



<http://www.diva-portal.org>

This is the published version of a paper published in *IEEE Transactions on Wireless Communications*.

Citation for the original published paper (version of record):

Kant, S., Bengtsson, M., Göransson, B., Fodor, G., Fischione, C. (2021)
Efficient Optimization for Large-Scale MIMO-OFDM Spectral Precoding
IEEE Transactions on Wireless Communications

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-294584>

good OOB suppression and low complexity as it is based on data-independent optimization problem. Unfortunately, the employment of AIC-type (or GSS) OOB reduction methods is impractical for 5G use cases to comply with NR specifications.

Spectral Precoding-type OOB Reduction Schemes: Spectral precoding is one of the promising contenders for OOB reduction that meets the stringent spectrum efficiency because it only exploits the active data subcarriers without sacrificing the cyclic prefix. It spectrally precodes the data symbols before OFDM modulation [12], [17], [20], which reduces the OOB without increasing the delay/time dispersion or penalizing the cyclic prefix of the transmitted signal. In [12], the authors propose a least-squares notching spectral precoder scheme that nulls the OOB at predefined frequency points while minimizing the Euclidean distance between the precoded and the original data vector. The notching spectral precoder (NSP) method renders high TxEVM, discernibly at the band edge subcarriers, leading to increased block error rate at the receiver. Recently, a mask-compliant spectral precoder (MSP) scheme was proposed that complies with a predefined spectral emission mask [14], [16], [18], [19], rather than nulling the undesired OOB at predefined frequency points, offering improved performance at the receiver due to reduced TxEVM.

A. Spectral Precoding in MIMO-OFDM

In [19], a mask-compliant and wideband/frequency-selective TxEVM-constrained spectral precoding for multiple-input multiple-output (MIMO)-OFDM systems was proposed encompassing computationally efficient algorithms, which shows the inherent trade-off between TxEVM and OOB. In [24], several linear receivers are investigated when NSP is employed in MIMO system.

In [20, Paper F], the authors extend the single-input single-output (SISO)-OFDM spectral precoding to massive multi-user MIMO-OFDM for a joint spatial and notching spectral precoder, which exploits full downlink channel state information at the transmitter utilizing channel reciprocity in time division duplexing, to improve the in-band performance at the receiver. In particular, the approach in this prior art is to perform a joint spectral and spatial precoding of a MIMO-OFDM symbol that not only notches the spectrum at well-chosen frequency points, *i.e.*, achieved by NSP, but also simultaneously nulls the multi-user or inter-spatial-layer interference, *i.e.*, due to zero forcing (ZF) spatial precoding. Although the proposed scheme in [20, Paper F] has a closed-form solution—see Appendix A, it has poor in-band performance notably due to the utilization of the NSP constraint rendering high amount of TxEVM—see from Fig. 8 to Fig. 9. Typically, the base station has limited transmit power budget per antenna branch. Therefore, high distortion power consumes relatively high proportion of the total transmit power budget, which consequently reduces the power of the useful transmit signal due to rescaling the composite transmit signal to maintain the power budget. Further, their problem formulation employs the ZF spatial precoder. In practice, different types of linear and/or non-linear spatial precoding schemes can be employed, *e.g.*, minimum mean squared error

precoding [25]. Thus, prior art is quite restrictive and yields poor in-band performance at the receiver.

Hence, we seek a computationally-affordable mask-compliant spectral precoding that not only improves the performance at the receiver(s) in single-user MIMO (SU-MIMO) or multi-user MIMO (MU-MIMO) OFDM systems, including massive ones, but also meets the OOB performance at the transmitter. In contrast to [20, Paper F], we endeavour “one-size-fits-all” precoding, *i.e.*, unaware of any (non-linear) spatial precoding employed at the transmitter, while complying with the mask and mitigating the resulting distortion substantially at the receiver.

B. Large-Scale Optimization for Signal Processing in Communications

Many problems for signal processing in communications can be posed as convex optimization problems. Due to the trend of large-scale (“big data”) problems, first-order optimization problems are becoming extremely popular thanks to their low-computational complexity.

Proximal splitting methods, also referred to as operator splitting methods [26]–[29], and in the recent decade, alternating direction method of multipliers (ADMM) [30] has emerged as one class of such methods, which has enjoyed its renaissance due to its wide applicability in large-scale machine learning problems by breaking down a massive problem into easy-to-solve subproblems.

ADMM can be described in a distributed fashion to solve the appropriate optimization problems, which is also known as consensus ADMM—see, *e.g.*, [29, Chapter 7]. This classical consensus ADMM can be seen as a parallel/consensus two-block or two-operator comprising, loosely-speaking, a sum of two functions/blocks/operators that converges for convex problems under mild conditions [29], [30]. Additionally, there is an exhaustive literature on the ADMM and its variants, see, *e.g.*, [27], [28], [31]–[34].

The operator splitting with more than two composite terms in the objective has been an open research problem without resorting to problem reformulations or product space reformulations [33], [35]—but unfortunately those tricks are not straightforward and generally may be slow to converge or not be feasible for some problems. Very recently, the authors in [36] made a breakthrough by proposing an innovative three-operator splitting scheme to a long-standing problem, which requires at least one of the functions/operators to be cocoercive, *i.e.*, a function is differentiable with Lipschitz continuous gradient. Furthermore, classic two-operator Douglas-Rachford splitting² [30] is a special case of their three-operator splitting. However, our proposed problem formulation, cf. Section III, unfortunately, cannot directly employ their three-operator splitting as our problem also requires the parallel or consensus formulation of such three-operator. Therefore, in this paper, we introduce a new three-operator consensus ADMM scheme, referred to as three-operator ADMM (TOP-ADMM). Further, our proposed TOP-ADMM recovers well-known two-operator consensus ADMM as a special case.

²Classical two-operator ADMM is related to Douglas-Rachford splitting (DRS) when DRS is applied to the dual problem [30].

C. Contribution of the Paper

The core idea of the proposed methods are to not only perform the OOB power reduction at the transmit antenna ports but also to mitigate the incurred signal distortion or TxEVM seen at the receiver without additional receiver complexity—independently of the chosen beamforming/precoding technique at the transmitter. More specifically, the idea is to exploit the excess degrees-of-freedom inherent in large-scale MIMO-OFDM systems equipped with a large number of antenna branches such that the distortion rendered by the spectral precoding is mitigated at the receiver while simultaneously complying with the mask requirement at the transmitter. Our key contributions are as follows:

- We firstly formulate two new optimization problems for the TxEVM-mitigated and mask-compliant spectral precoding scheme for (massive/large-scale) MIMO-OFDM systems operating in time-division duplexing mode. The second proposal is more robust against the channel uncertainty than the first proposal at the cost of increased computational complexity.
- For the ease of implementation in the realistic radio system, in contrast to the prior art [20, Paper F], we separate the signal spatial precoding or beamforming from the spectral precoding. Furthermore, our spatial shaping constraint to the TxEVM is oblivious of any (possibly non-linear) precoding for MU-MIMO or SU-MIMO.
- We propose and establish a novel and yet simple TOP-ADMM algorithm with convergence guarantees under some given conditions—please see Theorem 1—which reduces to the classical two-operator consensus ADMM as a special case. Subsequently, we develop a “computational-friendly” TOP-ADMM-based algorithm that breaks down the proposed large-scale/massive optimization problems into smaller subproblems where each subproblem entails an efficient solution.
- We finally present extensive simulations that illustrate the superior performance of the proposed methods using a 5G NR (release 15 compliant) link-level simulator [18], [19].

The structure of the remaining paper is as follows. The next section briefly defines the system model including spatial and spectral precoding with distortion/TxEVM modelling and useful performance metrics for the algorithm design. Next, Section III formulates the TxEVM-mitigated and mask-compliant problem. Section IV firstly describes the novel TOP-ADMM algorithmic framework and subsequently develops an TOP-ADMM-based algorithm to solve the proposed optimization problem. Subsequently, we explicate the shortcomings of classical two-operator ADMM in tackling the proposed three-operator problems. Section V exhibits an extensive set of simulation results and Section VI concludes with a summary. Moreover, Appendix A briefly describes the prior art [20, Paper F] used in Section V for benchmarking purpose. Appendix B analyzes the basic convergence of TOP-ADMM scheme using tools from [29, Section 3.2] [37]. Appendix C presents background on proximal, projection operators, and Lipschitz definition.

D. Notation

Let the set of complex and real numbers be denoted by \mathbb{C} and \mathbb{R} , respectively. $\Re\{x\}$ denotes the real part of a complex number x . The i -th element of a boldface vector $\mathbf{a} \in \mathbb{C}^{m \times 1}$ is denoted by $a[i] \in \mathbb{C}$, and the element in the i -th row and j -th column of the boldface matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ is denoted by $A[i, j] \in \mathbb{C}$. The i -th row and j -th column vector of a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ are represented as $\mathbf{A}[i, :] \in \mathbb{C}^{1 \times n}$ and $\mathbf{A}[:, j] \in \mathbb{C}^{m \times 1}$, respectively. An i -th higher order vector and matrix are denoted as $\mathbf{x}[i] \in \mathbb{C}^{m \times 1}$ or $\mathbf{x}_i \in \mathbb{C}^{m \times 1}$ and $\mathbf{X}[i] \in \mathbb{C}^{m \times n}$. We form a matrix by stacking the set of higher order vectors $\{\mathbf{a}[n] \in \mathbb{C}^{M \times 1}\}_{n=1}^N$ and $\{\mathbf{b}[m] \in \mathbb{C}^{1 \times N}\}_{m=1}^M$ column-wise and row-wise as $\mathbb{C}^{M \times N} \ni \mathbf{A} = [\mathbf{a}[1], \dots, \mathbf{a}[N]]$ and $\mathbb{C}^{M \times N} \ni \mathbf{B} = [\mathbf{b}[1]; \dots; \mathbf{b}[M]]$, respectively. The transpose and conjugate transpose of a vector or matrix are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively. The complex conjugate is represented by $(\cdot)^*$. All ones column vector and matrix of length K and $K \times N$ is expressed as $\mathbf{1}_K$ and $\mathbf{1}_{K \times N}$, respectively. \mathbf{I}_K , $\mathbb{E}\{\cdot\}$, and $(\cdot)^{(i)}$ denote $K \times K$ identity matrix, the expectation operator, and an i -th iterative update, respectively. A weighted Frobenius norm is defined as $\|\mathbf{A}\|_{\mathbf{W}}^2 := \text{Tr}(\mathbf{A}^H \mathbf{W} \mathbf{A})$, where Tr is a trace operator, \mathbf{W} is a positive definite matrix, and $:=$ denotes definition. $\mathbf{D} = \text{Diag}(\mathbf{d})$ creates a diagonal matrix by stacking elements of \mathbf{d} along the main diagonal.

II. SYSTEM MODEL AND PRELIMINARIES

In this section, we introduce the downlink massive MIMO-OFDM system model including spatial precoding. In the sequel, we briefly describe the spectrally precoded transmit and receive signal model followed by the performance metrics useful for the spectral precoding design.

A. Spatial Precoding

We consider an OFDM-based single-user MIMO downlink, where the base station is equipped with N_T transmit (Tx) antennas, and the user equipment (UE) is equipped with N_R receive (Rx) antennas as depicted in Fig. 1. Additionally, we reckon a spatial multiplexing scheme with $N_L \leq \min\{N_T, N_R\}$ spatial layers. Particularly, the assumption is that the base station is equipped with a relatively large number of antennas such that $N_T \geq N_R$.

A (generalized) spatially precoded symbol vector $\mathbf{x}[k]$ at the k -th subcarrier for a given OFDM symbol can be formed by $\mathbb{C}^{N_T \times 1} \ni \mathbf{x}[k] = \mathcal{P}(\mathbf{s}[k])$, where $\mathbf{s}[k] \in \mathcal{S}^{N_L \times 1}$ belongs to a complex-valued finite-alphabet set \mathcal{S} , *e.g.*, corresponding to a 2^Q -quadrature amplitude modulation (QAM) constellation with $Q \in \{2, 4, 6, 8\}$. The spatial (possibly nonlinear) precoding $\mathcal{P}: \mathcal{S}^{N_L \times 1} \rightarrow \mathbb{C}^{N_T \times 1}$ maps the non-precoded symbol vector $\mathbf{s}[k]$ to spatially precoded symbol vector $\mathbf{x}[k] \in \mathbb{C}^{N_T \times 1}$.

We now introduce a spatially precoded data matrix $\mathbf{X} \in \mathbb{C}^{N_T \times N}$ by stacking $\mathbf{x}[k]$, for all $k = 1, \dots, N$ subcarriers, such that $\mathbf{X} := [\mathbf{x}[1], \dots, \mathbf{x}[N]] := [\mathbf{d}_1^T; \dots; \mathbf{d}_{N_T}^T]$, where we define a spatially precoded vector $\mathbb{C}^{N \times 1} \ni \mathbf{d}_j := (\mathbf{X}[j, :])^T$ for a j -th transmit antenna.

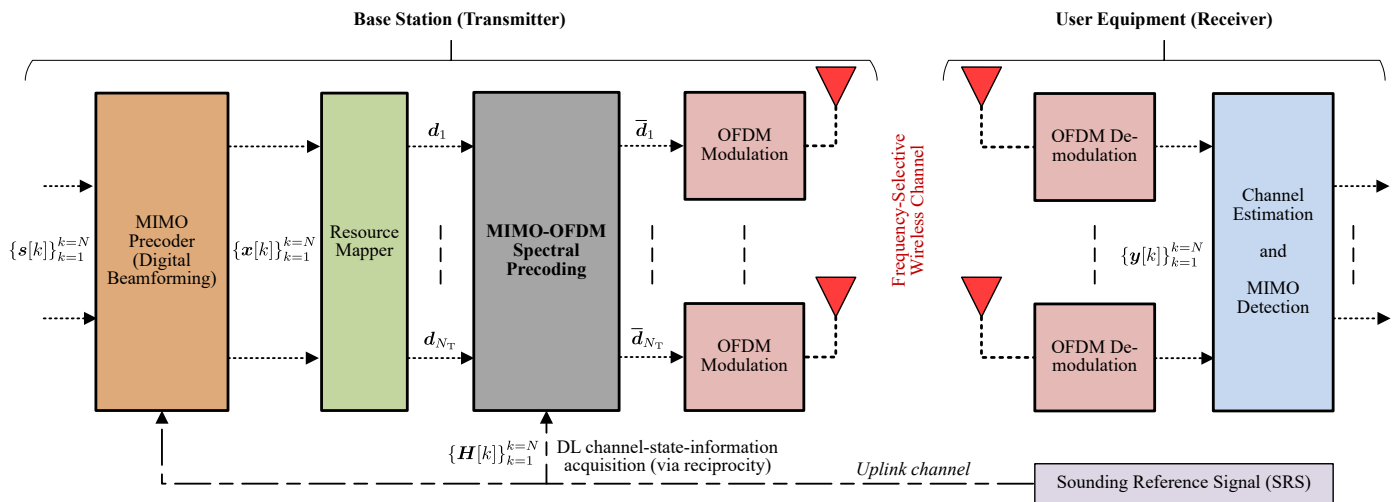


Fig. 1: Simplified block diagram of a single-user massive MIMO-OFDM transceiver with spectral precoding.

B. Transmit Spectrally Precoded Signal Model

Based on, *e.g.*, [14], [18], [19], we can model a *spectrally precoded* data matrix $\overline{\mathbf{X}}$ as

$$\mathbb{C}^{N_T \times N} \ni \overline{\mathbf{X}} := G(\mathbf{X}) := \begin{bmatrix} \overline{\mathbf{d}}_1^T; \dots; \overline{\mathbf{d}}_{N_T}^T \end{bmatrix}, \quad (1)$$

where a (possibly non-linear) function $G : \mathbb{C}^{N_T \times N} \rightarrow \mathbb{C}^{N_T \times N}$ manipulates the data symbols vector for each transmit antenna in the frequency-domain to comply with the mask constraints. Consequently, we define a column-vector $\bar{\mathbf{d}}_j$ of spectrally (and spatially) precoded vector corresponding to j -th transmit antenna branch as $\mathbb{C}^{N \times 1} \ni \bar{\mathbf{d}}_j := (\bar{\mathbf{X}}[j, :])^T$. To this end, we introduce a perturbation model to (1) such that the spectrally precoded symbol vector $\bar{\mathbf{X}}[:, k] \in \mathbb{C}^{N_T \times 1}$ at a given k -th subcarrier can be expressed as,

$$\overline{\mathbf{X}}[:, k] = \alpha[k] \mathbf{X}[:, k] + \underbrace{\boldsymbol{\epsilon}[k]}_{\text{TxEVM}} \approx \mathbf{X}[:, k] + \boldsymbol{\epsilon}[k], \quad (2)$$

where $\alpha[k] \in \mathbb{C}$ is a deterministic scalar and $\epsilon[k]$ is (effective) TxEVM³ or distortion that is uncorrelated with $\mathbf{x}[k]$ but statistically dependent⁴. However, appealing to the works in [39]–[42], $\epsilon[k]$ is correlated across antennas and consequently beamformed in the similar direction as the signal $\mathbf{x}[k]$, for instance, depending on the ratio of transmit spatial layers and the transmit antenna branches—*i.e.*, rank of the channel, frequency-granularity of the spatial precoding—*i.e.*, channel coherence bandwidth. We show in the numerical section later that the TxEVM emanating from spectral precoding is also beamformed in the similar direction as the signal, see, *e.g.*, Fig. 2(a).

It is worth to point out that although the scaling factor $\alpha[k]$ may vary across subcarriers, on average, the scaling is assumed to be real-valued and nearly unity—our extensive simulations, cf. Section V, show that the approximate additive TxEVM

³We use distortion and TxEVM interchangeably, and overload TxEVM as a performance metric.

⁴This frequency-domain TxEVM model can be employed for NSP [12], [20]. Moreover, one could invoke Bussgang decomposition [38] to other nonlinear devices [39]–[42], *e.g.*, power amplifier, resulting a similar TxEVM model.

model in (2) is accurate and ignoring $\alpha[k]$ has nearly negligible impact on the key performance metrics.

The spectrally precoded symbol matrix $\bar{\mathbf{X}}$ is an input to OFDM modulation, see Fig. 1, where each row of the matrix, *i.e.*, $\{\bar{d}_j\}$, is converted from frequency-domain to time-domain via inverse discrete Fourier transform followed by cyclic prefix addition for each transmit antenna branch. In practice, after OFDM modulation appropriate digital and analog processing of the signal is required prior to the transmission of the signals through antennas.

C. Received Spectrally Precoded Signal Model

The received symbol vector $\mathbf{y}[k] \in \mathbb{C}^{N_{\text{R}} \times 1}$ after OFDM demodulation, *i.e.*, after cyclic prefix removal followed by a discrete Fourier transformation (assuming perfect time and frequency synchronization, and without any inter-symbol interference), for the transmit model with an additive distortion (2), at a given k -th subcarrier can be shown as,

$$\begin{aligned} \mathbf{y}[k] &= \mathbf{H}[k] \overline{\mathbf{X}}[:, k] + \mathbf{n}[k] \\ &\approx \mathbf{H}[k] \mathbf{X}[:, k] + \underbrace{\mathbf{H}[k] \boldsymbol{\epsilon}[k]}_{\text{Received TxEVM}} + \mathbf{n}[k]. \end{aligned} \quad (3)$$

The vector $\mathbf{n}[k]$ corresponds to noise and other inter-cell interferences which is modelled as a zero-mean complex symmetric white Gaussian noise with $\mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{N_R})$, where N_0 is the noise variance at the receiver. It can be noted that the same received signal model applies to both cases with a single receiving terminal as well as multiple receiving terminals, *i.e.*, MU-MIMO. The received TxEVM at k -th subcarrier is mathematically expressed by $\mathbf{H}[k]\epsilon[k]$, which unfortunately correlates the total noise spatially at the receiver since the effective noise covariance matrix \mathbf{R} seen at the user side can be as following—ignoring the k -th index for brevity $\mathbf{R} = \mathbf{H}\mathbf{R}_{\epsilon\epsilon}\mathbf{H}^H + N_0\mathbf{I}_{N_R}$, where $\mathbf{R}_{\epsilon\epsilon} = \mathbb{E}\{\epsilon\epsilon^H\}$ is the TxEVM covariance. Thus, in general, the impact of TxEVM seen at the receiver, not only stemming from spectral precoding but also from other non-linear devices, *e.g.*, non-linear power amplifier, is deleterious both for (massive) SU-MIMO and MU-MIMO performance—see, *e.g.*, [39], [40], [43].

D. Performance Metrics for the Spectral Precoding Design

We develop a spectral precoding scheme characterized by two figure-of-merits, *i.e.*, OOB and in-band distortions.

1) *Out-of-Band Emissions*: The OOB is typically quantified in terms of the operating band unwanted emissions, *i.e.*, mask, and (conducted) adjacent channel leakage ratio (ACLR), whose definitions are given below.

Definition 1 (Operating band unwanted emissions [3, Section 6.6.4], referred to as mask). *Operating band unwanted emissions refer to unwanted emissions that are immediately outside the base station channel bandwidth resulting from the modulation process and non-linearity in the transmitter but excluding spurious emissions.*

The unwanted OOB due to the OFDM frequency-domain signal \mathbf{d}_j at M considered discrete frequency points $\mathbb{R}^{M \times 1} \ni \boldsymbol{\nu} = [\nu_1; \dots; \nu_M]$ can be described by $\mathbf{p}(\boldsymbol{\nu}) = \mathbf{A}\mathbf{d}_j$. We now define $\mathbf{A}[m, :] := \mathbf{a}(\nu_m)^T \in \mathbb{C}^{1 \times N}$, where $\mathbf{A}[m, k] := a(\nu_m, k)$ can be derived in discrete form as [12]:

$$a(\nu_m, k) = (1/\sqrt{N}) \exp\left(j\pi \frac{(\nu_m - k)}{N} (N_{\text{CP}} - N + 1)\right) \cdot \frac{\sin\left(\pi \frac{(\nu_m - k)}{N} (N + N_{\text{CP}})\right)}{\sin\left(\pi \frac{(\nu_m - k)}{N}\right)},$$

where N_{CP} corresponds to cyclic prefix length in samples.

It is worth highlighting that a set of well-chosen frequency points $\boldsymbol{\nu}$ would be sufficient to suppress OOB and consequently meet the regulating mask [12]. Therefore, a mask-compliant spectral precoding should ensure that the resulting OOB power, at given $\boldsymbol{\nu}$ discrete frequency points, is below some target mask $\mathbb{R}^{M \times 1} \ni \boldsymbol{\gamma} = [\gamma_1; \dots; \gamma_M]$, *i.e.*,

$$|\bar{\mathbf{p}}(\boldsymbol{\nu})|^2 = |\mathbf{A}\bar{\mathbf{d}}_j|^2 \preceq \boldsymbol{\gamma} \preceq \text{NR mask}, \quad (4)$$

to meet the overall NR stipulated mask, cf. Definition 1, and the inequality \preceq is element-wise.

Definition 2 (Adjacent channel leakage ratio (ACLR) [3, Section 6.6.3]). *ACLR is the ratio of the filtered mean power centred on the assigned channel frequency to the filtered mean power centred on an adjacent channel frequency. The (worst-case) ACLR can mathematically be expressed as*

$$\text{ACLR} := \frac{\int_{-\text{BW}/2}^{\text{BW}/2} S_{\text{desiredchannel}}(f) df}{\max \left\{ \int_{-\text{BW}/2}^{-\text{BW}/2} S_{\text{leftchannel}}(f) df, \int_{\text{BW}/2}^{\text{BW}/2} S_{\text{rightchannel}}(f) df \right\}},$$

where $S_{\text{desiredchannel}}$ is the power spectral density (PSD) in the desired carrier having BW bandwidth including the guard band; and similarly, $S_{\text{leftchannel}}$ and $S_{\text{rightchannel}}$ correspond to the PSD on the left and the right side of the desired carrier having same bandwidth BW as the desired carrier, respectively.

We would like to accentuate that we do not directly use ACLR for the spectral precoding design, but rather we use the OOB power at the considered discrete frequency points.

2) *In-Band Distortion*: The considered in-band distortion for the spectral precoding design is averaged TxEVM, which can be quantified as a loss in the demodulated signal quality, described mathematically per j -th transmit antenna as

$$\text{TxEVM}_j := \text{EVM}_j = \frac{\mathbb{E}\{\|\mathbf{d}_j - \bar{\mathbf{d}}_j\|_2^2\}}{\mathbb{E}\{\|\mathbf{d}_j\|_2^2\}}.$$

III. PROPOSED TXEVM-MITIGATION AND MASK-COMPLIANT SPECTRAL PRECODING IN MASSIVE MIMO-OFDM

In this section, we formulate a novel optimization problem that is aware of the channel state information and is capable to cater the massive spectral precoding problem. More specifically, we propose mask-compliant spectral precoding and show that such a convex problem formulation does not have a closed-form solution in general. Moreover, the proposed method is agnostic to any precoding or beamforming technique employed at the base station and yet mitigates the TxEVM, incurred due to spectral precoding, at the receiver(s).

A. Problem Formulations

Here, we formulate two mask-compliant spectral precoding problems that are channel state information aware, which capitalize on the knowledge of the full downlink channel knowledge available at the base station transmitter. The first proposal is oblivious of channel uncertainty, whereas the second proposal is more robust to channel imperfections than the first proposal at the expense of higher computational complexity. In contrast to the minimization of the signal energy posed in the prior art joint NSP and ZF precoding, cf. (23) in Appendix A, we minimize the (weighted) TxEVM, since the transmit power budget is fixed in practical systems. Furthermore, we impose two constraints, namely, spatial and spectral constraints, as described subsequently.

In the first proposal, consider that the base station transmitters are endowed with downlink channel knowledge, then the frequency-selective nulled received TxEVM-constrained with mask-compliant spectral precoding optimization problem is formulated as:

$$(\text{P1}) \quad \underset{\bar{\mathbf{X}} \in \mathbb{C}^{N_{\text{T}} \times N}}{\text{minimize}} \quad \|\mathbf{X} - \bar{\mathbf{X}}\|_{\mathbf{\Pi}}^2 \quad (5a)$$

$$\text{subject to} \quad \mathbf{H}[k](\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k]) = \mathbf{0} \quad \forall k \in \mathcal{T} \quad (5b)$$

$$|\mathbf{A}\bar{\mathbf{X}}^T|^2 \preceq \mathbf{\Gamma} \quad (5c)$$

where $\mathbf{\Pi}$ is any given positive definite matrix and the target mask is $\mathbb{R}^{M \times N_{\text{T}}} \ni \mathbf{\Gamma}$ for generality, but practically the same target mask constraint $\mathbf{\Gamma}[:, j] = \gamma$ is employed for all the antennas. A typical choice of $\mathbf{\Pi}$ is a real-valued diagonal matrix—inspired by [15], where larger weight would imply giving more priority to the appropriate subcarrier, *i.e.*, less incurred TxEVM due to the spectral precoding on that subcarrier. For instance, such a weight matrix can be dependent on the selected modulation alphabet, rank of the channel, and the link quality.

The spectral constraint (5c) complies with the mask requirements, as described by (4) per antenna, which indirectly

meet the target ACLR requirements at the transmitter. We have defined a spatial constraint to the received distortion, which mitigates or nulls out the TxEVM seen at the receiver(s) expressed by (5b). In other words, invoking our perturbed TxEVM model in (2), *i.e.*, $\epsilon[k] \approx \bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k]$, the received TxEVM in (3) is nulls out after passing through the propagation channel $\mathbf{H}[k]$, *i.e.*, $\mathbf{H}[k]\epsilon[k] = \mathbf{0}$.

In the second approach, a robust version of the problem P1 is proposed. We firstly relax the received TxEVM equality constraint (5b) to an inequality constraint

$$\|\mathbf{H}[k](\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k])\|_2^2 := \|\mathbf{H}[k] \Delta \mathbf{X}[:, k]\|_2^2 \leq \varsigma^2[k], \quad (6)$$

where $\varsigma^2[k]$ is a user defined parameter that allows some amount of received distortion power at the receiver instead of nulling and $\Delta \mathbf{X}[:, k] = (\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k])$. In order to make constraint (6) channel error aware, let the channel error model be described by $\mathbf{H} := \hat{\mathbf{H}} + \Delta \mathbf{H}$, where \mathbf{H} , $\hat{\mathbf{H}}$, and $\Delta \mathbf{H}$ denote true, estimated, and error, respectively. For simplicity, we assume that $\Delta \mathbf{H}$ takes values from the bounded set $\{\|\Delta \mathbf{H}\|_F^2 \leq \sigma_{ce}^2\}$, where $\sigma_{ce}^2 > 0$ describes the channel uncertainty that is assumed to be known to the transmitter. Applying the triangle inequality followed by the Cauchy-Schwarz inequality to (6) and ignoring k -th index on the channel matrices, we have

$$\begin{aligned} & \left\| (\hat{\mathbf{H}} + \Delta \mathbf{H}) \Delta \mathbf{X}[:, k] \right\|_2^2 \\ & \leq \left\| \hat{\mathbf{H}} \Delta \mathbf{X}[:, k] \right\|_2^2 + \left\| \Delta \mathbf{H} \Delta \mathbf{X}[:, k] \right\|_2^2 \\ & \leq \left\| \hat{\mathbf{H}} \Delta \mathbf{X}[:, k] \right\|_2^2 + \sigma_{ce}^2 \left\| \Delta \mathbf{X}[:, k] \right\|_2^2 - \sigma_{ce}^2 \\ & = \Delta \mathbf{X}[:, k]^H \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \sigma_{ce}^2 \mathbf{I} \right) \Delta \mathbf{X}[:, k] - \sigma_{ce}^2. \end{aligned} \quad (7)$$

Using the above inequalities (6) and (7), we propose the following robust scheme.

$$\begin{aligned} \text{(P2)} \quad & \underset{\bar{\mathbf{X}} \in \mathbb{C}^{N_T \times N}}{\text{minimize}} \quad \left\| \mathbf{X} - \bar{\mathbf{X}} \right\|_{\Pi}^2 \\ & \text{subject to} \quad \left\| \mathbf{Q}^{\frac{1}{2}}[k] (\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k]) \right\|_2^2 \\ & \quad \leq \varsigma^2[k] + \sigma_{ce}^2[k] \quad \forall k \\ & \quad \left| \mathbf{A} \bar{\mathbf{X}}^T \right|^2 \preceq \Gamma, \end{aligned} \quad (8a)$$

where $\mathbf{Q}[k] := \left(\hat{\mathbf{H}}[k]^H \hat{\mathbf{H}}[k] + \sigma_{ce}^2 \mathbf{I} \right)$ and $(\cdot)^{\frac{1}{2}}$ denotes the square-root. Specifically, spatial constraint (8a) is more sturdy against channel knowledge imperfections than the constraint (5b).

The proposed problems P1 (5) and P2 (8) are convex because the cost function and the constraints are convex. An optimal solution to problem P1/P2 can be obtained via a general-purpose optimization solver, *e.g.*, CVX [44]. However, such general-purpose algorithms typically employ interior-point-based methods which have prohibitive complexity [45] approximately $\mathcal{O}(N^{4.5} N_T)$ —using results in [16]—for state-of-the-art base station radio hardware systems. Hence, we seek a “hardware-friendly” algorithm rendering sufficiently mod-

est/low accuracy that can be employed in a typical base station hardware to solve such proposed optimization problems.

B. Reformulation of the Proposed Problems P1 and P2

We now unify the problems P1 (5) and P2 (8) as the following unconstrained problem amenable to the proposed efficient algorithm described in the following section,

$$\underset{\bar{\mathbf{X}} \in \mathbb{C}^{N_T \times N}}{\text{minimize}} \quad f(\bar{\mathbf{X}}) + \delta_{\mathcal{S}}(\bar{\mathbf{X}}) + \sum_{m=1}^M \delta_{\mathcal{C}_m}(\bar{\mathbf{X}}). \quad (9)$$

The function $f(\bar{\mathbf{X}})$ reads

$$f(\bar{\mathbf{X}}) := \left\| \mathbf{X} - \bar{\mathbf{X}} \right\|_{\Pi}^2, \quad (10)$$

that is a total (weighted) squared TxEVM over all the transmit antennas.

The indicator functions $\delta_{\mathcal{C}_m}(\cdot)$ and $\delta_{\mathcal{S}}(\cdot)$ correspond to the respective convex sets, namely m -th mask constraint set \mathcal{C}_m

$$\mathcal{C}_m := \left\{ \bar{\mathbf{X}} : \bar{\mathbf{d}}_j^H \bar{\mathbf{A}}_m \bar{\mathbf{d}}_j - \gamma_m \leq 0; \forall j = 1, \dots, N_T \right\}, \quad (11)$$

and spatial constraint set \mathcal{S} represents either (5b) for P1

$$\mathcal{S} \leftarrow \mathcal{S}_{P1} := \left\{ \bar{\mathbf{X}} : \mathbf{H}[k] (\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k]) = \mathbf{0}; \forall k \in \mathcal{T} \right\} \quad (12)$$

or (8a) for P2

$$\begin{aligned} \mathcal{S} \leftarrow \mathcal{S}_{P2} \\ := \left\{ \bar{\mathbf{X}} : \left\| \mathbf{Q}^{\frac{1}{2}}[k] (\bar{\mathbf{X}}[:, k] - \mathbf{X}[:, k]) \right\|_2^2 \leq \varsigma^2[k] + \sigma_{ce}^2[k]; \forall k \in \mathcal{T} \right\}. \end{aligned} \quad (13)$$

Towards our goal, in the following section we develop a computationally-efficient first-order method to solve the problems P1 and P2.

IV. THREE-OPERATOR ADMM ALGORITHM

In this section, we firstly present a new and simple yet stark three-operator consensus ADMM, which we call TOP-ADMM, optimization algorithm that is built upon the popular two-operator consensus ADMM algorithm [27], [29], [31]. Subsequently, we utilize TOP-ADMM to solve the proposed channel state information aware mask-compliant and TxEVM-mitigated spectral precoding problem P1 and P2 for the (massive) MIMO-OFDM systems. Afterwards, we accentuate the potential shortcomings of the solutions to P1 and P2 using classical two-operator ADMM.

A. General Algorithmic Framework

In the sequel, we introduce a simple and novel parallelized TOP-ADMM algorithmic framework to tackle large-scale optimization problems such as (9).

Theorem 1 (TOP-ADMM algorithm). *Consider a problem*

$$\begin{aligned} & \underset{\{\bar{\mathbf{y}}_m \in \mathbb{C}^n, \bar{\mathbf{x}} \in \mathbb{C}^n\}}{\text{minimize}} \quad \sum_{m=1}^M \mathcal{F}_m(\bar{\mathbf{y}}_m) + \mathcal{G}(\bar{\mathbf{x}}) + \mathcal{H}(\bar{\mathbf{x}}) \\ & \text{subject to} \quad \bar{\mathbf{y}}_m = \bar{\mathbf{x}}, \end{aligned} \quad (14)$$

where $\mathcal{F}_m(\cdot)$, $\mathcal{G}(\cdot)$, and $\mathcal{H}(\cdot)$ are closed convex proper functions. Let $\mathcal{H}(\cdot)$ be differentiable with an L -Lipschitz continuous gradient for some $L \in (0, \infty)$. Suppose (14) has at least

one solution. Now, assume all the subproblems have solutions, and so-called dual residual $\lim_{i \rightarrow +\infty} (\bar{\mathbf{x}}^{(i+1)} - \bar{\mathbf{x}}^{(i)}) = 0$ and primal residual $\lim_{i \rightarrow +\infty} (\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{x}}^{(i+1)}) = 0, \forall m = 1, \dots, M$; consider a suitable step-size $\tau \in \mathbb{R}_{\geq 0}$ and a relaxation/penalty parameter $\rho \in \mathbb{R}_{> 0}$ with some arbitrary initial $\{\bar{\mathbf{x}}^{(0)}, \bar{\mathbf{y}}_m^{(0)}, \bar{\mathbf{z}}_m^{(0)}\}$. Then, the following iterative scheme

$$\begin{aligned} \bar{\mathbf{x}}^{(i+1)} &:= \arg \min_{\bar{\mathbf{x}}} \mathcal{G}(\bar{\mathbf{x}}) \\ &+ \sum_{m=1}^M \rho \left\| \bar{\mathbf{y}}_m^{(i)} - \bar{\mathbf{x}} - \tau \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i)}) + \frac{\bar{\mathbf{z}}_m^{(i)}}{\rho} \right\|_2^2 \end{aligned} \quad (15a)$$

$$\begin{aligned} \bar{\mathbf{y}}_m^{(i+1)} &:= \arg \min_{\bar{\mathbf{y}}_m} \mathcal{F}_m(\bar{\mathbf{y}}_m) \\ &+ \rho \left\| \bar{\mathbf{y}}_m - \bar{\mathbf{x}}^{(i+1)} + \frac{\bar{\mathbf{z}}_m^{(i)}}{\rho} \right\|_2^2 \quad \forall m = 1, \dots, M \end{aligned} \quad (15b)$$

$$\bar{\mathbf{z}}_m^{(i+1)} := \bar{\mathbf{z}}_m^{(i)} + \rho (\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{x}}^{(i+1)}) \quad \forall m = 1, \dots, M, \quad (15c)$$

at any limit point, $\{\bar{\mathbf{x}}^{(i)}\}$ converges to a Karush-Kuhn-Tucker (KKT) point of (14).

Proof. See Appendix B. ■

Observation 1. Let step-size $\tau=0$, the proposed TOP-ADMM becomes classic ADMM.

B. TOP-ADMM-based Efficient Solution to Problem P1 and P2

We now employ the TOP-ADMM algorithm, cf. Theorem 1, to solve (9) such that

$$\begin{aligned} &\underset{\bar{\mathbf{X}}, \bar{\mathbf{Y}}_m \in \mathbb{C}^{N_T \times N}}{\text{minimize}} \quad f(\bar{\mathbf{X}}) + \delta_{\mathcal{S}}(\bar{\mathbf{X}}) + \sum_{m=1}^M \delta_{\mathcal{C}_m}(\bar{\mathbf{Y}}_m) \\ &\text{subject to} \quad \bar{\mathbf{Y}}_m = \bar{\mathbf{X}} \quad \forall m = 1, \dots, M, \end{aligned} \quad (16)$$

where $f(\bar{\mathbf{X}})$ is a differentiable function with Lipschitz continuous gradient—cf. Definition 5; $\delta_{\mathcal{S}}(\cdot)$ and $\delta_{\mathcal{C}_m}(\cdot)$ are indicator functions to the set (12)/(13) and (11), respectively.

Hence, the TOP-ADMM algorithm employed for the above problem (16) can be expressed as

$$\bar{\mathbf{X}} \leftarrow \arg \min_{\bar{\mathbf{X}}} \delta_{\mathcal{S}}(\bar{\mathbf{X}}) + \rho \sum_{m=1}^M \left\| \bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m - \tau \nabla f \right\|_F^2 \quad (17a)$$

$$\bar{\mathbf{Y}}_m \leftarrow \arg \min_{\bar{\mathbf{Y}}_m} \delta_{\mathcal{C}_m}(\bar{\mathbf{Y}}_m) + \rho \left\| \bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m \right\|_F^2 \quad \forall m \quad (17b)$$

$$\bar{\mathbf{Z}}_m \leftarrow \bar{\mathbf{Z}}_m + \bar{\mathbf{Y}}_m - \bar{\mathbf{X}} \quad \forall m = 1, \dots, M. \quad (17c)$$

In the first step of our proposed TOP-ADMM mask-compliant spectral precoding algorithm, taking the derivative with respect to $\bar{\mathbf{X}}$ and setting to zero yields orthogonal projection (20c), namely $\text{proj}_{\mathcal{S}}$ onto the spatial constraint either (12) or (13). The $\text{proj}_{\mathcal{S}_{p1}}$ for each k -th subcarrier

Algorithm 1 TOP-ADMM to solve P1 and P2—(9) or (16)

Inputs: $\mathbf{X} \in \mathbb{C}^{N_T \times N}$, $\{\gamma_m \in \mathbb{R}; \mathbf{a}(\nu_m)\}_{m=1}^M$

Output(s): $\bar{\mathbf{X}}^{(I)} \in \mathbb{C}^{N_T \times N}$

- 1: Initialization: $\bar{\mathbf{X}}^{(0)} = \mathbf{X}$, $\bar{\mathbf{Y}}_m^{(0)} = \mathbf{0}_{N_T \times N}$, and $\bar{\mathbf{Z}}_m^{(0)} = \mathbf{0}_{N_T \times N}$
- 2: for $i = 1, 2, \dots, I$ do

$$\nabla f(\bar{\mathbf{X}}^{(i-1)}) := -\Pi(\mathbf{X} - \bar{\mathbf{X}}^{(i-1)}) \quad (20a)$$

$$\bar{\mathbf{U}} = \frac{1}{M} \sum_{m=1}^M (\bar{\mathbf{Y}}_m^{(i-1)} + \bar{\mathbf{Z}}_m^{(i-1)}) - \tau \nabla f(\bar{\mathbf{X}}^{(i-1)}) \quad (20b)$$

$$\begin{aligned} \bar{\mathbf{X}}^{(i)} &= \text{proj}_{\mathcal{S}}(\bar{\mathbf{U}}) \\ &\equiv \text{select} \begin{cases} \text{proj}_{\mathcal{S}_{p1}}(\bar{\mathbf{U}}) & \% \text{ use Theorem 3} \\ \text{proj}_{\mathcal{S}_{p2}}(\bar{\mathbf{U}}) & \% \text{ use Theorem 4} \end{cases} \end{aligned} \quad (20c)$$

- 3: parfor $m = 1, \dots, M$ do ▷ % run parallel

$$\bar{\mathbf{Y}}_m^{(i)} = \text{proj}_{\mathcal{C}_m}(\bar{\mathbf{X}}^{(i)} - \bar{\mathbf{Z}}_m^{(i-1)}) \quad \% \text{ use Theorem 2} \quad (20d)$$

$$\bar{\mathbf{Z}}_m^{(i)} = \bar{\mathbf{Z}}_m^{(i-1)} + \bar{\mathbf{Y}}_m^{(i)} - \bar{\mathbf{X}}^{(i)} \quad (20e)$$

- 4: end parfor
- 5: end for
- 6: return $\bar{\mathbf{X}}^{(I)}$

corresponding to the set (12) reads by appealing to Theorem 3 and Theorem 5—cf. Appendix C:

$$\bar{\mathbf{X}}^{(i)}[:, k] = (\mathbf{I}_{N_T} - \widehat{\mathbf{H}}^+ [k] \widehat{\mathbf{H}} [k]) \bar{\mathbf{U}}[:, k] + \widehat{\mathbf{H}}^+ [k] \widehat{\mathbf{H}} [k] \mathbf{X}[:, k], \quad (18)$$

where matrix $\bar{\mathbf{U}} \in \mathbb{C}^{N_T \times N}$ is given in (20b) and $\widehat{\mathbf{H}}^+ := \widehat{\mathbf{H}}^H (\widehat{\mathbf{H}} \widehat{\mathbf{H}}^H)^{-1}$ if $N_R < N_T$ and full rank. The $\text{proj}_{\mathcal{S}_{p2}}$ can efficiently be computed using Theorem 4 (and Theorem 5 to exploit separability):

$$\begin{aligned} \bar{\mathbf{X}}^{(i)}[:, k] &= \mathbf{X}[:, k] \\ &+ \mathbf{U}(\mathbf{I}_{N_T} + \mu \mathbf{\Lambda})^{-1} \mathbf{U}^H (\bar{\mathbf{X}}^{(i-1)}[:, k] - \mathbf{X}[:, k]), \end{aligned} \quad (19)$$

where $\mathbf{Q}[k] = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U} := (\widehat{\mathbf{H}} [k]^H \widehat{\mathbf{H}} [k] + \sigma_{\text{ce}}^2 [k] \mathbf{I}_{N_T})$, $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_{N_T})$, and $\mu \geq 0$ is a unique positive root of $\varsigma^2 [k] + \sigma_{\text{ce}}^2 [k] - \sum_{j=1}^{N_T} \frac{\lambda_j}{(1+\mu\lambda_j)^2} |v[j]|^2 = 0$; $\mathbf{v} = \mathbf{U}^H \bar{\mathbf{X}}^{(i-1)}[:, k]$.

The second step is a projection operator onto the rank 1 quadratic constraint (cf. Theorem 2) yielding (20d). Algorithm 1 summarizes the proposed recipe for the TOP-ADMM based mask-compliant spectral precoding, where I denotes the total number of iterations.

Practical Convergence: In practice, we do not require to obtain the solution at very high accuracy in particular for the problem of interest in this paper. Therefore, the proposed TOP-ADMM-based solution can have early-stopping as per the desired modest accuracy of required key performance indicators—as further discussed in the numerical Section V.

C. Shortcomings of Classical ADMM-based Solutions to P1 and P2

In this section, we highlight the demerits of the possible solutions using classical two-operator ADMM and contrast

TABLE I: Complexity comparison between TOP-ADMM and general-purpose solvers typically used in CVX [44]

Method	Complexity for P1	Complexity for P2
General-purpose Interior-point-based Solver	$\mathcal{O}(N^{4.5}N_T)$	
TOP-ADMM	$\mathcal{O}(N(N_T^2N_R + N_R^3 + N_R^2N_T + N_T^2) + I(NN_T^2 + MNN_T))$	$\mathcal{O}(NN_T^3 + I(NI'N_T + NN_T^3 + MNN_T))$

them with the capability of the proposed TOP-ADMM-based solution. There can be at least two approaches to find solutions for P1 and P2 using classical ADMM. It can be achieved by reducing three operators to effectively two operators problem, namely

ADMM approach 1 (referred to as ADMM1):

$$\bar{\mathbf{X}} \leftarrow \arg \min_{\bar{\mathbf{X}}} f(\bar{\mathbf{X}}) + \delta_{\mathcal{S}}(\bar{\mathbf{X}}) + \rho \sum_{m=1}^M \|\bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m\|_F^2 \quad (21a)$$

$$\bar{\mathbf{Y}}_m \leftarrow \arg \min_{\bar{\mathbf{Y}}_m} \delta_{\mathcal{C}_m}(\bar{\mathbf{Y}}_m) + \rho \|\bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m\|_F^2 \quad \forall m \quad (21b)$$

$$\bar{\mathbf{Z}}_m \leftarrow \bar{\mathbf{Z}}_m + \bar{\mathbf{Y}}_m - \bar{\mathbf{X}} \quad \forall m = 1, \dots, M. \quad (21c)$$

ADMM approach 2 (referred to as ADMM2):

$$\bar{\mathbf{X}} \leftarrow \arg \min_{\bar{\mathbf{X}}} f(\bar{\mathbf{X}}) + \rho \sum_{m=1}^{M+1} \|\bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m\|_F^2 \quad (22a)$$

$$\bar{\mathbf{Y}}_m \leftarrow \arg \min_{\bar{\mathbf{Y}}_m} \delta_{\tilde{\mathcal{C}}_m}(\bar{\mathbf{Y}}_m) + \rho \|\bar{\mathbf{Y}}_m - \bar{\mathbf{X}} + \bar{\mathbf{Z}}_m\|_F^2 \quad \forall m \quad (22b)$$

$$\bar{\mathbf{Z}}_m \leftarrow \bar{\mathbf{Z}}_m + \bar{\mathbf{Y}}_m - \bar{\mathbf{X}} \quad \forall m = 1, \dots, M+1, \quad (22c)$$

where $\delta_{\tilde{\mathcal{C}}_m} = \delta_{\mathcal{C}_m}$ for $m = 1, \dots, M$, and $\delta_{\tilde{\mathcal{C}}_{M+1}} = \delta_{\mathcal{S}}$ for $m = M+1$.

On one hand, considering ADMM1, *i.e.*, (21), we can compute the projection operator corresponding to step (21b) in closed-form utilizing Theorem 2. However, an efficient solution to step (22a) is cumbersome because of the proximal operator of a composite function $f(\bar{\mathbf{X}}) + \delta_{\mathcal{S}}(\bar{\mathbf{X}})$, *i.e.*, $\text{prox}_{(f+\delta_{\mathcal{S}})}$. In general, there is no closed-form solution to such composite function unless one resorts to another iterative method, see, *e.g.*, [28]. This composite problem in hand implies some inner iterative scheme, in addition to iterative method to compute proximal operator for \mathcal{S}_{P2} (cf. (13)), within outer ADMM loop. Considering these unwieldy array of options, we do not pursue ADMM1 approach explicitly. Observe if we ignore f in (21a), then TOP-ADMM with $\tau = 0$ becomes ADMM1 (without f). On the other hand, ADMM2, *i.e.*, (22), may seem enticing to solve both P1 and P2, but as we see later in the numerical section, cf. Fig. 5, this approach is painfully slow to converge. On the stark contrast to both classical ADMM-based solutions, TOP-ADMM overcomes the shortcomings by splitting three operators elegantly—see Algorithm 1—since it capitalizes on the differentiability property of $f(\bar{\mathbf{X}})$.

D. Complexity Analysis

We now analyze the run-time complexity, in terms of required complex multiplications but ignore the offline complexity, complex additions, and real-valued multiplications/additions. Notice, we assume that all the subcarriers

are allocated, which will give the worst-case complexity analysis—please see the summary in Table I.

Interior-point-based P1/P2: Using complexity results in [16], the computational complexity of interior-point-based off-the-shelf solver is almost similar to MSP $\mathcal{O}(N^{4.5}N_T)$ since $N_T < N$.

TOP-ADMM-based P1/P2: The initialization step requires no multiplications. In gradient computation step (20a) and (20b), there will only be N real multiplications assuming $\mathbf{\Pi}$ as a diagonal matrix. In the orthogonal projection step onto the spatial constraint expressed by (20c), for P1, the computation of $\mathbf{H}^+[k]\mathbf{H}[k]$ in (18) and consequently the computation of $\mathbf{H}^+[k]\mathbf{H}[k]\mathbf{X}[:,k]$ for each subcarrier can be done outside the iteration loop once, whose complexities are $\mathcal{O}(N(N_T^2N_R + N_R^3 + N_R^2N_T))$ and $\mathcal{O}(NN_T^2)$ for all the N data carrying subcarriers. Thus, within an iteration cycle, the online algebraic complexity for the update corresponding to P1 spatial constraint is in the matrix vector product $\mathbf{H}^+[k]\mathbf{H}[k]\mathbf{U}[:,k]$, which is $\mathcal{O}(INN_T^2)$ for all I given iterations and N subcarriers (that can be done in parallel). Similarly, for P2, the projection computations in (19) require eigenvalue decomposition of $\mathbf{Q}[k]$ for each subcarrier that can be done once outside the iteration loop, whose complexity for all the subcarriers is $\mathcal{O}(NN_T^3)$. The computation of root μ needs to be computed within iteration cycle, which can efficiently be obtained using, *e.g.*, bisection method, see [37]—whose complexity for all the subcarriers and all I given iterations is $\mathcal{O}(INI'N_T)$, where I' denotes the number of bisection-like-method iterations. Additionally, the cost of computation of $\mathbf{U}(\mathbf{I}_{N_T} + \mu\mathbf{\Lambda})^{-1}\mathbf{U}^H$ is $\mathcal{O}(INN_T^3)$ for all the subcarriers and I given iterations. Further, within an iteration cycle, the orthogonal projection onto the spectral constraint in step (20d) requires $\mathcal{O}(MN)$ complex multiplications per iteration and transmit antenna. However, due to distributed nature of TOP-ADMM, the M subiterations can run in parallel per iteration cycle at the expense of increased memory requirements. Thus, ignoring parallelization possibility across N subcarriers, N_T antennas, and M subiterations, the total run-time complexity for I iterations and N_T transmit antennas is in the order of $\mathcal{O}(IMNN_T)$. Hence, the total run-time complexity to solve P1 and P2 are $\mathcal{O}(N(N_T^2N_R + N_R^3 + N_R^2N_T + N_T^2) + I(NN_T^2 + MNN_T))$ and $\mathcal{O}(NN_T^3 + I(NI'N_T + NN_T^3 + MNN_T))$, respectively.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed TOP-ADMM-based algorithm for the optimization problems P1 and P2 utilizing a 5G NR (Rel-15) compliant link-level simulator. Moreover, we compare the performance of the proposed algorithms with the conventional spectral precoders. For the perfect channel knowledge at the transmitter, we evaluate P1 only since P2 yield the same performance for $\varsigma[k] = 2\%$, $\forall k \in \mathcal{T}$, which we denote as “P1/P2”. However, for

TABLE II: Simulation Parameters for TDD NR (Rel-15) PDSCH Type-A

Parameters	Test 1	Test 2
Subcarrier Spacing		15 kHz
Carrier Bandwidth (PRB allocation)		5 MHz (25 PRBs, <i>i.e.</i> , active subcarriers $N_{SC}=300$)
Carrier Spacing for ACLR		5 MHz upper and lower adjacent channels [3]
DL SU-MIMO N_T, N_R		8/64Tx, 2Rx
Spatial Layers (rank)	Fixed rank 1	adaptive (10% BLER)
Spatial Precoding		RZF precoding with regularization $\alpha=0.001$ and 1 PRB granularity
Modulation	256QAM	adaptive (10% BLER)
Code-rate	$1/2$ $5/6$	adaptive (10% BLER)
Channel Model		TDL-A (300ns, 10Hz) & spatial correlation low [48]
Channel & Noise power		Practical linear minimum mean squared error (LMMSE) based
Hybrid Automatic Repeat Request (HARQ) max transmissions		4 (3 max retransmissions with rv {0, 2, 3, 1}) [49]
Other Information		LMMSE Rx with interference rejection combining (IRC); $\varsigma[k]=2\%$ $\forall k$; no other impairments

the imperfect channel knowledge, we evaluate the robustness of P2 against channel uncertainty.

A. Performance Measures

We analyze the spectral precoding performance in terms of two figure-of-merits namely, OOB and in-band distortions, in particular assuming a base station supporting sub-6 GHz, *e.g.*, frequency ranges between 410 MHz and 7.125 GHz [3, Section 5.1].

1) *Out-of-Band Distortion*: As mentioned in Section II-D1, mask and (conducted) ACLR are typically the performance metrics to quantify the operating band unwanted emissions.

In practical systems, the (digital) spectrum shaping is followed by other (non-linear) digital and analog processing, as illustrated in Fig. 1. Consequently, there is some spectral regrowth phenomenon due to such (non-linear) components in the transmitter after spectrum shaping. Thus, an implementation margin in terms of ACLR and mask requirements are necessary to cope with spectral regrowth. Hence, OOB performance by spectrum shaping, *e.g.*, spectral precoding, should be better than the stipulated mask as a margin to account for potential spectrum regrowth.

We have considered ACLR corresponding to the 1st adjacent carrier in both upper and lower frequencies, where the minimum requirement is 45 dB—worst-case of measured ACLR in the upper and lower channels [3, Section 6.6.3]. It is worth highlighting that these ACLR requirements are for the complete radio chain, *i.e.*, measurements need to be performed at the antenna connector. Thus, spectrum shaping may have some aggressive mask and ACLR requirements to meet the minimum requirements at the antenna connector. Therefore, our target mask, cf. (5c), $\Gamma[:, j] = \gamma \preceq \text{NR mask}; \forall j = 1, \dots, N_{\text{T}}$, is lower than the regulating mask—see Definition 1—which indirectly offers implementation margin to the stipulated ACLR requirements.

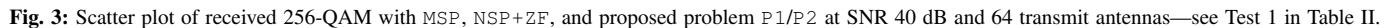
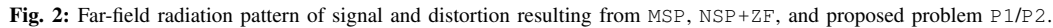
2) *In-Band Distortion*: In these simulations, the in-band distortion is not only quantified in terms of TxEVM but also in terms of block error rate (BLER) [46] and throughput [47]. We present normalized throughput, *i.e.*, normalizing the throughput results by the maximum achievable throughput without any spectral precoding or OOB reduction, referred to as NO_OOB, and other hardware impairments or imperfections for Test 2.

B. Simulation Parameters and Assumptions

The key simulation parameters for the physical downlink shared channel (PDSCH) with type-A⁵ and the three investigated test scenarios are summarized in Table II, see, *e.g.*, [3], [23], for the detailed NR physical layer and performance requirements. We have considered 15 kHz subcarrier spacing for the NR numerology unless otherwise mentioned. Furthermore, no supporting signals are transmitted besides PDSCH along with the demodulation reference signal for the practical channel and noise variance estimation at the UE side. Note that for simulations purpose, we have considered a 5 MHz channel bandwidth with 25 physical resource block (PRB) allocation, even though the proposed methods can be employed for arbitrary bandwidths.

For all the considered test scenarios, we have employed regularized ZF (RZF)-based MIMO (spatial) precoding in the downlink, where the downlink channel $\{\mathbf{H}[k]\}$ was acquired perfectly, unless otherwise stated, by sounding the whole bandwidth in the uplink every slot, *i.e.*, periodically 1 ms for 15 kHz subcarrier spacing, via sounding reference signals [1], [23]. However, the acquired channel estimates for RZF precoding are used after one downlink slot delay. Subsequently, we also evaluate the impact of channel uncertainty considering both proposed problems, *i.e.*, P1 and P2. The linear MIMO precoding can be expressed as $\mathbf{x}[k] = \mathbf{W}[k]\mathbf{s}[k]$, where the linear RZF precoder reads $\mathbf{W}[k] = \Xi \odot \widetilde{\mathbf{W}}[k]$. The un-normalized RZF is given by, see, *e.g.* [51], $\mathbb{C}^{N_T \times N_L} \ni \widetilde{\mathbf{W}}[k] = \mathbf{H}[k]^H \left(\mathbf{H}[k]\mathbf{H}[k]^H + \widetilde{\mathbf{R}} + \alpha \mathbf{I}_{N_R} \right)^{-1} := [\widetilde{\mathbf{w}}_1[k], \dots, \widetilde{\mathbf{w}}_{N_L}[k]]$. The choice of suitable regularizing parameters, namely Hermitian positive semidefinite $\widetilde{\mathbf{R}}$ and $\alpha \in \mathbb{R}_{\geq 0}$, are typically tunable and can be found in, *e.g.*, [51], [52], to maximize the desired performance metrics of the system such as sum-rate. Moreover, the power normalization matrix can be [25, Section 7.1] $\mathbb{C}^{N_T \times N_L} \ni \Xi = \left[\frac{\sqrt{\eta_1}}{\|\widetilde{\mathbf{w}}_1[k]\|_2} \mathbf{1}_{N_T}, \dots, \frac{\sqrt{\eta_{N_L}}}{\|\widetilde{\mathbf{w}}_{N_L}[k]\|_2} \mathbf{1}_{N_T} \right]$, where $\{\sqrt{\eta_s}\}$ is a set of powers that can be assigned to respective layers or users. Consequently, the total power allocated to the k -th subcarrier must satisfy $\|\mathbf{W}[k]\|_F^2 = \sum_{s=1}^{N_L} \eta_s$. Furthermore, for brevity, we have assumed the weight matrix $\mathbf{\Pi} = \mathbf{I}$ and left other possibilities as a potential future work.

⁵These data types refer to different PDSCH demodulation reference signals allocation [50].



Conspicuously, Fig. 2(a) and Fig. 2(b) show that the distortion/TxEVM $\epsilon[k]$ emanating from MSP , cf. (2), is beamformed along the same direction as the signal $\mathbf{X}[:, k]$. Moreover, in Fig. 2(b), there is some smearing effect of the beamforming of the distortion as well due to the weighted combination of the beamformed distortions from other sub-carriers as can be comprehended from (2). Other authors have observed a similar trend but in the context of the impact of non-linear devices, *e.g.*, non-linear power amplifier, due

1536-1276 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 11, 2021 at 09:45:12 UTC from IEEE Xplore. Restrictions apply.

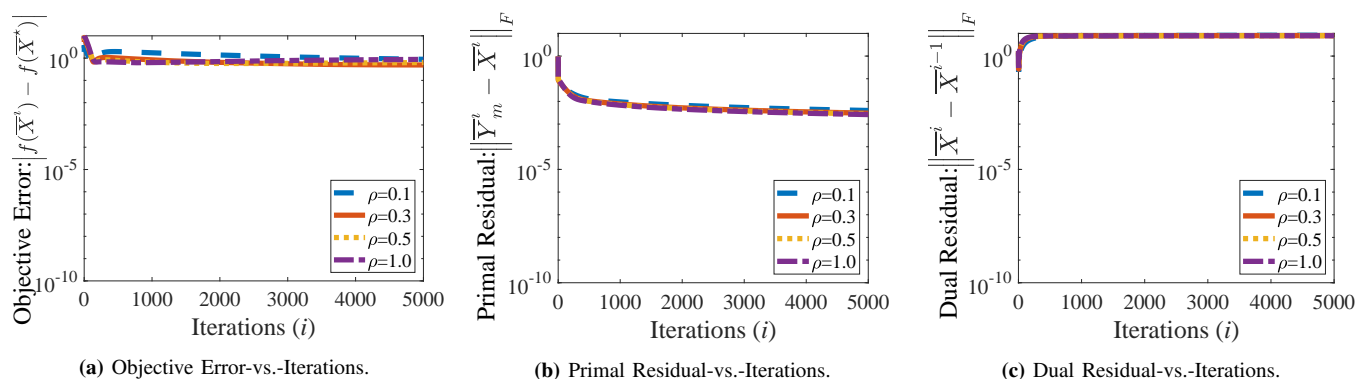
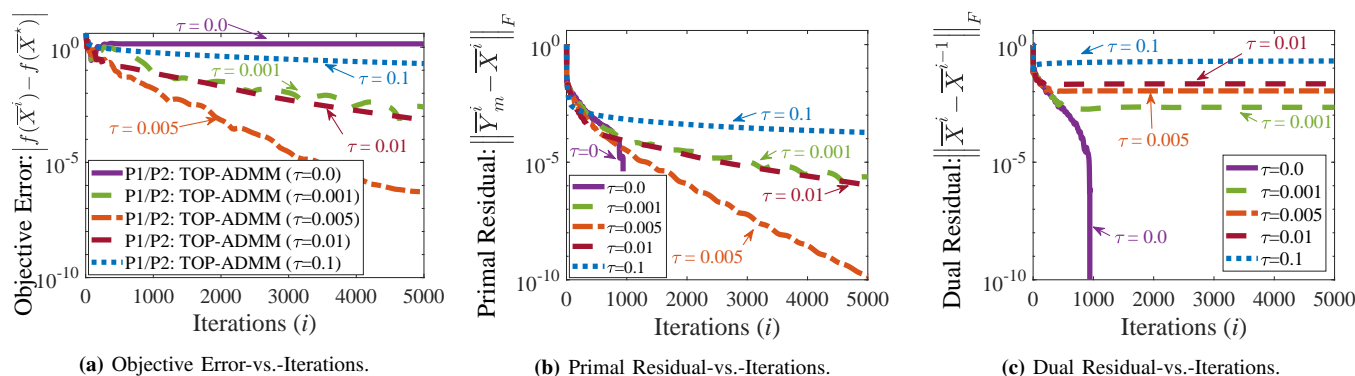


Figure 4 exhibits convergence behaviour of the proposed

Figure 6 shows the practical convergence of some key performance metrics, namely ACLR and the normalized throughput, of the proposed TOP-ADMM-based Algorithm 1 for various step-sizes $\tau \in \{0, 0.001, 0.005\}$ considering Test 2—cf. Table II. This figure illustrates the convergence behaviour

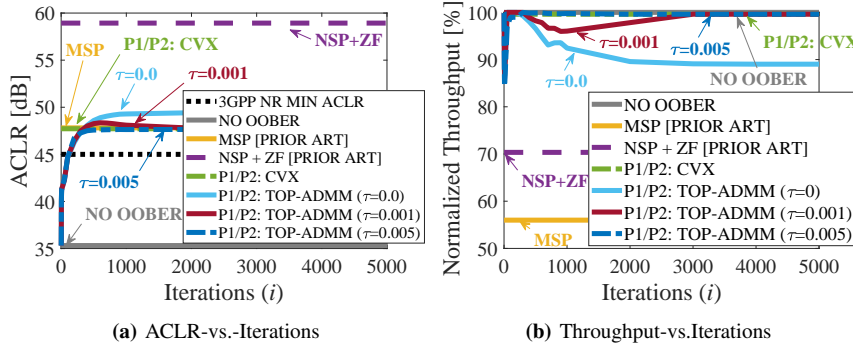


Fig. 6: Convergence of two key metrics using TOP-ADMM for P1/P2.

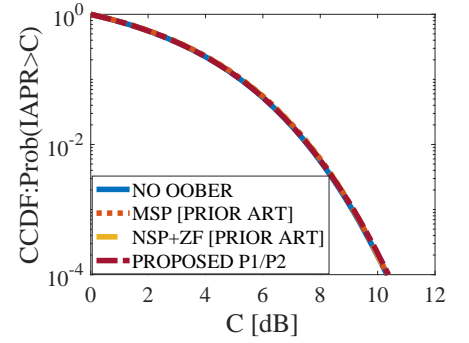


Fig. 7: IAPR distribution per antenna.

TABLE III: Comparison of achieved metrics among the various methods

Method	Iteration	ACLR [dB]	Throughput [%]
NO OOB (no OOB reduction)	-	35.32	100.00
MSP [14]	-	47.81	55.99
NSP+ZF [20, Paper F]	-	58.93	67.34
CVX-based P1 solution	-	47.72	99.59
TOP-ADMM $\tau = 0.0$ (ADMM1 w/o f)	400/5000	48.02/49.41	98.03/89.01
TOP-ADMM $\tau = 0.001$	400/5000	48.01/47.76	98.90/99.59
TOP-ADMM $\tau = 0.005$	400/5000	47.51/47.67	99.86/99.59

of the algorithm in terms of out-of-band ACLR and in-band normalized throughput, where the reference solution to the problem P1/P2 is obtained from the CVX [44]. Notice that both metrics, *i.e.*, ACLR and the normalized throughput, are not the direct results obtained from solving the optimization problem P1/P2, but they are the derived performance metrics which are more meaningful. All the presented methods meet the minimum 3GPP NR ACLR requirement of 45 dB for the first adjacent channel. Moreover, Table III summarizes and compares the achieved performance metrics by following the results shown in Fig. 6. At iteration 5000, the solution obtained with step-sizes $\tau \in \{0.001, 0.005\}$ converge nearly to the solution rendered by the CVX solver in terms of both considered metrics. However, the solution delivered by $\tau = 0.005$ at iteration 400 seems to be good enough for the considered test scenario. The subsequent presented TOP-ADMM-based $\tau = 0.005$ results are with fixed 400 iterations since it has converged to the desired result. Furthermore, for the completeness, we have illustrated complementary cumulative distribution function (CCDF) of instantaneous peak to average power ratio (IAPR), *i.e.*, $IAPR = |t_j[n]|^2 / \mathbb{E}\{|t_j|^2\}$ [53], where t_j are time domain OFDM signal corresponding to j -th antenna after spectral precoding—no impact on IAPR due to spectral precoding has been observed.

Figure 8(a) depicts the average TxEVM distribution per PRB in the frequency-domain [23]. The edge PRB have relatively high distortion power compared to the central PRB. Notice that the NR bandwidth is slightly asymmetric with respect to the direct-current carrier. Moreover, the discrete selected frequency points (ν) are not symmetric with respect to the direct-current carrier. Therefore, one could consequently observe that the TxEVM distribution in the frequency-domain is asymmetric. The TOP-ADMM-based solution, notably with $\tau = 0.005$ and fixed iterations $I = 400$, renders almost similar

TxEVM distribution as the solution obtained by CVX. The TOP-ADMM with $\tau = 0.0$, *i.e.*, ADMM1 without f , and fixed iterations $I = 1000$ renders higher TxEVM than the solution with $\tau = 0.005$ or CVX. Note that, in practical systems, the total transmission signal power budget is fixed per transmit antenna branch, which effectively means that the distortion and actual/true signal share the given power per subcarrier. In other words, after spectral precoding, we renormalize the transmitted signal (with distortion) to meet the given power budget. Since NSP+ZF (prior art) has relatively high distortion power, the actual signal power transmitted from most of the subcarriers is low, which explicates the poor performance of NSP+ZF. Note that one of the problems of interest in this work is to mitigate or null out the received TxEVM, *i.e.*, in-band TxEVM seen at the (far-field) receiver, while meeting the OOB requirement. Thus, the proposed method in this paper is suitable for base station classes supporting over-the-air requirements in NR [3, Section 9].

Figure 8(b) exhibits the average PSD versus frequency for the proposed algorithms. The NR mask corresponding to a medium-range BS with the maximum output of 38 dBm [3, Section 6.6.4] is also shown for the completeness. However, the mask is normalized according to the normalized transmit signal power of 0 dBm, *i.e.*, approximately -21.5 dBm/100 kHz, in the link simulations. All the proposed methods in addition to the prior art fulfil the 3GPP NR mask (except the original OFDM signal without any spectrum shaping).

In Fig. 9(a) and Fig. 9(b), we depict the in-band performance, BLER versus received signal-to-noise ratio (SNR), of the proposed method for both 8 and 64 transmit antennas for the Test 1 (cf. Table II), *i.e.*, fixed reference channel scenario with fixed rank 1 in SU-MIMO—there is no link adaptation in this test except for the spatial (RZF) precoding adaptation in every slot. The proposed TOP-ADMM-based P1/P2 solution

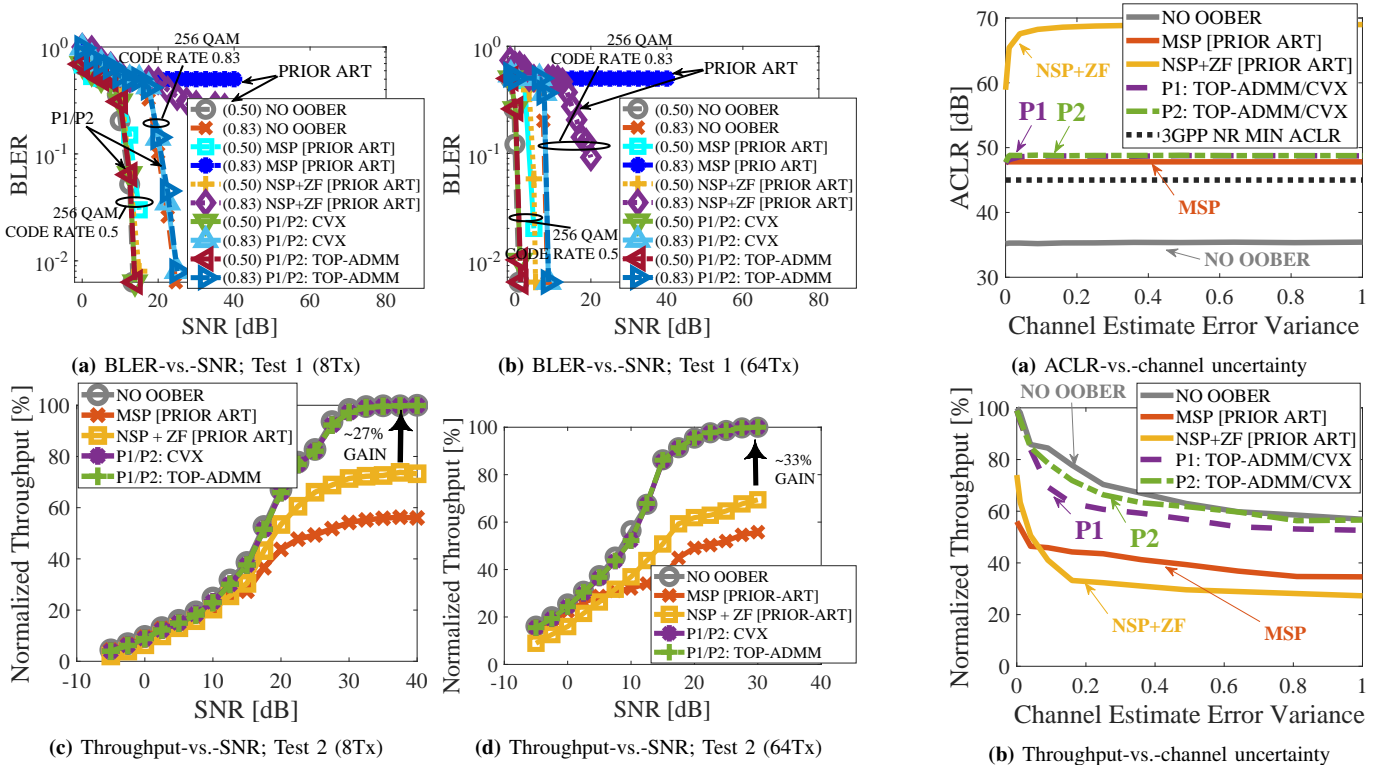
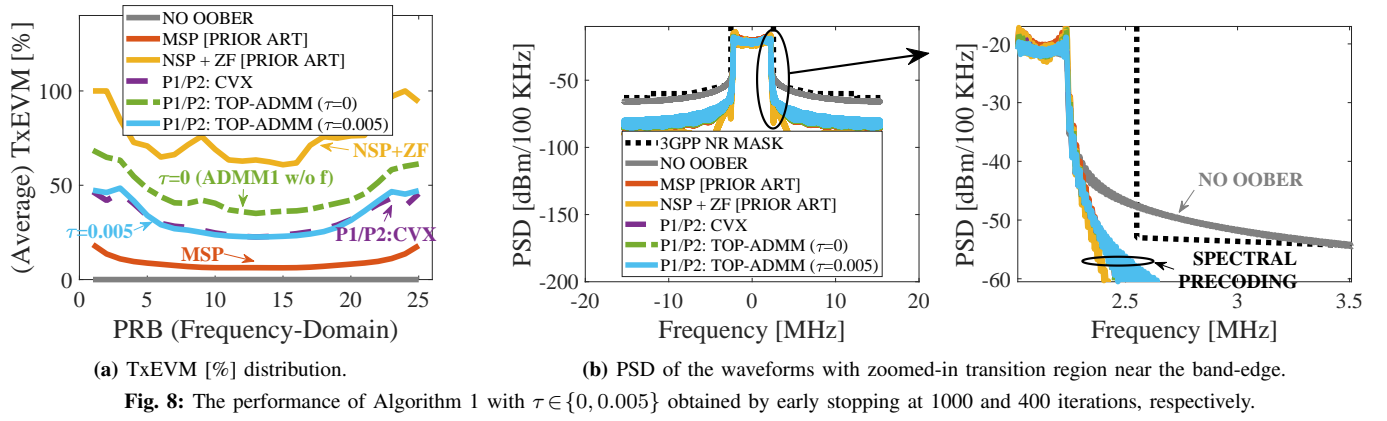


Fig. 10: Impact of channel uncertainty.

with $\tau=0.005$ and fixed iterations $I=400$ shows the similar performance as obtained by CVX, whereas NSP+ZF has poor performance. Figure 9(c) and Fig. 9(d) exhibit normalized throughput versus received SNR for the Test 2 with 8 and transmit antennas. Since this test employs fast link adaptation, *i.e.*, not only adaptive spatial precoding, but also adaptive rank/layers, modulation, and code rate considering 10% target BLER on long-term average. From these figures, we can deduce that increasing number of the transmit antennas from 8 to 64 for the low-rank scenario does not improve the in-band performance gain of MSP—as observed in [19]. However, NSP+ZF precoding fails to harness the potential gain offered in massive MIMO-based systems due to the availability of the large degrees of freedom in asymmetric MIMO systems, *i.e.*, low-rank system. In other words, NSP+ZF method does not support the flexibility to adapt the threshold/mask in the OOB

reduction as it only notches those desired frequencies, which consequently produces a large amount of distortion power and thereby little amount of useful signal power can be transmitted. Hence, our proposed method offers throughput gain by approximately 27% and 33% compared to the NSP+ZF in 8 and 64 transmit antennas, respectively, yet meeting OOB requirements.

Figure 10 manifests OOB and in-band performance against channel estimation error variance. In this set of simulations, we assume that the channel knowledge available at the transmitter for MIMO and spectral precoding is imperfect. The estimation error model is considered as $\mathbf{H} = \hat{\mathbf{H}} + \Delta\mathbf{H}$, where $\Delta\mathbf{H}$ have entries that are independent and identically distributed and follow zero mean circularly symmetric complex Gaussian distributed with given variance $\tilde{\sigma}_{ce}^2$. Furthermore, we have shown that the results obtained by CVX and TOP-ADMM

algorithm are same previously, we evaluate the performance of P1 and P2 using TOP-ADMM ($\tau = 0.005$ and 400 iterations) to avoid clutter look; thus, we use the legend as “TOP-ADMM/CVX”. Spectacularly, all the spectral precoding techniques have no loss on the OOB, *i.e.*, ACLR, performance due to channel uncertainty. Noticeably, ACLR performance of NSP+ZF has increased but at the significant cost of in-band, *i.e.*, throughput, performance. The throughput performance of P2, as expected, is more robust to channel uncertainty than P1. Strikingly, both proposed problems P1 and P2 render better throughput than the prior art NSP+ZF and MSP yet fulfilling OOB performance requirements. Note that P2 has some loss compared to NO OOB, which can be explained because MIMO precoding utilizes an averaged channel estimate within a considered PRB granularity (cf. Table II). Thus, MIMO precoding uses denoised channel estimate whereas the proposed spectral precoding in these simulations operates by using noisy channel estimate every subcarrier (but can employ averaged channel estimates of any granularity). This further substantiates that our proposed schemes work independently from any MIMO precoding.

VI. CONCLUSIONS

We formulated the TxEVM-mitigated and mask-compliant spectral precoding problems for a massive MIMO-OFDM-based system as convex optimization problems offering a trade-off between out-of-band emissions suppression and in-band performance for the massive MIMO-OFDM systems. More concretely, our proposed approaches not only meet the desired spectral emission mask requirement, *i.e.*, reduces the out-of-band emissions, at the transmitter but also mitigate the resulting distortion/TxEVM seen at the receiver. The second approach offers robustness against channel uncertainty at the expense of increased computational complexity. A general-purpose solver for such a large-scale problem suffers from high computational complexity. Hence, we have developed a novel, efficient, and simple three-operator consensus ADMM, so-called TOP-ADMM, scheme that recovers the classical two-operator consensus ADMM as a special case. The computationally efficient TOP-ADMM-based solution to the proposed optimization problem essentially decomposes the large-scale optimization problem into subproblems, where each subproblem can be solved efficiently. Numerical results corroborate that the proposed efficient algorithm achieves comparable performance to the solution obtained via a general-purpose solver using CVX [44] within hundreds of iterations. The presented numerical results indicate superior in-band performance over state-of-the-art methods, yet fulfilling the out-of-band emissions requirement, *i.e.*, mask and ACLR requirements.

APPENDIX A

JOINT NSP WITH ZF SPATIAL PRECODING (NSP+ZF)

Recently, channel state information aware spectral precoding, in particular joint spatial (ZF precoding) and NSP, has been proposed in [20, Paper F], which can be expressed as

$$\underset{\mathbf{X} \in \mathbb{C}^{N_T \times N}}{\text{minimize}} \quad \|\mathbf{X}\|_F^2 \quad (23a)$$

$$\text{subject to} \quad \mathbf{H}[k] \mathbf{X}[:, k] = \mathbf{S}[:, k] \quad \forall k \in \mathcal{T} \quad (23b)$$

$$\mathbf{A} \mathbf{X}^T = \mathbf{0}, \quad (23c)$$

where $\mathbb{C}^{N_L \times N} \ni \mathbf{S} := [\mathbf{s}[1], \dots, \mathbf{s}[N]]$ is neither spatially nor spectrally precoded.

Since (23) is a minimum Frobenius norm problem (23a) with ZF spatial constraint (23b) and NSP spectral constraint (23c), a closed-form solution reads $\mathbf{X} = \text{unvec}(\mathbb{A}^+ \mathbf{b})$, where $\mathbb{C}^{(N N_R + M N_T) \times N N_T} \ni \mathbb{A} := [\mathbf{e}_1 \otimes \mathbf{H}[1]; \dots; \mathbf{e}_N \otimes \mathbf{H}[N]; \mathbf{A} \otimes \mathbf{I}_{N_T}]$ and $\mathbf{b} := [\text{vec}(\mathbf{S}); \mathbf{0}_{M N_T}]$ such that $\mathbf{b} \in \mathbb{C}^{(N N_R + M N_T) \times 1}$. Moreover, $\mathbf{e}_k \in \mathbb{R}^{N \times 1}$ denotes standard basis vector and unvec creates a matrix by stacking the appropriate number of vector elements columnwise. Observe that we have a different notation from [20] for the benchmarking purpose.

APPENDIX B PROOF OF THEOREM 1

We show the basic convergence result of proposed TOP-ADMM scheme inspired by the convergence proof of classical two-operator (consensus) ADMM given in [29, Section 3.2] [37].

We form the Lagrangian to the problem (14),

$$\begin{aligned} \mathcal{L}(\{\bar{\mathbf{y}}_m\}_{m=1}^M, \bar{\mathbf{x}}, \{\bar{\mathbf{z}}_m\}_{m=1}^M) \\ := \sum_{m=1}^M \mathcal{F}_m(\bar{\mathbf{y}}_m) + \mathcal{G}(\bar{\mathbf{x}}) + \mathcal{H}(\bar{\mathbf{x}}) + \sum_{m=1}^M 2\Re\{\bar{\mathbf{z}}_m^H (\bar{\mathbf{y}}_m - \bar{\mathbf{x}})\}. \end{aligned}$$

Now, according to KKT optimality conditions—see, *e.g.*, [54], the necessary and sufficient optimality conditions for the problem (14) are dual feasibility, *i.e.*,

$$\mathbf{0} \in \partial \mathcal{G}(\bar{\mathbf{x}}^*) + \nabla \mathcal{H}(\bar{\mathbf{x}}^*) - \sum_{m=1}^M \bar{\mathbf{z}}_m^* \quad (24)$$

$$\mathbf{0} \in \partial \mathcal{F}_m(\bar{\mathbf{y}}_m^*) + \bar{\mathbf{z}}_m^*, \quad (25)$$

and the primal feasibility, *i.e.*,

$$\bar{\mathbf{y}}_m^* - \bar{\mathbf{x}}^* = \mathbf{0} \quad \forall m = 1, \dots, M. \quad (26)$$

Towards this end, we now analyze the proposed iterative algorithm, which asymptotically, $i \rightarrow \infty$, satisfies the above-mentioned optimality conditions.

In the first step of the algorithm (15a), $\bar{\mathbf{x}}^{(i+1)}$ minimizes that update step-1, *i.e.*,

$$\begin{aligned} \mathbf{0} \in \partial \mathcal{G}(\bar{\mathbf{x}}^{(i+1)}) \\ - \sum_{m=1}^M \rho \left(\bar{\mathbf{y}}_m^{(i)} - \bar{\mathbf{x}}^{(i+1)} - \tau \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i)}) + \frac{\bar{\mathbf{z}}_m^{(i)}}{\rho} \right) \quad (27) \end{aligned}$$

$$\begin{aligned} = \partial \mathcal{G}(\bar{\mathbf{x}}^{(i+1)}) \\ + \rho M \tau \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i)}) - \sum_{m=1}^M \left(\rho (\bar{\mathbf{y}}_m^{(i)} - \bar{\mathbf{x}}^{(i+1)}) + \bar{\mathbf{z}}_m^{(i)} \right) \quad (28) \end{aligned}$$

$$\begin{aligned} = \partial \mathcal{G}(\bar{\mathbf{x}}^{(i+1)}) \\ + \rho M \tau \left(\nabla \mathcal{H}(\bar{\mathbf{x}}^{(i)}) + \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i+1)}) - \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i+1)}) \right) \end{aligned}$$

$$-\sum_{m=1}^M \left(\rho \left(\bar{\mathbf{y}}_m^{(i+1)} - \mathbf{y}_m^{(i+1)} + \mathbf{y}_m^{(i)} - \mathbf{x}^{(i+1)} \right) + \mathbf{z}_m^{(i)} \right). \quad (29)$$

Plugging dual variable update (15c) in (29), then rearranging terms yield

$$\begin{aligned} \mathbf{0} \in & \partial \mathcal{G} \left(\bar{\mathbf{x}}^{(i+1)} \right) + \rho M \tau \nabla \mathcal{H} \left(\bar{\mathbf{x}}^{(i+1)} \right) - \sum_{m=1}^M \bar{\mathbf{z}}_m^{(i+1)} \\ & - \rho M \tau \left(\nabla \mathcal{H} \left(\bar{\mathbf{x}}^{(i+1)} \right) - \nabla \mathcal{H} \left(\bar{\mathbf{x}}^{(i)} \right) \right) \\ & + \sum_{m=1}^M \rho \left(\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{y}}_m^{(i)} \right). \end{aligned} \quad (30)$$

Now, assuming dual residual $(\bar{\mathbf{x}}^{(i+1)} - \bar{\mathbf{x}}^{(i)}) \rightarrow 0$ and primal residual $(\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{x}}^{(i)}) \rightarrow 0$, it implies $(\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{y}}_m^{(i)}) \rightarrow 0$ when $i \rightarrow \infty$. Since we assume that the gradient of \mathcal{H} is L -Lipschitz and $(\bar{\mathbf{x}}^{(i+1)} - \bar{\mathbf{x}}^{(i)}) \rightarrow 0$ for sufficiently large iterations, then the error $(\nabla \mathcal{H}(\bar{\mathbf{x}}^{(i+1)}) - \nabla \mathcal{H}(\bar{\mathbf{x}}^{(i)})) \rightarrow 0$ by using Definition 5—such that (30) satisfies the stationarity condition (24).

In the second step (15b), each $\bar{\mathbf{y}}_m^{(i+1)}$ minimizes the update in step-2 of the algorithm, *i.e.*,

$$\mathbf{0} \in \partial \mathcal{F}_m(\bar{\mathbf{y}}_m^{(i+1)}) + \rho \left(\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{x}}^{(i+1)} + \frac{\bar{\mathbf{z}}_m^{(i)}}{\rho} \right). \quad (31)$$

Now, using dual variable update (15c), (31) is $\mathbf{0} \in \partial \mathcal{F}_m(\bar{\mathbf{y}}_m^{(i+1)}) + \bar{\mathbf{z}}_m^{(i+1)}$ which always satisfies the stationarity condition (25) for sufficiently large iteration number $i \rightarrow \infty$.

Finally, primal feasibility (26) is satisfied by the assumption $\lim_{i \rightarrow +\infty} (\bar{\mathbf{y}}_m^{(i+1)} - \bar{\mathbf{x}}^{(i+1)}) = 0$.

APPENDIX C

BACKGROUND ON PROXIMAL, PROJECTIONS, AND LIPSCHITZ

We present the following definitions/theorems related to proximal, projections, and Lipschitz.

Definition 3 (proximal mapping [28], [31], [36]). *Given a proper closed convex function $f : \text{dom}_f \mapsto (-\infty, +\infty]$, then the proximal mapping of f is the operator given by*

$$\begin{aligned} \mathbf{prox}_{\lambda f}(\mathbf{X}) &:= (\mathbf{I} + \lambda \partial f)^{-1}(\mathbf{X}) \\ &= \arg \min_{\mathbf{Z} \in \text{dom}_f} \left\{ f(\mathbf{Z}) + \frac{1}{\beta \lambda} \|\mathbf{X} - \mathbf{Z}\|_F^2 \right\} \end{aligned}$$

for any $\mathbf{X} \in \text{dom}_f$, where dom_f corresponds to the domain of a function f , ∂f denotes the subdifferential of f , and $\lambda > 0$. If \mathbf{Z} is complex-valued or real-valued, $\beta = 1$ or $\beta = 2$, respectively.

Definition 4 (proximal mapping of the indicator function [31] [28]). *Let $f : \text{dom}_f \mapsto (-\infty, +\infty]$ be an indicator function, $f(\mathbf{x}) = \delta_C(\mathbf{x})$ is 0 if $\mathbf{x} \in C$ otherwise $+\infty$, where C is a nonempty set, then the proximal mapping of a given set C is an orthogonal projection operator, i.e.,*

$$\begin{aligned}\mathbf{prox}_{\lambda\delta_C}(\mathbf{x}) &= \arg \min_{\mathbf{z} \in \text{dom}_f} \left\{ \delta_C(\mathbf{x}) + \frac{1}{2\lambda} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{z} \in C} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\} = \mathbf{proj}_C(\mathbf{x}).\end{aligned}$$

Theorem 2 (projection onto the rank 1 quadratic constraint [19], [37]). *Let $C \subseteq \mathbb{C}^{N \times 1}$ and $C \neq \emptyset$ be given by $C = \{x \in \mathbb{C}^{N \times 1} : x^H \tilde{A} x - b \leq 0\}$, where $\mathbb{C}^{N \times N} \ni \tilde{A} = uu^H$ is rank one matrix and $b \in \mathbb{R}_{>0}$, then the proximal operator*

$$\begin{aligned} \mathbf{prox}_{x_c}(x) &= \mathbf{proj}_C(x) \\ &= \begin{cases} x + \left(\frac{\sqrt{b} - \|\mathbf{u}^H x\|}{\|\mathbf{u}\|_2^2 \|\mathbf{u}^H x\|} \right) \mathbf{u} (\mathbf{u}^H x), & \text{if } x^H \tilde{\mathbf{A}} x > b \\ x, & \text{if } x^H \tilde{\mathbf{A}} x < b. \end{cases} \end{aligned}$$

Theorem 3 (projection onto the affine set [28], [31]). *Let $S \subseteq \mathbb{C}^{N_T \times 1}$ and $S \neq \emptyset$ given by $S = \{\mathbf{x} \in \mathbb{C}^{N_T \times 1} : \mathbf{H}(\mathbf{x} - \mathbf{a}) = 0\}$, where $\mathbb{C}^{N_R \times N_T} \ni \mathbf{H}$ and $\mathbf{a} \in \mathbb{C}^{N_T \times 1}$, then $\text{prox}_{\chi_S}(\mathbf{x}) = \text{proj}_S(\mathbf{x}) = (\mathbf{I}_{N_T} - \mathbf{H}^+ \mathbf{H}) \mathbf{x} + \mathbf{H}^+ \mathbf{H} \mathbf{a}$, where \mathbf{H}^+ is Moore-Penrose pseudoinverse of \mathbf{H} .*

Theorem 4 (projection onto the rank ≥ 1 ellipsoid [37]). *Let $\mathcal{S} \subseteq \mathbb{C}^{N_T \times 1}$ and $\mathcal{S} \neq \emptyset$ given by $\mathcal{S} = \left\{ \mathbf{z} \in \mathbb{C}^{N_T \times 1} : \left\| \mathbf{Q}^{\frac{1}{2}} (\mathbf{z} - \mathbf{a}) \right\|_2^2 - b \leq 0 \right\}$, where $\text{rank}(\mathbf{Q}) = \text{rank}(\mathbf{Q}^{\frac{H}{2}} \mathbf{Q}^{\frac{1}{2}}) \geq 1$, $\mathbf{Q}^{\frac{1}{2}} \in \mathbb{C}^{N_T \times N_T}$, and $b \in \mathbb{R}_{\geq 0}$. If the eigenvalue decomposition of $\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_{N_T})$, and $\mathbf{D} = \text{Diag}\left(\frac{\lambda_1}{(1+\mu\lambda_1)^2}, \dots, \frac{\lambda_{N_T}}{(1+\mu\lambda_{N_T})^2}\right)$, then a (semi-closed-form) solution is $\mathbf{prox}_{\mathcal{S}}(\mathbf{z}) = \mathbf{proj}_{\mathcal{S}}(\mathbf{z}) = \mathbf{a} + \mathbf{U}(\mathbf{I} + \mu \mathbf{\Lambda})^{-1} \mathbf{U}^H (\mathbf{z} - \mathbf{a})$, where $\mu \geq 0$ is the unique positive root of $b - \mathbf{y}^H \mathbf{D} \mathbf{y} = b - \sum_{i=1}^{N_T} \frac{\lambda_i}{(1+\mu\lambda_i)^2} |\mathbf{y}[i]|^2 = 0$; $\mathbf{y} = \mathbf{U}^H \mathbf{z}$ —i.e., μ can be obtained via any root finding algorithm.*

Theorem 5. *If a function $f(\mathbf{X}) = \sum_{k=1}^n f_i(\mathbf{x}_k)$ is separable across the variables column-wise $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ or row-wise $\mathbf{X} = [\mathbf{x}_1; \dots; \mathbf{x}_{N_T}]$, then the respective prox operators can be shown as $\text{prox}_f(\mathbf{X}) = [\text{prox}_{f_1}(\mathbf{x}_1), \dots, \text{prox}_{f_n}(\mathbf{x}_n)]$ or $\text{prox}_f(\mathbf{X}) = [\text{prox}_{f_1}(\mathbf{x}_1); \dots; \text{prox}_{f_{N_T}}(\mathbf{x}_{N_T})]$.*

Proof. Following the proximal operator Definition 3, the minimization of the separable function is equivalent to minimization of respective functions $\{f_k\}$ independently [28], [31]. ■

Definition 5 (Appendix E, [55]). A differentiable function $f(\mathbf{x}) \in \mathbb{R}$, where $\mathbf{x} \in \mathbb{C}^n$, has L -Lipschitz continuous gradient (for $L \geq 0$) if $\|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\| \leq L \|\mathbf{x}_1 - \mathbf{x}_2\| \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$.

REFERENCES

- [1] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The next generation wireless access technology*. Academic Press, 2018.
- [2] R. Prasad, *OFDM for wireless communications systems*. Artech House, 2004.
- [3] 3GPP TS 38.104, “NR; Base station (BS) radio transmission and reception,” V15.9.0, 2020.
- [4] X. Huang, J. A. Zhang, and Y. J. Guo, “Out-of-band emission reduction and a unified framework for precoded OFDM,” *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 151–159, Jun. 2015.
- [5] M. Faulkner, “The effect of filtering on the performance of OFDM systems,” *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1877–1884, Sep. 2000.

- [6] E. Bala, J. Li, and R. Yang, "Shaping spectral leakage: A novel low-complexity transceiver architecture for cognitive radio," *IEEE Veh. Technol. Mag.*, vol. 8, no. 3, pp. 38–46, Jul. 2013.
- [7] H. Yamaguchi, "Active interference cancellation technique for MB-OFDM cognitive radio," in *Proc. IEEE Eur. Microw. Conf.*, 2004, pp. 1105–1108.
- [8] S. Brandes, I. Cosovic, and M. Schnell, "Sidelobe suppression in OFDM systems by insertion of cancellation carriers," in *Proc. IEEE Veh. Technol. Conf.*, 2005, pp. 152–156.
- [9] 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.*, vol. 59, no. 4, pp. 1689–1695, May 2010.
- [10] 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.*, vol. 67, no. 5, pp. 3541–3551, May 2019.
- [11] I. Cosovic, S. Brandes, and M. Schnell, "Subcarrier weighting: a method for sidelobe suppression in OFDM systems," *IEEE Commun. Lett.*, vol. 10, no. 6, pp. 444–446, Jun. 2006.
- [12] J. van de Beek, "Sculpting the multicarrier spectrum: A novel projection precoder," *IEEE Commun. Lett.*, vol. 13, no. 12, pp. 881–883, 2009.
- [13] H. Chen, W. Chen, and C. Chung, "Spectrally precoded OFDM and OFDMA with cyclic prefix and unconstrained guard ratios," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1416–1427, May 2011.
- [14] A. Tom, A. Sahin, and H. Arslan, "Mask compliant precoder for OFDM spectrum shaping," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 447–450, 2013.
- [15] R. Kumar and A. Tyagi, "Weighted least squares based spectral precoder for OFDM cognitive radio," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 641–644, Dec. 2015.
- [16] R. Kumar and A. Tyagi, "Computationally efficient mask-compliant spectral precoder for OFDM cognitive radio," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 2, no. 1, pp. 15–23, Jun. 2016.
- [17] M. Mohamad, R. Nilsson, and J. van de Beek, "A novel transmitter architecture for spectrally-precoded OFDM," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 65, no. 8, pp. 2592–2605, Aug. 2018.
- [18] 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.
- [19] S. Kant, M. Bengtsson, G. Fodor, B. Göransson, and C. Fischione, "EVM-constrained and mask-compliant MIMO-OFDM spectral precoding," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 590–606, 2021.
- [20] M. Mohamad, "Spectrally precoded OFDM design and analysis," PhD dissertation, Department of Computer Science, Electrical and Space Engineering, Luleå University, Luleå, Sweden, 2019.
- [21] Samsung, et al. R4-1709075, "Way forward on spectrum utilization," 3GPP TSG RAN WG4 #84 Meeting, Berlin, Germany, 21–25 Aug., 2017.
- [22] 3GPP RAN WG1, R1-1715184, "LS response on spectrum utilization," 3GPP TSG RAN WG1 #90 Meeting, Prague, Czech Republic, 21–25 Aug., 2017.
- [23] 3GPP TS 38.211, "NR; Physical channels and modulation," V15.8.0, 2020.
- [24] R. Pitaval and B. M. Popovi, "Linear receivers for spectrally-precoded MIMO-OFDM," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1269–1272, Jun. 2017.
- [25] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017.
- [26] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, no. 5–6, pp. 475–504, Oct. 2004.
- [27] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point algorithms for inverse problems in science and engineering*. Springer, New York, NY, 2011, pp. 185–212.
- [28] A. Beck, *First-order methods in optimization*. SIAM, 2017.
- [29] 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*, vol. 3, no. 1, pp. 1–122, 2011.
- [30] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.
- [31] N. Parikh, S. Boyd, N. Parikh, and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.
- [32] W. Deng, M. Lai, Z. Peng, and W. Yin, "Parallel multi-block ADMM with $o(1/k)$ convergence," *Journal of Scientific Computing*, vol. 71, no. 2, pp. 712–736, May 2017.
- [33] G. Stathopoulos, H. Shukla, A. Szűcs, Y. Pu, and C. N. Jones, "Operator splitting methods in control," *Foundations and Trends in Systems and Control*, vol. 3, no. 3, pp. 249–362, 2016.
- [34] R. Glowinski, S. J. Osher, and W. Yin, *Splitting methods in communication, imaging, science, and engineering*, ser. Scientific Computation. Cham, Switzerland: Springer International Publishing, 2016.
- [35] G. Pierra, "Decomposition through formalization in a product space," *Mathematical Programming*, vol. 28, no. 1, pp. 96–115, Jan. 1984.
- [36] D. Davis and W. Yin, "A three-operator splitting scheme and its optimization applications," *Set-Valued and Variational Analysis*, vol. 25, no. 4, pp. 829–858, Dec. 2017.
- [37] K. Huang and N. D. Sidiropoulos, "Consensus-ADMM for general quadratically constrained quadratic programming," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, Oct. 2016.
- [38] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," *RLE, MIT, Tech. Rep. 216*, 1952.
- [39] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7208–7221, Nov. 2018.
- [40] E. G. Larsson and L. Van Der Perre, "Out-of-band radiation from antenna arrays clarified," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 610–613, Aug. 2018.
- [41] N. N. Moghadam, P. Zetterberg, P. Händel, and H. Hjalmarsson, "Correlation of distortion noise between the branches of MIMO transmit antennas," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2012, pp. 2079–2084.
- [42] E. Sienkiewicz, N. McGowan, B. Göransson, T. Chapman, and T. Elfström, "Spatially dependent ACLR modelling," in *Proc. IEEE Conf. Antenna Meas. Appl. (CAMA)*, Nov. 2014, pp. 1–4.
- [43] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. ITG Workshop Smart Antennas*, Feb. 2010, pp. 189–196.
- [44] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [45] J. H. Johansson and A. Hansson, "An inexact interior-point method for system analysis," *International Journal of Control*, vol. 83, no. 3, pp. 601–616, Mar. 2010.
- [46] 3GPP TS 34.121, "Terminal conformance specification, radio transmission and reception (FDD)," V6.4.0, 2006.
- [47] 3GPP TS 38.306, "NR; User equipment (UE) radio access capabilities," V15.9.0, 2019.
- [48] 3GPP TS 38.101-4, "User equipment (UE) radio transmission and reception; part 4: Performance requirements," V15.5.0, 2020.
- [49] 3GPP TS 38.212, "NR; Multiplexing and channel coding," V15.8.0, 2020.
- [50] 3GPP TS 38.214, "NR; Physical layer procedures for data," V15.9.0, 2020.
- [51] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018.
- [52] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [53] Q. Liu, X. Ma, G. T. Zhou, and J. Wu, "Peak-to-average power ratio versus instantaneous-to-average power ratio for OFDM," in *Proc. ACSSC*, Oct. 2008, pp. 938–942.
- [54] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [55] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4–5, pp. 311–801, 2014.