

Joint Design of Beam Selection and Precoding Matrices for mmWave MU-MIMO Systems Relying on Lens Antenna Arrays

Rongbin Guo, Yunlong Cai, Minjian Zhao, Qingjiang Shi, Benoit Champagne, and Lajos Hanzo

Abstract—Wireless transmission relying on lens antenna arrays is becoming more and more attractive for millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems using a limited number of radio frequency (RF) chains due to the lens' energy-focusing capability. In this paper, we consider the joint design of the beam selection and precoding matrices in order to maximize the sum-rate of a downlink single-sided lens MU-MIMO mmWave system under transmit power constraints. We first formulate the optimization problem into a tractable form using the popular weighted minimum mean squared error (WMMSE) approach. To solve this problem, we then propose an efficient joint beam selection and precoding design algorithm based on the innovative penalty dual decomposition (PDD) method. To reduce the design complexity, we also propose a simplified algorithm by combining the interference-aware beam selection (IA-BS) scheme with the WMMSE approach. Simulation results demonstrate that our proposed algorithms can converge in a few iterations and achieve near-optimal performance when compared to the fully digital precoding scheme, thus enabling them to outperform the competing methods.

Index Terms—mmWave, MU-MIMO, lens antenna array, beam selection.

I. INTRODUCTION

Recently, due to the rapid consumer adoption of mobile devices such as tablets and smartphones, the tele-traffic has experienced a tremendous growth, resulting in a dramatic spectrum shortage at the centimeter-wave wireless frequencies [1]. As a remedy, communications over millimeter wave (mmWave) frequencies is expected to find its way into next generation cellular networks, allowing the latter to support unprecedented data rates in the 30-300GHz band [2], [3]. In [4], it has been demonstrated that a mmWave mobile broadband system could achieve gigabit per second data rates

The work of Y. Cai was supported in part by the National Natural Science Foundation of China under Grant 61471319 and in part by the Fundamental Research Funds for the Central Universities. The work of M. Zhao was supported by the National Natural Science Foundation of China under Grant 91538103. The work of Q. Shi was supported by the National Nature Science Foundation of China under Grant 61671411. (*Corresponding author: Yunlong Cai*)

R. Guo, Y. Cai and M. Zhao are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: rbguoieee@zju.edu.cn; ylcai@zju.edu.cn; mjzhao@zju.edu.cn).

Q. Shi is with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: qing.j.shi@gmail.com).

B. Champagne is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada (e-mail: benoit.champagne@mcgill.ca).

L. Hanzo is with the Department of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

at distances up to one kilometer in an urban environment, which indicates that mmWave communications can be used in future cellular networks for small-cell coverage. According to the fundamental Friss transmission theorem, mmWave signals suffer much higher free-space path loss than centimeter-wave for a given propagation distance and antenna gain. Fortunately, the significantly reduced wavelength makes it possible to realize massive MIMO systems [5]–[7] that use a large number of antennas within a small physical size, and yet achieve high array gain for directional communications by exploiting precoding techniques [8].

However, in the case of massive MIMO, the conventional fully digital precoding techniques lead to unaffordable costs in terms of RF chains and power consumption. To address this limitation, there have been extensive studies devoted to various cost-aware mmWave communication techniques, such as antenna selection [9]–[11], load-controlled parasitic antenna arrays (LC-PAAAs) [12], [13] and hybrid analog/digital processing [14]–[17]. Antenna selection techniques still need additional power in order to compensate for the attenuation of RF switches [9] and also lead to reduced average signal-to-noise ratios (SNR) [11]. LC-PAAAs exploit the mutual coupling among the antennas to reduce the number of required RF modules for given spatial degrees of freedom. However, this technique is hampered by the complex design of the parasitic antennas.

Hybrid analog/digital precoding has been proposed as an effective technique, allowing a trade-off between cost and performance by judiciously adjusting the number of required RF chains. In this structure, the signal processing is performed in two stages: digital precoding at the baseband with reduced output dimensionality, followed by analog beamforming through a phase shifter network in the RF-band, where the two processing modules are connected via a reduced number of RF chains to allow a trade-off. Nevertheless, a major limitation of hybrid analog/digital processing is the need for a large number of phase shifters, which leads to a significant increase in the power budget. Furthermore, the hybrid structure suffers from the constant-amplitude constraints on the analog beamformer weights. These limitations make the design of hybrid processing subsystems more complicated than that of their conventional fully digital counterparts.

To simplify the design of hybrid precoding, a number of studies have proposed the concept of *beam-space MIMO* based on discrete lens arrays (DLA) [18]. A typical DLA includes two main components: an electromagnetic (EM) lens and a matching antenna array whose elements are located in the

focal region of the lens. The fundamental principle of EM lenses is to provide variable phase shifts for the EM rays at different points on the lens aperture so as to achieve an angle-dependent energy focusing property. In effect, this approach can transform the traditional MIMO spatial channels into beamspace channels with angle-dependent energy-focusing capabilities [19]. In practice, only a small number of beams are needed due to the sparse nature of beamspace channels. Since each beam corresponds to a single RF chain, this effectively reduces the cost of RF chains in mmWave massive MIMO systems. Furthermore, the phase shifters needed in the hybrid structure are replaced by a switching network, which decreases the cost and complexity of the RF hardware as well as improving the power budget. In [20], beamspace MIMO is studied under multi-path propagation conditions, while its extensions to multiuser scenarios are considered in [21], [22]. Within this context, a critical problem of mmWave lens array systems is the design of effective beam selection and digital precoding schemes.

Recent studies [19], [21], [23], [24] of the beam selection problem for DLAs concentrate on choosing the beams with maximum magnitude (termed as “MM-BS” in the sequel) so as to obtain as much power from each user as possible. Subsequently, [25] considers the potential multiuser interference and proposes an interference-aware beam selection (termed as “IA-BS”) strategy, which outperforms the MM-BS schemes. However, all the aforementioned schemes are based on fixed digital precoding methods – such as zero forcing (ZF), minimum mean square error (MMSE) and maximum ratio combining (MRC), etc. – which might suffer from significant performance degradation, since the beam selection and precoding modules are designed separately.

Hence in this work, we consider the joint design of beam selection and precoding matrices with the aim of maximizing the sum-rate of a downlink single-sided lens mmWave MU-MIMO system using a limited number of RF chains. To this end, we first formulate the optimization problem into a mathematically tractable form by using the WMMSE approach of [26]. Then, to solve the resulting problem, we propose an efficient joint beam selection and precoding design algorithm based on the innovative PDD method [27]. Simulation results demonstrate that our proposed algorithm approaches the optimal performance of the fully digital precoding scheme and thus outperforms the competing methods.

Furthermore, since the IA-BS scheme of [25] achieves a good performance at a low complexity, we also design a simplified scheme where the beams are selected by considering the potential multiuser interference based on [25]. Subsequently, we optimize the precoding matrix for the selected beams by using the WMMSE approach for maximizing the system sum-rate. This simplified algorithm achieves a performance between that of the aforementioned MM-BS/IA-BS schemes using a fixed precoding matrix and the proposed PDD-based joint design algorithm, while its complexity is much less than the latter.

The specific contributions of this paper are summarized as follows.

- (1) For a downlink single-sided lens mmWave MU-MIMO

system, we propose an efficient joint beam selection and precoding design algorithm for maximizing the system sum-rate based on the PDD optimization method. To this end, we first introduce a number of auxiliary variables and equality constraints to handle the coupling power constraints and beam selection constraints. We then reformulate the optimization problem as an augmented Lagrangian (AL) problem by penalizing and dualizing the equality constraints. The AL problem is solved by partitioning the variables into independent blocks, thereby leading to various subproblems which are approached via the block coordinate descent (BCD) method [28]. The PDD-based algorithm exhibits an embedded double loop structure, where the inner loop serves to address the AL subproblems, while the outer loop aims to update the dual variables or penalty parameter based on the constraint violation. This algorithm ensures convergence to a stationarity point of the original problem.

- (2) We design a simplified algorithm which optimizes the precoding matrix based on beams selected by an IA-BS scheme, along with the aid of the WMMSE method. Subsequently, we analyze the computational complexities of the proposed algorithms and other existing schemes, showing in particular that our proposed simplified algorithm can achieve a practical trade-off between complexity and performance.
- (3) We provide exhaustive simulation results of the proposed algorithms for demonstrating their advantages over the competing methods.

The rest of this paper is organized as follows. Section II describes the downlink single-sided lens mmWave MU-MIMO system model and formulates the constrained sum-rate maximization problem. The proposed PDD-based joint beam selection and precoding design algorithm is developed in Section III, while the simplified algorithm is presented in Section IV. In Section V, we analyze the results of our experiments, while Section VII concludes the paper. The proof of Theorem 1 is relegated to Appendix A. Relevant background on the PDD method is given in Appendix B.

The notations used in this paper are summarized in Table I.

TABLE I: List of notations.

\mathbf{a}/\mathbf{A}	Vector \mathbf{a} /Matrix \mathbf{A}
a_{ij}	(i, j) -th element of \mathbf{A}
\mathbf{A}^T	Transpose of \mathbf{A}
\mathbf{A}^H	Conjugate transpose of \mathbf{A}
\mathbf{A}^{-1}	Inverse of \mathbf{A}
\mathbf{A}^\dagger	Pseudo-inverse of \mathbf{A}
$\text{Tr}(\mathbf{A})$	Trace of \mathbf{A}
$\ \mathbf{A}\ _2$	Frobenius norm of \mathbf{A}
$[\mathbf{X}]_{mm}$	m -th diagonal element of \mathbf{X}
$\text{Card}(\mathcal{A})$	Cardinality of set \mathcal{A}
\mathbf{I}_K	$K \times K$ identity matrix
$E\{\cdot\}$	Expectation operation
$\mathcal{A} \setminus \mathcal{B}$	Subset of elements from set \mathcal{A} that are not in \mathcal{B}
$\mathcal{CN}(m, \sigma^2)$	Complex circular Gaussian distribution with mean m and covariance σ^2

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the downlink of a wireless mmWave subnetwork where a base station (BS) serves K simultaneous users.

The BS is equipped with a single-sided DLA MU-MIMO transmitter system, comprising M_s antennas and N_{RF} RF chains, while the end users are equipped with single antenna receiver. To guarantee spatial multiplexing gain for the K users with this configuration, the number of RF chains should satisfy $N_{RF} \geq K$. Without loss of generality, we choose $N_{RF} = K$ in this paper.

A. System Model

As shown in Fig. 1, the BS transmits K data streams carrying independent messages, each intended for a specific user. The baseband data vector precoded at the BS can be expressed as

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \sum_{k=1}^K \mathbf{p}_k s_k, \quad (1)$$

where $\mathbf{s} = [s_1, s_2, \dots, s_K]^T$, s_k is the complex data symbol intended for user k with zero mean and normalized power of $\mathbb{E}\{|s_k|^2\} = 1$, with $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K] \in \mathbb{C}^{N_{RF} \times K}$ is the precoding matrix and \mathbf{p}_k is the digital precoding vector for user k . Under the narrow-band (flat fading) assumption, the $K \times 1$ received signal vector \mathbf{y} of all the K users can be expressed as

$$\mathbf{y} = \mathbf{H}^H \mathbf{F} \mathbf{P} \mathbf{s} + \mathbf{n}, \quad (2)$$

in which $\mathbf{H} \in \mathbb{C}^{M_s \times K}$ is the beamspace channel matrix, $\mathbf{F} \in \mathbb{C}^{M_s \times N_{RF}}$ is the beam selection matrix whose entries f_{ij} are either 0 or 1, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_K)$ is a $K \times 1$ additive zero-mean circular complex Gaussian noise vector with covariance matrix $E\{\mathbf{n}\mathbf{n}^H\} = \sigma^2 \mathbf{I}_K$, where σ^2 denotes the noise variance. The beamspace channel model is developed below.

B. Beamspace Channel Model

The beamspace channel matrix \mathbf{H} is obtained from the physical spatial MIMO channel by Fourier transformation:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K] = [\mathbf{U}\mathbf{g}_1, \mathbf{U}\mathbf{g}_2, \dots, \mathbf{U}\mathbf{g}_K], \quad (3)$$

where $\mathbf{U} \in \mathbb{C}^{M_s \times M_s}$ is a discrete Fourier transformation (DFT) matrix corresponding to a carefully designed DLA [21] at the BS and $\mathbf{g}_k \in \mathbb{C}^{M_s \times 1}$ is the spatial domain channel vector between the BS and user k . The DFT matrix \mathbf{U} consists of the array steering vectors of M_s orthogonal directions (beams) spread over the entire angular domain, i.e.:

$$\mathbf{U} = [\mathbf{a}(\varphi_1), \mathbf{a}(\varphi_2), \dots, \mathbf{a}(\varphi_{M_s})]^H, \quad (4)$$

in which $\varphi_m = \frac{1}{M_s}(m - \frac{M_s+1}{2})$ with $m = 1, 2, \dots, M_s$ are the normalized spatial directions [25], $\mathbf{a}(\varphi_m) = \frac{1}{\sqrt{M_s}} [e^{-j2\pi\varphi_m i}]_{i \in \mathcal{I}}$ are the corresponding $M_s \times 1$ array steering vectors, where $\mathcal{I} = \{j - (M_s - 1)/2 | j = 0, 1, \dots, M_s - 1\}$ is an index set of array elements. With this choice of steering vectors, the columns of matrix \mathbf{U} are orthonormal, i.e. $\mathbf{U}^H \mathbf{U} = \mathbf{I}$. The normalized spatial direction φ_m is related to the physical direction (angle) of propagation θ_m through the relation $\varphi_m = \frac{d}{\lambda} \sin \theta_m$, where λ is the signal wavelength at the operating mmWave frequency, and d is the antenna spacing, typically chosen as $d = \lambda/2$.

In this paper, we employ the well-known Saleh-Valenzuela channel model for the mmWave system considered [19], [21], [24], [25]:

$$\mathbf{g}_k = \beta_k^{(0)} \mathbf{a}(\phi_k^{(0)}) + \sum_{l=1}^L \beta_k^{(l)} \mathbf{a}(\phi_k^{(l)}), \quad (5)$$

where the terms $\beta_k^{(0)} \mathbf{a}(\phi_k^{(0)})$ and $\beta_k^{(l)} \mathbf{a}(\phi_k^{(l)})$ represent the line-of-sight (LoS) and the l -th non-line-of-sight (NLoS) channel vectors between the BS and user k , respectively. Furthermore, $\beta_k^{(0)}$ and $\beta_k^{(l)}$ represent the complex gains of the LoS and NLoS channels, whilst $\phi_k^{(0)}$ and $\phi_k^{(l)}$ denote the corresponding spatial directions. For convenience, we consider a 2D formulation, i.e., only the azimuth angle of departure (AoD) is considered¹, but the extension to a 3D scenario is straightforward and does not affect the nature of the problem. We assume that the beamspace channel matrix \mathbf{H} is perfectly estimated by the BS through the application of efficient channel estimation schemes [23].

The number of dominant scatters in a mmWave channel is quite limited [2]. Thus, the number of NLoS components L in (5) is typically much less than M_s , which implies that the channel matrix \mathbf{H} has a sparse structure.

C. Problem Formulation

Let us concentrate on the joint design of the digital precoding matrix \mathbf{P} and the beam selection matrix \mathbf{F} in order to maximize the downlink system sum-rate. The signal-to-interference-plus-noise ratio (SINR) of user k can be expressed as:

$$\gamma_k = \frac{|\mathbf{h}_k^H \mathbf{F} \mathbf{p}_k|^2}{\sum_{i \neq k}^K |\mathbf{h}_k^H \mathbf{F} \mathbf{p}_i|^2 + \sigma^2}. \quad (6)$$

Assuming that the user data symbols s_k are independent with identical distribution $\mathcal{CN}(0, 1)$, and that the noise vector \mathbf{n} and the data vector \mathbf{s} are independent, the achievable rate of user k can be written as $\mathcal{R}_k = \log(1 + \gamma_k)$. Then our sum-rate maximization problem can be mathematically formulated as

$$\max_{\mathbf{F}, \mathbf{P}} \sum_{k=1}^K \mathcal{R}_k \quad (7a)$$

$$\text{s.t. } \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \quad (7b)$$

$$\sum_{i=1}^{M_s} f_{ij} = 1, \sum_{j=1}^{N_{RF}} f_{ij} \leq 1, \quad (7c)$$

$$f_{ij} \in \{0, 1\}, \forall (i, j) \in \mathcal{S}, \quad (7d)$$

where P_s is the transmit power upper bound of the BS and $\mathcal{S} \triangleq \{(i, j) | i = 1, 2, \dots, M_s, j = 1, 2, \dots, N_{RF}\}$. Note that the identity $\mathbf{U}^H \mathbf{U} = \mathbf{I}_{M_s}$ was employed to simplify the constraint (7b). The constraints $\sum_{i=1}^{M_s} f_{ij} = 1, j = 1, 2, \dots, N_{RF}$ ensure that each RF chain feeds a single beam, while the constraints $\sum_{j=1}^{N_{RF}} f_{ij} \leq 1, i = 1, 2, \dots, M_s$ guarantee that each beam is selected for at most one RF chain. These constraints ensure that N_{RF} beams are selected to properly serve all K users.

¹For simplicity, we assume that the elevation AoDs are all zeros, which is practically valid if the height difference between the BS and users is much smaller than their separation distance.

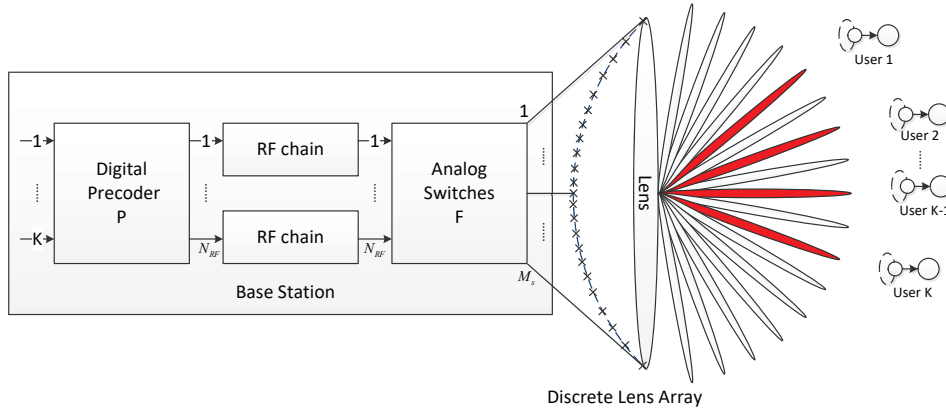


Fig. 1: A downlink mmWave MU-MIMO system with single-sided lens antenna array.

Note that problem (7) is nonconvex and very challenging to solve due to the beam selection constraints over a discrete space. In the next section, we propose an efficient joint beam selection and precoding design algorithm for solving problem (7).

III. THE PROPOSED PDD BASED ALGORITHM

We first use the WMMSE approach to transform problem (7) into a mathematically tractable form. We then propose an efficient joint beam selection and precoding design algorithm to solve the resulting equivalent problem based on the PDD method.

A. Reformulation of Problem (7)

Based on (2), the received signal of user k can be expressed as $y_k = \mathbf{h}_k^H \mathbf{F} \mathbf{P} \mathbf{s} + n_k = \sum_{i=1}^K \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i s_i + n_k$. Based on the observation of y_k , the estimated source signal by the same user is $\hat{s}_k = u_k y_k$, where u_k denotes the corresponding receiver. Under the independence assumption of s_k and n_k , the mean square error (MSE) of user k can be written as

$$\begin{aligned}
 e_k &\triangleq E\{|\hat{s}_k - s_k|^2\} = E\{|u_k y_k - s_k|^2\} \\
 &= E\left\{\left|\sum_{i=1}^K u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i s_i + u_k n_k - s_k\right|^2\right\}, \\
 &= E\{|u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k|^2 - u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k s_k^* - |s_k|^2\}, \\
 &+ s_k (u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k)^* + |u_k n_k|^2 + \sum_{i \neq k} |u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i|^2\}, \quad (8) \\
 &= |u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k|^2 - 2\Re\{u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k\}, \\
 &+ 1 + \sigma^2 |u_k|^2 + \sum_{i \neq k} |u_k \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i|^2,
 \end{aligned}$$

The following theorem establishes the equivalence between the sum-rate maximization problem (7) and a weighted sum-MSE minimization problem, formulated as (9).

Theorem 1: Let $w_k > 0, k = 1, \dots, K$ be a set of weighting

factor, then problem

$$\begin{aligned}
 \min_{\mathbf{F}, \mathbf{P}, \{w_k, u_k\}} & \sum_{k=1}^K w_k e_k - \log w_k \\
 \text{s.t.} & \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \sum_{i=1}^{M_s} f_{ij} = 1, \quad (9) \\
 & \sum_{j=1}^{N_{RF}} f_{ij} \leq 1, f_{ij} \in \{0, 1\}, \forall (i, j) \in \mathcal{S}
 \end{aligned}$$

is equivalent to problem (7), in the sense that the global optimal solution \mathbf{F}, \mathbf{P} for the two problems are identical. The proof of Theorem 1 is presented in the Appendix A.

Hence, the sum-rate maximization problem (7) can be solved via the weighted MSE minimization (9). The latter problem, which is defined over the parameter space $\mathbf{F}, \mathbf{P}, \{w_k, u_k\}$, is easier to handle since it has a convex objective function.

Before proceeding to the derivation of the proposed algorithm, let us reformulate problem (9) into an equivalent, yet mathematically more tractable form. To this end, we introduce the auxiliary variables $\{\hat{f}_{ij}\}$ and \mathbf{V} subject to the constraints of $f_{ij} = \hat{f}_{ij}, f_{ij}(1 - \hat{f}_{ij}) = 0, 0 \leq \hat{f}_{ij} \leq 1, \mathbf{F} \mathbf{P} = \mathbf{V}$ and $\text{Tr}(\mathbf{V} \mathbf{V}^H) \leq P_s$, in order to deal with the nonconvex and coupling constraints of problem (9). We also rewrite the constraint $\sum_{j=1}^{N_{RF}} f_{ij} \leq 1$ as $\mathbf{f}_i^T \mathbf{1} \leq 1$ with $\mathbf{f}_i^T \in \mathbb{C}^{1 \times N_{RF}}$ being the i -th row of \mathbf{F} and $\mathbf{1} \in \mathbb{C}^{N_{RF} \times 1}$ denoting a vector whose all elements are 1. Similarly, f_{ij} can be expressed as $\mathbf{f}_i^T \mathbf{b}_j$ with $\mathbf{b}_j \in \mathbb{C}^{N_{RF} \times 1}$ denoting the j -th column of $\mathbf{I}_{N_{RF}}$. Given these notations, problem (9) can be equivalently written as

$$\min_{\mathbf{V}, \mathbf{P}, \{w_k, u_k, \mathbf{f}_i, \hat{f}_{ij}\}} \sum_{k=1}^K w_k e_k - \log w_k \quad (10a)$$

$$\text{s.t. } \mathbf{F} \mathbf{P} = \mathbf{V}, \mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij}) = 0, \quad (10b)$$

$$\mathbf{f}_i^T \mathbf{b}_j = \hat{f}_{ij}, \sum_{i=1}^{M_s} \mathbf{f}_i^T \mathbf{b}_j = 1, \quad (10c)$$

$$\text{Tr}(\mathbf{V} \mathbf{V}^H) \leq P_s, \mathbf{f}_i^T \mathbf{1} \leq 1, \quad (10d)$$

$$0 \leq \hat{f}_{ij} \leq 1, \forall (i, j) \in \mathcal{S}. \quad (10e)$$

Problem (10) is defined over the parameter space

$\mathbf{V}, \mathbf{P}, \{w_k, u_k, \mathbf{f}_i, \hat{f}_{ij}\}$, optimizing each variable separately while holding the others fixed leads to a subproblem that can be easily solved (e.g., in closed-form). In the following, we develop the proposed algorithm based on the PDD method to solve problem (10).

B. The Proposed PDD Based Algorithm

In this subsection, we propose a PDD-based algorithm, which is characterized by an embedded double loop structure, where the inner loop solves the AL subproblem while the outer loop aims to update the dual variables or the penalty parameter based on the constraint violation. The conceptual framework of the PDD method is exposed in Appendix B.

We first convert problem (10) into its AL form as follows:

$$\begin{aligned} & \min_{\mathbf{V}, \mathbf{P}, \{f_i, \hat{f}_{ij}, w_k, u_k\}} \sum_{k=1}^K (w_k e_k - \log w_k) \\ & + \frac{1}{2\rho} \sum_{i=1}^{M_s} \sum_{j=1}^{N_{RF}} (\mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij}) + \rho \lambda_{ij})^2, \\ & + \frac{1}{2\rho} \sum_{i=1}^{M_s} \sum_{j=1}^{N_{RF}} (\mathbf{f}_i^T \mathbf{b}_j - \hat{f}_{ij} + \rho \hat{\lambda}_{ij})^2, \\ & + \frac{1}{2\rho} \sum_{j=1}^{N_{RF}} \left(\sum_{i=1}^{M_s} \mathbf{f}_i^T \mathbf{b}_j - 1 + \rho \mu_j \right)^2, \\ & + \frac{1}{2\rho} \|\mathbf{F}\mathbf{P} - \mathbf{V} + \rho \boldsymbol{\xi}\|_2^2, \\ & \text{s.t. } \text{Tr}(\mathbf{V}\mathbf{V}^H) \leq P_s, \mathbf{f}_i^T \mathbf{1} \leq 1, 0 \leq \hat{f}_{ij} \leq 1, \forall (i, j) \in \mathcal{S}, \end{aligned} \quad (11)$$

where $\{\lambda_{ij}\}, \{\hat{\lambda}_{ij}\}, \{\mu_j\}, \boldsymbol{\xi} \in \mathbb{C}^{M_s \times K}$ denote the dual variables associated with the equality constraints in (10b) and (10c) while ρ is a penalty factor. We address the AL problem (11) in the inner loop with the BCD method [28], where we partition the design variables into five blocks: $\{u_k\}, \{w_k\}, \{\hat{f}_{ij}, \mathbf{P}\}, \{\mathbf{V}\}$ and $\{\mathbf{f}_i\}$. In the following, we detail the methods used to solve the subproblems for each one of these blocks.

In **Step 1**, we optimize $\{u_k\}$ parallelly by fixing the remaining variables. In this case, (11) simplifies to the unconstrained problem:

$$\min_{\{u_k\}} \sum_{k=1}^K w_k e_k. \quad (12)$$

By examining the first-order optimality condition of (12) along with (8), we can obtain the closed-form solution of $\{u_k\}$ as:

$$u_k^{opt} = J_k^{-1} \mathbf{v}_k^H \mathbf{h}_k, \forall k, \quad (13)$$

where we have $J_k \triangleq \sum_{i=1}^K \mathbf{h}_k^H \mathbf{v}_i \mathbf{v}_i^H \mathbf{h}_k + \sigma^2$ and \mathbf{v}_i is the i -th column of matrix \mathbf{V} .

In **Step 2**, we optimize $\{w_k\}$ simultaneously by fixing the remaining variables. Following the approach in **Step 1**, we can obtain the closed-form solution of $\{w_k\}$:

$$w_k^{opt} = (1 - \mathbf{v}_k^H \mathbf{h}_k J_k^{-1} \mathbf{h}_k^H \mathbf{v}_k)^{-1}, \forall k. \quad (14)$$

In **Step 3**, we optimize $\{\hat{f}_{ij}\}, \mathbf{P}$ by fixing the remaining variables. The subproblem of optimizing $\{\hat{f}_{ij}\}$ parallelly can

be expressed as

$$\begin{aligned} & \min_{\hat{f}_{ij}} \frac{1}{2\rho} (\mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij}) + \rho \lambda_{ij})^2 \\ & + \frac{1}{2\rho} (\mathbf{f}_i^T \mathbf{b}_j - \hat{f}_{ij} + \rho \hat{\lambda}_{ij})^2, \\ & \text{s.t. } 0 \leq \hat{f}_{ij} \leq 1. \end{aligned} \quad (15)$$

Problem (15) features a scalar quadratic objective function of $\hat{f}_{ij}, \forall (i, j) \in \mathcal{S}$. We can obtain the unconstrained solution by enforcing the first-order optimality condition:

$$\hat{f}_{ij}^{oc} = \frac{f_{ij}^2 + f_{ij} \rho \lambda_{ij} + f_{ij} + \rho \hat{\lambda}_{ij}}{1 + f_{ij}^2}. \quad (16)$$

Recalling that \hat{f}_{ij} satisfies $0 \leq \hat{f}_{ij} \leq 1, \forall (i, j) \in \mathcal{S}$, we can obtain the optimal solution of the constrained problem (15) as follows:

$$\hat{f}_{ij}^{opt} = \begin{cases} 1, & 1 \leq \hat{f}_{ij}^{oc}, \\ \hat{f}_{ij}^{oc}, & 0 < \hat{f}_{ij}^{oc} < 1, \\ 0, & \hat{f}_{ij}^{oc} \leq 0. \end{cases} \quad (17)$$

Similarly, the subproblem of optimizing \mathbf{P} can be stated as

$$\min_{\mathbf{P}} \frac{1}{2\rho} \|\mathbf{F}\mathbf{P} - \mathbf{V} + \rho \boldsymbol{\xi}\|_2^2. \quad (18)$$

Since this is a quadratic programming problem without constraint, we can obtain the following closed-form solution by enforcing the first-order optimality condition of the objective function:

$$\mathbf{P}^{opt} = (\mathbf{F}^T \mathbf{F})^\dagger \mathbf{F}^T (\mathbf{V} - \rho \boldsymbol{\xi}). \quad (19)$$

As the variables \mathbf{P} and $\{\hat{f}_{ij}\}$ are uncoupled in the block, we should update them simultaneously.

In **Step 4**, we optimize \mathbf{V} by fixing the remaining variables. The subproblem of optimizing \mathbf{V} can be expressed as

$$\min_{\mathbf{V}} P_1(\mathbf{V}) \quad (20a)$$

$$\text{s.t. } \text{Tr}(\mathbf{V}\mathbf{V}^H) \leq P_s, \quad (20b)$$

where

$$\begin{aligned} P_1(\mathbf{V}) & \triangleq \sum_{k=1}^K w_k (|u_k \mathbf{h}_k^H \mathbf{V} \mathbf{b}_k|^2 - 2 \Re e(u_k \mathbf{h}_k^H \mathbf{V} \mathbf{b}_k) \\ & + 1 + \sigma^2 |u_k|^2 + \sum_{i \neq k} |u_k \mathbf{h}_k^H \mathbf{V} \mathbf{b}_i|^2), \\ & + \frac{1}{2\rho} \|\mathbf{F}\mathbf{P} - \mathbf{V} + \rho \boldsymbol{\xi}\|_2^2. \end{aligned} \quad (21)$$

This is a convex quadratic optimization subproblem subject to a quadratic constraint. By introducing the Lagrange multiplier λ for the constraint (20b), we define the Lagrangian function associated with problem (20) as follows

$$\mathcal{L}(\mathbf{V}, \lambda) \triangleq P_1(\mathbf{V}) + \lambda (\text{Tr}(\mathbf{V}\mathbf{V}^H) - P_s). \quad (22)$$

The first-order optimality condition of (22) with respect to \mathbf{V} yields:

$$\mathbf{V}^{opt} = \left(\sum_{k=1}^K w_k \mathbf{h}_k u_k^* u_k \mathbf{h}_k^H + \frac{1}{2\rho} \mathbf{I} + \lambda \mathbf{I} \right)^{-1} \mathbf{A}, \quad (23)$$

where we have $\mathbf{A} \triangleq \left[\frac{1}{2\rho}(\mathbf{F}\mathbf{P} + \rho\xi) + \sum_{k=1}^K w_k u_k^* \mathbf{h}_k \mathbf{b}_k^H \right]$, and $\lambda \geq 0$ should be chosen so that the complementarity slackness condition of the power budget constraint is satisfied. If the solution \mathbf{V} associated with $\lambda = 0$ satisfies the constraint (20b), the optimal λ (denoted by λ^{opt}) is zero. Otherwise we can obtain the solution of λ through the slackness condition:

$$\text{Tr}(\mathbf{V}\mathbf{V}^H) - P_s = 0, \quad (24)$$

which is equivalent to

$$\text{Tr}((\mathbf{\Lambda} + \lambda\mathbf{I})^{-2}\mathbf{\Theta}) = P_s, \quad (25)$$

where $\mathbf{D}\mathbf{\Delta}\mathbf{D}^H$ is the eigen-decomposition of $\sum_{k=1}^K w_k \mathbf{h}_k u_k^* u_k \mathbf{h}_k^H + \frac{1}{2\rho}\mathbf{I}$ and $\mathbf{\Theta} \triangleq \mathbf{D}^H \mathbf{A} \mathbf{A}^H \mathbf{D}$. (25) can also be expressed as

$$\sum_{m=1}^{M_s} \frac{[\mathbf{\Theta}]_{mm}}{([\mathbf{\Delta}]_{mm} + \lambda)^2} = P_s. \quad (26)$$

Note that λ^{opt} must be positive in this case and the left-hand side of (26) is a monotonously decreasing function of λ for $\lambda > 0$. Hence, we can find λ^{opt} by solving problem (26) with one dimensional search techniques such as the bisection method [29]. Finally, we can obtain the solution of \mathbf{V} by substituting λ^{opt} into (23).

In **Step 5**, we optimize $\{\mathbf{f}_i\}$ by fixing the remaining variables. The subproblem of optimizing $\{\mathbf{f}_i\}$ can be expressed as

$$\min_{\{\mathbf{f}_i\}} P_2(\mathbf{f}_i) \quad (27a)$$

$$\text{s.t. } \mathbf{f}_i^T \mathbf{1} \leq 1, i = 1, 2, \dots, M_s, \quad (27b)$$

where we have:

$$\begin{aligned} P_2(\mathbf{f}_i) &\triangleq \frac{1}{2\rho} \sum_{i=1}^{M_s} \sum_{j=1}^{N_{RF}} (\mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij}) + \rho \lambda_{ij})^2 \\ &+ \frac{1}{2\rho} \sum_{i=1}^{M_s} \sum_{j=1}^{N_{RF}} (\mathbf{f}_i^T \mathbf{b}_j - \hat{f}_{ij} + \rho \hat{\lambda}_{ij})^2, \\ &+ \frac{1}{2\rho} \sum_{j=1}^{N_{RF}} \sum_{i=1}^{M_s} (\sum_{i=1}^{M_s} \mathbf{f}_i^T \mathbf{b}_j - 1 + \rho \mu_j)^2, \\ &+ \frac{1}{2\rho} \|\mathbf{F}\mathbf{P} - \mathbf{V} + \rho\xi\|_2^2. \end{aligned} \quad (28)$$

We find that the subproblems associated with each \mathbf{f}_i are convex with an affine constraint. Similar to (20), each variable can be optimized based on the Lagrangian multiplier method. Therefore, the first-order optimality condition of the Lagrangian function of (27) with respect to \mathbf{f}_i yields:

$$\begin{aligned} \mathbf{f}_i^{opt} &= \left(3\mathbf{I} + \sum_{j=1}^{N_{RF}} (\hat{f}_{ij}^2 - 2\hat{f}_{ij}) \mathbf{b}_j \mathbf{b}_j^T + \Re\{\mathbf{P}\mathbf{P}^H\} \right)^{-1} \\ &\times \left[-\Re\{\mathbf{c}_i\} - \rho\theta_i \mathbf{1} - \sum_{j=1}^{N_{RF}} (\rho \lambda_{ij} (1 - \hat{f}_{ij}) \right. \\ &\left. + (\rho \hat{\lambda}_{ij} - \hat{f}_{ij}) + \left(\sum_{k \neq i}^{M_s} \mathbf{f}_k^T \mathbf{b}_j + \rho \mu_j - 1 \right) \mathbf{b}_j \right], \end{aligned} \quad (29)$$

where $\theta_i \geq 0$ is the Lagrangian multiplier, $\mathbf{C} \triangleq \mathbf{P}(\rho\xi - \mathbf{V})^H$ and \mathbf{c}_i denotes the i -th column of \mathbf{C} . If the solution of \mathbf{f}_i

TABLE II: One-iteration BCD method for solving (27)

-
- | |
|--|
| 1 For $i=1$ to M_s |
| 1.1 Determine θ_i from (31). |
| 1.2 Update \mathbf{f}_i based on (29). |
| 1.3 Assign \mathbf{f}_i^T to the i -th row of \mathbf{F} . |
-

associated with $\theta_i = 0$ satisfies constraint (27b), the optimal θ_i (denoted by θ_i^{opt}) is zero. Otherwise we can obtain θ_i from the slackness condition:

$$\mathbf{f}_i^T \mathbf{1} = 1. \quad (30)$$

Let us define $\mathbf{R} \triangleq (3\mathbf{I} + \sum_{j=1}^{N_{RF}} (\hat{f}_{ij}^2 - 2\hat{f}_{ij}) \mathbf{b}_j \mathbf{b}_j^T + \Re\{\mathbf{P}\mathbf{P}^H\})^{-1}$ and $\mathbf{d} \triangleq [-\Re\{\mathbf{c}_i\} - \sum_{j=1}^{N_{RF}} (\rho \lambda_{ij} (1 - \hat{f}_{ij}) - (\rho \hat{\lambda}_{ij} - \hat{f}_{ij}) - (\sum_{k \neq i}^{M_s} \mathbf{f}_k^T \mathbf{b}_j + \rho \mu_j - 1)) \mathbf{b}_j]$. Then (29) can be rewritten as $\mathbf{f}_i^{opt} \triangleq \mathbf{R}(\mathbf{d} - \rho\theta_i \mathbf{1})$ and we can then directly obtain a closed-form solution of θ_i by substituting \mathbf{f}_i^{opt} into (30):

$$\theta_i = \frac{\sum_{i=1}^{N_{RF}} \sum_{k=1}^{N_{RF}} d_k r_{ki} - 1}{\sum_{i=1}^{N_{RF}} \sum_{k=1}^{N_{RF}} \rho r_{ki}}, \quad (31)$$

where r_{ki} denotes the (k, i) -th element of \mathbf{R} and d_k is the k -th element of \mathbf{d} . Finally, this subproblem can be solved by using the algorithm in Table II, which can be viewed as a one-iteration BCD method.

Besides, in the outer iteration of the PDD method, the dual variables $\{\lambda_{ij}, \hat{\lambda}_{ij}, \mu_j, \xi\}^m$ can be updated by the following expressions:

$$\lambda_{ij}^{m+1} = \lambda_{ij}^m + \frac{1}{\rho^m} (\mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij})), \forall (i, j) \in \mathcal{S}, \quad (32a)$$

$$\hat{\lambda}_{ij}^{m+1} = \hat{\lambda}_{ij}^m + \frac{1}{\rho^m} (\mathbf{f}_i^T \mathbf{b}_j - \hat{f}_{ij}), \forall (i, j) \in \mathcal{S}, \quad (32b)$$

$$\mu_j^{m+1} = \mu_j^m + \frac{1}{\rho^m} \left(\sum_{i=1}^{M_s} (\mathbf{f}_i^T \mathbf{b}_j - 1) \right), \forall (i, j) \in \mathcal{S}, \quad (32c)$$

$$\xi^{m+1} = \xi^m + \frac{1}{\rho^m} (\mathbf{F}\mathbf{P} - \mathbf{V}), \quad (32d)$$

where m denotes the outer iteration index. We define the constraint violation indicator h as:

$$h = \max_{\forall (i,j) \in \mathcal{S}} \left\{ \begin{array}{l} |\mathbf{f}_i^T \mathbf{b}_j (1 - \hat{f}_{ij})|, |\mathbf{f}_i^T \mathbf{b}_j - \hat{f}_{ij}|, \\ \left| \sum_{i=1}^{M_s} (\mathbf{f}_i^T \mathbf{b}_j - 1) \right|, \|\mathbf{F}\mathbf{P} - \mathbf{V}\|_2. \end{array} \right\}. \quad (33)$$

The proposed joint design algorithm based on the PDD method for problem (10) is summarized in Table III, where we set I_n, ρ, c, η , and τ empirically. The algorithm exhibits an embedded double loop structure, where the inner loop solves the AL subproblems while the outer loop updates the dual variables and penalty parameter. Following [27], it can be shown that the proposed PDD-based algorithm converges

to the set of stationary solutions² of problem (10). Relevant background on the PDD method is given in Appendix B.

TABLE III: The proposed PDD-based algorithm

-
- 1 Initialize dual variables $\{\lambda_{ij}, \hat{\lambda}_{ij}, \mu_j, \xi_j\}^0$, primal variables $\{\mathbf{V}, \mathbf{P}, \mathbf{f}_i, \hat{f}_{ij}, w_k, u_k\}$, $\epsilon, I_n, \rho^0 > 0, \tau_0, c$ and $\eta^0 = \tau_0^{1/6}$. Set $m = 0$.
 - 2 **Repeat**
 - 2.1 **Repeat**
 - 2.1.1 Update $\{u_k\}$ by (13) with fixed $\{\mathbf{V}, \mathbf{P}, \mathbf{f}_i, \hat{f}_{ij}, w_k\}$.
 - 2.1.2 Update $\{w_k\}$ by (14) with fixed $\{\mathbf{V}, \mathbf{P}, \mathbf{f}_i, \hat{f}_{ij}, u_k\}$.
 - 2.1.3 Update $\{\hat{f}_{ij}, \mathbf{P}\}$ by (17), (19) with fixed $\{\mathbf{V}, \mathbf{f}_i, u_k, w_k\}$.
 - 2.1.4 Update $\{\mathbf{V}\}$ by (23) with fixed $\{\mathbf{f}_i, \mathbf{P}, \hat{f}_{ij}, u_k, w_k\}$.
 - 2.1.5 Update $\{\mathbf{f}_i\}$ by (29) with fixed $\{\mathbf{V}, \mathbf{P}, \hat{f}_{ij}, u_k, w_k\}$.
 - 2.2 **Until** the termination criterion is met.
 - 2.3 Calculate the constraint violation h by (33).
 - 2.4 **if** $h \leq \eta^m$ **then** update dual variables by (32),
 - 2.5 **else** set $\rho^{m+1} = c\rho^m$ **end if**.
 - 2.6 Set $\tau_{m+1} = 0.6\tau_m$, $\eta^{m+1} = \tau_{m+1}^{1/6}$ and $m = m+1$.
 - 3 **Until** $h < \epsilon$.
-

IV. THE SIMPLIFIED ALGORITHM

The PDD-based algorithm proposed in the previous section requires a joint optimization of the beam selection and precoding matrices. In this section, we present a simplified suboptimal algorithm which designs the beam selection and precoding matrices separately to achieve lower complexity. We first select the beams by using an IA-BS scheme [25] which takes into account the potential multiuser interference. Then, we obtain an optimal precoding matrix with the aid of WMMSE method to maximize the system sum-rate.

A. Beam Selection Algorithm

The beam selection algorithm first classifies all users into two groups, i.e., the interfering users (IUs) and non-interfering users (NIUs). For the NIUs, beams with large power are selected, while for IUs, the beams are selected by a low-complexity incremental algorithm based on the sum-rate maximization criterion. The beam selection process can be naturally divided into two stages:

- (1) First, we sort the elements (beams) of the beamspace channel \mathbf{h}_k in a descending order of magnitude, and denote the strongest beam index of user k as

²In the absence of a better alternative, it is readily seen that our proposed PDD-based algorithm is currently the best choice for solving problem (7). Due to the NP-hard nature of the problem, it does not seem possible to provide a globally optimal solution based on the current optimization techniques. At the present time, providing stationary point convergence for the extended, continuous type problem (10) is the best we can do in terms of convergence analysis for this kind of constrained optimization problem.

$b_k^* \in \{1, 2, \dots, M_s\}$. We identify user k as a NIU if its strongest beam b_k^* differs from the strongest beams of any other users, i.e. we have $b_k^* \notin \{b_1^*, \dots, b_{k-1}^*, b_{k+1}^*, \dots, b_K^*\}$. The group of all NIUs is denoted by \mathcal{G}_{NIU} . For a NIU, we directly select the strongest beam b_k^* , since this beam not only contains most of the power of the beamspace channel but also causes little interference to others.

- (2) Alternatively, we identify user k as an IU if its strongest beam coincides with that of another user, i.e., $b_k^* \in \{b_1^*, \dots, b_{k-1}^*, b_{k+1}^*, \dots, b_K^*\}$, and denote by \mathcal{G}_{IU} as the set of all IUs. For the IUs, we search a number $\text{Card}(\mathcal{G}_{IU})$ of appropriate beams from the set $\{1, 2, \dots, M_s\} \setminus \{b_k^* | k \in \mathcal{G}_{NIU}\}$ one by one. Specifically, in each step, the beam making the greatest contribution to the system sum-rate is selected. Additional details about this stage can be found in [25].

B. Optimization of the Precoding Matrix

The IA-BS scheme considers the potential multiuser interferences in the mmWave massive MIMO system. However, it suffers from performance degradation without the optimal digital precoding matrix based on the selected beams. This motivates us to optimize the precoding matrix \mathbf{P} . Once the K beams have been selected, we can form the beam selection matrix \mathbf{F} . Then the sum-rate maximization problem (7) can be reformulated as

$$\begin{aligned} \max_{\mathbf{P}} \quad & \sum_{k=1}^K \log(1 + \gamma_k) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s. \end{aligned} \quad (34)$$

We can solve this problem with the WMMSE approach. Specifically, and similar to (9), problem (34) can be equivalently recast as

$$\begin{aligned} \min_{\mathbf{P}, w_k, u_k} \quad & \sum_{k=1}^K w_k e_k - \log w_k \\ \text{s.t.} \quad & \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \end{aligned} \quad (35)$$

where u_k , w_k , and e_k are defined in the same way as before. We can then use the BCD method to solve (35) where we minimize the weighted sum-MSE cost function of (34) by sequentially fixing two of the three variables u_k , w_k , \mathbf{P} and then updating the third. The updates of u_k and w_k are obtained in closed forms which have been shown in **Step 1** and **Step 2** of the PDD-based algorithm in Section III-B. The subproblem of optimizing \mathbf{P} can be expressed as

$$\min_{\mathbf{P}} P_3(\mathbf{P}) \quad (36a)$$

$$\text{s.t.} \quad \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \quad (36b)$$

where we have

$$\begin{aligned} P_3(\mathbf{P}) \triangleq & \sum_{k=1}^K w_k (|u_k \mathbf{h}_k^H \mathbf{F} \mathbf{P} \mathbf{b}_k|^2 - 2\Re\{u_k \mathbf{h}_k^H \mathbf{F} \mathbf{P} \mathbf{b}_k\}) \\ & + 1 + \sigma^2 |u_k|^2 + \sum_{i \neq k} |u_k \mathbf{h}_k^H \mathbf{F} \mathbf{P} \mathbf{b}_i|^2. \end{aligned} \quad (37)$$

TABLE IV: The WMMSE algorithm proposed for solving problem (35)

-
- 1 Initialize \mathbf{P} such that $\text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) = P_s$. Set $\varrho > 0$.
 - 2 **Repeat**
 - 2.1 Store w_k into $w'_k, \forall k = 1, 2, \dots, K$.
 - 2.2 Update $\{u_k\}$ by (13) with fixed $\{\mathbf{P}, w_k\}, \forall k = 1, 2, \dots, K$.
 - 2.3 Update $\{w_k\}$ by (14) with fixed $\{\mathbf{P}, u_k\}, \forall k = 1, 2, \dots, K$.
 - 2.4 Update $\{\mathbf{P}\}$ by (39) with fixed $\{u_k, w_k\}$.
 - 3 **Until** $|\sum_{k=1}^K \log(w_k) - \sum_{k=1}^K \log(w'_k)| < \varrho$.
-

This is a convex quadratic optimization subproblem. Similar to **Step 4** of the PDD-based algorithm, by introducing the Lagrange multiplier δ for constraint (36b), we define the Lagrangian function associated with problem (36) as follows

$$\mathcal{L}(\mathbf{P}, \delta) \triangleq P_3(\mathbf{P}) + \delta(\text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) - P_s). \quad (38)$$

The first-order optimality condition of (38) with respect to \mathbf{P} when $\delta \geq 0$ yields:

$$\begin{aligned} \mathbf{P}^{opt} = & \left(\sum_{k=1}^K w_k \mathbf{F}^T \mathbf{h}_k u_k^* u_k \mathbf{h}_k^H \mathbf{F} + \delta \mathbf{F}^T \mathbf{F} \right)^{-1} \\ & \times \left(\sum_{k=1}^K w_k \mathbf{F}^T \mathbf{h}_k u_k b_k^H \right). \end{aligned} \quad (39)$$

If the solution \mathbf{P} with $\delta = 0$ satisfies the constraint (36b), the optimal δ (denoted by δ^{opt}) is zero. Otherwise we can obtain the solution of δ through the slackness condition:

$$\text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) - P_s = 0. \quad (40)$$

Then, we can find δ^{opt} by solving the problem (40) with the bisection method [29]. Finally, we can obtain the solution of \mathbf{P} by substituting δ^{opt} into (39). The WMMSE algorithm for problem (35) is summarized in Table IV. It is shown in [26] that the WMMSE algorithm is guaranteed to converge to a stationary point of problem (34).

V. COMPUTATIONAL COMPLEXITY

In Sections III and IV, we proposed a PDD-based algorithm and a simplified algorithm for solving problem (7), respectively. In this section, we compare the computational complexity of the proposed algorithms by evaluating the number of required multiplications per symbol time instance.

Let us focus on the proposed PDD-based joint design algorithm. The inner loop in Table III aims to solve the AL subproblems. Specifically, in each inner iteration, we solve the subproblems for the five blocks of variables in five respective steps:

In **Step 1**, the calculation of $u_k^{opt}, \forall k$ in (13) and $J_k, \forall k$ need $O(M_s K)$ and $O(M_s K^2)$ multiplications, respectively; thus the complexity of **Step 1** is

$$C1 \triangleq O(M_s K^2). \quad (41)$$

In **Step 2**, we only need $C2 \triangleq O(K)$ multiplications to solve w_k^{opt} in (14), since J_k and $\mathbf{v}_k^H \mathbf{h}_k$ have already been calculated in **Step 1**.

In **Step 3**, the complexity for solving $\hat{\mathbf{F}}$ is $O(M_s N_{RF})$, and the complexity of solving \mathbf{P} is

$$O(M_s N_{RF}^2 + M_s N_{RF} K + N_{RF}^3), \quad (42)$$

since the complexity for the matrix pseudo-inversion of $\mathbf{F}^T \mathbf{F}$ is $O(N_{RF}^3)$ [30].

In **Step 4**, when utilizing the bisection method to find the Lagrangian parameter λ in (26), the number of iterations is $\log_2(\frac{Q_{0,s}}{Q_s})$, where $Q_{0,s}$ is the initial interval size and Q_s is the tolerance, since we need to perform M_s multiplications in each iteration, the complexity of solving \mathbf{V} is

$$O(M_s^3 + M_s^2 K + M_s N_{RF} K + M_s \log(\frac{Q_{0,s}}{Q_s})). \quad (43)$$

In **Step 5**, the complexity of solving all \mathbf{f}_i is

$$O(M_s (M_s N_{RF} K + N_{RF}^3 + N_{RF}^2 K)). \quad (44)$$

Finally, the complexity of each inner iteration is the sum of these five steps:

$$\begin{aligned} C_{in} \triangleq & O(M_s^2 N_{RF} K + M_s N_{RF}^3 + M_s N_{RF}^2 K, \\ & + M_s^3 + M_s K^2 + M_s \log(\frac{Q_{0,s}}{Q_s})). \end{aligned} \quad (45)$$

It is of interest to investigate the asymptotic complexity of the proposed algorithms when M_s, N_{RF} and K are large, i.e., when we let $M_s \geq N_{RF} \geq K \rightarrow \infty$. Under this condition, we can obtain the asymptotic complexity of each inner iteration as $C_{ain} = O(M_s^2 N_{RF} K)$. Based on the above analysis, we can obtain the asymptotic complexity of the proposed PDD-based algorithm as:

$$C_{aPDD} = I_m I_n C_{ain} = O(I_m I_n M_s^2 N_{RF} K), \quad (46)$$

where I_m and I_n are the numbers of iterations in the outer and inner loops, respectively.

Similarly, as shown in Table IV, the complexity of the simplified algorithm is dominated by solving problem (35) I_w times, where I_w denotes the number of iterations. We note that in each iteration the solution of u_k and w_k are the same as that of the PDD-based algorithm, and the complexity of solving \mathbf{P} is

$$\begin{aligned} C_p \triangleq & O(M_s N_{RF}^2 + M_s N_{RF} K + N_{RF}^3, \\ & + N_{RF}^2 K + N_{RF} \log(\frac{Q_{0,s}}{Q_s})). \end{aligned} \quad (47)$$

In the beam selection stage, the complexity is $C_{bs} \triangleq O(K^3)$. Hence, the complexity of the simplified algorithm is

$$\begin{aligned} & I_w (C1 + C2 + C_p) + C_{bs} \\ & = O[I_w (M_s N_{RF}^2 + M_s N_{RF} K + M_s K^2 + N_{RF}^3, \\ & + N_{RF}^2 K + N_{RF} \log(\frac{Q_{0,s}}{Q_s})) + K^3]. \end{aligned} \quad (48)$$

Under same conditions $M_s \geq N_{RF} \geq K \rightarrow \infty$, we can obtain the asymptotic complexity of this algorithm as $C_{aSIM} = O(I_w M_s N_{RF}^2)$. The proposed PDD-based algorithm

TABLE V: Computational complexities of the proposed algorithms and other schemes

Algorithms	Asymptotic complexity
PDD based algorithm	$O(I_m I_n M_s^2 N_{RF} K)$
Simplified algorithm	$O(I_w M_s N_{RF}^2)$
Fully digital precoding based on PDD	$O(I_m I_n M_s^3 K)$
IA-BS	$O(K^3)$
MM-BS	$O(K^3)$

provides a higher computational complexity than the simplified algorithm, as a price for better performance.

Besides, in Table V, we summarize the asymptotic complexities of the proposed PDD-based algorithm, the simplified algorithm and other schemes used for mmWave MIMO systems design, namely: fully digital precoding (based on the PDD algorithm), IA-BS and MM-BS (the complexities of these schemes can be analyzed in a similar way). As seen from Table V, the fully digital precoding based on the PDD algorithm consumes more computational resources than other schemes, since it needs the same number of RF chains as antenna elements (which also leads to excessive hardware costs). In contrast, the proposed algorithms achieve lower complexity but can still maintain performance levels that are close to the fully digital precoding, as will be shown in the next section. While the complexity of the IA-BS and MM-BS scheme is generally lower than the proposed algorithms, their performance is in general not as good. Consequently, our proposed algorithms therefore offer a practical trade-off between complexity and performance.

VI. SIMULATION RESULTS

In order to evaluate the performance of the proposed algorithms, numerical results have been obtained by performing computer simulations. In this section, we present the performance of the PDD-based algorithm and the simplified algorithm, and compare them with other schemes from the literature. The system configuration is defined by the following choice of parameters: the BS is equipped with a DLA consisting of $M_s = 128$ antennas and $N_{RF} = 16$ RF chains to serve $K = 16$ users. The channel model parameters of user k are set according to [21]: 1) one LoS link and $L = 2$ NLoS links; 2) $\phi_k^{(0)}$ and $\phi_k^{(l)}$ obey the uniform distribution within $[-\frac{1}{2}, \frac{1}{2}]$; 3) $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$, $\beta_k^{(l)} \sim \mathcal{CN}(0, 10^{-1})$ with $l = 1, 2$; 4) the parameters $\phi_k^{(0)}$, $\phi_k^{(l)}$, $\beta_k^{(0)}$ and $\beta_k^{(l)}$ ($l = 1, 2$) are statistically independent. For the PDD-based algorithm, we set the initial penalty factor $\rho^0 = 10^{-2}$ and the control parameter $c = 0.1$. In addition, we set $\tau^0 = 1$ and the maximum number of inner iterations $I_n = 100$, which is large enough to observe convergence. For the simplified algorithm, we set the threshold of the WMMSE iteration $\varrho = 10^{-4}$ to guarantee convergence.

We first present the convergence performance of the proposed PDD-based algorithm and the simplified algorithm. Without loss of generality, we choose SNR = 25dB to present the convergence performance of the PDD-based algorithm in Fig. 2 and Fig. 3. The achievable system sum-rate converges rapidly in less than 5 outer iterations, and the constraint violation h reduces to a threshold $\epsilon = 10^{-7}$ in less than 200

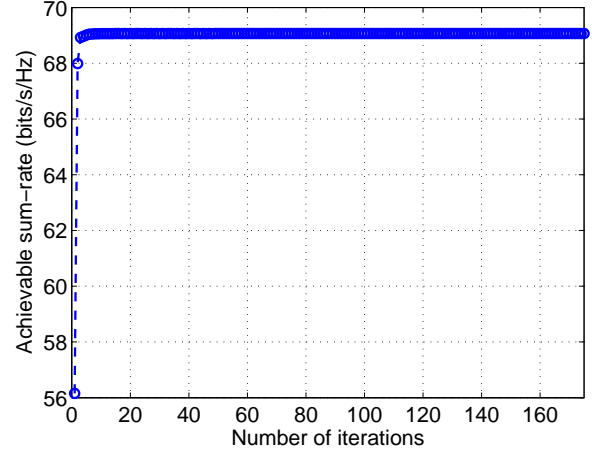


Fig. 2: Convergence performance of the PDD-based algorithm (SNR = 25dB).

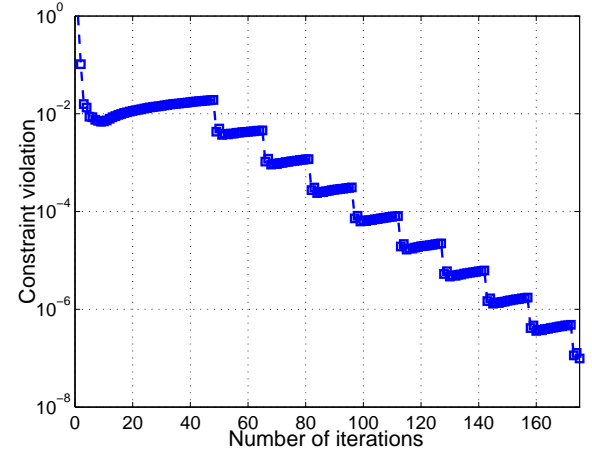


Fig. 3: Convergence performance of the PDD-based algorithm (SNR = 25dB).

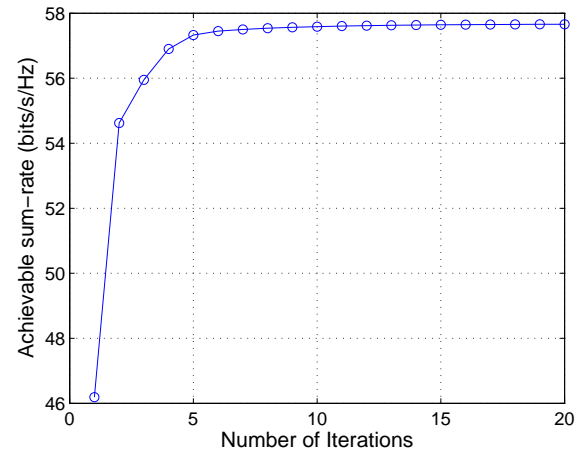


Fig. 4: Convergence performance of the simplified algorithm (SNR = 25dB).

outer iterations, which means that the solution has essentially

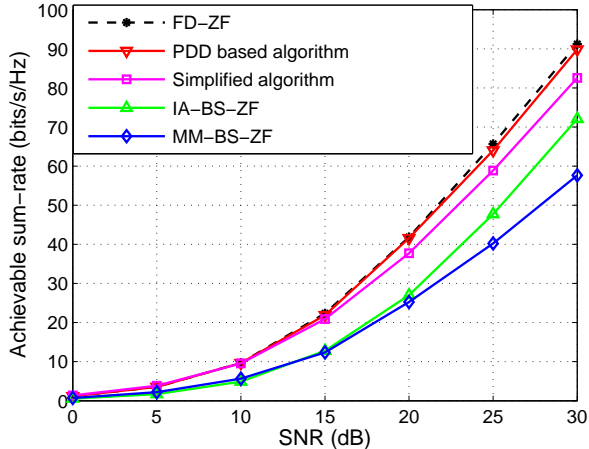


Fig. 5: Achievable sum-rate comparison versus the SNR.

met the equality constraints for problem (10). Fig. 4 illustrates the convergence performance of the simplified algorithm for the case of SNR = 25dB, which shows that the simplified algorithm converges monotonically in a few steps.

Fig. 5 compares the system sum-rate of the proposed PDD-based and simplified algorithms with the IA-BS [25] and MM-BS [19], [21], [24] schemes. The sum-rate is averaged over 100 channel realizations, while both the IA-BS and MM-BS schemes employ a ZF precoding matrix. We also simulate the fully digital ZF precoding scheme as a benchmark (labeled by “FD-ZF” in Fig. 5). From the results, we can see that the best performance is achieved by the fully digital precoding scheme, followed by the proposed PDD-based algorithm, the simplified algorithm, the conventional IA-BS scheme and the MM-BS scheme. In particular, the PDD-based algorithm can lead to a 3.5dB gain in SNR in comparison with the conventional IA-BS scheme at the system sum-rate level of 40bits/s/Hz. This demonstrates the merits of the proposed joint design algorithm for the beam selection and digital precoding matrices. We conclude that the PDD-based algorithm offers an efficient and attractive means of solution for problem (7). Besides, the simplified algorithm can achieve a sub-optimal performance between that of the PDD-based algorithm and other competing schemes, since it selects the same beams as the IA-BS while designing an optimal precoding matrix.

Fig. 6 compares the achievable system sum-rate when using different numbers of RF chains in the aforementioned schemes. In this experiment, we set $M_s = 128$, $K = 8$, SNR = 30dB and the number of RF chains N_{RF} increases from 8 to 24. Simulation results demonstrate that the PDD algorithm can achieve performance close to fully digital precoding with a small number of RF chains, while the performance of the simplified algorithm and the IA-BS scheme increase monotonically with the number of RF chains and converge to a sub-optimal value. Based on these and other similar results, we may conclude that the PDD-based algorithm can take full advantage of available RF chains. Note that here we do not consider the MM-BS scheme because it does not use a fixed number of RF chains, which means that some RF chains may be wasted since they do not contribute to the system sum-rate

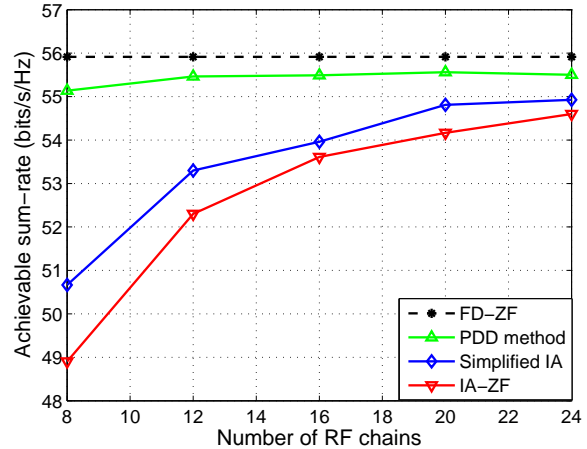


Fig. 6: Achievable system sum-rate comparison versus the number of RF chains (SNR = 30dB).

performance.

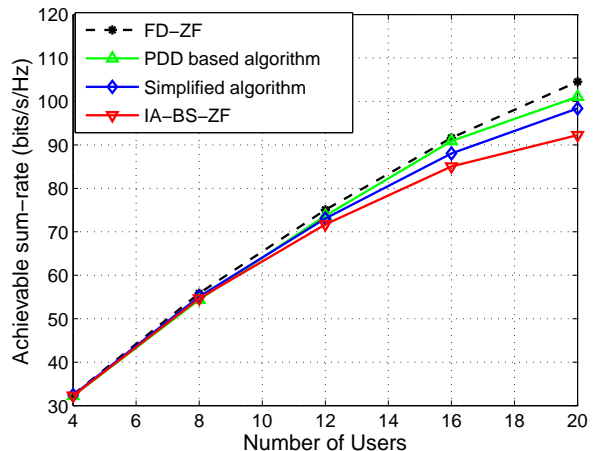


Fig. 7: Achievable sum-rate comparison versus the number of users (SNR = 30dB).

Finally, in Fig. 7, we analyze the aforementioned schemes for different numbers of users. In this scenario, we set $M_s = 128$, $N_{RF} = 24$, SNR = 30dB and the number of users K increases from 4 to 20. Simulation results demonstrate that the performance of all these schemes increase monotonically with the number of users. However, the PDD-based and simplified algorithms deliver better performance than the IA-BS scheme, pointing to their superior ability to mitigate the multi-user interference. Besides, the performance gap between the proposed algorithms and IA-BS escalates upon increasing number of users, which demonstrates that our proposed algorithms have more potential for applications with large number of users.

VII. CONCLUSION

In this paper, we have considered the joint design of the beam selection and precoding matrices in order to maximize the sum-rate of a downlink single-sided lens mmWave MU-MIMO system. We reformulated the sum-rate maximization

problem into a mathematically tractable form by using the WMMSE approach. Then, we proposed an efficient joint beam selection and precoding design algorithm based on the PDD method. To reduce design complexity, we also proposed a simplified algorithm based on the IA-BS scheme and the WMMSE approach. Simulation results have demonstrated that our proposed algorithms could converge in a few iterations and approach the optimal performance of the fully digital precoding scheme, thus enabling them to outperform the competing methods. We also evaluated all the aforementioned schemes for different numbers of RF chains and users, thereby showing that our algorithms could take full advantage of available RF chains and mitigate the multiuser interference efficiently.

APPENDIX A PROOF OF THEOREM 1

We first optimize $\{u_k\}$ by fixing the remaining variables of problem (9). In this case, (9) simplifies to the unconstrained problem:

$$\min_{\{u_k\}} \sum_{k=1}^K w_k e_k, \quad (49)$$

which is a weighted sum-MSE minimization problem. By examining the first-order optimality condition of (49), we can obtain the closed-form solution of $\{u_k\}$, which is the well-known MMSE receiver:

$$u_k^{opt} = u_k^{mmse} = J_k^{-1} \mathbf{p}_k^H \mathbf{F}^H \mathbf{h}_k, \forall k, \quad (50)$$

where $J_k \triangleq \sum_{i=1}^K \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i \mathbf{p}_i^H \mathbf{F}^H \mathbf{h}_k + \sigma^2$ is the covariance of the received signal at user k . Using the MMSE receiver u_k , the corresponding MSE e_k is given by:

$$e_k^{mmse} = (1 - \mathbf{p}_k^H \mathbf{F}^H \mathbf{h}_k J_k^{-1} \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k), \forall k. \quad (51)$$

Furthermore, fixing the remaining variables, the objective function in (9) is convex with respect to $\{w_k\}$. Therefore, by checking the first order optimality condition for $\{w_k\}$, we can obtain:

$$w_k^{opt} = (1 - \mathbf{p}_k^H \mathbf{F}^H \mathbf{h}_k J_k^{-1} \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k)^{-1} = (e_k^{mmse})^{-1}, \forall k. \quad (52)$$

Substituting $u_k^{opt}, w_k^{opt}, \forall k$ into (9), we have the following equivalent optimization problem:

$$\begin{aligned} \max_{\mathbf{F}, \mathbf{P}} \sum_{k=1}^K \log((e_k^{mmse})^{-1}) \\ \text{s.t. } \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \sum_{i=1}^{M_s} f_{ij} = 1, \\ \sum_{j=1}^{N_{RF}} f_{ij} \leq 1, f_{ij} \in \{0, 1\}, \forall (i, j) \in \mathcal{S}. \end{aligned} \quad (53)$$

Defining $\Gamma_k \triangleq \sum_{i \neq k} \mathbf{h}_k^H \mathbf{F} \mathbf{p}_i \mathbf{p}_i^H \mathbf{F}^H \mathbf{h}_k + \sigma^2$, we can obtain

$$\begin{aligned} \log((e_k^{mmse})^{-1}) &= \log(1 + \mathbf{p}_k^H \mathbf{F}^H \mathbf{h}_k \Gamma_k^{-1} \mathbf{h}_k^H \mathbf{F} \mathbf{p}_k) \\ &= \log(1 + |\mathbf{h}_k^H \mathbf{F} \mathbf{p}_k|^2 \Gamma_k^{-1}), \\ &= \log(1 + \gamma_k). \end{aligned} \quad (54)$$

Combining (54) with (53) we obtain

$$\begin{aligned} \max_{\mathbf{F}, \mathbf{P}} \sum_{k=1}^K \log(1 + \gamma_k) \\ \text{s.t. } \text{Tr}(\mathbf{P}^H \mathbf{F}^T \mathbf{F} \mathbf{P}) \leq P_s, \sum_{i=1}^{M_s} f_{ij} = 1, \\ \sum_{j=1}^{N_{RF}} f_{ij} \leq 1, f_{ij} \in \{0, 1\}, \forall (i, j) \in \mathcal{S}. \end{aligned} \quad (55)$$

This completes the proof. We note that the equivalence between problems (9) and (7) has nothing to do with their constraints.

APPENDIX B FRAMEWORK OF THE PDD METHOD

The PDD method can address nonconvex nonsmooth problems with coupling constraints. Consider the following problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \\ \text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \end{aligned} \quad (56)$$

where $f(\mathbf{x})$ is a scalar continuously differentiable function, the feasible set \mathcal{X} is the Cartesian product of n simple closed convex sets, and $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^{p \times 1}$ is a continuously differentiable vector function. The PDD method's framework for solving this problem is summarized in Table VI, which reveals a double-loop algorithm structure where the inner loop aims to solve the AL subproblem while the outer loop serves to update the dual variable or the penalty parameter in terms of the constraint violation. The main effort of the PDD method lies in Step 2.2, which updates the primal variables by applying the BCD algorithm [28] with initialization \mathbf{x}^0 to solve the AL problem $\mathcal{P}(\rho^m, \boldsymbol{\lambda}_m)$ to some accuracy ϵ^m , where $\mathcal{P}(\rho^m, \boldsymbol{\lambda}_m)$ is defined by

$$\min_{\mathbf{x} \in \mathcal{X}} \{\mathcal{L}^m(\mathbf{x}) \triangleq f(\mathbf{x}) + \boldsymbol{\lambda}_m^T \mathbf{h}(\mathbf{x}) + \frac{1}{2\rho^m} \|\mathbf{h}(\mathbf{x})\|^2\}. \quad (57)$$

$\mathcal{L}^m(\mathbf{x})$ is the augmented Lagrange function with dual variable $\boldsymbol{\lambda}_m$ and penalty parameter ρ^m . In the simulation, we set $\eta^m = (\tau^m)^{1/6}$ and $\tau^m = 0.6\tau^{m-1}$ empirically, and the termination criterion can be set as $\|\mathbf{h}(\mathbf{x})\| \leq 10^{-7}$ since the penalty term disappears finally. Furthermore, it can be shown that the limit point of the sequence generated by the PDD framework is a stationary point of problem (56) under suitable constraint qualification condition if $\mathcal{P}(\rho^m, \boldsymbol{\lambda}_m)$ can be solved to the accuracy ϵ^m in each inner iteration with $\epsilon^m \rightarrow 0$. The reader may consult [27] for further details.

REFERENCES

- [1] "Cisco visual networking index: Global mobile data traffic forecast update 2012-2017," Cisco, San Jose, CA, USA, White_paper_c11-520862, 2013.
- [2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335-349, May 2013.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.
- [4] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101-107, Jun. 2011.

TABLE VI: Framework of the PDD method for problem (56)

-
- 1 Initialize $\mathbf{x}^0, \rho^0 > 0, \lambda^0$, set $0 < c < 1, m = 1$.
 - 2 **Repeat**
 - 2.2 $\mathbf{x}^k = \text{optimize}(\mathcal{P}_{\rho^m, \lambda^m}, \mathbf{x}^{m-1}, \epsilon^m)$
 - 2.3 **if** $\|\mathbf{h}(\mathbf{x}^m)\| \leq \eta^m$ **then**

$$\lambda^{m+1} = \lambda^m + \frac{1}{\rho^m} \mathbf{h}(\mathbf{x}^m)$$

$$\rho^{m+1} = \rho^m$$
 - 2.4 **else**

$$\lambda^{m+1} = \lambda^m$$

$$\rho^{m+1} = c\rho^m$$
 - 2.5 **end if**
 - 2.6 $m = m + 1$
 - 3 **Until** some termination criterion is met.
-

- [5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
- [6] Rusek, F., Persson, D., Lau, B. K., Larsson, E. G., Marzetta, T. L., Edfors, O., and Tufvesson, F., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.
- [7] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742-758, Oct. 2014.
- [8] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun and F. Aryanfar, "Millimeter wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106-113, Feb. 2014.
- [9] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68-73, Oct. 2004.
- [10] X. Zhang, A. F. Molish, and S. Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.
- [11] A. Molisch and M. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, vol. 5, no. 1, pp. 46-56, Mar. 2004.
- [12] K. Ntougias, D. Ntaikos and C. B. Papadias, "Coordinated MIMO with Single-fed Load-Controlled Parasitic Antenna Arrays," in *Proc. Signal Process. Adv. Wireless Commun. (SPAWC)*, July 3-6 2016, pp. 1-5.
- [13] K. Ntougias, D. Ntaikos and C. B. Papadias, "Single- and multiple-RF load controlled parasitic antenna arrays operating at Cm-wave frequencies: Design and applications for 5G wireless access / backhaul," in *Proc. FITCE Congress (FITCE)*, Sept. 14-15 2017, pp. 28-34.
- [14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
- [15] A. Alkhateeb, G. Leus and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481-6494, Nov. 2015.
- [16] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.
- [17] A. Alkhateeb, O. El Ayach, G. Leus and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.
- [18] J. Brady, N. Behdad, and A. Sayeed, "Beamspace MIMO for millimeter wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814-3827, Jul. 2013.
- [19] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: a new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557-1571, Apr. 2016.
- [20] G.-H. Song, J. Brady and A. Sayeed, "Beamspace MIMO transceivers for low-complexity and near-optimal communication at mmWave frequencies," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 26-31, 2013, pp. 4394-4398.
- [21] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter wave frequencies," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3679-3684.
- [22] J. Brady and A. Sayeed, "Beamspace MU-MIMO for high-density gigabit small cell access at millimeter wave frequencies," in *Proc. Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 22-25, 2014, pp. 80-84.
- [23] Y. Zeng, L. Yang, and R. Zhang, "Multi-user millimeter wave MIMO with full-dimensional lens antenna array," to appear in *IEEE J. Sel. Areas Commun.*, 2017
- [24] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212-2222, Jun. 2015.
- [25] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054-1057, May 2016.
- [26] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Sept. 2011.
- [27] Q. Shi, M. Hong, X. Fu and T-H Chang, "Penalty Dual Decomposition Method For Nonsmooth Nonconvex Optimization." [Online]. Available: <https://arxiv.org/abs/1712.04767>
- [28] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo. "A block successive upper bound minimization method of multipliers for linearly constrained convex optimization," 2013. Preprint, [Online]. Available: arXiv:1401.7079.
- [29] Q. Shi, W. Xu, J. Wu, E. Song, and Y. Wang, "Secure beamforming for MIMO broadcasting with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2841-2853, May 2015.
- [30] S. Verdú, *Multiuser Detection*. Cambridge, 1998.