KTH ROYAL INSTITUTE
OF TECHNOLOGY

Doctoral Thesis in Electrical Engineering

Optimization and Learning for Large-Scale MIMO-OFDM Wireless Systems

Theory, Algorithms, and Applications

SHASHI KANT

Stockholm, Sweden 2022

Optimization and Learning for Large-Scale MIMO-OFDM Wireless Systems

Theory, Algorithms, and Applications

SHASHI KANT

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 2nd December 2022, at 2:00 p.m. in F3, Lindstedtsvägen 26, Stockholm

Doctoral Thesis in Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden 2022

© Shashi Kant

ISBN 978-91-8040-386-3
TRITA-EECS-AVL-2022:66

Printed by: Universitetsservice US-AB, Sweden 2022

Abstract

The requirements for next-generation wireless communications networks, particularly fifth-generation (5G) and beyond, are driven by at least three broad use cases. These include enhanced mobile broadband services to support extremely high data rates in terms of network or per user in both uplink and downlink, massive machine-type communications to accommodate massive internet-of-things applications, and critical machine-type communications to handle mission-critical applications that require ultra-high reliability and low latency.

These new-generation wireless communication systems adopt orthogonal frequency division multiplexing (OFDM) with cyclic prefix and multiple antennas at the transmitter and receiver (MIMO). There are many attractive characteristics of OFDM, namely robustness to the adverse effects of time dispersion due to multipath fading, simplicity in equalization, and flexibility in supporting both low and high symbol rates—thereby supporting a variety of various quality-of-service requirements.

It has been known for a long time that OFDM has problems with high out-of-band emissions (OOBE) and high peak-to-average-power ratio (PAPR). The OOBE must be adequately suppressed since high OOBE causes significant interference in the adjacent channels. Furthermore, high PAPR typically requires expensive linear radio frequency (RF) transmitter components and consequently costly digital predistortion to manage and mitigate OOBE resulting from the distortion caused by RF components, *e.g.*, power amplifiers. Additionally, there are practical 5G standard constraints, which necessitate using only data-carrying subcarriers for OOBE and PAPR reduction. Hence, it is of utmost importance to reduce OOBE and PAPR for MIMO-OFDM-based systems and mitigate/minimize the signal distortion at the receiver(s) to meet the new generation systems' requirements encompassing various use cases.

In this thesis, we seek principled approaches to tackle the distortion-based OOBE and PAPR reduction problems. More specifically, we present optimization formulations for these well-known issues in large-scale MIMO-OFDM-based systems, such as 5G New Radio (NR), and future extensions thereof. Unfortunately, these problems cannot be solved via a general-purpose optimization solver since these off-the-shelf solvers typically employ interior-point-based methods, which have prohibitive complexity for state-of-the-art radio hardware systems. Hence, we propose large-scale optimization techniques to tackle these problems resulting in implementation-friendly algorithms. More concretely, we develop (near) optimal and computationally-efficient data-dependent solutions by proposing a type of three-operator alternating direction method of multipliers (ADMM) method that essentially employs a divide-and-conquer approach to solve the huge and cumbersome OOBE and PAPR reduction problems in large-scale MIMO-OFDM-based systems. Moreover, in the last part of the thesis, we also investigate the application of our proposed three-operator ADMM (TOP-ADMM) for federated learning (FL) over networks that capitalize on the potentially rich datasets generated at the physical layer and/or RF hardware of a base station located near an edge server.

In summary, this thesis develops principled, implementation-friendly, and standards-

agnostic algorithms for distortion-based OOBE and PAPR reduction algorithms using first-order optimization algorithms which provides insights into the trade-off between computational complexities and in-band and out-of-band performance. Furthermore, we develop a novel yet simple TOP-ADMM first-order algorithm suitable for tackling centralized and distributed optimization problems. Additionally, this thesis studies the feasibility of the TOP-ADMM algorithm for edge FL exploiting rich datasets available at the base station(s) besides (private) datasets at the users. Finally, this thesis may provide input to the systemization and implementation of large-scale MIMO-OFDM-based wireless communication systems.

Sammanfattning

Kraven för nästa generations trådlösa kommunikationsnätverk, särskilt femte generationens (5G) och senare, drivs av minst tre breda användningsfall. Dessa inkluderar förbättrade mobila bredbandstjänster för att stödja extremt höga datahastigheter i form av systemkapacitet eller per användare i både upplänk och nedlänk, kommunikation av maskintyp för att tillgodose tillgodose storskaliga internet-of-things-applikationer, och kritisk kommunikation av maskintyp för att hantera verksamhetskritiska applikationer som kräver ultrahög tillförlitlighet och låg latens.

Dessa nya generationens trådlösa kommunikationssystem använder ortogonal frekvensdelningsmultiplexering (OFDM) med cykliskt prefix och flera antenner vid både sändaren och mottagaren (MIMO). Det finns många attraktiva egenskaper hos OFDM, nämligen robusthet mot de negativa effekterna av tidsspridning på grund av flervägsutbredning, enkelhet i utjämning och flexibilitet när det gäller att stödja både låga och höga symbolhastigheter. OFDM stöder därigenom en mängd olika servicekvalitetskrav.

De välkända nackdelarna med OFDM är höga störningsnivåer utanför huvudbandet (OOBE) och högt topp-till-medeleffektförhållande (PAPR). OOBE måste undertryckas eftersom hög OOBE orsakar betydande störningar på de intilliggande kanalerna. Vidare kräver hög PAPR typiskt dyra linjära radiofrekvenskomponenter (RF) och följaktligen kostsam digital fördistorsion för att hantera och mildra OOBE som genereras av distorsion från icke-linjära sändarkomponenter, t.ex. effektförstärkare. Dessutom finns det praktiska begränsningar i 5G-standarden, som kräver att endast databärande underbärvågor används för OOBE- och PAPR-reduktion. Därför är det av yttersta vikt att minska OOBE och PAPR för MIMO-OFDM-baserade system och mildra/minimera signalförvrängningen vid mottagaren/mottagarna för att möta den nya generationens systemkrav för olika användningsfall.

I denna avhandling söker vi principiella tillvägagångssätt för att tackla de distorsionsbaserade OOBE- och PAPR-reduceringsproblemen. Mer specifikt ställer vi upp optimeringsproblem för dessa välkända problem i storskaliga MIMO-OFDM-baserade system såsom 5G NR och framtida vidareutveckling därav. Tyvärr kan dessa problem inte lösas via generella optimeringslösare eftersom dessa standardlösare vanligtvis använder inrepunktsmetoder, som har alltför hög komplexitet för moderna radiohårdvarusystem. Därför föreslår vi storskaliga optimeringstekniker för att ta itu med dessa problem vilket resulterar i implementeringsvänliga algoritmer. Närmare bestämt utvecklar vi (nästan) optimala och beräkningseffektiva databeroende lösningar genom att föreslå en typ av multiplikatormetoder (ADMM) med tre operatorer som i huvudsak använder en söndra och härska metod för att lösa de stora och besvärliga OOBE- och PAPR-reduktionsproblemen i storskaliga MIMO-OFDM-baserade system. I den sista delen av avhandlingen undersöker vi dessutom tillämpningen av vår föreslagna ADMM-metod med tre operatorer för federerad inlärning över nätverk som drar nytta av den potentiellt innehållsrika datamängden som genereras i det fysiska lagret och/eller radiokomponenterna hos en basstation som ligger nära en randserver.

Sammanfattningsvis utvecklar denna avhandling grundläggande, implementeringsvän-

liga och standardagnostiska algoritmer för distorsionsbaserad reduktion av OOBE och PAPR, med hjälp av första ordningens optimeringsalgoritmer vilket ger insikter i avvägningen mellan beräkningskomplexitet och prestandan inomband och utomband. I avhandlingen presenterar vi en ny men ändå enkel TOP-ADMM första ordningens algoritm lämplig för att hantera centraliserade och distribuerade optimeringsproblem. Dessutom studeras hur TOP-ADMM-algoritmen kan användas för edge FL med hjälp av datamängder tillgängliga vid basstationen(erna) förutom (privata) data hos användarna. Slutligen kan denna avhandling ge vägledning vid systemisering och implementering av storskaliga MIMO-OFDM-baserade trådlösa kommunikationssystem.

To my family

Acknowledgments

Although it is time to culminate the industrial PhD journey, it is not an excuse to acknowledge the supervisors, managers, and everyone who, maybe involved knowingly or unknowingly, helped in the journey.

Most of all, I would like to express my utmost gratitude to my supervisor Prof. Carlo Fischione and my co-supervisors, Prof. Gabor Fodor, Prof. Mats Bengtsson, Prof. Bo Göransson, for their continuous support, encouragement, and guidance for the last four years and nearly six months. Thank you very much for improving me as a researcher. I will always cherish this journey with your blessings. I would also like to thank my other co-author, Dr. José Mairton B. da Silva Jr., for the fruitful discussions on my last crusade during this journey. Moreover, I would like to pay final respects to co-supervisor Prof. Peter Händel who prematurely passed away on September 15, 2019.

I am also sincerely thankful to the management team at Ericsson for the support in pursuing an industrial PhD, namely Ulf Norholm, Ingela Ericsson, Dr. Hannes Medelius, Preeti Nagarajan, Dr. Anders Caspár, Patrik Eriksson, Milan Basic, Pamela Gonzalez, and Mikael Svahn among many others.

I am genuinely grateful to Prof. Christoph Studer for acting as the opponent, Prof. Jaap van de Beek, Dr. Jakob Hoydis, and Prof. Sarah Kate Wilson for participating in the grading committee, and Prof. Joakim Jaldén for acting as an advanced reviewer for this thesis. Moreover, I want to thank Prof. Mathias Ekstedt for acting as a chairperson for the defence.

I want to thank many colleagues and friends at Ericsson. In particular, special thanks go to Dr. David Astely, Dr. Rohit Chandra, and Dr. Vidit Saxena. Also, I want to thank my other (ex)colleagues and friends Amit Singh, Goutam Ramamurthy, Harshendrakumar Nagaraj, Eric Nordström Pettigrew, Tomas Snitting, Goran Andrijevic, Dr. Hossein Shokri, Dr. Rong Du, Dr. Xiaolin Jiang, Dr. Mark Rollins, Pierre-André Laporte, Dr. Bilel Fehri, Dr. Mark Wyville, Mats Klingberg, and Dr. Ulf Gustavsson among many others at Ericsson and KTH. I thank many colleagues and friends who are not in the list as they have inspired me directly and/or indirectly.

Last but never least, I want to thank my mother, my parents in law, and all of my family members. Particularly, I want to thank Litta Marak (wife) and Shaun Kant (son) who are pillars of my life. I cannot thank my family enough for their love and affection. Without their support, I would not have made it this far.

Shashi Kant
Stockholm, October 2022

Contents

Contents xi

List of Papers xv

List of Acronyms xvii

I Thesis Overview 1

1 Introduction 3
1.1 Research Questions . . . 7
1.2 Research Challenges . . . 8
1.2.1 OOBE Research Challenges for 5G NR and Beyond . . . 8
1.2.2 PAPR Research Challenges for 5G NR and Beyond . . . 10
1.3 Literature Review . . . 11
1.3.1 OOBE Reduction Techniques . . . 11
1.3.2 PAPR Reduction Techniques . . . 14
1.3.3 First-Order Optimization For OOBE and PAPR Reduction . . . 18
1.3.4 Federated Learning . . . 20
1.4 Contributions of the Thesis . . . 20
1.4.1 EVM-Constrained & Mask-Compliant Spectral Precoding . . . 22
1.4.2 Efficient Optimization for LS-MIMO-OFDM Spectral Precoding 22
1.4.3 EVM Mitigation with PAPR and ACLR in LS-MIMO-OFDM Using TOP-ADMM . . . 23
1.4.4 Federated Learning Using Three-Operator ADMM . . . 23
1.4.5 Contributions not Covered in the Thesis . . . 23

References 25

2 Preliminaries 33
2.1 Some Notations and Definitions . . . 33
2.2 Background . . . 35
2.3 Projection Onto Convex Sets . . . 37

2.4 Two-Operator Splitting Methods 38
2.4.1 Two-Operator-Based Primal-Dual-Type Algorithms 39
2.4.2 Two-Operator-Based ADMM-Type Algorithms 42
2.5 Three-Operator Splitting Methods 44
2.5.1 Three-Operator-Based Primal-Dual-Type Algorithms 45
2.5.2 Three-Operator-Based ADMM-Type Algorithms 46
2.6 More Than Three Operators Splitting 48
2.6.1 Four-Operator-Based Function-Linearized Proximal ADMM . . 48
2.6.2 Multi-Block ADMM 49

References **50**

II Included Papers **55**

A EVM-Constrained and Mask-Compliant MIMO-OFDM Spectral Precoding **57**
A.1 Introduction 59
A.2 Related Works and Contributions 61
A.3 Preliminaries 64
A.4 EVM-Unconstrained LS-MSP in MIMO-OFDM 67
A.5 EVM-Constrained LS-MSP in MIMO-OFDM 71
A.6 Performance Evaluation 77
A.7 Conclusions 87
Ax.1 Convergence Analysis: ADMM and EADMM 88
Ax.2 Proof of (A.13) projection onto rank 1 ellipsoid 90
Ax.3 Derivation of SSP Algorithm 90
Ax.4 Useful Lemma 92

References **93**

B Efficient Optimization for Large-Scale MIMO-OFDM Spectral Precoding **99**
B.1 Introduction 101
B.2 System Model and Preliminaries 106
B.3 Proposed TxEVM-Mitigation and Mask-Compliant Spectral Precoding in Massive MIMO-OFDM 110
B.4 Three-Operator ADMM Algorithm 112
B.5 Performance Evaluation 117
B.6 Conclusions 130
Bx.1 Joint NSP with ZF Spatial Precoding (`NSP+ZF`) 130
Bx.2 Proof of Theorem 6 130
Bx.3 Background on Proximal, Projections, and Lipschitz 132

References **135**

C EVM Mitigation with PAPR and ACLR in LS-MIMO-OFDM Using TOP-ADMM **139**
C.1 Introduction . . . 141
C.2 Related Works and Contributions . . . 142
C.3 Preliminaries . . . 148
C.4 Proposed EVM Mitigation with PAPR and ACLR Constraints In MIMO-OFDM . . . 152
C.5 Simulation Results . . . 164
C.6 Conclusions . . . 178
Cx.1 Optimality Conditions of TOP-ADMM for P3/P4 . . . 179

References **180**

D Federated Learning Using Three-Operator ADMM **185**
D.1 Introduction . . . 187
D.2 State of the art . . . 190
D.3 Introduction to the TOP-ADMM Algorithm . . . 192
D.4 Federated Learning using TOP-ADMM . . . 194
D.5 Numerical Results . . . 200
D.6 Conclusions . . . 209
Dx.1 Some Useful Lemmas and Definitions . . . 209
Dx.2 Convergence Analysis of TOP-ADMM . . . 210
Dx.3 Proof of Theorem 14 . . . 221

References **225**

III Conclusions and Future Works **229**

1 Conclusions and Future Works **231**
1.1 Summary . . . 231
1.2 Future Works . . . 233

References **236**

List of Papers

The main body of this thesis comprises the following papers:

[J1] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "EVM-Constrained and Mask-Compliant MIMO-OFDM Spectral Precoding", *IEEE Transactions on Wireless Communications*, Vol. 20, No. 1, pp. 590–606, January 2021.

[J2] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "Efficient Optimization for Large-Scale MIMO-OFDM Spectral Precoding", *IEEE Transactions on Wireless Communications*, Vol. 20, No. 9, pp. 5496–5513, September 2021.

[J3] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "EVM Mitigation with PAPR and ACLR Constraints in Large-Scale MIMO-OFDM Using TOP-ADMM", accepted to *IEEE Transactions on Wireless Communications*, 2022.

[J4] **Shashi Kant**, José Mairton B. da Silva Jr., Gabor Fodor, Bo Göransson, Mats Bengtsson, Carlo Fischione, "Federated Learning using Three-Operator ADMM", submitted to *IEEE Journal of Selected Topics in Signal Processing*, 2022.

The other papers which are related to the main core of this thesis but not explicitly included in this thesis are:

[C1] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "Robust PAPR reduction in large-scale MIMO-OFDM using three-operator ADMM-type techniques", in 2021 55th Asilomar Conference on Signals, Systems, and Computers, 2021, pp. 616–622.

[C2] **Shashi Kant**, Gabor Fodor, Mats Bengtsson, Bo Göransson, Carlo Fischione, "Low-Complexity OFDM Spectral Precoding", in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Cannes, France, 2019.

List of Acronyms

1G	First Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
6G	Sixth Generation
ACLR	Adjacent Channel Leakage Ratio
ADMM	Alternating Direction Method of Multipliers
AI	Artificial Intelligence
AIC	Active Interference Cancellation
BADMM	Bregman Alternating Direction Method of Multipliers (ADMM)
BLER	BLock Error Rate
BS	Base Station
CCDF	Complementary Cumulative Distribution Function
cMTC	Critical Machine Type Communications
CPS	Cyber-Physical Systems
CSI	Channel State Information
DFT	Discrete Fourier Transform
DMRS	Demodulation Reference Signal
DRS	Douglas-Rachford Splitting
DYS	Davis-Yin Splitting
eMBB	enhanced Mobile BroadBand
EADMM	Error Vector Magnitude (EVM)-constrained ADMM
EVM	Error Vector Magnitude
EMSP	EVM-constrained Mask-compliant Spectral Precoder
ENSP	EVM-constrained Notching Spectral Precoder
ESSP	EVM-constrained Semi-analytical Spectral Precoding

FL	Federated Learning
FLiP-ADMM	Function-Linearized Proximal ADMM
ICF	Iterative Clipping and Filtering
IDFT	Inverse Discrete Fourier Transform
i.i.d.	independent and identically distributed
IIoT	Industrial Internet-of-Things
IPAPR	Instantaneous Peak-to-Average Power Ratio
IoT	Internet-of-Things
JCS	Joint Communications and Sensing
KKT	Karush-Kuhn-Tucker
LOS	Line-Of-Sight
LS	Large-Scale
LS-MSP	Large-Scale Mask-compliant Spectral Precoder
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-output
ML	Machine Learning
MSP	Mask-compliant Spectral Precoder
MU-MIMO	Multi-User Multiple-Input Multiple-output
mMTC	massive Machine Type Communications
NB-IoT	NarrowBand Internet-of-Things
NR	New Radio
NSP	Notching Spectral Precoder
OFDM	Orthogonal Frequency Division Multiplexing
OOB	Out-Of-Band
OOBE	Out-Of-Band Emissions
PA	Power Amplifier
PAPR	Peak-to-Average Power Ratio
PDSCH	Physical Downlink Shared Channel
POCS	Projection Onto Convex Sets
PRB	Physical Resource Block
PRG	Physical Resource Block Group
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation

RF	Radio Frequency
Rx	Receive(r)
RZF	Regularized Zero Forcing
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SSP	Semi-analytical Spectral Precoding
SU-MIMO	Single-User Multiple-Input Multiple-output
TDD	Time Division Duplexing
TDL	Tapped Delay Line
TOP-ADMM	Three-Operator ADMM
TSN	Time Sensitive Networking
Tx	Transmit(ter)
TxEVM	Transmit-EVM
UE	User Equipment
URLLC	Ultra-Reliable and Low-Latency Communications
V2X	Vehicle-to-everything
XR	eXtended Reality
ZF	Zero Forcing

Part I

Thesis Overview

Chapter 1

Introduction

The development and evolution of wireless communication networks began in the 1960s [TV05]. The past decades have witnessed a tremendous surge in the mobile communication industry with the underlying technologies from first generation (1G) to existing fifth generation (5G) networks. 5G New Radio (NR) has now been commercially and successfully deployed around the world by many network operators and used in millions of smartphones supporting the 5G-based 3rd generation partnership project (3GPP) standard, among other standards. More interestingly, since the incarnation of the fourth generation (4G) long-term evolution (LTE) cellular systems, the requirements for these next-generation wireless communication networks, notably 5G NR and beyond, are driven by at least three broad use cases [ITU17, DPS18, LSJ+21]. These use-cases—see Figure 1.1—include 1) enhanced mobile broadband (eMBB) services to support extremely high data rates in terms of network or per user in both uplink and downlink, 2) massive machine type communications (mMTC) to cater to massive internet-of-things applications, and 3) critical machine type communications (cMTC) to address mission-critical applications that require ultra-high reliability and low latency.

With the success of 5G NR catering to many use cases, naturally, 5G is entering into a new phase of evolution from release 18 of 3GPP, which is addressed as 5G-advanced [CML+22] toward the sixth generation (6G) that will support additional applications, *e.g.*, using artificial intelligence (AI)/machine learning (ML) for channel state information (CSI) reporting, see, *e.g.*, [RP-21]. 6G related activities in 3GPP are expected to begin from release 19—see the timeline in Figure 1.2 from Ericsson's perspective [RLR+22]. Moreover, 6G requirements in 3GPP are expected to begin from release 19, which can be envisioned to support AI/ML over networks such as federated learning (FL). Therefore, in essence, besides eMBB, mMTC, and cMTC use-cases, the additional use-cases envisioned for beyond 5G or 6G are, *e.g.*, 1) interplay of AI/ML with wireless communications, 2) joint communications and sensing (JCS), 3) cyber-physical systems (CPS), and 4) extended reality (XR)—cf. Figure 1.1.

Modern cellular/wireless communication systems, such as 5G NR and possibly 6G, adopt orthogonal frequency division multiplexing (OFDM) with cyclic prefix [Pra04]—

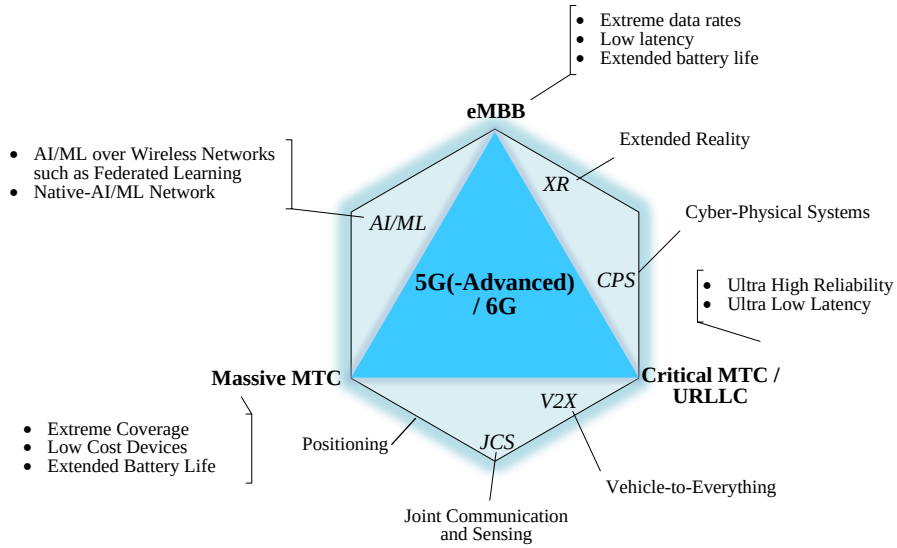


Figure 1.1: *5G, 5G-advanced, and possibly 6G application categories [RLR+22].*

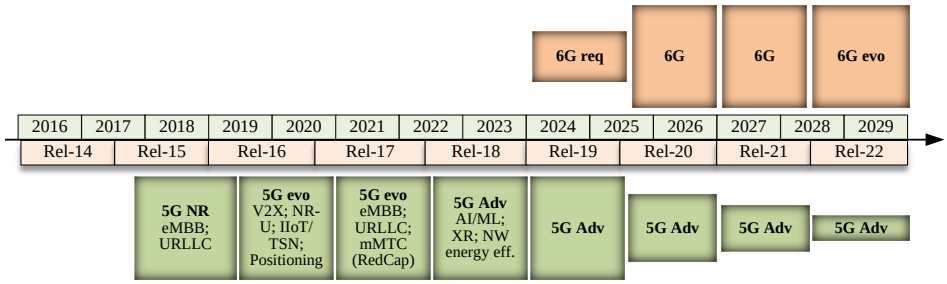


Figure 1.2: *3GPP timeline for 5G-advanced and 6G from Ericsson's perspective [RLR+22]. The size of the boxes of 5G and 6G for each release indicates the (possible) amount of study/work items.*

see Figure 1.3. In OFDM, multiple subcarriers are orthogonally and closely placed in the frequency domain, where each subcarrier carries a (typically complex-valued) data symbol. Subsequently, an inverse discrete Fourier transform is applied to these discrete-frequency domain block of data symbols rendering a discrete-time domain OFDM symbol block. Thereafter, a cyclic prefix is cascaded to the discrete-time domain OFDM symbol. OFDM with cyclic prefix has many appealing characteristics, such as robustness against the negative effects of time dispersive channels and multipath fading, simplicity in terms of equalization, and flexibility in terms of supporting both low and high symbol rates and thereby supporting a variety of various quality of service requirements. Moreover, these

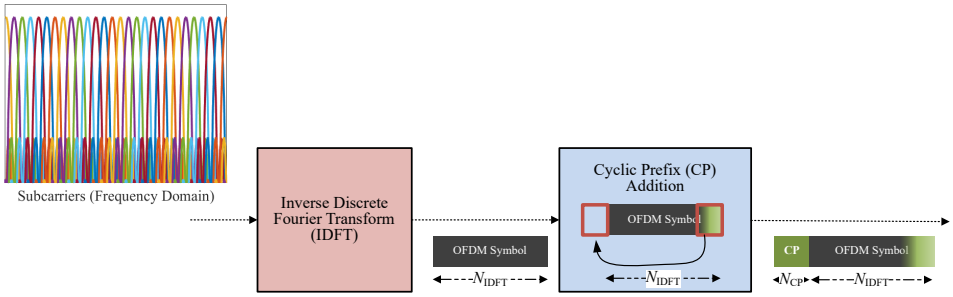


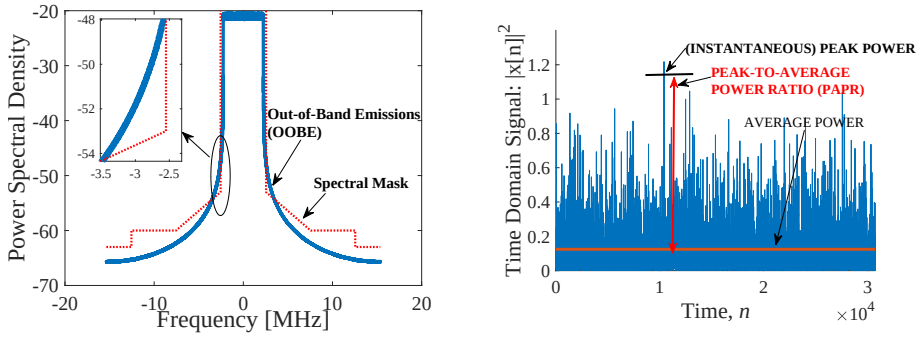
Figure 1.3: *Simplified illustration of OFDM modulation with cyclic prefix addition.*

latest wireless communication systems employ multiple antennas at both the ends of the communication link, *i.e.*, transmitter and receiver—referred to as multiple-input multiple-output (MIMO)—to improve spectral efficiency and/or link reliability over OFDM-based systems with single antennas at both the transmitter and the receiver, *i.e.*, single-input single-output (SISO).

Unfortunately, in practical wireless deployments, the hardware imperfections have a deleterious impact on the system-wide throughput performance of high-data rate achieving MIMO-OFDM-based systems [Sch08, GGLF08, SWB10, SWB11, BHS17, BHKD14, BA19], where these hardware imperfections emanate from both the digital and the analog components. These hardware imperfections comprise so-called clip noise, filter ripple or distortion, in-phase and quadrature (IQ) mismatch, nonlinearity of the hardware components, phase noise from a local oscillator, sampling clock offsets, timing, and frequency errors, see, *e.g.*, [FLP+05, Sch08]. However, some of these hardware imperfections are fundamentally rooted and/or related to the OFDM waveform. Concretely, OFDM suffers from high peak-to-average power ratio (PAPR) and out-of-band emissions (OOBE), see Figure 1.4, which have been well-known issues for many decades. High PAPR arises due to the multicarrier nature of the OFDM. Specifically, high PAPR is a consequence of the linear combination of the subcarriers carrying (pseudo-random) M-ary quadrature amplitude modulation (M-QAM), symbols in the inverse discrete (fast) Fourier transform (IDFT) operation. The reason for OOBE is due to the discontinuities at the boundaries of the implicit rectangular window and the high sidelobes associated with the sinc functions of the OFDM symbol.

Low PAPR allows the power amplifier (PA) to operate at the so-called bias-point to maximize the efficiency of PA operation and consequently minimizes the (possible) nonlinear signal distortion stemming from the PA-like nonlinear devices in the transmit radio frequency (RF) chain. Low OOBE is vital to meet 3GPP adjacent channel leakage ratio (ACLR) requirements and unwanted emissions mask stipulated by the standard (for the considered operating frequency bands). Hence, it is of utmost importance to minimize both the PAPR and OOBE of cyclic prefix-based OFDM waveform.

In the literature, there are a plethora of techniques to tackle these well-known OOBE and PAPR issues in OFDM-based systems. However, there are many practical



(a) *OOBE issue* **(b)** *PAPR issue*

Figure 1.4: *At least two well-known disadvantages of OFDM signals: OOBE and PAPR.*

research challenges in employing or selecting OOBE and PAPR reduction methods for realistic base station radios. More specifically, these reduction techniques are vendor specific (and typically not standardized). Consequently, to be practically useful, these reduction techniques should not only be computationally efficient but also compliant with regulatory requirements and standards such as 3GPP 5G NR (with a possibility of forward compatibility if the system will be MIMO-OFDM based). Additionally, it is highly desirable that these techniques should be backward compatible with previous generation standards if the same radios support multiple standards. We will discuss more of these research challenges in the subsequent section.

Many OOBE reduction methods, see, *e.g.*, [HZG15] and its references, are proposed in the literature, which can be classified into two main categories, namely time domain and frequency domain techniques. In the sequel, we will briefly survey the literature for OOBE reduction techniques. To break the ice, (non-orthogonal) spectral precoding—a signal distortion-based and frequency domain technique—is one of the promising techniques in meeting the research challenges because it is transparent to the standard. Among other motivational reasons, spectral precoding uses only the active[1] useful subcarriers without penalizing the cyclic prefix. It essentially precodes the data symbols over the subcarriers (or in the frequency domain) before the OFDM modulation [vdB09, MNJ18, Moh19]. Consequently, this spectral precoding reduces the OOBE without sacrificing the cyclic prefix of the transmitted signal. It is worth pointing out that these non-orthogonal spectral precoding techniques do not impact the receiver signal processing. However, it distorts the quality of the transmitted signal—typically quantified in terms of error vector magnitude (EVM).

Similar to OOBE reduction methods, there are a myriad number of PAPR reduction techniques that exist in the literature, see, *e.g.*, overview papers [HL05, RM13, WFB+13]. In the sequel, we will overview some PAPR reduction methods. To take-the-plunge, for our study, the distortion-based approach is more interesting, such as classical iterative clipping and filtering [Arm02, WL11]. These distortion-based techniques are widely employed in

[1] One could also exploit inactive subcarriers as long as spectral precoding meet the desired constraints.

the industry because they are transparent to the standards and the receiver, see [RM13]. However, these distortion-based approaches also render signal distortion, measured in terms of EVM as well.

In this chapter, we give an overview of some known methods for the PAPR and OOBE reductions, which are typically treated disjointly with each other even though these two problems are tightly coupled with each other, especially for distortion-based techniques. Before we review some existing studies on OOBE and PAPR reduction methods, we first discuss some practical challenges imposed by the standards such as 5G NR and possibly beyond 5G to motivate the employment of distortion-based OOBE and PAPR techniques. Furthermore, most of the principled approaches for distortion-based OOBE and PAPR reductions can naturally be posed as optimization problems. Fortunately, the optimization problem tackling mask-compliant spectral precoding, for OOBE reduction, is convex. However, the optimal solution to such an OOBE reduction problem with spectral precoding cannot be computed in closed form. Consequently, we seek a computationally efficient and implementation-friendly iterative technique to tackle the mask-compliant spectral precoding optimization problem for both SISO- and large-scale (LS) MIMO-OFDM-based systems. Contrary to the mask-compliant spectral precoding problem, the distortion-based PAPR reduction problem is unfortunately not convex. Therefore, similar to solving mask-compliant spectral precoding problems, we investigate a computationally friendly dedicated first-order optimization[2] solver for the posed PAPR reduction problem. Observe that one could envisage the usage of a generic optimization solver, such as `CVX` [GB14] wrapper with supported solvers, *e.g.*, SDPT3, in tackling the posed optimization problems for both OOBE and (at least convexified[3]) PAPR, but such generic solvers have prohibitive computational complexity notably for the large-scale problem. Consequently, we describe some powerful and efficient first-order optimization methods to tackle OOBE and PAPR reduction problems, notably for LS-MIMO-OFDM-based systems.

1.1 Research Questions

In this thesis, we are interested in the design of computationally efficient and implementation friendly algorithms for solving optimization problems for OOBE and PAPR reduction. Specifically, our work seeks to answer the following research questions:

RQ1. How can we optimally design “hardware-friendly" algorithms for mask-compliant spectral precoding to reduce OOBE in LS-MIMO-OFDM while constraining either wideband average or frequency-selective transmit-EVM?

RQ2. Can we exploit the excess spatial dimension inherent in the LS-MIMO system to mitigate the transmit-EVM—incurred due to mask-compliant spectral precoding—seen at the receiver(s) while meeting the desired spectral emission mask at the

[2]In Section 1.3.3, we briefly motivate the first-order optimization methods over the second-order methods for tackling the OOBE and PAPR reduction problems in the LS-MIMO-OFDM-based systems. Moreover, in Chapter 2, we provide an overview of some interesting first-order optimization methods.

[3]More information about possible convexification will be expounded in the subsequent chapters; more specifically, see Chapter C.4.2.

transmitter? Additionally, if the answer is yes to the initial question, how can we design an efficient algorithm for such a formidable task at hand?

RQ3. How can we develop efficient first-order algorithms to solve the distortion-based nonconvex PAPR reduction problem near optimally and simultaneously mitigate the transmit-EVM at the receiver(s) with robustness to wireless channel estimation uncertainty?

RQ4. Can we use the distributed first-order method not only for the abovementioned tasks but also for FL purposes with theoretically guaranteed convergence, at least under the convex settings?

In the remainder of this chapter, we firstly discuss the OOBE and PAPR reduction challenges imposed by the stringent 3GPP NR-like system performance requirements. Next, we review some existing studies on the OOBE and PAPR reduction, where we also motivate the employment of distortion-based OOBE and PAPR reduction schemes for MIMO-OFDM-based 5G NR and possibly beyond. Many of these distortion-based OOBE and PAPR reduction schemes can be explained via mathematical optimization such that the resulting solution can be (near) optimal according to the problem design. In the sequel, we provide an overview of first-order optimization methods that can possibly render computationally efficient methods, which can also be implementation friendly. Finally, we highlight our original contributions in answering some of the abovementioned research questions.

1.2 Research Challenges

In this section, we highlight some research challenges to reduce OOBE and PAPR, particularly, due to the requirements set by the existing and possibly upcoming releases of the standards, such as 5G NR. Moreover, the new releases of these standards may also support new features, posing additional potential challenges for the OOBE and PAPR reduction.

1.2.1 OOBE Research Challenges for 5G NR and Beyond

With the evolution of new generations of wireless communications, the same base station radio can support multiple radio access technologies, *e.g.*, 5G NR, long term evolution (LTE), narrowband internet-of-things (NB-IoT) and/or multiple carriers of the same/different technologies, by sharing the same band—see Figure 1.5. Therefore, all carriers (of respective technologies) should satisfy appropriate OOBE requirements in terms of ACLR [3GP21] such that all these carriers can coexist in the same band. Observe that 5G NR has $\sim$98% [Sam17, 3GP21] bandwidth efficiency compared to LTE, which was designed with relatively poor bandwidth efficiency ($\sim$90%). The reason for increasing the bandwidth efficiency is that the spectral resources are costly for network operators, among other capital and operating expenses. Interestingly, from LTE release 13 onward, NB-IoT carriers can also be placed within the LTE's guard bands—see Figure 1.6. Hence, one

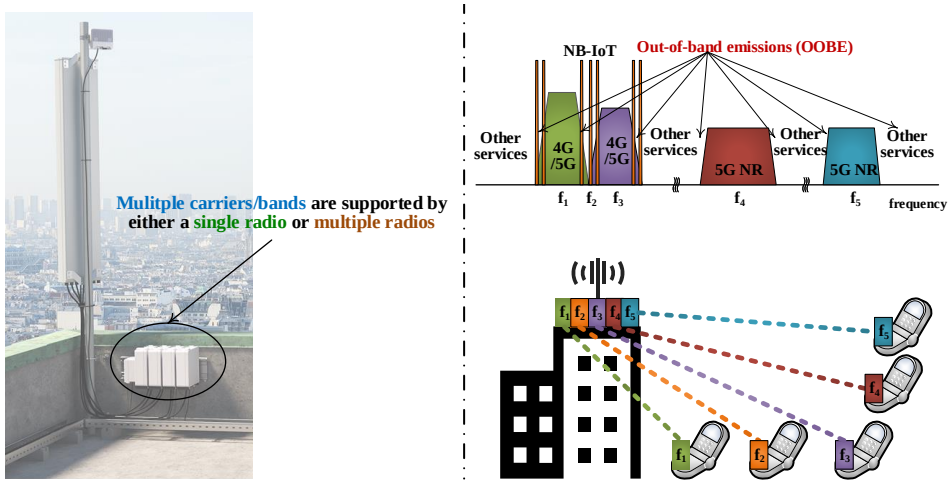
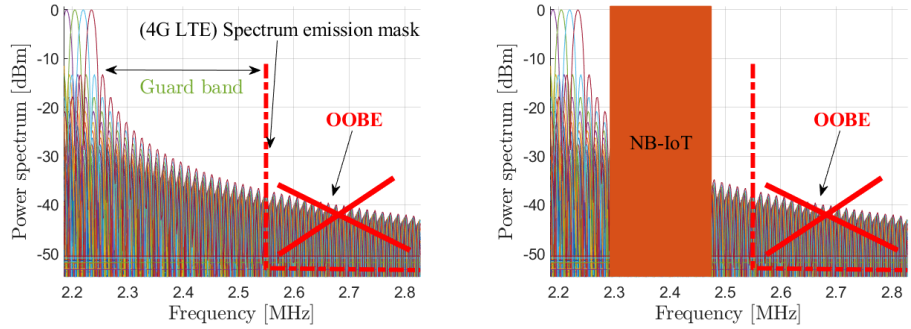


Figure 1.5: *On the left-hand side, we exemplify a "typical" radio head or antenna site with multiple radios supporting multiple carriers and/or bands [Eri]. On the right-hand side—top figure—we illustrate a band where multiple access technologies must coexist while meeting appropriate regulatory requirements in terms of OOBE among other standard-specific requirements. On the right hand side—bottom figure—we show multiple carriers at the same site (of the same/different technologies) catering to their users.*

could argue that the effective guard bands of LTE are also minimized, and consequently, the bandwidth efficiency of overall LTE has increased, compared to the LTE release 8–13. Although the 3GPP RAN4 specifies lower bandwidth/spectrum efficiency of NR carriers than the 3GPP RAN1 specifications [3GP17, 3GP20], it is expected that mobile network operators will want to increase further bandwidth efficiency (and consequently improve the overall throughput of the system). Thus, over time, new-generation cellular wireless communications will have higher (effective) bandwidth efficiency with almost no guard bands. Furthermore, the cyclic prefix must be spared to avoid intersymbol/intercarrier interference, particularly if the effective channel delay spread is longer than the cyclic prefix length. Additionally, network operators desire a large cell range meeting suitable key performance indicators. Therefore, the serving cell range may be reduced if the cyclic prefix is penalized (or effectively shortened). Hence, our OOBE research challenge is to meet the desired bandwidth efficiency (without using guard bands) and the minimum OOBE and in-band performance requirements stipulated by the standards, such as 5G NR or LTE, without sacrificing the cyclic prefix. Furthermore, it is worth mentioning that the OOBE reduction technique should be transparent to standard specifications, such as 5G NR.



(a) *Without utilization of guard band* **(b)** *NB-IoT placed in the guard band*

Figure 1.6: *Illustrates the right edge of unfiltered LTE (5 MHz bandwidth) without and with utilization of guard band, and the spectrum emission mask where OOBE must be below the limits.*

1.2.2 PAPR Research Challenges for 5G NR and Beyond

To meet the perpetual demands on high-data rates, lower end-to-end latencies, and connectivity among various (emerging and new) applications of eMBB, massive and critical Internet-of-Things (IoT) including AI/ML, *e.g.*, such as XR, CPS, and V2X—as illustrated in Figure 1.1, 5G and beyond, require some key enablers, comprising such as LS-MIMO technologies, support of wider carrier bandwidths, and the usage of higher frequency bands. Consequently, the implementation of green networks and radio base stations faces many challenges, notably in terms of power consumption and size—primarily dominated by PAs. Thus, the PAPR of the transmit signals must significantly be reduced for each transmit antenna to reduce the power consumption of radio base stations or PAs while simultaneously supporting high-data rate transmissions. Additionally, the same radio base station may support multiple radio access technologies, *e.g.*, such as 5G NR, 4G LTE, and NB-IoT—as illustrated in Figure 1.6. Therefore, the support of multiple radio access technologies, large carrier bandwidths, multiple concurrent bands, and many physical antennas pose enormous challenges in reducing PAPR for each antenna branch. The high average output power of the transmit signals is necessary for better cell coverage in some classes of base stations. If the required average transmit signal power and PAPR are high, efficient, and smaller size PA can render nonlinear signal distortions causing both OOBE and in-band distortions. Consequently, the base station with a high PAPR may fail to comply with the 3GPP standard and regulatory requirements unless inefficient energy measures are considered, such as a significant amount of power back off. In other words, to meet regulatory and standard performance requirements, a power hungry—inefficient—large size PA with power back-off should be employed to accommodate the high PAPR of transmit signals such that the PA can operate in a linear region. Hence, the PAPR of OFDM waveform and, more importantly, for LS-MIMO systems must be reduced to contribute toward greener and better energy-efficient radio base stations. It is worth pointing out that

a PA linearizer such as a digital predistorter is also an important constituent of a base station for improving the efficiency of the PA by complementing the contribution of reduced PAPR of transmit signals[4]. The reduced PAPR with the PA linearizer can allow the PA to operate near its so-called bias point.

1.3 Literature Review

This section briefly reviews some existing methods for reducing OOBE and PAPR. Subsequently, we provide a very short introduction to first-order optimization methods, notably proximal/operator splitting techniques. However, in Chapter 2, we introduce more background on optimization theory, particularly for proximal splitting methods. Some of these proximal splitting techniques can be used to solve the optimization problems posed to tackle OOBE and PAPR reduction problems.

1.3.1 OOBE Reduction Techniques

In this section, we briefly survey some known spectrum shaping or OOBE reduction techniques for the cyclic-prefix based OFDM (assuming rectangular window unless otherwise stated). There are numerous techniques to suppress/reduce OOBE—see, *e.g.*, [HZG15] and its references, and these techniques can be grouped into time and frequency domain methods, as illustrated in Figure 1.7 with some techniques that will briefly be discussed in the sequel.

Before we proceed further, we would like to highlight that the OOBE performance requirements are typically quantified in terms of mask and ACLR, *e.g.*, defined in NR-like standards [3GP21]. We present these requirements in more detail in the subsequent chapters—see Section A.3.2 or Section B.2.4.

1.3.1.1 Guard Band Insertion

This is one of the most straightforward approaches that allow some headroom in the frequency domain because the OOBE power of subcarriers decays on the order of $1/f^2$ [Cha66, Chu08], where f is the distance between the mainlobe of a considered subcarrier and observing frequency. However, 3GPP-like standards stipulate the minimum allowed guard band—depending on the operating band and carrier bandwidth, among other criteria—for coexistence with the neighbouring (typically non-orthogonal) carriers, see Figure 1.5 on the top right-hand side. Consequently, one can deactivate more useful subcarriers from the edges if they do not violate standards' requirements but penalize the achievable throughput (or spectral efficiency). Nevertheless, from a 3GPP-like standard perspective, the guard band alone is insufficient to meet OOBE performance and regulatory requirements. Therefore, other suitable OOBE reduction techniques are necessary on top of standard-given guard band inclusion with a minimal penalty to the spectral efficiency.

[4]In this thesis, we will focus on computationally efficient methods to reduce PAPR in LS-MIMO-OFDM-based systems while meeting some desired objectives. We will discuss this more in the subsequent sections and Chapter C.

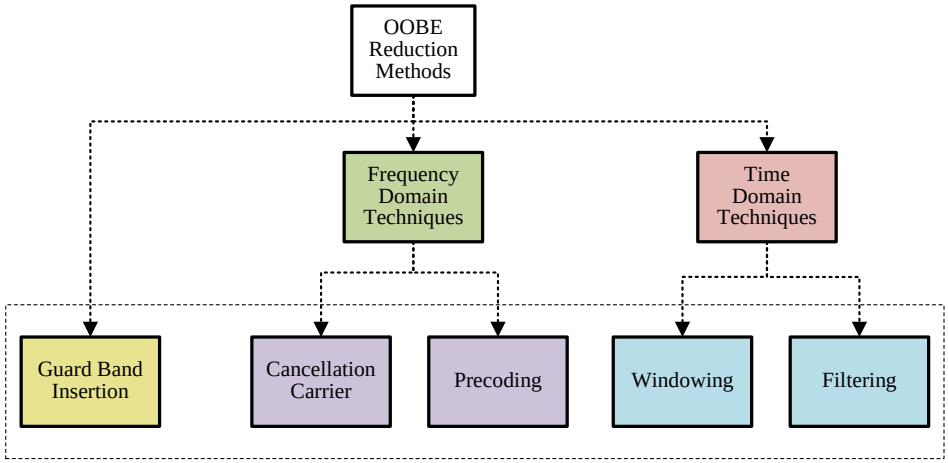


Figure 1.7: *Taxonomy of OOBE reduction techniques.*

1.3.1.2 Time Domain Techniques: Filtering and Windowing

Time domain OOBE reduction methods—filtering [Fau00] or windowing [BLY13]—are well-known and widely employed in the industry mainly because it is transparent to the standards, have low computational complexity, and do not necessarily impose any signal processing requirements at the receiver side. The windowing/filtering technique essentially smoothens the discontinuity between adjacent time-domain OFDM symbols (implicitly having a rectangular window). In windowing, a candidate window, *e.g.*, a simple triangular window, is element-wise multiplied in the discrete-time domain typically with the part of the cyclic prefix samples to taper the OFDM symbol transitions, which is equivalent to circular convolution in the discrete frequency domain, see, *e.g.*, [PM07, YKLP+20]. Consequently, filtering can be seen as a dual of windowing because of the circular convolution of an impulse response of the filter with the OFDM signal samples in the discrete-time domain implies element-wise multiplication with a frequency response of a filter in the frequency domain. One of the significant disadvantages of these time domain techniques is a penalty to cyclic prefixes—as discussed in the research challenges of OOBE reduction methods. Additionally, windowing/filtering cannot achieve strict/desired OOBE performance requirements when OFDM-based systems must support a non-contiguous spectrum for coexistence with other systems [HZG15, ZHCG12], *e.g.*, for dynamic spectrum sharing between two non-orthogonal carriers sharing the same spectrum.

1.3.1.3 Frequency Domain Technique: Cancellation Carrier

The frequency-domain cancellation carrier techniques, such as active interference cancellation (AIC) [Yam04], AIC with power constraint [BCS05], and extended AIC [QWJ10], exploit the inactive subcarriers/tones at the edges of a carrier as shown in Figure 1.8.

Although these methods are signal distortion-less and do not penalize the cyclic prefix of an OFDM signal, the useful signal transmit power in practice needs to be shared with these non-data-bearing subcarriers, which essentially reduces the transmit power of the useful signal and thereby penalizes the received signal-to-noise ratio (SNR) and spectral efficiency. Specifically, unconstrained AIC [Yam04] requires a significant amount of energy in the cancellation carriers to meet strict OOBE requirements, which consequently penalizes the useful transmit signal energy. Therefore, power-constrained AIC has been proposed in [BCS05] such that the “peaky" power spectral density (PSD) on the non-useful data-carrying cancellation carriers can be ameliorated; thus, constrained AIC can have almost similar PSD as the useful data-carrying subcarriers. However, to achieve strict OOBE performance requirements of 5G NR, more AIC carrying (sub)carriers would be required and leaving almost no room for placement of NB-IoT-like carriers in the guard band besides penalty to the useful transmit signal power. In addition to received SNR degradation, in particular, the extended AIC [QWJ10] causes intersymbol and intercarrier interferences to the desired/useful signal due to the placement of the non-orthogonal carriers. Hence, these abovementioned AIC schemes fail to meet many of our practical objectives discussed in Section 1.2.1.

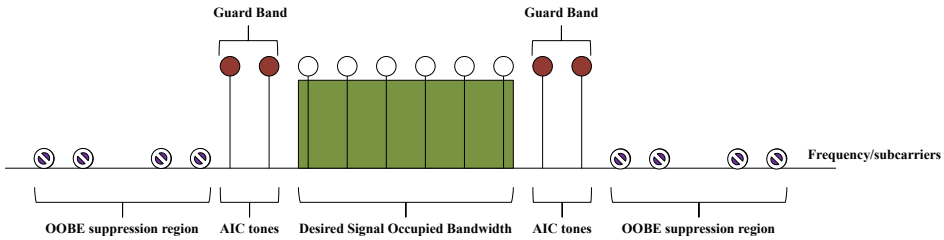


Figure 1.8: *Simplified pictograph of frequency-domain (active) cancellation carriers-based OOBE reduction technique.*

1.3.1.4 Frequency Domain Technique: Spectral Precoding

Spectral precoding is one of the promising techniques for its superior OOBE reduction performance in cyclic-prefix-based OFDM as it can comply with OOBE performance requirements in terms of ACLR and mask. This spectral precoder performs the precoding of data symbols over subcarriers before OFDM modulation [vdB09, MNJ18, Moh19], which essentially generalizes the abovementioned AIC scheme because it can operate on data subcarriers (and one can easily extend it to utilize inactive subcarriers). This precoding can be categorized into two groups, namely non-orthogonal spectral precoding [vdB09, TSA13, KT16, KFB$^+$19, KBF$^+$21] and orthogonal [J. 10, MHJ$^+$11, Chu06, KHLV21]. Non-orthogonal spectral precoding, has no impact on the receiver signal processing, whereas the latter, *i.e.*, orthogonal spectral precoding, requires some form of decoding at the receiver in order to undo the effect of transmit spectral precoding. Consequently, and

unfortunately, the orthogonal spectral precoding[5] does not meet our research objectives for the OOBE reduction mainly because it impacts the receiver signal processing. Thus, non-orthogonal spectral precoding is propitious, even though it causes signal distortions. It can meet our practical research objectives discussed in Section 1.2.1 such as no sacrifice to the cyclic prefix and no penalty to the transmit power, among others. We will discuss only non-orthogonal spectral precoding for OOBE reduction from now onward unless otherwise stated.

One of the popular data-independent schemes is notching spectral precoder (NSP) [vdB09] as it nulls the OOBE at the chosen discrete frequency points. This NSP can be posed as an equality constrained least-squares optimization problem that minimizes the Euclidean distance between the desired precoded symbol vector and the original data symbol vector with an equality constraint that zero-out the OOBE energy at the considered discrete frequency points in the OOBE region. However, NSP causes high (in-band) signal distortion—quantified in terms of EVM at the transmitter—for relatively low carrier bandwidth. Consequently, this high EVM at the transmitter penalizes the received SNR, and thereby, it has a detrimental impact on the system-wide throughput performance. Recently, mask-compliant spectral precoder (MSP) technique was proposed and investigated to generalize NSP, which is posed as a constrained least-squares problem that has the same objective as NSP but with an inequality constraint to meet a predefined mask rather than notching at the given discrete frequency points [TSA13, KT16, KFB$^+$19, KBF$^+$21]. This MSP offers a good trade-off[6] between EVM and out-of-band (OOB) suppression performance at the expense of increased computational complexity because it becomes data dependent and does not yield a closed-form solution. Consequently, a generic optimization solver, such as CVX [GB14] is suggested in [TSA13]. However, these generic solvers are difficult to deploy on application-specific integrated circuits and typically employ interior-point methods [BV04]. Therefore, [KT16] has proposed a suboptimal approach. In contrast to these known techniques, we seek not only implementation-friendly and memory-efficient methods but also (near) optimal methods for the large-scale system that meets the desired OOBE and in-band performance requirements—see our research questions 1 and 2 in Section 1.1.

1.3.2 PAPR Reduction Techniques

In this section, we briefly survey some known PAPR reduction techniques for the cyclic-prefix-based OFDM. There are plentitude techniques to suppress/reduce PAPR—see the taxonomy of some of the techniques in Figure 1.9, which will briefly be discussed next.

At this point, we would like to highlight that the PAPR reduction performance requirements are typically quantified in terms of PAPR metrics, in-band, *i.e.*, EVM, and OOBE requirements. We present these metrics and requirements in more detail

[5] One can ignore the decoding process at the receiver such that the orthogonal spectral precoding can become non-orthogonal.

[6] One can combine the strengths of these frequency domain spectral precoding (and/or AIC-like technique) with other time domain OOBE reduction methods, such as pulse shaping or windowing for better flexibility—see, *e.g.*, [DCCn$^+$19, HLV22]. However, in this thesis, we focus on the non-orthogonal MSP scheme unless otherwise stated.

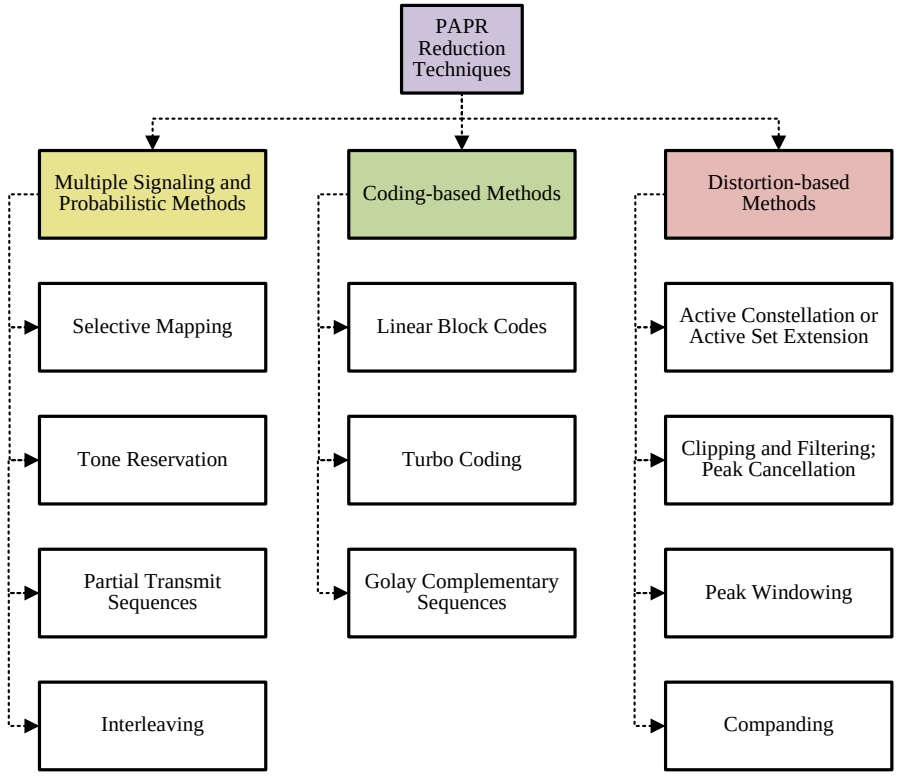


Figure 1.9: *Taxonomy of PAPR reduction techniques.*

in Chapter C—see Section C.3.2. Moreover, since many PAPR reduction schemes are essentially performed in the discrete-time domain, the discrete-time domain OFDM signal must be oversampled so as not to miss high peaks within the given discrete-time samples as this discrete-time domain signal is converted to an analog signal via a digital-to-analog converter. Naturally, the analog signal is at the end input to the PA. Therefore, the chosen PAPR scheme should ensure that the analog input signal to the PA has the desired PAPR.

1.3.2.1 Multiple Signaling and Probabilistic PAPR Reduction

Selective mapping [MS96, HL05, RM13, WFB$^+$13] generates a set of multiple copies of input data sequences, where each sequence is generated with different phases but represents the same useful information. Consequently, multiple statistically independent OFDM signals are generated, where the signal with the minimum PAPR is chosen for the transmission. In other words, PAPR reduction is performed by generating multiple signals carrying the same useful information and selecting the signal with minimum PAPR from the set. However, to successfully demodulate the received signal, this selective mapping impacts the receiver signal processing because the receiver must know the selected mapping or the set of phase sequences employed at the transmitter. However, some works

have investigated/proposed blind selective mapping [BMWH01, LGASK+09, JHJ+12], *i.e.*, the receiver does not need to know the side information explicitly; the extra signal processing at the receiver is required to decode the transmitted signal blindly.

The interleaver-based PAPR reduction technique, see, *e.g.*, overview paper [RM13] and references therein, is similar to selective mapping, where multiple OFDM signals are generated by various choices of interleavers either at the bit or symbol level. Subsequently, a suitable OFDM signal is selected, with the lowest PAPR among the generated signals. Similarly, the receiver must know the set of interleavers for de-interleaving the demodulated received signal.

Closely related to selective mapping, in partial transmit sequences, see, *e.g.*, overview paper [RM13] and references therein, the (frequency domain) symbol vector is partitioned into several non-overlapping sub-blocks. These sub-blocks are transformed into the time domain via IDFT. Subsequently, each time domain signal vector is elementwise multiplied with a suitable phase factor, and then all the weighted time domain signals of each sub-block are combined to form an OFDM signal with reduced PAPR. Hence, a suitable set of phase factors need to be found via optimization to reduce the PAPR of the OFDM signal formed by combining multiple weighted time domain signals of sub-blocks.

Similar to the AIC concept for OOBE reduction, briefly described in Section 1.3.1.3, tone reservation [TC98, WXX18, HL05, RM13, WFB+13] also utilizes non-useful data carrying unused/inactive subcarriers for PAPR reduction. Typically, these subcarriers are on the edges of a data-carrying carrier. Consequently, the tone reservation does not cause any in-band signal distortions. However, similar to shortcomings of the AIC, a large number of subcarriers are required to reduce the PAPR of the signal. Furthermore, such approaches penalize bandwidth efficiency.

In summary, these aforementioned multiple signaling, including tone reservation techniques, are unfortunately not a winning candidates for 5G NR, and possibly beyond, considering the objectives and practical challenges discussed in Section 1.2.2.

1.3.2.2 Coding-Based PAPR Reduction

Coding-based PAPR reduction methods, *e.g.*, linear block codes, Golay complementary sequences, and turbo coding—see [HL05, RM13, WFB+13] and references therein—generate encoded data bits such that the resulting OFDM symbol has low PAPR. However, these methods become intractable for relatively large-scale OFDM-based systems with more than a dozen subcarriers due to an exhaustive search for a good code. Clearly, these techniques are practically not viable for a 5G NR-like system.

1.3.2.3 Distortion-Based PAPR Reduction

In active constellation or active set extension [KJ03] scheme, to reduce PAPR, the data-carrying symbols modulated on the useful subcarriers are predistorted or modified with a minimal signal distortion energy or penalty in terms of EVM. However, this method has limited application to a system with a large modulation alphabet, such as an alphabet size larger than 64QAM [HL05, RM13]. Hence, this scheme is unsuitable for a 5G NR-like system with the support of more than 64QAM, *e.g.*, 256/1024QAM modulation alphabet.

Clipping and filtering [Arm02, WL11] is a widely used technique for PAPR reduction because it is transparent to the standards. In this method, the (oversampled[7]) discrete-time domain signal amplitude is clipped if it is above a given target peak but keeps the same phase of as the original signal. Consequently, this clipping process renders in-band signal distortions and OOBE. Therefore, this induced OOBE must be filtered to meet the desired 3GPP ACLR requirements. However, if the OOBE is filtered out, then time domain signal peaks are regrown. Thus, one could view clipping and filtering from a numerical convex optimization perspective, where few iterations among clipping and filtering are required for PAPR reduction while removing unwanted OOBE at the expense of increased in-band signal distortion (or EVM) depending on the chosen target or threshold. Note that there is an inherent trade-off between an achievable PAPR level and EVM, see, *e.g.*, [SL13]. Similar to clipping and filtering, peak cancellation [KMG15, KMG20], peak windowing [PK98, vNdW98], and companding [HLCZ01, HLZ+04] are yet another signal distortion schemes that offer a similar trade-off as clipping and filtering.

Since the clipping- and filtering-like technique has been one of the suitable candidates that can meet the practical research objectives laid out in Section 1.2.2, recently, massive/large-scale MIMO has been capitalized to potentially further mitigate the distortion seen at the receiver [SL13, BFW+18, YCS+19, ZSR19, TS21]. Many distortion mitigation-based clipping and filtering-like techniques are initially posed as a nonconvex optimization problem mainly due to the structure/definition of the PAPR. More importantly, finding a solution to such a nonconvex problem is hard and practically infeasible unless some convex relaxation is performed to the PAPR definition—see in Chapter C.3.2 for the definition of PAPR. In the existing literature, there are at least two known ways to relax the nonconvexity of a problem: 1) instead of the actual PAPR reduction, one can only clip peaks resulting in a convex problem—similar to the abovementioned clipping step—see, *e.g.*, [SL13], which renders a suboptimal solution, and 2) semidefinite relaxation, see, *e.g.*, [YCS+19], which can yield a near-optimal solution at the cost of high computational complexity. Moreover, in many of these channel state information (CSI)-aware PAPR reduction methods [SL13, BFW+18, ZSR19, TS21], perfect CSI is assumed to be available at the transmitter. Clearly, in practice, we do not have access to perfect and/or complete[8] CSI at the transmitter. Consequently, targeting zero-EVM at the receiver is practically infeasible. Furthermore, it is worth adapting achievable PAPR and/or OOBE level, *e.g.*, according to the modulation and coding scheme used for the transmission, because the effective received EVM is dependent on the desired PAPR and/or OOBE level target. Additionally, in 3GPP-like standards, besides user-specific data channels, the downlink transmissions carry broadcast channels (such as cell-specific control channels), where the base station does not have the possibility to acquire CSI in those time or frequency resources. To the best of our knowledge,

[7]The signal is typically oversampled by a factor of 4, where this oversampling factor is shown to perform well in practice, see, *e.g.*, [OL95].

[8]In 5G NR, time division duplexing (TDD), the users sound the uplink channel via sounding reference signals [3GP20]. However, due to limited transmit power resources at the users' transmit side; the users may not be able to sound the full carrier bandwidth. Thus, the base station may not be able to acquire the complete downlink channel knowledge of that given user.

one of the works [YCS+19] has investigated several robust methods to cope with channel uncertainty. However, this work seemingly omitted the constraint on OOBE or ACLR in their optimization problem, which is equally essential for PAPR reduction techniques because the PAPR constraint causes OOBE. Furthermore, [YCS+19] employed semidefinite programming-based algorithms, which are typically solved by using interior-point-based methods [BV04] and have prohibitive computational complexity for realistic base station radio hardware systems. Hence, we seek a principled and computationally efficient technique for PAPR reduction while imposing ACLR constraints and handling both CSI-aware and CSI-unaware accordingly—see our research question 3 in Section 1.1.

1.3.3 First-Order Optimization For OOBE and PAPR Reduction

Convex optimization methods have become ubiquitous in addressing and solving many problems of interest in ML and signal processing for communication systems. As briefly discussed in Sections 1.3.1.4 and 1.3.2.3, the promising distortion-based OOBE reduction, *i.e.*, mask-compliant (non-orthogonal) spectral precoding, and distortion-based PAPR reduction techniques can be posed as (non-)convex optimization problems. Clearly, one cannot employ off-the-shelf general purpose solvers such as well-known CVX wrapper [GB14] with dedicated solvers, such as MOSEK, Gurobi, or SeDuMi to solve these problems for LS-MIMO-OFDM-based systems in realistic base station radio hardware systems. Thus, we seek tailored and computationally efficient or "implementation friendly" algorithms that can be deployed in realistic radios with application-specific integrated circuits.

Second-order optimization methods utilize second-order derivatives (assuming the function is at least twice differentiable), whereas first-order methods exploit first-order information, *i.e.*, (sub)gradient. Consequently, second-order methods yield higher accuracy and take fewer iterations to reach (near) optimal solutions than the first-order, see, *e.g.*, [BPC+11, Bec17, RY22]. Therefore, second-order methods are suited for small to medium-scale problems even though they are typically not easy to parallelize. However, for large-scale optimization problems, second-order methods are not scalable and have prohibitive computation and memory requirements. Hence, for large-scale optimization problems, first-order optimization methods are prevalent in this "big" (good) data era, owing to their low computational complexity and the possibility for parallelization in contrast to second-order optimization. Unfortunately, first-order methods can be too slow to converge to accurate solutions. Nevertheless, for many applications, modest accuracy suffices the requirements in practice.

In this thesis, we utilize and focus on first-order methods to tackle the large-scale optimization problems at hand, *i.e.*, for OOBE and PAPR reduction methods, particularly using proximal/operator splitting techniques such as alternating direction method of multipliers (ADMM)-like methods, see, overview paper [BPC+11]. In Chapter 2, we give an overview of many ADMM-like methods, including the background theory for proximal splitting techniques for completeness. Additionally, in the subsequent section, we briefly introduce FL, which is rooted in the distributed optimization techniques, typically first-order-based methods.

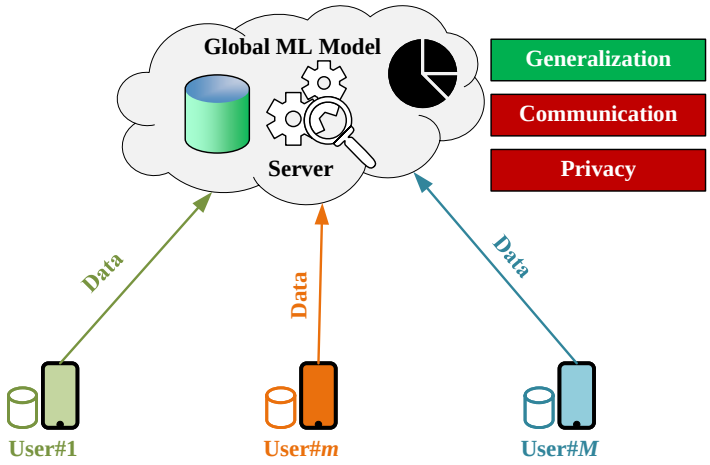


Figure 1.10: *Centralized learning: ML training is performed at the server (in the cloud), where dataset is collected from all the connected users. Subsequently, the server sends a global model to all the users.*

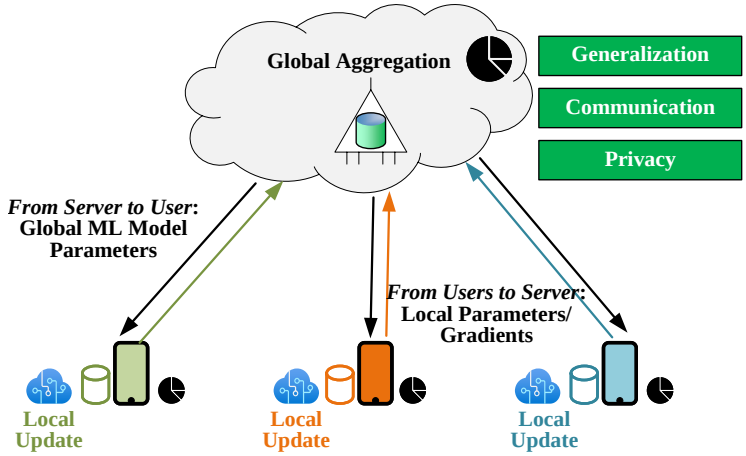


Figure 1.11: *Federated learning: ML training is performed locally at every connected user such that a global ML model is collaboratively computed at the server. The private raw data of each user resides locally, i.e., data is not shared with the server.*

1.3.4 Federated Learning

Traditionally, in centralized ML, see Figure 1.10, data from distributed users are uploaded to a central server. Consequently, a global model is learned at the server, and thereby the global model is shared with all the connected users to answer a concrete problem associated with the data, for instance, the power consumption of the user equipment. A global model learned from the users' data can be generalized very well such that the model can instantly be deployed to other compatible users/devices. However, training of centralized ML becomes exorbitantly costly in terms of communication when users must transfer an extensive dataset to the server. Moreover, for some users, sharing a (big) dataset with the server may not be achievable due to not only communication resource bottlenecks but also privacy concerns. Therefore, FL, a collaborative approach-based ML, see Figure 1.11, has recently been popularized for global ML model training without sending the private data of the users to the central/edge server [KMY+16, MMR+17]. More concretely, the FL framework entails the following steps: 1) local update, where each connected user trains the ML model using the data available to the user locally; 2) local updated weights (or parameters or gradients) are uploaded to the server; 3) global aggregation, where the server aggregates the weights received from the users to render a global, typically "average", model; and 4) send the global ML model to all the users in the network.

This FL can be seen as a synergy between distributed optimization techniques and ML. Many works, see, *e.g.*, [MMR+17, LSZ+20, TDPPN21, PW20, LSTS20, GLF22, ZL22], have been proposed in the literature to address FL related issues such as communication and computational efficiencies, because FL is one of the hottest topics in ML discipline. We explore the application of distributed proximal/operator splitting techniques in the context of FL capitalizing on the rich data available at the server because the base stations in cellular systems are expected to be colocated to the edge server to support various services [3GP22, S. 20, 5G 19], see research question 4 in Section 1.1. In Chapter 2, we overview some distributed optimization methods.

1.4 Contributions of the Thesis

In this thesis, we utilize key performance measures such as in-band and out-of-band metrics to optimize the signal distortion-based OOBE and PAPR reduction techniques for 5G NR and next-generation large-scale wireless communication systems. More concretely, we develop several computationally efficient algorithms for OOBE and PAPR reduction in large-scale MIMO-OFDM-based new generation cellular networks. Additionally, we investigate the feasibility of FL capitalizing on the rich dataset available at the edge server, where the base station is expected to be colocated near the edge server. Subsequently, each chapter is based on the following published or submitted manuscripts:

[J1] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "EVM-Constrained and Mask-Compliant MIMO-OFDM Spectral Precoding", *IEEE Transactions on Wireless Communications*, Vol. 20, No. 1, pp. 590–606, January 2021.

Table 1.1: *Performance objectives considered in the manuscripts reported in this thesis.*

	[J1]	**[J2]**	**[J3]**	**[J4]**
Objective	OOBE reduction—MSP		PAPR reduction	FL capitalizing on data at the edge server
In-band	Tx EVM constraint	Tx EVM mitigation observed at Rx		-
Out-of-band	ACLR improvement & meeting the spectral emission mask		ACLR constraint	-
Research questions	RQ1	RQ2	RQ3	RQ4

[J2] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "Efficient Optimization for Large-Scale MIMO-OFDM Spectral Precoding", *IEEE Transactions on Wireless Communications*, Vol. 20, No. 9, pp. 5496–5513, September 2021.

[J3] **Shashi Kant**, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "EVM Mitigation with PAPR and ACLR Constraints in Large-Scale MIMO-OFDM Using TOP-ADMM", accepted to *IEEE Transactions on Wireless Communications*, 2022.

[J4] **Shashi Kant**, José Mairton B. da Silva Jr., Gabor Fodor, Bo Göransson, Mats Bengtsson, Carlo Fischione, "Federated Learning using Three-Operator ADMM", submitted to *IEEE Journal of Selected Topics in Signal Processing*, 2022.

Table 1.1 explicitly presents the overview of the performance objectives of the aforementioned manuscripts used in this thesis. In this thesis, we consider OOBE and PAPR reduction of OFDM-based system to improve the desired performance measures in terms of in-band and out-of-band key performance metrics without impacting 3GPP 5G NR standards and receiver signal processing. All these publications consider an optimization problem and develop/investigate "implementation friendly" algorithm(s), where [J1] and [J2] tackle distortion-based OOBE reduction problems via MSP, *i.e.*, mask-compliant spectral precoding, and propose computationally efficient algorithms using powerful yet simple ADMM-type techniques, [J3] attempts to solve non-convex distortion-based PAPR reduction problem using three-operator ADMM-type methods, and [J4] performs the feasibility study on using our proposed three-operator ADMM-type algorithm to exploit the rich dataset available at the edge server in 5G NR and beyond. In the remaining part of this chapter, we elucidate our contributions to each manuscript accordingly.

Contributions by the Author of this Thesis

The author of this thesis has contributed significantly to all included manuscripts, especially in problem formulation, solution proposition, simulator development, result analysis, representation, and preparation of the first draft and updated version. The contributions of co-authors were purely advisory, such as how to formulate the problem correctly and paper writing. Moreover, in this thesis, the proposed solutions arguably pave the way for addressing sustainable development goals (SDGs) [Uni15]—established in 2015 by the United Nations—such as Goal 9 (Industry, Innovation, and Infrastructure) and Goal 13 (Climate Action).

1.4.1 EVM-Constrained & Mask-Compliant Spectral Precoding

Chapter A is based on [J1] for distortion-based OOBE reduction via non-orthogonal MSP. As discussed in Section 1.3.1.4, non-orthogonal MSP renders in-band signal distortion—quantified in terms of (transmit) EVM—that has a detrimental impact on the performance of high data-rate achieving MIMO-OFDM-based systems. Consequently, we formulate convex optimization problems with wideband and frequency-selective (transmit) EVM constraints for MSP that are amenable to proximal splitting algorithms. Subsequently, we develop two scalable and computationally efficient algorithms by dividing the large-scale MSP optimization problems with unconstrained and constrained EVM into smaller subproblems, where each subproblem is easy to tackle or solve efficiently. Undoubtedly, the resulting algorithms offer lower computational complexity than the general purpose solvers. More specifically, we propose solutions for both unconstrained and constrained EVM for MSP using ADMM-based algorithms. Moreover, we present extensive numerical results using a 5G NR (release-15 compliant) link-level simulator. Simulation results indicate that both the proposed schemes, namely unconstrained and constrained EVM for MSP, outcompete the previously developed techniques in terms of in-band and OOB key performance indicators.

1.4.2 Efficient Optimization for LS-MIMO-OFDM Spectral Precoding

Chapter B is based on [J2] for distortion-based OOBE reduction via non-orthogonal MSP. In [J2], we evolve [J1] by mitigating the signal distortion observed at the receiver(s) while performing MSP, because in [J1] we have observed spatially-coloured distortion, particularly in LS-MIMO-OFDM spectral precoding. In [J2], we pose optimization problems for the distortion (or EVM) mitigation incurred due to the MSP scheme in LS-MIMO-OFDM systems. To mitigate the distortion at the receiver, we capitalize on the excess degrees of freedom inherent in LS-MIMO-OFDM systems equipped with a large number of (physical) antennas. Additionally, we propose a robust convex constraint on distortion seen at the receiver (or predicted received EVM) to cope with the channel uncertainty or CSI imperfections but at the expense of increased computational complexity. In contrast to the existing joint spatial and spectral precoding method, we propose separating the spectral precoding from the signal spatial precoding or beamforming for practical deployment in realistic radio systems. Consequently, our spatial shaping constraint to signal distortion is agnostic to any (possibly non-linear) single-user multiple-input multiple-output (SU-MIMO) or multi-user multiple-input multiple-output (MU-MIMO) precoding. We formulate the convex optimization problem that takes the form of a sum of three functions or “operators"[9]. To develop an “implementation friendly" algorithm for such a cumbersome problem at hand, we propose a novel three-operator ADMM (TOP-ADMM) technique that can naturally consider the three-operator problem without resorting to product space and/or problem reformulations. In this work, we also show that the classical two-operator ADMM performs poorly in terms of convergence for solving the resulting three-operator problem. Subsequently, we apply our

[9] We discuss more of the problem formulations as one, two, and three operators in Chapter 2 of this thesis.

new TOP-ADMM algorithm to tackle the proposed three-operator problem, which shows promising convergence results. Furthermore, we present detailed simulation results that demonstrate the superior performance of our proposed MSP compared to existing methods, such as the joint spectral and spatial precoding method.

1.4.3 EVM Mitigation with PAPR and ACLR in LS-MIMO-OFDM Using TOP-ADMM

Chapter C is based on [J3] for distortion-based PAPR reduction in LS-MIMO-OFDM system. Analogous to [J1] and [J2], we observe a similar pattern that the distortions emanating from simple clipping and filtering-like schemes are also spatially coloured in nature for LS-MIMO-OFDM systems, such as transmitting data with user-specific spatial beamforming. Consequently, we formulate a robust problem that mitigates the transmit distortion seen at the receiver and/or minimizes the transmit signal distortion under CSI imperfection and/or incompleteness while flexibly offering to adapt the desired PAPR and ACLR targets. Therefore, our proposed problem formulation can cater to both beamforming and broadcasting carrying resources, where the frequency resources for the user-specific and common resources may be allocated in the same OFDM symbol. Unfortunately, the resulting proposed problem is non-convex. Nevertheless, we tackle this non-convex optimization problem using the majorization-minimization principle. We also solve the optimization problem using our proposed TOP-ADMM algorithm. To benchmark the performance of our TOP-ADMM, we also develop algorithms using other existing three-operator splitting algorithms. Numerical results corroborate the capability of our proposed TOP-ADMM-based algorithm.

1.4.4 Federated Learning Using Three-Operator ADMM

Chapter D is based on [J4], where we study the feasibility aspects of FL using our recently proposed TOP-ADMM algorithm. More concretely, in contrast to many existing FL schemes, we show the viability of a practical edge learning scenario. In addition to private datasets on the devices, a private dataset is also available on the edge server or base station. Subsequently, we model this edge learning scenario using our recently proposed TOP-ADMM technique that capitalizes on the private datasets on both edge servers and edge devices. Our numerical results indicate that our FL using TOP-ADMM has significant gain in terms of communication efficiency to reach the desired test accuracy with respect to existing FL that capitalizes only on the dataset available at the edge devices. In addition to proposing a novel FL using the TOP-ADMM algorithm, we extend our work published in [J2]/[J3] by establishing complete convergence proof of TOP-ADMM technique under general convex settings.

1.4.5 Contributions not Covered in the Thesis

In addition to the manuscripts listed above, we have published other papers during my PhD studies. The following publications are not explicitly included in the thesis to maintain

consistency and length. The papers are the following:

[C1] Shashi Kant, Mats Bengtsson, Gabor Fodor, Bo Göransson, Carlo Fischione, "Robust PAPR reduction in large-scale MIMO-OFDM using three-operator ADMM-type techniques", in 2021 55th Asilomar Conference on Signals, Systems, and Computers, 2021, pp. 616–622.

[C2] Shashi Kant, Gabor Fodor, Mats Bengtsson, Bo Göransson, Carlo Fischione, "Low-Complexity OFDM Spectral Precoding", in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Cannes, France, 2019.

References

[3GP17] 3GPP RAN WG1, R1-1715184. LS response on spectrum utilization. 3GPP TSG RAN WG1 #90 Meeting, Prague, Czech Republic, 21–25 Aug., 2017.

[3GP20] 3GPP TS 38.211. NR; Physical channels and modulation. V15.8.0, 2020.

[3GP21] 3GPP TS 38.104. NR; Base station (BS) radio transmission and reception. V15.14.0, 2021.

[3GP22] 3GPP TS 23.558. Architecture for enabling edge applications. V17.4.0, 2022.

[5G 19] 5G Americas, Whitepaper. 5G at the edge. `https://www.5gamericas.org/wp-content/uploads/2019/10/5G-Americas-EDGE-White-Paper-FINAL.pdf`, October 2019.

[Arm02] J. Armstrong. Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering. *Electron. Lett.*, 38(5):246–247, February 2002.

[BA19] A. A. A. Boulogeorgos and A. Alexiou. Analytical performance evaluation of beamforming under transceivers hardware imperfections. In *Proc. IEEE Wireless Commun. and Networking Conf.*, pages 1–7, 2019.

[BCS05] S. Brandes, I. Cosovic, and M. Schnell. Sidelobe suppression in OFDM systems by insertion of cancellation carriers. In *Proc. IEEE Veh. Technol. Conf.*, pages 152–156, 2005.

[Bec17] A. Beck. *First-order methods in optimization*. SIAM, 2017.

[BFW+18] H. Bao, J. Fang, Q. Wan, Z. Chen, and T. Jiang. An ADMM approach for PAPR reduction for large-scale MIMO-OFDM systems. *IEEE Trans. Veh. Technol.*, 67(8):7407–7418, August 2018.

[BHKD14] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah. Massive MIMO Systems With Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits. *IEEE Trans. Inf. Theory*, 60(11):7112–7139, September 2014.

[BHS17] E. Björnson, J. Hoydis, and L. Sanguinetti. Massive MIMO networks: Spectral, energy, and hardware efficiency. *Foundations and Trends®in Signal Processing*, 11(3-4):154–655, 2017.

[BLY13] E. Bala, J. Li, and R. Yang. Shaping spectral leakage: A novel low-complexity transceiver architecture for cognitive radio. *IEEE Veh. Technol. Mag.*, 8(3):38–46, July 2013.

[BMWH01] H. Breiling, S. Müller-Weinfurtner, and J. Huber. SLM peak-power reduction without explicit side information. *IEEE Commun. Lett.*, 5(6):239–241, June 2001.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[BV04] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[Cha66] R. W. Chang. Synthesis of band-limited orthogonal signals for multichannel data transmission. *The Bell System Technical Journal*, 45(10):1775–1796, 1966.

[Chu06] C.-D. Chung. Spectrally precoded OFDM. *IEEE Trans. Commun.*, 54(12):2173–2185, December 2006.

[Chu08] C.-D. Chung. Spectral precoding for rectangularly pulsed OFDM. *IEEE Trans. Commun.*, 56(9):1498–1510, September 2008.

[CML+22] W. Chen, J. Montojo, J. Lee, M. Shafi, and Y. Kim. The standardization of 5G-advanced in 3GPP. *IEEE Communications Magazine*, pages 1–7, 2022.

[DCCn+19] L. Díez, J. A. Cortés, F. J. Cañete, E. Martos-Naya, and S. Iranzo. A generalized spectral shaping method for OFDM signals. *IEEE Trans. Commun.*, 67(5):3541–3551, May 2019.

[DPS18] E. Dahlman, S. Parkvall, and J. Sköld. *5G NR: The next generation wireless access technology*. Academic Press, 2018.

[Eri] Ericsson. Products and Solutions: 5G RAN. `https://www.ericsson.com/en/5g-ran`. Accessed on Jan 16, 2020.

[Fau00] M. Faulkner. The effect of filtering on the performance of OFDM systems. *IEEE Trans. Veh. Technol.*, 49(5):1877–1884, September 2000.

[FLP+05] G. Fettweis, M. Lohning, D. Petrovic, M. Windisch, P. Zillmann, and W. Rave. Dirty RF: a new paradigm. In *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, volume 4, pages 2347–2355 Vol. 4, 2005.

[GB14] M. Grant and S. Boyd. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[GGLF08] B. Göransson, S. Grant, E. Larsson, and Z. Feng. Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA. In *Proc. IEEE Works. on Sign. Proc. Adv. in Wireless Commun.*, pages 496–500, 2008.

[GLF22] Y. Gong, Y. Li, and N. M. Freris. FedADMM: A robust federated deep learning framework with adaptivity to system heterogeneity. *arXiv e-prints*, page arXiv:2204.03529, April 2022.

[HL05] S. H. Han and J. H. Lee. An overview of peak-to-average power ratio reduction techniques for multicarrier transmission. *IEEE Wireless Commun.*, 12(2):56–65, April 2005.

[HLCZ01] X. Huang, J. Lu, J. Chuang, and J. Zheng. Companding transform for the reduction of peak-to-average power ratio of OFDM signals. In *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*, volume 2, pages 835–839, 2001.

[HLV22] K. Hussain and R. López-Valcarce. Joint Precoder and Window Design for OFDM Sidelobe Suppression. *IEEE Commun. Lett.*, pages 1–1, 2022.

[HLZ+04] X. Huang, J. Lu, J. Zheng, K. Letaief, and J. Gu. Companding transform for reduction in peak-to-average power ratio of OFDM signals. *IEEE Trans. Wireless Commun.*, 3(6):2030–2039, 2004.

[HZG15] X. Huang, J. A. Zhang, and Y. J. Guo. Out-of-band emission reduction and a unified framework for precoded OFDM. *IEEE Commun. Mag.*, 53(6):151–159, June 2015.

[ITU17] ITU-R. Guidelines for evaluation of radio interface technologies for IMT-2020. Report M.2412-0, 2017.

[J. 10] J. van de Beek. Orthogonal multiplexing in a subspace of frequency well-localized signals. *IEEE Commun. Lett.*, 14(10):882–884, 2010.

[JHJ+12] H.-S. Joo, S.-J. Heo, H.-B. Jeon, J.-S. No, and D.-J. Shin. A New Blind SLM Scheme With Low Decoding Complexity for OFDM Systems. *IEEE Transactions on Broadcasting*, 58(4):669–676, December 2012.

[KBF+21] S. Kant, M. Bengtsson, G. Fodor, B. Göransson, and C. Fischione. EVM-constrained and Mask-Compliant MIMO-OFDM Spectral Precoding. *IEEE Trans. Wireless Commun.*, 20(1):590–606, January 2021.

[KFB+19] S. Kant, G. Fodor, M. Bengtsson, B. Göransson, and C. Fischione. Low-complexity OFDM spectral precoding. In *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, 2019.

[KHLV21] R. Kumar, K. Hussain, and R. López-Valcarce. Mask-compliant orthogonal precoding for spectrally efficient OFDM. *IEEE Trans. Commun.*, 69(3):1990–2001, 2021.

[KJ03] B. Krongold and D. Jones. PAR reduction in OFDM via active constellation extension. *IEEE Trans. Broadcast.*, 49(3):258–268, 2003.

[KMG15] T. Kageyama, O. Muta, and H. Gacanin. An adaptive peak cancellation method for linear-precoded MIMO-OFDM signals. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 271–275, August 2015.

[KMG20] T. Kageyama, O. Muta, and H. Gacanin. Performance analysis of OFDM With peak cancellation under EVM and ACLR restrictions. *IEEE Trans. Veh. Technol.*, 69(6):6230–6241, June 2020.

[KMY+16] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

[KT16] R. Kumar and A. Tyagi. Computationally efficient mask-compliant spectral precoder for OFDM cognitive radio. *IEEE Trans. on Cogn. Commun. Netw.*, 2(1):15–23, June 2016.

[LGASK+09] S. Y. Le Goff, S. S. Al-Samahi, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif. Selected mapping without side information for PAPR reduction in OFDM. *IEEE Trans. Wireless Commun.*, 8(7):3320–3325, July 2009.

[LSJ+21] W. Lei, A. C. Soong, L. Jianghua, W. Yong, B. Classon, W. Xiao, D. Mazzarese, Z. Yang, and T. Saboorian. *5G system design: an end to end perspective*. Springer Cham, second edition, 2021.

[LSTS20] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.*, 37(3):50–60, 2020.

[LSZ+20] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith. Federated Optimization in Heterogeneous Networks. In *Machine Learning and Systems*, pages 429–450, 2020.

[MHJ+11] M. Ma, X. Huang, B. Jiao, Y. J. Guo, and S. Member. Optimal orthogonal precoding for power leakage suppression in DFT-based systems. *IEEE Trans. Commun.*, 59(3):844–853, March 2011.

[MMR+17] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, pages 1273–1282, 2017.

[MNJ18] M. Mohamad, R. Nilsson, and J. van de Beek. A novel transmitter architecture for spectrally-precoded OFDM. *IEEE Trans. Circuits Syst. I: Regular Papers*, 65(8):2592–2605, August 2018.

[Moh19] M. Mohamad. *Spectrally precoded OFDM design and analysis*. PhD dissertation, Department of Computer Science, Electrical and Space Engineering, Luleå University, Luleå, Sweden, 2019.

[MS96] D. Mestdagh and P. Spruyt. A method to reduce the probability of clipping in DMT-based transceivers. *IEEE Trans. Commun.*, 44(10):1234–1238, 1996.

[OL95] R. O'Neill and L. N. Lopes. Envelope variations and spectral splatter in clipped multicarrier signals. In *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, pages 71–75, Sep. 1995.

[PK98] M. Pauli and P. Kuchenbecker. On the reduction of the out-of-band radiation of OFDM-signals. In *ICC '98. 1998 IEEE International Conference on Communications. Conference Record. Affiliated with SUPERCOMM'98 (Cat. No.98CH36220)*, volume 3, pages 1304–1308, 1998.

[PM07] J. G. Proakis and D. G. Manolakis. *Digital signal processing: Principles, algorithms, and applications*. Pearson, fourth edition, 2007.

[Pra04] R. Prasad. *OFDM for wireless communications systems*. Artech House, 2004.

[PW20] R. Pathak and M. J. Wainwright. FedSplit: An algorithmic framework for fast federated optimization. In *Advances in Neural Information Processing Systems*, volume 33, pages 7057–7066, 2020.

[QWJ10] D. Qu, Z. Wang, and T. Jiang. Extended active interference cancellation for sidelobe suppression in cognitive radio OFDM systems with cyclic prefix. *IEEE Trans. Veh. Technol.*, 59(4):1689–1695, May 2010.

[RLR+22] I. Rahman, O. Liberg, S. M. Razavi, C. Hoynmann, S. Parkvall, G. Rune, R. Keller, and P. Persson. 5G Advanced: Evolution towards 6G. Technical Report BNEW-22:024836 Uen, Ericsson, June 2022.

[RM13] Y. Rahmatallah and S. Mohan. Peak-To-Average Power Ratio Reduction in OFDM Systems: A Survey And Taxonomy. *IEEE Commun. Surveys Tuts.*, 15(4):1567–1592, 4th Quart., 2013.

[RP-21] RP-213468. Summary for RAN Rel-18 package. `http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_94e/Docs/RP-213469.zip`, 3GPP RAN#94-e, December 2021. Accessed on October 2, 2022.

[RY22] E. K. Ryu and W. Yin. *Large-scale convex optimization via monotone operators*. Cambridge University Press (to be published), 2022.

[S. 20] S. Chitturi, 3GPP Working Group SA6 Chair. Enabling Edge Computing Applications in 3GPP. `https://www.3gpp.org/ftp/Information/Highlights/2020_Issue01/Flip_v01/mobile/index.html#p=8`, September 2020.

[Sam17] Samsung et al. R4-1709075. Way forward on spectrum utilization. 3GPP TSG RAN WG4 #84 Meeting, Berlin, Germany, 21–25 Aug., 2017.

[Sch08] T. Schenk. *RF imperfections in high-rate wireless systems*. Springer, The Netherlands, 2008.

[SL13] C. Studer and E. G. Larsson. PAR-Aware Large-Scale Multi-User MIMO-OFDM Downlink. *IEEE Journal on Selected Areas in Communications*, 31(2):303–313, February 2013.

[SWB10] C. Studer, M. Wenk, and A. Burg. MIMO transmission with residual transmit-RF impairments. In *Proc. ITG Workshop Smart Antennas*, pages 189–196, February 2010.

[SWB11] C. Studer, M. Wenk, and A. Burg. System-level implications of residual transmit-RF impairments in MIMO systems. In *Proc. European Conf. on Antennas and Propagation*, pages 2686–2689, April 2011.

[TC98] J. Tellado and J. M. Cioffi. Peak power reduction for multicarrier transmission. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, Sydney, Australia, November 1998.

[TDPPN21] Q. Tran-Dinh, N. H. Pham, D. T. Phan, and L. M. Nguyen. FedDR – Randomized Douglas-Rachford splitting algorithms for nonconvex federated composite optimization. In *Advances in Neural Information Processing Systems*, volume 34, pages 4447–4458, 2021.

[TS21] S. Taner and C. Studer. $\ell^p - \ell^q$-norm minimization for joint precoding and peak-to-average-power ratio reduction. *arXiv:2107.06986*, October 2021.

[TSA13] A. Tom, A. Sahin, and H. Arslan. Mask compliant precoder for OFDM spectrum shaping. *IEEE Commun. Lett.*, 17(3):447–450, 2013.

[TV05] D. Tse and P. Viswanath. *Fundamentals of wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[Uni15] United Nations, Department of Economic and Social Affairs, Sustainable Development. The 17 goals. `https://sdgs.un.org/goals#`, September 2015.

[vdB09] J. van de Beek. Sculpting the multicarrier spectrum: A novel projection precoder. *IEEE Commun. Lett.*, 13(12):881–883, 2009.

[vNdW98] R. van Nee and A. de Wild. Reducing the peak-to-average power ratio of OFDM. In *VTC '98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No.98CH36151)*, volume 3, pages 2072–2076, 1998.

[WFB+13] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J. No. The PAPR problem in OFDM transmission: New directions for a long-lasting problem. *IEEE Signal Process. Mag.*, 30(6):130–144, November 2013.

[WL11] Y.-C. Wang and Z.-Q. Luo. Optimized iterative clipping and filtering for PAPR reduction of OFDM signals. *IEEE Trans. Commun.*, 59(1):33–37, January 2011.

[WXX18] Y. Wang, S. Xie, and Z. Xie. FISTA-based PAPR reduction method for tone reservation's OFDM system. *IEEE Wireless Commun. Lett.*, 7(3):300–303, 2018.

[Yam04] H. Yamaguchi. Active interference cancellation technique for MB-OFDM cognitive radio. In *Proc. IEEE Eur. Microw. Conf.*, pages 1105–1108, 2004.

[YCS+19] M. Yao, M. Carrick, M. M. Sohul, V. Marojevic, C. D. Patterson, and J. H. Reed. Semidefinite Relaxation-Based PAPR-Aware Precoding for Massive MIMO-OFDM Systems. *IEEE Trans. Veh. Technol.*, 68(3):2229–2243, October 2019.

[YKLP+20] J. Yli-Kaakinen, T. Levanen, A. Palin, M. Renfors, and M. Valkama. Generalized fast-convolution-based filtered-OFDM: Techniques and application to 5G New Radio. *IEEE Trans. Signal Process.*, 68:1213–1228, 2020.

[ZHCG12] J. A. Zhang, X. Huang, A. Cantoni, and Y. J. Guo. Sidelobe suppression with orthogonal projection for multicarrier systems. *IEEE Trans. Commun.*, 60(2):589–599, February 2012.

[ZL22] S. Zhou and G. Y. Li. Federated learning via inexact ADMM. *arXiv e-prints*, page arXiv:2204.10607, April 2022.

[ZSR19] R. Zayani, H. Shaiek, and D. Roviras. PAPR-aware massive MIMO-OFDM downlink. *IEEE Access*, 7:25474–25484, February 2019.

Chapter 2

Preliminaries

This chapter summarizes the background theory on convex optimization and first-order methods, notably proximal/operator splitting techniques used in the thesis—see Figure 2.1. Observe that we present the background theory more from a complex-valued perspective, even though they are straightforwardly applicable to real-valued variables unless stated otherwise. Additionally, we focus only on deterministic optimization. For stochastic optimization, we refer the interested readers to, *e.g.*, [RY22] and references therein.

The remainder of this chapter is organized as follows. We first introduce some notations and well-known definitions, and then Section 2.2 provides some background on optimization with a focus on proximal/operator splitting methods. In Section 2.3, we briefly overview one of the popular methods projection onto convex sets (POCS), which can contestably be seen as a forerunner of proximal splitting algorithms. Section 2.4 and Section 2.5 briefly describe some two and three operators splitting methods.

2.1 Some Notations and Definitions

Let the set of complex and real numbers be denoted by $\mathbb{C}$ and $\mathbb{R}$, respectively. Let the extended real numbers be given by $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup +\infty \equiv (-\infty, +\infty]$. $\Re\{x\}$ denotes the real part of a complex number x. The vectors and matrices are denoted by boldface lowercase and uppercase symbols, *i.e.*, $\boldsymbol{x}$ and $\boldsymbol{X}$, respectively. The transpose, complex conjugate, and conjugate transpose of a vector or matrix are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, $(\cdot)^{\mathrm{H}}$, respectively. An i-th iterative update is defined as $(\cdot)^{(i)}$. All the norms $\|\cdot\|$ are ℓ_2 norm unless mentioned otherwise. The $\operatorname{dom} f$ represents the domain of a function f. f is said to be proper if its domain is nonempty. $\Gamma_0\left(\mathbb{C}^n\right)$ denotes a set of all proper lower semi-continuous convex functions from $\mathbb{C}^n$ to $\overline{\mathbb{R}}$ such that $\operatorname{dom} f \neq \varnothing$.

Definition 1 (Convex set [BV04, CLL17])**.** A set $\mathcal{X} \subseteq \mathbb{C}^n$ is called convex if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ and $\alpha \in [0,1]$, we have,

$$\alpha \boldsymbol{x} + (1-\alpha)\, \boldsymbol{y} \in \mathcal{X}.$$

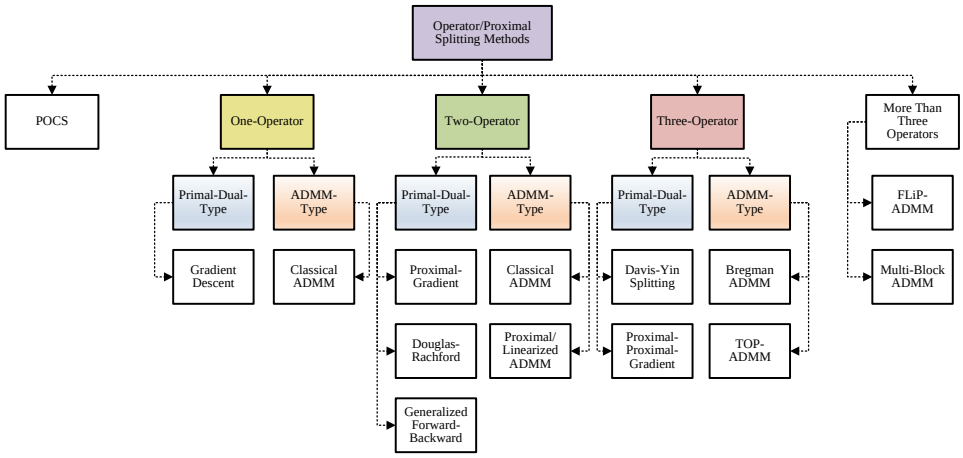


Figure 2.1: *Taxonomy of proximal/operator splitting techniques.*

Observe that i) convex set can be closed, *e.g.*, similar to the line interval $[0,1]$ in $\mathbb{R}$, or open, like $(0,1)$; ii) the intersection of convex sets is convex but the union of convex sets is not necessarily convex, see, *e.g.*, [BV04, Bec17, RY22, CLL17, CKCH21].

Definition 2 (Convex function [BV04,CLL17])**.** An extended real-valued function $f\colon \mathbb{C}^n \mapsto \mathbb{R}\cup\{+\infty\}$ is said to be convex if

- $\mathrm{dom} f$ is convex, *i.e.*, $\mathrm{dom}\, f = \{\boldsymbol{x} \in \mathbb{C}^n \colon f(\boldsymbol{x}) \neq +\infty\}$ and
- $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f$ and $\alpha \in [0,1]$.

Definition 3 (L-smooth function [BC11, Bec17])**.** A differentiable function $f(\boldsymbol{z}) \in \mathbb{R}$, where $\boldsymbol{z} \in \mathbb{C}^n$, is L-smooth, *i.e.*, has L-Lipschitz continuous gradient (for $L > 0$) if $\|\nabla f(\boldsymbol{z}_1) - \nabla f(\boldsymbol{z}_2)\| \leq L\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|\ \ \forall \boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{C}^n$.

Definition 4 (Subgradient [BC11, Definition 16.1], [LWW31])**.** A vector $\boldsymbol{g} \in \mathbb{C}^n$ is a subgradient of a proper function $f(\boldsymbol{z}) \in \mathbb{R}$ at some point $\boldsymbol{x}$ if

$$f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + 2\Re\left\{\boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z} - \boldsymbol{x})\right\} \quad \forall \boldsymbol{z} \in \mathbb{C}^n. \tag{2.1}$$

Definition 5 (Subdifferential [BV04, BC11])**.** Some useful properties of subdifferential:

- A proper function f is called subdifferential at some point $\boldsymbol{x}$ if there exists at least one subgradient (cf. Definition 4) at point $\boldsymbol{x}$.
- A function f is called subdifferential if it is subdifferential at all $\boldsymbol{x} \in \mathrm{dom} f$.

- The set of subgradients of f (cf. Definition 4) is called the subdifferential of f at x and denoted by ∂f.

Definition 6 (Proximal mapping [PBPB13, Bec17])**.** Given a proper closed convex function $f\colon \operatorname{dom} f \mapsto (-\infty\,,\,+\infty]$, then the proximal mapping of f is the operator given by

$$\mathbf{prox}_{\lambda f}(\boldsymbol{x}) \coloneqq (\boldsymbol{I}+\lambda\partial f)^{-1}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}\in\operatorname{dom} f}\left\{ f(\boldsymbol{z}) + \frac{1}{\beta\lambda}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \right\} \tag{2.2}$$

for any $\boldsymbol{x} \in \operatorname{dom} f$, where $\operatorname{dom} f$ corresponds to the domain of a function f, ∂f is a subdifferential of f, and $\lambda > 0$. If $\boldsymbol{z}$ is complex-valued or real-valued, $\beta = 1$ or $\beta = 2$, respectively.

Definition 7 (Proximal mapping of the indicator function [PBPB13, Bec17])**.** Let $f \coloneqq \delta_C : \operatorname{dom} f \mapsto \overline{\mathbb{R}}$ be an indicator function (also known as characteristic function),

$$f(\boldsymbol{z}) = \delta_{\mathcal{C}}(\boldsymbol{z}) \coloneqq \begin{cases} 0, & \boldsymbol{z} \in \mathcal{C} \\ +\infty, & \boldsymbol{z} \notin \mathcal{C}, \end{cases} \tag{2.3}$$

then the proximal mapping of a given set $\mathcal{C} \neq \varnothing$ is the orthogonal projection operator onto $\mathcal{C}$, *i.e.*,

$$\begin{aligned} \mathbf{prox}_{\lambda\delta_{\mathcal{C}}}(\boldsymbol{z}) &= \arg\min_{\boldsymbol{y}\in\operatorname{dom} f}\left\{ \delta_{\mathcal{C}}(\boldsymbol{y}) + \frac{1}{\beta\lambda}\|\boldsymbol{z}-\boldsymbol{y}\|_2^2 \right\} \\ &\equiv \arg\min_{\boldsymbol{y}\in\mathcal{C}}\left\{ \|\boldsymbol{z}-\boldsymbol{y}\|_2^2 \right\} \\ &= \mathbf{proj}_{\mathcal{C}}(\boldsymbol{z})\,. \end{aligned}$$

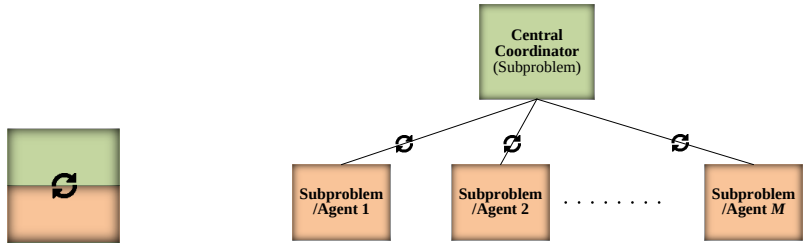
2.2 Background

Mathematical optimization finds a point in a given function f that achieves its minimum (or maximum) value, which can be expressed as

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}}\; f(\boldsymbol{x}), \tag{2.4}$$

where $\mathcal{X} \subseteq \mathbb{C}^n$ is a (convex) constraint set and $\boldsymbol{x} \in \mathbb{C}^n$ is a decision variable. To find such a (near) optimal point, an algorithm typically exhibits iterative nature. More specifically, if a function is convex (or concave), its local minimizers (or maximizers) are global, *i.e.*, in the whole search space, an optimal point attains a minimum (or maximum) value in its neighbourhood. Therefore, with respect to some criteria, optimization can be said as the search of a "best" point in a whole search space. More formally,

Definition 8. A point $\boldsymbol{x}^\star \in \mathcal{X}$ is locally optimal if $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x}$ in a neighbourhood of $\boldsymbol{x}^\star$. Globally optimal if $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$.



(a) *Centralized Optimization* **(b)** *Distributed Optimization*

Figure 2.2: *Illustrates centralized and distributed optimization. In case of centralized optimization, one can say that all the subproblems of a "big" problem are solved at one "location". However, in case of distributed optimization, the subproblems of a "big" problem are solved at different "locations" but jointly. In other words, the subproblems of a big problem are distributed, where agents tackle some subproblem and the central coordinator solves yet another subproblem by utilizing the updates/results from the agents.*

For example, let f in (2.4) be convex and differentiable at least once and $\mathcal{X} = \mathbb{C}^n$. Then, according to Fermat's rule, see, *e.g.*, [Bec17], the minimizers of f will be at the critical points, *i.e.*, at $\boldsymbol{x} \in \mathbb{C}^n$ when the gradient $\nabla f(\boldsymbol{x}) = \mathbf{0}$. Let f in (2.4) be non-smooth and $\mathcal{X} = \mathbb{C}^n$, but closed, convex, and proper, then an optimal solution is found by

$$\mathbf{0} \in \partial f\left(\boldsymbol{x}\right), \tag{2.5}$$

where ∂f is a subdifferential—see Definition 5. Moreover, from monotone operators theory[1] perspective, this above example can be expressed as a (one-operator) monotone inclusion problem, *i.e.*,

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{find}}\ \mathbf{0} \in \mathbb{A}\boldsymbol{x}, \tag{2.6}$$

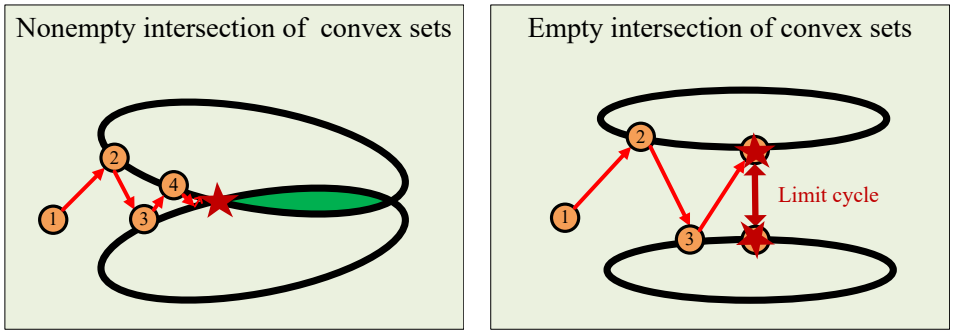
where $\mathbb{A} \equiv \partial f$ is called as maximally monotone operator.

A generic problem formulation can be more cumbersome than the abovementioned in (2.4). More broadly, many problems in wireless communications, machine learning, and signal processing among other disciplines can be written in the following form:

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}}\ f_1(\boldsymbol{A}_1\boldsymbol{x}) + f_2(\boldsymbol{A}_2\boldsymbol{x}) + \ldots + f_M(\boldsymbol{A}_M\boldsymbol{x}) \equiv \underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \sum_{m=1}^{M} f_m(\boldsymbol{A}_m\boldsymbol{x}), \tag{2.7}$$

where $\{f_m\}$ are convex (and possibly non-differentiable/non-smooth) functions, $\{\boldsymbol{A}_m\}$ are bounded linear operators, and $M \geq 1$. In the era of "big data" or large-scale problems, proximal splitting-based techniques—typically first-order optimization algorithms—are becoming extremely popular thanks to their low-computational complexity. There is a plethora of proximal splitting methods to tackle the generic composite problems,

[1] We refer the interested readers to [RY22, BC11] for monotone operators theory.



(a) *For* $\mathcal{C}_1 \bigcap \mathcal{C}_2 \neq \varnothing$. **(b)** *For* $\mathcal{C}_1 \bigcap \mathcal{C}_2 = \varnothing$.

Figure 2.3: *Exemplification of POCS with two convex sets.*

also referred to as operator splitting techniques [SNW11, DR56, EB92a, Com04, CP11b, CP11a, BPC$^+$11, HY12, PBPB13, RFP13, BST14, HA14, KP15, SSS$^+$16, GOY16, CHYY16, CHY17, Bec17, DY17, DLPY17, Yan18, Rag19, ACT20, BBC20, RY22].

In the sequel, we first present POCS technique, which can arguably be seen as a precursor of proximal splitting. Subsequently, we give brief overview of some proximal/operator splitting techniques tackling two-operator problem. Additionally, we comment on methods suitable for centralized and/or distributed[2] optimization where ever necessary—see simplified illustration on centralized and distributed optimization in Figure 2.2. Notice that FL typically tackles a distributed[3] optimization problem for machine learning. Lastly, we provide a short introduction to some three-operator-based splitting methods with a focus on ADMM-type techniques.

2.3 Projection Onto Convex Sets

One of the popular and widely used proximal splitting methods is POCS, see, *e.g.*, overview paper [CP11b, CP21], which is also known as alternating projections. In (2.7), assume each $f_m = \delta_{\mathcal{C}_m}$ and $\boldsymbol{A}_m = \boldsymbol{I}$, where $\{\mathcal{C}_m\}$ are nonempty closed convex sets. Consequently, the POCS algorithm attempts to solve the feasibility problem, *i.e.*,

$$\text{find} \qquad \boldsymbol{x} \in \mathbb{C}^n \tag{2.8a}$$

$$\text{subject to} \quad \boldsymbol{x} \in \bigcap_{m=1}^{M} \mathcal{C}_m. \tag{2.8b}$$

[2]We follow the terminology of [RY22]. Notably, distributed optimization requires a central coordination/server, whereas decentralized method do not require any central server. Observe that distributed optimization techniques can also be used to solve centralized optimization problems.

[3]Some FL techniques are also based on decentralized optimization, see, *e.g.*, [LSTS20], where there is no dedicated parameter server but all the agents solve the learning problem jointly. However, in this thesis, we focus on deterministic distributed optimization-based FL work.

The POCS-based iterative solution of (2.8) reads [CP11b, CP21]

$$\boldsymbol{x}^{(i+1)} = \mathbf{proj}_{\mathcal{C}_M}\left(\cdots \mathbf{proj}_{\mathcal{C}_1}\left(\boldsymbol{x}^{(i)}\right)\right), \quad \text{for } i = 0, 1, \ldots, \tag{2.9}$$

with appropriate initialization $\boldsymbol{x}^{(0)}$ and $\mathbf{proj}_{\mathcal{C}_m}(\cdot) = \mathbf{prox}_{\delta_{\mathcal{C}_m}}(\cdot)$, cf. Definition 7.

This POCS converges to a common point of an intersection of the closed convex sets (assuming intersection is non-empty), see, *e.g.*, [CP11b]. For instance, Figure 2.3a illustrates a simplified scenario where the intersection of a two convex sets is nonempty (shown in green coloured). Moreover, it shows some arbitrary point, denoted as encircled 1, which after the cyclic projection onto two sets eventually converges to a fixed-point. However, if the intersection is empty, then POCS may converge to a limit cycle in between the sets as depicted in Figure 2.3b.

There is so-called barycentric projection algorithm, see, *e.g.*, [CP21], whose solution reads

$$\boldsymbol{x}^{(i+1)} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{proj}_{\mathcal{C}_m}\left(\boldsymbol{x}^{(i)}\right), \quad \text{for } i = 0, 1, \ldots, \tag{2.10}$$

which can be seen as the distributed/parallel version of the POCS algorithm.

2.4 Two-Operator Splitting Methods

Operator splitting for two functions (or loosely speaking two-operator) has been widely studied by various researchers, see, *e.g.*, [SNW11, EB92a, Com04, CP11b, BPC$^+$11, PBPB13, KP15, SSS$^+$16, GOY16, Bec17, ACT20], to deal with the composite (two-operator) problem

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}}\; f(\boldsymbol{x}) + g(\boldsymbol{A}\boldsymbol{x}), \tag{2.11}$$

where $\boldsymbol{A}$ is a bounded linear operator, $f(\boldsymbol{x}),\ g(\boldsymbol{A}\boldsymbol{x}) \in \Gamma_0(\mathcal{X})$ and $\Gamma_0(\mathcal{X})$ denote a set of all proper lower semi-continuous convex functions from $\mathcal{X}$ to $\overline{\mathbb{R}}$. For simplicity, assuming $\boldsymbol{A} = \boldsymbol{I}$ in (2.11), a solution to such an optimization problem is

$$\mathbf{0} \in \partial f(\boldsymbol{x}) + \partial g(\boldsymbol{x}) \tag{2.12}$$

under some qualification conditions, which can equivalently be seen as a solution to a monotone inclusion problem, *i.e.*,

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{find}}\; \mathbf{0} \in \mathbb{A}\boldsymbol{x} + \mathbb{B}\boldsymbol{x}, \tag{2.13}$$

where $\mathbb{A} \equiv \partial f$ and $\mathbb{B} \equiv \partial g$ are called as maximally monotone operators—we refer the interested reader to, *e.g.*, [BC11,RY22] for further details. The splitting algorithms to solve such a composite problem by employing a divide-and-conquer approach can roughly be grouped into two classes, namely primal-dual-type and ADMM-type algorithms [Yan18].

2.4.1 Two-Operator-Based Primal-Dual-Type Algorithms

Within the primal-dual class, some well-known algorithms are forward-backward—also known as proximal-gradient descent—see, *e.g.*, [CP11b], Douglas-Rachford splitting (DRS)—see, *e.g.*, [DR56,EB92a,CP11b,ACT20], Chambolle-Pock [CP11a]. Furthermore, generalized forward-backward splitting [RFP13] can be seen as the distributed/parallel variant of Douglas-Rachford. Next, we briefly introduce some of these algorithms.

2.4.1.1 Proximal-Gradient Method

Consider problem (2.11) with $h \coloneqq g$ being differentiable having L-Lipschitz continuous gradient, *i.e.*,

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad f(\boldsymbol{x}) + h(\boldsymbol{A}\boldsymbol{x}). \tag{2.14}$$

The proximal-gradient method (also known as forward-backward splitting algorithm) reads, see, *e.g.*, [CP11b],

$$\begin{aligned}
&\text{For } i = 0, 1, \dots \\
&\qquad \boldsymbol{y}^{(i+1)} = \boldsymbol{x}^{(i)} - \tau^{(i)} \boldsymbol{A}^{\mathrm{H}} \nabla h \left(\boldsymbol{A}\boldsymbol{x}^{(i)} \right) && \text{(2.15a)} \\
&\qquad \boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \mu^{(i)} \left[\mathbf{prox}_{\tau^{(i)} f} \left(\boldsymbol{y}^{(i+1)} \right) - \boldsymbol{x}^{(i)} \right], && \text{(2.15b)}
\end{aligned}$$

where step-size $\tau^{(i)} \in \left(0, 2/\left(L \left\|\boldsymbol{A}\right\|^2\right)\right)$ and $\mu^{(i)} \in \left(0, 1\right]$. The generated sequence $\left\{\boldsymbol{x}^{(i)}\right\}$ converges under mild conditions—see, *e.g.*, [CP11b]. Observe that 1) if $f = \delta$ and $\mu^{(i)} = 1$ for all i then proximal-gradient method boils down to projection gradient descent, and 2) if $h = 0$ and $\mu^{(i)} = 1$ for all i then proximal-gradient method becomes classical gradient descent. Furthermore, there are some techniques to accelerate the convergence of proximal-gradient descent, see, *e.g.*, [RY22, LLF20], one of them is given below.

Nesterov's acceleration-based proximal-gradient descent

The accelerated variant of proximal-gradient method, see, *e.g.*, [RY22] and references therein, using so-called Nesterov's technique can be expressed as follows:

$$\begin{aligned}
&\text{For } i = 0, 1, \dots \\
&\qquad \boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau^{(i)} f} \left(\boldsymbol{y}^{(i)} - \tau^{(i)} \boldsymbol{A}^{\mathrm{H}} \nabla h \left(\boldsymbol{A}\boldsymbol{y}^{(i)} \right) \right) && \text{(2.16a)} \\
&\qquad \boldsymbol{y}^{(i+1)} = \boldsymbol{x}^{(i+1)} + \frac{i-1}{i+2} \left(\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)} \right). && \text{(2.16b)}
\end{aligned}$$

The sequence $\left\{\boldsymbol{x}^{(i)}\right\}$ generated by the above algorithm converges under given conditions—see, *e.g.*, [RY22] and references therein.

2.4.1.2 Proximal-Gradient Method for Distributed Optimization

A distributed proximal-gradient method, see, *e.g.*, [RY22], solves a problem of the form

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad f(\boldsymbol{x}) + \sum_{m=1}^{M} h_m(\boldsymbol{A}\boldsymbol{x}), \tag{2.17}$$

where h_m have L-Lipschitz continuous gradient; f is closed, convex, and proper functions; and $M \geq 1$. The solution can be seen as a combination of forward-backward and consensus trick at the central entity. More specifically, the algorithm can be shown as [RY22]

For $i = 0, 1, \ldots$

$$\boldsymbol{y}_m^{(i+1)} = \boldsymbol{x}^{(i)} - \tau \boldsymbol{A}^{\mathrm{H}} \nabla h_m \left(\boldsymbol{A}\boldsymbol{x}^{(i)} \right), \ \forall m = 1, \ldots, M \tag{2.18a}$$

$$\boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau f} \left(\frac{1}{M} \sum_{m=1}^{M} \boldsymbol{y}_m^{(i+1)} \right), \tag{2.18b}$$

where step size $\tau \in (0, 2/L)$. There are also accelerated variants of distributed proximal-gradient method, see, *e.g.*, [CO12]. Notice that one may arguably view (2.17) with more than two operators. However, we treat it as a two-operator problem.

2.4.1.3 Douglas-Rachford Splitting Method

Consider problem (2.11) with $\boldsymbol{A} = \boldsymbol{I}$, *i.e.*,

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad f(\boldsymbol{x}) + g(\boldsymbol{x}). \tag{2.19}$$

Consider $\tau \in (0, \infty)$, and a sequence of relaxation parameters $\rho_i \in (0, 2)$ $\forall i \geq 0$ and satisfy $\sum_i \rho_i (2 - \rho_i) = +\infty$. Then, every sequence $\left\{ \boldsymbol{x}^{(i)} \right\}$ generated by the following iterative scheme

For $i = 0, 1, \ldots$

$$\boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau f} \left(\boldsymbol{y}^{(i)} \right) \tag{2.20a}$$

$$\boldsymbol{y}^{(i+1)} = \boldsymbol{y}^{(i)} + \rho_i \left(\mathbf{prox}_{\tau g} \left(2\boldsymbol{x}^{(i+1)} - \boldsymbol{y}^{(i)} \right) - \boldsymbol{x}^{(i+1)} \right) \tag{2.20b}$$

converges to a solution of (2.19) with some arbitrary initial $\boldsymbol{y}^{(0)}$—see, *e.g.*, [DR56] [Com04, Corollary 5.2] [EB92b].

Note that if $\boldsymbol{A} \neq \boldsymbol{I}$ then DRS cannot be employed unless $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \nu \boldsymbol{I}$ for $\nu \in \mathbb{R}_{>0}$. The reason is that it is difficult to find a computationally efficient proximal operator for $g(\boldsymbol{A}\boldsymbol{x})$ without $\boldsymbol{A} = \boldsymbol{I}$ or holding $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \nu \boldsymbol{I}$. However, there are other two-operator-based primal-dual-type methods that can work for any bounded linear operator $\boldsymbol{A}$ and are referred to as Chambolle-Pock, Loris-Verhoeven, and Condat-Vu algorithms—see details in, *e.g.*, [CKCH21, RY22]. These methods typically exploit the conjugate function of f in the above composite function (2.11) without assuming g being differentiable.

2.4.1.4 Generalized Forward-Backward Splitting for Distributed Optimization

The generalized forward-backward splitting [RFP13] solves a problem of the form

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad \sum_{m=1}^{M} f_m(\boldsymbol{x}) + h(\boldsymbol{x}), \tag{2.21}$$

where h has L-Lipschitz continuous gradient; f_m are closed, convex, and proper functions; and $M \geq 1$. The solution can be seen as a combination of forward-backward and DRS. More specifically, the algorithm can be shown as [RFP13]

For $i = 0, 1, \ldots$

$$\boldsymbol{y}_m^{(i+1)} = \boldsymbol{y}_m^{(i)} + \mathbf{prox}_{M\tau^{(i)} f_m}\Big(2\boldsymbol{x}^{(i)} - \boldsymbol{y}_m^{(i)} - \tau^{(i)}\nabla h\Big(\boldsymbol{x}^{(i)}\Big)\Big) - \boldsymbol{x}^{(i)} \;\forall m = 1, \ldots, M, \tag{2.22a}$$

$$\boldsymbol{x}^{(i+1)} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{y}_m^{(i+1)}, \tag{2.22b}$$

where step sizes $\big\{\tau^{(i)} \in (0, 2/L)\big\}$. Observe that this method can also be useful for tackling two-operator-based centralized optimization problem. Moreover, similar to problem (2.17), problem (2.21) can contestably be seen as more than two operators problem. But, in this thesis, we consider it as a two-operator problem.

2.4.1.5 Douglas-Rachford Splitting for Distributed Optimization

In [CKCH21], the authors have evolved classical DRS—briefly described in Section 2.4.1.3—to solve distributed optimization problem. More concretely, this parallel version of DRS attempts to solve the problem of the form

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad \sum_{m=1}^{M} f_m(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{2.23}$$

where $\{f_m\}$ and g are closed, convex, and proper functions. One of the variants of this distributed DRS algorithm reads [CKCH21]

For $i = 0, 1, \ldots$

$$\boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau g}\left(\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{y}_m^{(i+1)}\right) \tag{2.24a}$$

$$\boldsymbol{y}_m^{(i+1)} = \boldsymbol{y}_m^{(i)} + \rho^{(i)}\left[\mathbf{prox}_{M\tau f_m}\Big(2\boldsymbol{x}^{(i+1)} - \boldsymbol{y}_m^{(i)}\Big) - \boldsymbol{x}^{(i+1)}\right] \;\forall m = 1, \ldots, M, \tag{2.24b}$$

where $\big\{\rho^{(i)} \in \mathbb{R}_{\geq 0}\big\}$. Notice that, similar to problem (2.17) and (2.21), problem (2.23) can arguably be viewed as more than two operators problem. However, in this thesis, we treat it as a two-operator problem.

2.4.2 Two-Operator-Based ADMM-Type Algorithms

In this section, we briefly introduce some ADMM-type methods that generally employs similar concept as primal-dual-type, *i.e.*, divide and conquer, to decompose the large-scale composite problem into easy-to-solve subproblems. Among ADMM-type algorithms, some popular techniques are (two-block and "two-operator"[4]) classical ADMM, see, *e.g.*, [GM75, EB92a, CP11b, BPC$^+$11, PBPB13, KP15, LWW31, SSS$^+$16, GOY16, Bec17] and proximal ADMM [BST14]—which is a generalized version of the so-called linearized ADMM [Bec17]. There is yet another variant of classical ADMM known as symmetric ADMM, see, *e.g.*, [GOY16, HMY16].

2.4.2.1 Classical ADMM

Consider a problem, with a two-block and a separable objective,

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{C}^n,\boldsymbol{y}\in\mathbb{C}^m}{\text{minimize}} \quad & f(\boldsymbol{x})+g(\boldsymbol{y}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}=\boldsymbol{c}, \end{aligned} \tag{2.25}$$

where f and g are closed, convex, and proper functions; $\boldsymbol{A}\in\mathbb{C}^{p\times n}$; $\boldsymbol{B}\in\mathbb{C}^{p\times m}$; and $\boldsymbol{c}\in\mathbb{C}^p$. Toward the goal to solve (2.25) via ADMM, we first form the augmented Lagrangian

$$\mathcal{L}_\rho\left(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}\right) \coloneqq f(\boldsymbol{x})+g(\boldsymbol{y})+2\Re\left\{\boldsymbol{\lambda}^{\mathrm{H}}\left(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}\right)\right\}+\rho\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}\right\|_2^2. \tag{2.26}$$

ADMM iterative algorithm is, see, *e.g.*, [EB92a, CP11b, BPC$^+$11, PBPB13, KP15, LWW31, SSS$^+$16, GOY16, Bec17]:

$$\boldsymbol{x}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{x}} \mathcal{L}_\rho\left(\boldsymbol{x},\boldsymbol{y}^{(i)},\boldsymbol{\lambda}^{(i)}\right) \tag{2.27a}$$

$$\equiv \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \rho\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2 \tag{2.27b}$$

$$\boldsymbol{y}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{y}} \mathcal{L}_\rho\left(\boldsymbol{x}^{(i+1)},\boldsymbol{y},\boldsymbol{\lambda}^{(i)}\right) \tag{2.27c}$$

$$\equiv \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \rho\left\|\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2 \tag{2.27d}$$

$$\boldsymbol{\lambda}^{(i+1)} \coloneqq \boldsymbol{\lambda}^{(i)} + \gamma\rho\left(\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}^{(i+1)}-\boldsymbol{c}\right), \tag{2.27e}$$

where relaxation parameter $\rho\in\mathbb{R}_{>0}$ and $\gamma\in(0,2)$ is so-called parameter[5] of Fortin and Glowinski, see, *e.g.*, [FG83]. The generated sequence $\left[\boldsymbol{x}^{(i)},\boldsymbol{y}^{(i)},\boldsymbol{\lambda}^{(i)}\right]$ converges to the solution of (2.25) under mild conditions, see, *e.g.*, [BPC$^+$11]. Observe that the

[4] We overload operator terminology for ADMM-type algorithms.

[5] Typically, for ADMM γ can be chosen between zero and the golden ratio, *i.e.*, $\gamma\in\left(0,(1+\sqrt{5})/2\right)$, as proposed by Fortin and Glowinski. However, for proximal ADMM, γ can be further relaxed and can be chosen between zero and two, *i.e.*, $\gamma\in(0,2)$—see, *e.g.*, [Ma20].

sequence updates of ADMM are performed in Gauss-Seidel fashion. Therefore, one can swap $\boldsymbol{x}^{(i+1)}$ (2.27a)/(2.27b) and $\boldsymbol{y}^{(i+1)}$ (2.27c)/(2.27d) updates without penalizing the convergence guarantee but the generated sequences over iterations may be different in general [GOY16, Chapter 5].

There is yet another popular variant of ADMM, which is called as proximal ADMM by adding proximity terms $\left\|\boldsymbol{x}-\boldsymbol{x}^{(i)}\right\|_{\boldsymbol{Q}_x}^2$ and $\left\|\boldsymbol{y}-\boldsymbol{y}^{(i)}\right\|_{\boldsymbol{Q}_y}^2$ with positive definite matrices $\boldsymbol{Q}_x$ and $\boldsymbol{Q}_y$ to (2.27a)/(2.27b) and (2.27c)/(2.27d), respectively. This variant of ADMM is also called as generalized ADMM, see, *e.g.*, [RY22] and references therein. More concretely, the special case of proximal ADMM is sometimes referred to as linearized ADMM if the subproblems are not easy to solve because the primal updates may contain a computationally inconvenient quadratic term. Therefore, one can "linearize"[6] the subproblem by choosing appropriate

$$\mathbb{C}^{n\times n} \ni \boldsymbol{Q}_x = \alpha\boldsymbol{I}-\rho\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \tag{2.28a}$$

$$\mathbb{C}^{m\times m} \ni \boldsymbol{Q}_y = \beta\boldsymbol{I}-\rho\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B} \tag{2.28b}$$

with $\alpha > \rho\left\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right\|$ and $\beta > \rho\left\|\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right\|$, where $\left\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right\|$ and $\left\|\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}\right\|$ are operator norms—*i.e.*, maximum eigenvalue of $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ and $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$, respectively. Consequently, the proximal linearized ADMM updates are given below:

$$\boldsymbol{x}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \rho\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2 + \left\|\boldsymbol{x}-\boldsymbol{x}^{(i)}\right\|_{\boldsymbol{Q}_x}^2 \tag{2.29a}$$

$$\overset{\text{(a)}}{=} \mathbf{prox}_{\frac{1}{\alpha}f}\left(\boldsymbol{x}^{(i)} - \frac{\rho}{\alpha}\boldsymbol{A}^{\mathrm{H}}\left[\boldsymbol{A}\boldsymbol{x}^{(i)}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right]\right) \tag{2.29b}$$

$$\boldsymbol{y}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \rho\left\|\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2 + \left\|\boldsymbol{y}-\boldsymbol{y}^{(i)}\right\|_{\boldsymbol{Q}_y}^2 \tag{2.29c}$$

$$\overset{\text{(b)}}{=} \mathbf{prox}_{\frac{1}{\beta}g}\left(\boldsymbol{y}^{(i)} - \frac{\rho}{\beta}\boldsymbol{B}^{\mathrm{H}}\left[\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right]\right) \tag{2.29d}$$

$$\boldsymbol{\lambda}^{(i+1)} \coloneqq \boldsymbol{\lambda}^{(i)} + \gamma\rho\left(\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}^{(i+1)}-\boldsymbol{c}\right),$$

where in (a) and (b) we have plugged in (2.28).

For function linearization: in some cases, if a function is differentiable but computing a proximal operator is hard such as in logistic regression, then one can perform inexact update by considering first-order Taylor expansion of a function, *e.g.*, in (2.27b) with a first-order approximation yields

$$\begin{aligned}\boldsymbol{x}^{(i+1)} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \rho\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2 \\ &\approx \arg\min_{\boldsymbol{x}} f\left(\boldsymbol{x}^{(i)}\right) + \left\langle\nabla f\left(\boldsymbol{x}^{(i)}\right), \boldsymbol{x}-\boldsymbol{x}^{(i)}\right\rangle + \rho\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}+\frac{1}{\rho}\boldsymbol{\lambda}^{(i)}\right\|_2^2.\end{aligned}$$

[6] Observe that the linearization term in ADMM-type algorithms is overloaded [RY22]. In particular, one can linearize a subproblem and a function.

2.4.2.2 Classical Consensus ADMM for Distributed Optimization

The two-operator-based consensus ADMM [BPC+11, PBPB13] is one of the popular methods among (distributed) optimization and machine learning communities to solve problems of the form

$$\begin{aligned} \underset{\{\boldsymbol{x}_m\in\mathbb{C}^n\},\boldsymbol{y}\in\mathbb{C}^n}{\text{minimize}} \quad & \sum_{m=1}^{M} f_m(\boldsymbol{x}_m)+g(\boldsymbol{y}) \\ \text{subject to} \quad & \boldsymbol{x}_m-\boldsymbol{y}=\boldsymbol{0}, \quad \forall m=1,\dots,M, \end{aligned} \tag{2.30}$$

where $\{f_m\}$ and $g(\cdot)$ are closed, convex, and proper functions. The classical ADMM algorithm that solves problem (2.30) can be summarized, following [BPC+11, Chapter 7],

$$\boldsymbol{x}_m^{(i+1)} \coloneqq \arg\min_{\boldsymbol{x}_m} f_m(\boldsymbol{x}_m) + \rho \left\| \boldsymbol{x}_m-\boldsymbol{y}^{(i)}+\frac{\boldsymbol{\lambda}_m^{(i)}}{\rho} \right\|_2^2 \quad \forall m=1,\dots,M, \tag{2.31a}$$

$$\boldsymbol{y}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \sum_{m=1}^{M} \rho \left\| \boldsymbol{x}_m^{(i+1)}-\boldsymbol{y}+\frac{\boldsymbol{\lambda}_m^{(i)}}{\rho} \right\|_2^2 \tag{2.31b}$$

$$\boldsymbol{\lambda}_m^{(i+1)} \coloneqq \boldsymbol{\lambda}_m^{(i)}+\gamma\rho\left(\boldsymbol{x}_m^{(i+1)}-\boldsymbol{y}^{(i+1)}\right) \ \forall m=1,\dots,M, \tag{2.31c}$$

where $\boldsymbol{\lambda}_m^{(i)}\in\mathbb{C}^n$ is the Lagrange multiplier, $\rho\in\mathbb{R}_{>0}$ is a penalty parameter, and $\gamma\in(0,2)$ is a Glowinski's parameter.

If the functions f_m and g contain a computationally inconvenient quadratic term, then one could employ proximal linearized ADMM, see, *e.g.*, [RY22] and references therein, to cancel out such a quadratic term. More specifically, in linearized ADMM, one adds a so-called proximal term $\left\|\boldsymbol{x}_m-\boldsymbol{x}_m^{(i)}\right\|_{\boldsymbol{Q}_x}$ and $\left\|\boldsymbol{y}-\boldsymbol{y}^{(i)}\right\|_{\boldsymbol{Q}_y}$ with positive definite matrices $\boldsymbol{Q}_x$ and $\boldsymbol{Q}_y$—see (2.28)—to (2.31a) and (2.31b), respectively, using definitions.

2.5 Three-Operator Splitting Methods

In general, operator/proximal splitting with more than two functions (or operators) in the objective, such as problem (2.7), are typically handled by applying the product space and/or problem reformulation techniques [Pie84, DY17]. Consequently, these techniques result in a block-wise two-operator problem, which can be tackled by any suitable two-operator splitting methods—cf. Section 2.4. Subsequently, one can invoke the separability of grouped functions within each block. Unfortunately, such reformulations are not straightforward, or may exhibit slow convergence in practice, or result in memory inefficiency, see, *e.g.*, [DY17, Yan18, KBG+21, CKCH21].

In contrast to a rich literature on the two-operator splitting methods, to the best of our knowledge, there are a handful of techniques in the literature that can tackle three (or more) operators in a composite problem for centralized and/or distributed

optimization. Concretely, some authors have extended the primal-dual-type [DY17, Yan18, RY19, CKCH21] and ADMM-type [BBC20, KBG+21, HA14, RY22] splitting algorithms for solving three (or more) operators problem. It is worth to point out that the suitable two-operator problem can be solved by three-operator problem. In the sequel, we briefly describe some of these three-operator splitting algorithms.

2.5.1 Three-Operator-Based Primal-Dual-Type Algorithms

This section briefly introduces some methods, including distributed versions, within the primal-dual class, tackling a problem with three operators.

2.5.1.1 Davis-Yin Splitting Method

Consider a problem of the following form

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad f(\boldsymbol{x}) + g(\boldsymbol{x}) + h(\boldsymbol{x}), \tag{2.32}$$

where f and g are closed, convex, and proper functions; and h is L-smooth function.

The Davis-Yin splitting (DYS) [DY17] algorithm reads

For $i = 0, 1, \ldots$

$$\boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau f}\left(\boldsymbol{y}^{(i)}\right) \tag{2.33a}$$

$$\boldsymbol{y}^{(i+1)} = \boldsymbol{y}^{(i)} + \rho^{(i)}\left(\mathbf{prox}_{\tau g}\left(2\boldsymbol{x}^{(i+1)} - \boldsymbol{y}^{(i)} - \tau\nabla h\left(\boldsymbol{x}^{(i+1)}\right)\right) - \boldsymbol{x}^{(i+1)}\right), \tag{2.33b}$$

where step size $\tau \in (0, 2/L)$. Consider some arbitrary initial $\boldsymbol{y}^{(0)} \in \mathscr{X}$. Set $\varrho = 2 - (\tau L)/2$. Let $\left(\rho^{(i)}\right)_{i\in\mathbb{N}}$ be a sequence in $[0, \varrho]$ such that $\sum_{i\in\mathbb{N}} \rho^{(i)}\left(\varrho - \rho^{(i)}\right) = +\infty$. Then, the sequence $\left\{\boldsymbol{x}^{(i+1)}\right\}$ generated by (2.33) converges to a Karush-Kuhn-Tucker (KKT) stationary point of (2.32).

2.5.1.2 Proximal-Proximal-Gradient Method for Distributed Optimization

Consider an optimization problem

$$\underset{\boldsymbol{x}\in\mathcal{X}}{\text{minimize}} \quad f(\boldsymbol{x}) + \sum_{m=1}^{M}\left(g_m(\boldsymbol{x}) + h_m(\boldsymbol{x})\right), \tag{2.34}$$

where f and $\{g_m\}$ are closed, convex, and proper functions; $\{h_m\}$ are L-smooth functions.

The updates of proximal-proximal-gradient method [RY19] are as follows:

For $i = 0, 1, \ldots$

$$\boldsymbol{x}^{(i+1)} = \mathbf{prox}_{\tau f}\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{z}_m^{(i)}\right) \tag{2.35a}$$

$$\boldsymbol{y}_m^{(i+1)} = \mathbf{prox}_{\tau g_m}\left(2\boldsymbol{x}^{(i+1)} - \boldsymbol{y}_m^{(i)} - \tau\nabla h_m\left(\boldsymbol{x}^{(i+1)}\right)\right) \tag{2.35b}$$

$$\boldsymbol{z}_m^{(i+1)} = \boldsymbol{z}_m^{(i)} + \boldsymbol{y}_m^{(i)} - \boldsymbol{x}^{(i+1)}, \tag{2.35c}$$

where $\tau > 0$ is a step size parameter. Interestingly, this proximal-proximal-gradient method can also be seen as a distributed version of DYS [CKCH21].

2.5.2 Three-Operator-Based ADMM-Type Algorithms

In this section, we briefly overview some three-operator-based ADMM-type algorithms including distributed versions.

2.5.2.1 Generalized Bregman ADMM

Bregman ADMM (BADMM) [HA14] generalizes classical two-operator ADMM by replacing the quadratic penalty term in the augmented Lagrangian (2.26) with a Bregman divergence or distance.

Definition 9 (Bregman distance [Bre16, CT93a])**.** Let ϕ be a convex and continuously differentiable function on the interior of dom_ϕ, the Bregman distance between two points $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\,(\phi)$ is given by

$$B_\phi(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \phi(\boldsymbol{x}) - \phi(\boldsymbol{y}) - \langle\nabla\phi(\boldsymbol{y}), (\boldsymbol{x} - \boldsymbol{y})\rangle.$$

The Bregman distance is nonnegative, *i.e.*, $B_\phi(\boldsymbol{x}, \boldsymbol{y}) \geq 0$, because ϕ is convex; $B_\phi(\boldsymbol{x}, \boldsymbol{y}) = 0$ implies that $\boldsymbol{x} = \boldsymbol{y}$; $B_\phi(\boldsymbol{x}, \boldsymbol{y}) = B_\phi(\boldsymbol{y}, \boldsymbol{x})$ is not necessary—see, *e.g.*, [CT93b] for more details on the Bregman distance. For example, if $\phi(x) = \|\boldsymbol{x}\|^2$, then $B_\phi(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.

The generalized BADMM has the following updates:

$$\begin{aligned}\boldsymbol{x}^{(i+1)} = &\arg\min_{\boldsymbol{x}\in\mathbb{C}^n} f(\boldsymbol{x}) + 2\Re\left\{\left(\boldsymbol{\lambda}^{(i)}\right)^{\mathrm{H}}\left(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(i)} - \boldsymbol{c}\right)\right\} \\ &+ \rho B_\phi\left(\boldsymbol{c} - \boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{y}^{(i)}\right) + \rho_x B_{\varphi_x}\left(\boldsymbol{x}, \boldsymbol{x}^{(i+1)}\right)\end{aligned} \tag{2.36a}$$

$$\begin{aligned}\boldsymbol{y}^{(i+1)} = &\arg\min_{\boldsymbol{y}\in\mathbb{C}^p} g(\boldsymbol{y}) + 2\Re\left\{\left(\boldsymbol{\lambda}^{(i)}\right)^{\mathrm{H}}\left(\boldsymbol{A}\boldsymbol{x}^{(i+1)} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{c}\right)\right\} \\ &+ \rho B_\phi\left(\boldsymbol{B}\boldsymbol{y}, \boldsymbol{c} - \boldsymbol{A}\boldsymbol{x}^{(i+1)}\right) + \rho_y B_{\varphi_y}\left(\boldsymbol{y}, \boldsymbol{y}^{(i+1)}\right)\end{aligned} \tag{2.36b}$$

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} + \tau\left(\boldsymbol{A}\boldsymbol{x}^{(i+1)} + \boldsymbol{B}\boldsymbol{y}^{(i+1)} - \boldsymbol{c}\right), \tag{2.36c}$$

where $\rho, \tau \in \mathbb{R}_{>0}$ and $\rho_x, \rho_y \in \mathbb{R}_{\geq 0}$, and Bregman distances $B_\phi(\cdot,\cdot)$, $B_{\varphi_x}(\cdot,\cdot)$, and $B_{\varphi_y}(\cdot,\cdot)$ with appropriate choices of convex functions ϕ, φ_x, and φ_y, respectively. The generated sequences $\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}, \boldsymbol{\lambda}^{(i)}\right)$ converge to a KKT point of (2.25)—see convergence analysis in [HA14].

At first glance, the application of BADMM to solve three-operator problem such as

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{C}^n, \boldsymbol{y}\in\mathbb{C}^p}{\text{minimize}} \quad & f(\boldsymbol{x})+g(\boldsymbol{y})+h(\boldsymbol{y}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}=\boldsymbol{c}, \end{aligned} \tag{2.37}$$

may not be obvious. However, with an appropriate choice of Bregman distance one can tackle problem (2.37). Since h is L-smooth function, then following the trick described in [HA14, Section 2.1] results in the following BADMM updates

$$\begin{aligned} \boldsymbol{x}^{(i+1)} \coloneqq & \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \left\langle \nabla h\left(\boldsymbol{x}^{(i)}\right), \boldsymbol{x}\right\rangle + \rho_x \left\|\boldsymbol{x}-\boldsymbol{x}^{(i)}\right\|_2^2 \\ & + \left\langle \boldsymbol{\lambda}^{(i)}, \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}\right\rangle + \rho \left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}^{(i)}-\boldsymbol{c}\right\|_2^2 \qquad (2.38a) \\ \boldsymbol{y}^{(i+1)} \coloneqq & \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \left\langle \nabla h\left(\boldsymbol{y}^{(i)}\right), \boldsymbol{y}\right\rangle + \rho_y \left\|\boldsymbol{y}-\boldsymbol{y}^{(i)}\right\|_2^2 \\ & + \left\langle \boldsymbol{\lambda}^{(i)}, \boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}\right\rangle + \rho \left\|\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}-\boldsymbol{c}\right\|_2^2 \qquad (2.38b) \\ \boldsymbol{\lambda}^{(i+1)} \coloneqq & \boldsymbol{\lambda}^{(i)} + \rho\left(\boldsymbol{A}\boldsymbol{x}^{(i+1)}+\boldsymbol{B}\boldsymbol{y}^{(i+1)}-\boldsymbol{c}\right). \qquad (2.38c) \end{aligned}$$

2.5.2.2 A Three-Operator ADMM Algorithm

In this section, we briefly introduce our recently proposed TOP-ADMM algorithm [KBG+21, KBF+22], which evolves and generalizes the popular two-operator or two-block (consensus) ADMM algorithm [BPC+11, PBPB13, CP11b], yet keeps the simplicity of traditional two-operator consensus ADMM.

Consider an optimization problem

$$\underset{\{\boldsymbol{x}_m\in\mathbb{C}^n\}, \boldsymbol{y}\in\mathbb{C}^n}{\text{minimize}} \quad \sum_{m=1}^{M} f_m(\boldsymbol{x}_m)+g(\boldsymbol{y})+\beta h(\boldsymbol{y}) \tag{2.39a}$$

$$\text{subject to} \quad \boldsymbol{x}_m - \boldsymbol{y} = \boldsymbol{0}, \forall m=1,\ldots,M, \tag{2.39b}$$

where $\{f_m\}$ and g are closed, convex, and proper functions; h is an L-smooth function; $\beta \in \mathbb{R}_{\geq 0}$.

Our proposed TOP-ADMM algorithm is an iterative scheme that solves (2.39a), which

is summarized below.

$$\boldsymbol{x}_m^{(i+1)} = \arg\min_{\boldsymbol{x}_m} f_m\left(\boldsymbol{x}_m\right) + \rho \left\| \boldsymbol{x}_m - \boldsymbol{y}^{(i)} + \frac{\boldsymbol{\lambda}_m^{(i)}}{\rho} \right\|_2^2 \quad \forall m = 1, \ldots, M \tag{2.40a}$$

$$\boldsymbol{y}^{(i+1)} = \arg\min_{\boldsymbol{y}} g\left(\boldsymbol{y}\right) + \sum_{m=1}^{M} \rho \left\| \boldsymbol{x}_m^{(i+1)} - \boldsymbol{y} - \tau \nabla h\left(\boldsymbol{y}^{(i)}\right) + \frac{\boldsymbol{\lambda}_m^{(i)}}{\rho} \right\|_2^2 \tag{2.40b}$$

$$\boldsymbol{\lambda}_m^{(i+1)} = \boldsymbol{\lambda}_m^{(i)} + \gamma\rho \left(\boldsymbol{x}_m^{(i+1)} - \boldsymbol{y}^{(i+1)} \right) \quad \forall m = 1, \ldots, M, \tag{2.40c}$$

where step-size $\tau \coloneqq \beta/(\rho M) \in \mathbb{R}_{\geq 0}$, $\gamma \in (0, 2)$, and a penalty parameter $\rho \in \mathbb{R}_{>0}$. The convergence proof of TOP-ADMM is established in Chapter D, *i.e.*, the generated sequences $\left(\left\{ \boldsymbol{x}_m^{(i)} \right\}, \boldsymbol{y}^{(i)}, \left\{ \boldsymbol{\lambda}_m^{(i)} \right\} \right)$ converge to a KKT point under mild conditions.

2.6 More Than Three Operators Splitting

The techniques seem to be scarce in the literature that can perform operator/proximal splitting for more than three operators such as function-linearized proximal ADMM (FLiP-ADMM) [RY22], majorized ADMM [LST16], multi-block ADMM, see, *e.g.*, [RY22] and references therein. Nevertheless, as discussed previously, for more than three operators problem one can resort to problem reformulation in product space [Pie84] such that three-/two-operator splitting methods can be employed block wise.

2.6.1 Four-Operator-Based Function-Linearized Proximal ADMM

Consider an optimization problem of the following form.

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{C}^n, \boldsymbol{y} \in \mathbb{C}^m}{\text{minimize}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{y}) + h(\boldsymbol{y}) + r(\boldsymbol{y}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{c}, \end{aligned} \tag{2.41}$$

where $\boldsymbol{A} \in \mathbb{C}^{p \times n}$, $\boldsymbol{B} \in \mathbb{C}^{p \times m}$, and $\boldsymbol{c} \in \mathbb{C}^p$. Let f and g be closed, convex, and proper functions; and h and r be L-smooth functions.

The function-linearized proximal ADMM [RY22, Chapter 8], referred to as FLiP-ADMM, has following updates.

$$\begin{aligned} \boldsymbol{x}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{x} \in \mathbb{C}^n} & f\left(\boldsymbol{x}\right) + \left\langle \nabla h\left(\boldsymbol{x}^{(i)}\right) + \boldsymbol{A}^{\mathrm{H}} \boldsymbol{\lambda}^{(i)}, \boldsymbol{x} \right\rangle \\ & + \rho \left\| \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(i)} - \boldsymbol{c} \right\|_2^2 + \left\| \boldsymbol{x} - \boldsymbol{x}^{(i)} \right\|_{\boldsymbol{Q}_x}^2 \end{aligned} \tag{2.42a}$$

$$\begin{aligned} \boldsymbol{y}^{(i+1)} \coloneqq \arg\min_{\boldsymbol{y} \in \mathbb{C}^m} & g\left(\boldsymbol{y}\right) + \left\langle \nabla r\left(\boldsymbol{y}^{(i)}\right) + \boldsymbol{B}^{\mathrm{H}} \boldsymbol{\lambda}^{(i)}, \boldsymbol{y} \right\rangle \\ & + \rho \left\| \boldsymbol{A}\boldsymbol{x}^{(i+1)} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{c} \right\|_2^2 + \left\| \boldsymbol{y} - \boldsymbol{y}^{(i)} \right\|_{\boldsymbol{Q}_y}^2 \end{aligned} \tag{2.42b}$$

$$\boldsymbol{\lambda}^{(i+1)} \coloneqq \boldsymbol{\lambda}^{(i)} + \gamma\rho \left(\boldsymbol{A}\boldsymbol{x}^{(i+1)} + \boldsymbol{B}\boldsymbol{y}^{(i+1)} - \boldsymbol{c} \right), \tag{2.42c}$$

where $\rho \in \mathbb{R}_{>0}$, $\gamma \in (0, 2)$, $\boldsymbol{Q}_x \in \mathbb{C}^{n\times n}$, and $\boldsymbol{Q}_y \in \mathbb{C}^{m\times m}$ are positive definite matrices.

2.6.2 Multi-Block ADMM

There has been some works to extend from two-block (and two-operator/-function) ADMM to multi-block (and multi-operator/-function) ADMM [HY12, CHYY16, CHY17] to solve problems of the form

$$\underset{\{\boldsymbol{x}_m \in \mathcal{X}\}}{\text{minimize}} \quad \sum_{m=1}^{M} f_m(\boldsymbol{x}_m) \tag{2.43a}$$

$$\text{subject to} \quad \sum_{m=1}^{M} \boldsymbol{A}_m \boldsymbol{x}_m = \boldsymbol{b} \tag{2.43b}$$

for some given bounded linear operators $\{\boldsymbol{A}_m\}$ and $\boldsymbol{b}$. However, the authors in [CHYY16] have shown that the direct extension of ADMM with sequential updates may not converge without additional modifications/assumptions [HY12, CHY17].

References

[ACT20] F. J. Aragón Artacho, R. Campoy, and M. K. Tam. The Douglas–Rachford algorithm for convex and nonconvex feasibility problems. *Mathematical Methods of Operations Research*, 91(2):201–240, April 2020.

[BBC20] S. Banert, R. I. Boţ, and E. R. Csetnek. Fixing and extending some recent results on the ADMM algorithm. *Numerical Algorithms*, pages 1–23, May 2020.

[BC11] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer New York, New York, NY, 2011.

[Bec17] A. Beck. *First-order methods in optimization*. SIAM, 2017.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[Bre16] L. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.*, 7(3):200–217, January 2016.

[BST14] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1–2):459–494, 2014.

[BV04] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[CHY17] X. Cai, D. Han, and X. Yuan. The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. *Comput. Optim. Appl.*, 66(1):39–73, November 2017.

[CHYY16] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1):57–79, January 2016.

[CKCH21] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms: A tour of recent advances, with new twists! *arXiv:1912.00137*, December 2021.

[CLL17] C.-Y. Chi, W.-C. Li, and C.-H. Lin. *Convex optimization for signal processing and communications: from fundamentals to applications*. CRC Press, 1st edition, 2017.

[CO12] A. I. Chen and A. Ozdaglar. A fast distributed proximal-gradient method. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 601–608, 2012.

[Com04] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, October 2004.

[CP11a] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1):120–145, 2011.

[CP11b] 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, New York, NY, 2011.

[CP21] P. L. Combettes and J.-C. Pesquet. Fixed Point Strategies in Data Science. *IEEE Trans. Signal Process.*, pages 1–1, 2021.

[CT93a] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.*, 3(3):538–543, 1993.

[CT93b] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3:538–543, 1993.

[DLPY17] W. Deng, M. Lai, Z. Peng, and W. Yin. Parallel multi-block ADMM with $o(1/k)$ convergence. *J. Sci. Comput.*, 71(2):712–736, May 2017.

[DR56] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two or three space variables. *Trans. Amer. Math. Soc.*, 82(2):421–439, July 1956.

[DY17] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*, 25(4):829–858, December 2017.

[EB92a] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(1–3):293–318, April 1992.

[EB92b] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1–3):293–318, April 1992.

[FG83] M. Fortin and R. Glowinksi. *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems*. North-Holland, Amsterdam, The Netherlands: Elsevier Science Publishers B. V., 1983.

[GM75] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.

[GOY16] R. Glowinski, S. J. Osher, and W. Yin. *Splitting methods in communication, imaging, science, and engineering*. Scientific Computation. Springer International Publishing, Cham, Switzerland, 2016.

[HA14] H. Wang and A. Banerjee. Bregman alternating direction method of multipliers. In *Proc. Adv. Neural Inf. Process. Syst.*, pages 2816–2824, 2014.

[HMY16] B. He, F. Ma, and X. Yuan. Convergence Study on the Symmetric Version of ADMM with Larger Step Sizes. *SIAM J. Imaging Sciences*, 9(3):1467–1501, 2016.

[HY12] D. Han and X. Yuan. A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.*, 155(1):227–238, February 2012.

[KBF+22] S. Kant, M. Bengtsson, G. Fodor, B. Göransson, and C. Fischione. EVM mitigation with PAPR and ACLR constraints in large-scale MIMO-OFDM using TOP-ADMM. *accepted to IEEE Trans. Wireless Commun.*, 2022.

[KBG+21] S. Kant, M. Bengtsson, B. Göransson, G. Fodor, and C. Fischione. Efficient optimization for large-scale MIMO-OFDM spectral precoding. *IEEE Trans. Wireless Commun.*, 20(9):5496–5513, September 2021.

[KP15] N. Komodakis and J. Pesquet. Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.*, 32(6):31–54, October 2015.

[LLF20] Z. Lin, H. Li, and C. Fang. *Accelerated optimization for machine learning: First-order algorithms*. Springer Nature Singapore Pte Ltd., 2020.

[LST16] M. Li, D. Sun, and K.-C. Toh. A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.*, 26(2):922–950, 2016.

[LSTS20] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process. Mag.*, 37(3):50–60, May 2020.

[LWW31] L. Li, X. Wang, and G. Wang. Alternating direction method of multipliers for separable convex optimization of real functions in complex variables. *Math. Probl. Eng.*, 2015:1–14, November 2015, Art. no. 104531.

[Ma20] F. Ma. Convergence study on the proximal alternating direction method with larger step size. *Numer. Algor.*, 85(2):399–425, 2020.

[PBPB13] N. Parikh, S. Boyd, N. Parikh, and S. Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):123–231, 2013.

[Pie84] G. Pierra. Decomposition through formalization in a product space. *Mathematical Programming*, 28(1):96–115, January 1984.

[Rag19] H. Raguet. A note on the Forward-Douglas–Rachford splitting for monotone inclusion and convex optimization. *Optim. Lett.*, 13:717–740, May 2019.

[RFP13] H. Raguet, J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, January 2013.

[RY19] E. K. Ryu and W. Yin. Proximal-proximal-gradient method. *J. Comput. Math.*, 37(6):778–812, 2019.

[RY22] E. K. Ryu and W. Yin. *Large-scale convex optimization via monotone operators*. Cambridge University Press (to be published), 2022.

[SNW11] S. Sra, S. Nowozin, and S. J. Wright. *Optimization for machine learning*. The MIT Press, 2011.

[SSS+16] G. Stathopoulos, H. Shukla, A. Szűcs, Y. Pu, and C. N. Jones. Operator splitting methods in control. *Foundations and Trends® in Systems and Control*, 3(3):249–362, 2016.

[Yan18] M. Yan. A New Primal-Dual Algorithm for Minimizing the Sum of Three Functions with a Linear Operator. *J. Sci. Comput.*, 76:1698–1717, 2018.

Part II

Included Papers