# Codebook Based Precoding and Power Allocation for MU-MIMO Systems for Sum Rate Maximization

Sai Subramanyam Thoota, Prabhu Babu, and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—In this paper, we study the problem of downlink (DL) sum rate maximization in codebook based multiuser (MU) multiple input multiple output (MIMO) systems. The user equipments (UEs) estimate the DL channels using pilot symbols sent by the access point (AP) and feedback the estimates to the AP over a control channel. We present a closed form expression for the achievable sum rate of the MU-MIMO broadcast system with codebook constrained precoding based on the estimated channels, where multiple data streams are simultaneously transmitted to all users. Next, we present novel, computationally efficient, minorization-maximization (MM) based algorithms to determine the selection of beamforming vectors and power allocation to each beam that maximizes the achievable sum rate. Our solution involves multiple uses of MM in a nested fashion. Based on this approach, we propose and contrast two algorithms, which we call the square-root-MM (SMM) and inverse-MM (IMM) algorithms. The algorithms are iterative and converge to a locally optimal beamforming vector selection and power allocation solution from any initialization. We evaluate the performance and complexity of the algorithms for various values of the system parameters, compare them with existing solutions, and provide further insights into how they can be used in system design.**

***Index Terms*—** Minorization-Maximization, MU-MIMO, Beamforming, Precoding.

## I. Introduction

The design of downlink precoding and beamforming schemes for multiuser muliple input multiple output (MU-MIMO) systems with a large number of antennas at the base station (BS) or access point (AP) has attracted significant research interest in recent years [2]–[8]. In typical frequency division duplex (FDD) systems, the channel state information (CSI) is first obtained at the UEs using downlink (DL) training, i.e., from pilot symbols transmitted by the AP. Then, the UEs send their channel estimates back to the AP over an uplink (UL) control channel. The AP, upon receiving the channel estimates, computes a precoding matrix for data transmission to each of the users. In this paper, we consider an approach where the AP selects the columns of the precoding matrix from a predetermined codebook of beamforming vectors. This allows the AP the flexibility of either conveying the selected codebook indices to the UEs over a DL control channel, or using dedicated pilots to enable the UEs to estimate their respective effective channels. Codebook based precoding is also relevant because it is employed (for example) in the IEEE 802.15.3 and IEEE 802.15ac standards [9], [10]. Our goal in this context is to determine the optimal selection of beamforming vectors and power allocation across users, with possibly *multi-stream* data transmission to each of the UEs. This is a non-convex and combinatorial problem, and therefore hard to solve. We present two novel algorithms based on the minorization-maximization (MM) framework for maximizing the sum rate under the codebook constraint. In the process, we also develop new matrix inequalities that facilitate the use of the MM approach for optimization. These latter results could be of independent interest in many other non-convex optimization problems.

Most of the existing studies on sum rate maximization in MU-MIMO systems do not consider the problem when the transmitter is constrained to select its precoding vectors from a codebook of candidate vectors [11]–[18]. In codebook based transmission, the columns of the precoding matrices need to be selected from the codebook. This makes our problem fundamentally different from, and intrinsically harder, than the (unconstrained) design of precoding matrices, as the underlying problem becomes one of allocating beamforming vectors to users, i.e., an integer optimization problem.

A codebook based approach for beamforming and power allocation in multiuser multiple input *single* output (MU-MISO) systems is considered in [19], where the authors transform the underlying mixed integer optimization problem into a structured mixed integer second-order cone program. They also customize a convex continuous relaxation based branch-and-cut algorithm to compute an optimal solution to the beamforming problem. Considering the MU-MIMO system (where multiple data streams are transmitted to each user) significantly changes the problem, because interference between streams assigned to the same UE can be handled via joint processing of the signals received at the UE antennas. This is unlike the single antenna UE case, where all inter-stream interference negatively impacts the data rate.

In [20], the authors consider the beamforming assignment and power allocation (BAPA) problem for MU-MISO systems. They introduce a virtual uplink (VUL) to decouple the power allocations across different UEs, thereby admitting an iterative

Sai Subramanyam Thoota and Chandra R. Murthy are with the Department of ECE, Indian Institute of Science, Bangalore, India (email: thoota@iisc.ac.in; cmurthy@iisc.ac.in).

Prabhu Babu is with the Centre for Applied Research in Electronics (CARE), Indian Institute of Technology, Delhi, India (email: prabhubabu@care.iitd.ac.in).

This work was financially supported in part by a research grant from Intel India Inc., and in part by the Young Faculty Research Fellowship, Ministry of Electronics and Information Technology, Govt. of India.

A part of this work has been published in the National Conference on Communications 2019 [1].

solution. In the case of imperfect CSI at the transmitter, rate-splitting is shown to be sum-rate optimal [21]–[24], but it requires successive interference cancellation to remove the interference caused by the common messages and decode the private messages of all the users. In contrast, we consider linear receivers and linear precoding of data using a beamforming vector codebook, which leads to a solution that is easy to implement in practical systems.

In this paper, we approach the problem of codebook based precoding for sum rate maximization using the iterative technique of minorization-maximization (MM). We bound the original objective function in multiple stages, which simplifies the optimization problem and helps in finding a closed-form analytical solution. Note that, when using an MM approach for optimizing a non-convex objective function, the key novelty is to bound the cost function by a surrogate function that is tight at the current iterate and is easy to optimize. Different bounds can lead to different convergence and complexity tradeoffs. We present and compare two alternatives for bounding the cost function. The resulting algorithms are computationally simple (e.g., they do not involve any matrix inversion operations), making them attractive for implementation. Further, as they are based on the MM principle, they are guaranteed to converge to a local optimum from any initialization. Our main contributions in this paper are as follows:

1) We present a closed-form expression for the achievable sum rate of a codebook based precoding MU-MIMO broadcast system with minimum mean squared error (MMSE) channel estimation at the receiver and the feedback of imperfect CSI to the transmitter via an error-free control channel. The achievable sum rate expression provides us with the objective function for beamforming vector selection and power allocation.
2) We propose two algorithms for solving the sum rate maximization problem in MU-MIMO systems. The two algorithms differ in the way they bound the non-convex sum rate cost function to arrive at the surrogate cost that needs to be optimized.
   a) Square root MM (SMM) algorithm: Here, we consider the square-root of the power allocation as the optimization variable, and apply minorization three times to lower-bound the objective function with a quadratic-form cost function. The surrogate cost function so obtained admits a closed-form optimal solution.
   b) Inverse MM (IMM) algorithm: We use a matrix inequality to deal with the matrix inverse term in the objective function. After two rounds of minorization, this again leads to a surrogate quadratic lower bound, and admits a closed-form optimal solution.

   We analytically show that the closed-form solutions of the SMM and IMM algorithms are optimal with respect to their corresponding surrogate optimization problems.
3) We empirically study the performance of the SMM and IMM algorithms with respect to the number of users, codebook size, data SNR, pilot SNR etc. Further, we illustrate the performance advantage offered by the SMM and IMM algorithms compared to the WMMSE [12] and WSRMax [16] approaches as well as a single-user-optimal codebook based precoding approach, where the IMM algorithm is used to select the beamforming vector and power allocation on a per-user basis. The results demonstrate that jointly choosing beamforming vectors is necessary to realize the full potential of MU-MIMO transmission. We compare the sum rate performance and run times of the MM algorithms with that of CVX [25], [26], a convex optimization package available online. The IMM algorithm has significantly lower run time compared to CVX in the interference-limited regime, which is the primary domain of interest of our work.

We note that, even in the single-user context, our solution to the problem of selecting *multiple* beamforming vectors in a codebook along with their corresponding power allocation is novel, and a similar solution does not exist in the literature, to the best of our knowledge. Moreover, our approach can easily accommodate additional constraints such as a minimum rate per user for a selected subset of users, etc. Also, codebook based precoding will necessarily play a role in the next generation wireless systems like mmWave massive MIMO, where it is customary to adopt a hybrid precoding architecture [27] due to the high cost and power consumption in the power amplifiers and ADCs. For instance, in the analog precoding stage, due to the finite angular resolution of the analog phase shifters, codebook based precoding arises naturally. The solution presented in this paper can be easily adapted to mmWave hybrid analog-digital beamforming based systems, by imposing a constraint on the total number of spatial streams to which nonzero power is allocated. Finally, the guaranteed convergence and simple implementation makes the novel bounding technique developed in this paper a potentially attractive approach for a variety of optimization problems which arise in MU-MIMO systems.

## II. System Model & Problem Statement

We consider a MU-MIMO system comprised of a AP equipped with $N_t$ antennas and $K$ users each equipped with $N_r$ antennas. The UEs and the AP share a codebook $\mathbf{C} \in \mathbb{C}^{N_t \times N}$, whose columns consist of $N$ unit-norm beamforming vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$, with $\mathbf{c}_j \in \mathbb{C}^{N_t}$. The complex baseband channel between the AP and the $k^{\text{th}}$ UE is denoted by $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$. The AP sends the data symbol $s_k(l)$ to the $k^{\text{th}}$ UE by precoding it using the $l^{\text{th}}$ beamforming vector $\mathbf{c}_l$, and the composite signal $\mathbf{x} \in \mathbb{C}^{N_t}$ transmitted by the AP is

$$\mathbf{x} = \sqrt{\rho_{dl}} \sum_{k=1}^{K} \sum_{l=1}^{N} \sqrt{P_k(l)}\, \mathbf{c}_l s_k(l), \tag{1}$$

where $\rho_{dl}$ is the data signal to noise ratio (SNR). In the sequel, all powers are normalized with respect to the noise variance, and we use the SNR and transmit power interchangeably.

Note that this general model allows multiple users to receive data on the same beamforming vector or multiple beamforming vectors to be assigned to a given user. Ultimately, the beamforming vector selection and power allocation solution will ensure that the objective function, namely, the sum rate, is maximized. Therefore, there is no need to explicitly impose

constraints such as each beamforming vector should be allotted to at most one user, or that a user should not be allocated more than a given number of beamforming vectors. In (1), the data symbols $\{s_k(l)\}$ for $k = 1, \ldots, K$, $l = 1, \ldots, L$ are assumed to be independent and identically distributed (i.i.d.) Gaussian, with zero mean and unit variance. Let $\mathbf{\Phi}_k \triangleq \mathrm{diag}(P_k(1), P_k(2), \ldots, P_k(N))$ denote a diagonal matrix whose entries contain the fraction of the available power at the AP that is allocated to $k^{\text{th}}$ user on the $N$ beamforming vectors (hence, $\mathrm{tr}\left(\sum_{k=1}^{K} \mathbf{\Phi}_k\right) = 1$). Then, the goal at the AP is to determine $\mathbf{\Phi}_k$, based on $\mathbf{H}_k$, $k = 1, 2, \ldots, K$, to maximize the achievable sum rate in the system. Note that, $P_k(l) = 0$ is equivalent to *not allotting* the $l^{\text{th}}$ beamforming vector in the codebook to the $k^{\text{th}}$ user.

Past works in the area, e.g., [19], [20], assume that $\mathbf{H}_k$ is perfectly known at the AP. However, in practice, channel is estimated using training symbols, which results in imperfect CSI. Therefore, we first describe the MMSE channel estimation at the UEs using common pilots transmitted by the AP.

### A. Downlink Training and Channel Estimation

In the downlink training phase, the AP transmits $\tau_p$ orthogonal pilot symbols ($\tau_p \geq N_t$) over its $N_t$ antennas. The pilot signal $\mathbf{X}_p \in \mathbb{C}^{N_t \times \tau_p}$ satisfies $\mathbf{X}_p \mathbf{X}_p^H = \mathbf{I}_{N_t}$. The received pilot sequence at the $k^{\text{th}}$ user, $\mathbf{Y}_k^{(p)} \in \mathbb{C}^{N_r \times \tau_p}$, is given by

$$\mathbf{Y}_k^{(p)} = \sqrt{\rho_{dl}^{(p)} \tau_p} \mathbf{H}_k \mathbf{X}_p + \mathbf{W}_k, \tag{2}$$

where $\rho_{dl}^{(p)}$ is the pilot signal to noise ratio (SNR), and $\mathbf{W}_k \in \mathbb{C}^{N_r \times \tau_p}$ is the complex additive white Gaussian noise (AWGN) whose columns are i.i.d. with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_{N_r}$, denoted $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$. The multiplication of the transmit symbols by $\sqrt{\tau_p}$ above is to ensure that the total energy expended over the entire pilot duration is $\tau_p$. Also, $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix of the $k^{\text{th}}$ user, which contains i.i.d. entries drawn from $\mathcal{CN}(0, \beta_k)$, where $\beta_k$ denotes the combined effect of long term pathloss and large scale shadowing between the AP and user $k$.

By multiplying (2) by $\mathbf{X}_p^H$ on the right, we get $\mathbf{Y}_k^{(p)'} \triangleq \mathbf{Y}_k^{(p)} \mathbf{X}_p^H = \sqrt{\rho_{dl}^{(p)} \tau_p} \mathbf{H}_k + \mathbf{W}_k'$, where $\mathbf{W}_k' = \mathbf{W}_k \mathbf{X}_p^H \in \mathbb{C}^{N_r \times N_t}$ is the effective noise whose columns are also distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$. The MMSE estimate of the channel is given by [28] $\widehat{\mathbf{H}}_k \triangleq \sqrt{\rho_{dl}^{(p)} \tau_p} \beta_k \mathbf{Y}_k^{(p)'} / \left(1 + \rho_{dl}^{(p)} \tau_p \beta_k\right)$, and the mean square value of each entry of $\widehat{\mathbf{H}}_k$ is given by $\gamma_k \triangleq \rho_{dl}^{(p)} \tau_p \beta_k^2 / \left(1 + \rho_{dl}^{(p)} \tau_p \beta_k\right)$. Note that, $\widehat{\mathbf{H}}_k$ is Gaussian distributed and is uncorrelated with the channel estimation error $\widetilde{\mathbf{H}}_k \triangleq \mathbf{H}_k - \widehat{\mathbf{H}}_k$. This is useful in computing the noise plus interference covariance matrix, in the next subsection.

### B. Derivation of the Achievable Rate

Consider a power allocation matrix $\mathbf{\Phi}_k$, $k = 1, 2, \ldots, K$. From (1), the composite signal transmitted by the AP to all the users can be written compactly as $\sqrt{\rho_{dl}} \sum_{j=1}^{K} \mathbf{C} \mathbf{\Phi}_j^{\frac{1}{2}} \mathbf{s}_j \in \mathbb{C}^{N_t}$ where $\mathbf{s}_j = [s_j(1), s_j(2), \ldots, s_j(N)]^T$ is the data transmitted data to the $j^{\text{th}}$ user using the beamforming codebook $\mathbf{C}$, and $\rho_{dl}$ is the downlink SNR. The received signal $\mathbf{y}_k \in \mathbb{C}^{N_r}$ at the $k^{\text{th}}$ user is given by

$$\mathbf{y}_k = \mathbf{H}_k \left( \sqrt{\rho_{dl}} \sum_{j=1}^{K} \mathbf{C} \mathbf{\Phi}_j^{\frac{1}{2}} \mathbf{s}_j \right) + \mathbf{w}_k,$$

where $\mathbf{w}_k \in \mathbb{C}^{N_r}$ is the complex AWGN at the $k^{\text{th}}$ user with distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$.

Given the channel estimate $\widehat{\mathbf{H}}_k$ at the receiver, the received signal can be rewritten as

$$\mathbf{y}_k = \underbrace{\sqrt{\rho_{dl}} \widehat{\mathbf{H}}_k \mathbf{C} \mathbf{\Phi}_k^{\frac{1}{2}} \mathbf{s}_k}_{\text{Desired signal}} + \sqrt{\rho_{dl}} \widehat{\mathbf{H}}_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{C} \mathbf{\Phi}_j^{\frac{1}{2}} \mathbf{s}_j + \sqrt{\rho_{dl}} \widetilde{\mathbf{H}}_k \sum_{j=1}^{K} \mathbf{C} \mathbf{\Phi}_j^{\frac{1}{2}} \mathbf{s}_j + \mathbf{w}_k.$$

In order to compute the achievable rate from the above equation, we need to compute the signal and noise plus interference covariance matrices, find the signal to interference plus noise ratio (SINR) and then use the worst case noise theorem [29]. The covariance of the desired signal is $\rho_{dl} \widehat{\mathbf{H}}_k \mathbf{C} \mathbf{\Phi}_k \mathbf{C}^H \widehat{\mathbf{H}}_k^H$. We denote the covariance matrix of the noise and interference of the $k^{\text{th}}$ user by $\mathbf{V}_k$. Using the fact that the terms involved are uncorrelated, it is easy to show that

$$\mathbf{V}_k = \mathbf{I}_{N_r} + \rho_{dl} \widehat{\mathbf{H}}_k \mathbf{C} \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{\Phi}_j \mathbf{C}^H \widehat{\mathbf{H}}_k^H + \underbrace{\rho_{dl} \mathbb{E}\left[ \widetilde{\mathbf{H}}_k \mathbf{x} \mathbf{x}^H \widetilde{\mathbf{H}}_k^H \right]}_{\text{Due to channel est. errors}}. \tag{3}$$

It is shown in Appendix A that

$$\mathbb{E}\left[ \widetilde{\mathbf{H}}_k \mathbf{x} \mathbf{x}^H \widetilde{\mathbf{H}}_k^H \right] = (\beta_k - \gamma_k) \mathbf{I}_{N_r}. \tag{4}$$

Substituting (4) in (3) and simplifying, we get

$$\mathbf{V}_k = \sigma_k^2 \mathbf{I}_{N_r} + \widehat{\widehat{\mathbf{H}}}_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{\Phi}_j \widehat{\widehat{\mathbf{H}}}_k^H, \tag{5}$$

where $\widehat{\widehat{\mathbf{H}}}_k \triangleq \sqrt{\rho_{dl}} \widehat{\mathbf{H}}_k \mathbf{C}$ and $\sigma_k^2 \triangleq (1 + \rho_{dl} (\beta_k - \gamma_k))$. Now, since the interference terms are uncorrelated with the desired signal by virtue of MMSE estimation, using the worst case noise theorem [29], the achievable rate of the $k^{\text{th}}$ user and the downlink sum rate are given by

$$R_k = \log \left| \mathbf{I}_{N_r} + \mathbf{V}_k^{-1} \widehat{\widehat{\mathbf{H}}}_k \mathbf{\Phi}_k \widehat{\widehat{\mathbf{H}}}_k^H \right|, \tag{6}$$

Our goal is to maximize the sum rate $R_{\text{tot}} = \sum_{k=1}^{K} R_k$ under a total power constraint:

$$\underset{\substack{\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_K \\ \mathbf{\Phi}_k \text{ diagonal, p.s.d.}}}{\text{maximize}} \sum_{k=1}^{K} \log \left| \mathbf{I}_{N_r} + \mathbf{V}_k^{-1} \widehat{\widehat{\mathbf{H}}}_k \mathbf{\Phi}_k \widehat{\widehat{\mathbf{H}}}_k^H \right|, \tag{7}$$

$$\text{subject to } \mathrm{tr}\Big( \sum_{k=1}^{K} \mathbf{\Phi}_k \Big) = 1.$$

The optimization problem in (7) is nonconvex in $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K$ due to the $\mathbf{V}_k^{-1}$ term, and cannot be solved

in closed-form. Note that, we restrict the power allocation matrices to be diagonal in order to be implementable under codebook based precoding. This constrains the precoding matrices to belong to the finite set of matrices that can be expressed as the sum of outer products of codebook vectors weighted by the corresponding power allocation, and makes the problem significantly harder than unconstrained designs of precoding matrices [11], [12].

In this work, we propose two algorithms based on the MM principle, which proceeds by finding a surrogate function that is a lower bound on the objective function, followed by maximizing the surrogate cost function, iteratively, until convergence to a local optimum. An excellent tutorial on the MM principle can be found in [30].

## III. Minorization-Maximization Algorithms for Sum Rate Maximization

In this section, we present our solutions to the beamforming vector selection and power allocation problem stated in (7). We propose two algorithms, namely, the square root MM (SMM) and inverse MM (IMM) algorithms. These algorithms start with a common minorization step, and then solve the resulting optimization problem by two different approaches.

The first step in finding a computationally efficient solution to (7) is to find a surrogate function which is a lower bound on the sum rate, and is tight at the current iterate. To this end, consider the function $f(\mathbf{Z},\mathbf{Y}) = \log\left|\mathbf{Z}^{-1}\mathbf{Y}\right|$, for $\mathbf{Z},\ \mathbf{Y} \succeq 0$. This function is convex in $\mathbf{Z},\ \mathbf{Y}^{-1}$. Hence, we can bound it from below using the first order Taylor series expansion, as given by the following Lemma:

*Lemma 1:* For matrices $\mathbf{Z},\ \mathbf{Y} \succeq 0$, the function

$$f(\mathbf{Z},\mathbf{Y}) = \log\left|\mathbf{Z}^{-1}\mathbf{Y}\right|$$

can be lower bounded by

$$f(\mathbf{Z},\mathbf{Y}) \geq -\Big(\log\left|\mathbf{Z}^{(m)}\right| + \mathrm{tr}\left(\mathbf{Z}^{(m)^{-1}}\left(\mathbf{Z}-\mathbf{Z}^{(m)}\right)\right) + \log\left|\mathbf{Y}^{(m)^{-1}}\right| + \mathrm{tr}\left(\mathbf{Y}^{(m)}\left(\mathbf{Y}^{-1}-\mathbf{Y}^{(m)^{-1}}\right)\right)\Big)$$

with equality at $\mathbf{Z} = \mathbf{Z}^{(m)}$ and $\mathbf{Y} = \mathbf{Y}^{(m)}$. (Later, $m$ will be used to denote the iteration index).

Returning to our problem, we define an intermediate matrix

$$\mathbf{B}_k \triangleq \sigma^2\mathbf{I}_{N_r} + \sum_{j=1}^{K}\hat{\hat{\mathbf{H}}}_k\mathbf{\Phi}_j\hat{\hat{\mathbf{H}}}_k^H. \tag{8}$$

The rate of the $k^{\text{th}}$ user in (6) can then be written as $R_k = \log\left|\mathbf{V}_k^{-1}\mathbf{B}_k\right|$. Using Lemma 1, we get the following surrogate optimization problem for (7):

$$\begin{aligned}&\{\mathbf{\Phi}_1^{(m+1)},\ldots,\mathbf{\Phi}_K^{(m+1)}\}\\ &= \underset{\mathbf{\Phi}_1,\ldots,\mathbf{\Phi}_K}{\mathrm{argmax}}\sum_{k=1}^{K}\Bigg\{-\mathrm{tr}\left(\mathbf{V}_k^{(m)^{-1}}\Big(\sigma^2\mathbf{I}_{N_r}+\sum_{\substack{j=1\\j\neq k}}^{K}\hat{\hat{\mathbf{H}}}_k\mathbf{\Phi}_j\hat{\hat{\mathbf{H}}}_k^H\Big)\right)\\ &\qquad -\mathrm{tr}\left(\mathbf{B}_k^{(m)}\Big[\sigma^2\mathbf{I}_{N_r}+\sum_{j=1}^{K}\hat{\hat{\mathbf{H}}}_k\mathbf{\Phi}_j\hat{\hat{\mathbf{H}}}_k^H\Big]^{-1}\right)\Bigg\},\\ &\text{subject to}\quad \mathrm{tr}\Big(\sum_{k=1}^{K}\mathbf{\Phi}_k\Big)\leq 1,\end{aligned} \tag{9}$$

where $m$ is the iteration index. Here, we omit the $\log_e 2$ term in the denominator, as it does not affect the solution. In (9), the quantities $\mathbf{V}_k^{(m)}$ and $\mathbf{B}_k^{(m)}$ are computed by substituting $\mathbf{\Phi}_k^{(m)}$ for $\mathbf{\Phi}_k$ in (5) and (8), respectively. Now, if we are able to solve the surrogate problem in (9), then, starting from an arbitrary initialization for $\mathbf{\Phi_k}$, the MM procedure iterates between solving (9) and updating $\mathbf{V}_k$ and $\mathbf{B}_k$. By virtue of the fact that the cost function increases in each iteration and is bounded above (for example, by the sum of the best rates achievable by each individual user), such a procedure is guaranteed to converge to a local optimum from any initialization.

Now, the optimization problem in (9) is a semidefinite program (SDP). However, the matrices $\{\mathbf{\Phi}_k\}_{k=1}^{k=K}$ are coupled in the objective function and constraints, making it a large dimensional problem. Due to this, SDP based methods such as `sdpsol` to solve (9) can quickly become computationally prohibitive as the number of users, the size of the codebook, and/or number of antennas gets large. Hence, there is a need to find alternative, computationally inexpensive approaches to solving (9). The proposed SMM and IMM algorithms employ two different surrogate functions to further lower bound the objective function, in turn, leading to a surrogate cost function that is more amenable to optimization. In fact, we are able to solve the final surrogate problem in closed-form.

Before discussing the SMM and IMM algorithms further, we define some notation and simplify the first term in the objective function in (9). Let

$$\mathbf{\Phi} \triangleq \mathrm{diag}\left(\mathbf{\Phi}_1,\ldots,\mathbf{\Phi}_K\right), \tag{10}$$

$$\mathbf{\Psi}_k \triangleq \left[\hat{\hat{\mathbf{H}}}_k,\ldots,\hat{\hat{\mathbf{H}}}_k\right], k = 1,\ldots,K \tag{11}$$

denote the augmented power allocation and the $k^{\text{th}}$ user's channel matrices, respectively. In (11), $\hat{\hat{\mathbf{H}}}_k$ is repeated $K$ times. Also, let

$$\mathbf{Q} \triangleq \sum_{k=1}^{K}\mathrm{diag}\left(\hat{\hat{\mathbf{H}}}_k^H\mathbf{V}_k^{-1}\hat{\hat{\mathbf{H}}}_k,\ldots,\mathbf{0}_N,\ldots,\hat{\hat{\mathbf{H}}}_k^H\mathbf{V}_k^{-1}\hat{\hat{\mathbf{H}}}_k\right). \tag{12}$$

In the above, the $N\times N$ all zero matrix $\mathbf{0}_N$ is in the $k^{\text{th}}$ block diagonal position of $\mathbf{Q}$. Excluding the constant noise variance part, we can rewrite the first term of (9) as

$$\sum_{k=1}^{K}\mathrm{tr}\left(\mathbf{V}_k^{(m)^{-1}}\Big(\sum_{\substack{j=1\\j\neq k}}^{K}\hat{\hat{\mathbf{H}}}_k\mathbf{\Phi}_j\hat{\hat{\mathbf{H}}}_k^H\Big)\right) = \mathrm{tr}\left(\mathbf{Q}^{(m)}\mathbf{\Phi}\right), \tag{13}$$

where the superscript $m$ denotes the iteration index, and $\mathbf{Q}^{(m)}$ is obtained by substituting $\mathbf{V}_k^{(m)}$ for $\mathbf{V}_k$ in (12). We are now ready to describe the SMM and IMM algorithms in detail.

### A. Square-Root Minorization Maximization Procedure

The square root MM procedure involves working with the square root of the power allocation matrix $\mathbf{\Phi}$. It also involves two stages of minorization. The result is a surrogate objective function that is a lower bound on the cost function in (9), is

tight at the current iterate, and is easy to optimize. First, note that, with the notation defined in (10) and (11), the second term in (9) can be written as

$$-\sum_{k=1}^{K}\mathrm{tr}\left(\mathbf{F}_k^{(m)}\left(\sigma^2\mathbf{I}_{N_r}+\mathbf{\Psi}_k\mathbf{\Phi}\mathbf{\Psi}_k^H\right)^{-1}\mathbf{F}_k^{(m)^H}\right), \quad (14)$$

where $\mathbf{F}_k$ is such that $\mathbf{B}_k=\mathbf{F}_k^H\mathbf{F}_k$, and can be computed via the Cholesky decomposition of $\mathbf{B}_k$. The above cost function cannot be directly optimized due to the matrix inversion involved. Hence, we minorize it using the following lemma.

*Lemma 2:* Let $\mathbf{R}$ denote a diagonal p.s.d. square matrix, and consider the function

$$f(\mathbf{R})\triangleq -\mathrm{tr}\left(\mathbf{A}\left(\mathbf{B}+\mathbf{C}\mathbf{R}\mathbf{C}^H\right)^{-1}\mathbf{A}^H\right), \quad (15)$$

where $\mathbf{A},\mathbf{B}$ and $\mathbf{C}$ are matrices of compatible dimensions, and $\mathbf{B}\succ 0$, so that $\mathbf{B}+\mathbf{C}\mathbf{R}\mathbf{C}^H$ is invertible. Then, for a given diagonal p.s.d. matrix $\mathbf{R}^{(m)}$, $f(\mathbf{R})$ can be lower bounded by

$$\begin{aligned} &f(\mathbf{R})\\ &\geq g(\mathbf{R}|\mathbf{R}^{(m)})\triangleq -\mathrm{tr}(\hat{\mathbf{K}})\\ &\quad+\mathrm{tr}\left(\left(\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\mathbf{A}\mathbf{B}^{-1}\mathbf{C}+\mathbf{C}^H\mathbf{B}^{-H}\mathbf{A}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\right)\mathbf{R}^{\frac{1}{2}}\right.\\ &\quad\left.-\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\right), \end{aligned} \quad (16)$$

where

$$\begin{aligned} &\hat{\mathbf{X}}\triangleq\mathbf{A}\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{(m)\frac{1}{2}},\quad \hat{\mathbf{Y}}\triangleq\mathbf{I}+\mathbf{R}^{(m)\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{(m)\frac{1}{2}},\\ &\hat{\mathbf{K}}\triangleq\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^H+\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}-\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\hat{\mathbf{Y}}\\ &\qquad+\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}+\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H. \end{aligned}$$

Also, $g(\mathbf{R}^{(m)}|\mathbf{R}^{(m)})=f(\mathbf{R}^{(m)})$.

*Proof:* See Appendix B. ∎

The objective function in (14) is in the same form as the function in Lemma 2. Applying Lemma 2 to (14), we get

$$-\mathrm{tr}\left(\mathbf{W}_{1,k}^{(m)}\mathbf{\Phi}^{\frac{1}{2}}+\mathbf{W}_{2,k}^{(m)}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right), \quad (17)$$

where

$$\mathbf{W}_{1,k}\triangleq-\left\{\frac{\mathbf{Y}_k^{-1}\mathbf{X}_k^H\mathbf{F}_k\mathbf{\Psi}_k+\mathbf{\Psi}_k^H\mathbf{F}_k^H\mathbf{X}_k\mathbf{Y}_k^{-1}}{\sigma^2}\right\}, \quad (18)$$

$$\mathbf{W}_{2,k}\triangleq\mathbf{Y}_k^{-1}\mathbf{X}_k^H\mathbf{X}_k\mathbf{Y}_k^{-1}, \quad (19)$$

and $\mathbf{S}_k\in\mathbb{C}^{KN\times KN},\mathbf{X}_k\in\mathbb{C}^{N_r\times KN}$ and $\mathbf{Y}_k\in\mathbb{C}^{KN\times KN}$ are defined as

$$\mathbf{S}_k\triangleq\frac{\mathbf{\Psi}_k^H\mathbf{\Psi}_k}{\sigma^2},\ \mathbf{X}_k\triangleq\frac{\mathbf{F}_K\mathbf{\Psi}_k\mathbf{\Phi}^{\frac{1}{2}}}{\sigma^2},\ \mathbf{Y}_k\triangleq\mathbf{I}_{KN}+\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}.$$

Note that, $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ in (18) and (19), are negative and p.s.d. matrices, and hence, their diagonal entries are non-positive and non-negative, respectively. Also, $\mathbf{Y}_k^{-1}$ can be computed with low complexity using Woodbury matrix identity (requiring only a $N_r\times N_r$ matrix inverse instead of a $KN\times KN$ matrix inverse). Substituting (17) into (9), we get the surrogate cost function that needs to be maximized. Notice that the matrix inversion in (14) has been circumvented by the use of the lower bound. However, the surrogate cost function is not yet amenable to a closed-form solution due to the $\mathbf{W}_{2,k}^{(m)}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}$ term in (17). Hence, we minorize the second term in (17) again to get a cost function that is easy to optimize. To this end, we need the following Lemma.

*Lemma 3:* Suppose $\mathbf{R}$ is a p.s.d. diagonal matrix, and $\mathbf{A}$ and $\mathbf{B}$ are symmetric p.s.d. square matrices. Then, the function $f(\mathbf{R})\triangleq-\mathrm{tr}\left(\mathbf{A}\mathbf{R}\mathbf{B}\mathbf{R}\right)$ can be lower bounded by

$$\begin{aligned} f(\mathbf{R})\geq&-\mathrm{tr}\left(\mathbf{A}\mathbf{R}^{(m)}\mathbf{B}\mathbf{R}^{(m)}-\left(\left(\mathbf{B}-\lambda\mathbf{I}\right)\mathbf{R}^{(m)}\mathbf{A}\right.\right.\\ &\left.\left.+\mathbf{A}\mathbf{R}^{(m)}\left(\mathbf{B}-\lambda\mathbf{I}\right)\right)\mathbf{R}^{(m)}\right)\\ &-\mathrm{tr}\left(\left(\left(\mathbf{B}-\lambda\mathbf{I}\right)\mathbf{R}^{(m)}\mathbf{A}\right.\right.\\ &\left.\left.+\mathbf{A}\mathbf{R}^{(m)}\left(\mathbf{B}-\lambda\mathbf{I}\right)\right)\mathbf{R}\right)-\lambda\mathrm{tr}\left(\mathbf{A}\mathbf{R}^2\right), \end{aligned} \quad (20)$$

where $\lambda$ is the largest eigenvalue of $\mathbf{B}$. Further, we have equality in (20) at $\mathbf{R}=\mathbf{R}^{(m)}$.

*Proof:* See Appendix C. ∎

Applying Lemma 3 to (17), we get the final lower bound for (14) as follows:

$$\begin{aligned} &-\sum_{k=1}^{K}\mathrm{tr}\left(\mathbf{W}_{1,k}^{(m)}\mathbf{\Phi}^{\frac{1}{2}}+\mathbf{W}_{2,k}^{(m)}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)\\ &\qquad\geq-\mathrm{tr}\left(\mathbf{W}_A^{(m)}\mathbf{\Phi}^{\frac{1}{2}}+\mathbf{W}_B^{(m)}\mathbf{\Phi}\right), \end{aligned} \quad (21)$$

where

$$\mathbf{W}_A\triangleq\sum_{k=1}^{K}\left(\mathbf{W}_{1,k}+\left(\mathbf{S}_k-\lambda_{\max}\left(\mathbf{S}_k\right)\mathbf{I}_{KN}\right)\mathbf{\Phi}^{\frac{1}{2}}\mathbf{W}_{2,k}\right), \quad (22)$$

$$\mathbf{W}_B\triangleq\sum_{k=1}^{K}\lambda_{\max}\left(\mathbf{S}_k\right)\mathbf{W}_{2,k}, \quad (23)$$

and $\lambda_{\max}(\mathbf{S}_k)$ is the largest eigenvalue of $\mathbf{S}_k$. Note that the superscript $m$ in (17) and (21) is the iteration index. Also, we can compute the eigenvalues of $\mathbf{S}_k$ by multiplying the eigenvalues of the smaller dimensional matrix $\hat{\hat{\mathbf{H}}}_k^H\hat{\hat{\mathbf{H}}}_k/\sigma^2$ by the number of users in the system. Thus, $\lambda_{\max}(\mathbf{S}_k)$ needs to be computed only once and stored in the memory. Combining (21) with $\mathrm{tr}\left(\mathbf{Q}^{(m)}\mathbf{\Phi}\right)$ in (13), the optimization problem we wish to solve becomes

$$\begin{aligned} &\{\mathbf{\Phi}^{(m+1)}\}=\\ &\qquad\underset{\mathbf{\Phi}}{\mathrm{argmax}}\left\{-\mathrm{tr}\left(\mathbf{Q}^{(m)}\mathbf{\Phi}+\mathbf{W}_A^{(m)}\mathbf{\Phi}^{\frac{1}{2}}+\mathbf{W}_B^{(m)}\mathbf{\Phi}\right)\right\}\\ &\qquad\text{subject to}\quad\mathrm{tr}\left(\mathbf{\Phi}\right)\leq 1. \end{aligned} \quad (24)$$

*Lemma 4:* The optimization problem in (24) has a locally optimal solution given by

$$P(i)=\left(\frac{\left[\mathbf{W}_A^{(m)}\right]_{(i,i)}}{2\left(\left[\mathbf{W}_B^{(m)}\right]_{(i,i)}+\left[\mathbf{Q}^{(m)}\right]_{(i,i)}+\eta\right)}\right)^2,\quad\forall i, \quad (25)$$

where $\eta$ is chosen to satisfy $\sum_{i=1}^{KN}P(i)=1$.

*Proof:* See Appendix D. ∎

Using the solution for $P(i)$, one can construct the new surrogate function that needs to be optimized in the next

iteration. Iterating the process of computing $P(i)$, we arrive at a locally optimal joint power and beamforming vector allocation solution for maximizing the sum rate.

We next present an alternative bounding approach which leads to a different MM procedure for sum rate maximization.

### *B. Inverse Minorization Maximization Procedure*

We now return to the original optimization problem in (9). Recall that $\mathbf{\Phi} \in \mathbb{R}^{KN \times KN}$ is the augmented transmit power allocation matrix defined in (10). For convenience, let us define an augmented covariance matrix $\widetilde{\mathbf{\Phi}} \in \mathbb{R}^{(KN+N_r)\times(KN+N_r)}$, an augmented channel matrix $\widetilde{\mathbf{\Psi}}_k \in \mathbb{C}^{N_r \times (KN+N_r)}$ and the matrix $\mathbf{\Xi}_k \in \mathbb{C}^{N_r \times N_r}$ as follows:

$$\widetilde{\mathbf{\Phi}} \triangleq \text{diag}\left(\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K, \sigma^2 \mathbf{I}_{N_r}\right), \tag{26}$$

$$\widetilde{\mathbf{\Psi}}_k \triangleq \left[\widehat{\widehat{\mathbf{H}}}_k, \ldots, \widehat{\widehat{\mathbf{H}}}_k, \mathbf{I}_{N_r}\right], \tag{27}$$

$$\mathbf{\Xi}_k \triangleq \widetilde{\mathbf{\Psi}}_k \widetilde{\mathbf{\Phi}} \widetilde{\mathbf{\Psi}}_k^H, \tag{28}$$

$k = 1, \ldots, K$. In the definition of $\widetilde{\mathbf{\Phi}}$ above, the matrix $\widehat{\widehat{\mathbf{H}}}_k$ is repeated $K$ times. Then, we can rewrite the term inside the square brackets in (9) as $\mathbf{B}_k^{(m)} \mathbf{\Xi}_k^{-1}$. Note that the matrix $\mathbf{\Xi}_k \, \forall \, k$ is p.s.d., which will be useful in showing that the optimization problem has a feasible solution.

In order to develop the IMM procedure, we start with the following proposition from [31].

*Proposition 1:* Let $\mathbf{R}$ be an $m \times n$ matrix and $\mathbf{A}$ be an $m \times m$ p.s.d. matrix. We can upper bound the function $f(\mathbf{U}) \triangleq \text{tr}\left(\mathbf{A}\left(\mathbf{R}\mathbf{U}\mathbf{R}^H\right)^{-1}\right)$ as $f(\mathbf{U}) \leq \text{tr}\Big(\mathbf{A}\big(\mathbf{R}\mathbf{U}^{(m)}\mathbf{R}^H\big)^{-1}\mathbf{R}\mathbf{U}^{(m)}\mathbf{U}^{-1}\mathbf{U}^{(m)}\mathbf{R}^H\big(\mathbf{R}\mathbf{U}^{(m)}\mathbf{R}^H\big)^{-1}\Big)$, with equality at $\mathbf{U} = \mathbf{U}^{(m)}$.

Since $\mathbf{B}_k^{(m)} \succeq 0 \; \forall k$, we can apply proposition 1 to $\text{tr}\left(\mathbf{B}_k^{(m)} \mathbf{\Xi}_k^{-1}\right)$, which leads to

$$\sum_{k=1}^{K} \text{tr}\left(\mathbf{B}_k^{(m)} \mathbf{\Xi}_k^{-1}\right) \leq \sum_{k=1}^{K} \text{tr}\left(\mathbf{B}_k^{(m)} {\mathbf{\Xi}_k^{(m)}}^{-1} \widetilde{\mathbf{\Psi}}_k \widetilde{\mathbf{\Phi}}^{(m)} \widetilde{\mathbf{\Phi}}^{-1} \widetilde{\mathbf{\Phi}}^{(m)} \widetilde{\mathbf{\Psi}}_k^H {\mathbf{\Xi}_k^{(m)}}^{-1}\right) \tag{29}$$

$$= \text{tr}\left(\sum_{k=1}^{K} \widetilde{\mathbf{\Phi}}^{(m)} \widetilde{\mathbf{\Psi}}_k^H {\mathbf{\Xi}_k^{(m)}}^{-1} \widetilde{\mathbf{\Psi}}_k \widetilde{\mathbf{\Phi}}^{(m)} \widetilde{\mathbf{\Phi}}^{-1}\right), \tag{30}$$

where (30) is obtained by recognizing that $\mathbf{B}_k^{(m)} {\mathbf{\Xi}_k^{(m)}}^{-1}$ is the identity matrix, cyclically permuting the terms, and pulling the summation over $k$ into the trace function. In (30), the matrix $\tilde{\mathbf{\Phi}}$ is diagonal and positive semi-definite, which may become singular. On the other hand, Proposition 1 assumes it to be an invertible matrix for deriving the upper bound to the objective function of the optimization problem and obtain a closed form solution. However, this does not pose a problem in practice, because if some of the diagonal entries of $\tilde{\mathbf{\Phi}}$ become 0, we can remove the corresponding columns and rows of the matrices in the left hand side and right hand side of $\tilde{\mathbf{\Phi}}$ in (29), and form a new nonsingular $\tilde{\mathbf{\Phi}}$. During initialization, we allocate equal or random powers to all the users across all the beamforming vectors, which makes $\tilde{\mathbf{\Phi}}$ invertible. Therefore, without loss of generality, we can assume that $\tilde{\mathbf{\Phi}}$ is invertible. Further, in the final iterative algorithm, we compute the power allocations using closed form expressions, which converge to a stationary point of the optimization problem. Now, let

$$\mathbf{Z} \triangleq \sum_{k=1}^{K} \widetilde{\mathbf{\Phi}} \widetilde{\mathbf{\Psi}}_k^H \mathbf{\Xi}_k^{-1} \widetilde{\mathbf{\Psi}}_k \widetilde{\mathbf{\Phi}}. \tag{31}$$

Substituting $\text{tr}\left(\mathbf{Q}^{(m)}\mathbf{\Phi}\right)$ (from (13)) and (30) into (7), we get the following surrogate optimization problem:

$$\mathbf{\Phi}^{(m+1)} = \underset{\mathbf{\Phi} \succeq 0}{\arg\max} \left\{-\text{tr}\left(\mathbf{Q}^{(m)}\mathbf{\Phi} + \mathbf{Z}^{(m)}\widetilde{\mathbf{\Phi}}^{-1}\right)\right\} \quad \text{subject to} \quad \text{tr}\left(\mathbf{\Phi}\right) \leq 1, \tag{32}$$

where $m$ is the iteration index.

*Lemma 5:* The optimization problem in (32) has a locally optimal solution given by

$$P(i) = \left(\frac{\left[\mathbf{Z}^{(m)}\right]_{(i,i)}}{\left[\mathbf{Q}^{(m)}\right]_{(i,i)} + \eta}\right)^{\frac{1}{2}}, \quad i = 1, \ldots, KN, \tag{33}$$

where $\eta$ is chosen to satisfy $\sum_{i=1}^{KN} P(i) = 1$.

*Proof:* See Appendix E. ∎

In each iteration, we bound the cost function using proposition 1, and maximize the corresponding surrogate cost function using the solution in (33). Then, we recompute the bounding function, and the process repeats till convergence. Since (33) is strictly decreasing in $\eta$, we can determine the value of $\eta$ for which the solution satisfies the power constraint using a line search or bisection method [32].

The outcome of both the SMM and IMM procedures is the matrix $\mathbf{\Phi}$, which gives the individual users' powers across all the beamforming vectors. The power allocated to the $k^{\text{th}}$ user on the $j^{\text{th}}$ beamforming vector can be written using the solutions from the SMM or IMM procedure as

$$P_k(j) = P\left((k-1)\,N + j\right). \tag{34}$$

Pseudo-codes for SMM and IMM are shown in Appendix F.

TABLE I.
FLOP COUNT ORDER OF SMM PER ITERATION

| Matrix | Size | Flop Count |
|---|---|---|
| $\mathbf{S}_k$ | $KN \times KN$ | $K^2N^2N_r$ |
| $\mathbf{X}_k$ | $N_r \times KN$ | $KNN_r^2$ |
| $\mathbf{Y}_k^{-1}$ | $KN \times KN$ | $KNN_r^3$ |
| $\mathbf{W}_{1,k}$ | $KN \times KN$ | $KN$ |
| $\mathbf{W}_{2,k}$ | $KN \times KN$ | $K^2N^2N_r$ |
| $\mathbf{W}_{A,k}$ | $KN \times KN$ | $K^3N^3$ |
| $\mathbf{W}_{B,k}$ | $KN \times KN$ | $KN$ |
| $\mathbf{W}_A$ | $KN \times KN$ | $K^2N$ |
| $\mathbf{W}_B$ | $KN \times KN$ | $K^2N$ |

TABLE II.
FLOP COUNT ORDER OF IMM PER ITERATION

| Matrix | Size | Flop Count |
|---|---|---|
| $\mathbf{\Xi}_k$ | $N_r \times N_r$ | $KNN_r^2$ |
| $\mathbf{Z}$ | $(KN + N_r)$ $\times (KN + N_r)$ | $KNN_r^2$ |
| $\mathbf{Q}$ | $KN \times KN$ | $(N_r + K)N^2$ |

### C. Computational Complexity

We use floating point operations (flops) to quantify the computational complexity of the algorithms. We assume that the multiplication of a $p \times q$ matrix with a $q \times r$ matrix requires $\mathcal{O}(pqr)$ flops. The per-iteration computational complexity of the SMM and IMM algorithms are provided in Tables I and II, respectively. The flop counts account for the structural properties of the various matrices. For example, while computing the flop counts for the matrix $\mathbf{Z}$ or $\mathbf{Q}$, we only consider the computations involved in finding the diagonal entries of those matrices. Also, although the flop count for computing the matrix $\mathbf{Q}$ is mentioned only in Table I, it is common for both the SMM and IMM algorithms. The overall computational complexities of the SMM and IMM algorithms are of the order $\mathcal{O}(K^4N^3)$ and $\mathcal{O}(K^2N^2)$, respectively. Thus, the per-iteration complexity of IMM is lower than that of the SMM algorithm. Finally, we note that the complexity of SMM and IMM are independent of the number of transmit antennas. In practice, one would typically scale the size of the codebook with the number of transmit antennas, for example, as $N = 2^{N_t B}$ when a codebook with resolution $B$-bits per antenna is used. This can be substituted in the tables to infer the dependence of the complexity on the number of antennas.

## IV. Simulation Results

In this section, we evaluate the performance of the SMM and IMM algorithms using Monte Carlo simulations. We consider $N_t = \{8, 16, 32, 64, 128\}$ transmit antennas at the AP and $N_r = \{1, 2\}$ antennas at each UE. The number of UEs is varied from $K = 1$ to 10. The channel coefficients are drawn i.i.d. from $\mathcal{CN}(0, 1)$. The AWGN at the receivers is also distributed as $\mathcal{CN}(0, 1)$ and is independent across receive antennas. For channel state feedback, we quantize the CSI using 4 to 8 bits per channel coefficient. The dynamic range of the channel coefficients typically vary around three times the standard deviation from the mean, and we set the maximum and minimum values of the quantization levels accordingly, and use uniform quantization. We consider the size of the beamforming codebook varying from $N = 64$ (6 bits) to 1024 (10 bits), uniformly distributed on the $N$ dimensional complex unit sphere. The algorithms are initialized randomly for all the users across all the beams, such that the total power constraint is satisfied with equality. The algorithms are run till the normalized increase in the sum rate between two consecutive iterations is less than $10^{-4}$.

We compare the sum rate of the IMM algorithm against the WMMSE [12] and WSRMax [16] algorithms. These algorithms use perfect CSI to design the precoder matrix, and do not consider the codebook constraint in the optimization. Hence, for comparison with our work, we quantize the precoding vectors output by the above algorithms to the nearest vector in the codebook, and compare the sum rates achieved. We also compare against the eigen-mode beamforming (EBF), a heuristic precoding approach (User-BFVec Selection) in which each UE maximizes its achievable sum rate using the IMM algorithm and feeds back the selected beamforming vectors to the AP, and the use of CVX [25], [26] to solve (9).

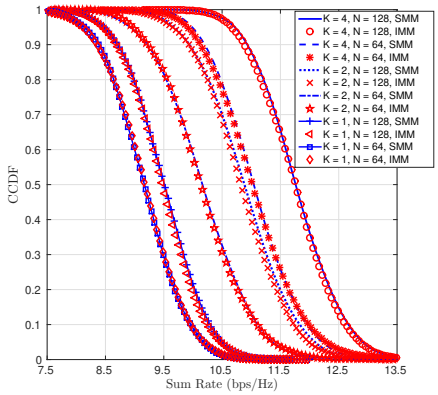
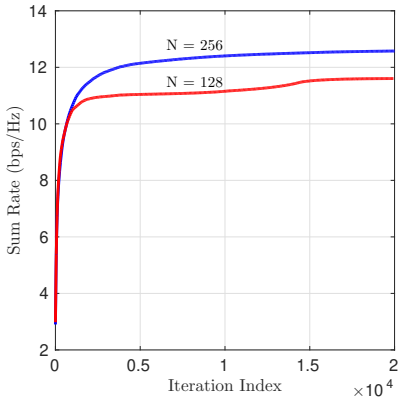
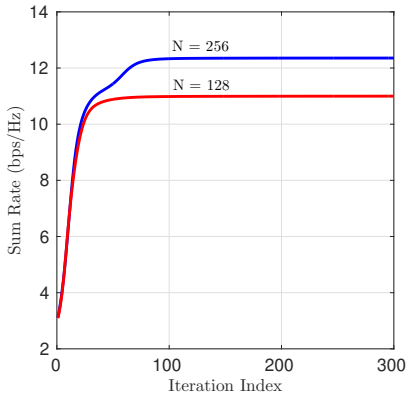



Fig. 1. CCDF comparison between SMM and IMM for data SNR = 10 dB, $N_r = 2, N_t = 16$. The distribution of the sum rates achieved by SMM and IMM are almost the same.




(a) SMM convergence.




(b) IMM convergence.

Fig. 2. Convergence behavior of the SMM and IMM procedures, $K = 4, N_r = 2, N_t = 16$, Data SNR = 10 dB.

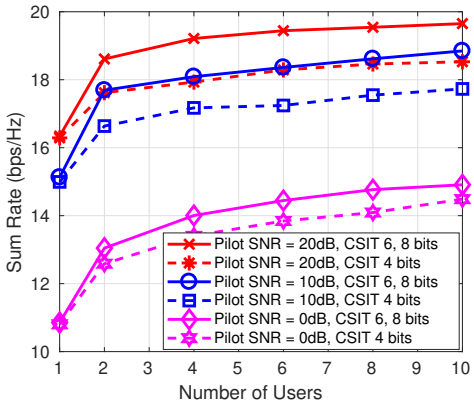



Fig. 3. Sum rate vs. $K$, $N = 512, N_r = 2, N_t = 16$, data SNR = 20 dB. The sum rate improves with the number of CSI quantization bits, but beyond 6 bits, the performance improvement is negligible.

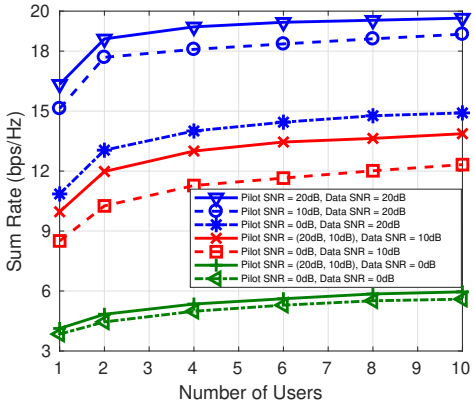



Fig. 4. Sum rate vs. $K$, $N = 512, N_r = 2, N_t = 16$, CSI quantized to 6 bits. The sum rate improves with pilot SNR, but the improvement is marginal once the pilot SNR exceeds the data SNR.

Figure 1 shows the complementary cumulative distribution function (CCDF) of the achieved sum rates for the SMM and IMM algorithms at an SNR of 10 dB. We see that the two approaches offer similar sum rates. Figure 2 illustrates that although both SMM and IMM algorithms exhibit monotonic convergence, IMM converges much faster than SMM. This highlights the impact of the choice of the surrogate function on the rate of convergence [31]. In SMM, we apply the minorization three times to lower bound the objective function, whereas, in IMM, the minorization is applied twice. Also, SMM makes use of the first order Taylor series expansion, whereas, IMM uses a matrix inequality to find the surrogate function. These differences result in the different rates of convergence of the two procedures. We note that the associated matrix inequalities (Lemmas 2, 3, and Proposition 1) are potentially be useful in other problem scenarios. Given their similar performance, in the sequel, we do not include the SMM algorithm in the performance plots, to avoid clutter.

Figure 3 shows the achievable sum rate vs. the number of users, with $N = 512, N_r = 2, N_t = 16$ and data SNR 20 dB. The sum rate improves with finer quantization of channel estimates, as expected. However, increasing the number bits of quantization beyond 6 has a negligible impact on the performance. So, we have merged the curves for 6 bits and 8 bits to avoid clutter.

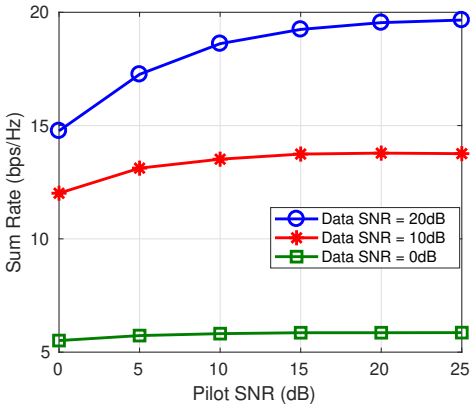



Fig. 5. Sum rate vs. pilot SNR for $K = 8, N = 512, N_r = 2, N_t = 16$, CSI quantized to 6 bits. We observe a monotonic increase in the sum rate with pilot and data SNR. Eventually, the sum rate saturates beyond a certain level of the pilot SNR but the saturation point moves to the right as the data SNR increases.

Figure 4 shows the achievable sum rate vs. the number of users, with $N = 512, N_r = 2, N_t = 16$ and CSI quantized to 6 bits per channel coefficient. We have merged the 10 dB and 20 dB pilot SNR plots, as their performances were very similar. While the sum rate increases, the rate of increase decreases with the number of users. As $K$ is increased from $K = 1$, the algorithm has a larger number of choices to assign resources, which improves the sum rate. However, this only offers a marginal benefit at a given pilot/data SNR for large $K$. Also, the sum rate improvement with pilot SNR becomes marginal once it exceeds the data SNR. Thus, the AP can use these results to determine the pilot and data transmission powers in order to, for example, maximize energy efficiency while achieving a desired rate.

Figure 5 shows the sum rate vs. the pilot SNR, with $N = 512, K = 8, N_r = 2, N_t = 16$, and CSI quantized to 6 bits. The sum rate monotonically increases with the pilot SNR, but saturates as the pilot SNR exceeds the data SNR, as observed earlier. From the achievable sum rate expression, the tipping point occurs when the channel estimation error variance is of the same order as that of the AWGN. The channel estimation error decreases linearly with the pilot SNR, while the residual interference caused due to channel estimation error increases linearly with the data SNR. Hence, if the pilot power scales linearly with the data power, it results in a roughly constant interference in the denominator of the SINR. Interestingly, this intuition continues to hold for the sum rate even after the beams and corresponding data powers are optimally chosen by the IMM algorithm. Hence, increasing the pilot SNR beyond the data SNR only marginally improves the sum rate, and the system becomes noise and multiuser interference limited.

Figure 6 shows the sum rate vs. the data SNR, with $N = 512$, $K = 8, N_r = 2, N_t = 16$ and CSI quantized to 6 bits. The sum rate initially increases linearly with the data SNR, but begins to saturate once the data SNR exceeds the pilot SNR. For example, the sum rates achieved with pilot SNR $= 10$ dB and 20 dB match till a data SNR of 10 dB, with the sum rate increasing linearly with the data SNR. Beyond a data SNR of 10 dB, the performance with pilot SNR $= 10$ dB becomes limited by channel estimation errors, and the sum rate starts

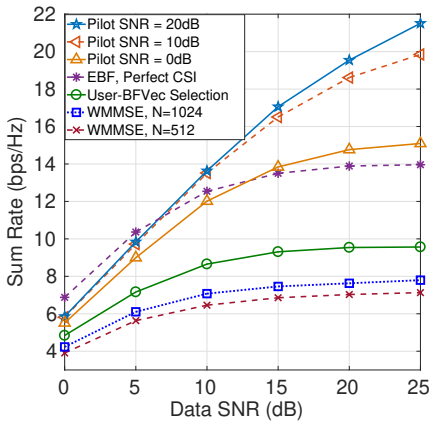



Fig. 6. Sum rate vs. data SNR for $K = 8, N = 512, N_r = 2, N_t = 16$, CSI quantized to 6 bits. We observe a monotonic increase in the sum rate with pilot and data SNR. The sum rate performance for the values of pilot SNRs 10 dB and 20 dB are almost the same till the data SNR reaches 10 dB beyond which they diverge.

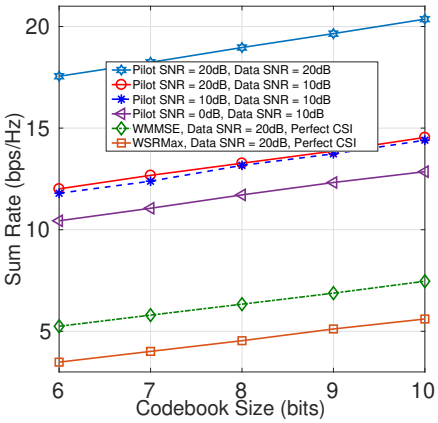



Fig. 7. Sum rate vs. codebook size for $K = 10, N_r = 2, N_t = 16$, CSI quantized to 6 bits. We observe a monotonic increase in the sum rate with pilot and data SNR. As the codebook size increases, the AP has more number of beamforming vectors to choose resulting in an increase in the sum rate performance.

to saturate. This behavior is consistent with the observations made for Fig. 5. We also illustrate the sum rate performance of the WMMSE algorithm Fig. 6. We see that the sum rate performance of WMMSE is far inferior than that achieved by the MM based approaches when its precoder outputs are quantized to the nearest vectors in the codebook. The sum rate obtained using the non-codebook based WMMSE is higher than that of the MM algorithms, but it comes with the high overhead of conveying the precoding matrices to the users. For example, at a data SNR of 10 dB, the sum rate achieved by the unconstrained WMMSE is 36 bps/Hz (not shown in the figure), but when the precoders are quantized to the nearest beamforming vectors in the codebook, it drastically reduces to 6.46 bps/Hz. This illustrates the importance of considering the codebook constraint while solving the beamforming vector assignment and power allocation problem.

In Fig. 6, we also show the performance of the eigen mode beamforming (EBF) and a heuristic precoding approach (User-BFVec Selection) in which each UE maximizes its achievable sum rate using the IMM algorithm and feeds back the selected beamforming vectors to the AP. In the EBF method, each UE chooses the beamforming vectors as the $N_r$ dominant right singular vectors, and feeds back their quantized versions. In our simulations, we use 32 bits to quantize each selected beamforming vector. Note that, the feedback overhead associated with this is higher than the feedback overhead associated with the IMM approaches. The EBF approach outperforms the IMM algorithm at low SNRs. This is because the noise dominates the interference, and the multiuser interference term does not significantly affect the achievable rate. At higher SNRs, the multiuser interference terms dominate the noise terms, and the IMM algorithm performs better than the other approaches as it is able to mitigate the multiuser interference through the joint selection of BF vectors. Also, User-BFVec selection approach performs worse than the IMM algorithm because of its inability to suppress the multiuser interference. Note that the feedback associated with this approach is the same as that of the IMM algorithm.

Figure 7 shows the sum rate performance vs. the codebook size (in bits), with $K = 10, N_r = 2$ and $N_t = 16$. The sum

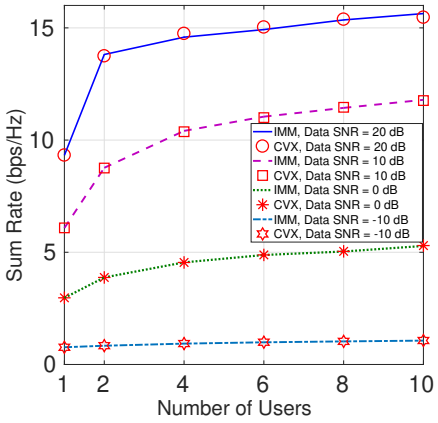



Fig. 8. Sum Rate vs. $K$ with $N_r = 1$, $N_t = 32$, Pilot SNR=20 dB, $N = 1024$ for SMM, IMM and CVX. We see that the MM based algorithms and CVX converge to almost same sum rate for various values of data SNR.

rate increases linearly with the codebook size, as the AP has more choices to select the beamforming vectors, which helps in canceling the multiuser interference. Note that, although the complexity of the problem increases with the codebook size, if the computational resources are limited, the IMM algorithm can be stopped at any iteration, resulting in a correspondingly effective solution. The figure also compares the the SMM and IMM algorithms with the WMMSE and WSRMax procedures. The sum rates achieved by WMMSE and WSRMax when the precoding vectors are quantized to the nearest vectors in the codebook are far inferior compared to that achieved by the proposed algorithms. Once again, this illustrates the importance of accounting for the codebook constraint while solving the sum rate optimization problem.

In Figure 8, we compare the sum rates achieved by directly solving (9) using CVX with the MM based algorithms. We see that both solutions converge to almost the same sum rates. However, the IMM algorithm exhibits faster run time, as shown in Figures 9 and 10. In these figures, we plot the ratio of the run times of the IMM and the convex solver CVX for various values of data SNRs and number of transmit antennas and number of users. The pilot SNR and the codebook size are set to 20 dB and $\{512, 1024\}$, respectively. We observe

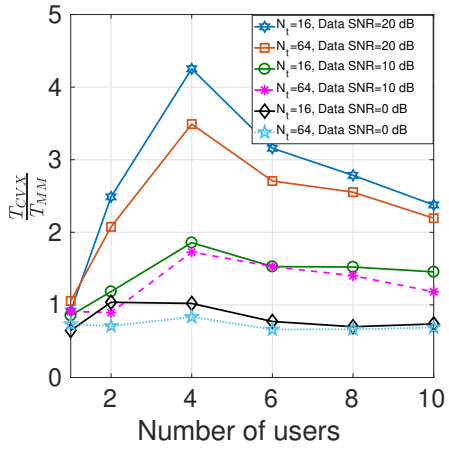



Fig. 9. Ratio of run times of CVX and IMM with respect to $K$. Pilot SNR = 20 dB, $N_r = 1, N = 1024$. We see that IMM converges much faster than CVX in the interference limited regime.

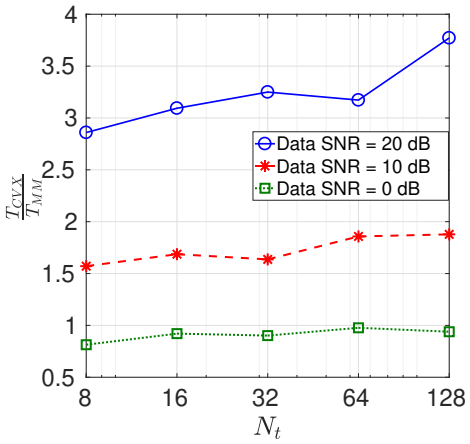



Fig. 10. Ratio of run times of CVX and IMM with respect to $N_t$. Pilot SNR = 20 dB, $N_r = 1, N = 512, K = 8$.

that as the data SNR increases, IMM converges much faster than the CVX package. This is thanks to the closed form expressions obtained for the final optimization problem. Also, as the number of transmit antennas increases, the ratio of the run times of CVX and IMM increases, which shows that IMM converges faster for large MIMO systems.

## V. Conclusions

We proposed two procedures, named square root MM (SMM) and inverse MM (IMM), to solve the problem of codebook based DL sum rate maximization in a FDD MU-MIMO broadcast system. Both SMM and IMM procedures find a locally optimal allocation of beamforming vectors and data signal powers to each user, so as to maximize the DL sum rate. These procedures are based on a nested application of the MM principle to lower bound the objective function, which is then maximized in an iterative fashion. The novelty of the algorithms lies in the choice of the surrogate functions used to bound the objective function. We also proved the optimality of the solutions to the surrogate optimization problems, as a consequence of which, the SMM and IMM algorithms reach a local optimum of the overall sum rate from any initialization.

We empirically illustrated the dependence of the achieved sum rate on the number of antennas, users, pilot power, and size of the codebook. We compared the sum rate of the SMM and IMM algorithms with that of WMMSE and WSRMax algorithms from the existing literature. The comparisons illustrated the importance of accounting for the codebook constraint while solving the sum rate optimization problem. Future work could include a theoretical analysis of the dependence of the optimal sum rate on parameters such as the codebook size, pilot/data SNR, and number of antennas.

## Appendix A
## Derivation of (4)

We simplify the covariance matrix of interest as follows:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\mathbf{H}}_k \mathbf{x}\mathbf{x}^H \widetilde{\mathbf{H}}_k^H\right] &= \mathbb{E}\left[\widetilde{\mathbf{H}}_k \mathbf{C} \sum_{j=1}^{K} \mathbf{\Phi}_j \mathbf{C}^H \widetilde{\mathbf{H}}_k^H\right] \\
&= \mathbb{E}\left[\widetilde{\mathbf{H}}_k \mathbf{C} \mathbf{\Phi}^{\frac{1}{2}} \mathbf{\Phi}^{\frac{1}{2}} \mathbf{C}^H \widetilde{\mathbf{H}}_k^H\right] \\
&= \mathbb{E}\left( \begin{bmatrix} \widetilde{\mathbf{h}_k^{(1)}}^T \\ \vdots \\ \widetilde{\mathbf{h}_k^{(N_r)}}^T \end{bmatrix} \left[\sqrt{P^{(1)}}\mathbf{c}_1, \ldots, \sqrt{P^{(N)}}\mathbf{c}_N\right] \right. \\
&\qquad \left. \times \begin{bmatrix} \sqrt{P^{(1)}}\mathbf{c}_1^H \\ \vdots \\ \sqrt{P^{(N)}}\mathbf{c}_N^H \end{bmatrix} \left[\widetilde{\mathbf{h}_k^{(1)}}^*, \ldots, \widetilde{\mathbf{h}_k^{(N_r)}}^*\right] \right) \\
&= \mathbb{E}\left( \begin{bmatrix} \sqrt{P^{(1)}}\mathbf{c}_1^T \widetilde{\mathbf{h}_k^{(1)}} & \ldots & \sqrt{P^{(N)}}\mathbf{c}_N^T \widetilde{\mathbf{h}_k^{(1)}} \\ \vdots & \vdots & \vdots \\ \sqrt{P^{(1)}}\mathbf{c}_1^T \widetilde{\mathbf{h}_k^{(N_r)}} & \ldots & \sqrt{P^{(N)}}\mathbf{c}_N^T \widetilde{\mathbf{h}_k^{(N_r)}} \end{bmatrix} \right. \\
&\qquad \left. \times \begin{bmatrix} \sqrt{P^{(1)}}\widetilde{\mathbf{h}_k^{(1)}}^H \mathbf{c}_1^* & \ldots & \sqrt{P^{(1)}}\widetilde{\mathbf{h}_k^{(N_r)}}^H \mathbf{c}_1^* \\ \vdots & \vdots & \vdots \\ \sqrt{P^{(N)}}\widetilde{\mathbf{h}_k^{(1)}}^H \mathbf{c}_N^* & \ldots & \sqrt{P^{(N)}}\widetilde{\mathbf{h}_k^{(N_r)}}^H \mathbf{c}_N^* \end{bmatrix} \right) \\
&= (\beta_k - \gamma_k) \sum_{n=1}^{N} P^{(n)} \mathbf{I}_{N_r} \\
&= (\beta_k - \gamma_k) \operatorname{tr}\left(\sum_{j=1}^{K} \mathbf{\Phi}_j\right) \mathbf{I}_{N_r} = (\beta_k - \gamma_k)\, \mathbf{I}_{N_r}. \qquad (35)
\end{aligned}$$

In the above derivation, $P^{(n)} = \sum_{j=1}^{N} P_j(n)$, i.e., the sum of powers allocated to all the users in the $n^{\text{th}}$ beamforming vector. Also, $\widetilde{\mathbf{h}_k^{(l)}}$ denotes the $l^{\text{th}}$ row of the matrix $\widetilde{\mathbf{H}}_k$.

## Appendix B
## Proof of Lemma 2

Using the Woodbury identity for the inverse term in the function $f$ in (15), we get

$$-\operatorname{tr}\left(\mathbf{A}\left(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\right)^{-1} \times \mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\right)\mathbf{A}^H\right)$$

$$= -\operatorname{tr}\left(\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^H\right) + \operatorname{tr}\left(\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\right), \quad (36)$$

where $\mathbf{X} \triangleq \mathbf{A}\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}$ and $\mathbf{Y} \triangleq \mathbf{I} + \mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}$. The function $\operatorname{tr}\left(\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\right)$ is jointly convex in $\mathbf{X}$ and $\mathbf{Y}$, and, can be minorized using a first order Taylor series. The complex matrix differential of $\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H$ is computed as follows [33]:

$$\begin{aligned}&\operatorname{tr}\left(\mathrm{d}\left(\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\right)\right)\\ &= \operatorname{tr}\left(\mathbf{Y}^{-1}\mathbf{X}^H\mathrm{d}\mathbf{X} - \mathbf{Y}^{-1}\mathbf{X}^H\mathbf{X}\mathbf{Y}^{-1}\mathrm{d}\mathbf{Y} + \mathbf{X}\mathbf{Y}^{-1}\mathrm{d}\mathbf{X}^H\right).\end{aligned}$$

Thus, around the point $(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$, (36) can be lower bounded as

$$\begin{aligned}&-\mathrm{tr}(\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^H) + \operatorname{tr}\left(\mathbf{X}\mathbf{Y}^{-1}\mathbf{X}^H\right)\\ &\geq -\mathrm{tr}(\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^H) + \operatorname{tr}\Big(\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\left(\mathbf{X} - \hat{\mathbf{X}}\right)\\ &\qquad -\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\left(\mathbf{X} - \hat{\mathbf{X}}\right)^H\Big)\\ &= -\mathrm{tr}(\hat{\mathbf{K}}) + \operatorname{tr}\Big(\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\mathbf{A}\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\\ &\qquad - \hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\\ &\qquad + \hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\left(\mathbf{A}\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\right)^H\Big)\\ &= -\mathrm{tr}(\hat{\mathbf{K}}) + \operatorname{tr}\Big(\hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\mathbf{A}\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\\ &\qquad + \mathbf{C}^H\mathbf{B}^{-H}\mathbf{A}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\mathbf{R}^{\frac{1}{2}}\\ &\qquad - \hat{\mathbf{Y}}^{-1}\hat{\mathbf{X}}^H\hat{\mathbf{X}}\hat{\mathbf{Y}}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{C}^H\mathbf{B}^{-1}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\Big),\end{aligned}$$

where $\hat{\mathbf{K}}$ is as defined in Lemma 2. Finally, grouping the constant matrices together, we get (16).

## Appendix C
## Proof of Lemma 3

We lower bound the function $f(\mathbf{R})$ using $\lambda$, the largest eigenvalue of the matrix $\mathbf{B}$, as follows:

$$\begin{aligned}f(\mathbf{R}) &= -\operatorname{tr}\left(\mathbf{A}\mathbf{R}(\mathbf{B} - \lambda\mathbf{I})\mathbf{R} + \lambda\mathbf{A}\mathbf{R}^2\right)\\ &= -\operatorname{tr}\left(\mathbf{A}\mathbf{R}\mathbf{C}\mathbf{R} + \lambda\mathbf{A}\mathbf{R}^2\right), \quad (37)\end{aligned}$$

where $\mathbf{C} \triangleq (\mathbf{B} - \lambda\mathbf{I})$. The complex matrix differential of the first term in (37) is [33]

$$\operatorname{tr}\left(\mathrm{d}(\mathbf{A}\mathbf{R}\mathbf{C}\mathbf{R})\right) = \operatorname{tr}\left(\mathbf{C}\mathbf{R}\mathbf{A}(\mathrm{d}\mathbf{R}) + \mathbf{A}\mathbf{R}\mathbf{C}(\mathrm{d}\mathbf{R})\right).$$

Hence, around the previous iterate $\mathbf{R}^{(m)}$, a lower bound on $f$ can be written as

$$\begin{aligned}f(\mathbf{R}) \geq &-\operatorname{tr}\left(\mathbf{A}\mathbf{R}^{(m)}\mathbf{B}\mathbf{R}^{(m)}\right)\\ &- \operatorname{tr}\left(\left(\mathbf{C}\mathbf{R}^{(m)}\mathbf{A} + \mathbf{A}\mathbf{R}^{(m)}\mathbf{C}\right)\left(\mathbf{R} - \mathbf{R}^{(m)}\right)\right)\\ &- \lambda\mathrm{tr}\left(\mathbf{A}\mathbf{R}^2\right). \quad (38)\end{aligned}$$

Grouping the constant terms in (38) together and substituting for $\mathbf{C}$, we get (20).

## Appendix D
## Proof of Lemma 4

Note that (24) is a separable convex optimization problem, which can be solved in closed form using the Lagrangian method. The Lagrangian for (24) is given by

$$\begin{aligned}&\sum_{i=1}^{KN}\left([\mathbf{Q}^{(m)}]_{(i,i)}P(i) + [\mathbf{W}_A^{(m)}]_{(i,i)}P(i)^{\frac{1}{2}} + [\mathbf{W}_B^{(m)}]_{(i,i)}P(i)\right)\\ &\qquad + \eta\left(\sum_{i=1}^{KN}P(i) - 1\right), \quad (39)\end{aligned}$$

where $P(i), i = 1, 2, \ldots, KN$ denote the diagonal entries of $\mathbf{\Phi}$. Note that the sign of the objective function in (24) has been flipped while forming the Lagrangian, which makes it a minimization problem. By straightforward differentiation with respect to $P(i)$ in (39), we obtain the closed form solution

$$P(i) = \left(\frac{\left[\mathbf{W}_A^{(m)}\right]_{(i,i)}}{2\left(\left[\mathbf{W}_B^{(m)}\right]_{(i,i)} + \left[\mathbf{Q}^{(m)}\right]_{(i,i)} + \eta\right)}\right)^2, \quad (40)$$

where $\eta$ is chosen to satisfy $\sum_{i=1}^{KN}P(i) = 1$. Since the $P(i)$ in (40) is strictly decreasing in $\eta$, it can be found using a simple line search or the bisection method [32].

Now, we show that the solution in (40) satisfies the second order sufficiency condition for optimality. That is, we show that the Hessian matrix of the Lagrangian in (39) is positive definite. The Hessian matrix of the Lagrangian in (39) is diagonal, with the $i^{\text{th}}$ diagonal entry $-[\mathbf{W}_A^{(m)}]_{(i,i)}/(4P^{3/2}(i))$, where $P(i)$ is given by (40). Thus, we need to show that the diagonal entries of $\mathbf{W}_A$ are strictly negative.

For simplicity and without loss of generality, we assume that $\sigma^2 = 1$. First, we simplify the first term, $\mathbf{W}_{1,k}$, of $\mathbf{W}_A$ in (22). Note that, $\mathbf{W}_{1,k}$, defined in (19), is a negative semidefinite matrix, as it is the sum of the matrix $\left(-\mathbf{Y}_k^{-1}\mathbf{X}_k^H\mathbf{F}_k\mathbf{\Psi}_k\right)$ and its conjugate transpose. The first term $\mathbf{Y}_k^{-1}\mathbf{X}_k^H\mathbf{F}_k\mathbf{\Psi}_k$ becomes

$$\begin{aligned}&\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}^H\mathbf{F}_k^H\mathbf{F}_k\mathbf{\Psi}\\ &= \left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}^H\left(\mathbf{I}_{N_r} + \mathbf{\Psi}_k\mathbf{\Phi}\mathbf{\Psi}_k^H\right)\mathbf{\Psi}_k\\ &= \left(\mathbf{\Phi}^{-\frac{1}{2}} + \mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\left(\mathbf{I}_{KN} + \mathbf{S}_k\mathbf{\Phi}\right)\mathbf{S}_k\\ &= \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k.\end{aligned}$$

Thus, $\mathbf{W}_{1,k}$ can be written as

$$\mathbf{W}_{1,k} = -\left(\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k + \mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right). \quad (41)$$

Define $\hat{\lambda}_k \triangleq \lambda_{\max}(\mathbf{S}_k)$. The term $\left(\mathbf{S}_k - \hat{\lambda}_k\mathbf{I}_{KN}\right)\mathbf{\Phi}^{\frac{1}{2}}\mathbf{W}_{2,k}$ in the expression for $\mathbf{W}_A$ in (22) becomes

$$\begin{aligned}&\left(\mathbf{S}_k - \hat{\lambda}_k\mathbf{I}_{KN}\right)\mathbf{\Phi}^{\frac{1}{2}}\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}_k^H\\ &\qquad\times\left(\mathbf{I}_{N_r} + \mathbf{\Psi}_k\mathbf{\Phi}\mathbf{\Psi}_k^H\right)\mathbf{\Psi}_k\mathbf{\Phi}^{\frac{1}{2}}\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\\ &= \left(\mathbf{S}_k - \hat{\lambda}_k\mathbf{I}_{KN}\right)\mathbf{\Phi}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}. \quad (42)\end{aligned}$$

Combining (41) and (42), we get the term inside the summation in the expression for $\mathbf{W}_A$ in (22) as follows:

$$-\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k - \mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}} + \left(\mathbf{S}_k - \hat{\lambda}_k\mathbf{I}_{KN}\right)\mathbf{\Phi}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}$$

$$\times \left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}$$
$$= -\left(\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k + \mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}} + \hat{\lambda}_k\mathbf{\Phi}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right) \times \left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}$$
$$= -\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k - \left(\mathbf{I}_{KN} + \hat{\lambda}_k\mathbf{\Phi}\right)\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}. \quad (43)$$

Now, we simplify the term $\left(\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\left(\mathbf{I}_{KN} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right)^{-1}\right)$ in (43). This becomes

$$\mathbf{S}_k\left(\mathbf{\Phi}^{-\frac{1}{2}} + \mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k\right)^{-1} = \mathbf{S}_k\left(\mathbf{\Phi}^{\frac{1}{2}} - \mathbf{\Phi}\left(\mathbf{I}_{KN} + \mathbf{S}_k\mathbf{\Phi}\right)^{-1}\mathbf{S}_k\mathbf{\Phi}^{\frac{1}{2}}\right) \quad (44)$$
$$= \left(\mathbf{S}_k - \mathbf{S}_k\left(\mathbf{\Phi}^{-1} + \mathbf{S}_k\right)^{-1}\mathbf{S}_k\right)\mathbf{\Phi}^{\frac{1}{2}}. \quad (45)$$

Here, the right hand side of (44) is obtained by applying the Woodbury matrix identity to the inverse term in the left hand side. Substituting (45) in (43), we get

$$\mathbf{W}_A = -\sum_{k=1}^{K}\left\{\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k + \left(\mathbf{I}_{KN} + \lambda_{\max}\left(\mathbf{S}_k\right)\mathbf{\Phi}\right)\mathbf{W}_{A1,k}\mathbf{\Phi}^{\frac{1}{2}}\right\},$$

where $\mathbf{W}_{A1,k} \triangleq \mathbf{S}_k - \mathbf{S}_k\left(\mathbf{\Phi}^{-1} + \mathbf{S}_k\right)^{-1}\mathbf{S}_k$. Since $\mathbf{S}_k$ is symmetric and p.s.d., the diagonal entries of $\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k$ are non-negative. Moreover, since $\mathbf{H}_k$ is drawn from a continuous valued distribution, the diagonal entries of $\mathbf{\Phi}^{\frac{1}{2}}\mathbf{S}_k$ are strictly positive with probability 1. Also, $(\mathbf{I}_{KN} + \lambda_{\max}\left(\mathbf{S}_k\right)\mathbf{\Phi})$, and $\mathbf{\Phi}^{\frac{1}{2}}$ are diagonal matrices with strictly positive entries on the diagonal. Finally, it is easy to show that the diagonal entries of $\mathbf{W}_{A1,k}$ are also non-negative. For this, it suffices to show that the diagonal entries of $\left\{\mathbf{S}_k - \mathbf{S}_k\left(\frac{1}{\lambda_{\max}(\mathbf{\Phi})}\mathbf{I}_{KN} + \mathbf{S}_k\right)^{-1}\mathbf{S}_k\right\}$ are non-negative, where $\lambda_{\max}(\mathbf{\Phi})$ is the largest eigenvalue of the matrix $\mathbf{\Phi}$. The eigenvalues of this matrix are given by $\lambda(\mathbf{S}_k)/\left(1 + \lambda_{\max}\left(\mathbf{\Phi}\right)\lambda\left(\mathbf{S}_k\right)\right)$, where $\lambda(\mathbf{S}_k)$ is an eigenvalue of $\mathbf{S}_k$. These eigenvalues are non-negative, and hence the eigenvalues of $\mathbf{W}_{A1,k}$ are also non-negative. Moreover, since $\mathbf{H}_k$ is drawn from a continuous valued distribution, the diagonal entries of $\mathbf{W}_{A1,k}$ are strictly positive with probability 1. Therefore, the diagonal entries of $\mathbf{W}_A$ are strictly negative, thus satisfying the second order sufficient conditions for optimality of the surrogate optimization problem in (24).

## Appendix E
## Proof of Lemma 5

The objective function is quadratic in $\mathbf{\Phi}$, and is therefore amenable to optimization. Further, the problem in (32) is separable in the optimization variables, and can be solved using the Lagrangian method to obtain a closed form solution. The Lagrangian is given by

$$\sum_{i=1}^{KN}\left(\left[\mathbf{Q}^{(m)}\right]_{(i,i)}P(i) + \left[\mathbf{Z}^{(m)}\right]_{(i,i)}\frac{1}{P(i)}\right) + \eta\left(\sum_{i=1}^{KN}P(i) - 1\right). \quad (46)$$

Note that, the negative sign in the optimization problem in (32) is removed while forming the Lagrangian, making it a minimization problem. By differentiating (46) with respect to $P(i)$, we obtain (33).

Now, we show that the solution to the surrogate convex optimization problem in (33) satisfies the second order sufficiency condition for optimality. The Hessian matrix of the Lagrangian in (46) is diagonal, with the $i^{\text{th}}$ diagonal entry $\{2[\mathbf{Z}^{(m)}]_{(i,i)}/P(i)^3\}$, where $P(i)$ is given by (33). Thus, we need to show that the diagonal entries of $\mathbf{Z}$ are strictly positive.

Without loss of generality, we assume that $\sigma^2 = 1$. We can simplify each term inside the summation in (31) as follows:

$$\widetilde{\mathbf{\Phi}}\left(\widetilde{\mathbf{S}}_k - \widetilde{\mathbf{S}}_k\left(\widetilde{\mathbf{\Phi}}^{-1} + \widetilde{\mathbf{S}}_k\right)^{-1}\widetilde{\mathbf{S}}_k\right)\widetilde{\mathbf{\Phi}},$$

where $\widetilde{\mathbf{S}}_k \triangleq \widetilde{\mathbf{\Psi}}_k^H\widetilde{\mathbf{\Psi}}_k$. By following a similar procedure as done for showing the optimality of the SMM solution in (40), we can show that the diagonal entries of $\mathbf{Z}$ are strictly positive with probability 1, and hence it satisfies the second order sufficient conditions for optimality of the solution (33) to the surrogate optimization problem (32).

## Appendix F

The pseudo codes for the SMM and IMM algorithms are provided in Algorithm 1 and 2, respectively.

**Algorithm 1** SMM

**Input:** $\widehat{\mathbf{H}}_1, \ldots, \widehat{\mathbf{H}}_K, \mathbf{C}, K, \rho_{dl}$
**Output:** $P_1(1), \ldots, P_1(N), \ldots, P_K(1), \ldots, P_K(N)$

```
Initialize P_1(1),...,P_1(N),...,P_K(1),...,P_K(N) to satisfy the total power constraint.
Compute Ĥ̂_k = √ρ_dl Ĥ_k C, k = 1,2,...,K.
Compute Ψ_1,...,Ψ_K using (11).
repeat
    Compute Φ using (10).
    Compute Q using (12).
    Compute W_A, W_B using (22), (23).
    Calculate Lagrange multiplier η using line search to satisfy maximum power constraint.
    Compute P(i) using (25), i = 1,2,...,KN.
    for k = 1 to K do
        for i = 1 to N do
            Compute P_k(i) using (34).
        end for
    end for
until convergence
```

## References


[1] S. S. Thoota, P. Babu, and C. R. Murthy, "Codebook based precoding for multiuser MIMO broadcast systems: An MM approach," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2019, pp. 1–6.

[2] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2532–2547, May 2015.

[3] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

**Algorithm 2** IMM

**Input:** $\widehat{\mathbf{H}}_1, \ldots, \widehat{\mathbf{H}}_K, \mathbf{C}, K, \rho_{dl}$
**Output:** $P_1(1), \ldots, P_1(N), \ldots, P_K(1), \ldots, P_K(N)$

Initialize $P_1(1), \ldots, P_1(N), \ldots, P_K(1), \ldots, P_K(N)$ to satisfy the total power constraint.
Compute $\widehat{\widehat{\mathbf{H}}}_k = \sqrt{\rho_{dl}}\widehat{\mathbf{H}}_k\mathbf{C}$, $k = 1, 2, \ldots, K$.
Compute $\widetilde{\mathbf{\Psi}}_1, \ldots, \widetilde{\mathbf{\Psi}}_K$ using (27).
**repeat**
  Compute $\mathbf{\Phi}, \widetilde{\mathbf{\Phi}}$ using (10), (26), respectively.
  Compute $\mathbf{Q}$ and $\mathbf{Z}$ using (12) and (31) respectively.
  Calculate Lagrange multiplier $\eta$ using line search to satisfy maximum power constraint.
  Compute $P(i)$ using (33), $i = 1, 2, \ldots, KN$.
  **for** $k = 1$ to $K$ **do**
    **for** $i = 1$ to $N$ **do**
      Compute $P_k(i)$ using (34).
    **end for**
  **end for**
**until** convergence


[4] E. G. Larsson and H. V. Poor, "Joint beamforming and broadcasting in massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 3058–3070, Apr. 2016.

[5] M. Schubert and H. Boche, "Solution of the multi-user downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Tech.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[6] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. Allerton Conf. on Commun., Control and Comput.*, Sep. 1999, pp. 987–996.

[7] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[8] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[9] "IEEE Std 802.15.3c-2009, Part 15.3: Wireless Medium Access Control and Physical Layer Specifications for High Rate Wireless Personal Area Networks Amendment 2: Millimeterwave-based Alternative Physical Layer Extension," *IEEE Computer Society*, 2009.

[10] "IEEE standard for information technology–telecommunications and information exchange between systems, local and metropolitan area networks–specific requirements–part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications–amendment 4: Enhancements for very high throughput for operation in bands below 6 GHz." *IEEE Std 802.11ac(TM)-2013*, pp. 1–425, Dec 2013.

[11] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[12] 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, Sep. 2011.

[13] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2362–2373, Jun. 2013.

[14] S. Zarei, W. Gerstacker, R. Schober, "A low-complexity linear precoding and power allocation scheme for downlink massive MIMO systems," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2013, pp. 285–290.

[15] D. H. N. Nguyen and T. Le-Ngoc, "Sum-rate maximization in the multicell MIMO broadcast channel with interference coordination," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1501–1513, Mar. 2014.

[16] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[17] S. Hu, Y. Zhang, X. Wang, and G. B. Giannakis, "Weighted sum-rate maximization for MIMO downlink systems powered by renewables," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5615–5625, Aug. 2016.

[18] H. V. Nguyen, V. D. Nguyen, and O. S. Shin, "Low-complexity precoding for sum rate maximization in downlink massive MIMO systems," *IEEE Wireless Comm. Letters*, vol. 6, no. 2, pp. 186–189, 2017.

[19] Y. Cheng and M. Pesavento, "Robust codebook-based downlink beamforming using mixed integer conic programming," in *Proc. ICASSP*, 2013.

[20] Y. Cheng, and M. Pesavento, "An optimal iterative algorithm for codebook-based downlink beamforming," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 775–778, Aug. 2013.

[21] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[22] ——, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

[23] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, Jul. 2016.

[24] E. Piovano and B. Clerckx, "Optimal DoF region of the $k$-user MISO BC with partial CSIT," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2368–2371, Nov. 2017.

[25] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[26] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[27] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[28] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[29] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[30] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[31] Y. Sun, P. Babu and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, Jul. 2016.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[33] Are Hjørungnes, *Complex-Valued Matrix Derivatives*. Cambridge University Press, 2011.


**Sai Subramanyam Thoota** (S'16) received the B. E. degree in ECE from the College of Engg., Guindy, Anna University, Chennai, India, in 2003, and the M. Tech. degree in EE from the Indian Institute of Technology, Madras, Chennai, India, in 2006. Since 2016, he is pursuing Ph. D at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. His research interests include massive MIMO and mmWave MIMO communication systems, sparse signal recovery and statistical signal processing.

**Prabhu Babu** Biography not available at the time of publication.

**Chandra R. Murthy** Biography not available at the time of publication.