Multicast Beamforming Design in Multicell Networks With Successive Group Decoding

Mehdi Ashraphijuo, Xiaodong Wang, *Fellow, IEEE*, and Meixia Tao, *Senior Member, IEEE*

***Abstract*— We consider a generic problem of multicast beamforming design in multicell networks where each base station (BS) has multiple independent messages to multicast and each user intends to decode an arbitrary subset of messages sent from all BSs using successive group decoding (SGD). We first formulate the total transmit power minimization problem subject to the constraints that a target rate vector is achievable by the SGDs at all receivers. This problem is a non-convex quadratically constrained quadratic program and NP-hard. We propose a new method based on solving a sequence of linearly regularized semi-definite programming (SDP) relaxation of the original problem that yields feasible and near-optimal solutions with high probability. Moreover, we propose a decentralized algorithm based on the alternating direction method of multipliers to solve each linearly regularized SDP, which consists of solving a quadratic program at the central controller, and closed-form analytic computations at each BS. Finally, we propose an iterative procedure for joint beamformer and rate optimization under the SGD framework. Numerical results confirm the superiority of the proposed beamformer design in both performance and complexity. It is also demonstrated that, compared with the traditional linear receivers, the SGD receivers achieve both significant rate improvement and energy savings.**

***Index Terms*— Successive group decoding, multicell multicast beamforming, semi-definite program, decentralized algorithm, ADMM.**

I. Introduction

THE physical layer multicast beamforming is a promising technology to increase the efficiency of transmitting a common message to multiple users in wireless communications. The multicast beamforming problem for quality-of-service (QoS) guarantee was first considered in [1]. The similar problem of multicasting to multiple co-channel groups was later studied in [2]. The design of coordinated multicast beamforming in multicell networks was studied in [3]. The design in several other scenarios such as massive MIMO systems, relay networks, and cache-enabled cloud radio access networks and under different practical constraints was investigated in [4]–[11]. Some other issues such as outage analysis and capacity limits were studied later in [12] and [13]. It is shown that the multicast beamforming problem is in general a non-convex quadratically constrained quadratic program (QCQP) and NP-hard [1]. In addition, the problem can be infeasible in multi-group and multicell scenarios due to co-channel interference [2], [3].

Interference is known to be a limiting factor in wireless networks. One venue of mitigating interference and increasing data rate is the deployment of advanced receiver techniques by exploiting the structure of interfering signals. In particular, the successive group decoding (SGD) technique [14] chooses an optimal subset of interferers to decode, in addition to decoding the desired signals, to maximize the transmission rate or to minimize the outage probability. The efficient rate allocation in a single-antenna interference channel with single-codebook and fixed power per user has been investigated in [15] and [16]. In particular, the work [16] considers a K-user interference channel where each user employs the successive interference cancellation-based decoder and obtains a max-min fair decentralized rate allocation algorithm. The work [15] considers a K-user Gaussian interference channel and solves the problem of maximizing the desired user's rate at a particular receiver given the transmission rates of the other users. In addition, [15] proposes sequential and iterative rate allocation algorithms, which yield pareto-optimal rate-vectors albeit without a fairness guarantee. Later, [17] studied this approach of rate allocation in multiple-input single-output (MISO) cognitive radio channel when each secondary user has only one message to decode. Moreover, [18], [19] address the rate allocation problem where each user has multiple intended messages and there is a constraint on the number of users that can be jointly decoded at any stage. Recently, [6]–[8] applied rate splitting for multigroup multicasting problems.

This paper considers a generic multicast beamforming design problem in multicell networks where each base station (BS) transmits multiple messages and each user performs the SGD to decode multiple intended messages. We formulate the problem as a sum power minimization problem subject to the constraint that a target rate vector is achievable by the SGDs at all users. For the degenerative case that the SGD becomes the simple linear receiver and hence each user only desires one single message, our formulation unifies several existing works on multicast and unicast beamforming design such as [3] and [20]–[22].

In general, the problem of multicast beamforming design is non-convex and NP-hard. A common approach is to first solve

Manuscript received May 12, 2016; revised October 25, 2016 and February 12, 2017; accepted February 27, 2017. Date of publication March 17, 2017; date of current version June 8, 2017. This work was supported in part by the U.S. National Science Foundation under Grant CCF-1526215, and in part by the National Natural Science Foundation of China under Grant 61329101 and Grant 61571299. The associate editor coordinating the review of this paper and approving it for publication was Y. Jing.

M. Ashraphijuo and X. Wang are with the Electrical Engineering Department, Columbia University, New York, NY 10027 USA (e-mail: mehdi@ee.columbia.edu; wangx@ee.columbia.edu).

M. Tao is with the Electrical Engineering Department, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: mxtao@sjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2683478

a convex problem through semi-definite relaxation (SDR).[1] If the solution to the relaxed problem consists of all rank-one matrices, then we obtain the optimal solution to the original problem. However, this is not the case in general and in the literature, randomization methods have been proposed to obtain suboptimal rank-one solutions. Unfortunately, these methods yield infeasible solutions most of the time for the problem considered here due to the large number of constraints. We therefore propose a new approach to obtaining rank-one feasible solutions by solving a sequence of linearly regularized semi-definite programs, which is termed as successive linearly regularized semi-definite relaxation (SLR-SDR). The proposed linear regularization promotes low-rank solutions through simultaneous penalization of the trace function and uplift of off-diagonal elements of the non rank-one matrices. Numerical examples confirm that the proposed method yields rank-one solutions with very high probability.

Another contribution of this work is that we propose a decentralized algorithm to solve the linearly regularized semi-definite program (SDP), based on the alternating direction method of multipliers (ADMM) [23]. Each iteration of the proposed algorithm only involves a computationally cheap convex program at the central controller and evaluations of closed-form expressions at each BS locally. Thus our algorithm is very computationally efficient and highly scalable. In addition, the algorithm is guaranteed to converge to the optimal solution of the considered SDP problem.

Finally, we propose an iterative procedure for joint beamformer and rate optimization under the SGD framework. Simulation results show that the procedure converges usually in only 2 iterations.

The main contributions of this work are summarized as follows:

- We consider a very generic multicast network model where each cell (or BS) has multiple messages to send and each user aims to decode an arbitrary subset of the messages sent from all BSs. Our network model includes the conventional multi-cell multicast, multi-group multicast, hybrid multicast/unicast transmission, as well as multi-user unicast transmission as special cases. To facilitate the transmission and reception in such generic network model, each message is encoded using multiple codebooks (rate splitting) and each user employs the successive group decoding receiver to decode its intended messages (as opposed to the conventional linear receiver assumed by most works in the literature.)
- We propose a new and efficient rank-reduction method using linearly regularized SDP that significantly outperforms the existing methods of SDR with randomization.
- We propose an ADMM-based low complexity algorithm for solving the linearly regularized SDP. Compared with the traditional SDP solver, the proposed algorithm is decentralized without loss of any optimality.
- An iterative joint beamformer and rate optimization method is also proposed to successively improve the rate and reduce the transmission power, starting from the simple channel-matched beamformers.

The remainder of the paper is organized as follows. Section II describes the system under consideration. Section III provides the problem formulation and gives the proposed centralized solution. The proposed decentralized solution based on ADMM is given in Section IV. Section V proposes an iterative procedure for joint beamformer and rate optimization for SGD. Section VI provides simulation results and finally, the concluding remarks are made in Section VII.

II. System Descriptions

A. System Model

Consider a generic multicast network comprising M BSs and K mobile users. Assume that the j^{th} BS is equipped with M_j transmit antennas, $\forall j \in \{1, \ldots, M\}$, and each mobile user has one receive antenna. Each BS j, for $j = 1, \ldots, M$, has S_j $(\leq M_j)$ independent messages to transmit. Let us define $\mathcal{M}$ as the set of indices of all transmitted messages, i.e., $\mathcal{M} \triangleq \{(j, s) | j \in \{1, \ldots, M\}, s \in \{1, \ldots, S_j\}\}$. Also, denote $\mathcal{A}(k) \subseteq \mathcal{M}$ as the set of messages that the k^{th} user for $k = 1, \ldots, K$ aims to decode. Let $\mathbf{h}_{j,k} \in \mathbb{C}^{M_j}$ denote the frequency-flat quasi-static complex channel vector from the j^{th} BS to the k^{th} mobile user, $\forall j \in \{1, \ldots, M\}$, $k \in \{1, \ldots, K\}$. Let $\mathbf{w}_{j,s} \in \mathbb{C}^{M_j}$ denote a beamforming vector applied to the s^{th} message of the j^{th} BS, $\forall j \in \{1, \ldots, M\}, s \in \{1, \ldots, S_j\}$. Note that the message s from BS j can be desired by one or multiple users in the network. As a result, $\mathbf{w}_{j,s}$ is a general multicast beamformer by including unicast as a special case.

Moreover, we define L as the number of codebooks used per message and denote the set of codebooks of the s^{th} message of the j^{th} BS by $\mathcal{C}_{j,s} \triangleq \{C_{j,s,1}, \ldots, C_{j,s,L}\}$ where $C_{j,s,\ell}$ is a Gaussian codebook with rate $r_{j,s,\ell}$. Increasing L provides the receivers with more freedom to decide what fraction of the interfering messages to decode and what fraction to treat as noise. This is essentially the rate splitting method [24] and was considered in [17] and [18].

Define $x_{j,s,\ell}$ as the unit-power input from codebook $C_{j,s,\ell}$ to the channel. Then, we have

$$x_{j,s} = \sqrt{\frac{1}{L}} \sum_{\ell=1}^{L} x_{j,s,\ell}, \tag{1}$$

as the transmitted signal for the s^{th} message of the j^{th} BS which satisfies the unit-power constraint for all $j \in \{1, \ldots, M\}$, $s \in \{1, \ldots, S_j\}$, and consequently, the rate of the s^{th} message of the j^{th} BS is

$$r_{j,s} = \sum_{\ell=1}^{L} r_{j,s,\ell} \ , \quad \text{for all } j \in \{1, \ldots, M\}, s \in \{1, \ldots, S_j\}. \tag{2}$$

Note that although here for simplicity we assumed that each message is divided into L parts with equal power, all results can be extended to the case of unequal power. In that case,

[1] In general, there are several common methods to approximate non-convex QCQPs, including (a) SDR (prevailing); (b) reformulation linearization technique (RLT); (c) successive convex approximation (SCA). The drawback with approaches (b) and (c) is that they need a feasible point as initialization, which is difficult to obtain in general.

Eqn. (1) can be replaced with $x_{j,s} = \sqrt{\frac{P_{j,s,\ell}}{L}} \sum_{\ell=1}^{L} x_{j,s,\ell}$ such that $\sum_{\ell=1}^{L} P_{j,s,\ell} = 1$, $P_{j,s,\ell} \geq 0$, $\forall j, s$ and the rest of the paper can be updated accordingly.

Denote $\mathcal{B}(k) = \mathcal{M} \setminus \mathcal{A}(k)$ as the set of messages that the k^{th} user does not intend to decode, $\forall k \in \{1, \ldots, K\}$. Moreover for every set $\mathcal{D} \subseteq \mathcal{M}$, define $\mathcal{D}^* = \{(j, s, \ell)|(j, s) \in \mathcal{D}\}$.

The discrete-time baseband signal received by the k^{th} user, $\forall k \in \{1, \ldots, K\}$, is given by

$$\begin{aligned} y_k &= \sum_{j=1}^{M} \sum_{s=1}^{S_j} \mathbf{h}_{j,k}^H \mathbf{w}_{j,s} x_{j,s} + z_k \\ &= \sqrt{\frac{1}{L}} \sum_{(j,s,\ell) \in \mathcal{A}^*(k)} \mathbf{h}_{j,k}^H \mathbf{w}_{j,s} x_{j,s,\ell} \\ &\quad + \sqrt{\frac{1}{L}} \sum_{(j,s,\ell) \in \mathcal{B}^*(k)} \mathbf{h}_{j,k}^H \mathbf{w}_{j,s} x_{j,s,\ell} + z_k, \end{aligned} \tag{3}$$

where $z_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2)$ is an independent circularly symmetric complex Gaussian noise. The first term in (3) is the desired signal and the second term is the received interference.

B. Successive Group Decoding

Each user k performs the successive group decoding (SGD) [14], [18] to decode the messages in $\mathcal{A}^*(k)$. Specifically, given a set of beamformers $\{\mathbf{w}_{j,s}\}$, a target rate vector $\mathbf{r}^0 = \{r_{j,s,\ell}^0\}$, and a fairness vector $\mathbf{t} = \{t_{j,s,\ell}\}$, it first implements SGD(a)-(b) in the Appendix to obtain the optimal SGD schedule $\underline{Q}(k) = [Q_1(k), \ldots, Q_{p_k}(k)]$, and the final achievable rate vector $\mathbf{r} = \mathbf{r}^0 + x\mathbf{t}$, where $Q_m(k)$ denotes the set of messages to be jointly decoded during the m^{th} stage of SGD, p_k represents the number of SGD stages for the k^{th} user, and x is a scalar. Once $\{\underline{Q}(k)\}$ and $\mathbf{r}$ are computed, each user k then implements SGD(c) to decode the messages in $\mathcal{A}^*(k)$. We next briefly explain these two steps.

1) Computing the Optimal SGD Schedule and Achievable Rate Vector: For the k^{th} user, we define $\mathbf{r}_k^0 = \{r_{j,s,\ell}^0 : (j, s, \ell) \in \mathcal{A}^*(k)\}$, i.e., the sub-vector of $\mathbf{r}^0$ corresponding to the messages to be decoded by user k. Similarly define $\mathbf{t}_k$ and $\mathbf{r}_k$. Denote $\mathcal{H}_k = \{\mathbf{h}_{1,k}, \ldots, \mathbf{h}_{M,k}\}$ as the channel parameters from all transmitters to the k^{th} receiver. Each user k then implements SGD(a) independently to compute its SGD schedule $\underline{Q}(k) = \{Q_m(k)\}$ with $|Q_m(k)| \leq \mu_k$ where μ_k is the maximum size of the group of messages that can be jointly decoded by user k. We then have that $\forall u \in \mathcal{A}^*(k), \exists\ 1 \leq m \leq p_k$ such that $u \in Q_m(k)$. We also define $\overline{Q_i(k)} \triangleq \mathcal{B}^*(k) \cup_{j=i+1}^{p_k} Q_j(k)$.

The group decoding theory, including the optimality properties, is developed in [18]. It follows from [18] that for a set of given beamformers and a target rate vector, SGD obtains the optimal transmission rates and the corresponding decoding orders.

As it is detailed in [18], a low-rate feedback is sent to the BSs to inform them the rate. BSs need to know the $\min_k\{x_k\}$. There are two possible approaches (we used the first one): 1) One approach is that each user k sends the scalar value of x_k to all the BSs; 2) Another approach is that all users send the scalar values of x_k's to only one specific user, and that user calculates $\min_k\{x_k\}$ and sends this scalar value to all the BSs. This is the only "coordination" that is needed.

2) Performing the SGD: For any subset $\mathcal{D} \subseteq \mathcal{M}^*$, let $\mathbf{x}_{\mathcal{D}}$ denote the set of the signals with indices $(j, s, \ell) \in \mathcal{D}$ and $\mathbf{r}_{\mathcal{D}}$ denote the vector of the corresponding rates. Also, for any two disjoint subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{M}^*$, let $\mathcal{C}_k(\mathcal{H}_k, \mathcal{U}, \mathcal{V})$ denote the instantaneous achievable rate region of the k^{th} user for the sub-messages in $\mathcal{U}$ decoded using the maximum-likelihood (ML) decoder, after assuming the messages in $\mathcal{V}$ to be additive Gaussian interferers. In particular, we have that

$$\begin{aligned} &\mathcal{C}_k(\mathcal{H}_k, \mathcal{U}, \mathcal{V}) \\ &\quad = \Bigg\{ \mathbf{r}_{\mathcal{U}} \in \mathbb{R}_+^{|\mathcal{U}|} : \forall \mathcal{D} \subseteq \mathcal{U}, \mathcal{D} \neq \emptyset, \sum_{(j,s,\ell) \in \mathcal{D}} r_{j,s,\ell} < \\ &\qquad \log\left(1 + \frac{\frac{1}{L}\sum_{(j,s,\ell) \in \mathcal{D}} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell) \in \mathcal{V}} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2 + \sigma_k^2}\right) \Bigg\}. \end{aligned} \tag{4}$$

In SGD we start by decoding $Q_1(k)$ by treating all other messages as noise, and then subtracting $Q_1(k)$ and decoding $Q_2(k)$ and so on. SGD(c) shows this procedure.

The achievable rate region of SGD is $\cap_{k=1}^{K} \cap_{i=1}^{p_k} \mathcal{C}_k\left(\mathcal{H}_k, Q_i(k), \overline{Q_i(k)}\right)$. That is, the following constraints should be satisfied:

$$\begin{aligned} &\log\left(1 + \frac{\frac{1}{L}\sum_{(j,s,\ell) \in J_i} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell) \in \overline{Q_i(k)}} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2 + \sigma_k^2}\right) \\ &\quad \geq \sum_{(j,s,\ell) \in J_i} r_{j,s,\ell}, \quad \forall k \in \{1, 2, ..., K\}, \\ &\quad \forall i \in \{1, 2, ..., p_k\}, \quad \forall J_i \subseteq Q_i(k), \; J_i \neq \emptyset. \end{aligned} \tag{5}$$

III. Beamforming Design With SGD for Given Rates

A. Problem Formulation

The objective of beamforming design is to minimize the total transmit power while maintaining the target transmission rates for all signals transmitted by the BSs. For a given set of SGD schedules $\underline{Q}(k) = \{Q_1(k), \ldots, Q_{p_k}(k)\}, k \in \{1, \ldots, K\}$, and a given set of rates $\mathbf{r} \in \mathbb{R}_+^{|\mathcal{M}^*|}$ of the messages that we seek to be satisfied, by using the rate region of SGD, the problem can be formulated as

$$\begin{aligned} P(\mathbf{r}) : \quad &\min_{\{\mathbf{w}_{j,s}\}_{j=1,s=1}^{M,\ S_j}} \sum_{j=1}^{M} \sum_{s=1}^{S_j} ||\mathbf{w}_{j,s}||^2 \\ &\text{subject to (5).} \end{aligned} \tag{6}$$

Define

$$\begin{aligned} r_{j,s,\ell} &= \log(1 + \gamma_{j,s,\ell}), \\ &\quad \forall j \in \{1, \ldots, M\}, s \in \{1, \ldots, S_j\}, \ell \in \{1, \ldots, L\}. \end{aligned} \tag{7}$$

Then, the problem in (6) can be rewritten in terms of the target SINR values $\boldsymbol{\gamma} = \{\gamma_{j,s,\ell}\}$ as the following

$$\begin{aligned} P(\boldsymbol{\gamma}): \quad & \min_{\{\mathbf{w}_{j,s}\}_{j=1,s=1}^{M,\ S_j}} \sum_{j=1}^{M}\sum_{s=1}^{S_j} ||\mathbf{w}_{j,s}||^2 \\ & \text{s.t. } \frac{\frac{1}{L}\sum_{(j,s,\ell)\in J_i} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in \overline{Q_i(k)}} |\mathbf{h}_{j,k}^H \mathbf{w}_{j,s}|^2 + \sigma_k^2} \geq f(\boldsymbol{\gamma}_{J_i}), \\ & \forall k \in \{1,2,...,K\}, \quad \forall i \in \{1,2,\ldots,p_k\}, \\ & \forall J_i \subseteq Q_i(k), \quad J_i \neq \emptyset, \end{aligned} \tag{8}$$

where $f(\boldsymbol{\gamma}_{J_i}) \triangleq \Pi_{(j,s,\ell)\in J_i}(\gamma_{j,s,\ell}+1) - 1$.

Remark 1: The problem $P(\boldsymbol{\gamma})$ in (8) is non-convex and NP-hard[2]. In general, determining its feasibility is as difficult as solving the problem itself. However, as long as the target rate vector $\{r_{j,s,l}\}$ is obtained through SGD(a)-(b) and hence achievable by using SGD, the problem is always feasible.

Remark 2: Minimizing the total power is a common objective in multicell systems, see, e.g., [25], [26]. Moreover, we can also use the weighted sum power as the objective similar to [10], [21] and the algorithm can be modified accordingly to take into account the weights. Note that the peak-antenna or peak-BS power constraint can be easily incorporated in our sum-power minimization problem and proposed algorithms since these constraints are all linear.

In what follows, we propose a suboptimal method for solving this problem.

B. Semi-Definite Programming Relaxation

Here, we employ an SDR approach. By introducing new variables $\mathbf{W}_{j,s} = \mathbf{w}_{j,s}\mathbf{w}_{j,s}^H$, the problem $P(\boldsymbol{\gamma})$ in (8) can be written in terms of $\{\mathbf{W}_{j,s}\}$ with the added constraints $\text{rank}\{\mathbf{W}_{j,s}\} = 1, \mathbf{W}_{j,s} \succeq 0$, and $\mathbf{W}_{j,s}$ is a Hermitian matrix, $\forall j \in \{1,2,\ldots,M\}, s \in \{1,2,\ldots,S_j\}$.

For the new formulation, both the objective function and all constraints except for the rank-one constraints, are convex. Therefore we can consider the following relaxed version of $P(\boldsymbol{\gamma})$:

$$\begin{aligned} P_1(\boldsymbol{\gamma}): \quad & \min_{\{\mathbf{W}_{j,s}\}_{j=1,s=1}^{M,\ S_j}} \sum_{j=1}^{M}\sum_{s=1}^{S_j} \text{tr}\{\mathbf{W}_{j,s}\} \\ & \text{s.t. } \sum_{(j,s,\ell)\in J_i} \text{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\} - f(\boldsymbol{\gamma}_{J_i})L\sigma_k^2 \\ & \quad - f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \text{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\} \geq 0, \\ & \forall k \in \{1,2,\ldots,K\}, \quad \forall i \in \{1,2,\ldots,p_k\}, \\ & \forall J_i \subseteq Q_i(k), \quad J_i \neq \emptyset, \\ & \mathbf{W}_{j,s} \succeq 0, \\ & \forall j \in \{1,2,...,M\}, \quad \forall s \in \{1,2,...,S_j\}, \end{aligned} \tag{9}$$

where $\tilde{\mathbf{H}}_{j,k} = \mathbf{h}_{j,k}\mathbf{h}_{j,k}^H$. Problem $P_1(\boldsymbol{\gamma})$ is known as the SDR of problem $P(\boldsymbol{\gamma})$, where the name stems from the fact that $P_1(\boldsymbol{\gamma})$ is an instance of SDP, which can be solved, to any arbitrary accuracy, in a numerically stable and efficient fashion.

Due to the relaxation, the matrices $\{\mathbf{W}_{j,s}\}$ obtained by solving $P_1(\boldsymbol{\gamma})$ will not be rank-one necessarily. If they are all rank-one, then we can write $\mathbf{W}_{j,s} = \mathbf{w}_{j,s}\mathbf{w}_{j,s}^H$, and $\mathbf{w}_{j,s}$ will be a feasible – and in fact optimal – solution to problem $P(\boldsymbol{\gamma})$. If not, then $\text{tr}(\mathbf{W}_{j,s})$ is a lower bound on the optimal solution to $P(\boldsymbol{\gamma})$. If the rank of any $\mathbf{W}_{j,s}$ is larger than 1, then we must extract from it, in an efficient manner, a vector $\mathbf{w}_{j,s}$ that is feasible for $P(\boldsymbol{\gamma})$.

In the following, after a review of the previous works, in Section III-D we will propose an efficient method for obtaining near-optimal rank-one solutions.

C. Relationship With Previous Works

1) Formulations: The existing works on multicast beamforming design do not consider advanced interference management techniques, and only treat interference as noise. Therefore they can be considered as degenerative cases of our general formulation with SGD. One case is where the number of users is equal to the total number of transmitted messages (at all BSs), i.e., $K = \sum_{j=1}^{M} S_j$, and each user desires a distinct messages. Then, the degenerative case of our model[3] reduces to the unicast transmission model, which has been considered in [20] and [21] for single-cell and multi-cell networks, respectively. In particular, in [20] the beamforming design problem is solved using the standard conic optimization package and in [21] since the strong duality holds, the problem is solved by a convex solver. In addition, in [22] the case of only one transmitter with multiple messages (one for each individual user) is investigated where the SDR for beamforming design is employed. Due to the simplicity of the problem, the solution is always rank-one and therefore optimal. Another case is where there is only one message to be transmitted from each BS, i.e., $S_j = 1$, $\forall j \in \{1,2,\ldots,M\}$, each user intends to decode only one message and multiple users can intend to decode the same message. Such degenerative case becomes the multi-cell multicast model considered in [3], where a decentralized algorithm based on SDR is proposed.

In [7] a rate splitting method is applied for multicast beamforming design, where the message of each user is divided into only two parts of public and private. More specifically, part of the interference is broadcasted to all groups such that it is decoded and canceled. In this paper, each BS divides its message into multiple parts, and each unintended user decides if it wants to decode any of these parts.

2) Rank-One Solution Extraction: Existing approaches to extracting rank-one solutions from the solution to the SDR problem is based on "randomization" [27]–[29]. In the first method (rand-A), it was proposed in [1] to perform eigen-decomposition on $\mathbf{W}_{i,s} = \mathbf{U}_{i,s}\Sigma_{i,s}\mathbf{U}_{i,s}$ and then form $\mathbf{w}_{i,s} = \mathbf{U}_{i,s}\Sigma_{i,s}^{\frac{1}{2}}\mathbf{e}_{i,s}$, where $[\mathbf{e}_{i,s}]_\ell = e^{j\theta_{i,s,\ell}}$, with $\theta_{i,s,\ell}$'s being independent and uniformly decentralized on $[0, 2\pi)$. In the second method (rand-B), inspired by [28], $\mathbf{w}_{i,s}$ is formed according

[2]Our problem is a QCQP problem and any 0–1 integer program (which is NP-hard) can be formulated as a QCQP problem. Also in [2] it is discussed that a degenerative case of our problem is NP-hard.

[3]For example, if the k^{th} user aims to decode the s^{th} message transmitted from the j^{th} BS, then we have $\underline{Q}(k) = \{(j,s)\}$.

to $[\mathbf{w}_{i,s}]_\ell = \sqrt{[\mathbf{W}_{i,s}]_{\ell\ell}}[\mathbf{e}_{i,s}]_\ell$. The third method (rand-C) [30] forms $\mathbf{w}_{i,s} = \mathbf{U}_{i,s}\Sigma_{i,s}^{\frac{1}{2}}\mathbf{v}_{i,s}$, where $\mathbf{v}_{i,s} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

These randomization methods are applicable to a special case of $P(\boldsymbol{\gamma})$ where on the left-hand-sides of the constraints in (8) (which are in the form of fractions), $\mathbf{w}_{i,s}$'s do not appear in the denominators. In other words, randomization methods are applicable to maximum-likelihood decoding. For this case, after randomization, a set of feasible beamforming vectors can be found by simply scaling to the minimum norm necessary so that all constraints are satisfied. For example, for the case where there is only one transmitter with only one message and there are multiple receivers [1], randomization gives a feasible solution. However, such randomization can lead to infeasible solution for the general formulation $P(\boldsymbol{\gamma})$ considered in this paper, because scaling can increase both the numerator and denominator on the left-hand-side of the constraints in (8) and does not necessarily increase the value of the fraction.

3) Algorithm Design: There are also recent methods other than SDR for solving multicast beamforming problems in the literature [5], [11], [31], [32]. In specific, in [5], [11], and [31], iterative algorithms are proposed by solving a sequence of convex approximations of the original non-convex QCQP problem, which is known as successive convex approximation (SCA) or convex-concave procedure (CCP). These convex approximation methods need a feasible point as initialization. The SLR-SDR we shall propose in the next subsection can provide effective initial feasible solution when employed together with RLT or SCA methods which need a feasible point as initialization. Also, [32] proposed an algorithm for non-convex QCQPs, where any QCQP with only one constraint can be optimally solved. Moreover, the consensus ADMM can be used to solve general QCQPs, in such a way that each update requires to solve a number of QCQPs with only one constraint. For the ADMM proposed in this paper, in contrast to [32], all local problems have closed-form solutions which makes the algorithm very fast.

D. Successive Linearly Regularized Semi-Definite Relaxation (SLR-SDR)

The basic idea of our proposed approach is to use an alternative objective function and the same set of constraints in the SDR, so that the chance of obtaining rank-one solutions is higher. In particular, we propose to solve a sequence of SDR of the following form by introducing the new set of linear regularization matrices $\{\mathbf{A}\}$:

$$\begin{aligned} P_2(\boldsymbol{\gamma}, \mathbf{A}): \quad & \min_{\{\mathbf{W}_{j,s}\}_{j=1,s=1}^{M,\ S_j}} \sum_{j=1}^{M}\sum_{s=1}^{S_j} \mathrm{tr}\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\} \\ & \text{s.t.} \sum_{(j,s,\ell)\in J_i} \mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\} - f(\boldsymbol{\gamma}_{J_i})L\sigma_k^2 \\ & \qquad -f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\} \geq 0, \\ & \forall k \in \{1,2,\ldots,K\}, \quad \forall i \in \{1,2,\ldots,p_k\}, \\ & \forall J_i \subseteq Q_i(k), \quad J_i \neq \emptyset, \\ & \mathbf{W}_{j,s} \succeq 0, \\ & \forall j \in \{1,2,...,M\}, \quad \forall s \in \{1,2,...,S_j\}. \end{aligned} \tag{10}$$

At the beginning, we have $\mathbf{A}_{j,s} = \mathbf{I}_{M_j}$ and $P_2(\boldsymbol{\gamma}, \mathbf{A})$ is the same as $P_1(\boldsymbol{\gamma})$. We then iteratively adjust $\{\mathbf{A}_{j,s}\}$ and solve the corresponding $P_2(\boldsymbol{\gamma}, \mathbf{A})$. Since the constraints are always the same, every solution of $P_2(\boldsymbol{\gamma}, \mathbf{A})$ is feasible for $P_1(\boldsymbol{\gamma})$. We propose a method to adjust the matrices $\{\mathbf{A}_{j,s}\}$ in order to make the solution of $P_2(\boldsymbol{\gamma}, \mathbf{A})$ rank-one, which would also provide a rank-one, feasible and near-optimal solution for $P_1(\boldsymbol{\gamma})$, if the deviation in the objective function from $P_1(\boldsymbol{\gamma})$ is small.

Consider an optimal solution to the relaxed problem $P_1(\boldsymbol{\gamma})$. Since the mapping from $\{\mathbf{W}_{j,s}\}$ to $\mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\}$ is not bijective, there can be many optimal matrix solutions with different rank values. In this case, a solution found by a generic SDP solver would normally have a high rank value, although there may exist a rank-one solution.

To this end, SLR-SDR aims to search for a linear regularization term of the form $\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathrm{tr}\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\}$ that produces rank-one solutions. Observe that choosing $\mathbf{A}_{j,s} = \mathbf{I}_{M_j}$ leads to the simple trace regularizer and alternatively choosing $\mathbf{A}_{j,s} = \mathbf{I}_{M_j} - \mathbf{1}_{M_j}$ leads to the maximization of the sum of off-diagonal values of $\mathbf{W}_{j,s}$, where $\mathbf{1}_m$ is an all-1 matrix of size m. In SLR-SDR we incorporated two distinct heuristics in choosing $\{\mathbf{A}_{j,s}\}$ as we will elaborate in the following.

1) Penalizing the Trace Function: We propose a regularization term that promotes low-rank solutions by penalizing the trace function which is in fact the convex envelope of the nonsmooth rank function [33], [34].

Successive Linearly Regularized Semi-Definite Relaxation (SLR-SDR) for Solving $P_1(\boldsymbol{\gamma})$

1: **Choose** the thresholds $T_a, T_b > 0$ and constants $\alpha > 1$ and $\beta > 0$.
2: **Initialize** $a_{j,s} = 1$ and $b_{j,s} = 0$ for all $j \in \{1,\ldots,M\}$ and $s \in \{1,\ldots,S_j\}$.
3: **Set** $\mathbf{A}_{j,s} \leftarrow a_{j,s}\mathbf{I}_{M_j} - b_{j,s}(\mathbf{1}_{M_j} - \mathbf{I}_{M_j})$.
4: **Solve** the relaxed problem $P_2(\boldsymbol{\gamma}, \mathbf{A})$ in a centralized way using a general SDP solver, or in a decentralized way using ADMM-BF in Section IV.
5: If all matrices $\{\mathbf{W}_{j,s}\}$ are rank-one, then exit.
6: For every (j, s) such that $\mathrm{rank}(\mathbf{W}_{j,s}) > 1$,
   If $a_{j,s} < T_a$, set $a_{j,s} \leftarrow \alpha a_{j,s}$; go to step 3,
   Else if $b_{j,s} < T_b$, set $b_{j,s} \leftarrow \beta + b_{j,s}$; go to step 3,
   Else terminate and declare failure.

Example: Consider a simple case where the objective function in the relaxed problem is $\min\{a.\mathrm{tr}\{\mathbf{W}_1\}+b.\mathrm{tr}\{\mathbf{W}_2\}\}$ where a and b are some positive values. If the solution to this relaxed problem is of rank-one for $\mathbf{W}_1$ and of higher rank for $\mathbf{W}_2$, by increasing the value of b, we put a stronger emphasize on minimizing $\mathrm{tr}\{\mathbf{W}_2\}$ (in comparison with $\mathrm{tr}\{\mathbf{W}_1\}$) which increases the chance of getting a rank-one solution for $\mathrm{tr}\{\mathbf{W}_2\}$ since the trace function is the convex envelope of the rank function.[4]

In SLR-SDR, the term $a_{j,s}$ is to increase the diagonal entries of $\mathbf{A}_{j,s}$ which magnifies the minimization of $\mathrm{tr}\{\mathbf{W}_{j,s}\}$. This is due to the fact that the trace of the product of two

[4]We should be careful to not to increase b too much which may increase the rank of $\mathbf{W}_1$ as a consequence.

matrices can be interpreted as the inner product of their entries, and changing an element in $\mathbf{A}_{j,s}$ changes the corresponding component in $\mathbf{W}_{j,s}$ in the opposite direction.

2) Penalizing the Off-Diagonal Elements: An alternative technique for obtaining low rank solutions is to uplift the off-diagonal values of matrix variables in addition to minimizing the diagonal values. To be more specific, we apply a regularization term that is a combination of both diagonal and off-diagonal entries in this work. We start the off-diagonal method with an illustrative example.

Example: Consider a 2×2 positive semi-definite matrix $\mathbf{X}$ with fixed diagonal entries a and b and variable off-diagonal entries $X_{12} = X_{21} = x$. If we maximize x subject to $\mathbf{X} \succeq \mathbf{0}$, then it would admit the value $x = \sqrt{ab}$, which makes the matrix $\mathbf{X}$ rank-one. Motivated by this, we apply the idea of lifting the off-diagonal entries of positive semi-definite matrix variables to obtain a low-rank solution.

It is known that every low-rank matrix solution to an SDP problem belongs to the boundary of the feasible set as opposed to the interior, c.f. [35, Sec. 2.9]. Consider a matrix solution $\mathbf{W}^*_{j,s} \succeq 0$, with at least one zero eigenvalue. Let $\mathbf{v}$ belong to the null space of $\mathbf{W}^*_{j,s}$. Then for any $\epsilon > 0$, the matrix $\mathbf{W}^*_{j,s} - \epsilon \times \mathbf{v}\mathbf{v}^T \not\succeq 0$ which means that $\mathbf{W}^*_{j,s}$ belongs to the boundary of feasible set. Having these explanations, suppose that the entry $[\mathbf{W}_{j,s}]_{i,k}$ of a matrix solution does not appear in the objective function of the relaxed problem $P_1(\boldsymbol{\gamma})$ where $i \neq k$. In this case, $[\mathbf{W}_{j,s}]_{i,k}$ is a free parameter and the relaxed problem $P_1(\boldsymbol{\gamma})$ may have infinitely many solutions with different values of $[\mathbf{W}_{j,s}]_{i,k}$. Note that only a subset of these solutions belong to the boundary and the rest lie within the interior of the feasible set. In short, the motivation behind our proposed off-diagonal regularization term is to incorporate all of the entries (not only the diagonal entries) in the objective and to direct the solution to the boundary of the feasible set.

In SLR-SDR, the term $b_{j,s}$ is to decrease the off-diagonal entries of $\mathbf{A}_{j,s}$ which causes the off-diagonal entries of $\mathbf{W}_{j,s}$ to increase.

Remark 3: It is not difficult to find a set of appropriate parameters. We can start with some initial values for α, β, T_a, and T_b. Then, if the step sizes α and β are chosen too large, the ranks may even increase; and if they are too small, there may not be any change in ranks. (We only need to try a few values for α and β to find some proper values.) The parameters T_a and T_b determine the total number of iterations and can be determined when more iterations do not decrease the rank any more.

IV. Decentralized Algorithm

In this section, we develop a decentralized algorithm to solve $P_2(\boldsymbol{\gamma}, \mathbf{A})$ such that each BS j calculates its beamformers $\{\mathbf{W}_{j,s}, s = 1, \ldots, S_j\}$ locally, with the coordination of a central controller. To that end, we make use of the alternating direction method of multipliers (ADMM), a first-order convex optimization algorithm [36], [37] to solve large-scale problems in a decentralized or parallel way [38], [39].

A. Background on ADMM

Consider the following optimization problem with a separable objective function and linear constraints:

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} f(x) + g(z) \quad \text{s.t. } Ax + Bz = c, \tag{11}$$

where $f(x)$ and $g(z)$ are convex functions, A, B are known matrices, and c is a given vector of appropriate dimension.

The first method for solving (11) is "dual decomposition," which uses the Lagrangian function

$$\begin{aligned} \mathcal{L}(x, z, y) &= f(x) + g(z) + y^T(Ax + Bz - c) \\ &= \underbrace{f(x) + y^T Ax}_{F(x,y)} + \underbrace{g(z) + y^T Bz}_{G(z,y)} - y^T c, \end{aligned} \tag{12}$$

where y is the Lagrange multiplier corresponding to the constraint (11). The above Lagrangian function can be separated into two distinct functions $F(x, y)$ and $G(z, y)$. The dual decomposition method updates x, z and y separately, according to the following iterations

$$x^{t+1} \leftarrow \arg\min_x F(x, y^t), \tag{13}$$

$$z^{t+1} \leftarrow \arg\min_z G(z, y^t), \tag{14}$$

$$y^{t+1} \leftarrow y^t + \rho^t(Ax^{t+1} + Bz^{t+1} - c), \tag{15}$$

for $t = 0, 1, 2, \ldots,$ with an arbitrary initialization (x^0, z^0, y^0), where ρ^t is a step size.

Despite its decomposability, the dual decomposition method has robustness and convergence issues. The "method of multipliers" can be used to remedy these difficulties, which is based on the augmented Lagrangian function

$$\begin{aligned} \mathcal{L}_\rho(x, z, y) &= f(x) + g(z) \\ &\quad + y^T(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|^2, \end{aligned} \tag{16}$$

where ρ is a nonnegative constant. Notice that (16) is obtained by augmenting the Lagrangian function in (12) with a quadratic penalty term in order to increase the smallest eigenvalue of the Hessian of the Lagrangian with respect to (x, z). However, this augmentation creates a coupling between x and z. The method of multipliers performs the following iterations

$$(x^{t+1}, z^{t+1}) \rightarrow \arg\min_{(x,z)} \mathcal{L}_\rho(x, z, y^t), \tag{17}$$

$$y^{t+1} \rightarrow y^t + \rho(Ax^{t+1} + Bz^{t+1} - c). \tag{18}$$

In order to avoid solving a joint optimization with respect to x and z at every iteration, the ADMM first updates x by freezing z at its latest value, and then updates z based on the most recent value of x. This leads to the following 2-block ADMM [39]:

$$\text{block 1 update: } x^{t+1} \leftarrow \arg\min_x \mathcal{L}_\rho(x, z^t, y^t), \tag{19}$$

$$\text{block 2 update: } z^{t+1} \leftarrow \arg\min_z \mathcal{L}_\rho(x^t, z, y^t), \tag{20}$$

$$\text{dual variable update: } y^{t+1} \leftarrow y^t + \rho(Ax^{t+1} + Bz^{t+1} - c). \tag{21}$$

B. Decentralized Beamforming Based on ADMM

We now consider the ADMM approach to solving the relaxed beamforming problem $P_2(\boldsymbol{\gamma}, \mathbf{A})$ in (10). Since the standard ADMM does not allow inequalities, we add slack variables $\Gamma = \{\Gamma_{j,s,k} \in \mathbb{R}\}$ and $\boldsymbol{\delta} = \{\delta_{k,J_i} \in \mathbb{R}\}$ to only have equality constraints. We also introduce another set of slack variables $\widetilde{\mathbf{W}} = \{\widetilde{\mathbf{W}}_{j,s} \in \mathbb{C}^{M_j \times M_j}$ and is Hermitian$\}$, to make the local computation at each BS in closed-form, rather than solving an SDP (See Remark 4). Moreover, define $I^+(\mathbf{A}) = 0$ if $\mathbf{A} \succeq 0$ and $+\infty$ otherwise. Then, $P_2(\boldsymbol{\gamma}, \mathbf{A})$ can be rewritten equivalently in the form of (11) as

$$
\begin{aligned}
&P_3(\mathbf{W}, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}):\\
&\min_{\mathbf{W}, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}} \sum_{j=1}^{M}\sum_{s=1}^{S_j} \mathrm{tr}\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\}\\
&\qquad + \sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} I^+(\delta_{k,J_i}) + \sum_{j=1}^{M}\sum_{s=1}^{S_j} I^+(\widetilde{\mathbf{W}}_{j,s})\\
&\text{s.t. } \mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\} = \Gamma_{j,s,k},\\
&\quad \forall j \in \{1,\ldots,M\}, \quad \forall s \in \{1,\ldots,S_j\},\\
&\quad \forall k \in \{1,2,...,K\}, \qquad (22)\\
&\quad \delta_{k,J_i} = \sum_{(j,s,\ell)\in J_i} \Gamma_{j,s,k} - f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \Gamma_{j,s,k}\\
&\qquad - f(\boldsymbol{\gamma}_{J_i}) L\sigma_k^2,\\
&\quad \forall k \in \{1,\ldots,K\}, \quad \forall i \in \{1,\ldots,p_k\},\\
&\quad \forall J_i \subseteq Q_i(k), J_i \neq \emptyset, \qquad (23)\\
&\quad \widetilde{\mathbf{W}}_{j,s} = \mathbf{W}_{j,s},\\
&\quad \forall j \in \{1,\ldots,M\}, \quad \forall s \in \{1,\ldots,S_j\}. \qquad (24)
\end{aligned}
$$

Now, by introducing the Lagrange multipliers $\boldsymbol{\lambda} = \{\lambda_{j,s,k} \in \mathbb{R}\}$, $\boldsymbol{\nu} = \{\nu_{k,J_i} \in \mathbb{R}\}$, and $\boldsymbol{\kappa} = \{\boldsymbol{\kappa}_{j,s} \in \mathbb{C}^{M_j \times M_j}$ and is Hermitian$\}$ of the constraints (22), (23) and (24), respectively, we can write the augmented Lagrangian of $P_3(\mathbf{W}, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta})$ as in (16), given by

$$
\begin{aligned}
&\mathcal{L}_\rho(\mathbf{W}, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{\nu})\\
&= \sum_{j=1}^{M}\sum_{s=1}^{S_j} \mathrm{tr}\left\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\right\}\\
&\quad + \sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} I^+(\delta_{k,J_i}) + \sum_{j=1}^{M}\sum_{s=1}^{S_j} I^+(\widetilde{\mathbf{W}}_{j,s})\\
&\quad + \sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K} \lambda_{j,s,k}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\} - \Gamma_{j,s,k}\right)\\
&\quad + \sum_{j=1}^{M}\sum_{s=1}^{S_j} \mathrm{tr}\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}(\widetilde{\mathbf{W}}_{j,s} - \mathbf{W}_{j,s})\}\\
&\quad + \sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} \nu_{k,J_i}\Bigg(\delta_{k,J_i} + f(\boldsymbol{\gamma}_{J_i}) L\sigma_k^2\\
&\quad + f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \Gamma_{j,s,k} - \sum_{(j,s,\ell)\in J_i} \Gamma_{j,s,k}\Bigg)\\
&\quad + \frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\} - \Gamma_{j,s,k}\right)^2\\
&\quad + \frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j} \|\widetilde{\mathbf{W}}_{j,s} - \mathbf{W}_{j,s}\|^2\\
&\quad + \frac{\rho}{2}\sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} \Bigg(\delta_{k,J_i} + f(\boldsymbol{\gamma}_{J_i}) L\sigma_k^2\\
&\quad + f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \Gamma_{j,s,k} - \sum_{(j,s,\ell)\in J_i} \Gamma_{j,s,k}\Bigg)^2. \qquad (25)
\end{aligned}
$$

Then following the two-block ADMM as in (19)-(21), we alternatively minimize $\mathcal{L}_\rho$ in terms of two sets of variables $\{\widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}\}$ and $\{\mathbf{W}\}$, as well as the dual variables $\{\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\kappa}\}$. In particular we have

$$
\begin{aligned}
&\text{block 1 update:} (\widetilde{\mathbf{W}}^{t+1}, \Gamma^{t+1}, \boldsymbol{\delta}^{t+1})\\
&\quad \leftarrow \mathrm{argmin}_{\widetilde{\mathbf{W}}, \boldsymbol{\delta}, \Gamma}\ \mathcal{L}_\rho(\mathbf{W}^t, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}, \boldsymbol{\lambda}^t, \boldsymbol{\nu}^t, \boldsymbol{\kappa}^t), \qquad (26)\\
&\text{block 2 update:} \mathbf{W}^{t+1}\\
&\quad \leftarrow \mathrm{argmin}_{\mathbf{W}}\ \mathcal{L}_\rho(\mathbf{W}, \widetilde{\mathbf{W}}^{t+1}, \Gamma^{t+1}, \boldsymbol{\delta}^{t+1}, \boldsymbol{\lambda}^t, \boldsymbol{\nu}^t, \boldsymbol{\kappa}^t). \qquad (27)
\end{aligned}
$$

In what follows, we solve the two optimization problems (26) and (27).

1) Block 1 Update: Define $\mathcal{L}_\rho^1$ as the terms of $\mathcal{L}_\rho$ in (25) that includes $\{\widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}\}$, i.e.,

$$
\begin{aligned}
&\mathcal{L}_\rho^1(\mathbf{W}, \widetilde{\mathbf{W}}, \Gamma, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{\nu})\\
&= \sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} I^+(\delta_{k,J_i})\\
&\quad + \sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K} \lambda_{j,s,k}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\} - \Gamma_{j,s,k}\right)\\
&\quad + \sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i \subseteq Q_i(k)\\ J_i \neq \emptyset}} \nu_{k,J_i}\Bigg(\delta_{k,J_i} + f(\boldsymbol{\gamma}_{J_i}) L\sigma_k^2\\
&\quad + f(\boldsymbol{\gamma}_{J_i}) \sum_{(j,s,\ell)\in\overline{Q_i(k)}} \Gamma_{j,s,k} - \sum_{(j,s,\ell)\in J_i} \Gamma_{j,s,k}\Bigg)\\
&\quad + \frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\} - \Gamma_{j,s,k}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&+\underbrace{\frac{\rho}{2}\sum_{k=1}^{K}\sum_{i=1}^{p_k}\sum_{\substack{J_i\subseteq Q_i(k)\\ [4pt] J_i\neq\emptyset}}\left(\delta_{k,J_i}+f(\boldsymbol{\gamma}_{J_i})L\sigma_k^2 +f(\boldsymbol{\gamma}_{J_i})\sum_{(j,s,\ell)\in\overline{Q_i(k)}}\Gamma_{j,s,k}-\sum_{(j,s,\ell)\in J_i}\Gamma_{j,s,k}\right)^2}_{\triangleq\mathcal{L}_\rho^{1,1}(\mathbf{W},\Gamma,\boldsymbol{\delta},\boldsymbol{\lambda},\boldsymbol{\nu})}\\
&+\underbrace{\sum_{j=1}^{M}\sum_{s=1}^{S_j}I^{+}(\widetilde{\mathbf{W}}_{j,s})+\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathrm{tr}\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s})\} +\frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j}\|\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}\|^2.}_{\triangleq\mathcal{L}_\rho^{1,2}(\mathbf{W},\widetilde{\mathbf{W}},\boldsymbol{\kappa})} \qquad (28)
\end{aligned}
$$

Note that $\mathcal{L}_\rho^{1,1}$ includes only the terms with Γ and $\boldsymbol{\delta}$, and $\mathcal{L}_\rho^{1,2}$ includes only the terms with $\widetilde{\mathbf{W}}$. Therefore, (26) involves two decoupled optimizations and $\{\Gamma,\boldsymbol{\delta}\}$ and each $\{\widetilde{\mathbf{W}}\}$ can be solved separately which we discuss in the following.

Solving $\min_{\Gamma,\boldsymbol{\delta}}\mathcal{L}_\rho^{1,1}(\mathbf{W},\Gamma,\boldsymbol{\delta},\boldsymbol{\lambda},\boldsymbol{\nu})$*:* Note that $\mathcal{L}_\rho^{1,1}(\mathbf{W},\Gamma,\boldsymbol{\delta},\boldsymbol{\lambda},\boldsymbol{\kappa},\boldsymbol{\nu})$ contains only linear and positive quadratic terms of Γ and $\boldsymbol{\delta}$ and therefore its minimization over Γ and $\boldsymbol{\delta}\geq\mathbf{0}$ can be easily solved, e.g., using a standard convex solver.

Solving $\min_{\widetilde{\mathbf{W}}}\mathcal{L}_\rho^{1,2}(\mathbf{W},\widetilde{\mathbf{W}},\boldsymbol{\kappa})$*:* Note that the optimization of $\{\widetilde{\mathbf{W}}_{j,s}\}$ can be carried out independently and in parallel, since

$$
\mathcal{L}_\rho^{1,2}(\mathbf{W},\widetilde{\mathbf{W}},\boldsymbol{\kappa})=\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathcal{L}_{\rho,\,j,s}^{1,2}(\mathbf{W}_{j,s},\widetilde{\mathbf{W}}_{j,s},\boldsymbol{\kappa}), \qquad (29)
$$

where

$$
\begin{aligned}
&\mathcal{L}_{\rho,\,j,s}^{1,2}(\mathbf{W}_{j,s},\widetilde{\mathbf{W}}_{j,s},\boldsymbol{\kappa})\\
&\triangleq I^{+}(\widetilde{\mathbf{W}}_{j,s})+\frac{\rho}{2}\|\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}\|^2+\mathrm{tr}\left\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s})\right\}\\
&=I^{+}(\widetilde{\mathbf{W}}_{j,s})+\frac{\rho}{2}\mathrm{tr}\left\{\left(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}+\frac{1}{\rho}\boldsymbol{\kappa}_{j,s}\right)^{\mathrm{H}}\left(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}+\frac{1}{\rho}\boldsymbol{\kappa}_{j,s}\right)\right\}-\frac{1}{2\rho}\mathrm{tr}\left\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}\boldsymbol{\kappa}_{j,s}\right\}. \qquad (30)
\end{aligned}
$$

Then the optimal $\widetilde{\mathbf{W}}_{j,s}$ is obtained by performing an eigenvalue decomposition on $\mathbf{W}_{j,s}-\frac{1}{\rho}\boldsymbol{\kappa}_{j,s}$ and replacing all negative eigenvalues by zero, i.e.,

$$
\begin{aligned}
\widetilde{\mathbf{W}}_{j,s}^{*}&=\arg\min_{\widetilde{\mathbf{W}}_{j,s}}\mathcal{L}_{\rho,\,j,s}^{1,2}(\mathbf{W}_{j,s},\widetilde{\mathbf{W}}_{j,s},\boldsymbol{\kappa})\\
&=\left(\mathbf{W}_{j,s}-\frac{1}{\rho}\boldsymbol{\kappa}_{j,s}\right)_{+}. \qquad (31)
\end{aligned}
$$

2) Block 2 Update: Define $\mathcal{L}_\rho^2$ as the terms of $\mathcal{L}_\rho$ in (25) that includes $\mathbf{W}$, i.e.,

$$
\begin{aligned}
&\mathcal{L}_\rho^2(\mathbf{W},\widetilde{\mathbf{W}},\Gamma,\boldsymbol{\lambda},\boldsymbol{\kappa})\\
&=\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathrm{tr}\left\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\right\}\\
&\quad+\sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K}\lambda_{j,s,k}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\}-\Gamma_{j,s,k}\right)\\
&\quad+\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathrm{tr}\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s})\}\\
&\quad+\frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j}\sum_{k=1}^{K}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\}-\Gamma_{j,s,k}\right)^2\\
&\quad+\frac{\rho}{2}\sum_{j=1}^{M}\sum_{s=1}^{S_j}\|\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}\|^2\\
&=\sum_{j=1}^{M}\sum_{s=1}^{S_j}\mathcal{L}_{\rho,j,s}^{2}(\mathbf{W}_{j,s},\widetilde{\mathbf{W}}_{j,s},\Gamma,\boldsymbol{\lambda},\boldsymbol{\kappa}), \qquad (32)
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathcal{L}_{\rho,\,j,s}^{2}(\mathbf{W}_{j,s},\widetilde{\mathbf{W}}_{j,s},\Gamma,\boldsymbol{\lambda},\boldsymbol{\kappa})\\
&\triangleq\mathrm{tr}\left\{\mathbf{A}_{j,s}\mathbf{W}_{j,s}\right\}\\
&\quad+\sum_{k=1}^{K}\lambda_{j,s,k}\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\}-\Gamma_{j,s,k}\right)\\
&\quad+\mathrm{tr}\{\boldsymbol{\kappa}_{j,s}^{\mathrm{T}}(\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s})\}+\frac{\rho}{2}\|\widetilde{\mathbf{W}}_{j,s}-\mathbf{W}_{j,s}\|^2\\
&\quad+\frac{\rho}{2}\sum_{k=1}^{K}(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\right\}-\Gamma_{j,s,k})^2. \qquad (33)
\end{aligned}
$$

Note that block 2 update is an unconstrained optimization and (33) is quadratic in terms of $\mathbf{W}_{j,s}$. Taking the derivative of (33) with respect to $\mathbf{W}_{j,s}$ and setting it zero, we have (Recall that for complex valued matrices $\mathbf{A}$ and $\mathbf{X}$ of appropriate dimensions, $\frac{d\ \mathrm{tr}(\mathbf{AX})}{d\ \mathbf{X}}=\mathbf{A}^T$ [40])

$$
\begin{aligned}
&\mathbf{A}_{j,s}^{T}+\sum_{k=1}^{K}\lambda_{j,s,k}\tilde{\mathbf{H}}_{j,k}^{T}-\boldsymbol{\kappa}_{j,s}-\rho\sum_{k=1}^{K}\Gamma_{j,s,k}\tilde{\mathbf{H}}_{j,k}^{T}\\
&\quad+\rho\sum_{k=1}^{K}\mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\}\tilde{\mathbf{H}}_{j,k}^{T}+\rho\left(\mathbf{W}_{j,s}-\widetilde{\mathbf{W}}_{j,s}\right)=0. \qquad (34)
\end{aligned}
$$

Define

$$
\begin{aligned}
\mathbf{K}_{j,s}\triangleq-\left(\frac{1}{\rho}\right)&\\
\times\Bigg(&\mathbf{A}_{j,s}^{T}+\sum_{k=1}^{K}\lambda_{j,s,k}\tilde{\mathbf{H}}_{j,k}^{T}-\boldsymbol{\kappa}_{j,s}-\rho\sum_{k=1}^{K}\Gamma_{j,s,k}\tilde{\mathbf{H}}_{j,k}^{T}\\
&-\rho\widetilde{\mathbf{W}}_{j,s}\Bigg). \qquad (35)
\end{aligned}
$$

Then (34) simplifies to

$$\mathbf{W}_{j,s} + \sum_{k=1}^{K} \mathrm{tr}\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}\}\tilde{\mathbf{H}}_{j,k}^{T} = \mathbf{K}_{j,s}. \tag{36}$$

Define the row concatenation operator $\widetilde{\mathrm{vec}}(\mathbf{H}) = [\mathrm{vec}(\mathbf{H}^T)]^T$ and the inverse operator $\widetilde{\mathrm{vec}}^{-1}(.)$ such that $\widetilde{\mathrm{vec}}^{-1}(\widetilde{\mathrm{vec}}(\mathbf{H})) = \mathbf{H}$. Then (36) can be written as

$$\widetilde{\mathrm{vec}}\left(\mathbf{W}_{j,s}\right) + \sum_{k=1}^{K} \mathrm{diag}\left(\widetilde{\mathrm{vec}}\left(\left(\tilde{\mathbf{H}}_{j,k}^{T}\right)\right)\right)\mathbf{1}_{M_j^2} \times \widetilde{\mathrm{vec}}^{T}(\tilde{\mathbf{H}}_{j,k}^{T})\widetilde{\mathrm{vec}}(\mathbf{W}_{j,s}) = \widetilde{\mathrm{vec}}\left(\mathbf{K}_{j,s}\right). \tag{37}$$

$$\Longrightarrow \mathbf{W}_{j,s}^{*} = \widetilde{\mathrm{vec}}^{-1}\left(\mathbf{L}_{j,s}\widetilde{\mathrm{vec}}\left(\mathbf{K}_{j,s}\right)\right). \tag{38}$$

where

$$\mathbf{L}_{j,s} \triangleq \left(\mathbf{I}_{M_j^2} - \sum_{k=1}^{K} \mathrm{diag}\left(\widetilde{\mathrm{vec}}\left(\left(\tilde{\mathbf{H}}_{j,k}^{T}\right)\right)\right)\mathbf{1}_{M_j^2}\widetilde{\mathrm{vec}}^{T}\left(\tilde{\mathbf{H}}_{j,k}^{T}\right)\right)^{-1}. \tag{39}$$

Note that the term $\mathbf{L}_{j,s}$ is constant and therefore can be pre-computed. The decentralized beamforming algorithm is summarized in a table as ADMM-BF.

Theorem 1: ADMM-BF solves the problem $P_2(\boldsymbol{\gamma}, \mathbf{A})$ in (10) exactly.

Proof: The conditions for the 2-block ADMM to solve the optimization problem in (11) exactly are given in [39] and they hold true here, i.e., the cost function of $P_2(\boldsymbol{\gamma}, \mathbf{A})$ and those in block 1 and block 2 are all convex. □

Remark 4: Introducing the slack variables $\{\widetilde{\mathbf{W}}_{j,s}\}$ into the ADMM helped us to solve the local problems, i.e., (30) and (33), in a closed-form manner rather than needing to solve an SDP for each agent (j, s) in each iteration. Specifically, without introducing $\{\widetilde{\mathbf{W}}_{j,s}\}$, the block 2 update, i.e., (27) would have the constraints $\mathbf{W}_{j,s} \succeq 0$ for each (j, s) which does not admit a closed-form solution and hence an SDP needs to be solved for each agent (j, s) in each iteration.

Remark 5: We note the following implementation details regarding ADMM-BF:

- *Channel information:* The local agents in each BS need to know only the channels from that BS to all users.
- *Decoding orders:* The local agents do not need to know the decoding orders. Only the central agent knows it.
- *Information exchange:* The central agent shares each $\Gamma_{j,s,k}^{t+1}$ with local agent (j, s), and each local agent shares each $\mathbf{W}_{j,s}^{t+1}$ and $\lambda_{j,s,k}^{t+1}$ with the central agent.

Remark 6: Here we briefly compare our decentralized method with the centralized one. In the centralized method, the beamformers for all BSs must be calculated all together by a central agent and then the corresponding beamformers of each BS are given to it. In contrast, in the decentralized method each BS calculates its own beamformers with some coordination of the central agent. Another important feature of the decentralized algorithm is that all calculations in each individual BS are in closed-form (rather than solving an SDP as in the centralized method). It takes several iterations for the decentralized algorithm to converge to the centralized solution, and the convergence is guaranteed.

ADMM-BF for Solving $P_2(\boldsymbol{\gamma}, \mathbf{A})$

1: Choose the step size ρ, the stopping criterion ϵ and the maximum number of iterations $t_{\max}$.
2: Set the initial values $\mathbf{W}^0$, $\boldsymbol{\lambda}^0$, $\boldsymbol{\kappa}^0$, $\boldsymbol{\nu}^0$ and $t \leftarrow 0$.
3: Calculate $\{\mathbf{L}_{j,s}\}$ in (38).
4: **Repeat**
5: Block 1 update:
-The central agent solves the quadratic problem
$(\Gamma^{t+1}, \boldsymbol{\delta}^{t+1}) \leftarrow \arg\min_{\boldsymbol{\delta}, \Gamma} \mathcal{L}_{\rho}^{1,1}(\mathbf{W}^t, \Gamma, \boldsymbol{\delta}, \boldsymbol{\lambda}^t, \boldsymbol{\nu}^t)$
and shares each $\Gamma_{j,s,k}^{t+1}$ with local agent (j, s).
-Every local agent (j, s) computes
$\widetilde{\mathbf{W}}_{j,s}^{t+1} \leftarrow \left(\mathbf{W}_{j,s}^{t} - \frac{1}{\rho}\boldsymbol{\kappa}_{j,s}^{t}\right)_{+}$
6: Block 2 update:
-Every local agent (j, s) computes $\mathbf{K}_{j,s}$ in (35) and $\mathbf{W}_{j,s}^{t+1} \leftarrow \widetilde{\mathrm{vec}}^{-1}\left(\mathbf{L}_{j,s}\ \widetilde{\mathrm{vec}}\left(\mathbf{K}_{j,s}\right)\right)$
7: Dual variable update:
-Every local agent (j, s) computes
$\lambda_{j,s,k}^{t+1} \leftarrow \lambda_{j,s,k}^{t} + \rho\left(\mathrm{tr}\left\{\tilde{\mathbf{H}}_{j,k}\mathbf{W}_{j,s}^{t+1}\right\} - \Gamma_{j,s,k}^{t+1}\right)$.
$\boldsymbol{\kappa}_{j,s}^{t+1} \leftarrow \boldsymbol{\kappa}_{j,s}^{t} + \rho\left(\widetilde{\mathbf{W}}_{j,s}^{t+1} - \mathbf{W}_{j,s}^{t+1}\right)$ and
shares each $\mathbf{W}_{j,s}^{t+1}$ and $\lambda_{j,s,k}^{t+1}$ with central agent.
-Central agent computes
$\nu_{k,J_i}^{t+1} \leftarrow \nu_{k,J_i}^{t} + \rho\left(\delta_{k,J_i}^{t+1} + f(\boldsymbol{\gamma}_{J_i})L\sigma_k^2 + f(\boldsymbol{\gamma}_{J_i})\sum_{(j,s,\ell)\in\overline{Q_i(k)}}\Gamma_{j,s,k}^{t+1} - \sum_{(j,s,\ell)\in J_i}\Gamma_{j,s,k}^{t+1}\right)$.
8: $t \leftarrow t + 1$.
9: **Until** $\|\mathbf{W}^t - \mathbf{W}^{t-1}\| \leq \epsilon$ or $t > t_{\max}$.
10: **If** $\|\mathbf{W}^t - \mathbf{W}^{t-1}\| \leq \epsilon$ output $\mathbf{W}^t$.
Else declare failure.

V. Joint Beamforming and Rate Optimization With SGD

In the previous two sections, we discussed beamformer optimization with SGD given the achievable rate vector. Once the beamformers are optimized for the given rate vector, if we run SGD(a)-(b) again, we will obtain an improved rate vector, for which we can further optimize the beamformers. Such a process can iterate until it converges, as shown in Fig. 1. The first two initialization steps are done locally at each BS. The rate calculation step involves computing the local rate at each user using algorithm SGD(a) and then the user coordination using algorithm SGD(b) through the BSs. More details can be found in [18]. The beamformer optimization step through the SLR-SDR algorithm can be implemented in either centralized or decentralized way, as discussed in Sections III and IV, respectively.

Note that since the initial rate vector $\mathbf{r}^0$ is computed based on the initial beamformers $\{\mathbf{w}_{j,s}^0\}$, it is achievable. Then by running SGD(a)-(b) we obtain an improved rate vector $\mathbf{r}$ that is also achievable by $\{\mathbf{w}_{j,s}^0\}$. Next for the achievable rate vector $\mathbf{r}$, by running SLR-SDR we obtain an improved set of beamformers $\{\mathbf{w}_{j,s}\}$ that results in less total transmission

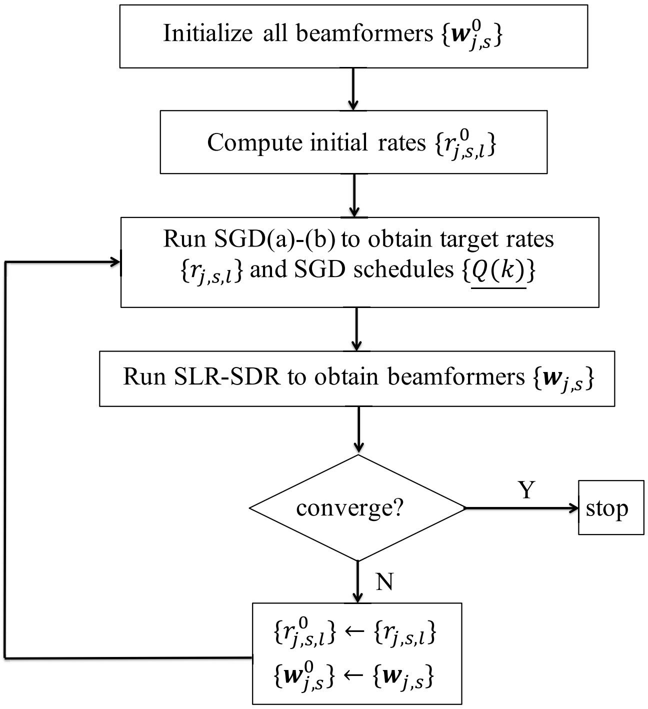


Fig. 1. Iterative procedure for joint beamformer and rate optimization for SGD.

power. Hence in each iteration, the rate increases and the total power decreases and the procedure will converge.

In this work, we initialize the beamformers as the average of matched filters. Specifically we set $\mathbf{w}^0_{j,s} = \sum_{k \in \mathcal{K}_{j,s}} \mathbf{h}_{j,k}/|\mathcal{K}_{j,s}|$ where $\mathcal{K}_{j,s}$ represents the set of receivers that intend to decode the signal (j,s). Then assuming all interference is treated as noise, the initial rate is given by

$$\begin{aligned} r^0_{j,s,\ell} &= \frac{1}{L} \\ &\times \min_{k \in \mathcal{K}_{j,s}} \left\{ \log\left(1 + \frac{|\mathbf{h}^H_{j,k}\mathbf{w}_{j,s}|^2}{\sum_{(j',s') \in \mathcal{M} \setminus (j,s)} |\mathbf{h}^H_{j',k}\mathbf{w}_{j',s'}|^2 + \sigma^2_k}\right)\right\}, \\ &\text{for all } j \in \{1,\ldots,M\}, s \in \{1,\ldots,S_j\}, \ell \in \{1,\ldots,L\}. \end{aligned} \tag{40}$$

VI. Simulation Results

In this section, we provide simulation results to illustrate the performance of the proposed centralized and decentralized beamforming design algorithms, and also illustrate the advantage of using SGD over traditional linear receivers. After performing an SDR followed by solving a convex optimization $P_1(\boldsymbol{\gamma})$ in (9) over $\{\mathbf{W}_{j,s}\}_{j=1,s=1}^{M,\ S_j}$, we compute the eigenvalues of each $\mathbf{W}_{j,s}$, $\lambda^{j,s}_1 \geq \lambda^{j,s}_2 \geq \cdots \geq \lambda^{j,s}_{M_j}$. Then, we check to see whether or not $\frac{\lambda^{j,s}_1}{\lambda^{j,s}_2} \geq 10^8$ holds for all (j,s). If it holds, we declare all $\{\mathbf{W}_{j,s}\}$ are rank-one and optimal; otherwise we implement SLR-SDR (the linear regularized SDP) to decrease the ranks of the non-rank-one matrices $\mathbf{W}_{j,s}$. Throughout this section, we use the following parameters in SLR-SDR: $\alpha = 10$, $\beta = 1$, $T_a = \alpha^7$ and $T_b = 7\beta$. The group size $\mu_k = \mu$ for all k in SGD. In all simulations, all channel vectors $\{\mathbf{h}_{j,k}\}$ are independent and each contains i.i.d. $\mathcal{CN}(0,1)$ entries. For all simulations, 200 channel realizations are simulated and 100 randomizations are generated if any of the randomization methods are needed.

A. Comparison With Randomization Methods

As noted in Section III-C.2, there are three randomization methods to obtain a rank-one solution from a non-rank-one solution to the SDR [1]. However, for the problem considered in this paper, i.e., $P_1(\boldsymbol{\gamma})$ in (9), the rank-one solutions given by these methods may not be feasible. In fact, since the number of constraints in (9) is usually large, the randomization methods, together with scaling, yield infeasible solutions most of the time, as will be shown in Fig. 4.

To show the advantage of our proposed reduced-rank method using linear regularized SDP, we consider two simple systems that do not employ SGD. The first system consists of $M = 2$ BSs, each with $M_j = 5$ antennas and transmitting $S_j = 1$ message, $j = 1, 2$, and $K = 3$ ML users each intending to decode both messages. We assume the same predetermined desired rate for both signals, i.e., $r_1 = r_2 = r$. Also, we assume that the noise variance at each user is $\sigma^2_k = 1$, $k = 1, 2, 3$.

For this example there are 3 constraints for each receiver $k \in \{1, 2, 3\}$:

$$\begin{aligned} \log\left(1 + |\mathbf{h}^H_{1,k}\mathbf{w}_1|^2\right) &\geq r_1, \\ \log\left(1 + |\mathbf{h}^H_{2,k}\mathbf{w}_2|^2\right) &\geq r_2, \\ \log\left(1 + |\mathbf{h}^H_{1,k}\mathbf{w}_1|^2 + |\mathbf{h}^H_{2,k}\mathbf{w}_2|^2\right) &\geq r_1 + r_2. \end{aligned} \tag{41}$$

Then, if the solution to the SDR, or that returned by SLR-SDR does not meet the rank-one condition, by applying one of the randomization methods followed by scaling, we always obtain a feasible solution. Fig. 2 depicts a comparison of total power consumption between SLR-SDR and simple SDP (i.e., solving P_1 directly without linear regularization) together with different randomization methods. It is seen that SLR-SDR results in ~5 dB reduction in total power consumption than the simple SDP. This is mostly due to the fact that in many of the channel realizations that the SDP is not exact, applying SLR-SDR results in a close to optimal solution for the total power (usually more optimal than the randomization plus scaling). Here we also compared the solutions with the lower bound, i.e., the solution to $P_1(\boldsymbol{\gamma})$, where the rank-one constraints are dropped.

The second system consists of $M = 2$ transmitters and $K = 6$ receivers. Each transmitter has $M_j = 10$ antennas and sends $S_j = 1$ message, $j = 1, 2$. We assume that the sets of messages to be decoded by the receivers are $\mathcal{A}(1) = \mathcal{A}(2) = \mathcal{A}(3) = \{1\}$, $\mathcal{A}(4) = \mathcal{A}(5) = \mathcal{A}(6) = \{2\}$ and also $L = 1$. Each receiver simply treats the unintended message as

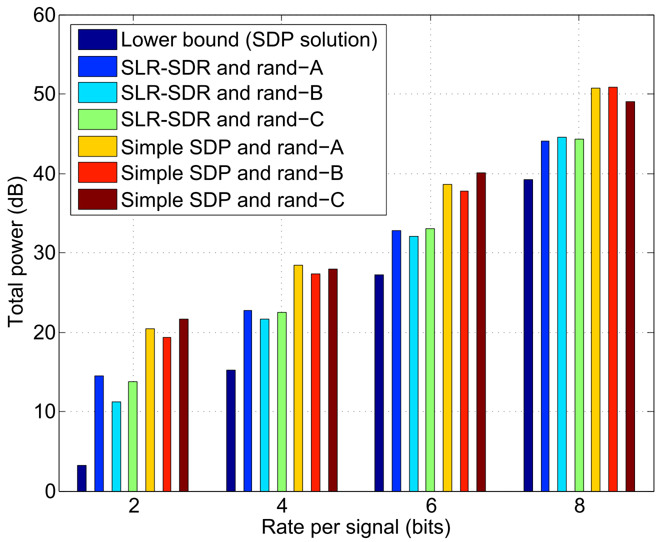


Fig. 2. Comparison of the power consumption between SLR-SDR and simple SDP together with different randomization methods.

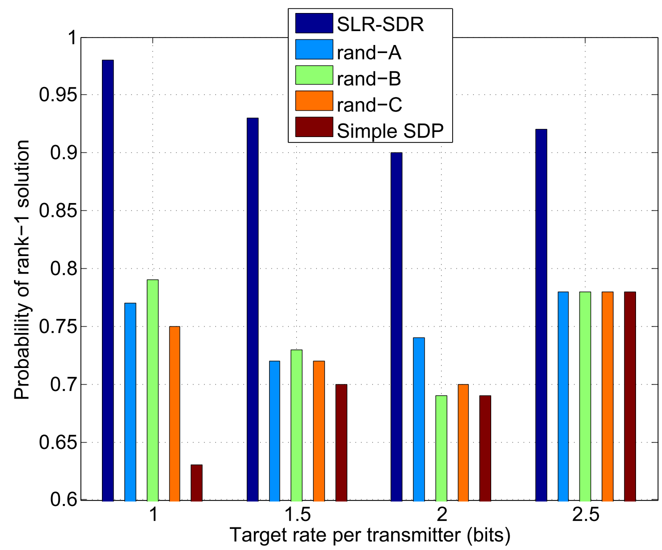


Fig. 3. Comparison of the probability of feasible rank-one solution of different randomization methods and SLR-SDR.

noise. Similar to the previous example, we consider the same predetermined desired rate for both signals, i.e., $r_1 = r_2 = r$. Also, we assume that the noise variance at each receiver is $\sigma_k^2 = 1$, $\forall k$. Therefore, there are six constraints, one for the decodability of each user, as below:

$$\log\left(1+\frac{|\mathbf{h}_{j,k}^H\mathbf{w}_j|^2}{|\mathbf{h}_{j',k}^H\mathbf{w}_{j'}|^2+1}\right) \geq r_j,$$
$$(j, j', k) \in \left\{(1,2,k')|k' \in \{1,2,3\}\right\} \cup \left\{(2,1,k'')|k'' \in \{4,5,6\}\right\}. \quad (42)$$

Fig. 3 depicts a comparison of probability of obtaining feasible rank-one solution by the randomization methods and SLR-SDR. It is seen that when the target rate r is low, randomization can increase the probability of feasible rank-one solutions compared with the simple SDP. But as the rate increases, the improvement becomes smaller[5]. However, SLR-SDR can significantly increase the probability of rank-one solutions compared with randomization methods for all rate ranges.

B. Performance of Centralized Algorithm for SGD (SLR-SDR)

Throughout the rest of this section, we assume the following parameters: $M = 2$, $K = 4$, $S_j = 2$, $M_j = 4$, $\forall j \in \{1, \ldots, M\}$. We also set $\mathcal{A}(1) = \{(1, 1), (2, 1)\}$, $\mathcal{A}(2) = \{(1, 2), (2, 2)\}$, $\mathcal{A}(3) = \{(1, 1), (2, 1)\}$, and $\mathcal{A}(4) = \{(1, 1), (2, 2)\}$. Also, we assume that the noise variance in each receiver is $\sigma_k^2 = \frac{1}{2}$, $\forall k$. First the beamforming and rate

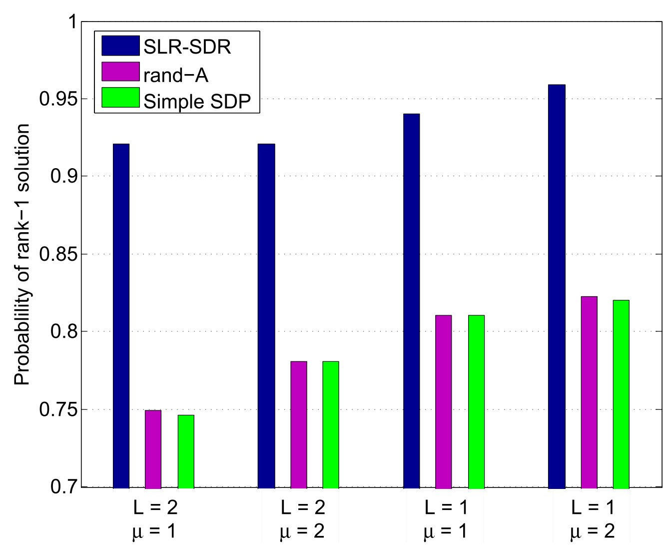


Fig. 4. Comparison of the probability of feasible rank-one solutions between SLR-SDR, simple SDP and rand-A.

initialization methods in Section V are implemented and then using $\{\mathbf{w}_{j,s}^0\}$ and $\{r_{j,s,\ell}^0\}$ by applying SGD(a)-(b), we obtain the rate vector $\{r_{j,s,\ell}\}$ and the group decoding schedules $\{\underline{Q}(k)\}$ which are used to formulate problem $P_1(\boldsymbol{\gamma})$ in (9) and $P_2(\boldsymbol{\gamma}, \mathbf{A})$ in (10).

Fig. 4 illustrates the probability of getting feasible rank-one solutions by SLR-SDR in comparison with the simple SDP and the rand-A method. It is seen that in this case, randomization has little effect in getting feasible rank-one solutions compared with the simple SDP[6]; whereas SLR-SDR provides significant improvement over the simple SDP.

[5]This could be due to the fact that by increasing the target rates, the feasible set becomes smaller and, consequently, the randomization methods have lower chances to find a feasible solution in realizations that SDP is not exact (because of their randomized nature).

[6]This is due to the fact that we have a lot of constraints in (8) because of SGD(a)-(b), and randomization methods become less effective in finding a feasible solution.

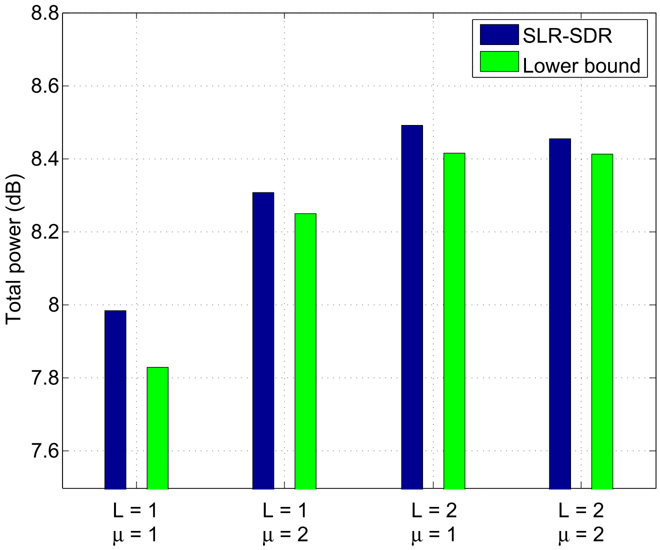


Fig. 5. Comparison of the total power of SLR-SDR with the lower bound.

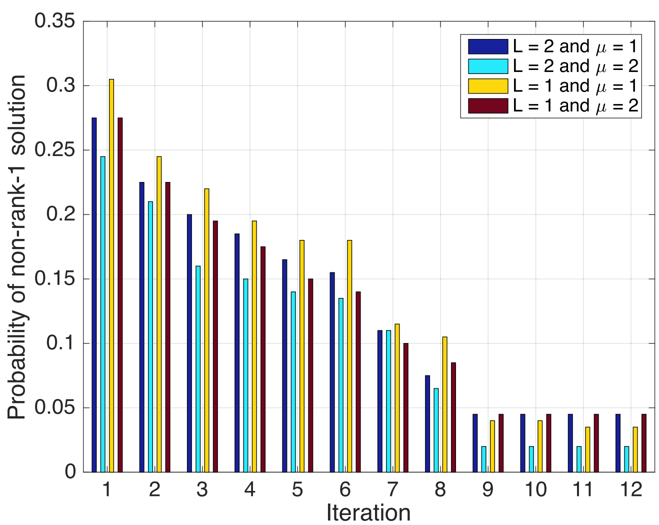


Fig. 6. Probability of non-rank-1 solution by SLR-SDR over iterations.

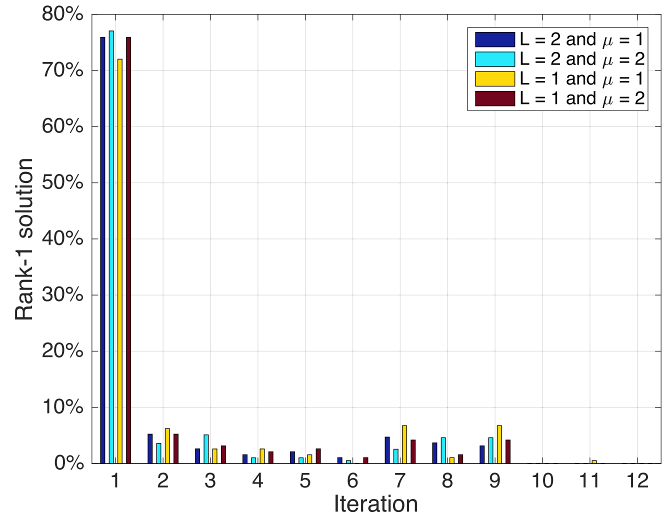


Fig. 7. Percentages of different number of iterations to reach rank-1 solutions.

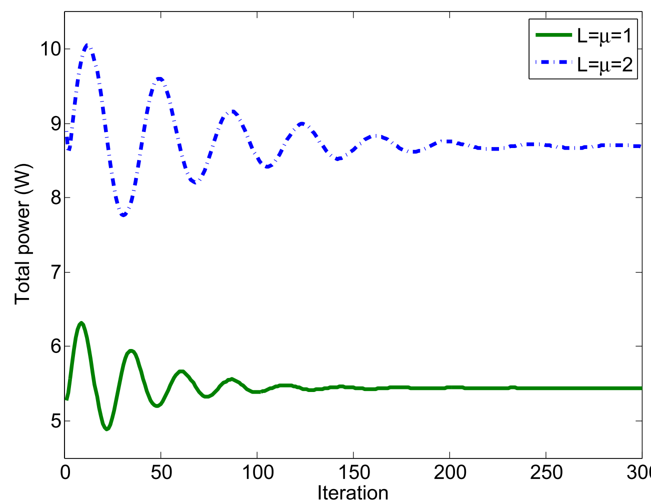


Fig. 8. Convergence behavior of the decentralized beamformer design algorithm (ADMM-BF).

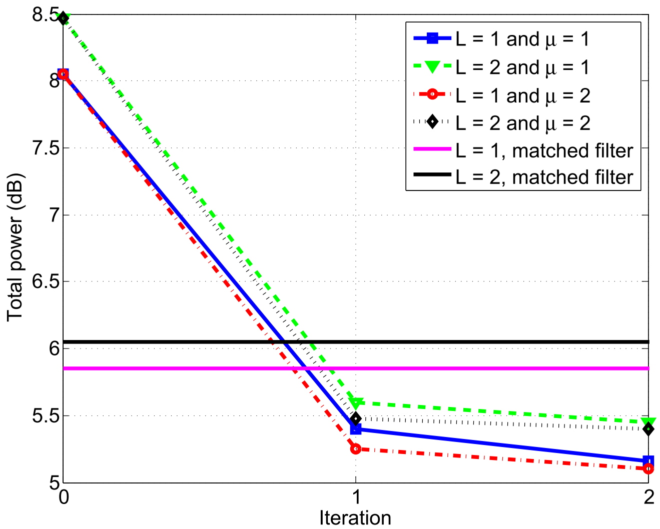
Fig. 5 illustrates the efficiency of SLR-SDR for the cases of $L \in \{1, 2\}$ and $\mu \in \{1, 2\}$, by comparing its solution with the lower bound. The lower bound is the SDP solution to the relaxed problem, i.e., the problem without the rank-1 constraints. In this plot we only consider the channel realizations where the solutions obtained by SLR-SDR are rank-one. It can be seen that SLR-SDR achieves a performance that is very close to the optimum.

Fig. 6 depicts the probability of non-rank-1 solution by SLR-SDR over iterations for different values of (L, μ). It is seen that the probability of non-rank-1 solution drops steadily over iterations and it stabilizes after 9 iterations.

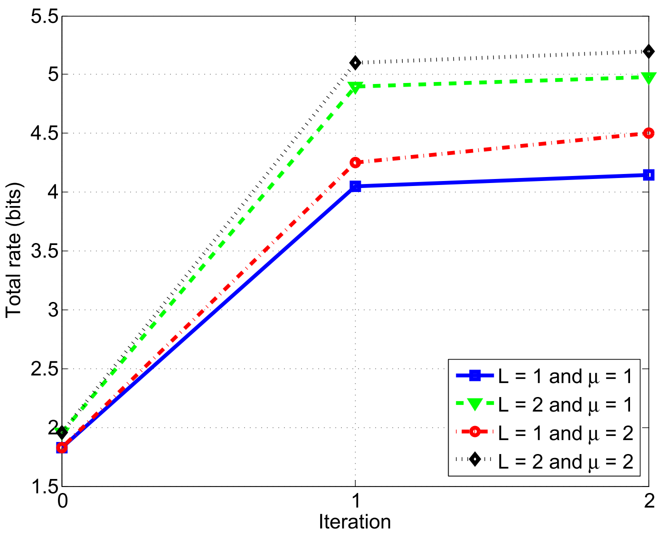
Fig. 7 depicts the number of iterations to reach rank-1 solutions (up to 12 iterations). For each pair of (L, μ), the percentages of channel realizations that take different number of iterations to reach rank-1 solutions are plotted. It can be observed that around 75% of the rank-1 solutions are obtained in the first iteration and 25% required additional iterations. The average number of iterations to obtain rank-1 solutions for all pairs of (L, μ), is only 2.13.

C. Convergence of Decentralized Algorithm (ADMM-BF)

Fig. 8 depicts the convergence of ADMM-BF to the solution of the centralized SDR, i.e., solution to problem $P_2(\boldsymbol{\gamma}, \mathbf{A})$ over iterations, for one channel realization, for $L = \mu = 1$ and $L = \mu = 2$ assuming $\mathbf{A} = \mathbf{I}$. The group decoding schedules and the rate initialization algorithm described in Section VI-B is applied here, as well. Also, we set $\rho = 2$ in the augmented Lagrangian. As it was discussed in Section IV, the proposed decentralized algorithm always converges to its optimal value and the simulation result corroborates this fact. Also we can see that at the first few iterations,



(a) Total power over iterations.



(b) Total rate over iterations.

Fig. 9. Performance of the iterative procedure for joint beamformer and rate optimization for SGD.

the algorithm converges very fast and achieves the major part of the optimal value. Note that since in ADMM-BF the local problems are solved in closed-form each iteration runs very fast.

D. Performance of Joint Beamformer and Rate Optimization for SGD

In Fig. 9, we show the rate and power improvements by the iterative procedure for joint beamformer and rate optimization given in Section V. The system setup is the same as that in Section VI-B and the stopping criterion is that the total rate increment is smaller than 0.01. Fig. 9(a) depicts the total power in each iteration and Fig. 9(b) depicts the total rate in each iteration. The 0^{th} iteration corresponds to the initialization in Fig. 1. It is seen that the major improvement is due to a single step of beamformer optimization and rate optimization in the 1^{st} iteration. For almost all channel realizations the stopping criterion is satisfied after the second iteration.

Note that the 0^{th} iteration corresponds to the conventional linear receivers typically assumed in existing works on multicast beamforing with channel-matched-filter precoders. We can employ our proposed beamformer design SLR-SDR to optimize the beamformers based on the initial rates. Then we obtain the power corresponding to optimized multicast beamforing with conventional linear receivers [3], [20]–[22]. These values are shown as the two horizontal lines in Fig. 9(a) for $L = 1$ and $L = 2$. It can be seen that the performance with SGD and optimal beamformer is superior to that of linear receiver with optimal beamformer. Hence employing the successive group decoding, both significant rate improvement and energy savings can be achieved.

VII. Conclusions

We have treated the problem of multicast beamforming design in multicell networks where advanced interference mitigation techniques are employed. In particular, each BS transmits multiple messages, each using a distinct beamformer and destined for several users. Each message is composed of several sub-messages that enables the receivers to perform partial interference decoding. Each receiver performs successive group decoding (SGD) where the desired message and a subset of interferers are decoded. The goal of beamformer design is to minimize the total power consumption while guaranteeing that all the receivers can decode their desired signals. This problem is non-convex and we have proposed an efficient algorithm based on solving a sequence of linearly regularized SDR of the problem that provides feasible and near-optimal solution with high probability. We have also proposed a very fast decentralized algorithm based on ADMM with provable convergence to the centralized solution. Finally, we have provided an iterative procedure for joint beamformer and rate optimization under the SGD framework.

Appendix
Sequential Group Decoding

The search complexity for identifying the best set of sub-messages to be decoded at each step (Steps 4, 8 and 13 in SGD(a)) can be made polynomial in the total number of codebooks $|\mathcal{M}^*|$, by a successive decoding approach and leveraging the matroid structure of the achievable rate regions [18, Lemma 2]. The total number of iterations within SGD(a) is at most $|\mathcal{M}^*|$. In SGD(a), $\mathbb{1}_{\mathcal{D}}$ denotes a vector in which all entries corresponding to the elements in set $\mathcal{D}$ are equal to 1 and the rest are 0, and $\circ$ denotes the Hadamard product of two vectors.

In SGD(a), the k^{th} user computes the rate increment x_k for all transmitted signals such that the new rate vector $\mathbf{r}_k = \mathbf{r}_k^0 + x_k \mathbf{t}_k$ for $\mathcal{A}^*(k)$ remains decodable at the k^{th} user. Based on such rate increment suggestions by all users, SGD(b) determines the final rate increment by taking the minimum of $\{x_k\}$.

SGD(a) - Computing the optimal SGD schedule and local rate increment for the k^{th} user

1: Input $\mathcal{H}_k$, $\{\mathbf{w}_{j,s}\}$, $\{r_{j,s,\ell}^0\}$ and $\mathcal{A}(k)$.

2: Initialize $\mathcal{U}=\mathcal{M}^*$, $\mathcal{V}=\emptyset$, $i=1$.

3: **Repeat**

4: Find $\mathcal{V}_i(k)=\arg\min_{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right),$$

and $\delta_i(k)=\min_{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right).$$

5: **If** $\nexists(j,s,\ell)\in\mathcal{A}^*(k)$ such that $(j,s,\ell)\in\mathcal{V}_i(k)$
$\mathcal{U}\leftarrow\mathcal{U}\backslash\mathcal{V}_i(k)$ and $\mathcal{V}\leftarrow\mathcal{V}\cup\mathcal{V}_i(k)$ and $i\leftarrow i+1$.
End if

6: **Until** $\exists(j,s,\ell)\in\mathcal{A}^*(k)$ such that $(j,s,\ell)\in\mathcal{V}_i(k)$

7: **Repeat**

8: Find $\mathcal{V}_i(k)=\arg\min_{\substack{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}\\|\mathcal{D}|\leq\mu_k}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right),$$

and $\delta_i(k)=\min_{\substack{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}\\ [1pt]|\mathcal{D}|\leq\mu_k}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right).$$

9: **If** $\nexists(j,s,\ell)\in\mathcal{A}^*(k)$ such that $(j,s,\ell)\in\mathcal{V}_i(k)$
$\mathcal{U}\leftarrow\mathcal{U}\backslash\mathcal{V}_i(k)$ and $\mathcal{V}\leftarrow\mathcal{V}\cup\mathcal{V}_i(k)$ and $i\leftarrow i+1$.
End if

10: **Until** $\exists(j,s,\ell)\in\mathcal{A}^*(k)$ such that $(j,s,\ell)\in\mathcal{V}_i(k)$

11: $q=i$.

12: **Repeat**

13: Find $Q_{i-q+1}(k)=\arg\min_{\substack{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}\\|\mathcal{D}|\leq\mu_k}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right),$$

and $\delta_i(k)=\min_{\substack{\mathcal{D}\neq\emptyset,\mathcal{D}\subseteq\mathcal{U}\\|\mathcal{D}|\leq\mu_k}}$

$$\left(\frac{\log\left(1+\frac{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{D}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2}{\frac{1}{L}\sum_{(j,s,\ell)\in\mathcal{V}}|\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}|^2+\sigma_k^2}\right)-\sum_{(j,s,\ell)\in\mathcal{D}}r_{j,s,\ell}^0}{\|\mathbf{t}\circ\mathbb{1}_{\mathcal{D}}\|_1}\right).$$

14: $\mathcal{U}\leftarrow\mathcal{U}\backslash Q_{i-q+1}(k)$ and $i\leftarrow i+1$.

15: **Until** $\nexists(j,s,\ell)\in\mathcal{A}^*(k)$ such that $(j,s,\ell)\in\mathcal{U}$.

16: Set $x_k=\delta_q(k)$ and $p_k=i-q$.

17: Output $\underline{Q}(k)=\{Q_1(k),\ldots,Q_{p_k}(k)\}$ and x_k.

SGD(b) - User coordination to obtain final rates

1: Initialize $\mathbf{r}^0$.

2: Each user k runs SGD(a) to determine $\underline{Q}(k)$ and x_k.

3: Output $\mathbf{r}=\mathbf{r}^0+\min_{1\leq k\leq K}\{x_k\}\mathbf{t}$ and $\{\underline{Q}(k)\}_{k=1}^K$.

SGD(c) - Performing the SGD for the k^{th} User

1: Input $\mathcal{H}_k$, $\{\mathbf{w}_{j,s,\ell}\}$, $\{r_{j,s,\ell}\}$ and $\underline{Q}(k)=\{Q_1(k),\ldots,Q_{p_k}(k)\}$.

2: **For** $i=1$ to p_k:

3: Initialize $\mathcal{U}=Q_i(k)$ and $\mathcal{V}=\mathcal{M}^*\backslash\cup_{j=1}^i Q_j(k)$.

4: **Check** condition $\mathbf{r}_{\mathcal{U}}\in\mathcal{C}_k(\mathcal{H}_k,\mathcal{U},\mathcal{V})$.

5: **If** check is **true**,
a) $\hat{\mathbf{x}}_{\mathcal{U}}=\arg\min_{\mathbf{X}_{\mathcal{U}}}$
$$\left|y_k-\sqrt{\frac{1}{L}}\sum_{(j,s,\ell)\in\mathcal{U}}\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}x_{j,s,\ell}\right|;$$
b) $y_k\leftarrow y_k-\sqrt{\frac{1}{L}}\sum_{(j,s,\ell)\in\mathcal{U}}\mathbf{h}_{j,k}^H\mathbf{w}_{j,s}x_{j,s,\ell}$.

6: **Else**,
a) declare outage and **quit**.

7: **End if**

8: **End for**

ACKNOWLEDGEMENT

The authors are greatly appreciative of the suggestions made by the referees and the Associate Editor, which helped us to substantially improve the presentation of this paper.

REFERENCES

[1] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[2] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[3] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, Jan. 2013.

[4] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO multicasting in noncooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1180–1193, Jun. 2014.

[5] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[6] H. Joudeh and B. Clerckx, "Sum rate maximization for MU-MISO with partial CSIT using joint multicasting and broadcasting," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 4733–4738.

[7] H. Joudeh and B. Clerckx, "A rate-splitting strategy for max-min fair multigroup multicasting," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2016, pp. 1–5.

[8] 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.

[9] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, Oct. 2014.

[10] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup beamforming for per-antenna power constrained large-scale arrays," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun./Jul. 2015, pp. 271–275.

[11] A. Schad, K. L. Law, and M. Pesavento, "Rank-two beamforming and power allocation in multicasting relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3435–3447, Jul. 2015.

[12] V. Ntranos, N. D. Sidiropoulos, and L. Tassiulas, "On multicast beamforming for minimum outage," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3172–3181, Jun. 2009.

[13] S. Y. Park and D. J. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, Jun. 2008.

[14] N. Prasad, G. Yue, X. Wang, and M. Varanasi, "Optimal successive group decoders for MIMO multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4821–4846, Nov. 2008.

[15] A. S. Motahari and A. K. Khandani, "M-user Gaussian interference channels: To decode the interference or to consider it as noise," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 2196–2200.

[16] M. A. Maddah-Ali, H. Mahdavi-Doost, and A. K. Khandani, "Optimal order of decoding for max-min fairness in K-user memoryless interference channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 2866–2870.

[17] A. Tajer, N. Prasad, and X. Wang, "Beamforming and rate allocation in MISO cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 362–377, Jan. 2010.

[18] M. Ashraphijuo, A. Tajer, C. Gong, and X. Wang, "A receiver-centric approach to interference management: Fairness and outage optimization," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5619–5642, Oct. 2016.

[19] C. Gong, A. Tajer, and X. Wang, "Group decoding for multi-relay assisted interference channels," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1489–1499, Sep. 2012.

[20] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[21] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[22] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.

[23] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[24] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[25] H. Pennanen, A. Tölli, J. Kaleva, P. Komulainen, and M. Latva-Aho, "Decentralized linear transceiver design and signaling strategies for sum power minimization in multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1729–1743, Apr. 2016.

[26] S. Lakshminarayana, M. Assaad, and M. Debbah, "Coordinated multicell beamforming for massive MIMO: A random matrix approach," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3387–3412, Jun. 2015.

[27] W.-K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, Apr. 2002.

[28] P. Tseng, "Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation," *SIAM J. Optim.*, vol. 14, no. 1, pp. 268–283, Jul. 2003.

[29] S. Zhang, "Quadratic maximization and semidefinite relaxation," *Math. Program.*, vol. 87, no. 3, pp. 453–465, 2000.

[30] Z.-Q. Luo, *Lecture Notes for EE8950: Engineering Optimization*. Minneapolis, MN, USA: Univ. of Minnesota, 2004.

[31] L.-N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.

[32] K. Huang and N. D. Sidiropoulos, "Consensus-ADMM for general quadratically constrained quadratic programming," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, Oct. 2016.

[33] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Control Conf.*, vol. 6. 2001, pp. 4734–4739.

[34] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. Amer. Control Conf.*, vol. 3. 2003, pp. 2156–2162.

[35] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*. Raleigh, NC, USA: Lulu.com, Jun. 2010.

[36] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Optim. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[37] R. Glowinski and A. Marrocco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires," *Rev. Française d'autom., Inf., Recherche Opér.*, vol. 9, no. 2, pp. 41–76, 1975.

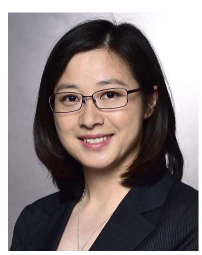
[38] J. Eckstein, "Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," Rutgers Center Oper. Res. Rutgers Univ., Piscataway, NJ, USA, RUTCOR Res. Rep. 32, 2012.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[40] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Lyngby, Denmark: Tech. Univ. Denmark, 2012.

Mehdi Ashraphijuo received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Columbia University, New York, NY, USA. His research interests lie in the general areas of network information theory and statistical signal processing with applications in wireless communication. His current research focuses primarily on developing fundamental principles for communication network design, with emphasis on developing interference management and understanding the role of feedback and user cooperation. He was a recipient of Qualcomm Innovation Fellowship in 2014.

Xiaodong Wang (S'98–M'98–SM'04–F'08) received the Ph.D. degree in electrical engineering from Princeton University Princeton, NJ, USA. He is currently a Professor of Electrical Engineering with Columbia University, New York NY, USA. He authored the book *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice Hall, 2003). His research interests include computing, signal processing and communications. He has authored extensively in these areas. His current research interests include wireless communications, statistical signal processing, and genomic signal processing. He received the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY. He is listed as an ISI highly-cited author.



Meixia Tao (S'00–M'04–SM'10) received the B.S. degree in electronics engineering from Fudan University, Shanghai, China, in 1999, and the Ph.D. degree in electrical and electronic engineering from The Hong Kong University of Science and Technology in 2003. She was a member of the Professional Staff with The Hong Kong Applied Science and Technology Research Institute from 2003 to 2004, and a Teaching Fellow and then an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from 2004 to 2007. She is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. Her current research interests include content-centric wireless networks, resource allocation, interference management, and physical layer security.

Dr. Tao serves as a member of the Executive Editorial Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. She was a recipient of the IEEE Heinrich Hertz Award for Best Communications Letters in 2013, the IEEE/CIC ICCC Best Paper Award in 2015, the International Conference on Wireless Communications and Signal Processing Best Paper Award in 2012, and the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2009. She also served as the TPC Chair of the IEEE/CIC International Conference on Communications, China, in 2014 and the Symposium Co-Chair of the IEEE ICC in 2015. She was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2010, the IEEE COMMUNICATIONS LETTERS from 2009 to 2012, and the IEEE WIRELESS COMMUNICATIONS LETTERS from 2011 to 2015.