On Cooperative Beamforming Based on Second-Order Statistics of Channel State Information

Jiangyuan Li, Athina P. Petropulu, H. Vincent Poor*

Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104

*School of Engineering and Applied Science, Princeton University, Princeton, NJ 08544

Abstract

[1] Cooperative beamforming in relay networks is considered, in which a source transmits to its destination with the help of a set of cooperating nodes. The source first transmits locally. The cooperating nodes that receive the source signal retransmit a weighted version of it in an amplify-and-forward (AF) fashion. Assuming knowledge of the second-order statistics of the channel state information, beamforming weights are determined so that the signal-to-noise ratio (SNR) at the destination is maximized subject to two different power constraints, i.e., a total (source and relay) power constraint, and individual relay power constraints. For the former constraint, the original problem is transformed into a problem of one variable, which can be solved via Newton's method. For the latter constraint, the original problem is transformed into a homogeneous quadratically constrained quadratic programming (QCQP) problem. In this case, it is shown that when the number of relays does not exceed three the global solution can always be constructed via semidefinite programming (SDP) relaxation and the matrix rank-one decomposition technique. For the cases in which the SDP relaxation does not generate a rank one solution, two methods are proposed to solve the problem: the first one is based on the coordinate descent method, and the second one transforms the QCQP problem into an infinity norm maximization problem in which a smooth finite norm approximation can lead to the solution using the augmented Lagrangian method.

Index Terms

Cooperative beamforming, channel uncertainty, relay networks, fractional programming, semidefinite programming.

I. INTRODUCTION

Cooperative beamforming (CB), also called distributed beamforming has attracted considerable research interest recently, due to its potential for improving communication reliability. One form of distributed beamforming, the so-called distributed transmit beamforming, is a form of cooperative communications in which a network of multiple transmitters cooperate to transmit a common message coherently to a Base Station (BS). The distributed transmit beamforming can provide energy efficiency and reasonable directional gain for ad hoc sensor networks [1], [2]. The challenges and recent progress of distributed transmit beamforming are discussed in [3]. Another form of distributed beamforming is the distributed relay beamforming, in which a set of cooperating nodes act as a virtual antenna array and adjust their transmission weights to form a beam to the destination. This can result in diversity gains similar to those of multiple-antenna systems [7], [10]. Various effective cooperation schemes have been proposed in the literature, such as amplify-and-forward (AF), decode-and-forward (DF) [4], coded-cooperation [5], and compress-and-forward [6]. The AF protocol, due to its simplicity, is of particular interest [10].

In distributed relay beamforming, the objective is to determine source power and beamforming weights according to some optimality criterion. Existing results for this problem can be classified into those that rely on channel state information (CSI) availability at the relays [7], [8], [9], and those that allow for channel uncertainly, i.e., that rely on statistics of CSI, such as the covariance of channel coefficients, or imperfect CSI feedback [10], [11], [12], as opposed to explicit CSI. The latter class of techniques is particularly important because CSI is never perfectly known at the transmitter. This work picks up on some important results presented in [10], in which a source transmits a signal to a destination with the assistance of a set of AF relay nodes In [10], the problem of obtaining the beamforming weights so that the signal-to-noise ratio (SNR) at the destination is maximized subject to certain power constraints is considered, i.e., individual relay power constraints and a total power relay constraint. For the case of individual relay power constraints, a semidefinite programming (SDP) relaxation plus bisection search technique was proposed in [10]. When the SDP relaxation generates a rank-one solution, then this is the exact solution of the original problem; otherwise, the exact solution cannot be guaranteed, and the authors of [10] proposed a Gaussian random procedure (GRP) to search for an approximate solution based on the SDP relaxation solution. However, GRP is time-consuming and sometimes ineffective.

In this paper, we investigate the same scenario as in [10], i.e., cooperative beamforming under the assumption that the second-order statistics of the channel state information (CSI) are available. The beamforming weights are determined so that the SNR at the destination is maximized subject to two

different power constraints: (i) a total (source plus relay) power constraint, and (ii) individual relay constraints. The differences of this work as compared to [10], are the following.

- Our first kind of power constraint includes the source power as well as the power of the relays. In a wireless network all nodes have power constraints, therefore, placing a constraint on the source is more realistic. However, this results in a more difficult optimization problem. A similar constraint was also used in [13]. For this case, we transform the original problem into a problem of one variable, which can then be solved via Newton's method.

- The second kind of power constraint is exactly the same as that of [10], but our work contributes new results and more efficient algorithms to reach the solution. In particular,
  - We show that when the number of relays does not exceed three, the global solution can always be constructed via SDP relaxation and the matrix rank-one decomposition technique.
  - For the case in which the SDP relaxation solution has rank greater than one, we propose two methods to obtain an approximate solution that is more effective than the Gaussian random procedure employed in [10]. The first method is based on the coordinate descent method. The second method transforms the original problem into an infinity norm maximization problem, for which a smooth finite norm approximation results in a solution using the augmented Lagrangian method.

- For both types of constraints, we obtain exact solutions for the special cases in which the channel coefficients between different node pairs are uncorrelated and follow a Rayleigh fading model. These cases were not discussed in [10].

The remainder of the paper is organized as follows. The mathematical model is introduced in §II. In §III, the SNR maximization subject to a total power constraint is presented. The SNR maximization subject to individual relay power constraints is developed in Section §IV. Numerical results are presented in §V to illustrate the proposed algorithms. Finally, §VI provides concluding remarks.

A. Notation

Upper case and lower case bold symbols denote matrices and vectors, respectively. Superscripts $*$, T and $\dagger$ denote respectively conjugate, transposition and conjugate transposition. $|\cdot|$ denotes the amplitude of a complex number. $\det(\mathbf{A})$ and $\mathrm{Tr}(\mathbf{A})$ denote determinant and trace of matrix $\mathbf{A}$, respectively. $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the smallest and largest eigenvalues of $\mathbf{A}$, respectively. $\mathbf{A} \succeq 0$ and $\mathbf{A} \succ 0$ mean that matrix $\mathbf{A}$ is Hermitian positive semidefinite, and positive definite, respectively. $\mathbf{A} \succeq \mathbf{B}$ denotes

that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. $\text{rank}(\mathbf{A})$ denotes the rank of matrix $\mathbf{A}$. $\text{diag}(\mathbf{v})$ denotes a diagonal matrix with diagonal entries consisting of the elements of $\mathbf{v}$. $\|\mathbf{a}\|$ denotes Euclidean norm of vector $\mathbf{a}$. $\mathbf{I}_n$ denotes the identity matrix of order n (the subscript is dropped when the dimension is obvious). $\mathbb{E}(\cdot)$ denotes expectation.
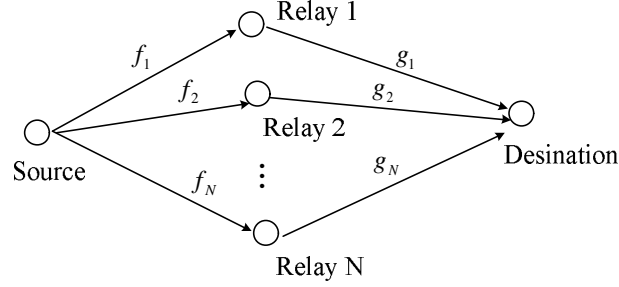

Fig. 1. System model.

II. SYSTEM MODEL AND PROBLEM STATEMENT

The system model is the same as in [10] and is depicted in Fig. 1. It consists of a source node, a destination node and N relay nodes, each node equipped with a single antenna. The source transmits signals to the destination with the help of relay nodes. We assume that the direct link between the source and destination is very weak and thus ignored. The channel gains from the source to the ith relay, and from the ith relay to the destination, are denoted respectively by f_i and g_i.

Communication between source and destination occurs in two stages (slots). During the first stage, the source broadcasts its signal to the relays. During the second stage, the relays working in AF fashion transmit a weighted version of the signal that they received during the first stage. Let $\sqrt{P_s}\, s$ be the source signal, where P_s is the source transmit power and s is the information symbol with $\mathbb{E}(|s|^2) = 1$. The received signal at the ith relay is given by

$$x_i = \sqrt{P_s}\, f_i s + v_i \tag{1}$$

where v_i represents the noise at the ith relay having zero mean and variance σ^2. The ith relay weights the received signal and transmits $z_i = w_i x_i$ where w_i is the weight. The received signal at the destination equals

$$y = \sum_{i=1}^{N} g_i z_i + \nu = \sqrt{P_s} \sum_{i=1}^{N} w_i f_i g_i s + \sum_{i=1}^{N} w_i g_i v_i + \nu \tag{2}$$

where ν is the noise at the destination having zero mean and variance σ^2.

Let us assume that the second-order statistics of the channel gains f_i's and g_i's are known. We also assume that f_i and g_j, $\forall i, j$ are statistically independent. Define

$$\mathbf{w} = [w_1, \cdots, w_N]^T,$$

$$\mathbf{h} = [f_1 g_1, \cdots, f_N g_N]^T,$$

$$\mathbf{g} = [g_1, \cdots, g_N]^T,$$

$$\mathbf{R} = \mathbb{E}\{\mathbf{h}\mathbf{h}^\dagger\},$$

$$\mathbf{Q} = \mathbb{E}\{\mathbf{g}\mathbf{g}^\dagger\},$$

$$\text{and } \mathbf{D} = \text{diag}(\mathbb{E}\{|f_1|^2\}, \cdots, \mathbb{E}\{|f_N|^2\}). \tag{3}$$

In general, $\mathbf{Q}$ and $\mathbf{R}$ are full matrices. In case of uncorrelated Rayleigh fading, in holds that $\mathbb{E}(f_i^* f_j) = 0$, and $\mathbb{E}(g_i^* g_j) = 0$, $\forall i \neq j$, in which case $\mathbf{R}$ and $\mathbf{Q}$ both are diagonal.

From (2), the signal component power is given by

$$P_d = \mathbb{E}\left\{\left|\sqrt{P_s}\sum_{i=1}^{N} w_i f_i g_i s\right|^2\right\} = P_s \mathbf{w}^\dagger \mathbf{R} \mathbf{w} \tag{4}$$

and the total noise power P_n equals

$$P_n = \mathbb{E}\left\{\left|\sum_{i=1}^{N} w_i g_i v_i + \nu\right|^2\right\} = \sigma^2 + \sigma^2 \mathbf{w}^\dagger \mathbf{Q} \mathbf{w}. \tag{5}$$

The SNR at the destination is given by

$$\Gamma_d = \frac{P_d}{P_n} = \frac{P_s}{\sigma^2} \frac{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}}{1 + \mathbf{w}^\dagger \mathbf{Q} \mathbf{w}}. \tag{6}$$

The total relay transmit power and transmit power at the ith relay are respectively given by

$$P_r = \sum_{i=1}^{N} \mathbb{E}\{|z_i|^2\} = P_s \mathbf{w}^\dagger \mathbf{D} \mathbf{w} + \sigma^2 \mathbf{w}^\dagger \mathbf{w} \tag{7}$$

$$P_{r,i} = \mathbb{E}\{|z_i|^2\} = (P_s D_{ii} + \sigma^2)|w_i|^2 \tag{8}$$

where D_{ii} is the (i, i)th entry of $\mathbf{D}$.

Our goal in this paper is to determine the beamforming weights w_i's such that Γ_d is maximized subject to certain power constraints. In this paper, we consider two kinds of power constraints. The first kind corresponds to the case in which the total power of the source and all relays is constrained, i.e.,

$$P_s + P_r \leq P_0. \tag{9}$$

where P_0 is the maximum allowable total transmit power of the source and all relays. The second kind is the individual relay power constraints in which each relay node is restricted in its transmit power, i.e,

$$P_{r,i} \leq P_i \tag{10}$$

where P_i is the maximum allowable transmit power of the ith relay.

III. SNR Maximization Under Total Power Constraint

From (6) and (9), the SNR maximization problem subject to a total power constraint is expressed as

$$\max_{P_s, \mathbf{w}} \frac{P_s}{\sigma^2} \frac{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}}{1 + \mathbf{w}^\dagger \mathbf{Q} \mathbf{w}} \tag{11}$$

$$\text{s.t.} \quad P_s + P_s \mathbf{w}^\dagger \mathbf{D} \mathbf{w} + \sigma^2 \mathbf{w}^\dagger \mathbf{w} \leq P_0.$$

We give the following lemma, the proof of which can be found in Appendix A.

Lemma 1: *Let P_s° be the solution of the following*

$$\max_{P_s} \frac{P_s}{\sigma^2} \frac{P_0 - P_s}{\lambda_{\min}(P_s \mathbf{S}_1 + (P_0 - P_s)\mathbf{S}_2)} \tag{12}$$

$$\text{s.t.} \quad 0 \leq P_s \leq P_0$$

where

$$\mathbf{S}_1 = \mathbf{R}^{-1/2} \mathbf{D} \mathbf{R}^{-1/2} + (\sigma^2/P_0)\mathbf{R}^{-1}, \tag{13}$$

$$\text{and } \mathbf{S}_2 = \mathbf{R}^{-1/2} \mathbf{Q} \mathbf{R}^{-1/2} + (\sigma^2/P_0)\mathbf{R}^{-1}. \tag{14}$$

Let $\mathbf{w}^\circ$ be the eigenvector associated with the smallest eigenvalue of $P_s^\circ \mathbf{S}_1 + (P_0 - P_s^\circ)\mathbf{S}_2$. Then $(P_s^\circ, \mathbf{w}^\circ)$ is the solution to the problem of (11).

Remarks: Here we assume that $\mathbf{R} \succ 0$. If $\mathbf{R} \not\succ 0$, the methodology is similar. In fact, from Appendix A, the problem of (11) is also equivalent to

$$\max_{P_s} P_s(P_0 - P_s)\lambda_{\max} \left([P_s \mathbf{D} + \sigma^2 \mathbf{I} + (P_0 - P_s)\mathbf{Q}]^{-\frac{1}{2}} \mathbf{R} [P_s \mathbf{D} + \sigma^2 \mathbf{I} + (P_0 - P_s)\mathbf{Q}]^{-\frac{1}{2}} \right) \tag{15}$$

$$\text{s.t.} \quad 0 \leq P_s \leq P_0.$$

A similar procedure can be used to solve the above problem.

Let us normalize P_s by letting $x = P_s/P_0$, $0 \leq x \leq 1$. With this, the problem of (12) is equivalent to

$$\max_{x} \frac{P_0}{\sigma^2} \frac{x(1-x)}{\lambda_{\min}(x\mathbf{S}_1 + (1-x)\mathbf{S}_2)} \tag{16}$$

$$\text{s.t.} \quad 0 \leq x \leq 1.$$

A. $\mathbf{S}_1$ and $\mathbf{S}_2$ are both diagonal

In case of uncorrelated Rayleigh fading, $\mathbf{R}$ and $\mathbf{Q}$ are diagonal matrices. Then, $\mathbf{S}_1$ and $\mathbf{S}_2$ are both diagonal, and as it will be shown next the exact solution can be obtained analytically.

By denoting the (k,k)-th entry of $\mathbf{S}_1$ and $\mathbf{S}_2$ as a_k and b_k, respectively, the problem of (16) becomes

$$
\begin{aligned}
&\min_{0<x<1} \ \lambda_{\min}\left(\frac{1}{1-x}\mathbf{S}_1 + \frac{1}{x}\mathbf{S}_2\right)\\
=&\min_{0<x<1}\ \min_{k=1,\cdots,N}\ \left\{\frac{a_k}{1-x} + \frac{b_k}{x}\right\}\\
=&\min_{k=1,\cdots,N}\ \min_{0<x<1}\ \left\{\frac{a_k}{1-x} + \frac{b_k}{x}\right\}\\
=&\min_{k=1,\cdots,N}\ \left(\sqrt{a_k} + \sqrt{b_k}\right)^2\\
=&\left(\sqrt{a_{k_0}} + \sqrt{b_{k_0}}\right)^2.
\end{aligned}
\tag{17}
$$

The above minimum is attained for

$$
x = \frac{\sqrt{b_{k_0}}}{\sqrt{a_{k_0}} + \sqrt{b_{k_0}}}
\tag{18}
$$

where

$$
k_0 = \arg\min_{k=1,\cdots,N}\ \left(\sqrt{a_k} + \sqrt{b_k}\right)^2.
\tag{19}
$$

B. $\mathbf{S}_1$ or $\mathbf{S}_2$ is not diagonal

Lemma 2: *The optimal x of (16) lies in $[x_l, x_u]$ where*

$$
x_l = \frac{\sqrt{c}}{1 + \sqrt{c}},
\tag{20}
$$

$$
\text{and } x_u = \frac{\sqrt{d}}{1 + \sqrt{d}}
\tag{21}
$$

where $c = \lambda_{\min}(\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2})$, and $d = \lambda_{\max}(\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2})$.
The proof is given in Appendix B.

From Lemma 2, to solve the problem of (16) is equivalent to solving the problem of

$$
\min_x\ \lambda_{\min}\left(\frac{1}{1-x}\mathbf{S}_1 + \frac{1}{x}\mathbf{S}_2\right)
\tag{22}
$$

$$
\text{s.t.}\quad x_l \leq x \leq x_u.
$$

The objective in (22) is in general not a convex function over $[x_l, x_u]$. We will use Newton's method to search for the stationary points. Let us start by denoting

$$
\mathbf{G}(x) = \frac{1}{1-x}\mathbf{S}_1 + \frac{1}{x}\mathbf{S}_2, \ x \in [x_l, x_u].
\tag{23}
$$

Note that $\mathbf{G}(x)$ depends smoothly on $x \in [x_l, x_u]$ as any order derivative of $\mathbf{G}(x)$ exists. We assume that $\mathbf{G}(x)$ has a simple spectrum for $x \in [x_l, x_u]$. This is a reasonable assumption for general $\mathbf{S}_1$ and $\mathbf{S}_2$ (see [14], [15, §4]). Under this assumption, $\lambda_{\min}(\mathbf{G}(x))$ also depends smoothly on $x \in [x_l, x_u]$ [14]. First- and second- order necessary conditions for x to be a local minimizer are respectively [31, Theorem 2.2, 2.3]

$$\frac{\mathrm{d}}{\mathrm{d}x}\lambda_{\min}(\mathbf{G}(x)) = 0, \tag{24}$$

$$\text{and } \frac{\mathrm{d}^2}{\mathrm{d}x^2}\lambda_{\min}(\mathbf{G}(x)) \geq 0. \tag{25}$$

If (25) holds with strict inequality, then x is a strict local minimizer [31, Theorem 2.4]. In Newton's method, the $(k+1)$th iteration is given by [31, Ch. 3]

$$x_{k+1} = x_k - \alpha_k \frac{\frac{\mathrm{d}}{\mathrm{d}x}\lambda_{\min}(\mathbf{G}(x))}{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\lambda_{\min}(\mathbf{G}(x))}, \ \ k = 0, 1, \cdots \tag{26}$$

where $\alpha_k > 0$ is chosen such that x_{k+1} does not exceed $[x_l, x_u]$, and otherwise, $\alpha_k \leftarrow \alpha_k/2$.

In the iteration expression (26), we need to calculate the first- and second- order derivatives of $\lambda_{\min}(\mathbf{G}(x))$. Let $\mathbf{u}_0(x)$ be the eigenvector associated with $\lambda_{\min}(\mathbf{G}(x))$. Let $\mathbf{u}_k(x)$, $k = 1, \cdots, N-1$ be the eigenvectors associated with the other eigenvalues $\lambda_k(x)$ of $\mathbf{G}(x)$, respectively, where $\lambda_1(x) > \cdots > \lambda_{N-1}(x) > \lambda_{\min}(\mathbf{G}(x))$. The first- and second- order derivatives of $\lambda_{\min}(\mathbf{G}(x))$ (the so-called *Hadamard first variation formula* and *Hadamard second variation formula* [15, §4]) are respectively given by [16], [17]

$$\frac{\mathrm{d}}{\mathrm{d}x}\lambda_{\min}(\mathbf{G}(x)) = \mathbf{u}_0(x)^\dagger \frac{\mathrm{d}\mathbf{G}(x)}{\mathrm{d}x}\mathbf{u}_0(x), \tag{27}$$

$$\text{and } \frac{\mathrm{d}^2}{\mathrm{d}x^2}\lambda_{\min}(\mathbf{G}(x)) = \mathbf{u}_0(x)^\dagger \frac{\mathrm{d}^2\mathbf{G}(x)}{\mathrm{d}x^2}\mathbf{u}_0(x) - \sum_{j=1}^{N-1} \frac{2\left|\mathbf{u}_j(x)^\dagger \frac{\mathrm{d}\mathbf{G}(x)}{\mathrm{d}x}\mathbf{u}_0(x)\right|^2}{\lambda_j(x) - \lambda_{\min}(\mathbf{G}(x))} \tag{28}$$

where

$$\frac{\mathrm{d}\mathbf{G}(x)}{\mathrm{d}x} = \frac{1}{(1-x)^2}\mathbf{S}_1 - \frac{1}{x^2}\mathbf{S}_2, \tag{29}$$

$$\text{and } \frac{\mathrm{d}^2\mathbf{G}(x)}{\mathrm{d}x^2} = \frac{2}{(1-x)^3}\mathbf{S}_1 + \frac{2}{x^3}\mathbf{S}_2. \tag{30}$$

IV. SNR MAXIMIZATION UNDER INDIVIDUAL RELAY POWER CONSTRAINTS

From (6) and (10), the SNR maximization problem subject to individual relay power constraints is expressed as

$$\max_{\mathbf{w}} \frac{P_s}{\sigma^2}\frac{\mathbf{w}^\dagger\mathbf{R}\mathbf{w}}{1 + \mathbf{w}^\dagger\mathbf{Q}\mathbf{w}} \tag{31}$$

$$\text{s.t. } (P_s D_{kk} + \sigma^2)|w_k|^2 \leq P_k, \ k \in I$$

where $I = \{1, 2, \cdots, N\}$. The problem of (31) belongs to the class of quadratically constrained fractional programs. In [10], this problem was analyzed and an SDP relaxation plus bisection search technique was proposed. Here, we first consider the case of uncorrelated Rayleigh fading, and show that an exact solution can be obtained. Then, for the general fading case, we propose two methods that are more efficient than the search method of [10]. As it will be shown in the simulations section, the random search approach, in addition to being time consuming, can result in a noticeable performance gap as compared to the proposed approaches.

A. $\mathbf{R}$ and $\mathbf{Q}$ are both diagonal

By using the Dinkelbach-type method [18], we introduce the following function:

$$F(t) = \max_{\mathbf{w}} \left[f(t, \mathbf{w}) = \frac{P_s}{\sigma^2} \mathbf{w}^\dagger \mathbf{R} \mathbf{w} - t(1 + \mathbf{w}^\dagger \mathbf{Q} \mathbf{w}) \right] \tag{32}$$

$$\text{s.t.} \quad (P_s D_{kk} + \sigma^2)|w_k|^2 \le P_k, \quad k \in I.$$

The relation between $F(t)$ and the problem of (31) is given in the following property [18].

Property 1:

(i) $F(t)$ is strictly decreasing, and $F(t) = 0$ has a unique root, say $t^\star$;

(ii) Let $\mathbf{w}^\star$ be the solution of (32) corresponding to $t^\star$. Then $\mathbf{w}^\star$ is also the solution of (31) with the largest objective value $t^\star$ exactly.

According to Property 1, we aim to find $t^\star$ and the associated $\mathbf{w}^\star$, which is also the solution of (31). To this end, by denoting the (k, k)th entry of $\mathbf{R}$, $\mathbf{Q}$ as r_k, q_k, respectively, we rewrite

$$f(t, \mathbf{w}) = -t + \sum_{k=1}^{N} \left(\frac{P_s}{\sigma^2} r_k - t q_k \right) |w_k|^2 \tag{33}$$

to get that

$$F(t) = -t + \sum_{k=1}^{N} \frac{P_k}{P_s D_{kk} + \sigma^2} \varphi \left(\frac{P_s}{\sigma^2} r_k - t q_k \right) \tag{34}$$

associated with the optimal

$$|w_k|^2 = \begin{cases} \frac{P_k}{P_s D_{kk} + \sigma^2} & \frac{P_s}{\sigma^2} r_k - t q_k > 0 \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

where

$$\varphi(x) \triangleq \begin{cases} x & x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

To find the root of $F(t) = 0$, let us denote

$$t_k = \frac{P_s r_k}{\sigma^2 q_k}, \quad k = 1, \cdots, N \tag{37}$$

and their rearrangement $\tilde{t}_1 < \tilde{t}_2 < \cdots < \tilde{t}_N$ corresponding to $\tilde{r}_k$, $\tilde{q}_k$, $\tilde{P}_k$, and $\tilde{D}_{kk}$, respectively. With these, we rewrite (34) as

$$F(t) = -t + \sum_{k=1}^{N} \frac{\tilde{P}_k}{P_s \tilde{D}_{kk} + \sigma^2} \varphi\left(\frac{P_s}{\sigma^2}\tilde{r}_k - t\tilde{q}_k\right). \tag{38}$$

Note that $F(0) > 0$ and $F(\tilde{t}_N) = -\tilde{t}_N < 0$. Thus, it follows from Property 1 that $0 < t^\star < \tilde{t}_N$. The root $t^\star$ is determined based on the following theorem, the proof of which is given in Appendix C.

Theorem 1: *If $F(\tilde{t}_{k_0}) = 0$ for an integer k_0, then $t^\star = t_{k_0}$. Otherwise, let k_0 be the smallest integer such that $F(\tilde{t}_{k_0}) < 0$. Then*

$$t^\star = \left(1 + \sum_{k=k_0}^{N} \frac{\tilde{P}_k \tilde{q}_k}{P_s \tilde{D}_{kk} + \sigma^2}\right)^{-1} \sum_{k=k_0}^{N} \frac{\tilde{P}_k P_s \tilde{r}_k}{(P_s \tilde{D}_{kk} + \sigma^2)\sigma^2}. \tag{39}$$

Once $t^\star$ is obtained, we can obtain $\mathbf{w}^\star$ from (35).

B. $\mathbf{R}$ *or* $\mathbf{Q}$ *is not diagonal*

1) Equivalent QCQP and SDP relaxation: The problem of (31) is equivalent (*up to scaling*) to a QCQP, as stated in the following lemma. The proof of the lemma is given in Appendix D.

Lemma 3: *Let $\mathbf{w}^\circ$ be the solution of the following homogeneous QCQP problem:*

$$\max_{\mathbf{w}} \ \mathbf{w}^\dagger \mathbf{R}\mathbf{w} \tag{40}$$

$$\text{s.t.} \quad \mathbf{w}^\dagger \mathbf{A}_k \mathbf{w} \leq 1, \ k \in I$$

where

$$\mathbf{A}_k = \frac{P_s D_{kk} + \sigma^2}{P_k}\mathbf{J}_k + \mathbf{Q} \tag{41}$$

and $\mathbf{J}_k$ is a matrix with all zero entries except for the (k, k)th entry one. Let

$$\eta = \max_{k \in I} \frac{P_s D_{kk} + \sigma^2}{P_k}\mathbf{w}^{\circ\dagger}\mathbf{J}_k\mathbf{w}^\circ. \tag{42}$$

Then $\frac{1}{\sqrt{\eta}}\mathbf{w}^\circ$ is the solution to the problem of (31).

Remarks: In fact, Lemma 3 states that the QCQP of (40) and the problem of (31) are equivalent *up to scaling*.

Note that the constraint in (40) is convex but the objective is concave. Thus, the problem of (40) is not a convex problem. In fact, this problem belongs to the class of problems involving maximization of convex functions over a convex set [19].

The SDP relaxation is a popular method for QCQP problems. Let $\mathbf{X} = \mathbf{w}\mathbf{w}^\dagger$, and we can write $\mathbf{w}^\dagger\mathbf{R}\mathbf{w} = \mathrm{Tr}(\mathbf{R}\mathbf{X})$, $\mathbf{w}^\dagger\mathbf{A}_k\mathbf{w} = \mathrm{Tr}(\mathbf{A}_k\mathbf{X})$. With this, we can rewrite the problem of (40) as

$$\min_{\mathbf{X}} \ -\mathrm{Tr}(\mathbf{R}\mathbf{X}) \tag{43}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{A}_k\mathbf{X}) \le 1, \ k \in I$$

$$\mathbf{X} \succeq 0,$$

$$\mathrm{rank}(\mathbf{X}) = 1.$$

Dropping the non-convex constraint $\mathrm{rank}(\mathbf{X}) = 1$, we obtain the SDP relaxation [30]

$$\min_{\mathbf{X}} \ -\mathrm{Tr}(\mathbf{R}\mathbf{X}) \tag{44}$$

$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{A}_k\mathbf{X}) \le 1, \ k \in I$$

$$\mathbf{X} \succeq 0.$$

The SDP of (44) a convex problem which can be effectively solved by CVX software [32]. Let $\mathbf{X}^\star$ be such a solution. Obviously, if $\mathbf{X}^\star$ has rank one, then it is the solution to the problem of (43) and hence generates the solution to the problem of (40). Otherwise, a search technique may be used to obtain the suboptimal solution of the original problem, e.g., the Gaussian random procedure (GRP) [10]. For general $\mathbf{R}$ and $\mathbf{Q}$, the solution $\mathbf{X}^\star$ from CVX software does not necessarily have rank one (in fact, for general $\mathbf{R}$ and $\mathbf{Q}$ matrices, the SDP of (44) does not necessarily have a rank one solution). Some examples on the above claim will be given in the simulation section below.

The SDP relaxation problem of (44) has several advantages as compared to the SDP relaxation of [10]. First, it obtains the same objective value while avoiding the bisection search. Second, for $N = 2, 3$, it attains the global optimal solution in polynomial time. In other words, for $N = 2, 3$, one can ensure that the problem of (44) has a rank one solution. Moreover, one can construct a rank one solution from any non rank one $\mathbf{X}^\star$ in polynomial time. In fact, for $N = 2, 3$, the problem has been solved using the complex matrix rank-one decomposition [20, Theorem 2.1], as stated in the following theorem.

Theorem 2: *For $N = 2, 3$, the problem of (44) has a rank one solution. Let $\mathbf{X}^\star$ be any one of the solutions. If $\mathbf{X}^\star$ has a rank greater than one, one can construct a rank one solution from $\mathbf{X}^\star$ in polynomial time by using the complex matrix rank one decomposition.*

For the case in which the solution $\mathbf{X}^\star$ from the CVX software has a rank greater than one, the GRP can used, although it is in general time-consuming and sometimes ineffective. In the following, we give two more effective methods for that case.

2) Coordinate descent method: If the solution $\mathbf{X}^\star$ from CVX software has rank greater than one we can use the coordinate descent method [21, §8.9], [22, §2.7], [23], [24] to directly deal with the original problem of (31). Note that the constraints of the problem of (31) are some bounds for the elements of $\mathbf{w}$, i.e., a Cartesian product of some closed convex sets (see [22, §2.7]). The idea behind the coordinate descent method is the following. At each iteration, the objective is minimized with respect to one element of $\mathbf{w}$ while keeping the other elements fixed. The method is particularly attractive when the subproblem is easy to solve (e.g., there is a closed form solution) and also satisfies certain condition for convergence [22, Proposition 2.7.1], [23, Theorem 4.1], [24, §6]. The coordinate descent algorithm applied to our problem is as follows.

Algorithm 1:

1) *Set $\varepsilon = 10^{-3}$; Choose an initial point $\mathbf{w}^0$; Set $k = 0$.*
2) *For $p = 1 : N$, determine the optimal pth element while keeping the other elements fixed. This results in $\mathbf{w}_p^k$;*
3) *$\mathbf{w}^{k+1} = \mathbf{w}_N^k$;*
4) *If $\frac{\|\mathbf{w}^{k+1} - \mathbf{w}^k\|}{\|\mathbf{w}^k\|} < \varepsilon$, stop;*
5) *$k = k + 1$; Go to 2).*

In the following, we show that the subproblem stated in Step 2 has a closed form solution (see Theorem 3) and also study its convergence to a stationary point (see Theorem 4).

It is easy to verify that minimizing the objective with respect to the kth element of $\mathbf{w}$ while keeping the other elements fixed leads to the following optimization problem:

$$\max_y \quad \frac{a_1|y|^2 + b_1 y + b_1^* y^* + c_1}{a_2|y|^2 + b_2 y + b_2^* y^* + c_2} \tag{45}$$

$$\text{s.t.} \quad |y| \le \beta$$

where $\beta = \sqrt{P_k/(P_s D_{kk} + \sigma^2)}$, $a_1 = R_{kk}$, $a_2 = Q_{kk}$ and b_1, b_2, c_1, c_2 can be inferred from (31). For example, when $k = 1$, let $\mathbf{w} = [y, \widetilde{\mathbf{w}}^T]^T$ and

$$\mathbf{Q} = \begin{pmatrix} Q_{11} & \mathbf{l}_1^\dagger \\ \mathbf{l}_1 & \mathbf{Q}_1 \end{pmatrix}, \text{ and } \mathbf{R} = \begin{pmatrix} R_{11} & \mathbf{l}_2^\dagger \\ \mathbf{l}_2 & \mathbf{R}_1 \end{pmatrix}. \tag{46}$$

Then $b_1 = \widetilde{\mathbf{w}}^\dagger \mathbf{l}_1$, $c_1 = \widetilde{\mathbf{w}}^\dagger \mathbf{Q}_1 \widetilde{\mathbf{w}}$, $b_2 = \widetilde{\mathbf{w}}^\dagger \mathbf{l}_2$ and $c_2 = 1 + \widetilde{\mathbf{w}}^\dagger \mathbf{R}_1 \widetilde{\mathbf{w}}$.

For the solution of (45) we give the following theorem, the proof of which can be found in Appendix F.

Theorem 3: *If $a_1/a_2 = b_1/b_2 = c_1/c_2$, the objective in (45) is a constant, and the optimum y, i.e., $y^\star$, is any value satisfying $|y| \leq \beta$. Otherwise: If the equation $(a_1 - ta_2)\beta^2 + 2|b_1 - tb_2|\beta + c_1 - tc_2 = 0$ has a real root, i.e., t_1, such that $|b_1 - t_1 b_2| \geq (t_1 a_2 - a_1)\beta$, then the optimal y is given by*

$$y^\star = \beta e^{-i\theta_1} \tag{47}$$

where $\theta_1 \in (-\pi, \pi]$ is the argument of $b_1 - t_1 b_2$; Else, let t_2 be the root of $|b_1 - tb_2|^2 = (a_1 - ta_2)(c_1 - tc_2)$ such that $|b_1 - t_2 b_2| < (t_2 a_2 - a_1)\beta$, then the optimal y is given by

$$y^\star = \frac{|b_1 - t_2 b_2|}{t_2 a_2 - a_1} e^{-i\theta_2} \tag{48}$$

where θ_2 is the argument of $b_1 - t_2 b_2$.

Remarks: The roots t_1 and t_2 in Theorem 3 can both be obtained in closed form.

For the coordinate descent method, obviously the function value sequence converges. However, in general additional conditions for convergence to a stationary point (or fixed point used in [24, §6]) are needed.

Theorem 4: *The sequence $\{\mathbf{w}^k\}$ generated by Algorithm 1 converges globally to a stationary point.*

Proof: Our proof is based on [22, Proposition 2.7.1] and its proof. Let us denote the objective in (31) as $f(\mathbf{w})$. Let $\bar{\mathbf{w}} = (\bar{w}_1, \cdots, \bar{w}_N)$ be the limit point of the sequence $\{\mathbf{w}^k\}$. We first show

$$f(\bar{\mathbf{w}}) \geq f(w_1, \bar{w}_2, \cdots, \bar{w}_N), \ \forall w_1. \tag{49}$$

If $f(w_1, \bar{w}_2, \cdots, \bar{w}_N)$ is a constant, then obviously (49) holds. If $f(w_1, \bar{w}_2, \cdots, \bar{w}_N)$ is not a constant, to see why, let us assume that (49) does not hold. A verbatim repetition of the proof for [22, Proposition 2.7.1] results in

$$f(\bar{\mathbf{w}}) = f(\bar{w}_1 + \epsilon v_1, \bar{w}_2, \cdots, \bar{w}_N), \ \forall \epsilon \in [0, \epsilon_0] \tag{50}$$

for some $v_1 \neq 0$, $\epsilon_0 > 0$. But from Theorem 3, (50) does not hold for any $v_1 \neq 0$, $\epsilon_0 > 0$ if $f(w_1, \bar{w}_2, \cdots, \bar{w}_N)$ is not a constant. Thus, (49) holds. Similarly, we show

$$f(\bar{\mathbf{w}}) \geq f(\bar{w}_1, \cdots, \bar{w}_{j-1}, w_j, \bar{w}_{j+1}, \cdots, \bar{w}_N), \ \forall w_j \tag{51}$$

for $j = 1, \cdots, N$. This completes the proof. ■

3) p-norm approximation: If the solution $\mathbf{X}^\star$ from the CVX software has rank greater than one, we can also use p-norm approximation plus an augmented Lagrangian method to solve the problem of (40). The convergence of the augmented Lagrangian method can be found in [31]. First, we can show that the problem of (40) is equivalent (*up to scaling*) to

$$\min_{\mathbf{w}} \left[\max_{k \in I} \mathbf{w}^\dagger \mathbf{A}_k \mathbf{w} \right] \tag{52}$$

$$\text{s.t.} \quad \mathbf{w}^\dagger \mathbf{R} \mathbf{w} = 1.$$

To see why this is the case, let $\mathbf{w}^\star$ be the solution to the problem of (52) associated with the optimal objective value $\max_{k \in I} \mathbf{w}^{\star\dagger} \mathbf{A}_k \mathbf{w}^\star = C$. Then, $\frac{1}{\sqrt{C}}\mathbf{w}^\star$ is the solution to the problem of (40) associated with optimal objective value $(\frac{1}{\sqrt{C}}\mathbf{w}^\star)^\dagger \mathbf{R}(\frac{1}{\sqrt{C}}\mathbf{w}^\star) = \frac{1}{C}$. Otherwise, let us assume that the solution to the problem of (40) is $\mathbf{w}'$ with $\mathbf{w}'^\dagger \mathbf{R} \mathbf{w}' = C_1 > \frac{1}{C}$. Thus, $\mathbf{w}'' = \frac{1}{\sqrt{C_1}}\mathbf{w}'$ satisfies $\mathbf{w}''^\dagger \mathbf{R} \mathbf{w}' = 1$ and $\max_{k \in I} \mathbf{w}''^\dagger \mathbf{A}_k \mathbf{w}'' = \frac{1}{C_1} < C$. This contradicts the optimality of $\mathbf{w}^\star$ for the problem of (52). In fact, the two problems are equivalent *up to scaling*.

On denoting $\mathbf{D}_1 = \text{diag}(\sqrt{\frac{P_s D_{11} + \sigma^2}{P_1}}, \cdots, \sqrt{\frac{P_s D_{NN} + \sigma^2}{P_N}})$, $\mathbf{u} = \mathbf{D}_1 \mathbf{w}$, $\mathbf{R}_1 = \mathbf{D}_1^{-1} \mathbf{R} \mathbf{D}_1^{-1}$ and $\mathbf{Q}_1 = \mathbf{D}_1^{-1} \mathbf{Q} \mathbf{D}_1^{-1}$, we rewrite the problem of (52) as

$$\min_{\mathbf{u}} \quad \mathbf{u}^\dagger \mathbf{Q}_1 \mathbf{u} + \|\mathbf{u}\|_\infty^2 \tag{53}$$

$$\text{s.t.} \quad \mathbf{u}^\dagger \mathbf{R}_1 \mathbf{u} = 1$$

where $\|\mathbf{u}\|_\infty = \max_{k \in I} |u_k|$ is the infinity norm. Note that $\|\mathbf{u}\|_\infty$ is not smooth [26]. However, we can approximate $\|\mathbf{u}\|_\infty$ by (smooth) p-norm, i.e., $\|\mathbf{u}\|_p = (\sum_{k \in I} |u_k|^p)^{1/p}$, so that [27], [28]

$$\|\mathbf{u}\|_\infty = \lim_{p \to \infty} \|\mathbf{u}\|_p, \tag{54}$$

$$\text{and } \|\mathbf{u}\|_\infty \leq \|\mathbf{u}\|_p \leq N^{1/p}\|\mathbf{u}\|_\infty. \tag{55}$$

When p is sufficiently large, the approximation is good. In fact, from (55), it is easy to show that given a tolerance ε, the relative error does not exceed ε as long as $p \geq \log N / \log(1 + \varepsilon)$. For example, for $N = 10$, $\varepsilon = 1\%$, we get $p \geq 232$; for $N = 40$, $\varepsilon = 0.5\%$, we get $p \geq 740$.

Now, using $\|\mathbf{u}\|_{2p}^2$, $p \geq 1$ as a smooth approximation to $\|\mathbf{u}\|_\infty^2$, we turn to solve the following

$$\min_{\mathbf{u}} \quad \mathbf{u}^\dagger \mathbf{Q}_1 \mathbf{u} + \|\mathbf{u}\|_{2p}^2 \tag{56}$$

$$\text{s.t.} \quad \mathbf{u}^\dagger \mathbf{R}_1 \mathbf{u} = 1.$$

We use the augmented Lagrangian method [31, §17] to solve the problem of (56). Since the augmented Lagrangian method was originally proposed for real variables, we first modify our problem as follows.

Define [25]

$$\mathbf{z} = \begin{pmatrix} \mathrm{Re}(\mathbf{u}) \\ \mathrm{Im}(\mathbf{u}) \end{pmatrix}, \tag{57}$$

$$\mathbf{F} = \begin{pmatrix} \mathrm{Re}(\mathbf{Q}_1) & -\mathrm{Im}(\mathbf{Q}_1) \\ \mathrm{Im}(\mathbf{Q}_1) & \mathrm{Re}(\mathbf{Q}_1) \end{pmatrix}, \tag{58}$$

$$\mathbf{K} = \begin{pmatrix} \mathrm{Re}(\mathbf{R}_1) & -\mathrm{Im}(\mathbf{R}_1) \\ \mathrm{Im}(\mathbf{R}_1) & \mathrm{Re}(\mathbf{R}_1) \end{pmatrix}, \tag{59}$$

$$\text{and } \widetilde{\mathbf{J}}_k = \begin{pmatrix} \mathbf{J}_k & 0 \\ 0 & \mathbf{J}_k \end{pmatrix} \tag{60}$$

where $\mathbf{J}_k$ is defined in Lemma 3, and $\mathrm{Re}(\cdot)$, $\mathrm{Im}(\cdot)$ denote the real and imaginary part respectively, then

$$\mathbf{u}^\dagger \mathbf{Q}_1 \mathbf{u} = \mathbf{z}^T \mathbf{F} \mathbf{z}, \tag{61}$$

$$\mathbf{u}^\dagger \mathbf{R}_1 \mathbf{u} = \mathbf{z}^T \mathbf{K} \mathbf{z}, \tag{62}$$

$$\text{and } \|\mathbf{u}\|_{2p}^2 = \left(\sum_{k \in I} (\mathbf{z}^T \widetilde{\mathbf{J}}_k \mathbf{z})^p \right)^{1/p}. \tag{63}$$

With these, we rewrite the problem of (56) as

$$\min_{\mathbf{z}} \ \mathbf{z}^T \mathbf{F} \mathbf{z} + \phi_p(\mathbf{z}) \tag{64}$$

$$\text{s.t.} \quad \mathbf{z}^T \mathbf{K} \mathbf{z} - 1 = 0$$

where $\phi_p(\mathbf{z})$ is defined as the right hand side of (63).

Now we can apply the augmented Lagrangian method, given by

$$L(\mathbf{z}; \lambda; \mu) = \mathbf{z}^T \mathbf{F} \mathbf{z} + \phi_p(\mathbf{z}) - \lambda(\mathbf{z}^T \mathbf{K} \mathbf{z} - 1) + \frac{1}{2\mu}(\mathbf{z}^T \mathbf{K} \mathbf{z} - 1)^2 \tag{65}$$

where λ is the Lagrangian multiplier, and the fourth term in the right hand side of (65) is the penalty function. The algorithm is described as follows:

1) *Choose an initial estimate $\lambda^{(0)}$ of $\lambda^\star$ and $\mu = 0.001$. Set $k = 1$.*
2) *Determine $\mathbf{z}_k$ to be a minimizer of $L(\mathbf{z}; \lambda^{(k-1)}; \mu)$;*
3) *Compute $\lambda^{(k)} = \lambda^{(k-1)} - (\mathbf{z}_k^T \mathbf{K} \mathbf{z}_k - 1)/\mu$;*
4) *If a convergence test is satisfied, stop;*
5) *$k = k + 1$; Go to 2).*

For Step 2, we use the backtracking line search Newton's method with Hessian modification [31, Algorithm 3.2]. The iteration expression is

$$\mathbf{z}^{(i+1)} = \mathbf{z}^{(i)} + \alpha \mathbf{p}_i \tag{66}$$

$$\text{with } \mathbf{p}_i = -(\nabla^2 L + \beta \mathbf{I})^{-1} \nabla L \tag{67}$$

where β is chosen such that $\nabla^2 L + \beta \mathbf{I}$ is positive definite, e.g., $\beta = \lambda_{\min}(\nabla^2 L) + 10^{-6}$, and α is the step size determined by the backtracking line search described as follows [31, Algorithm 3.1]:

a) *Set $\alpha = 1$, $c_1 = 10^{-4}$, $\rho = 0.5$;*

b) *Repeat: if $L(\mathbf{z}^{(i)} + \alpha \mathbf{p}_i; \cdot; \cdot) > L(\mathbf{z}^{(i)}; \cdot; \cdot) + c_1 \alpha \mathbf{p}_i^T \nabla L$, then $\alpha \leftarrow \rho \alpha$.*

In the algorithm, we need to calculate ∇L and $\nabla^2 L$ given by

$$\nabla L = 2\mathbf{F}\mathbf{z} + \nabla \phi_p - 2\lambda \mathbf{K}\mathbf{z} + \frac{2}{\mu}(\mathbf{z}^T \mathbf{K}\mathbf{z} - 1)\mathbf{K}\mathbf{z} \tag{68}$$

$$\text{and } \nabla^2 L = 2\mathbf{F} + \nabla^2 \phi_p - 2\lambda \mathbf{K} + \frac{2}{\mu}(\mathbf{z}^T \mathbf{K}\mathbf{z} - 1)\mathbf{K} + \frac{4}{\mu}\mathbf{K}\mathbf{z}\mathbf{z}^T \mathbf{K}. \tag{69}$$

The calculation of $\nabla \phi_p$ and $\nabla^2 \phi_p$ is given in Appendix E.

Remarks: For the initial estimate $\lambda^{(0)}$ of $\lambda^\star$, note that when $p = 1$, then $\phi_p(\mathbf{z}) = \mathbf{z}^T \mathbf{z}$ and $\lambda^\star$ can be expressed in closed form as $\lambda_{\min}(\mathbf{K}^{-1/2}\mathbf{F}\mathbf{K}^{-1/2} + \mathbf{K}^{-1})$. We choose this as $\lambda^{(0)}$.

V. NUMERICAL RESULTS

In this section, we provide some examples illustrating the proposed algorithms. For more simulation results on beamforming itself the reader can refer to [10]. We consider a channel model as follows:

$$f_i = \bar{f}_i + \sqrt{\psi_i}\, \tilde{f}_i \tag{70}$$

$$\text{and } g_j = \bar{g}_j + \sqrt{\varphi_j}\, \tilde{g}_j \tag{71}$$

where $\bar{f}_i$ and $\bar{g}_j$ are means, ψ_i and φ_j are variances, $\tilde{f}_i$ and $\tilde{g}_j$ both are zero-mean random variables with unit variance. We assume that $\tilde{f}_i$, $\tilde{f}_j$, $\tilde{g}_i$ and $\tilde{g}_j$, $\forall i \neq j$ are independent. $\bar{f}_i = 0$ corresponds to the scenario in which there is no line-of-sight (LOS) path (Rayleigh fading), while $\bar{f}_i \neq 0$ corresponds the scenario in which there is an LOS path (Rician fading). Thus, the matrices $\mathbf{D}$, $\mathbf{R}$ and $\mathbf{Q}$ are given by

$$\mathbf{D} = \text{diag}(|\bar{f}_1|^2 + \psi_1, \cdots, |\bar{f}_N|^2 + \psi_N),$$

$$Q_{ij} = \bar{g}_i \bar{g}_j^* + \sqrt{\varphi_i \varphi_j}\, \delta_{ij},$$

$$\text{and } R_{ij} = (\bar{f}_i \bar{f}_j^* + \sqrt{\psi_i \psi_j}\, \delta_{ij})(\bar{g}_i \bar{g}_j^* + \sqrt{\varphi_i \varphi_j}\, \delta_{ij})$$

where δ_{ij} is the Kronecker function.

A. SNR maximization under total power constraint

Please refer to §III for details. First, we consider a network consisting of $N = 6$ relays with channel parameters given by

$$\bar{f}_1 = 0.2202 + 0.8130i, \ \bar{f}_2 = -0.4075 - 0.7644i,$$
$$\bar{f}_3 = -2.0107 + 0.4016i, \ \bar{f}_4 = -0.4503 + 0.0678i,$$
$$\bar{f}_5 = 0.8588 - 0.1130i, \ \bar{f}_6 = -0.1219 + 0.4260i;$$

$$\psi_1 = 3.8042, \ \psi_2 = 2.6326, \ \psi_3 = 4.7590,$$
$$\psi_4 = 0.4989, \ \psi_5 = 1.2576, \ \psi_6 = 1.2484;$$

$$\bar{g}_1 = -0.3726 + 0.8007i, \ \bar{g}_2 = 0.4592 - 0.2045i,$$
$$\bar{g}_3 = -0.8769 + 0.4671i, \ \bar{g}_4 = -0.9270 + 0.5430i,$$
$$\bar{g}_5 = -0.0063 - 0.4977i, \ \bar{g}_6 = -0.7783 - 0.7712i;$$

$$\varphi_1 = 0.3913, \ \varphi_2 = 0.4791, \ \varphi_3 = 0.0865,$$
$$\varphi_4 = 2.7813, \ \varphi_5 = 4.8960, \ \varphi_6 = 4.6789.$$

Fig. 2 plots $\lambda_{\min}(\mathbf{G}(x))$ for x in $[x_l, x_u] = [0.1711, 0.7077]$ with 100 uniform points. Using Newton's method for starting points $x_0 = x_l$, $x_0 = x_u$, the convergent points $(0.2156, 1.2191)$, $(0.5844, 1.2694)$ are also plotted in Fig. 2. Fig. 3 plots the iteration process under the stopping test: $\left|\frac{x_{k+1}-x_k}{x_k}\right| < 10^{-3}$ and $\left|\frac{\mathrm{d}}{\mathrm{d}x}\lambda_{\min}(\mathbf{G}(x))\right| < 10^{-3}$. It can be seen from Fig. 3 that Newton's method converges rapidly.

Second, we consider a network consisting of $N = 6$ relays with channel parameters given by

$$\bar{f}_1 = -0.4751 + 0.7340i, \ \bar{f}_2 = -0.0449 - 0.4609i,$$
$$\bar{f}_3 = 0.0239 - 1.5154i, \ \bar{f}_4 = 0.5130 - 0.1755i,$$
$$\bar{f}_5 = -0.2017 + 0.6717i, \ \bar{f}_6 = 1.0134 - 0.1985i;$$

$$\psi_1 = 2.4707, \ \psi_2 = 3.9193, \ \psi_3 = 2.4121,$$
$$\psi_4 = 3.8879, \ \psi_5 = 1.2050, \ \psi_6 = 3.0901;$$

$$\bar{g}_1 = 0.5360 - 1.2932i, \ \bar{g}_2 = 1.7471 - 0.8914i,$$
$$\bar{g}_3 = 0.0955 - 0.1577i, \ \bar{g}_4 = -0.6795 + 0.2479i,$$
$$\bar{g}_5 = 0.5815 + 0.5039i, \ \bar{g}_6 = -0.3090 + 0.8413i;$$

$$\varphi_1 = 3.9655, \ \varphi_2 = 0.2693, \ \varphi_3 = 0.9205,$$
$$\varphi_4 = 0.5567, \ \varphi_5 = 3.3901, \ \varphi_6 = 2.9367.$$

Fig. 4 plots $\lambda_{\min}(\mathbf{G}(x))$ for x in $[x_l, x_u] = [0.2754, 0.6392]$ with 100 uniform points. Using Newton's

method for starting points $x_0 = x_l$, $x_0 = x_u$, the same convergent point $(0.4087, 0.6060)$ is also plotted in Fig. 4. Fig. 5 plots the iteration process under the stopping test: $|\frac{x_{k+1}-x_k}{x_k}| < 10^{-3}$ and $|\frac{\mathrm{d}}{\mathrm{d}x}\lambda_{\min}(\mathbf{G}(x))| < 10^{-3}$. It can be seen from Fig. 5 that Newton's method converges rapidly.

B. SNR maximization under individual relay power constraints

Please refer to §IV for details. In [10], the authors stated that, based on their simulations, the SDP relaxation always has a rank one solution. However, no analytic proof was provided for that claim. However, although a rank one solution often occurs, for general $\mathbf{R}$ and $\mathbf{Q}$ the SDP relaxation does not necessarily have a rank one solution. This can be seen in the following examples, for which the SDP relaxation has a rank greater than one.

First, we consider a network consisting of $N = 4$ relays with

$$\mathbf{Q} = \begin{pmatrix} 2.1 & .73+.75\text{i} & .43+1.1\text{i} & .70-.33\text{i} \\ .73-.75\text{i} & 1.6 & -.20+.18\text{i} & .57-.71\text{i} \\ .43-1.1\text{i} & -.20-.18\text{i} & 2 & -.52-.45\text{i} \\ .70+.33\text{i} & .57+.71\text{i} & -.52+.45\text{i} & .98 \end{pmatrix} \tag{72}$$

$$\text{and } \mathbf{R} = \begin{pmatrix} 1.6 & -.74-.16\text{i} & .084-.57\text{i} & -.19+.67\text{i} \\ -.74+.16\text{i} & 1.1 & -.88+.31\text{i} & -.44-.24\text{i} \\ .084+.57\text{i} & -.88-.31\text{i} & 2 & .20-.14\text{i} \\ -.19-.67\text{i} & -.44+.24\text{i} & .20+.14\text{i} & 1.5 \end{pmatrix}. \tag{73}$$

For simplicity, we let $\mathbf{D}_1 = \mathbf{I}$ (defined in §IV-B3) and denote the SDP relaxation solution from CVX software by $\mathbf{X}^\star$. The eigenvalues of $\mathbf{X}^\star$ are

$$0.0000, 0.0000, 0.2064, 1.8148.$$

Thus, $\mathbf{X}^\star$ has rank two rather than rank one and can be eigen-decomposed as $0.2064\mathbf{u}_1\mathbf{u}_1^\dagger + 1.8148\mathbf{u}_2\mathbf{u}_2^\dagger$ where $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors associated with the eigenvalues 0.2064 and 1.8148 respectively. We obtain the objective values of the problem of (40) for SDP relaxation, GRP from [10], coordinate descent method from §IV-B2, and p-norm approximation from §IV-B3 (starting points: $\sqrt{1.8148}\,\mathbf{u}_2$ or some samples from $\mathcal{CN}(0, \mathbf{X}^\star)$) as, respectively:

SDP relaxation: 3.74112

GRP (10^6 samples from $\mathcal{CN}(0, \mathbf{X}^\star)$): 3.6970

Coordinate descent method: 3.7076

p-norm approximation: 3.7069 *(p = 1024)*

It can be seen that the objective values from GRP, coordinate descent method and p-norm approximation are close to each other (with a difference $< 0.3\%$) and close to the SDP relaxation solution (with a difference $< 2\%$). It can be seen that: although the GRP attains a close performance compared with the other two methods, it is time consuming in the sense that it needs much more time (processes 10^6 samples from $\mathcal{CN}(0, \mathbf{X}^\star)$). The augmented Lagrangian $L(\mathbf{z}; \lambda; \mu)$ (defined in (65)) during the iteration is plotted in Fig. 6. The objective value during the iteration for the coordinate descent method is plotted in Fig. 7. It can be seen that for these two algorithms the iteration converges rapidly.

Second, we consider a network consisting of $N = 6$ relays with

$$
\mathbf{Q} = \begin{pmatrix}
.778 & -.658 - .646\mathrm{i} & .135 + .269\mathrm{i} & -.273 + .005\mathrm{i} & .088 - .261\mathrm{i} & -.021 - .013\mathrm{i} \\
-.658 + .646\mathrm{i} & 2.20 & -.379 - 1.14\mathrm{i} & .253 - .872\mathrm{i} & -.337 + 1.02\mathrm{i} & .444 - .035\mathrm{i} \\
.135 - .269\mathrm{i} & -.379 + 1.14\mathrm{i} & 2. & .689 + .298\mathrm{i} & -.547 - .160\mathrm{i} & .373 + .693\mathrm{i} \\
-.273 - .005\mathrm{i} & .253 + .872\mathrm{i} & .689 - .298\mathrm{i} & 1. & -.655 + .192\mathrm{i} & .132 - .107\mathrm{i} \\
.088 + .261\mathrm{i} & -.337 - 1.02\mathrm{i} & -.547 + .160\mathrm{i} & -.655 - .192\mathrm{i} & 2.40 & -.721 - .276\mathrm{i} \\
-.021 + .013\mathrm{i} & .444 + .035\mathrm{i} & .373 - .693\mathrm{i} & .132 + .107\mathrm{i} & -.721 + .276\mathrm{i} & 1.09
\end{pmatrix}
\tag{74}
$$

and

$$
\mathbf{R} = \begin{pmatrix}
3.44 & -.263 + .054\mathrm{i} & .572 + 1.73\mathrm{i} & .490 - .276\mathrm{i} & -.613 - 1.62\mathrm{i} & -.014 + .375\mathrm{i} \\
-.263 - .054\mathrm{i} & 3.09 & -.342 - 1.49\mathrm{i} & .926 + 1.13\mathrm{i} & -.282 - .713\mathrm{i} & -.211 + .911\mathrm{i} \\
.572 - 1.73\mathrm{i} & -.342 + 1.49\mathrm{i} & 2.70 & -.493 + .865\mathrm{i} & -.396 + .826\mathrm{i} & .149 - .836\mathrm{i} \\
.490 + .276\mathrm{i} & .926 - 1.13\mathrm{i} & -.493 - .865\mathrm{i} & 3.09 & .541 + .330\mathrm{i} & -.552 - .221\mathrm{i} \\
-.613 + 1.62\mathrm{i} & -.282 + .713\mathrm{i} & -.396 - .826\mathrm{i} & .541 - .330\mathrm{i} & 2.75 & -.442 - .352\mathrm{i} \\
-.014 - .375\mathrm{i} & -.211 - .911\mathrm{i} & .149 + .836\mathrm{i} & -.552 + .221\mathrm{i} & -.442 + .352\mathrm{i} & 2.08
\end{pmatrix}.
\tag{75}
$$

The eigenvalues of $\mathbf{X}^\star$ are

$$0.0000, 0.0000, 0.0000, 0.0000, 0.8369, 2.3774.$$

Thus, $\mathbf{X}^\star$ has rank two rather than rank one and can be eigen-decomposed as $0.8369\mathbf{u}_1\mathbf{u}_1^\dagger + 2.3774\mathbf{u}_2\mathbf{u}_2^\dagger$ where $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors associated with the eigenvalues 0.8369 and 2.3774 respectively. We obtain the objective values of the problem of (40) for SDP relaxation, GRP, coordinate descent method, and p-norm approximation (starting points: $\sqrt{2.3774}\,\mathbf{u}_2$ or some samples from $\mathcal{CN}(0, \mathbf{X}^\star)$) as, respectively:

SDP relaxation: 9.33816

GRP (10^6 samples from $\mathcal{CN}(0, \mathbf{X}^\star)$): 8.1472

Coordinate descent method: 8.9428

p-norm approximation: 8.9409 *(p = 1024)*

It can be seen that: the objective value from GRP has a significant ($> 10\%$) difference from the SDP relaxation solution; p-norm approximation and coordinate descent method attain objective values close to each other; the improvement of objective value from coordinate descent method is $(8.9428 - 8.1472)/8.1472 = 9.77\%$ compared with GRP. It can be seen that for this example, GRP is time consuming and ineffective in the sense that it needs more time (processes 10^6 samples from $\mathcal{CN}(0, \mathbf{X}^\star)$) but attains worse performance compared with the other two algorithms. The augmented Lagrangian $L(\mathbf{z}; \lambda; \mu)$ (defined in (65)) during the iteration is plotted in Fig. 8. The objective value during the iteration for the coordinate descent method is plotted in Fig. 9. We can see that for these two algorithms the iteration converges rapidly.

VI. CONCLUSION

We have investigated the problem of cooperative beamforming under the assumption that the second-order statistics of the channel state information (CSI) are available. Beamforming weights are determined so that the SNR at the destination is maximized subject to two kinds of power constraints. The first kind of power constraint is a constraint on the total power, i.e., source plus relay power. The second kind of power constraint is a constraint on each relay's transmit power. For uncorrelated Rayleigh fading scenario, we attained the exact solution. For generic fading scenario, we focused on the case in which the SDP relaxation does not produce a rank-one solution and proposed two methods to solve it. The numerical simulations suggest that the proposed methods are more effective than the method of [10].

APPENDIX A
PROOF OF LEMMA 1

Let $(P_s^\circ, \mathbf{w}^\circ)$ be the solution to the problem of (11). We can show that $P_s^\circ + P_s^\circ \mathbf{w}^{\circ\dagger} \mathbf{D} \mathbf{w}^\circ + \sigma^2 \mathbf{w}^{\circ\dagger} \mathbf{w}^\circ = P_0$. Otherwise, let us assume that $P_s^\circ + P_s^\circ \mathbf{w}^{\circ\dagger} \mathbf{D} \mathbf{w}^\circ + \sigma^2 \mathbf{w}^{\circ\dagger} \mathbf{w}^\circ < P_0$. Let $\beta = (P_0 - P_s^\circ)/(P_s^\circ \mathbf{w}^{\circ\dagger} \mathbf{D} \mathbf{w}^\circ + \sigma^2 \mathbf{w}^{\circ\dagger} \mathbf{w}^\circ)$, and hence $\beta > 1$. It is easy to verify that $(P_s^\circ, \sqrt{\beta} \, \mathbf{w}^\circ)$ satisfies the constraint but results in a larger objective value. This violates the optimality of $(P_s^\circ, \mathbf{w}^\circ)$. With this, the problem of (11) is equivalent to

$$\max_{P_s, \mathbf{w}} \frac{P_s}{\sigma^2} \frac{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}}{1 + \mathbf{w}^\dagger \mathbf{Q} \mathbf{w}} \tag{76}$$

$$\text{s.t.} \quad P_s + P_s \mathbf{w}^\dagger \mathbf{D} \mathbf{w} + \sigma^2 \mathbf{w}^\dagger \mathbf{w} = P_0.$$

It follows from the constraint in (76) that

$$1 = \frac{\mathbf{w}^\dagger (P_s \mathbf{D} + \sigma^2 \mathbf{I})\mathbf{w}}{P_0 - P_s}. \tag{77}$$

By using (77), we rewrite the problem of (76) as

$$\max_{P_s, \mathbf{w}} \frac{P_s}{\sigma^2} \frac{(P_0 - P_s)\mathbf{w}^\dagger \mathbf{R}\mathbf{w}}{\mathbf{w}^\dagger [P_s \mathbf{D} + \sigma^2 \mathbf{I} + (P_0 - P_s)\mathbf{Q}]\mathbf{w}} \tag{78}$$

$$\text{s.t.} \quad P_s + P_s \mathbf{w}^\dagger \mathbf{D}\mathbf{w} + \sigma^2 \mathbf{w}^\dagger \mathbf{w} = P_0.$$

Note that the objective in (78) has the same value at $\mathbf{w}$ and $\beta_1 \mathbf{w}$, $\forall \beta_1 \neq 0$, $\mathbf{w} \neq 0$. Thus, the problem of (78) is equivalent to

$$\max_{P_s, \mathbf{w}} \frac{P_s}{\sigma^2} \frac{(P_0 - P_s)\mathbf{w}^\dagger \mathbf{R}\mathbf{w}}{\mathbf{w}^\dagger [P_s \mathbf{D} + \sigma^2 \mathbf{I} + (P_0 - P_s)\mathbf{Q}]\mathbf{w}} \tag{79}$$

$$\text{s.t.} \quad 0 \leq P_s \leq P_0, \ \mathbf{w} \neq 0.$$

Further, we rewrite

$$\sigma^2 \mathbf{I} = P_s \frac{\sigma^2}{P_0} \mathbf{I} + (P_0 - P_s)\frac{\sigma^2}{P_0} \mathbf{I}, \tag{80}$$

which enables us to write

$$P_s \mathbf{D} + \sigma^2 \mathbf{I} + (P_0 - P_s)\mathbf{Q} = P_s \left(\mathbf{D} + \frac{\sigma^2}{P_0} \right) + (P_0 - P_s) \left(\mathbf{Q} + \frac{\sigma^2}{P_0} \right). \tag{81}$$

With this, by using the fact that for $\mathbf{C}_1 \succ 0$ and $\mathbf{C}_2 \succ 0$ [29, p. 549]

$$\frac{1}{\lambda_{\min}(\mathbf{C}_1^{-1/2}\mathbf{C}_2\mathbf{C}_1^{-1/2})} = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^\dagger \mathbf{C}_1 \mathbf{x}}{\mathbf{x}^\dagger \mathbf{C}_2 \mathbf{x}}, \tag{82}$$

the problem of (79) is equivalent to the problem of (12).

APPENDIX B

PROOF OF LEMMA 2

Obviously, neither $x = 0$ nor $x = 1$ is the solution to the problem of (16). Define the function

$$K(x) = \frac{x\mathbf{S}_1 + (1-x)\mathbf{S}_2}{x(1-x)} = \frac{\mathbf{S}_1}{1-x} + \frac{\mathbf{S}_2}{x}, \ x \in (0,1). \tag{83}$$

Let $x \in (0,1)$ and $\Delta x \neq 0$ be an increment such that $x + \Delta x \in (0,1)$. Using Taylor series expansion, we approximate

$$\frac{1}{1 - (x + \Delta x)} = \frac{1}{1-x} + \frac{\Delta x}{(1-x)^2} + \frac{(\Delta x)^2}{(1-\xi_1)^3}, \tag{84}$$

$$\text{and } \frac{1}{x + \Delta x} = \frac{1}{x} - \frac{\Delta x}{x^2} + \frac{(\Delta x)^2}{\xi_2^3} \tag{85}$$

where ξ_1 and ξ_2 both lie between x and $x + \Delta x$. From (83), (84) and (85), we get

$$K(x + \Delta x) = K(x) + \Delta x \left(\frac{1}{(1-x)^2} \mathbf{S}_1 - \frac{1}{x^2} \mathbf{S}_2 \right) + (\Delta x)^2 \left(\frac{1}{(1-\xi_1)^3} \mathbf{S}_1 + \frac{1}{\xi_2^3} \mathbf{S}_2 \right). \tag{86}$$

Note that the third term in the right hand side of (86) is positive definite. By using the facts that [29, p. 549]

$$c\mathbf{S}_1 \preceq \mathbf{S}_2, \text{ and } d\mathbf{S}_1 \succeq \mathbf{S}_2 \tag{87}$$

it is not difficult to prove that

$$\frac{1}{(1-x)^2} \mathbf{S}_1 - \frac{1}{x^2} \mathbf{S}_2 \begin{cases} \preceq 0 & x \in (0, x_l] \\ \succeq 0 & x \in [x_u, 1). \end{cases} \tag{88}$$

With these, we know that: if $x \in (0, x_l]$ and $\Delta x < 0$, then $K(x + \Delta x) \succ K(x)$ and it follows from Weyl's inequality [33, p. 181] that $\lambda_{\min}(K(x+\Delta x)) > \lambda_{\min}(K(x))$; if $x \in [x_u, 1)$ and $\Delta x > 0$, similarly, $\lambda_{\min}(K(x + \Delta x)) > \lambda_{\min}(K(x))$. This completes the proof.

APPENDIX C

PROOF OF THEOREM 1

If $k_0 = 1$, then $0 < t^\star < \tilde{t}_1$, and

$$\frac{P_s}{\sigma^2} \tilde{r}_k - t^\star \tilde{q}_k > 0, \ k = 1, \cdots, N. \tag{89}$$

Thus, $F(t) = 0$ in (38) leads to

$$- t^\star + \sum_{k=1}^{N} \frac{\tilde{P}_k}{P_s \tilde{D}_{kk} + \sigma^2} \left(\frac{P_s}{\sigma^2} \tilde{r}_k - t^\star \tilde{q}_k \right) = 0. \tag{90}$$

The desired result can be obtained from the above equation.

If $k_0 > 1$, then $\tilde{t}_{k_0-1} < t^\star < \tilde{t}_{k_0}$, and

$$\frac{P_s}{\sigma^2} \tilde{r}_k - t^\star \tilde{q}_k \begin{cases} > 0 & k = k_0, \cdots, N \\ < 0 & k = 1, \cdots, k_0 - 1. \end{cases} \tag{91}$$

Thus, $F(t) = 0$ in (38) leads to

$$- t^\star + \sum_{k=k_0}^{N} \frac{\tilde{P}_k}{P_s \tilde{D}_{kk} + \sigma^2} \left(\frac{P_s}{\sigma^2} \tilde{r}_k - t^\star \tilde{q}_k \right) = 0. \tag{92}$$

The desired result can be obtained from the above equation.

APPENDIX D

PROOF OF LEMMA 3

Note that the constraints in (31) can be rewritten as

$$\frac{P_s D_{kk} + \sigma^2}{P_k} \mathbf{w}^\dagger \mathbf{J}_k \mathbf{w} \le 1, \ k \in I. \tag{93}$$

Note that the objective in (31) has a greater value at $\alpha \mathbf{w}$ than that at $\mathbf{w}$, $\forall \alpha > 1$, $\mathbf{w} \neq 0$. Thus, there exists $j \in I$ such that $((P_s D_{jj} + \sigma^2)/P_j)\mathbf{w}^\dagger \mathbf{J}_j \mathbf{w} = 1$, i.e., at least one constraint is active. With this, the constraints in (31) can be rewritten as

$$\max_{k \in I} \frac{P_s D_{kk} + \sigma^2}{P_k} \mathbf{w}^\dagger \mathbf{J}_k \mathbf{w} = 1. \tag{94}$$

By using (94), we rewrite the problem of (31) as

$$\max_{\mathbf{w}} \ \frac{P_s}{\sigma^2} \frac{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}}{\max_{k \in I} \mathbf{w}^\dagger \mathbf{A}_k \mathbf{w}} \tag{95}$$

$$\text{s.t.} \quad \max_{k \in I} \frac{P_s D_{kk} + \sigma^2}{P_k} \mathbf{w}^\dagger \mathbf{J}_k \mathbf{w} = 1.$$

Note that the objective in (95) has the same value at $\mathbf{w}$ and $\beta_1 \mathbf{w}$, $\forall \beta_1 \neq 0$, $\mathbf{w} \neq 0$. Thus, the problem of (95) is equivalent to

$$\max_{\mathbf{w}} \ \frac{P_s}{\sigma^2} \frac{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}}{\max_{k \in I} \mathbf{w}^\dagger \mathbf{A}_k \mathbf{w}} \tag{96}$$

$$\text{s.t.} \quad \mathbf{w} \neq 0.$$

Similarly, the problem of (96) is equivalent to

$$\max_{\mathbf{w}} \ \frac{P_s}{\sigma^2} \mathbf{w}^\dagger \mathbf{R} \mathbf{w} \tag{97}$$

$$\text{s.t.} \quad \max_{k \in I} \mathbf{w}^\dagger \mathbf{A}_k \mathbf{w} = 1.$$

Obviously, the problem of (97) is equivalent to the problem of (40). This completes the proof.

APPENDIX E

CALCULATION OF $\nabla \phi_p$ AND $\nabla^2 \phi_p$

We have

$$\nabla \phi_p = \frac{\partial \phi_p}{\partial \mathbf{z}} = 2 \sum_{k \in I} \left(\frac{\mathbf{z}^T \widetilde{\mathbf{J}}_k \mathbf{z}}{\phi_p(\mathbf{z})} \right)^{p-1} \widetilde{\mathbf{J}}_k \mathbf{z} \tag{98}$$

and

$$\nabla^2 \phi_p = \frac{\partial}{\partial \mathbf{z}^T} \left(\frac{\partial \phi_p}{\partial \mathbf{z}} \right)$$

$$= 2 \sum_{k \in I} \left(\frac{\mathbf{z}^T \widetilde{\mathbf{J}}_k \mathbf{z}}{\phi_p(\mathbf{z})} \right)^{p-1} \widetilde{\mathbf{J}}_k + \frac{1-p}{\phi_p(\mathbf{z})} \left(\frac{\partial \phi_p}{\partial \mathbf{z}} \right) \left(\frac{\partial \phi_p}{\partial \mathbf{z}} \right)^T + \frac{4(p-1)}{\phi_p(\mathbf{z})} \sum_{k \in I} \left(\frac{\mathbf{z}^T \widetilde{\mathbf{J}}_k \mathbf{z}}{\phi_p(\mathbf{z})} \right)^{p-2} \widetilde{\mathbf{J}}_k \mathbf{z} \mathbf{z}^T \widetilde{\mathbf{J}}_k. \tag{99}$$

APPENDIX F

PROOF OF THEOREM 3

By using the Dinkelbach-type method [18] (cf. §IV-A), we introduce the function

$$F(t) = \max_y \ f(t, y) \tag{100}$$

$$\text{s.t.} \quad |y| \leq \beta$$

where

$$f(t, y) = a_1 |y|^2 + b_1 y + b_1^* y^* + c_1 - t(a_2 |y|^2 + b_2 y + b_2^* y^* + c_2). \tag{101}$$

Similarly to Property 1 in §IV-A, $F(t)$ is a strictly decreasing function and the equation $F(t) = 0$ has a unique root $t^\star$. The optimal $y^\star$ associated with $F(t^\star)$ is also the solution for the problem of (45) with the optimal objective value $t^\star$.

To obtain the expression of $F(t)$, we denote $y = |y|e^{i\theta}$ and write

$$f(t, y) = (a_1 - ta_2)|y|^2 + \left[(b_1 - tb_2)e^{i\theta} + (b_1 - tb_2)^* e^{-i\theta} \right] |y| + c_1 - tc_2$$

$$\leq (a_1 - ta_2)|y|^2 + 2|b_1 - tb_2||y| + c_1 - tc_2. \tag{102}$$

The equality in (102) occurs when the argument of $b_1 - tb_2$ equals $-\theta$ (if $b_1 - tb_2 = 0$, then θ is arbitrary). With this, we let $r = |y|$ and write

$$F(t) = c_1 - tc_2 + \max_r \ (a_1 - ta_2)r^2 + 2|b_1 - tb_2|r \tag{103}$$

$$\text{s.t.} \quad 0 \leq r \leq \beta.$$

Further, it is easy to get:

1) When $a_1 - ta_2 > 0$, i.e., $t < a_1/a_2$, the optimal r is β (unique), and we wirte

$$F(t) = (a_1 - ta_2)\beta^2 + 2|b_1 - tb_2|\beta + c_1 - tc_2. \tag{104}$$

2) When $a_1 - ta_2 < 0$, i.e., $t > a_1/a_2$, the optimal r is given by (unique)

$$r^\star = \min\left\{\frac{|b_1 - tb_2|}{ta_2 - a_1}, \ \beta\right\}, \tag{105}$$

and we write: If $t > a_1/a_2$ and $|b_1 - tb_2| \geq (ta_2 - a_1)\beta$, then

$$F(t) = (a_1 - ta_2)\beta^2 + 2|b_1 - tb_2|\beta + c_1 - tc_2; \tag{106}$$

If $t > a_1/a_2$ and $|b_1 - tb_2| < (ta_2 - a_1)\beta$, then

$$F(t) = \frac{|b_1 - tb_2|^2}{ta_2 - a_1} + c_1 - tc_2. \tag{107}$$

3) When $a_1 - ta_2 = 0$, i.e., $t = a_1/a_2$, we know: If $b_1 - tb_2 \neq 0$ (i.e., $b_1/b_2 \neq a_1/a_2$), the optimal r is β (unique), and

$$F(a_1/a_2) = 2|b_1 - (a_1/a_2)b_2|\beta + c_1 - (a_1/a_2)c_2; \tag{108}$$

If $b_1 - tb_2 = 0$ (i.e., $b_1/b_2 = a_1/a_2$), the optimal r is arbitrary in $[0, \beta]$, and

$$F(a_1/a_2) = c_1 - (a_1/a_2)c_2. \tag{109}$$

Recall that $F(t)$ is a strictly decreasing function and the equation $F(t) = 0$ has a unique root $t^\star$. Thus, one and only one of the equations (104), (106), (107), (108), (109) satisfies $F(t^\star) = 0$.

Based on the analysis above, it is not difficult to obtain the desired result.

REFERENCES

[1] H. Ochiai, P. Mitran, H. V. Poor and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4110-4124, Nov. 2005.

[2] S. Pun, D. R. Brown III, and H. V. Poor, "Opportunistic collaborative beamforming with one-bit feedback," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2629-2641, May. 2009.

[3] R. Mudumbai, U. Madhow, D. R. Brown III, and H. V. Poor, "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102-110, Feb. 2009.

[4] J.N. Laneman, D.N.C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[5] M. Janani, A. Hedayat, T.E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, pp. 362-371, Feb. 2004.

[6] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorem for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037-3063, Sep. 2005.

[7] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Honolulu, HI, pp. III-473-III-476, Apr. 15-21, 2007.

[8] L. Dong, A.P. Petropulu, and H.V. Poor, "Weighted cross-layer cooperative beamforming for wireless networks," *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3240-3252, Aug. 2009.

[9] V.H. Nassab, S. Shahbazpanahi, A. Grami, "Optimal distributed beamforming for two-way relay networks," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1238-1250, Mar. 2010.

[10] V.H. Nassab, S. Shahbazpanahi, A. Grami, and Z.Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4306-4316, Sep. 2008.

[11] S. Fazeli-Dehkordy, S. Shahbazpanahi, S. Gazor, "Multiple peer-to-peer communications using a network of relays," *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3053-3062, Aug. 2009.

[12] E. Koyuncu, Y. Jing, H. Jafarkhani, "Distributed beamforming in wireless relay networks with quantized feedback," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1429-1439, Oct. 2008.

[13] L. Dong, Z. Han, A. Petropulu, and H. V. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Trans. Signal Processing*, accepted in 2009.

[14] G. Still, "How to split the eigenvalues of a one-parameter family of matrices," *Optimization: A Journal of Mathematical Programming and Operations Research*, vol. 49, no. 4, pp. 387-403, 2001.

[15] T. Tao, "254A, notes 3a: Eigenvalues and sums of hermitian matrices," [online]. Available: http://terrytao.wordpress.com/2010/01/12/254a-notes-3a-eigenvalues-and-sums-of-hermitian-matrices/.

[16] N.P. Van Der Aa, H.G. Ter Morsche, and R.R.M. Mattheij, "Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem," *Electronic Journal of Linear Algebra*, vol. 16, pp. 300-314, Oct. 2007.

[17] M.L. Overton and R.S. Womersley, "Second derivatives for optimizing eigenvalues of symmetric matrices," *SIAM J. Matrix Anal. Appl.*, vol. 16, no. 3, pp. 697-718, 1995.

[18] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492-498, 1967.

[19] C. Zalinescu, "On the maximization of (not necessarily) convex functions on convex sets," *Journal of Global Optimization*, vol. 36, no. 3, pp. 379-389, Nov. 2006.

[20] Y. Huang and S. Zhang, "Complex matrix decomposition and quadratic programming," *Mathematics of Operations Research*, vol. 32, no. 3, pp. 758-768, 2007.

[21] D.G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd ed., New York: Springer, 2008.

[22] D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Belmont, MA: Athena Scientific, 1999.

[23] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475-494, Jun. 2001.

[24] Jan de Leeuw, "Block-relaxation algorithms in statistics, " *UCLA Statistics*, 1994.

[25] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, Nov. 1999.

[26] C. Charalambous and A.R. Conn, "An efficient method to solve the minimax problem directly," *SIAM J. Numer. Anal.*, vol. 15, no. 1, pp. 162-187, Feb. 1978.

[27] E.W. Cheney, *Introduction to Approximation Theory*, 2nd ed., New York: Chelsea, 1982.

[28] R. Chen, "Solution of minimax problems using equivalent differentiable functions," *Computers and Mathematics with Applications*, vol. 11, no. 12, pp. 1165-1169, Dec. 1985.

[29] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA: SIAM, 2000.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, UK: Cambridge Univ. Press, 2004.

[31] J. Nocedal and S.J. Wright, *Numerical optimization*, New York: Springer, 2006.

[32] M. Grant, S. Boyd, cvx Users' Guide, 2009.

[33] R.A. Horn and C.A. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1990.

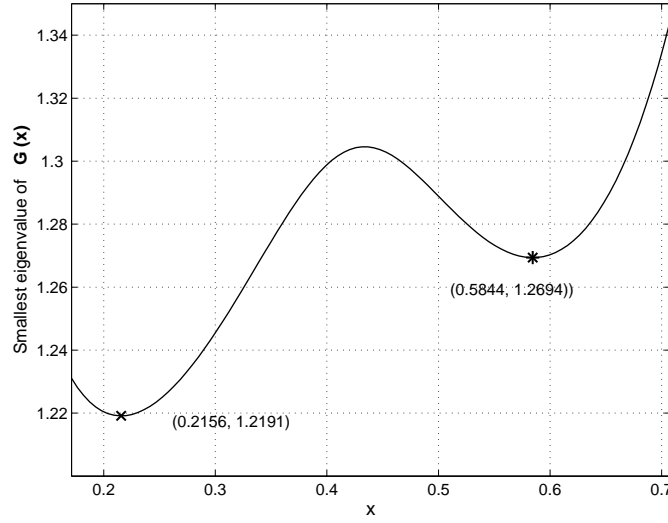[34] O.L. Mangasarian, *Nonlinear Programming*, Philadelphia, PA: SIAM, 1994.

Fig. 2. $\lambda_{\min}(\mathbf{G}(x))$ for x in $[x_l, x_u] = [0.2754, 0.6392]$ with 100 uniform points; The left point is for starting point $x_0 = x_l$ and the right point is for the starting point $x_0 = x_u$; SNR $= 10\,\text{dB}$; A total power constraint.
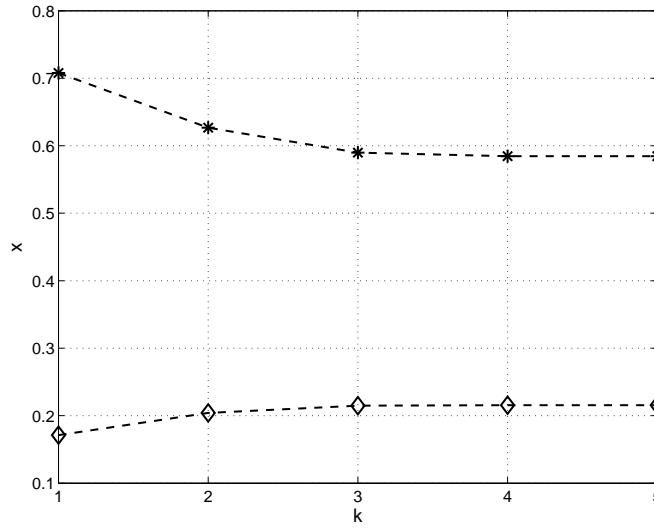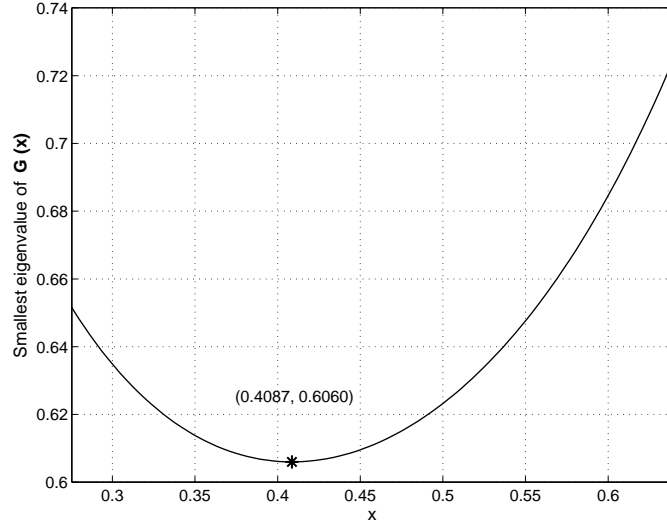

Fig. 3.   The iteration process; The upper line is for starting point $x_0 = x_l$ and the lower line is for the starting point $x_0 = x_u$; SNR $= 10\,\text{dB}$; A total power constraint.

Fig. 4. $\lambda_{\min}(\mathbf{G}(x))$ for x in $[x_l, x_u] = [0.2754, 0.6392]$ with 100 uniform points; The same convergent point is for starting point $x_0 = x_l$ and $x_0 = x_u$; SNR $= 10\,$dB; A total power constraint.
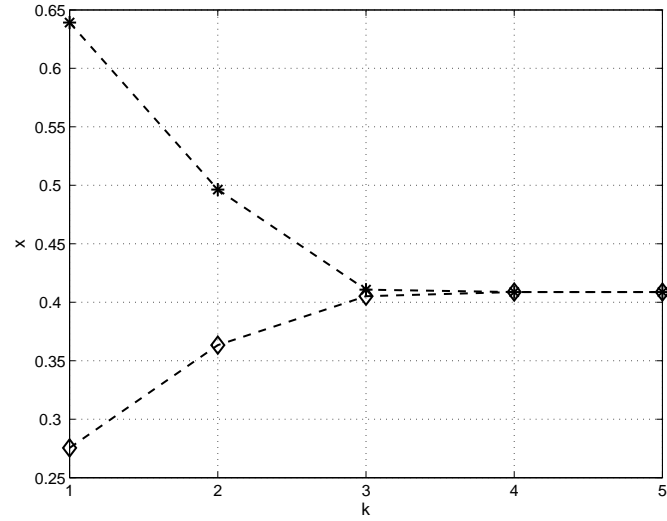

Fig. 5. The iteration process; The upper and lower lines converge to the same point for starting point $x_0 = x_l$ and $x_0 = x_u$; SNR $= 10\,$dB; A total power constraint.
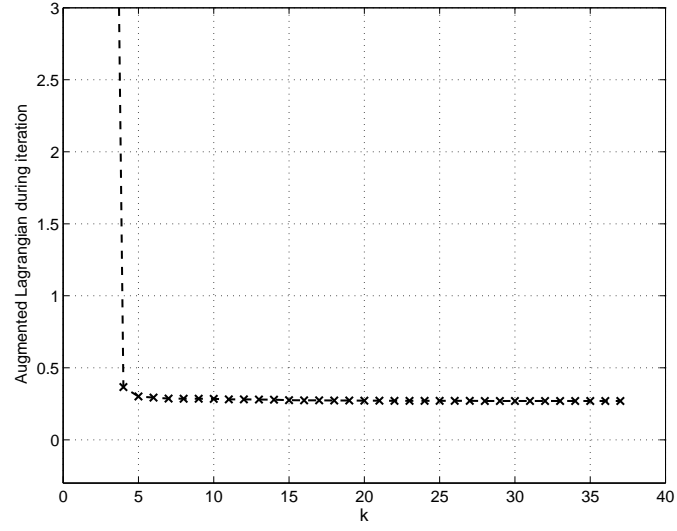
Fig. 6. The augmented Lagrangian $L(\mathbf{z}; \lambda; \mu)$ (defined in (65)) during the iteration process of the proposed algorithm; Individual relay power constraints; p-norm approximation.
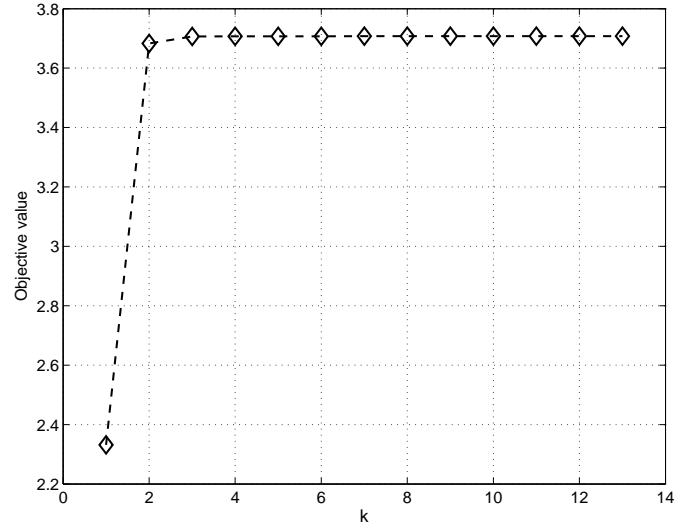

Fig. 7. The objective value during the iteration process of the proposed algorithm; Individual relay power constraints; Coordinate descent method.
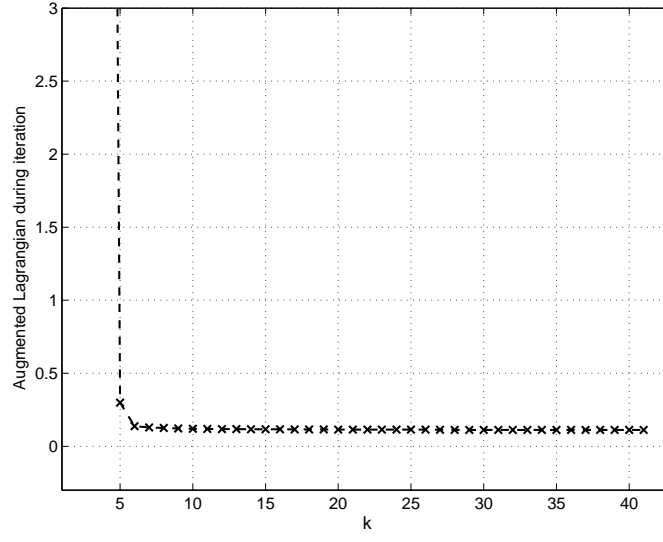
Fig. 8. The augmented Lagrangian $L(\mathbf{z}; \lambda; \mu)$ (defined in (65)) during the iteration process of the proposed algorithm; Individual relay power constraints; p-norm approximation.
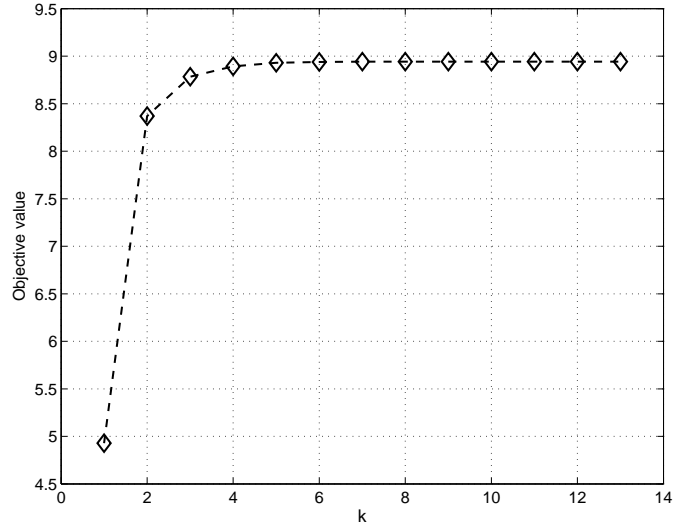

Fig. 9. The objective value during the iteration process of the proposed algorithm; Individual relay power constraints; Coordinate descent method.