subcarriers, which allows to exploit sparsity of the channel in both angular domain and delay domain. We note that this performance improvement is despite the fact that random delays and phase shifts are introduced in the simulation in Figure 5.3, rendering the problem more challenging, or even infeasible for batch methods. Moreover, joint estimation over all subcarriers using batch methods based on general-purpose solvers is likely to be prohibitive even for constant channels, due to the problem dimension of $N = N_{\mathrm{R}} N_{\mathrm{T}} N_{\mathrm{F}} = 65536$ complex variables.

5.5. Conclusion

In this chapter, we used the APSM in Theorem 2.3 to devise online algorithms for estimating and tracking time-varying wideband channels with hybrid-beamforming architectures. We introduced a phase and delay estimation scheme to compensate for random delay and phase variations, and we compared its performance to the theoretical lower bound. Furthermore, we proposed a data-driven analog combining policy that leverages the online nature of the proposed channel estimation technique to improve its performance. Simulations in a narrowband setting show that the proposed algorithms can outperform existing batch methods. Moreover, the proposed algorithms can exploit channel sparsity in the time domain by jointly processing multiple subcarriers. The numerical results show that jointly estimating the entire wideband channel reduces considerably the estimation error compared to the narrowband setting. While the parameters of the unperturbed APSM algorithm with ℓ_1-norm constraint need to be adapted depending on the number of subcarriers, the proposed heuristic, which is inspired by superiorization, can be applied to a wide range of subcarriers without changing its parameters.

6. PAPR Reduction

In this chapter, we consider the PAPR problem in OFDM systems. We present algorithmic solutions that were originally proposed in [1]. The content of this chapter is an extended version of [1] that regards heuristics proposed in [1] in light of the theoretical results derived in Chapter 2. In particular, we propose PAPR reduction techniques based on the POCS algorithm in Theorem 2.1, the GPR algorithm in 2.4, and the EAPM in Theorem 2.5. Some passages of the following text are borrowed from [1] without changes.

6.1. Introduction

In OFDM systems with a large number of subcarriers, the time domain transmit signals are known to exhibit a high PAPR [WFB+13, KJ03]. Because the constellation points in the frequency domain constitute random variables, their weighted sum (represented by the Fourier transform) causes the time-domain samples to approach a Gaussian distribution as the number of subcarriers increases [KJ03]. Amplifying an OFDM signal with N subcarriers without introducing nonlinear distortions requires a power amplifier with a linear region N times as large as the average power [WFB+13]. Since power amplifiers with a high dynamic range are inefficient in terms of both energy consumption and manufacturing costs, operating the power amplifier purely in the linear region typically results in a low SNR at the receiver. By contrast, operating the amplifier in the nonlinear region introduces undesired nonlinear signal distortions, which equally reduce the channel capacity [SA09].

To mitigate the PAPR problem, an enormous research effort has been directed towards the development of practical PAPR reduction techniques, including basic approaches such as coding, interleaving, or selective mapping [RM13]. In this work, we focus on set-theoretic approaches to the PAPR reduction problem because they open up the door to the development of flexible, low-complexity PAPR reduction techniques that can be easily implemented in state-of-the-art OFDM systems with large numbers of subcarriers and very strict latency requirements. The basic idea behind many approaches to the PAPR reduction problem is to modify the waveform of the transmission signal while satisfying certain constraints such as spectral mask and error vector magnitude (EVM) constraints.

In general, these constraints depend on parameters that can change from symbol to symbol (e.g., allocation of pilot and data subcarriers or the constellation size might change between subsequent OFDM symbols), so there is a strong need for PAPR reduction techniques that have enough flexibility to systematically incorporate fast-changing constraints. Against this background, set-theoretic approaches provide a natural framework for dealing with the PAPR reduction problem in modern OFDM systems.

The contribution of the work in this chapter can be summarized as follows. First, we develop a set-theoretic framework for implementing PAPR reduction algorithms that can flexibly combine different frequency-domain constraints and adapt constraints on a time scale of two consecutive OFDM symbols. In particular, we propose a PAPR reduction algorithm for tone reservation (TR) [TC98, CY09] and iterative clipping and filtering (ICF) [Arm02] (affine/linear frequency constraints) based on the EAPM in Theorem 2.5. In addition, we propose a generalization of this algorithm to non-affine frequency-domain constraints based on the GPR approach in Theorem 2.4. The proposed generalization can deal with arbitrary combinations of constraints related to EVM, spectral masks, active constellation extension (ACE) [KJ03], and compensation carriers. Moreover, we present a heuristic extension, which is motivated by superiorization, to incorporate a nonconvex equality constraint on the magnitude of compensation subcarriers. In this way, these subcarriers can be used simultaneously for peak compensation and for channel estimation with phaseless pilots [WBJ15a, WBJ15b].

6.2. System Model and Basic Definitions

Let $\underline{\mathbf{c}} \in \mathbb{C}^N$ be an OFDM symbol with N subcarriers in the frequency domain, each of them containing a complex valued constellation point. Since peaks in the time-domain signal, which is obtained by applying the inverse discrete Fourier transform to $\underline{\mathbf{c}}$, can increase owing to digital-to-analog conversion, we typically approximate the analog signal by digitally upsampling with a factor of $L \geq 4$ [SGAK03]. By doing so, we gain information about the magnitude of peaks in the analog time-domain signal. More precisely, upsampling can be captured by defining a zero-padded vector

$$\hat{\underline{\mathbf{c}}} = \left[\underline{\mathbf{c}}^T, \underline{\mathbf{0}}^T\right]^T \in \mathbb{C}^{NL}$$

and a DFT matrix $\underline{\mathbf{F}} \in \mathbb{C}^{NL \times NL}$ with entries

$$\underline{F}_{k,l} = \frac{1}{\sqrt{NL}} e^{-j2\pi \frac{\nu_k t_l}{NL}},$$

where $(\forall k, l \in \{1, \ldots, NL\})$ $\nu_k = k - N/2 - 1$, and $t_l = l - 1$. The digital approximation of the continuous time-domain signal is thus given by $\underline{\mathbf{F}}^H \hat{\underline{\mathbf{c}}}$. Note that $\underline{\mathbf{F}}$ is normalized such that $\underline{\mathbf{F}}\underline{\mathbf{F}}^H = \underline{\mathbf{F}}^H\underline{\mathbf{F}} = \mathbf{I}$.

A measure commonly used to express the distortion on the data subcarriers of a modified version $\underline{\mathbf{x}}$ of a given OFDM symbol $\hat{\underline{\mathbf{c}}}$ is the EVM, defined by

$$\epsilon(\underline{\mathbf{x}}) = \sqrt{\frac{\sum_{k \in \mathcal{I}_\mathrm{d}} |\underline{x}_k - \underline{c}_k|^2}{N_\mathrm{d} P_0}},$$

where $\mathcal{I}_\mathrm{d} = \{i_1, \ldots, i_{N_\mathrm{d}}\}$ is the set of data subcarrier indices. In real transceivers, distortions potentially increasing the EVM or bit error ratio (BER) are mainly introduced by peaks exceeding the linear regions of the amplifiers. Therefore, another important signal property of a frequency-domain OFDM symbol $\underline{\mathbf{x}}$ is the PAPR, defined as [KJ04, Erd06, CY09]

$$\mathrm{PAPR}(\underline{\mathbf{x}}) = \frac{\|\underline{\mathbf{F}}^H \underline{\mathbf{x}}\|_\infty^2}{\frac{1}{NL}\mathbb{E}\left[\|\mathbf{S}_\mathrm{d}\underline{\mathbf{c}}\|_2^2\right]}, \tag{6.1}$$

where $\mathbf{S}_\mathrm{d} = [\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{N_\mathrm{d}}}]^T \in \mathbb{R}^{N_\mathrm{d} \times N}$ is a row selection matrix for the rows indexed by $\mathcal{I}_\mathrm{d}$. In the following, we assume that expected power of all data subcarriers $i \in \mathcal{I}_\mathrm{d}$ is normalized to $E\left(|\underline{c}_i|^2\right) = P_0$, so that (6.1) simplifies to

$$\mathrm{PAPR}(\underline{\mathbf{x}}) = \tfrac{NL}{N_\mathrm{d} P_0}\|\underline{\mathbf{F}}^H \underline{\mathbf{x}}\|_\infty^2. \tag{6.2}$$

Amplifier non-linearities affect signals with high PAPR more severely, so PAPR reduction methods allow for the deployment of less costly amplifiers. These methods are the topic of the next sections.

6.3. PAPR Reduction by Set Theoretic Estimation

To apply theoretical results for real Hilbert spaces presented in Chapter 2, we pose the PAPR problem in a real Hilbert space $(\mathcal{H} = \mathbb{R}^{2NL}, \langle\cdot,\cdot\rangle)$ equipped with the standard Euclidean inner product

$$(\forall \mathbf{x}, \mathbf{y} \in \mathcal{H}) \quad \langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^T \mathbf{x}$$

inducing the Euclidean norm $(\forall \mathbf{x} \in \mathcal{H})$ $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \|\mathbf{x}\|_2$. To represent complex signal vectors $\underline{\mathbf{x}} \in \mathbb{C}^{NL}$ using elements $\mathbf{x} \in \mathcal{H}$ of this real Hilbert space, we introduce a mapping $\xi : \mathbb{C}^{NL} \to \mathcal{H}$ given by

$$(\forall \underline{\mathbf{x}} \in \mathbb{C}^{NL}) \quad \xi(\underline{\mathbf{x}}) := \left[\mathrm{Re}\{\underline{\mathbf{x}}^T\}, \mathrm{Im}\{\underline{\mathbf{x}}^T\}\right]^T,$$

and a matrix

$$\mathbf{F} := \begin{bmatrix} \mathrm{Re}\{\underline{\mathbf{F}}\} & -\mathrm{Im}\{\underline{\mathbf{F}}\} \\ \mathrm{Im}\{\underline{\mathbf{F}}\} & \mathrm{Re}\{\underline{\mathbf{F}}\} \end{bmatrix} \in \mathbb{R}^{2NL \times 2NL},$$

such that $\mathbf{F}^T\mathbf{F} = \mathbf{F}\mathbf{F}^T = \mathbf{I}$ and $(\forall \underline{\mathbf{x}} \in \mathbb{C}^{NL})\ \mathbf{F}\xi(\underline{\mathbf{x}}) = \xi(\underline{\mathbf{F}}\underline{\mathbf{x}})$. Moreover, we define $\hat{\mathbf{c}} := \xi(\underline{\hat{\mathbf{c}}})$. The maximum norm of a complex vector $\underline{\mathbf{v}} \in \mathbb{C}^{NL}$ is given by

$$\|\underline{\mathbf{v}}\|_\infty = \max_{k \in \{1,\ldots,NL\}} \sqrt{(\mathrm{Re}\{\underline{v}_k\})^2 + (\mathrm{Im}\{\underline{v}_k\})^2}.$$

Therefore we define a function[1] $\|\cdot\|_{\mathcal{H},\infty} : \mathcal{H} \to \mathbb{R}_+$ by $(\forall \mathbf{v} \in \mathcal{H})$

$$\|\mathbf{v}\|_{\mathcal{H},\infty} := \max_{k \in \{1,\ldots,NL\}} \sqrt{v_k^2 + v_{k+NL}^2}, \tag{6.3}$$

so that $(\forall \underline{\mathbf{v}} \in \mathbb{C}^{NL})\ \|\xi(\underline{\mathbf{v}})\|_{\mathcal{H},\infty} = \|\underline{\mathbf{v}}\|_\infty$.

6.3.1. Problem Statement

The idea of set theoretic PAPR reduction is to minimize time-domain peaks subject to constraints specifying certain desirable signal properties in the frequency-domain, such as the maximal EVM or spectral masks. If these constraint sets can be expressed by a closed convex set $\mathcal{F} \subset \mathcal{H}$, PAPR reduction can be posed in terms of a constrained convex minimization problem the objective of which is given by (6.2). Since scaling the objective function does not change the solution set, the feasible vector with lowest PAPR can be found by solving the problem

$$\begin{aligned} \underset{\mathbf{x} \in \mathcal{H}}{\text{minimize}} \quad & \|\mathbf{F}^T\mathbf{x}\|_{\mathcal{H},\infty} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{F} \end{aligned} \tag{6.4}$$

A solution $\mathbf{x}^\star$ to the Problem (6.4) can be obtained with interior point methods or projected subgradient methods, but the complexity of this approach may be prohibitive in real-time applications. Therefore, several simple suboptimal techniques have been proposed. Examples include ICF [Arm02], TR [TC98, CY09], or ACE methods [KJ03]. The idea of these techniques is to reduce the objective value in (6.4) to a pre-defined threshold value $\theta \geq \|\mathbf{F}^T\mathbf{x}^\star\|_{\mathcal{H},\infty}$. Formally, these methods replace the problem in (6.4) by the feasibility problem

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{find}}\ \mathbf{x} \in \mathcal{F} \cap \mathcal{T} \tag{6.5}$$

[1] It is straightforward to verify that the function in (6.3) satisfies all axioms of a norm on $\mathcal{H}$.

where $\mathcal{T} = \left\{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{F}^T\mathbf{x}\|_{\mathcal{H},\infty} \leq \theta\right\}$ is the set of signals bounded by magnitude θ in the time domain. All of the aforementioned techniques can be related to this problem by choosing the constraint set $\mathcal{F}$ accordingly. If the projection onto $\mathcal{F}$ is simple to compute, Problem (6.5) can be solved with projection methods [SY98], which only involve operations with low computational complexity, while typically achieving much of the progress towards the solution during the initial iterations [CCC+12].

6.3.2. Some Relevant Constraint Sets

This section shows how the set $\mathcal{F} \subset \mathcal{H}$ can be constructed as the intersection of multiple constraint sets, the projections onto which are simple to compute. To this end, we define exemplary realizations of constraint sets that are commonly used in the literature. To streamline the notation, we will use the following convention for finite sets indexing certain frequency bins. If $\mathcal{I} \subset \{1, \ldots, NL\}$ is a set of frequency bin indices of a vector $\underline{\mathbf{v}} \in \mathbb{C}^{NL}$, we use the superscript $\mathcal{H}$ to denote a finite set indexing the corresponding entries of a vector $\mathbf{v} = \xi(\underline{\mathbf{v}}) \in \mathcal{H}$ by defining

$$\mathcal{I}^{\mathcal{H}} = \mathcal{I} \cup \{k \in \mathbb{N} \mid k - NL \in \mathcal{I}\}.$$

Subspace of in-band signals

As the modified OFDM signal is downsampled to the original rate (with subcarriers indexed by the set $\mathcal{I}_{\text{in}} = \{1, \ldots, N\}$) before transmission, the out-of-band radiation (i.e nonzero values in frequency bins exceeding the original bandwidth) caused by clipping needs to be removed. Therefore, the transmit signal should be restricted to the subspace

$$\mathcal{C}_1 := \left\{\mathbf{x} \in \mathcal{H} \;\middle|\; (\forall k \notin \mathcal{I}_{\text{in}}^{\mathcal{H}}),\ x_k = 0\right\}$$

of in-band signals. The projection of a point $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{C}_1$ can be written entry-wise as

$$(\forall k \in \{1, \ldots, 2NL\}) \quad P_{\mathcal{C}_1}(\mathbf{x})|_k = \begin{cases} x_k & \text{if } k \in \mathcal{I}_{\text{in}}^{\mathcal{H}} \\ 0 & \text{otherwise.} \end{cases}$$

Affine subspace of compensation signals

In the TR method, a subset $\mathcal{I}_{\text{c}} \subset \mathcal{I}_{\text{in}}$ of the subcarriers is not used for data transmission. These subcarriers transmit dummy symbols that have the sole purpose of decreasing peaks in the time domain [TC98]. Formally, the corresponding frequency-domain constraint set

restricts all but the compensation subcarriers $\mathcal{I}_\mathrm{c}$ to their original values, by defining

$$\mathcal{C}_2 := \left\{ \mathbf{x} \in \mathcal{H} \;\middle|\; (\forall k \in \mathcal{I}_\mathrm{in}^\mathcal{H} \setminus \mathcal{I}_\mathrm{c}^\mathcal{H}),\ x_k = \hat{c}_k \right\}.$$

The projection of a point $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{C}_2$ can be expressed entry-wise as

$$(\forall k \in \{1, \ldots, 2NL\}) \quad P_{\mathcal{C}_2}(\mathbf{x})|_k = \begin{cases} \hat{c}_k & \text{if } k \in \mathcal{I}_\mathrm{in}^\mathcal{H} \setminus \mathcal{I}_\mathrm{c}^\mathcal{H} \\ x_k & \text{otherwise.} \end{cases}$$

EVM constraint set

Besides exclusively reserving bandwidth, a degree of freedom for peak cancellation can also be obtained by distorting the data subcarriers. For simplicity, we define $\mathcal{I}_\mathrm{d} = \mathcal{I}_\mathrm{in} \setminus \mathcal{I}_\mathrm{c}$. The set of signals with distortion on the data subcarriers bounded by the maximal EVM ε is given by [AM03]

$$\mathcal{C}_3 := \left\{ \mathbf{x} \in \mathcal{H} \;\middle|\; \sum_{k \in \mathcal{I}_\mathrm{d}^\mathcal{H}} (x_k - \hat{c}_k)^2 \leq N_\mathrm{d} P_0 \varepsilon^2 \right\}.$$

The projection of a point $\mathbf{x} \notin \mathcal{H}$ onto $\mathcal{C}_3$ is given entry-wise by [SY98, Theorem 3.4-1] $(\forall k \in \{1, \ldots, 2NL\})$

$$P_{\mathcal{C}_3}(\mathbf{x})|_k = \begin{cases} \hat{c}_k + \dfrac{\varepsilon\sqrt{N_\mathrm{d} P_0}}{\sqrt{\sum_{k \in \mathcal{I}_\mathrm{d}^\mathcal{H}} (x_k - \hat{c}_k)^2}} (x_k - \hat{c}_k) & \text{if } k \in \mathcal{I}_\mathrm{d}^\mathcal{H} \text{ and } \mathbf{x} \notin \mathcal{C}_3, \\ x_k & \text{otherwise.} \end{cases}$$

More compactly, we can express the projection of a point $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{C}_3$ as

$$P_{\mathcal{C}_3}(\mathbf{x})|_k = \begin{cases} \hat{c}_k + \rho(\mathbf{x})(x_k - \hat{c}_k) & \text{if } k \in \mathcal{I}_\mathrm{d}^\mathcal{H}, \\ x_k & \text{otherwise,} \end{cases} \tag{6.6}$$

where the function $\rho : \mathcal{H} \to [0, 1]$ is given by

$$\rho(\mathbf{x}) = \begin{cases} \left(\max \left(\dfrac{\sqrt{\sum_{k \in \mathcal{I}_\mathrm{d}^\mathcal{H}} (x_k - \hat{c}_k)^2}}{\varepsilon\sqrt{N_\mathrm{d} P_0}}, 1 \right) \right)^{-1} & \text{if } \varepsilon > 0 \\ 0 & \text{otherwise.} \end{cases}$$

ACE constraint set

The idea of active constellation extension is to compensate peaks by allowing boundary points of a square QAM constellation to be moved further outside, thereby increasing the margin to any other constellation point. While this increases the EVM, the BER can be reduced by this technique [KJ03]. The set of allowable modifications is denoted by

$$\mathcal{C}_4 := \bigcap_{k \in \mathcal{I}_{\mathrm{d}}^{\mathcal{H}}} \mathcal{C}_4^k,$$

where

$$(\forall k \in \mathcal{I}_{\mathrm{d}}^{\mathcal{H}}) \quad \mathcal{C}_4^k = \begin{cases} \{\mathbf{x} \in \mathcal{H} \mid \langle \mathbf{e}_k, \mathbf{x} \rangle = \hat{c}_k\}, & \text{if } |\hat{c}_k| \neq \gamma, \\ \{\mathbf{x} \in \mathcal{H} \mid \langle \operatorname{sgn}(\hat{c}_k)\mathbf{e}_k, \mathbf{x} \rangle \geq |\hat{c}_k|\}, & \text{otherwise.} \end{cases}$$

Here, $\mathbf{e}_k \in \mathcal{H}$ denotes the kth unit vector, and γ is a constant equal to the largest positive real part of all points of the respective QAM constellation. Consequently, for all $k \in \mathcal{I}_{\mathrm{d}}^{\mathcal{H}}$ the set $\mathcal{C}_4^k$ is either a hyperplane or a half space,

Proposition 6.1. *For all $k \in \mathcal{I}_{\mathrm{d}}^{\mathcal{H}}$, the projection of a point $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{C}_4^k$ is given by*

$$P_{\mathcal{C}_4^k}(\mathbf{x}) = \begin{cases} \mathbf{x} + (\hat{c}_k - x_k)\mathbf{e}_k & \text{if } \mathbf{x} \notin \mathcal{C}_4^k \\ \mathbf{x} & \text{otherwise.} \end{cases} \tag{6.7}$$

Proof. *For inner constellation points, i.e., if $|\hat{c}_k| \neq \gamma$, $\mathcal{C}_4^k$ is a hyperplane, the projection onto which is given by [SY98, Eq. (3.2-6)]* ($\forall \mathbf{x} \in \mathcal{H}$)

$$\begin{aligned} P_{\mathcal{C}_4^k}(\mathbf{x}) &= \mathbf{x} + \frac{\hat{c}_k - \langle \mathbf{e}_k, \mathbf{x} \rangle}{\|\mathbf{e}_k\|^2} \mathbf{e}_k \\ &= \mathbf{x} + (\hat{c}_k - x_k)\mathbf{e}_k. \end{aligned}$$

For outer constellation points, i.e., if $|\hat{c}_k| \neq \gamma$, $\mathcal{C}_4^k$ is a half space, the projection onto which is given by [SY98, Eq. (3.3-17)] ($\forall \mathbf{x} \in \mathcal{H}$)

$$\begin{aligned} P_{\mathcal{C}_4^k}(\mathbf{x}) &= \begin{cases} \mathbf{x} + \frac{|\hat{c}_k| - \langle \operatorname{sgn}(\hat{c}_k)\mathbf{e}_k, \mathbf{x} \rangle}{\|\operatorname{sgn}(\hat{c}_k)\mathbf{e}_k\|^2} \operatorname{sgn}(\hat{c}_k)\mathbf{e}_k & \text{if } \mathbf{x} \notin \mathcal{C}_4^k \\ \mathbf{x} & \text{otherwise.} \end{cases} \\ &= \begin{cases} \mathbf{x} + (\hat{c}_k - x_k)\mathbf{e}_k & \text{if } \mathbf{x} \notin \mathcal{C}_4^k \\ \mathbf{x} & \text{otherwise,} \end{cases} \end{aligned}$$

which completes the proof. □

In light of Proposition 6.1, it is easy to see that $(\forall \mathbf{x} \in \mathcal{H})\ \langle P_{\mathcal{C}_4^k}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_4^l}(\mathbf{x}) - \mathbf{x}\rangle = 0$, whenever $k \neq l$. Consequently, according to Proposition 1.1, we can compute the projection of a point $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{C}_4$ as

$$P_{\mathcal{C}_4}(\mathbf{x}) = \mathbf{x} + \sum_{k \in \mathcal{I}_{\mathrm{d}}^{\mathcal{H}}} \left(P_{\mathcal{C}_k^4}(\mathbf{x}) - \mathbf{x} \right). \tag{6.8}$$

Closed form expressions for the projections onto $\mathcal{F}$ and $\mathcal{T}$

As mentioned in Section 6.3.1, the solution to (6.5) can be found with projection methods if simple expressions for the projections onto $\mathcal{F}$ and $\mathcal{T}$ exist. Some potential definitions of the frequency domain constraint set $\mathcal{F}$ in (6.5), which have been used in previous studies, include the following:

- $\mathcal{F} = \mathcal{C}_1$ (iterative clipping and filtering) [Arm02]
- $\mathcal{F} = \mathcal{C}_1 \cap \mathcal{C}_2$ (tone reservation) [TC98]
- $\mathcal{F} = \mathcal{C}_1 \cap \mathcal{C}_4$ (active constellation extension) [KJ03], (if $\mathcal{I}_{\mathrm{c}} \neq \emptyset$ ACE + TR) [PWK09]

Further examples of constraint sets are given in [CY09]. For all combinations mentioned above, a closed form expression for the projection onto $\mathcal{F}$ exists. In Proposition 6.2, we derive a closed form expression for the projection onto the set $\mathcal{F} = \mathcal{C}_1 \cap \mathcal{C}_3 \cap \mathcal{C}_4$, which simultaneously enforces in-band-, EVM-, and ACE-constraints while allowing for compensation signals on subcarriers reserved for TR. To characterize the projection onto the intersection of EVM- and ACE constraints, which affect the same frequency bins, we will make use of the following lemma. Note that, in contrast to Proposition 1.1, the order in which the projections are performed is important.

Lemma 6.1. *Let $\mathcal{A} \subset \mathcal{H}$ and $\mathcal{B} \subset \mathcal{H}$ be two closed convex subsets of a real Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ such that $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. If $(\forall \mathbf{x} \in \mathcal{H})\ P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) \in \mathcal{A} \cap \mathcal{B}$ and $\langle P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{B}}(\mathbf{x}), P_{\mathcal{B}}(\mathbf{x}) - \mathbf{x}\rangle = 0$, then $P_{\mathcal{A}\cap\mathcal{B}} = P_{\mathcal{A}}P_{\mathcal{B}}$.*

Proof. *Fix $\mathbf{x} \in \mathcal{H}$ and define*

$$\mathcal{M}_{\mathcal{B}} := \{\mathbf{y} \in \mathcal{H} \mid \langle \mathbf{y} - P_{\mathcal{B}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{B}}(\mathbf{x})\rangle \leq 0\}$$
$$\mathcal{M}_{\mathcal{A}} := \{\mathbf{y} \in \mathcal{H} \mid \langle \mathbf{y} - P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x}), P_{\mathcal{B}}(\mathbf{x}) - P_{\mathcal{A}}P_{\mathcal{B}}(\mathbf{x})\rangle \leq 0\}.$$

The projection theorem in Fact 1.1 implies that $\mathcal{A} \subset \mathcal{M}_{\mathcal{A}}$ and $\mathcal{B} \subset \mathcal{M}_{\mathcal{B}}$, so $\mathcal{A} \cap \mathcal{B} \subset \mathcal{M}_{\mathcal{A}} \cap \mathcal{M}_{\mathcal{B}}$. By the premise of this lemma, the normal vectors of $\mathcal{M}_{\mathcal{A}}$ and $\mathcal{M}_{\mathcal{B}}$ are orthogonal, so $(\forall \mathbf{y} \in \mathcal{H})\ \langle P_{\mathcal{M}_{\mathcal{A}}}(\mathbf{y}) - \mathbf{y}, P_{\mathcal{M}_{\mathcal{B}}}(\mathbf{y}) - \mathbf{y})\rangle = 0$. Thus by Proposition 1.1,

$P_{\mathcal{M}_\mathcal{A} \cap \mathcal{M}_\mathcal{B}} = P_{\mathcal{M}_\mathcal{A}} P_{\mathcal{M}_\mathcal{B}}$. *Moreover, it holds by construction that* $P_{\mathcal{M}_\mathcal{A}} P_{\mathcal{M}_\mathcal{B}}(\mathbf{x}) = P_\mathcal{A} P_\mathcal{B}(\mathbf{x})$. *Consequently, given that* $(\forall \mathbf{x} \in \mathcal{H})\ P_\mathcal{A} P_\mathcal{B}(\mathbf{x}) \in \mathcal{A} \cap \mathcal{B}$, *we have that*

$$d(\mathbf{x}, P_\mathcal{A} P_\mathcal{B}(\mathbf{x})) \geq d(\mathbf{x}, \mathcal{A} \cap \mathcal{B}) \geq d(\mathbf{x}, \mathcal{M}_\mathcal{A} \cap \mathcal{M}_\mathcal{B}) = d(\mathbf{x}, P_\mathcal{A} P_\mathcal{B}(\mathbf{x})),$$

which shows that $P_\mathcal{A} P_\mathcal{B}(\mathbf{x})$ *is the point in* $\mathcal{A} \cap \mathcal{B}$ *with minimal distance to* $\mathbf{x}$, *i.e.,* $P_{\mathcal{A}\cap\mathcal{B}}(\mathbf{x}) = P_\mathcal{A} P_\mathcal{B}(\mathbf{x})$. □

Proposition 6.2. *The projection onto* $\mathcal{F} = \mathcal{C}_1 \cap \mathcal{C}_3 \cap \mathcal{C}_4$ *is given by the composition*

$$P_\mathcal{F} = P_{\mathcal{C}_1} P_{\mathcal{C}_3} P_{\mathcal{C}_4}.$$

Proof. *Let* $\mathbf{p} := P_{\mathcal{C}_4}(\mathbf{x})$ *and let* $\mathbf{y} := \mathbf{p} - \mathbf{x}$. *According to* (6.7) *and* (6.8), $\mathbf{p}$ *and* $\mathbf{y}$ *are given entry-wise by* $(\forall k \in \{1, \ldots, 2NL\})$

$$p_k = \begin{cases} \hat{c}_k & \text{if } \mathbf{x} \notin \mathcal{C}_4^k \\ x_k & \text{otherwise} \end{cases} \qquad \text{and} \qquad y_k = \begin{cases} \hat{c}_k - x_k & \text{if } \mathbf{x} \notin \mathcal{C}_4^k \\ 0 & \text{otherwise.} \end{cases}$$

Now, let $\mathbf{z} := P_{\mathcal{C}_3}(\mathbf{p}) - \mathbf{p}$. *From* (6.6) *we obtain*

$$\begin{aligned} z_k &= \begin{cases} \hat{c}_k + \rho(\mathbf{p})(p_k - \hat{c}_k) - p_k & \text{if } k \in \mathcal{I}_\mathrm{d}^\mathcal{H}, \\ 0 & \text{otherwise.} \end{cases} \\ &= \begin{cases} \hat{c}_k + \rho(\mathbf{x})(x_k - \hat{c}_k) - x_k & \text{if } k \in \mathcal{I}_\mathrm{d}^\mathcal{H} \text{ and } \mathbf{x} \in \mathcal{C}_4^k, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Further, we have that

$$z_k + p_k = \begin{cases} \hat{c}_k + \rho(\mathbf{x})(x_k - \hat{c}_k) & \text{if } k \in \mathcal{I}_\mathrm{d}^\mathcal{H} \text{ and } \mathbf{x} \in \mathcal{C}_4^k, \\ \hat{c}_k & \text{otherwise,} \end{cases}$$

so the convexity of $\mathcal{C}_4$ *and* $(\forall \mathbf{x} \in \mathcal{H})\ \rho(\mathbf{x}) \in [0, 1]$ *imply that* $\mathbf{z} + \mathbf{p} = P_{\mathcal{C}_3} P_{\mathcal{C}_4}(\mathbf{x}) \in \mathcal{C}_3 \cap \mathcal{C}_4$. *As the entries of* $\mathbf{y}$ *and* $\mathbf{z}$ *are nonzero on disjoint subsets of frequency bins, it is clear that* $\langle \mathbf{z}, \mathbf{y} \rangle = 0$. *Hence Lemma 6.1 implies that* $P_{\mathcal{C}_3 \cap \mathcal{C}_4}(\mathbf{x}) = \mathbf{x} + \mathbf{y} + \mathbf{z} = P_{\mathcal{C}_3} P_{\mathcal{C}_4}(\mathbf{x})$. *Note that the order of the projections onto* $\mathcal{C}_3$ *and* $\mathcal{C}_4$ *cannot be changed.*

Since the constraint sets $\mathcal{C}_1$ *and* $\mathcal{C}_3 \cap \mathcal{C}_4$ *affect disjoint subsets of frequency bins, it holds that* $(\forall \mathbf{x} \in \mathcal{H})\ \langle P_{\mathcal{C}_1}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{C}_3 \cap \mathcal{C}_4}(\mathbf{x}) - \mathbf{x} \rangle = 0$. *Thus according to Proposition 1.1,*

$P_{\mathcal{F}} = P_{\mathcal{C}_3 \cap \mathcal{C}_4} P_{\mathcal{C}_1} = P_{\mathcal{C}_1} P_{\mathcal{C}_3 \cap \mathcal{C}_4}$, *which concludes the proof.* □

The projection of a frequency-domain vector $\mathbf{x} \in \mathcal{H}$ onto the set of signals bounded by magnitude θ in the time domain is given by

$$P_{\mathcal{T}}(\mathbf{x}) = \mathbf{F} P_{\mathcal{D}} \left(\mathbf{F}^T \mathbf{x} \right),$$

where $\mathcal{D} := \mathrm{lev}_{\leq \theta} \| \cdot \|_{\mathcal{H},\infty}$ is the set of (complex) time domain signals with magnitude bounded by θ. This is an immediate consequence of Fact 1.2, because $\mathbf{F}$ is an orthogonal matrix and the projection onto a closed convex set is the proximal mapping associated with its indicator function. According to (6.3), we can express the set $\mathcal{D}$ as the intersection $\mathcal{D} = \bigcap_{k=1}^{NL} \mathcal{D}_k$, where

$$(\forall k \in \{1, \ldots, NL\}) \quad \mathcal{D}_k = \left\{ \mathbf{x} \in \mathcal{H} \mid x_k^2 + x_{k+NL}^2 \leq \theta^2 \right\},$$

the projection onto which is given entry-wise by [SY98, Theorem 3.4—1] $(\forall k \in \{1, \ldots, NL\})$ $(\forall \mathbf{x} \in \mathcal{H})$ $(\forall i \in \{1, \ldots, 2NL\})$

$$P_{\mathcal{D}_k}(\mathbf{x})|_i = \begin{cases} \frac{\theta}{\sqrt{x_k^2 + x_{k+NL}^2}} x_i & \text{if } i \in \{k, k+NL\} \text{ and } x_k^2 + x_{k+NL}^2 > \theta^2 \\ x_i & \text{otherwise.} \end{cases}$$

Again, it is clear that $\langle P_{\mathcal{D}_k}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{D}_l}(\mathbf{x}) - \mathbf{x} \rangle =$ whenever $k \neq l$, so by Proposition 1.1 we can write

$$P_{\mathcal{D}}(\mathbf{x}) = \mathbf{x} + \sum_{k=1}^{NL} \left(P_{\mathcal{D}_k}(\mathbf{x}) - \mathbf{x} \right).$$

In the following section, we use closed form expressions for the projections onto $\mathcal{T}$ and $\mathcal{F}$, which were derived above, to devise fixed point algorithms for the feasibility Problem (6.5).

6.4. Algorithmic Solutions

In this section, we propose PAPR techniques based on algorithms investigated in Chapter 2. In Subsection 6.4.1, we show that existing PAPR reduction techniques can be derived as special cases of the POCS algorithm in (2.3) for certain choices frequency constraint sets. In Subsection 6.4.2, we propose a PAPR reduction algorithm for affine frequency domain constraints based on the EAPM in (2.29). Subsection 6.4.3 generalizes this method to arbitrary closed convex frequency constraint sets based on the extrapolated projection method in (2.27). In Subsection 6.4.4 we introduce an additional nonconvex constraint set that allows us to reuse phaseless pilot subcarriers for PAPR reduction. This additional

constraint set is incorporated into the algorithm by exploiting its bounded perturbation resilience, which was established in Chapter 2. Practical aspects of the proposed algorithms are discussed in Subsection 6.4.5.

6.4.1. Previous Methods

Many existing PAPR reduction techniques approximate a solution to the feasibility problem in (6.5) by applying the recursion

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathcal{F}} P_{\mathcal{T}}(\mathbf{x}_n), \quad \mathbf{x}_0 \in \mathcal{H}, \tag{6.9}$$

where typically $\mathbf{x}_0 = \underline{\hat{\mathbf{c}}}$. This algorithm iteratively clips the time-domain representation of $\mathbf{x}_n$ and enforces the frequency-domain constraints specified by $\mathcal{F}$. Depending on the choice of $\mathcal{F}$ (see 6.3.2), (6.9) yields the ICF algorithm [Arm02], the Fourier projection algorithm (FPA) [GP97], or the ACE-POCS algorithm [KJ03]. It is easy to see that the algorithm in (6.9) is a particular case of the POCS algorithm in (2.3) with unit relaxation factors $\lambda_1 = \lambda_2 = 1$.

In the simplified clipping and filtering (SCF) method [WT05], the clipping noise $P_{\mathcal{T}}(\mathbf{x}) - \mathbf{x}$ is multiplied by a constant λ in order to achieve strong PAPR reduction within one iteration. For $\lambda \in (0, 2)$, SCF corresponds to the first iteration of a relaxed POCS algorithm

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathcal{F}} \left(\mathbf{x}_n + \lambda \left(P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \right) \right), \quad \mathbf{x}_0 \in \mathcal{H}, \tag{6.10}$$

with $\mathcal{F} = \mathcal{C}_1$. Clearly, this is also a particular case of the POCS algorithm in (2.3), where $\lambda_1 = \lambda$ and $\lambda_2 = 1$. For the remainder of this chapter, we refer to the algorithm in (6.10) (with arbitrary $\mathcal{F}$) as rPOCS. Theorem 2.1 guarantees that the sequences defined by the recursions in (6.9) and (6.10) converge to a point $\mathbf{x}^\star \in \mathcal{F} \cap \mathcal{T}$ given that $\mathcal{F} \cap \mathcal{T} \neq \emptyset$. However, the convergence of POCS can be slow in practice [CCC+12]. Therefore, we investigate the applicability of extrapolated projection methods for PAPR reduction in the following subsections.

6.4.2. Extrapolated Projection Method for Tone Reservation

In the TR setting, the constraint set $\mathcal{F}$ is an affine subspace. Therefore, we can apply the EAPM in (2.29) to solve (6.5). It has been shown in [CCC+12], that the EAPM converges considerably faster than the POCS algorithms in (6.9) and (6.10) for affine-convex feasibility problems. Given a point $\mathbf{x}_0 \in \mathcal{F}$ in an affine subspace $\mathcal{F} \subset \mathcal{H}$ of a real Hilbert space, and a closed convex set $\mathcal{T} \subset \mathcal{H}$, the EAPM generates a sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$

by

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = \mathbf{x}_n + \lambda_n K_n \left(P_{\mathcal{F}} P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \right), \tag{6.11}$$

with $\lambda_n \in (0, 2)$ and an extrapolation factor

$$(\forall n \in \mathbb{N}) \quad K_n = \frac{\| P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \|^2}{\| P_{\mathcal{F}} P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \|^2}.$$

Since $\mathcal{H}$ is finite dimensional, convergence to a point $\mathbf{x}^\star \in \mathcal{F} \cap \mathcal{T}$ is guaranteed by Theorem 2.5 (a), given that $\mathcal{F} \cap \mathcal{T} \neq \emptyset$.

6.4.3. PAPR Reduction with Arbitrary Constraints

Imposing EVM or ACE constraints results in a non-affine set $\mathcal{F}$. In this case, the EAPM cannot be used because extrapolation might cause the iterate $\mathbf{x}_n$ to violate the frequency-domain constraints. To avoid this issue, we propose a PAPR reduction technique for non-affine frequency-domain constraints based on the GPR approach in Theorem 2.4. Given two arbitrary closed convex sets $\mathcal{F}, \mathcal{T} \subset \mathcal{H}$, the GPR algorithm produces a sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ by

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathcal{F}} \left(\mathbf{x}_n + \lambda_n \sigma_n^{\mathrm{GPR}} \left(P_{\mathcal{F}} P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \right) \right), \quad \mathbf{x}_0 \in \mathcal{F}, \tag{6.12}$$

where

$$(\forall n \in \mathbb{N}) \quad \sigma_n^{\mathrm{GPR}} = \frac{\| P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \|^2}{\langle P_{\mathcal{F}} P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n, P_{\mathcal{T}}(\mathbf{x}_n) - \mathbf{x}_n \rangle}.$$

According to Theorem 2.4(b), the sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ produced by this algorithm is guaranteed to converge to a point $\mathbf{x}^\star \in \mathcal{F} \cap \mathcal{T}$ given hat $\mathcal{F} \cap \mathcal{T} \neq \emptyset$, because $\mathcal{H}$ is finite dimensional. As shown in Theorem 2.5, the EAPM in (2.29) is a particular case of the GPR algorithm in (2.26). Therefore, the algorithm in (6.12) can be seen as a generalization of the the algorithm in (6.11) for arbitrary closed convex sets $\mathcal{F}$. This allows us to use the algorithm in 6.12 for PAPR reduction with arbitrary combinations of the constraint sets defined in Section 6.3.2. The constraints can be altered without the necessity to change the structure of the algorithm. If the set $\mathcal{F}$ is affine, the performance is equivalent to that of the EAPM, whereas for non-affine constraints, the method is still applicable and yields rapid PAPR reduction. A further property of this approach is that the iterate $\mathbf{x}_n$ satisfies the frequency-domain constraints at each iteration, owing to the final unrelaxed projection onto $\mathcal{F}$. Note also that the computational complexity of one iteration is roughly equal for all algorithms considered in Section 6.4, since each iteration involves one projection onto $\mathcal{T}$, and consequently one IDFT/DFT-pair.

6.4.4. Phaseless Pilot Reuse by Superiorization

A recent branch of research in channel estimation for OFDM systems is directed towards the deployment of phaseless pilots [WBJ15a, WBJ15b], which allow the receiver to reconstruct the phase shift introduced by the wireless channel without requiring knowledge of the pilot phases. This channel estimation scheme enables the transmitter to use the phase of the pilot subcarriers for peak compensation. In this scheme, the set of pilot subcarriers is identical with the set $\mathcal{I}_\mathrm{c}$ of compensation subcarriers. While in the various definitions of $\mathcal{F}$ in Section 6.3.2 there was no constraint on the compensation subcarriers, simultaneously using them as phaseless pilots imposes an equality constraint on their magnitude. This is taken into account by defining the nonconvex set

$$\mathcal{P} := \left\{ \mathbf{x} \in \mathbb{R}^{2NL} \;\middle|\; (\forall k \in \mathcal{I}_\mathrm{c}),\ x_k^2 + x_{k+NL}^2 = p_k \right\}, \tag{6.13}$$

where p_k denotes the power of the kth subcarrier. The problem in (6.5) is then extended to the nonconvex feasibility problem

$$\operatorname*{find}_{\mathbf{x} \in \mathcal{H}} \mathbf{x} \in \mathcal{F} \cap \mathcal{T} \cap \mathcal{P}.$$

Note that, although $\mathcal{P}$ is a nonconvex set, a projection of $\mathbf{x} \in \mathcal{H}$ onto $\mathcal{P}$ always exists, and it is unique if $(\forall k \in \mathcal{I}_\mathrm{c})\ x_k^2 + x_{k+NL}^2 \neq 0$. In [1], we proposed heuristics that terminate each iteration of the algorithms in (6.10) and (6.12) by a projection onto the nonconvex set $\mathcal{P}$. However, as $\mathcal{P}$ is nonconvex, we cannot guarantee the convergence of these methods. Therefore, we follow a different approach below, which encorporates the nonconvex constraint set $\mathcal{P}$ by defining superiorized versions of the algorithms in (6.10) and (6.12). To simplify a comparison with the heuristics in [1], we begin by reformulating the algorithms in (6.10) and (6.12). By defining $T = I + \lambda(P_\mathcal{T} - I)$, we can equivalently write the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ in (6.10) as

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad \mathbf{y}_{n+1} &= T P_\mathcal{F}(\mathbf{y}_n), \quad \mathbf{y}_0 \in \mathcal{H} \\ \mathbf{x}_n &= P_\mathcal{F}(\mathbf{y}_n). \end{aligned}$$

Since T is a relaxed projection onto a closed convex set, the convergence of the superiorized version

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad \mathbf{y}_{n+1} &= T P_\mathcal{F}(\mathbf{y}_n + \beta_n \mathbf{v}_n), \quad \mathbf{y}_0 \in \mathcal{H} \\ \mathbf{x}_n &= P_\mathcal{F}(\mathbf{y}_n) \end{aligned} \tag{6.14}$$

of this algorithm is guaranteed by Theorem 2.1, given that $(\beta_n \mathbf{v}_n)_{n\in\mathbb{N}}$ are bounded perturbations. Similarly, we can define a superiorized version of the GPR algorithm in (6.12). Using the mapping $T^{\mathrm{GPR}}_{\lambda_n} : \mathcal{F} \to \mathcal{H}$ in (2.26), we can write (6.12) as $\mathbf{x}_0 \in \mathcal{F}$

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathcal{F}} T^{\mathrm{GPR}}_{\lambda_n}(\mathbf{x}_n).$$

Note that this algorithm is equivalent to

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathbf{y}_{n+1} &= T^{\mathrm{GPR}}_{\lambda_n} P_{\mathcal{F}}(\mathbf{y}_n), \quad \mathbf{y}_0 \in \mathcal{H} \\ \mathbf{x}_n &= P_{\mathcal{F}}(\mathbf{y}_n),\end{aligned}$$

so we can use Theorem 2.4 to define a superiorized version by

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathbf{y}_{n+1} &= T^{\mathrm{GPR}}_{\lambda_n} P_{\mathcal{F}}(\mathbf{y}_n + \beta_n \mathbf{v}_n), \quad \mathbf{y}_0 \in \mathcal{H} \\ \mathbf{x}_n &= P_{\mathcal{F}}(\mathbf{y}_n).\end{aligned} \tag{6.15}$$

In slight deviation from the superiorization methodology in [Cen15], we choose a nonconvex superiorization objective $f : \mathcal{H} \to \mathbb{R}_+$ given by $f(\mathbf{x}) = d(\mathbf{x}, \mathcal{P})$. To reduce the value of this objective function, we use perturbations of the form $(\forall n \in \mathbb{N})$ $\mathbf{v}_n := P_{\mathcal{P}}(\mathbf{y}_n) - \mathbf{y}_n$. We omit the proof that the perturbations are bounded. However, we note that boundedness can always be ensured by projecting the perturbations onto a ball $\mathcal{B}_r(\mathbf{0}) = \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x}\| \leq r\}$ for some $r > 0$.

Remark 6.1. If we choose $\beta_0 = 0$ and $(\forall n \in \mathbb{N} \setminus \{0\})$ $\beta_n = 1$, the algorithms in (6.14) and (6.15) yield the heuristics[a] proposed in [1], which terminate each iteration in (6.10) and (6.12), respectively, with a projection onto $\mathcal{P}$.

To see this fix $\mathbf{x} \in \mathcal{H}$, let $P_{\mathcal{P}}(\mathbf{x}) \in \Pi_{\mathcal{P}}(\mathbf{x})$ be a projection of $\mathbf{x}$ onto the nonconvex set $\mathcal{P}$, and define a hyperplane $\mathcal{V} := \{\mathbf{y} \in \mathcal{H} \mid \langle \mathbf{y} - P_{\mathcal{P}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{P}}(\mathbf{x})\rangle = 0\}$, the projection onto which satisfies $P_{\mathcal{V}}(\mathbf{x}) = P_{\mathcal{P}}(\mathbf{x})$ and

$$(\forall \mathbf{y} \in \mathcal{H}) \quad P_{\mathcal{V}}(\mathbf{y}) - \mathbf{y} = \frac{\langle \mathbf{y} - P_{\mathcal{P}}(\mathbf{x}), \mathbf{x} - P_{\mathcal{P}}(\mathbf{x})\rangle}{\|\mathbf{x} - P_{\mathcal{P}}(\mathbf{x})\|^2}(\mathbf{x} - P_{\mathcal{P}}(\mathbf{x})). \tag{6.16}$$

Note that the constraints $\mathcal{F}$ and $\mathcal{P}$ affect disjoint subsets of subcarriers, so $\langle P_{\mathcal{P}}(\mathbf{x}) - \mathbf{x}, P_{\mathcal{F}}(\mathbf{x}) - \mathbf{x}\rangle = 0$, whereby according to (6.16) $(\forall \mathbf{y} \in \mathcal{H})$ $\langle P_{\mathcal{V}}(\mathbf{y}) - \mathbf{y}, P_{\mathcal{F}}(\mathbf{y}) - \mathbf{y}\rangle = 0$. Thus by Proposition 1.1, $P_{\mathcal{P}} P_{\mathcal{F}}(\mathbf{x}) = P_{\mathcal{V}} P_{\mathcal{F}}(\mathbf{x}) = P_{\mathcal{V}\cap\mathcal{F}}(\mathbf{x}) = P_{\mathcal{F}} P_{\mathcal{V}}(\mathbf{x}) = P_{\mathcal{F}} P_{\mathcal{P}}(\mathbf{x})$. Consequently, substituting $\mathbf{v}_n = P_{\mathcal{P}}(\mathbf{y}_n) - \mathbf{y}_n$ in (6.14) and (6.15) yields the desired

result.

[a] Note that in this case $(\beta_n)_{n\in\mathbb{N}} \notin \ell_+^1(\mathbb{N})$.

Remark 6.1 shows the relation between the superiorized algorithms in (6.14) and (6.15) and the heuristic proposed in [1]. For rPOCS, this heuristic is very similar to the algorithm proposed in [MH09] for PAPR reduction in pulse amplitude modulated transform domain communication systems.

6.4.5. Practical Aspects: Feasibility, Computational Efficiency

As mentioned above, convergence of the extrapolated methods in (6.11) and (6.12) is only guaranteed if θ is chosen such that $\mathcal{F} \cap \mathcal{T} \neq \emptyset$. One way of achieving this is to use bisection search as proposed in [CCC+12]. However, the resulting increase in computation time required to find an appropriate value for θ can be prohibitive in practice, where typically only a small number of iterations can be computed before transmitting the symbol. It is therefore desirable to have a good choice for θ from the first iteration. This can be achieved by running several instances of the algorithm with different clipping thresholds θ in parallel, and returning the result with lowest PAPR after the last iteration. If the available computational resources do not permit such a parallel approach, it is also possible to use only one instance of the algorithm in (6.12) with a clipping threshold chosen such that the probability of the problem in (6.5) being infeasible is sufficiently small.

While the projection methods mentioned above produce sequences $(\mathbf{x}_n)_{n\in\mathbb{N}}$ that converge to a feasible point $\mathbf{x}^\star \in \mathcal{F} \cap \mathcal{T}$, as $n \to \infty$, the computational resources of the transmitter often only allow for computing a single iteration. This fact is taken into account by the SCF algorithm [WT05], which computes a single overrelaxed POCS iteration instead of multiple unrelaxed POCS iterations (as in the ICF algorithm). The SCF approach relies on a fixed overrelaxation/extrapolation parameter λ that may be larger than two, in which case the POCS algorithm would no longer be guaranteed to converge. A disadvantage of SCF is that it lacks a means of controlling the in-band distortion introduced by the peak compensation. The BER of the system will therefore greatly depend on the choice of clipping threshold θ and relaxation/extrapolation parameter λ. In contrast, the algorithm in (6.12) produces a sequence of symbols each of them satisfying all constraints in the frequency domain. For example, individual EVM constraints can be specified for subcarriers conveying constellations of different size. In this way, the distortion is restricted independently of the choice of clipping threshold and relaxation parameter. Furthermore, the extrapolation factor σ_n^{GPR} is chosen adaptively, so computing one iteration according to (6.12) can reduce the PAPR by an amount similar to that achieved by the SCF method, while guaranteeing convergence if multiple iterations can be

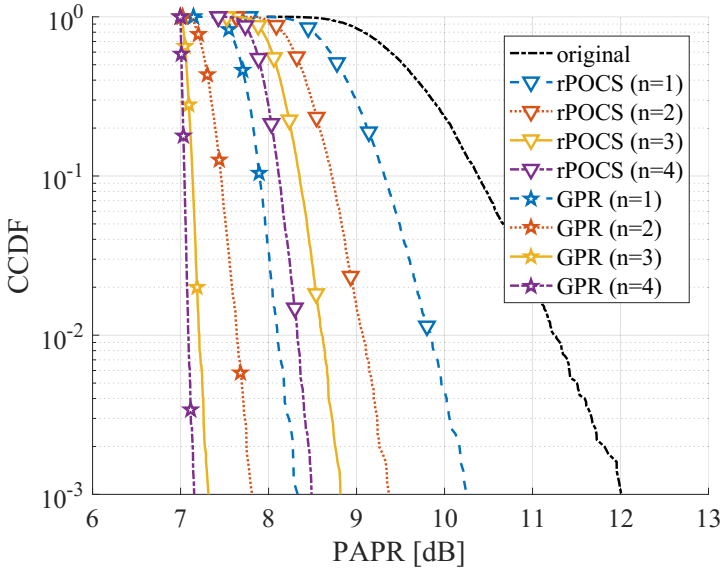


Figure 6.1.: CCDF of the PAPR for the first four iterations of `rPOCS` ($\lambda_{\text{POCS}} = 2$) and `GPR` ($\lambda_{\text{GPR}} = 1.4$) with ACE constaints, $N = 2048$, $L = 4$, CR = 7 dB, 5% compensation subcarriers, data subcarriers modulated with QPSK and 16-QAM, respectively, where $\text{EVM}^{\text{max}}_{\text{QPSK}} = 15\%$ and $\text{EVM}^{\text{max}}_{\text{QAM}} = 5\%$.

computed to refine the estimate.

6.5. Numerical Results

In this section, we compare the performance of the algorithms in (6.10) and (6.12), in the following denoted as `rPOCS` and `GPR`, respectively. The simulation is performed for randomly generated OFDM symbols with $N = 2048$ subcarriers. Negative frequency subcarriers convey QPSK constellation points, whereas positive frequency subcarriers convey 16-QAM constellation points. For each symbol, 5% of the subcarriers are randomly selected as compensation subcarriers. Data subcarriers are required to satisfy both EVM and ACE constraints, where $\text{EVM}^{\text{max}}_{\text{QPSK}} = 15\%$ and $\text{EVM}^{\text{max}}_{\text{QAM}} = 5\%$, respectively. Peaks in the analog signal were approximated by upsampling with a factor $L = 4$, and clipped at a clipping ratio (target PAPR) of CR = 7 dB. The relaxation parameter for `GPR` was set to $(\forall n \in \mathbb{N})\ \lambda_n = \lambda_{\text{GPR}} = 1.4$, although rapid decrease in PAPR was observed for any $\lambda_{\text{GPR}} \in [1, 2]$. As large overrelaxation yielded the fastest PAPR decrease for `rPOCS`, the relaxation parameter was set to $(\forall n \in \mathbb{N})\ \lambda_n = \lambda_{\text{POCS}} = 2$. Note that, by definition, all of the frequency-domain constraints are satisfied in every iteration.

Fig. 6.1 shows the CCDF of the PAPR throughout the first four iterations of both

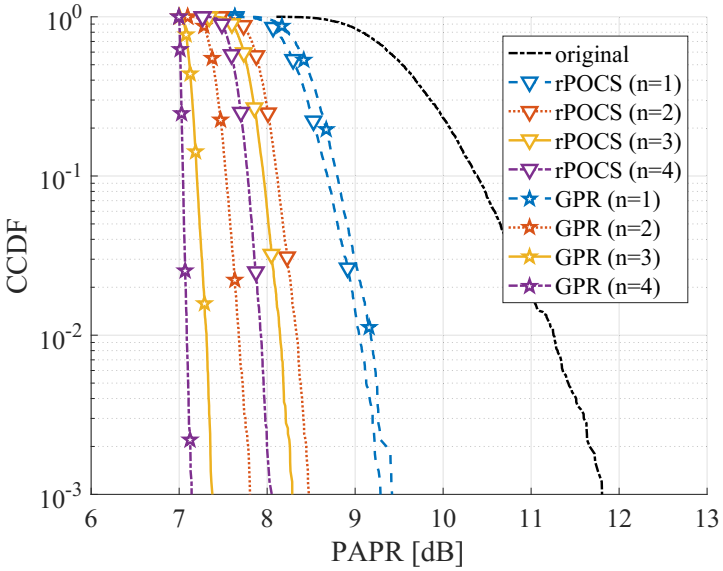


Figure 6.2.: CCDF for the same simulation setup as in Fig. 6.1, where each iteration of `rPOCS` and `GPR` is terminated by an additional projection onto the nonconvex set $\mathcal{P}$ in (6.13).

algorithms. It can be seen that, for `rPOCS`, the rate of convergence decreases at every iteration. The `GPR` algorithm achieves almost 2 dB lower PAPR than `rPOCS` in the first iteration, while almost reaching the target PAPR after four iterations. Next, we add the nonconvex constraint in (6.13), where $(\forall k \in \mathcal{I}_c)\ p_k = 1$. Fig. 6.2 shows the CCDF for the same scenario for the superiorized algorithms in (6.14) and (6.15), where we heuristically chose $\beta_0 = 0$ and $(\forall n \in \mathbb{N} \setminus \{0\})\ \beta_n = 1$ such that $(\beta_n)_{n\in\mathbb{N}} \notin \ell^1_+(\mathbb{N})$. In this way, we ensure that the nonconvex magnitude constraint on pilot/compensation subcarriers introduced in Section 6.4.4 is satisfied in each iteration. It can be seen that both algorithms still reduce the PAPR by a similar amount when this additional nonconvex constraint is added. The performance of `rPOCS` even improves slightly, which may result from the fact that the projection onto $\mathcal{P}$ scales up the compensation signal in the compensation subcarriers, acting similarly to an overrelaxation in the initial iteration. Nevertheless, the extrapolated `GPR` algorithm outperforms `rPOCS` for $n > 1$.

6.6. Conclusion

In this chapter, we proposed a PAPR reduction algorithm for tone reservation based on the EAPM in (2.29), as well as its generalization for non-affine frequency-domain constraints, which is based on the GPR approach in (2.27). In addition, we proposed a heuristic

extension for nonconvex constraints motivated by superiorization, which allows reusing the phase of pilot subcarriers for phaseless channel estimation to reduce the PAPR. Simulations showed that the proposed methods were able to decrease considerably the PAPR within very few iterations.

7. Discussion and Outlook

In this thesis, we applied fixed point algorithms to tackle optimization tasks in several parts of modern wireless communication systems. Owing to their simple structure, the proposed algorithms can be implemented with low effort in practical systems. Fixed point methods give rise to online algorithms that can easily be integrated with algorithmic subroutines, as shown in Chapter 5, where combined online channel estimation algorithms with a phase and delay compensation mechanism and a heuristic receive beamforming policy. In a similar way, future research could integrate more of the building blocks described in this thesis into a universal signal processing framework that jointly performs channel estimation, transmit beamforming, PAPR reduction, receive beamforming and signal detection, while potentially taking into account the interplay between the individual building blocks. As we showed in Chapter 4, deep unfolding of iterative algorithms naturally leads to neural network architectures that can achieve similar performance with much fewer iterations. Applying deep unfolding to signal processing chains comprising multiple algorithmic building blocks could guide the design process of deep neural networks that jointly perform several tasks in future wireless systems.

In the superiorization methodology, perturbations are typically computed based on subgradient steps for a convex superiorization objective. By contrast, we considered both convex and nonconvex superiorization objectives, and we defined the perturbations in closed form (e.g., using proximal mappings). From a theoretical point of view, superiorization cannot be used in online settings, as the bounded perturbations decay to zero by definition, whereas the algorithm runs indefinitely. The heuristic proposed in Chapter 5 avoids this problem by using a nondecaying sequence of perturbations, whereby its convergence can no longer be guaranteed, even if the solution set remains constant over time. Alternatively, one could use a summable sequence to scale the perturbations, and restart this sequence occasionally. In this way, quasi-Fejér monotonicity would be guaranteed until the sequence is restarted.

Yet another use of perturbation resilience was presented in [2]. In that publication, which was not presented in this thesis, we used perturbation resilience to interleave the iterations of different localization algorithms in order to combine their advantages. A more numerically stable algorithm based on a direct convex relaxation of the localization prob-

lem was used as a basic algorithm, and a more error prone algorithm based on a tighter semidefinite relaxation was used to produce perturbations. Similarly, future research could use heuristics or neural networks to accelerate the convergence and to enhance the performance of perturbation resilient basic algorithms, without losing convergence guarantees.

Most of the results on perturbation resilience and superiorization in the literature are restricted to finite dimensional Hilbert spaces. Although all problems considered in this theses are posed in finite dimensional spaces, the results in Chapter 2 remain valid in infinite dimensional spaces. Therefore, these results open up new possibilities for developing superiorized versions of the algorithmic frameworks covered in Chapter 2, which are applicable even in infinite dimensional spaces.

Appendices

A. Multicast Beamforming

A.1. Proof of Remark 3.1

Given a real vector space $\mathcal{V}$, a real inner product is a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ satisfying [JAA04]

(a) $(\forall \mathbf{x} \in \mathcal{V})\ \langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$

(b) $(\forall \mathbf{x}, \mathbf{y} \in \mathcal{V})\ \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

(c) $(\forall \mathbf{x}, \mathbf{y} \in \mathcal{V})(\forall \alpha \in \mathbb{R})\ \langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$

(d) $(\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V})\ \langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.

Note that $(\forall \mathbf{X} \in \mathcal{V})\ \mathrm{Re}\{\mathrm{tr}(\mathbf{X}^H \mathbf{X})\} = \mathrm{tr}(\mathbf{X}^H \mathbf{X}) = \|\mathbf{X}\|_F^2$. Consequently, (a) follows from the nonnegativity and positive-definiteness of a norm. The symmetry in (b) follows from the fact that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ for matrices $\mathbf{A}, \mathbf{B}$ with compatible dimensions, and $\mathrm{Re}\{\mathrm{tr}(\mathbf{X})\} = \mathrm{Re}\{\mathrm{tr}(\mathbf{X}^H)\}$ for $\mathbf{X} \in \mathcal{V}$. Moreover, (c) and (d) follow from the linearity of $\mathrm{Re}\{\cdot\}$ and $\mathrm{tr}(\cdot)$. □

A.2. Proof of Proposition 3.1

Denote the perturbed point for a given choice of τ by $\mathbf{Z}_\tau^\star := \mathbf{X} + \mathbf{Y}_\tau^\star$. By substituting $\mathbf{Y} = \mathbf{Z} - \mathbf{X}$ in (3.19), we can identify this point as $\mathbf{Z}_\tau^\star = \mathrm{prox}_{\tau f_1}(\mathbf{X})$, where the proximal mapping is given by

$$(\forall \mathbf{X} \in \mathcal{H}^M) \quad \mathrm{prox}_{\tau f_1}(\mathbf{X}) \in \underset{\mathbf{Z} \in \mathcal{H}^M}{\arg\min} \left(\tau f_1(\mathbf{Z}) + \frac{1}{2} |||\mathbf{X} - \mathbf{Z}|||^2 \right). \tag{A.1}$$

Note that the function

$$\tau f_1(\mathbf{Z}) + \frac{1}{2} |||\mathbf{X} - \mathbf{Z}|||^2 = \tau \sum_{m=1}^{M} \|\mathbf{Z}_m\|_* + \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{X}_m - \mathbf{Z}_m\|^2$$

is separable over m. Consequently, we can compute the proximal mapping in (A.1) by solving

$$(\forall m \in \mathcal{M}) \quad \mathbf{Z}_\tau^\star|_m \in \arg\min_{\mathbf{Z}\in\mathcal{H}^M} \tau\|\mathbf{Z}\|_* + \frac{1}{2}\|\mathbf{X}_m - \mathbf{Z}\|^2. \tag{A.2}$$

According to [CCS10, Theorem 2.1], the unique solution to (A.2) is given by $\mathbf{Z}_\tau^\star|_m = \chi_\tau(\mathbf{X}_m)$.[1] Substituting $\mathbf{Y}_\tau^\star = \mathbf{Z}_\tau^\star - \mathbf{X}$ yields (3.20), which is the desired result. □

A.3. Technical Lemmas Required to Prove Proposition 3.2(a)

The proof of Proposition 3.2(a) relies on the following lemmas:

Lemma A.1. *Let $\mathcal{C}_+ \subset \mathcal{H}^M$ be the set defined in (3.8) and let T_{P}^α Then $(\forall \mathbf{X} = (\mathbf{X}_m)_{m\in\mathcal{M}} \in \mathcal{H}^M)$*

$$d^2(\mathbf{X}, \mathcal{C}_+) = \sum_{m\in\mathcal{M}} \sum_{i\in\mathcal{N}} (-\lambda_i(\mathbf{X}_m))_+^2 ,$$

where $(\forall m \in \mathcal{M})$ $\lambda_1(\mathbf{X}_m), \ldots, \lambda_N(\mathbf{X}_m)$ are the (real) eigenvalues of $\mathbf{X}_m$.

Proof. *Denote by $(\forall m \in \mathcal{M})$ $\mathbf{X}_m = \mathbf{V}_m\mathbf{\Lambda}_m\mathbf{V}_m^H$ the eigendecomposition of $\mathbf{X}_m$. Using the definition of $P_{\mathcal{C}_+}$ in (3) and (3.16), we obtain*

$$d^2(\mathbf{X}, \mathcal{C}_+) = \left|\left|\left|\mathbf{X} - P_{\mathcal{C}_+}(\mathbf{X})\right|\right|\right|^2 = \sum_{m\in\mathcal{M}} \|\mathbf{V}_m(\mathbf{\Lambda}_m - \mathbf{\Lambda}_m^+)\mathbf{V}_M^H\|^2,$$

where $(\forall m \in \mathcal{M})$ $\mathbf{\Lambda}_m^+ = \Big((\lambda_1(\mathbf{X}_m))_+, \ldots, (\lambda_N(\mathbf{X}_m))_+\Big)$. Moreover, because the Frobenius norm is invariant under unitary transformation, we can write

$$\begin{aligned} d^2(\mathbf{X}, \mathcal{C}_+) = \sum_{m\in\mathcal{M}} \|\mathbf{\Lambda}_m - \mathbf{\Lambda}_m^+\|^2 &= \sum_{m\in\mathcal{M}} \sum_{i\in\mathcal{N}} (\lambda_i(\mathbf{X}_m) - (\lambda_i(\mathbf{X}_m))_+)^2 \\ &= \sum_{m\in\mathcal{M}} \sum_{i\in\mathcal{N}} (-\lambda_i(\mathbf{X}_m))_+^2 , \end{aligned}$$

which is the desired result. □

Lemma A.2. *Let $\mathbf{A} \in \mathcal{H} = \{\mathbf{A} \in \mathbb{C}^{N\times N} \mid \mathbf{A}^H = \mathbf{A}\}$ with SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Then an eigendecomposition is given by $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$.*

Proof. *Let $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$ be an eigendecomposition of $\mathbf{A}$. Then we have $\mathbf{A}^H\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^H\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^H$ and $\mathbf{A}^H\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^H$. Moreover, since*

[1] The proof in [CCS10] is for real matrices. However, the generalization to complex matrices is straightforward.

$\mathbf{A}$ is Hermitian, the eigenvalues can be ordered such that $\mathbf{\Sigma}^2 = \mathbf{\Lambda}^2$, so there exists an eigendecomposition $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$ with $\mathbf{Q} = \mathbf{V}$. □

Lemma A.3. *Let $\mathbf{A} \in \mathcal{H} = \{\mathbf{A} \in \mathbb{C}^{N\times N} \mid \mathbf{A}^H = \mathbf{A}\}$ be a Hermitian matrix with $\operatorname{rank}(\mathbf{A}) = 1$ and SVD $\mathbf{A} = \sigma_1(\mathbf{A})\mathbf{u}_1\mathbf{v}_1^H$. Then an eigendecomposition is given by $\mathbf{A} = \lambda_1(\mathbf{A})\mathbf{v}\mathbf{v}^H$, where $\lambda_1(\mathbf{A}) = \mathbf{v}_1^H\mathbf{u}_1\sigma_1(\mathbf{A})$.*

Proof. *According to Lemma A.2, an eigendecomposition is given by $\mathbf{A} = \lambda_1(\mathbf{A})\mathbf{v}_1\mathbf{v}_1^H$. Moreover, since $\mathbf{A}$ is Hermitian, it holds that $\sigma_1(\mathbf{A}) = |\lambda_1(\mathbf{A})|$, so we can write*

$$\begin{aligned}\lambda_1(\mathbf{A})\mathbf{v}_1\mathbf{v}_1^H &= \sigma_1(\mathbf{A})\mathbf{u}_1\mathbf{v}_1^H\\ \sigma_1(\mathbf{A})\operatorname{sgn}(\lambda_1(\mathbf{A}))\mathbf{v}_1\mathbf{v}_1^H &= \sigma_1(\mathbf{A})\mathbf{u}_1\mathbf{v}_1^H.\end{aligned}$$

Since $\operatorname{rank}(\mathbf{A}) = 1$ implies that $\sigma_1(\mathbf{A}) > 0$, we can divide both sides by $\sigma_1(\mathbf{A})$ to obtain

$$\begin{aligned}\operatorname{sgn}(\lambda_1(\mathbf{A}))\mathbf{v}_1\mathbf{v}_1^H &= \mathbf{u}_1\mathbf{v}_1^H\\ \operatorname{sgn}(\lambda_1(\mathbf{A}))\mathbf{v}_1^H\mathbf{v}_1\mathbf{v}_1^H\mathbf{v}_1 &= \mathbf{v}_1^H\mathbf{u}_1\mathbf{v}_1^H\mathbf{v}_1\\ \operatorname{sgn}(\lambda_1(\mathbf{A})) &= \mathbf{v}_1^H\mathbf{u}_1,\end{aligned}$$

where we used the fact that $\mathbf{v}_1^H\mathbf{v}_1 = 1$. Thus we have that $\lambda_1(\mathbf{A}) = \operatorname{sgn}(\lambda_1(\mathbf{A}))|\lambda_1(\mathbf{A})| = \mathbf{v}_1^H\mathbf{u}_1\sigma_1(\mathbf{A})$, which concludes the proof. □

B. MIMO Detection

To prove Proposition 4.1, we exploit the fact that the objective function is separable, and we make use of the following lemma:

Lemma B.1. *Let* $\mathcal{A} := \{a_0 + k\Delta \mid k \in \{1, \dots, K\}\}$, *where* $K \in \mathbb{N}$, $a_0 \in \mathbb{R}$ *and* $\Delta > 0$. *Then* $(\forall x \in \mathbb{R})$

$$\arg\min_{y\in\mathbb{R}} \left(\tau|y - P_{\mathcal{A}}(y)| + \frac{1}{2}(x-y)^2\right) = \arg\min_{y\in\mathbb{R}} \left(\tau|y - P_{\mathcal{A}}(x)| + \frac{1}{2}(x-y)^2\right).$$

Proof. *Fix* $x \in \mathbb{R}$ *and define*

$$y^\star \in \arg\min_{y\in\mathbb{R}} \left(f(y) + \frac{1}{2}(x-y)^2\right), \tag{B.1}$$

where the function $f : \mathbb{R} \to \mathbb{R}$ *is given by* $(\forall y \in \mathbb{R})$ $f(y) = \tau|y - P_{\mathcal{A}}(y)|$. *Now, suppose that for some* $y \in \mathbb{R}$, *it holds that* $P_{\mathcal{A}}(y) \neq P_{\mathcal{A}}(x)$. *Then by definition of the function* f, *which is illustrated in Figure B.1, there exists* $k \in \mathbb{Z}$ *such that:*

(a) $P_{\mathcal{A}}(y + k\Delta) = P_{\mathcal{A}}(x)$

(b) $|y + k\Delta - P_{\mathcal{A}}(y + k\Delta)| \leq |y - P_{\mathcal{A}}(y)|$

(c) $(x - (y + k\Delta))^2 < (x - y)^2$

Note that $P_{\mathcal{A}}(y) \neq P_{\mathcal{A}}(x)$ *and (a) imply that* $k \neq 0$. *Thus there exists* $y' := y + k\Delta \neq y$ *such that* $f(y') + \frac{1}{2}(x - y')^2 < f(y) + \frac{1}{2}(x - y)^2$, *which shows that* y *is not optimal for* (B.1). *Therefore, any solution* $y^\star$ *to* (B.1) *must satisfy* $P_{\mathcal{A}}(y^\star) = P_{\mathcal{A}}(x)$, *which concludes the proof.* □

Proof of Proposition 4.1

According to Definition 1.1, the proximal mapping associated with τf_{ℓ_1} satisfies

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \operatorname{prox}_{\tau f_{\ell_1}}(\mathbf{x}) \in \arg\min_{\mathbf{y}\in\mathcal{H}} \left(\tau f_{\ell_1}(\mathbf{y}) + \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2\right). \tag{B.2}$$

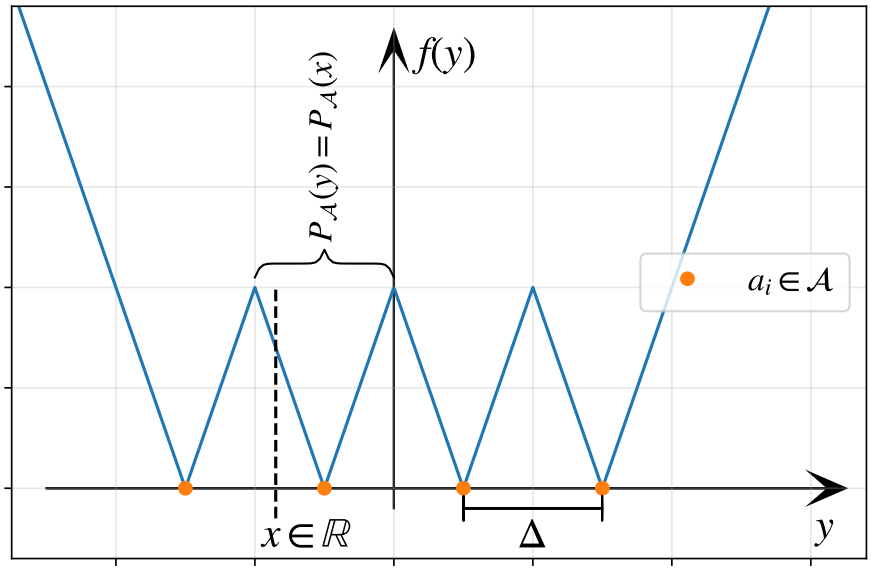


Figure B.1.: Illustration of the function f in Lemma B.1 for a set $\mathcal{A}$ with cardinality $K = 4$.

Note that

$$\begin{aligned}\tau f_{\ell_1}(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 &= \tau\|\mathbf{y} - P_{\mathcal{S}}(\mathbf{y})\|_1 + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 \\ &= \sum_{i=1}^{2N_{\mathrm{T}}} \left(\tau|y_i - P_{\mathcal{A}}(y_i)| + \frac{1}{2}(x_i - y_i)^2\right),\end{aligned}$$

so the objective in (B.2) is separable over i, and we can write

$$(\forall \mathbf{x} \in \mathcal{H})(\forall i \in \{1,\dots,2N_{\mathrm{T}}\}) \quad \mathrm{prox}_{\tau f_{\ell_1}}(\mathbf{x})|_i \in \arg\min_{y\in\mathbb{R}} \left(\tau|y - P_{\mathcal{A}}(y)| + \frac{1}{2}(x_i - y)^2\right).$$

According to Lemma B.1, we can write (B.2) as

$$(\forall \mathbf{x} \in \mathcal{H})(\forall i \in \{1,\dots,2N_{\mathrm{T}}\}) \quad \mathrm{prox}_{\tau f_{\ell_1}}(\mathbf{x})|_i \in \arg\min_{y\in\mathbb{R}} \left(\tau|y - P_{\mathcal{A}}(x)| + \frac{1}{2}(x_i - y)^2\right).$$

Note that the objective function is now convex in y. Thus, according to [Bec17, Example 6.8] and the translation property of the proximal mapping [Bec17, Theorem 6.11], [CP11, Table 10.1.i], a proximal mapping associated with τf_{ℓ_1} is given by

$$\mathrm{prox}_{\tau f_{\ell_1}}(\mathbf{x}) = \phi_\tau\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right) + P_{\mathcal{S}}(\mathbf{x}),$$

where $(\forall \tau \geq 0)$ $\phi_\tau : \mathcal{H} \to \mathcal{H}$ is the soft-thresholding operator

$$(\forall \mathbf{x} \in \mathcal{H})(\forall i \in \mathcal{I}_{\mathrm{T}}) \quad \phi_\tau(\mathbf{x})|_i := \mathrm{sgn}(x_i)(|x_i| - \tau)_+.$$

C. Online Channel Estimation

C.1. Proof of Proposition 5.4

Let ξ be the bijection between $\mathcal{H} = \mathbb{C}^N$ and $\mathbb{R}^{2N}$ defined in in (5.7), and define a function $f : \mathbb{R}^{2N} \to \mathbb{R}_+$ by

$$f(\bar{\mathbf{x}}) := \sum_{i=1}^{N} \omega_i \sqrt{\bar{x}_i^2 + \bar{x}_{i+N}^2}.$$

Note that this function satisfies $(\forall \mathbf{x} \in \mathcal{H})\ f(\xi(\mathbf{x})) = \|\mathbf{x}\|_1^{\boldsymbol{\omega}}$. Thus we can write

$$\begin{aligned}
&\xi\left(\mathrm{prox}_{\lambda\|\cdot\|_1^{\boldsymbol{\omega}}}(\mathbf{x})\right) \in \underset{\bar{\mathbf{y}}\in\mathbb{R}^{2N}}{\arg\min} f(\mathbf{y}) + \frac{1}{2\lambda}\|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|^2 \\
&= \underset{\bar{\mathbf{y}}\in\mathbb{R}^{2N}}{\arg\min} \sum_{i=1}^{N} \left(\omega_i \sqrt{\bar{y}_i^2 + \bar{y}_{i+N}^2} + \frac{1}{2\lambda}\left((\bar{x}_i - \bar{y}_i)^2 + (\bar{x}_{i+N} - \bar{y}_{i+N})^2\right)\right),
\end{aligned}$$

where $\bar{\mathbf{x}} := \xi(\mathbf{x})$. This problem can be decomposed into N independent problems $(\forall i \in \{1, \ldots, N\})$

$$\bar{\mathbf{p}}_i \in \underset{\mathbf{y}\in\mathbb{R}^2}{\arg\min}\ \Theta_i(\mathbf{y}); \qquad \Theta_i(\mathbf{y}) := \omega_i\|\mathbf{y}\|_2 + \frac{1}{2\lambda}\|\mathbf{y} - \bar{\mathbf{x}}_i\|_2^2,$$

where $\bar{\mathbf{x}}_i := [\bar{x}_i, \bar{x}_{i+N}]^T$. In case $\mathbf{y} \neq \mathbf{0}$, the gradient of Θ_i is given by

$$\Delta\Theta_i(\mathbf{y}) = \frac{\omega_i}{\|\mathbf{y}\|_2}\mathbf{y} + \frac{1}{\lambda}(\mathbf{y} - \bar{\mathbf{x}}_i).$$

Consequently, if $\bar{\mathbf{p}} \neq \mathbf{0}$, we have

$$\begin{aligned}
\frac{\omega_i}{\|\bar{\mathbf{p}}_i\|_2}\bar{\mathbf{p}}_i + \frac{1}{\lambda}(\bar{\mathbf{p}}_i - \bar{\mathbf{x}}_i) &\stackrel{!}{=} \mathbf{0} \\
\bar{\mathbf{p}}_i\left(\frac{\omega_i}{\|\bar{\mathbf{p}}_i\|_2} + \frac{1}{\lambda}\right) &= \frac{1}{\lambda}\bar{\mathbf{x}}_i \\
\bar{\mathbf{p}}_i &= \left(\frac{\lambda\omega_i}{\|\bar{\mathbf{p}}_i\|_2} + 1\right)^{-1}\bar{\mathbf{x}}_i, \qquad \text{(C.1)}
\end{aligned}$$

where

$$\|\bar{\mathbf{p}}_i\|_2 = \left\| \left(\frac{\lambda\omega_i}{\|\bar{\mathbf{p}}_i\|_2} + 1 \right)^{-1} \bar{\mathbf{x}}_i \right\|_2 = \left(\frac{\lambda\omega_i}{\|\bar{\mathbf{p}}_i\|_2} + 1 \right)^{-1} \|\bar{\mathbf{x}}_i\|_2 \,, = \|\bar{\mathbf{x}}_i\|_2 - \lambda\omega_i.$$

Substituting into (C.1) yields

$$\bar{\mathbf{p}}_i = \left(\frac{\lambda\omega_i}{\|\bar{\mathbf{x}}_i\|_2 - \lambda\omega_i} + 1 \right)^{-1} \bar{\mathbf{x}}_i = \left(1 - \frac{\lambda\omega_i}{\|\bar{\mathbf{x}}_i\|_2} \right) \bar{\mathbf{x}}_i,$$

where the condition $\mathbf{y} \neq \mathbf{0}$ is equivalent to $\|\bar{\mathbf{p}}_i\|_2 > 0$, which gives $\|\bar{\mathbf{p}}_i\|_2 + \lambda\omega_i = \|\bar{\mathbf{x}}_i\|_2 > \lambda\omega_i$.

For $\bar{\mathbf{p}}_i = \mathbf{0}$, the gradient of $f_i(\mathbf{y}) := \omega\|\mathbf{y}\|_2$ does not exist. However, we can consider a subgradient $\mathbf{g}_i \in \partial f_i(\mathbf{0})$, where

$$\partial f_i(\mathbf{y}_0) := \left\{ \mathbf{z} \in \mathbb{R}^2 |\ (\forall \mathbf{y} \in \mathbb{R}^2)\ \mathbf{z}^T(\mathbf{y} - \mathbf{y}_0) + f_i(\mathbf{y}_0) \leq f_i(\mathbf{y}) \right\}$$

is the subdifferential of f_i at $\mathbf{y}_0 \in \mathbb{R}^2$. Consequently, we have

$$\begin{aligned} \mathbf{g}_i \in \partial f_i(\mathbf{0}) &= \left\{ \mathbf{z} \in \mathbb{R}^2 |\ (\forall \mathbf{y} \in \mathbb{R}^2)\ \mathbf{z}^T\mathbf{y} \leq f_i(\mathbf{y}) - f_i(\mathbf{0}) \right\} \\ &= \left\{ \mathbf{z} \in \mathbb{R}^2 |\ (\forall \mathbf{y} \in \mathbb{R}^2)\ \mathbf{z}^T\mathbf{y} \leq \omega_i\|\mathbf{y}\| \right\}, \end{aligned}$$

which implies $\|\mathbf{g}_i\| \leq \omega_i$. Setting the subdgradient of Θ_i to zero yields

$$\partial\Theta_i(\mathbf{0}) \ni \mathbf{g}_i - \frac{1}{\lambda}\bar{\mathbf{x}}_i \stackrel{!}{=} \mathbf{0} \qquad \Rightarrow \qquad \mathbf{g}_i = \frac{1}{\lambda}\bar{\mathbf{x}}_i,$$

whereby $\|\bar{\mathbf{x}}_i\|_2 = \lambda\|\mathbf{g}_i\|_2 \leq \lambda\omega_i$. Putting both cases together, we have

$$\bar{\mathbf{p}}_i = \begin{cases} \left(1 - \frac{\lambda\omega_i}{\|\bar{\mathbf{x}}_i\|_2}\right) \bar{\mathbf{x}}_i, & \text{if } \|\bar{\mathbf{x}}_i\|_2 > \lambda\omega_i \\ \mathbf{0}, & \text{otherwise,} \end{cases}$$

or equivalently,[1]

$$\bar{\mathbf{p}}_i = \bar{\mathbf{x}}_i \left(1 - \frac{\lambda\omega_i}{\|\bar{\mathbf{x}}_i\|_2} \right)_+.$$

Now we can define a vector $\bar{\mathbf{p}} \in \mathbb{R}^{2N}$ by $(\forall n \in \mathbb{N})\ \bar{p}_i = \bar{\mathbf{p}}_i|_1$ and $\bar{p}_{N+i} = \bar{\mathbf{p}}_i|_2$, and apply the inverse mapping $\mathbf{p} = \xi^{-1}(\bar{\mathbf{p}}) = \bar{\mathbf{p}}_{1:N} + j\bar{\mathbf{p}}_{N+1:2N}$, resulting in $(\forall i \in \{1, \ldots, N\})$

$$p_i = x_i \left(1 - \frac{\lambda\omega_i}{|x_i|} \right)_+.$$

[1] For notational convenience, we use the convention that $\frac{1}{0} = \infty$ and $\frac{0}{0} = 1$.

Finally, we obtain the proximal mapping $\operatorname{prox}_{\lambda\|\cdot\|_1^\omega} : \mathcal{H} \to \mathcal{H} : \mathbf{x} \mapsto \mathbf{p}$ by rearranging N complex coefficients p_i into a vector

$$\mathbf{p} = \sum_{i=1}^{N} x_i \left(1 - \frac{\lambda\omega_i}{|x_i|}\right)_+ \mathbf{e}_i,$$

which completes the proof. □

C.2. Proof of Proposition 5.7

In the following, we derive the CRLBs in Proposition 5.7. A very similar derivation can be found in [RB74], where the goal is to estimate the frequency, phase, and magnitude of a single tone based on discrete time observations. The derivation below follows along the lines of [RB74].

We consider the general case in which pilot signals are transmitted on a subset $\mathcal{P} \subset \mathcal{I}_{\mathrm{sc}}$ of subcarriers. To simplify the subsequent analysis, we write the entries of $\hat{\mathbf{y}}_k$ in polar form as $(\forall k \in \mathcal{P})(\forall l \in \{1, \ldots, L_{\mathrm{R}}\})\ \hat{y}_{l,k} = r_{l,k} e^{j\psi_{l,k}}$ and perform a change of variables by a unitary transform $\mathbf{U}_k := \operatorname{diag}\left(e^{-j\psi_{1,k}}, \ldots, e^{-j\psi_{L_{\mathrm{R}},k}}\right)$, whereby

$$\begin{aligned}(\forall k \in \mathcal{P}) \qquad \mathbf{z}_k = \mathbf{a}_k + j\mathbf{b}_k := \mathbf{U}_k \tilde{\mathbf{y}}_k &= e^{j\varphi} e^{-j2\pi(k-1)\tau} \mathbf{U}_k \hat{\mathbf{y}}_k + \mathbf{U}_k \mathbf{n}_k \\ &= e^{j\varphi} e^{-j2\pi(k-1)\tau} \mathbf{r}_k + \tilde{\mathbf{n}}_k.\end{aligned}$$

Here, $(\forall k \in \mathcal{P})\ \mathbf{r}_k \in \mathbb{R}^{L_{\mathrm{R}}}$ is the vector comprised of the magnitudes of $\hat{\mathbf{y}}_k$, and $\tilde{\mathbf{n}}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, since $\mathbf{U}_k$ is a unitary matrix, i.e., $\mathbf{U}_k^H \mathbf{U}_k = \mathbf{I}$. Denoting the entries of $\mathbf{a}_k, \mathbf{b}_k, \mathbf{r}_k$ and $\tilde{\mathbf{n}}_k$ by $a_{l,k}, b_{l,k}, r_{l,k}$ and $\tilde{n}_{l,k}$, respectively, we can express the joint probability density function (pdf) of all entries of $\mathbf{Z} := [\mathbf{z}_{k_1}, \ldots, \mathbf{z}_{k_{|\mathcal{P}|}}] \in \mathbb{C}^{L_{\mathrm{R}} \times |\mathcal{P}|}$ for the unknown parameter vector $\boldsymbol{\alpha} = (\varphi, \theta)^T$ as

$$\begin{aligned} f(\mathbf{Z}, \boldsymbol{\alpha}) &= \prod_{l=1}^{L_{\mathrm{R}}} \prod_{k \in \mathcal{P}} \left(\frac{1}{\sigma^2\pi}\right) \exp\left(\frac{(a_{l,k} - \mu_{l,k}(\varphi,\tau))^2 + (b_{l,k} - \nu_{l,k}(\varphi,\tau))^2}{\sigma^2}\right) \\ &= \left(\frac{1}{\sigma^2\pi}\right)^{L_{\mathrm{R}}|\mathcal{P}|} \exp\left(\frac{\sum_{l=1}^{L_{\mathrm{R}}} \sum_{k \in \mathcal{P}} \left((a_{l,k} - \mu_{l,k}(\varphi,\tau))^2 + (b_{l,k} - \nu_{l,k}(\varphi,\tau))^2\right)}{\sigma^2}\right), \end{aligned} \tag{C.2}$$

where $(\forall l \in \{1, \ldots, L_{\mathrm{R}}\})(\forall k \in \mathcal{P})$

$$\begin{aligned} \mu_{l,k}(\varphi,\tau) &= r_{l,k} \cos(\varphi - 2\pi(k-1)\tau), \\ \nu_{l,k}(\varphi,\tau) &= r_{l,k} \sin(\varphi - 2\pi(k-1)\tau). \end{aligned}$$

The Cramér-Rao bounds for the estimation of φ and τ correspond to the diagonal entries of the inverse of the Fisher information matrix $\mathbf{J} \in \mathbb{R}^{2\times 2}$. For the parameter vector $\boldsymbol{\alpha} = (\varphi, \theta)^T$, the entries of $\mathbf{J}$ can be written as (see also [RB74])

$$(\forall i, j \in \{1, 2\}) \quad J_{ij} = \mathbb{E}_{\mathbf{Z}}[H_{\alpha_i} H_{\alpha_j}],$$

where the expectation is w.r.t. the sample matrix $\mathbf{Z}$ and

$$H_{\alpha_i} = \frac{\partial}{\partial \alpha_i} \ln f(\mathbf{Z}, \boldsymbol{\alpha})$$

is the derivative of the log-likelihood function w.r.t. α_i. According to (C.2) we can write

$$H_{\alpha_i} = \frac{2}{\sigma^2} \sum_{l=1}^{L_\mathrm{R}} \sum_{k\in\mathcal{P}} \left((\mu_{l,k}(\boldsymbol{\alpha}) - a_{l,k}) \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} + (\nu_{l,k}(\boldsymbol{\alpha}) - b_{l,k}) \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \right).$$

Applying the law of total expectation, i.e., $\mathbb{E}_{\mathbf{Z}} = \mathbb{E}_{\mathbf{R}}[\mathbb{E}_{\tilde{\mathbf{N}}}[\mathbf{Z}|\mathbf{R}]]$, where $\mathbf{R} = [\mathbf{r}_{k_1}, \ldots, \mathbf{r}_{k_{|\mathcal{P}|}}]$ and $\tilde{\mathbf{N}} = [\tilde{\mathbf{n}}_{k_1}, \ldots, \tilde{\mathbf{n}}_{k|\mathcal{P}|}]$, we can write the the entries of the Fisher information matrix as

$$\begin{aligned}
J_{ij} =& \frac{2}{\sigma^4} \sum_{l=1}^{L_\mathrm{R}} \sum_{k\in\mathcal{P}} \mathbb{E}_{\mathbf{Z}} \Bigg[(\mu_{l,k}(\boldsymbol{\alpha}) - a_{l,k})^2 \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} + (\nu_{l,k}(\boldsymbol{\alpha}) - b_{l,k})^2 \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} \\
&+ (\mu_{l,k}(\boldsymbol{\alpha}) - a_{l,k})(\nu_{l,k}(\boldsymbol{\alpha}) - b_{l,k}) \left(\frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} + \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \right) \Bigg] \\
=& \frac{4}{\sigma^4} \sum_{l=1}^{L_\mathrm{R}} \sum_{k\in\mathcal{P}} \mathbb{E}_{\mathbf{R}} \Bigg[\mathbb{E}_{\tilde{\mathbf{N}}} \Bigg[(\mathrm{Re}^2\{\tilde{n}_{l,k}\} \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} + \mathrm{Im}^2\{\tilde{n}_{l,k}\} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} \\
&+ \mathrm{Re}\{\tilde{n}_{l,k}\}\mathrm{Im}\{\tilde{n}_{l,k}\} \left(\frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} + \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \right) \Bigg] \Bigg| \mathbf{R} \Bigg] \\
=& \frac{2}{\sigma^2} \sum_{l=1}^{L_\mathrm{R}} \sum_{k\in\mathcal{P}} \mathbb{E}_{\mathbf{R}} \left[\frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} + \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_j} \right].
\end{aligned} \tag{C.3}$$

The derivatives of $\mu_{l,k}$ and $\nu_{l,k}$ with respect to the parameters φ and τ are given by $(\forall k \in \mathcal{P})(\forall l \in \{1, \ldots, L_\mathrm{R}\}$

$$\begin{aligned}
\frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_1} &= \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \varphi} = r_{l,k} \sin(\varphi - 2\pi(k-1)\tau) \\
\frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_1} &= \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \varphi} = -r_{l,k} \cos(\varphi - 2\pi(k-1)\tau)
\end{aligned}$$

and

$$\frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_2} = \frac{\partial \mu_{l,k}(\boldsymbol{\alpha})}{\partial \tau} = -2\pi(k-1) r_{l,k} \sin(\varphi - 2\pi(k-1)\tau)$$
$$\frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \alpha_2} = \frac{\partial \nu_{l,k}(\boldsymbol{\alpha})}{\partial \tau} = 2\pi(k-1) r_{l,k} \cos(\varphi - 2\pi(k-1)\tau).$$

Substituting into (C.3) and assuming i.i.d. Gaussian distributed entries $\hat{y}_{l,k}$ with variance γ/σ^2 yields

$$J_{11} = \frac{2}{\sigma^2} \sum_{l=1}^{L_\mathrm{R}} \sum_{k \in \mathcal{P}} \mathbb{E}\left[r_{l,k}^2\right] = 2\gamma L_\mathrm{R} |\mathcal{P}|$$
$$J_{12} = J_{21} = -\frac{2}{\sigma^2} \sum_{l=1}^{L_\mathrm{R}} \sum_{k \in \mathcal{P}} 2\pi(k-1) \mathbb{E}\left[r_{l,k}^2\right] = -4\gamma\pi L_\mathrm{R} C_1$$
$$J_{22} = \frac{2}{\sigma^2} \sum_{l=1}^{L_\mathrm{R}} \sum_{k \in \mathcal{P}} 4\pi^2(k-1)^2 \mathbb{E}\left[r_{l,k}^2\right] = 8\pi^2\gamma L_\mathrm{R} C_2,$$

where $C_1 := \sum_{k\in\mathcal{P}}(k-1)$ and $C_2 := \sum_{k\in\mathcal{P}}(k-1)^2$. By inverting the Fisher information matrix $\mathbf{J}$, we obtain the CRLBs

$$\mathrm{var}\{\varphi\} \geq \frac{J_{22}}{J_{11}J_{22} - J_{12}J_{21}} = \frac{1}{2\gamma L_\mathrm{R}\left(|\mathcal{P}| - \frac{C_1^2}{C_2}\right)} \tag{C.4}$$

and

$$\mathrm{var}\{\tau\} \geq \frac{J_{11}}{J_{11}J_{22} - J_{12}J_{21}} = \frac{1}{8\pi^2\gamma L_\mathrm{R}\left(C_2 - \frac{C_1^2}{|\mathcal{P}|}\right)}. \tag{C.5}$$

For the special case $\mathcal{P} = \mathcal{I}_\mathrm{sc}$, C_1 is a triangular number and C_2 is a square pyramidial number, i.e., we have $C_1 = \frac{1}{2}N_\mathrm{F}(N_\mathrm{F}-1)$ and $C_2 = \frac{1}{6}N_\mathrm{F}(N_\mathrm{F}-1)(2N_\mathrm{F}-1)$. Substituting C_1 and C_2 in (C.4) and (C.5) yields

$$\mathrm{var}\{\varphi\} \geq \frac{2N_\mathrm{F}-1}{\gamma L_\mathrm{R} N_\mathrm{F}(N_\mathrm{F}+1)} \qquad \text{and} \qquad \mathrm{var}\{\tau\} \geq \frac{3}{2\pi^2\gamma L_\mathrm{R} N_\mathrm{F}(N_\mathrm{F}^2-1)},$$

which is the desired result. □

Publication List

[1] J. Fink, R. L. G. Cavalcante, P. Jung, and S. Stańczak, "Extrapolated projection methods for PAPR reduction," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1810–1814, IEEE, 2018.

[2] J. Fink, D. Schäufele, M. Kasparick, R. L. G. Cavalcante, and S. Stańczak, "Cooperative localization by set-theoretic estimation," in *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, pp. 1–8, VDE, 2019.

[3] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Multicast beamforming using semidefinite relaxation and bounded perturbation resilience," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4749–4753, IEEE, 2019.

[4] J. Fink, M. Kasparick, and S. Stańczak, "Robust submodular RRH selection for joint multicast downlink transmission," in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, pp. 1–6, VDE, 2020.

[5] F. Göttsch, J. Fink, and M. Kasparick, "Proactive resource allocation for the coexistence of concurrent services," in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, pp. 1–6, VDE, 2020.

[6] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Online channel estimation for hybrid beamforming architectures," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4995–4999, IEEE, 2020.

[7] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Multi-group multicast beamforming by superiorized projections onto convex sets," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5708–5722, 2021.

[8] A. Palaios, P. Geuer, J. Fink, D. F. Külzer, F. Göttsch, M. Kasparick, D. Schäufele, R. Hernangómez, S. Partani, R. Sattiraju, A. Kumar, F. Burmeister, A. Weinand, C. Vielhaus, G. Fettweis, F. H. P. Fitzek, H. D. Schotten, and S. Stańczak, "Network under control: Multi-vehicle E2E measurements for AI-based QoS prediction," in

2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1432–1438, IEEE, 2021.

[9] A. Palaios, C. Vielhaus, D. F. Külzer, P. Geuer, R. Sattiraju, J. Fink, M. Kasparick, C. Watermann, G. Fettweis, F. H. P. Fitzek, H. D. Schotten, and S. Stańczak, "Effect of spatial, temporal and network features on uplink and downlink throughput prediction," in *2021 IEEE 4th 5G World Forum (5GWF)*, pp. 418–423, IEEE, 2021.

[10] J. Fink, R. L. G. Cavalcante, Z. Utkovski, and S. Stańczak, "A set-theoretic approach to MIMO detection," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022. (to appear).

[11] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Superiorized adaptive projected subgradient method with application to MIMO detection," *IEEE Transactions on Signal Processing*, 2022. (submitted).

Bibliography

[3GP17] 3GPP TR 36.873 v12.5.0. Study on 3D channel model for LTE, June 2017.

[ACYS18] D. A. Awan, R. L. G. Cavalcante, M. Yukawa, and S. Stańczak. Detection for 5G-NOMA: An online adaptive machine learning approach. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.

[ACYS20] D. A. Awan, R. L. G. Cavalcante, M. Yukawa, and S. Stańczak. Adaptive learning for symbol detection: a reproducing kernel Hilbert space approach. *Machine Learning for Future Wireless Communications*, pages 197–211, 2020.

[AEALH14] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, 2014.

[Agm54] S. Agmon. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:382–392, 1954.

[AJS19] M. A. Albreem, M. Juntti, and S. Shahabuddin. Massive MIMO detection techniques: A survey. *IEEE Communications Surveys & Tutorials*, 21(4):3109–3132, 2019.

[ALH15] A. Alkhateeb, G. Leus, and R. W. Heath. Compressed sensing based multi-user millimeter wave systems: How many measurements are needed? In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2909–2913. IEEE, 2015.

[AM03] A. Aggarwal and T. H. Meng. Minimizing the peak-to-average power ratio of OFDM signals via convex optimization. In *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, volume 4, pages 2385–2389. IEEE, 2003.

[AMB17] D. Ampeliotis, C. Mavrokefalidis, and K. Berberidis. Distributed dictionary learning via projections onto convex sets. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 2294–2298. IEEE, 2017.

[Arm02] J. Armstrong. Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering. *Electronics letters*, 38(5):246–247, 2002.

[Aro50] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[BB96] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM review*, 38(3):367–426, 1996.

[BC01] H. H. Bauschke and P. L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of operations research*, 26(2):248–264, 2001.

[BC11] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2 edition, 2011.

[BCK06] H. H. Bauschke, P. L. Combettes, and S. G. Kruk. Extrapolation algorithm for affine-convex feasibility problems. *Numerical Algorithms*, 41(3):239–274, 2006.

[BCL02] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization. *JOSA A*, 19(7):1334–1345, 2002.

[Bec17] A. Beck. *First-order methods in optimization*. SIAM, 2017.

[BH06] D. Blatt and A. O. Hero. Energy-based sensor network source localization via projection onto convex sets. *IEEE Transactions on Signal Processing*, 54(9):3614–3619, 2006.

[BK15] H. H. Bauschke and V. R. Koch. Projection methods: Swiss army knives for solving feasibility and best approximation problems with halfspaces. *Contemporary Mathematics*, 636:1–40, 2015.

[BLW15] T. E. Bogale, L. B. Le, and X. Wang. Hybrid analog-digital channel estimation and beamforming: Training-throughput tradeoff. *IEEE Transactions on Communications*, 63(12):5235–5249, 2015.

[Bre65] L. M. Bregman. Finding the common point of convex sets by the method of successive projection. In *Doklady Akademii Nauk*, volume 162, pages 487–490. Russian Academy of Sciences, 1965.

[BRZ18] C. Bargetz, S. Reich, and R. Zalas. Convergence properties of dynamic string-averaging projection methods in the presence of perturbations. *Numerical Algorithms*, 77(1):185–209, 2018.

[CC15] Y. Censor and A. Cegielski. Projection methods: an annotated bibliography of books and reviews. *Optimization*, 64(11):2343–2358, 2015.

[CCC+12] Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman. On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints. *Computational Optimization and Applications*, 51(3):1065–1088, 2012.

[CCO14a] D. Christopoulos, S. Chatzinotas, and B. Ottersten. Sum rate maximizing multigroup multicast beamforming under per-antenna power constraints. In *2014 IEEE Global Communications Conference*, pages 3354–3359. IEEE, 2014.

[CCO14b] D. Christopoulos, S. Chatzinotas, and B. Ottersten. Weighted fair multicast multigroup beamforming under per-antenna power constraints. *IEEE Transactions on Signal Processing*, 62(19):5132–5142, 2014.

[CCO15] D. Christopoulos, S. Chatzinotas, and B. Ottersten. Multicast multigroup beamforming for per-antenna power constrained large-scale arrays. In *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 271–275. IEEE, 2015.

[CCS10] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[CDH10] Y. Censor, R. Davidi, and G. T. Herman. Perturbation resilience and superiorization of iterative algorithms. *Inverse Problems*, 26(6):065008, 2010.

[CDH+14] Y. Censor, R. Davidi, G. T. Herman, R. W. Schulte, and L. Tetruashvili. Projected subgradient minimization versus superiorization. *Journal of Optimization Theory and Applications*, 160(3):730–747, 2014.

[Ceg12] A. Cegielski. *Iterative methods for fixed point problems in Hilbert spaces*, volume 2057. Springer, 2012.

[CEH01] Y. Censor, T. Elfving, and G. T. Herman. Averaging strings of sequential iterations for convex feasibility problems. In *Studies in Computational Mathematics*, volume 8, pages 101–113. Elsevier, 2001.

[Cen15] Y. Censor. Weak and strong superiorization: Between feasibility-seeking and minimization. *Analele Universitatii "Ovidius" Constanta-Seria Matematica*, 23(3):41–54, 2015.

[Cis14] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018, February 2014. [Online]. Available: `https://blog.acens.com/wp-content/images/cisco-global-mobile-data-2013\T1\textendash2018-blog-acens.pdf`, accessed Jan. 27, 2022.

[Cis19] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022, February 2019. [Online]. Available: `https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf`, accessed Jan. 27, 2022.

[CKT14] S. Chouvardas, Y. Kopsinis, and S. Theodoridis. An adaptive projected subgradient based algorithm for robust subspace tracking. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5497–5501. IEEE, 2014.

[Com93] P. L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, 1993.

[Com97] P. L. Combettes. Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. *IEEE Transactions on Image Processing*, 6(4):493–506, 1997.

[Com01] P. L. Combettes. Quasi-Fejérian analysis of some optimization algorithms. In *Studies in Computational Mathematics*, volume 8, pages 115–152. Elsevier, 2001.

[CP11] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[CSSK18] R. L. G. Cavalcante, S. Stańczak, D. Schupke, and J. Klaue. Low-complexity distributed set-theoretic decoders for analog fountain codes. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 1480–1484. IEEE, 2018.

[CST11] S. Chouvardas, K. Slavakis, and S. Theodoridis. Adaptive robust distributed learning in diffusion sensor networks. *IEEE Transactions on Signal Processing*, 59(10):4692–4707, 2011.

[CT17] E. Chen and M. Tao. ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems. *IEEE Transactions on Communications*, 65(6):2685–2698, 2017.

[CWB08] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[CY08] R. L. G. Cavalcante and I. Yamada. Multiaccess interference suppression in orthogonal space–time block coded MIMO systems by adaptive projected subgradient method. *IEEE Transactions on Signal Processing*, 56(3):1028–1042, 2008.

[CY09] R. L. G. Cavalcante and I. Yamada. A flexible peak-to-average power ratio reduction scheme for OFDM systems by the adaptive projected subgradient method. *IEEE Transactions on Signal Processing*, 57(4):1456–1468, 2009.

[CY15] P. L. Combettes and I. Yamada. Compositions and convex combinations of averaged nonexpansive operators. *Journal of Mathematical Analysis and Applications*, 425(1):55–70, 2015.

[CYM09] R. L. G. Cavalcante, I. Yamada, and B. Mulgrew. An adaptive projected subgradient approach to learning in diffusion networks. *IEEE Transactions on Signal Processing*, 57(7):2762–2774, 2009.

[CYX+19] X. Cheng, Y. Yang, B. Xia, N. Wei, and S. Li. Sparse channel estimation for millimeter wave massive MIMO systems with lens antenna array. *IEEE Transactions on Vehicular Technology*, 2019.

[CZ13] Y. Censor and A. J. Zaslavski. Convergence and perturbation resilience of dynamic string-averaging projection methods. *Computational Optimization and Applications*, 54(1):65–76, 2013.

[Don06] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[DPDM17] A. Del Pia, S. S. Dey, and M. Molinaro. Mixed-integer quadratic programming is in NP. *Mathematical Programming*, 162(1):225–240, 2017.

[DSSSC08] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, pages 272–279. ACM, 2008.

[EBL+15] D. Eustice, C. Baylis, C. Latham, R. J. Marks, and L. Cohen. Optimizing radar waveforms using generalized alternating projections. In *2015 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*, pages 1–6. IEEE, 2015.

[Erd06] A. T. Erdogan. A low complexity multicarrier PAR reduction approach based on subgradient optimization. *Signal Processing*, 86(12):3890–3903, 2006.

[Ere65] I. Eremin. A generalization of the Motzkin-Agmon relaxation method. *Uspekhi Mathematicheski Nauk*, 20(1):5, 1965.

[Eri21] Ericsson. Ericsson mobility report, June 2021. [Online]. Available: `https://www.ericsson.com/4a03c2/assets/local/reports-papers/mobility-report/documents/2021/june-2021-ericsson-mobility-report.pdf`, accessed Jan. 27, 2022.

[EY36] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[Fra19] Fraunhofer HHI. Quasi deterministic radio channel generator (QuaDRiGa), 2019. [Online]. Available: `https://quadriga-channel-model.de/`, accessed Aug. 08, 2019.

[FZ08] G. Fettweis and E. Zimmermann. ICT energy consumption-trends and challenges. In *Proceedings of the 11th international symposium on wireless personal multimedia communications.* Citeseer, 2008.

[GCL+10] J.-J. Gao, X.-H. Chen, J.-Y. Li, G.-C. Liu, and J. Ma. Irregular seismic data reconstruction based on exponential threshold model of POCS method. *Applied Geophysics*, 7(3):229–238, 2010.

[GDN18] G. Gao, C. Dong, and K. Niu. Sparsely connected neural network for massive MIMO detection. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pages 397–402. IEEE, 2018.

[Gin18] O. Ginat. The method of alternating projections. *arXiv preprint arXiv:1809.05858*, 2018.

[GMZ+16] H. Guo, X. Ma, Z. Zhang, B. Zhang, C. Yuan, and F. Huang. POCS-enhanced inherent correction of motion-induced phase errors (POCS-ICE) for high-resolution multishot diffusion MRI. *Magnetic resonance in medicine*, 75(1):169–180, 2016.

[GNR20] P. J. Goulart, Y. Nakatsukasa, and N. Rontsis. Accuracy of approximate projection to the semidefinite cone. *Linear Algebra and its Applications*, 594:177–192, 2020.

[GP97] A. Gatherer and M. Polley. Controlling clipping probability in DMT transmission. In *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, volume 1, pages 578–584. IEEE, 1997.

[GPR67] L. G. Gurin, B. T. Polyak, and È. V. Raik. The method of projections for finding the common point of convex sets. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(6):1211–1228, 1967.

[GWC+15] S. Gan, S. Wang, Y. Chen, Y. Zhang, and Z. Jin. Dealiased seismic data interpolation using seislet transform with low-frequency constraint. *IEEE Geoscience and remote sensing letters*, 12(10):2150–2154, 2015.

[Hal62] I. Halperin. The product of projection operators. *Acta Sci. Math.(Szeged)*, 23(1):96–99, 1962.

[HGPR+16] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE journal of selected topics in signal processing*, 10(3):436–453, 2016.

[HJ13] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2 edition, 2013.

[HS16] K. Huang and N. D. Sidiropoulos. Consensus-admm for general quadratically constrained quadratic programming. *IEEE Transactions on Signal Processing*, 64(20):5297–5310, 2016.

[HWJL18a] H. He, C.-K. Wen, S. Jin, and G. Y. Li. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems. *IEEE Wireless Communications Letters*, 7(5):852–855, 2018.

[HWJL18b] H. He, C.-K. Wen, S. Jin, and G. Y. Li. A model-driven deep learning network for MIMO detection. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 584–588. IEEE, 2018.

[HX17] H. He and H.-K. Xu. Perturbation resilience and superiorization methodology of averaged mappings. *Inverse Problems*, 33(4):044007, 2017.

[ITW19] D. Ito, S. Takabe, and T. Wadayama. Trainable ISTA for sparse signal recovery. *IEEE Transactions on Signal Processing*, 67(12):3113–3125, 2019.

[JAA04] P. K. Jain, O. P. Ahuja, and K. Ahmad. *Functional Analysis*. New Age International (P) Ltd., 2004.

[JGMS15] C. Jeon, R. Ghods, A. Maleki, and C. Studer. Optimality of large MIMO detection via approximate message passing. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1227–1231. IEEE, 2015.

[Kac37] S. Kaczmarz. Przyblizone rozwiazywanie ukladów równan liniowych. – Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles*, Série A:355–357, 1937.

[KAHF20] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming. Adaptive neural signal detection for massive MIMO. *IEEE Transactions on Wireless Communications*, 19(8):5635–5648, 2020.

[KB14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KCKP16] Y. Kim, O. Choi, Y. Kim, and J. Park. Performance analysis of LTE multi-antenna technology in live network. In *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, pages 1302–1305. IEEE, 2016.

[KCV+15] M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stańczak, and M. Yukawa. Kernel-based adaptive online reconstruction of coverage maps with side information. *IEEE Transactions on Vehicular Technology*, 65(7):5461–5473, 2015.

[KJ03] B. S. Krongold and D. L. Jones. PAR reduction in OFDM via active constellation extension. *IEEE Transactions on broadcasting*, 49(3):258–268, 2003.

[KJ04] B. S. Krongold and D. L. Jones. An active-set approach for OFDM PAR reduction via tone reservation. *IEEE transactions on signal processing*, 52(2):495–509, 2004.

[KKL+07] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE journal of selected topics in signal processing*, 1(4):606–617, 2007.

[KSL08] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo. Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups. *IEEE Transactions on Signal Processing*, 56(3):1268–1279, 2008.

[LGL16] J. Lee, G.-T. Gil, and Y. H. Lee. Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. *IEEE Transactions on Communications*, 64(6):2370–2386, 2016.

[LMS+10] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[Luk13] D. R. Luke. Prox-regularity of rank constraint sets and implications for algorithms. *Journal of Mathematical Imaging and Vision*, 47(3):231–238, 2013.

[LVdP17] E. G. Larsson and L. Van der Perre. Massive MIMO for 5G. *IEEE 5G Tech Focus*, 2017.

[LWP13] K. L. Law, X. Wen, and M. Pesavento. General-rank transmit beamforming for multi-group multicasting networks using OSTBC. In *2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 475–479. IEEE, 2013.

[LYSM17] H. Liu, M.-C. Yue, A. M.-C. So, and W.-K. Ma. A discrete first-order method for large-scale MIMO detection with provable guarantees. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2017.

[MCS18] L. Miretti, R. L. G. Cavalcante, and S. Stańczak. FDD massive MIMO channel spatial covariance conversion using projection methods. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3609–3613. IEEE, 2018.

[MCS21] L. Miretti, R. L. G. Cavalcante, and S. Stańczak. Channel covariance conversion and modelling using infinite dimensional Hilbert spaces. *IEEE Transactions on Signal Processing*, 69:3145–3159, 2021.

[MH99] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE transactions on Information Theory*, 45(1):139–157, 1999.

[MH09] R. K. Martin and M. Haker. Reduction of peak-to-average power ratio in transform domain communication systems. *IEEE Transactions on Wireless Communications*, 8(9), 2009.

[MHG+14] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos. Feasible point pursuit and successive approximation of non-convex QCQPs. *IEEE Signal Processing Letters*, 22(7):804–808, 2014.

[Mic01] D. Micciancio. The hardness of the closest vector problem with preprocessing. *IEEE Transactions on Information Theory*, 47(3):1212–1215, 2001.

[Mir60] L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, 11(1):50–59, 1960.

[Mol05] A. F. Molisch. Ultrawideband propagation channels-theory, measurement, and modeling. *IEEE Transactions on Vehicular Technology*, 54(5):1528–1545, 2005.

[MP17] J. Ma and L. Ping. Orthogonal AMP. *IEEE Access*, 5:2020–2033, 2017.

[MRRA+15] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath. Channel estimation and hybrid combining for mmWave: Phase shifters or switches? In *2015 Information Theory and Applications Workshop (ITA)*, pages 90–97. IEEE, 2015.

[MS54] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[Nak53] H. Nakano. *Spectral theory in the Hilbert space*, volume 4. Japan Society for the Promotion of Science, 1953.

[Nes18] Y. Nesterov. *Lectures on convex optimization*, volume 137. Springer, 2018.

[NL20] N. T. Nguyen and K. Lee. Deep learning-aided tabu search detection for large MIMO systems. *IEEE Transactions on Wireless Communications*, 19(6):4262–4275, 2020.

[NS06] A. Netyanun and D. C. Solmon. Iterated products of projections in Hilbert space. *The American Mathematical Monthly*, 113(7):644–648, 2006.

[Opi67] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73(4):591–597, 1967.

[OY02] N. Ogura and I. Yamada. Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping. *Numerical Functional Analysis and Optimization*, 23:113–137, 2002.

[Pie75] G. Pierra. Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique. In *IFIP Technical Conference on Optimization Techniques*, pages 200–218. Springer, 1975.

[Pie84] G. Pierra. Decomposition through formalization in a product space. *Mathematical Programming*, 28(1):96–115, 1984.

[Pol69] B. T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29, 1969.

[PWK09] M. Petermann, D. Wübben, and K.-D. Kammeyer. Joint constellation extension and tone reservation for PAPR reduction in adaptive OFDM systems. In *2009 IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, pages 439–443. IEEE, 2009.

[RB74] D. Rife and R. Boorstyn. Single tone parameter estimation from discrete-time observations. *IEEE Transactions on information theory*, 20(5):591–598, 1974.

[RM13] Y. Rahmatallah and S. Mohan. Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy. *IEEE communications surveys & tutorials*, 15(4):1567–1592, 2013.

[RS89] P. A. Regalia and M. K. Sanjit. Kronecker products, unitary matrices and signal processing applications. *SIAM review*, 31(4):586–613, 1989.

[RSS06] M. Rydstrom, E. G. Strom, and A. Svensson. Robust sensor network positioning based on projections onto circular and hyperbolic convex sets (POCS). In *2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, pages 1–5. IEEE, 2006.

[SA09] M. Senst and G. Ascheid. Optimal output back-off in OFDM systems with nonlinear power amplifiers. In *2009 IEEE International Conference on Communications*, pages 1–6. IEEE, 2009.

[SDL06] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo. Transmit beamforming for physical-layer multicasting. *IEEE Trans. Signal Processing*, 54(6-1):2239–2251, 2006.

[SDW17] N. Samuel, T. Diskin, and A. Wiesel. Deep MIMO detection. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2017.

[SDW19] N. Samuel, T. Diskin, and A. Wiesel. Learning to detect. *IEEE Transactions on Signal Processing*, 67(10):2554–2564, 2019.

[SGAK03] M. Sharif, M. Gharavi-Alkhansari, and B. H. Khalaj. On the peak-to-average power of OFDM signals based on oversampling. *IEEE Transactions on Communications*, 51(1):72–78, 2003.

[SKPJ04] A. A. Samsonov, E. G. Kholmovski, D. L. Parker, and C. R. Johnson. POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(6):1397–1406, 2004.

[SS84] M. I. Sezan and H. Stark. Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion. *IEEE transactions on medical imaging*, 3(2):91–98, 1984.

[STY08] K. Slavakis, S. Theodoridis, and I. Yamada. Online kernel-based classification using adaptive projection algorithms. *IEEE Transactions on Signal Processing*, 56(7):2781–2796, 2008.

[STY09] K. Slavakis, S. Theodoridis, and I. Yamada. Adaptive constrained learning in reproducing kernel Hilbert spaces: The robust beamforming case. *IEEE Transactions on Signal Processing*, 57(12):4744–4764, 2009.

[SY98] H. Stark and Y. Yang. *Vector space projections: a numerical approach to signal and image processing, neural nets, and optics*. John Wiley & Sons, Inc., 1998.

[SYCD18] B.-S. Shin, M. Yukawa, R. L. G. Cavalcante, and A. Dekorsy. Distributed adaptive learning with multiple kernels in diffusion networks. *IEEE Transactions on Signal Processing*, 66(21):5505–5519, 2018.

[TAXH16] C. Thrampoulidis, E. Abbasi, W. Xu, and B. Hassibi. BER analysis of the box relaxation for BPSK signal recovery. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3776–3780. IEEE, 2016.

[TBL+18] K. C. Tezcan, C. F. Baumgartner, R. Luechinger, K. P. Pruessmann, and E. Konukoglu. MR image reconstruction using deep density priors. *IEEE transactions on medical imaging*, 38(7):1633–1642, 2018.

[TC98] J. Tellado and J. M. Cioffi. Peak power reduction for multicarrier transmission. In *IEEE GLOBECOM*, volume 99, pages 5–9, 1998.

[Tel99] E. Telatar. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595, 1999.

[Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[TP20] D. Taleb and M. Pesavento. General rank beamforming using full rate real-value OSTBC for multicasting networks. In *WSA 2020; 24th International ITG Workshop on Smart Antennas*, pages 1–5. VDE, 2020.

[TRL01] P. H. Tan, L. K. Rasmussen, and T. J. Lim. Constrained maximum-likelihood detection in CDMA. *IEEE Transactions on Communications*, 49(1):142–153, 2001.

[TSY10] S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *IEEE Signal Processing Magazine*, 28(1):97–123, 2010.

[TTT99] K.-C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3—a MATLAB software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

[TW20a] S. Takabe and T. Wadayama. Convergence acceleration via chebyshev step: Plausible interpretation of deep-unfolded gradient descent. *arXiv preprint arXiv:2010.13335*, 2020.

[TW20b] S. Takabe and T. Wadayama. Theoretical interpretation of learned step size in deep-unfolded gradient descent. *arXiv preprint arXiv:2001.05142*, 2020.

[USMC19] M.-W. Un, M. Shao, W.-K. Ma, and P. C. Ching. Deep MIMO detection using ADMM unfolding. In *2019 IEEE Data Science Workshop (DSW)*, pages 333–337. IEEE, 2019.

[VAHP17] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic. Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6493–6497. IEEE, 2017.

[vEB81] P. van Emde Boas. Another NP-complete problem and the complexity of computing short vectors in a lattice. *Tecnical Report, Department of Mathmatics, University of Amsterdam*, 1981.

[Ver89] S. Verdú. Computational complexity of optimum multiuser detection. *Algorithmica*, 4(1):303–312, 1989.

[VLG83] C. F. Van Loan and G. H. Golub. *Matrix computations*. Johns Hopkins University Press Baltimore, 1983.

[VLH13] C. Voicu, E. S. Lohan, and S. Halunga. Multiuser detection in DS-CDMA systems using POCS algorithm. In *2013 21st Telecommunications Forum Telfor (TELFOR)*, pages 188–191. IEEE, 2013.

[vN33] J. von Neumann. Functional operators, vol. 2 (Annals of Mathematics Studies, no. 22), Princeton, NJ, 1950. *Reprinted from mimeographed lecture notes first distributed in*, 1933.

[WBJ15a] P. Walk, H. Becker, and P. Jung. OFDM channel estimation via phase retrieval. In *Signals, Systems and Computers, 2015 49th Asilomar Conference on*, pages 1161–1168. IEEE, 2015.

[WBJ15b] P. Walk, H. Becker, and P. Jung. Phaseless pilots for OFDM. In *Wireless Communication Systems (ISWCS), 2015 International Symposium on*, pages 86–90. IEEE, 2015.

[WDCS16] M. Wu, C. Dick, J. R. Cavallaro, and C. Studer. High-throughput data detection for massive MU-MIMO-OFDM using coordinate descent. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 63(12):2357–2367, 2016.

[WFB+13] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J.-S. No. The PAPR problem in OFDM transmission: New directions for a long-lasting problem. *IEEE Signal Processing Magazine*, 30(6):130–144, 2013.

[Wie55] N. Wiener. On the factorization of matrices. *Commentarii Mathematici Helvetici*, 29(1):97–111, 1955.

[WMS13] S. X. Wu, W.-K. Ma, and A. M.-C. So. Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time coding. *IEEE Transactions on Signal Processing*, 61(17):4230–4245, 2013.

[WSM12] S. X. Wu, A. M.-C. So, and W.-K. Ma. Rank-two transmit beamformed Alamouti space-time coding for physical-layer multicasting. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2793–2796. IEEE, 2012.

[WT05] L. Wang and C. Tellambura. A simplified clipping and filtering technique for PAR reduction in OFDM systems. *IEEE signal processing letters*, 12(6):453–456, 2005.

[WZQZ10] Q. Wang, L. Zhao, J. Qiao, and C. Zou. Acoustic feedback cancellation based on weighted adaptive projection subgradient method in hearing aids. *Signal processing*, 90(1):69–79, 2010.

[YCY05] M. Yukawa, R. L. G. Cavalcante, and I. Yamada. Efficient adaptive blind MAI suppression in DS/CDMA by embedded constraint parallel projection techniques. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 3, pages iii–929. IEEE, 2005.

[YGC12] P. Yang, J. Gao, and W. Chen. Curvelet-based POCS interpolation of nonuniformly sampled seismic records. *Journal of Applied Geophysics*, 79:90–99, 2012.

[YH15] S. Yang and L. Hanzo. Fifty years of MIMO detection: The road to large-scale MIMOs. *IEEE Communications Surveys & Tutorials*, 17(4):1941–1988, 2015.

[YO05] I. Yamada and N. Ogura. Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions. *Numerical Functional Analysis and Optimization*, 25(7-8):593–617, 2005.

[YSY02] I. Yamada, K. Slavakis, and K. Yamada. An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques. *IEEE Transactions on Signal Processing*, 50(5):1091–1101, 2002.

[YW82] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1 - theory. *IEEE transactions on medical imaging*, 1(2):81–94, 1982.

[YWJG18] J. Yang, C.-K. Wen, S. Jin, and F. Gao. Beamspace channel estimation in mmWave systems via cosparse image reconstruction technique. *IEEE Transactions on Communications*, 66(10):4767–4782, 2018.

[YY06] M. Yukawa and I. Yamada. Adaptive quadratic-metric parallel subgradient projection algorithm and its application to acoustic echo cancellation. In *2006 14th European Signal Processing Conference*, pages 1–5. IEEE, 2006.

[ZLW+15] Y. Zheng, S. E. Li, J. Wang, D. Cao, and K. Li. Stability and scalability of homogeneous vehicular platoon: Study on the influence of information flow topologies. *IEEE Transactions on intelligent transportation systems*, 17(1):14–26, 2015.

[ZZZ+20] H. Zhang, H. Zhang, J. Zhang, Y. Hao, and B. Wang. An anti-aliasing POCS interpolation method for regularly undersampled seismic data using curvelet transform. *Journal of Applied Geophysics*, 172:103894, 2020.