

Energy-Efficient Multiuser SIMO: Achieving Probabilistic Robustness with Gaussian Channel Uncertainty

Gan Zheng *Member, IEEE*, Kai-Kit Wong, *Senior Member, IEEE*, and Tung-Sang Ng, *Fellow, IEEE*

Abstract—This paper addresses the joint optimization of power control and receive beamforming vectors for a multiuser single-input multiple-output (SIMO) antenna system in the uplink in which mobile users are single-antenna transmitters and the base station receiver has multiple antennas. Channel state information at the receiver (CSIR) is exploited but the CSIR is imperfect with its uncertainty being modeled as a random Gaussian matrix. Our objective is to devise an energy-efficient solution to minimize the individual users' transmit power while meeting the users' signal-to-interference plus noise ratio (SINR) constraints, under the consideration of CSIR and its error characteristics. This is achieved by solving a sum-power minimization problem, subject to a collection of users' outage probability constraints on their target SINRs. Regarding the signal power minus the sum of inter-user interferences (SMI) power as Gaussian, an iterative and convergent algorithm which is proved to reach the global optimum for the joint power allocation and receive beamforming solution, is proposed, though the optimization problem is indeed non-convex. A systematic scheme to detect feasibility and find a feasible initial solution, if there exists any, is also devised. Simulation results verify the use of Gaussian approximation and robustness of the proposed algorithm in terms of users' probability constraints, and indicate a significant performance gain as compared to the zero-forcing (ZF) and minimum mean-square-error (MMSE) beamforming systems.

Index Terms—Convex optimization, MIMO antenna, multiple-access channel, robust design, SDP relaxation, S-procedure.

I. INTRODUCTION

MULTI-ANTENNA systems [or known as multiple-input multiple-output (MIMO)] are the key of improving the energy and spectral efficiency of communication over wireless channels. This capacity gain is achievable not only in a point-to-point communication system [1], but also in both downlink and uplink of a multiuser wireless network [2]–[7]. Previous studies on multiuser MIMO systems are mostly based on the assumption of having perfect channel state information (CSI), under which the optimal strategy

in achieving the channel sum-rate has already been found [3], [4] and some low-complexity beamforming approaches have also been proposed (e.g., [5]–[7]). Remarkably, due to duality theories, the referenced works apply for both up and downlinks. However, it is known that the performance gain depends greatly on the correctness of the CSI available at both the transmitters and receivers and an unusable CSI will degrade the system performance considerably [8]. Due to estimation errors, the results based on a model with perfect CSI at the receiver (CSIR) are rarely applicable in practice, and a technique robust to CSIR errors needs to be sought.

Robust systems against channel mismatches can be obtained by two approaches: i) worst-case optimization and ii) stochastic or statistical analysis. In worst-case approaches, the CSI error has to be bounded and the system is required to maintain a given quality-of-service (QoS) for every possible channel realizations and error conditions [9]. This technique is, however, impossible if the CSI error is unbounded, for instance, in the case when the CSI is estimated at the receiver from training, which results in an unbounded Gaussian uncertainty in the channel estimates. For this reason, statistical approaches have emerged to provide robustness, but in the form of confidence level measured by probability [10]. Motivated by the fact that CSI error tends to be Gaussian distributed at the receiver, this paper will adopt the stochastic approach. Some of the related works are now reviewed below.

In [11]–[14], robustness is realized via the outage probability or the probability of distortionless response being considered as either a constraint or cost. To be specific, [11] investigated a multiuser single-input multiple-output (SIMO) system in the uplink¹ assuming the knowledge of the density function of the spatial covariance uncertainty. A robust beamforming scheme to minimize each user's outage probability was presented. Later in [12], a multiple-input single-output (MISO) downlink channel (i.e., the reverse link of a SIMO uplink) was addressed and a zero-forcing (ZF) beamforming-based cross-layer method which maximizes the system goodput with users' outage probability constraints was devised. Recently, [14] studied an uplink space-time block-coding (STBC) system and the receive beamforming vectors were optimized in maximizing the users' probability of distortionless response. Nevertheless, in the uplink results [11], [14], power control

Paper approved by K. K. Leung, the Editor for Wireless Network Access and Performance of the IEEE Communications Society. Manuscript received November 12, 2007; revised July 15, 2008.

G. Zheng and K. K. Wong are with the Department of Electronic and Electrical Engineering, University College London, UK (e-mail: {g.zheng, k.wong}@ee.ucl.ac.uk).

T. S. Ng is with the Electrical and Electronic Engineering Department, The University of Hong Kong (e-mail: tsng@eee.hku.hk).

This work was supported in part by The Hong Kong Research Grants Council under grant HKU7175/03E and The University Research Committee of The University of Hong Kong, and in part by The Engineering and Physical Science Research Council under grant EP/D058716/1, United Kingdom.

Digital Object Identifier 10.1109/TCOMM.2009.06.070574

¹In a SIMO uplink, mobile transmitters employ single antennas and only the base station receiver has a multi-element antenna array, whereas for a multiuser MIMO, both the base and mobile stations use multiple antennas.

between users was not considered, which would have a severe impact on the users' power consumption, while the adoption of ZF beamforming in [12] also inevitably compromises the capacity performance considerably [15].

In this paper, we consider a multiuser SIMO uplink system with imperfect CSIR. In an uplink system, the (transmit) power consumption at the mobile stations, is crucial because of the limited battery life. For this reason, power control is particularly important since it can manage the interference levels at the users' output signals, thereby reducing the required transmit power from the users for a given QoS (e.g., [16]–[18] addressed the optimal power control in a multiuser environment based on statistical CSI). The objective of this paper is to minimize each mobile user's transmit power while attaining the users' given outage probability constraints, by jointly optimizing the power control and the receive beamforming vectors of the users based on imperfect CSIR and the statistical knowledge of the CSIR uncertainty. Using the result in [19], it will be shown in Section IV-C that in our setting, minimization of each user's power can be accomplished by minimizing the sum-power of the users. As such, mathematically, we study the *sum-power* minimization problem with individual users' outage probability constraints in the uplink. Note that for each user, outage occurs if the target signal-to-interference plus noise ratio (SINR) is not satisfied and is caused by the CSIR errors, which we model as complex Gaussian, as is typical in the CSI estimates from training based on minimizing the mean-square-error (MMSE).

The problem under investigation is challenging because first the joint power control and receive beamforming optimization is non-convex and secondly, the outage probability expression, which involves the distribution of the signal power minus the sum of inter-user interference power (SII) on the SINR, is not known. To overcome the latter, we propose to treat the SII as Gaussian,² which permits to express the probability constraints in closed form. We then tackle the problem by developing an iterative algorithm which converges to a joint solution after a number of iterations by optimizing one set of variables (i.e., the power control vector or the beamforming vectors) at a time while keeping the other one fixed. Remarkably, we shall prove, by the method of standard mapping, that the algorithm converges to the global optimum³ though the problem is not convex. Feasibility issues are also addressed by presenting a method that guarantees to obtain a feasible solution as long as there exists one. Simulation results will reveal that probabilistic robustness at the user-level is achieved with each mobile user's power consumption minimized in the uplink. Further, a significant performance gap between the optimal power control with fixed ZF beamforming vectors and the joint power and beamforming optimization is observed.

The remainder of the paper is structured as follows. In Section II, we introduce the system model of an uplink MISO system with imperfect CSIR. The robust optimization problem is also formulated. Section III describes how one set of

variables can be optimized with others being fixed. Section IV then proposes an iterative algorithm which makes use of the results in Section III to obtain a joint solution for the power control and receive beamforming vectors. Convergence analysis, a proof for global optimality, and a scheme to find a feasible initial solution will also be given. Section V discusses some practical issues of the algorithm. Simulation results are presented in Section VI and we conclude the paper in Section VII.

Throughout this paper, the following notations are adopted. Complex number field is denoted as \mathbb{C} . Scalar is represented by a lowercase letter and $|\cdot|$ denotes its modulus. Vectors and matrices are represented by bold lowercase and uppercase letters respectively, and $\|\cdot\|$ is the Frobenius norm. $\mathbb{E}[\cdot]$ denotes the mean of a random variable, which may be a scalar, vector or even matrix. The superscript \dagger is used to denote the Hermitian transposition. $\mathbf{X} \succeq \mathbf{0}$ means that the matrix \mathbf{X} is positive semi-definite, while $\text{trace}(\mathbf{A})$ denotes the trace of \mathbf{A} . Finally, $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{V})$ denotes a vector of complex Gaussian entries with mean vector of \mathbf{m} and covariance matrix of \mathbf{V} .

II. SYSTEM MODEL

A. Multiuser SIMO Uplink

Consider an M -user uplink system where each mobile user has single antenna and the base station receiver has n_R receive antennas, as shown in Figure 1. For a particular user, say user m , data symbol $s_m \in \mathbb{C}$ is transmitted in time with $\mathbb{E}[|s_m|^2] = p_m$ where the time index is omitted for conciseness. At each of the receive antennas of the base station, a perturbed version of the transmitted symbol caused by a multiplicative channel fading and an additive noise is received. In this multiuser environment, the received signals can be written in vector form as

$$\mathbf{x} = \sum_{m=1}^M \mathbf{h}_m s_m + \mathbf{n}, \quad (1)$$

where $\mathbf{h}_m \in \mathbb{C}^{n_R}$ is the channel vector from user m to the base station and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I})$ is the noise vector of independent and identically distributed (i.i.d.) complex Gaussian entries. We model \mathbf{h}_m as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ so that the channel is in Rayleigh flat-fading, as is typical for indoor environments without the direct line-of-sight. It is also possible to characterize the channel using a composite model $g_m \mathbf{h}_m$ where the scalar g_m can be used to account for the large-scale path loss and shadowing and is assumed to be a constant during the period of interest. As such, in the following, for simplicity, we assume $g_m = 1 \forall m$.

The soft-estimate of s_m , denoted by \tilde{s}_m , can be obtained by multiplying \mathbf{x} with a receive beamforming vector \mathbf{r}_m , i.e.,

$$\tilde{s}_m = \mathbf{r}_m^\dagger \mathbf{x} \quad \forall m. \quad (2)$$

The fidelity of the output signal is measured by the SINR

$$\Gamma_m = \frac{p_m |\mathbf{r}_m^\dagger \mathbf{h}_m|^2}{\sum_{\substack{n=1 \\ n \neq m}}^M p_n |\mathbf{r}_m^\dagger \mathbf{h}_n|^2 + N_0}. \quad (3)$$

²The Gaussian approximation improves with the number of users due to the central limit theorem. Remarkably, results in Section VI will show that this approximation is acceptable even for systems with small number of users.

³Note that the global optimality is claimed for the problem with Gaussian probability constraints. For this reason, the proposed method is suboptimal.

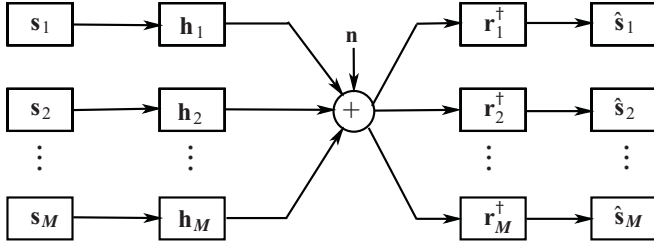


Fig. 1. The system model of the multiuser SIMO uplink system.

B. Gaussian CSIR Uncertainty

At each user, SINR depends on the power allocation of all users \$\{p_n\}_{\forall n}\$ and the receive beamforming of this particular user, \$\mathbf{r}_m\$. Ideally, both the power control and beamforming vectors should be jointly optimized in accordance with the CSIR \$\{\mathbf{h}_n\}_{\forall n}\$. However, CSI is usually estimated from training at finite received signal-to-noise ratio (SNR) based on the MMSE criterion and uncertainty in the CSI estimates will exist. To model this CSIR uncertainty, we assume

$$\mathbf{h}_m = \tilde{\mathbf{h}}_m + \Delta \mathbf{h}_m \quad \forall m, \quad (4)$$

where \$\tilde{\mathbf{h}}_m\$ is the CSI estimate known at the base station, and \$\Delta \mathbf{h}_m \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I})\$ corresponds to the CSIR error/uncertainty. This model is particularly suitable for the case when CSIR is learned from an MMSE estimator which results in a Gaussian error, independent of \$\tilde{\mathbf{h}}_m\$. In addition, the mean-square-error (MSE) in the CSI estimates is measured by

$$\frac{1}{n_R} \mathbb{E} \left[\left\| \mathbf{h}_m - \tilde{\mathbf{h}}_m \right\|^2 \right] = \frac{1}{n_R} \mathbb{E} \left[\left\| \Delta \mathbf{h}_m \right\|^2 \right] = \sigma_h^2. \quad (5)$$

In general, \$\sigma_h^2 \le 1\$ and typically, \$\sigma_h^2\$ should be smaller than 0.05 in order for the CSI estimates to be useful. This model has been widely used to characterize the CSI uncertainty due to channel estimation (e.g., [8], [12], [20], [21]). Henceforth, we assume that ‘‘imperfect CSIR’’ includes the knowledge of \$\{\tilde{\mathbf{h}}_n\}_{\forall n}\$ and \$\sigma_h^2\$, which will be exploited in the design of a robust system.

C. The Robust Design Problem

With unbounded CSIR uncertainty in (4), it is impossible that a given SINR of a user can be maintained for every possible channel error condition. For this reason, robustness can only be achieved in the probabilistic sense. In light of this, we consider that users are subject to individual probability constraints on their target SINRs \$\{\gamma_m\}_{\forall m}\$, and our aim is to minimize each user’s (transmit) power consumption for attaining the users’ service probability requirements \$\{\varepsilon_m\}_{\forall m}\$ in the SIMO uplink (1) with imperfect CSIR. Mathematically, that is,

$$\min_{\substack{\mathbf{p} \\ \{r_m^\dagger\}_{m=1}^M \\ \mathbf{p} > \mathbf{0}}} \mathbf{p} \quad \text{s.t.} \quad \mathcal{P}(\{\Gamma_m \geq \gamma_m\}) \geq \varepsilon_m \quad \forall m, \quad (6)$$

where \$\mathcal{P}(A)\$ denotes the probability of an event \$A\$, and \$\mathbf{p} \triangleq (p_1, \dots, p_M)\$ is the power control vector. The cost function in (6) is an element-wise minimization of \$\mathbf{p}\$. Notice that the service probability is conditioned on the channel

estimate \$\{\tilde{\mathbf{h}}_m\}\$ and the outage is caused by the estimation error only.

To begin, we note that the probability constraint is

$$\mathcal{P}(\{x_m \geq \gamma_m N_0\}) \geq \varepsilon_m, \quad (7)$$

where

$$x_m \triangleq p_m \left| \mathbf{r}_m^\dagger \mathbf{h}_m \right|^2 - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \left| \mathbf{r}_m^\dagger \mathbf{h}_n \right|^2. \quad (8)$$

In order to proceed further, we need to know the statistics of \$x_m\$ whose randomness is caused by the channel uncertainty. For given \$\{\mathbf{r}_m\}\$ and \$\{p_m\}\$, it is noted that \$|\mathbf{r}_m^\dagger \mathbf{h}_n|^2\$ follows a Chi-square distribution with 2 degrees of freedom and the nonlinear parameters \$s = \sqrt{p_n} |\mathbf{r}_m^\dagger \tilde{\mathbf{h}}_n|\$ and \$\sigma = \frac{p_n \sigma_h^2}{2}\$. As such, \$x_m\$ is the difference of non-central Chi-square random variables. This kind of distribution has been studied in [22], [23], but unfortunately, the cumulative distribution function (CDF) is too complicated to be useful for the optimization of the power and beamforming vectors. To resolve this, we note that \$x_m\$ is the sum of \$M^2\$ independent random variables and from the central limit theorem (CLT), if \$M\$ is large, then \$x_m\$ is anticipated to be approximately Gaussian with the parameters \$\mu_{x_m} = \mathbb{E}[x_m]\$ and \$\sigma_{x_m}^2 = \text{VAR}[x_m]\$, given as (9a) and (9b) (see top of next page and see Appendix A for the derivation) As a result, the service (or non-outage) probability for user \$m\$ can be rewritten as

$$\begin{aligned} \mathcal{P}(\{x_m \geq \gamma_m N_0\}) &\geq \varepsilon_m \\ \Leftrightarrow \frac{1}{2} - \frac{1}{2} \text{erf} \left(\frac{\gamma_m N_0 - \mu_{x_m}}{\sqrt{2} \sigma_{x_m}} \right) &\geq \varepsilon_m \\ \Leftrightarrow \frac{\mu_{x_m} - \gamma_m N_0}{\sqrt{2} \sigma_{x_m}} &\geq \text{erf}^{-1}(2\varepsilon_m - 1) \end{aligned} \quad (10)$$

where \$\text{erf}(\cdot)\$ denotes the error function, and \$\text{erf}^{-1}(\cdot)\$ denotes its inverse. Substituting (9a) and (9b) into (10), (7) becomes (11). For ease of exposition, we define the following parameters:

$$\eta_m = \sqrt{2} \text{erf}^{-1}(2\varepsilon_m - 1), \quad (12a)$$

$$\mathbf{A}_m = \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger + \sigma_h^2 \mathbf{I}, \quad (12b)$$

$$\mathbf{B}_m(\mathbf{p}) = \gamma_m \left(\sum_{\substack{n=1 \\ n \neq m}}^M p_n \mathbf{A}_n + N_0 \mathbf{I} \right), \quad (12c)$$

$$\begin{aligned} \mathbf{C}_m(\mathbf{p}) &= \eta_m^2 \left[\left(2\sigma_h^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I} \right) p_m^2 \right. \\ &\quad \left. + \gamma_m^2 \sum_{\substack{n=1 \\ n \neq m}}^M \left(2\sigma_h^2 \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_h^4 \mathbf{I} \right) p_n^2 \right], \end{aligned} \quad (12d)$$

$$\mathbf{D}_m(\mathbf{p}) = p_m \mathbf{A}_m - \mathbf{B}_m(\mathbf{p}). \quad (12e)$$

Further to the Gaussian approximation, (6) can be expressed as

$$\begin{aligned} \min_{\mathbf{p} > \mathbf{0}, \{\mathbf{r}_m\}} \mathbf{p} \quad \text{s.t.} \quad &F_m(\mathbf{p}, \mathbf{r}_m) \triangleq \mathbf{r}_m^\dagger \mathbf{D}_m(\mathbf{p}) \mathbf{r}_m \\ &- \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m(\mathbf{p}) \mathbf{r}_m} \geq 0 \quad \forall m. \end{aligned} \quad (13)$$

The cost function in (13) is an element-wise minimization of \$\mathbf{p}\$, which by the results in [19], [28], can be realized by minimizing the sum of the elements of \$\mathbf{p}\$ or \$\sum_m p_m\$ (more

$$\mu_{x_m} = \mathbf{r}_m^\dagger \left(p_m \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger \right) \mathbf{r}_m + \sigma_h^2 \left(p_m - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \right), \quad (9a)$$

$$\sigma_{x_m}^2 = \mathbf{r}_m^\dagger \left[\left(2\sigma_h^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I} \right) p_m^2 + \gamma_m^2 \sum_{\substack{n=1 \\ n \neq m}}^M \left(2\sigma_h^2 \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_h^4 \mathbf{I} \right) p_n^2 \right] \mathbf{r}_m, \quad (9b)$$

$$\mathbf{r}_m^\dagger \left(p_m \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger \right) \mathbf{r}_m + \sigma_h^2 \left(p_m - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \right) - \sqrt{2} \operatorname{erf}^{-1}(2\varepsilon_m - 1) \\ \sqrt{\mathbf{r}_m^\dagger \left[\left(2\sigma_h^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I} \right) p_m^2 + \gamma_m^2 \sum_{\substack{n=1 \\ n \neq m}}^M \left(2\sigma_h^2 \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_h^4 \mathbf{I} \right) p_n^2 \right] \mathbf{r}_m} \geq \gamma_m N_0 \quad (11)$$

details will be provided in Section IV-C). In what follows, we shall focus on

$$\min_{\mathbf{p} > \mathbf{0}, \{\mathbf{r}_m\}} \sum_{m=1}^M p_m \quad \text{s.t.} \quad F_m(\mathbf{p}, \mathbf{r}_m) \geq 0 \quad \forall m. \quad (14)$$

In the sequel, our efforts will be spent on solving (14) optimally whereas the accuracy of the Gaussian approximation will be tested and discussed in the numerical results section.

III. FIXED-POINT ANALYSIS AT THE OPTIMUM STATE

Since (14) is non-convex and the optimizing variables \mathbf{p} and $\{\mathbf{r}_m\}$ depend on each other, the joint optimization problem is extremely arduous. In this section, we look at how one set of variables can be optimized if the others are given and fixed. The results can be interpreted as the fixed-point analysis at the joint optimum state, and will facilitate an iterative optimization procedure to be presented in Section IV.

A. Optimal Receive Beamforming Vectors

If the optimal power vector \mathbf{p} is known, finding the corresponding optimal \mathbf{r}_m requires finding the vector that maximizes the individual user's performance metric $F_m(\mathbf{p}, \mathbf{r}_m)$ in (14), i.e.,

$$\mathbf{r}_{\text{opt}} = \arg \max_{\|\mathbf{r}\|=1} F(\mathbf{p}, \mathbf{r}) = \mathbf{r}^\dagger \mathbf{D} \mathbf{r} - \sqrt{\mathbf{r}^\dagger \mathbf{C} \mathbf{r}}, \quad (15)$$

where for convenience, we have omitted the user index m . This problem is not convex because of the equality norm constraint and an indefinite matrix \mathbf{D} . However, we shall show that the global-optimal solution can indeed be found using semi-definite programming (SDP) relaxation followed by an efficient one-dimensional search, such as the Dividing RECTangle (DIRECT) algorithm [24]. Before we move on, it is worth noting that with perfect CSIR, the optimal receive beamforming vector can be easily found by MMSE, but this is generally not the case here.

Step 1—Simplified Beamforming Optimization Using SDP:

We proceed by first considering that $\mathbf{r}^\dagger \mathbf{C} \mathbf{r} = t$ for some fixed and known t . Then, (15) reduces to

$$f(t) = \min_{\|\mathbf{r}\|=1, \mathbf{r}^\dagger \mathbf{C} \mathbf{r} = t} -\mathbf{r}^\dagger \mathbf{D} \mathbf{r}. \quad (16)$$

The optimal solution of (16) is summarized in Theorem 1.

Theorem 1: The globally optimal solution to the non-convex quadratic problem (16) can be obtained by solving the SDP:

$$\min_{\mathbf{R} \succeq \mathbf{0}} -\operatorname{trace}(\mathbf{D} \mathbf{R}) \quad \text{s.t.} \quad \begin{cases} \operatorname{trace}(\mathbf{R}) = 1, \\ \operatorname{trace}(\mathbf{C} \mathbf{R}) = t. \end{cases} \quad (17)$$

Also, there exists at least a rank-1 solution \mathbf{R} such that $\mathbf{R} = \mathbf{r} \mathbf{r}^\dagger$.

Proof: We outline the proof here with the technical details given in Appendix B.

- 1) An equivalent convex problem (60) is first found in the sense that both (16) and (60) achieve the same objective value using the S-Lemma [25], [26].
- 2) Due to the rank relaxation, the solution to (17) generally provides a lower bound for (16). However, it can be shown that (17) is exactly the dual of (60) and because of the Slater's condition, they have the same objective value. As a result, the relaxation problem (17) is equivalent to (16).
- 3) Finally, the optimal rank-1 solution to (17) in closed form is derived based on the Karush-Kuhn-Tucker (necessary and sufficient) condition, which gives \mathbf{r}_{opt} of (16).

This completes the proof. \blacksquare

Step 2—One-Dimensional Sampling Search: In Step 1, it has been shown that with a known t , an optimal solution to (16) can be obtained. Denoting $f(t)$ as the optimized objective value of (16), the original beamforming problem (15) is expressed as

$$\min_{\substack{t \geq 0 \\ \lambda_{\min}(\mathbf{C}) \leq t \leq \lambda_{\max}(\mathbf{C})}} f(t) + \sqrt{t}, \quad (18)$$

where $\lambda_{\min}(\mathbf{C})$ and $\lambda_{\max}(\mathbf{C})$ denote, respectively, the smallest and largest eigenvalues of \mathbf{C} . Note that (18) is a problem

with only one bounded variable t . Therefore, the globally optimal solution can be obtained efficiently by a one-dimensional sampling search. One possible method is DIRECT [24], which is a numerical sampling algorithm, and requires no knowledge of the objective function nor its gradient, and uses the information from the samples it has obtained to decide where to search next.

In summary, by using SDP relaxation together with DIRECT, the optimal receive beamforming vectors $\{\mathbf{r}_m\}$ can be obtained for any given \mathbf{p} . In other words, if the optimal power vector is known, the optimal beamforming vectors can be found.

B. Optimal Power Control Vector

With known $\{\mathbf{r}_m\}$, the optimization of \mathbf{p} is convex. In particular, we use the following simple fixed-point update [19], [28] for optimizing \mathbf{p} such that each user's performance requirement is satisfied. That is, at the $(n+1)$ th iteration, we have

$$p_m^{n+1} = \frac{\mathbf{r}_m^\dagger \mathbf{B}_m(\mathbf{p}^n) \mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m(\mathbf{p}^n) \mathbf{r}_m}}{\mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}, \quad (19)$$

where the superscript n denotes the variable at the n th iteration. It turns out that this iterative update on \mathbf{p} will converge to the optimal solution given $\{\mathbf{r}_m\}$ is known. Technical details on the optimality of this will be proved in the next section.

IV. ALGORITHM AND ANALYSIS

A. The Iterative Algorithm

The techniques above are now combined to jointly optimize the power and beamforming vectors iteratively as follows:

- 1) Initialize the iteration index $n = 0$, and choose a feasible joint solution $(\{\mathbf{r}_m^n\}, \mathbf{p}^n)$ (details on how a feasible initial solution may be found, to be addressed in Section IV-D).
- 2) For a known \mathbf{p}^n , the optimal receive beamforming vectors $\{\mathbf{r}_m^{n+1}\}$ are obtained by solving (18) using SDP relaxation with a one-dimensional sampling algorithm.
- 3) For a given set of $\{\mathbf{r}_m^{n+1}\}$, \mathbf{p}^{n+1} is updated using (19).
- 4) Update $n := n + 1$ and go back to Step 2 until it converges.

B. Convergence Analysis

Though the above algorithm emerges naturally, whether it converges is an important issue and this is addressed below.

Theorem 2: Given a feasible initial solution, the total transmit power in the iterative algorithm described in Section IV-A is monotonically decreasing and hence the algorithm converges.

Proof: Given an initial feasible solution $(\mathbf{p}^0, \{\mathbf{r}_m^0\})$, we have $F_m(\mathbf{p}^0, \mathbf{r}_m^0) \geq 0$. The receive beamforming vector at the $(n+1)$ th iteration is subsequently chosen such that $\mathbf{r}_m^{n+1} = \arg \max_{\|\mathbf{r}_m\|=1} F_m(\mathbf{p}^n, \mathbf{r}_m)$. Thus, $F_m(\mathbf{p}^n, \mathbf{r}_m^{n+1}) \geq 0$ and

$$p_m^{n+1} = \frac{(\mathbf{r}_m^{n+1})^\dagger \mathbf{B}_m(\mathbf{p}^n) \mathbf{r}_m^{n+1} + \sqrt{(\mathbf{r}_m^{n+1})^\dagger \mathbf{C}_m(\mathbf{p}^n) \mathbf{r}_m^{n+1}}}{(\mathbf{r}_m^{n+1})^\dagger \mathbf{A}_m \mathbf{r}_m^{n+1}} \leq p_m^n. \quad (20)$$

This means that each user's transmit power (and so does the total power) is monotonically decreasing. Furthermore, since $\mathbf{r}^\dagger \mathbf{B}_m(\mathbf{p}) \mathbf{r}$ and $\mathbf{r}^\dagger \mathbf{C}_m(\mathbf{p}) \mathbf{r}$ are increasing functions of \mathbf{p} ,

$$p_m^{n+1} \geq \frac{(\mathbf{r}_m^{n+1})^\dagger \mathbf{B}_m(\mathbf{p}^{n+1}) \mathbf{r}_m^{n+1} + \sqrt{(\mathbf{r}_m^{n+1})^\dagger \mathbf{C}_m(\mathbf{p}^{n+1}) \mathbf{r}_m^{n+1}}}{(\mathbf{r}_m^{n+1})^\dagger \mathbf{A}_m \mathbf{r}_m^{n+1}}, \quad (21)$$

which implies that

$$F_m(\mathbf{p}^{n+1}, \mathbf{r}_m^{n+1}) \geq 0. \quad (22)$$

It follows that all the constraints are satisfied in each iteration and the total transmit power decreases as iteration goes, so convergence follows. Finally, from the definition of the power update, all the constraints must be satisfied with equality at the fixed-point $(\mathbf{p}^*, \mathbf{r}_m^*)$, i.e., $F_m(\mathbf{p}^*, \mathbf{r}_m^*) = 0 \forall m$. ■

C. Proof of Global Optimality

Since (14) is non-convex, the above convergence proof does not generally guarantee the optimality of the steady-state solution. However, as shall be shown in Theorem 3 below, the proposed algorithm is indeed globally optimal.

Theorem 3: Given any feasible initial solution, the iterative algorithm above converges to the global optimum of (14).

Proof: To facilitate our analysis, we define

$$G_m(\mathbf{q}, \mathbf{s}_m) \triangleq \frac{\mathbf{s}_m^\dagger \mathbf{B}_m(\mathbf{q}) \mathbf{s}_m + \sqrt{\mathbf{s}_m^\dagger \mathbf{C}_m(\mathbf{q}) \mathbf{s}_m}}{\mathbf{s}_m^\dagger \mathbf{A}_m \mathbf{s}_m}. \quad (23)$$

Note that the proposed algorithm can be viewed as a mapping \mathcal{T} from \mathbf{p}^n to \mathbf{p}^{n+1} , or written as $\mathbf{p}^{n+1} = \mathcal{T}(\mathbf{p}^n)$ given by

$$\mathcal{T} : \begin{cases} \forall m : \mathbf{r}_m^{n+1} = \arg \min_{\|\mathbf{r}\|=1} -\mathbf{r}^\dagger \mathbf{D}_m(\mathbf{p}^n) \mathbf{r} + \sqrt{\mathbf{r}^\dagger \mathbf{C}_m(\mathbf{p}^n) \mathbf{r}} \\ \quad = \arg \min_{\|\mathbf{r}\|=1} -F_m(\mathbf{p}^n, \mathbf{r}), \\ \forall m : p_m^{n+1} = G_m(\mathbf{p}^n, \mathbf{r}_m^{n+1}). \end{cases} \quad (24)$$

Now, define another mapping \mathcal{S} from \mathbf{q}^n to \mathbf{q}^{n+1} , or $\mathbf{q}^{n+1} = \mathcal{S}(\mathbf{q}^n)$, which is given by

$$\mathcal{S} : \begin{cases} \forall m : \mathbf{s}_m^{n+1} = \arg \min_{\|\mathbf{s}\|=1} \frac{\mathbf{s}^\dagger \mathbf{B}_m(\mathbf{q}^n) \mathbf{s} + \sqrt{\mathbf{s}^\dagger \mathbf{C}_m(\mathbf{q}^n) \mathbf{s}}}{\mathbf{s}^\dagger \mathbf{A}_m \mathbf{s}} \\ \quad = \arg \min_{\|\mathbf{s}\|=1} G_m(\mathbf{q}^n, \mathbf{s}), \\ \forall m : q_m^{n+1} = G_m(\mathbf{q}^n, \mathbf{s}_m^{n+1}). \end{cases} \quad (25)$$

It can be easily shown that \mathcal{S} satisfies the following properties.

- 1) Positivity: $\mathcal{S}(\mathbf{p}) > 0$.
- 2) Monotonicity: if $\mathbf{p} \geq \mathbf{p}'$, then $\mathcal{S}(\mathbf{p}) \geq \mathcal{S}(\mathbf{p}')$.
- 3) Scalability: for any constant $\alpha > 1$, $\alpha \mathcal{S}(\mathbf{p}) \geq \mathcal{S}(\alpha \mathbf{p})$.

Although the solution of the mapping \mathcal{S} is unknown, it has been proved in [19] that if the problem is feasible, such properties can guarantee the convergence of \mathcal{S} to the global and unique optimum. Numerical results, however, indicate that the proposed algorithm (i.e., the mapping \mathcal{T}) does not have all the properties. Despite this, the proof here shows that the mapping \mathcal{T} actually converges to the same fixed-point as the mapping \mathcal{S} , thereby ensuring the optimality of \mathcal{T} . This proof goes as follows.

By definition, $(\mathbf{q}^*, \mathbf{s}_m^*)$ is the fixed-point of \mathcal{S} if and only if

$$\begin{cases} \forall m : \mathbf{s}_m^* = \arg \min_{\|\mathbf{s}\|=1} G_m(\mathbf{q}^*, \mathbf{s}), \\ \forall m : q_m^* = G_m(\mathbf{q}^*, \mathbf{s}_m^*) \\ = \frac{(\mathbf{s}_m^*)^\dagger \mathbf{B}_m(\mathbf{q}^*) \mathbf{s}_m^* + \sqrt{(\mathbf{s}_m^*)^\dagger \mathbf{C}_m(\mathbf{q}^*) \mathbf{s}_m^*}}{(\mathbf{s}_m^*)^\dagger \mathbf{A}_m \mathbf{s}_m^*}, \end{cases} \quad (26)$$

and the power vector \mathbf{q}^* is unique. Denoting the fixed-point solution of \mathcal{T} as $(\mathbf{p}^*, \mathbf{r}_m^*)$, we have also

$$\begin{cases} \forall m : \mathbf{r}_m^* = \arg \min_{\|\mathbf{r}\|=1} -F_m(\mathbf{p}^*, \mathbf{r}) \text{ and } F_m(\mathbf{p}^*, \mathbf{r}_m^*) = 0, \\ \forall m : p_m^* = G_m(\mathbf{p}^*, \mathbf{r}_m^*) \\ = \frac{\mathbf{r}_m^{*\dagger} \mathbf{B}_m(\mathbf{p}^*) \mathbf{r}_m^* + \sqrt{\mathbf{r}_m^{*\dagger} \mathbf{C}_m(\mathbf{p}^*) \mathbf{r}_m^*}}{\mathbf{r}_m^{*\dagger} \mathbf{A}_m \mathbf{r}_m^*}. \end{cases} \quad (27)$$

We prove the main result by the method of contradiction. First, we assume that for the fixed-point power vector \mathbf{p}^* in the steady state of \mathcal{T} , there exists a better vector \mathbf{c}_m than \mathbf{r}_m^* so

$$G_m(\mathbf{p}^*, \mathbf{c}_m) < G_m(\mathbf{p}^*, \mathbf{r}_m^*) = p_m^*. \quad (28)$$

Then, this implies that

$$\frac{\mathbf{c}_m^\dagger \mathbf{B}_m(\mathbf{p}^*) \mathbf{c}_m + \sqrt{\mathbf{c}_m^\dagger \mathbf{C}_m(\mathbf{p}^*) \mathbf{c}_m}}{\mathbf{c}_m^\dagger \mathbf{A}_m \mathbf{c}_m} < p_m^* \quad (29)$$

and

$$-F_m(\mathbf{p}^*, \mathbf{c}_m) = \mathbf{c}_m^\dagger [\mathbf{B}_m(\mathbf{p}^*) - p_m^* \mathbf{A}_m] \mathbf{c}_m + \sqrt{\mathbf{c}_m^\dagger \mathbf{C}_m(\mathbf{p}^*) \mathbf{c}_m} < 0, \quad (30)$$

which contradicts (27). Thus,

$$\mathbf{r}_m^* = \arg \min_{\|\mathbf{r}\|=1} -F_m(\mathbf{p}^*, \mathbf{r}) = \arg \min_{\|\mathbf{r}\|=1} G_m(\mathbf{p}^*, \mathbf{r}). \quad (31)$$

Therefore,

$$\begin{cases} \forall m : \mathbf{r}_m^* = \arg \min_{\|\mathbf{r}\|=1} G_m(\mathbf{p}^*, \mathbf{r}), \\ \forall m : p_m^* = G_m(\mathbf{p}^*, \mathbf{r}_m^*). \end{cases} \quad (32)$$

In other words, the fixed-point $(\mathbf{p}^*, \mathbf{r}_m^*)$ of \mathcal{T} is also the fixed-point of \mathcal{S} . As the fixed-point of \mathcal{S} is unique and optimal, \mathcal{T} is also optimal, which completes the proof. ■

Element-wise Minimum Power Vector—It has been proved in [19] that the fixed-point power vector from \mathcal{S} is element-wise minimum; i.e., for any feasible power allocation \mathbf{q}' , $\mathbf{q}^* \leq \mathbf{q}'$, and the same is true for \mathcal{T} . As a result, it can be concluded that the proposed algorithm is not only optimal in minimizing the sum-power, but also in minimizing the individual users' power.

D. Initialization and Feasibility Issue

Thus far, little is understood about the feasibility of linear multiuser MIMO antenna systems with imperfect CSI and even perfect CSI. The above analysis we have presented is based on the condition that (14) is feasible, and the proposed algorithm also works on the assumption that a feasible starting point exists and is known to begin the required iteration. Here, we address this critical issue by devising a method to check the feasibility and to find a feasible initial solution, if there exists any.

D.1 Feasibility Check: The main result for feasibility check is the QoS-balancing algorithm, which is based on solving the problem (14), written in the following form:

$$\min_{\mathbf{p} > \mathbf{0}, \{\mathbf{r}_m\}} \sum_{m=1}^M p_m \quad \text{s.t.} \quad \frac{p_m \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m}} \geq 1 \quad \forall m. \quad (33)$$

It is proved in Lemma 1 that at the optimum, all the constraints must be active, meaning that the constraints hold with equality.

Lemma 1: All constraints in (33) should be satisfied with equalities to achieve the optimum.

Proof: The proof uses the method of contradiction. To start with, without loss of generality, we assume that at the optimum \mathbf{p}^* , the m th user's requirement is over satisfied, i.e.,

$$\frac{p_m^* \mathbf{r}_m^{*\dagger} \mathbf{A}_m \mathbf{r}_m^*}{\mathbf{r}_m^{*\dagger} \mathbf{B}_m \mathbf{r}_m^* + \sqrt{\eta_m^2 \left[(2\sigma_h^2 \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I}) (p_m^*)^2 + c \right]}} \equiv \alpha > 1, \quad (34)$$

where we have defined

$$c \triangleq \gamma_m^2 \sum_{\substack{n=1 \\ n \neq m}}^M \left(2\sigma_h^2 \hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^\dagger + \sigma_h^4 \mathbf{I} \right) (p_n^*)^2 \quad (35)$$

for ease of composition. Then, it is always possible to choose a new power element $q_m = \frac{p_m^*}{\alpha}$ that gives

$$\begin{aligned} & \frac{q_m \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\eta_m^2 \left[(2\sigma_h^2 \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I}) q_m^2 + c \right]}} \\ &= \frac{\frac{p_m^*}{\alpha} \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\eta_m^2 \left[(2\sigma_h^2 \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I}) \frac{(p_m^*)^2}{\alpha^2} + c \right]}} \\ &> \frac{\frac{p_m^*}{\alpha} \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\eta_m^2 \left[(2\sigma_h^2 \hat{\mathbf{h}}_m \hat{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I}) (p_m^*)^2 + c \right]}} \\ &> \left(\frac{1}{\alpha} \right) \alpha = 1. \end{aligned} \quad (36)$$

As a result, the m th user's power can be further reduced to $\frac{p_m^*}{\alpha}$ without violating the users' requirements (i.e., the m th user's requirement is still satisfied and all the other users' requirements are over satisfied), which contradicts the optimality of \mathbf{p} and therefore completes the proof. ■

Now consider the following QoS-balancing problem, which has the physical meaning of balancing the values of the users' requirements with a total power constraint P_T :

$$\begin{aligned} & \max_{\mathbf{p} > \mathbf{0}, y > \mathbf{0}, \{\mathbf{r}_m\}} y \\ & \text{s.t.} \quad \begin{cases} \frac{p_m \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m}} \geq y \quad \forall m, \\ \sum_{m=1}^M p_m \leq P_T. \end{cases} \end{aligned} \quad (37)$$

(37) is equivalent to the original problem (14) if the optimal value of y equals one and in this case, the required P_T in

(37) gives the optimal objective value of (14). Using the same argument as in Lemma 1, we know that at the optimum, all the constraints are satisfied with equalities. For a given fixed \mathbf{r}_m , obtaining the optimal power allocation requires to solve

$$\begin{cases} \min_{\mathbf{p} > \mathbf{0}, y > 0} y^{-1} \text{ s.t.} \\ (\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m) y p_m^{-1} + (\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m)^{\frac{1}{2}} y p_m^{-1} \leq \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m, \\ \sum_{m=1}^M p_m \leq P_T. \end{cases} \quad (38)$$

Note that $(\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m)$ and $\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m$ are all posynomials of the power allocation $\{p_m\}$. Hence, the above power optimization problem is recognized as a geometric programming (GP) problem [27], and thus, the optimal power allocation can be readily found. To jointly optimize power and beamforming, the following QoS-balancing algorithm is proposed.

QoS-balancing algorithm:

- 1) Initialize $n = 0$ and $\{\mathbf{r}_m^n\}$ arbitrarily.
- 2) For known $\{\mathbf{r}_m^n\}$, find the optimal $\{\mathbf{p}^{n+1}, y^{n+1}\}$ using GP to solve (38).
- 3) For given $\{\mathbf{p}^{n+1}, y^{n+1}\}$, the optimal $\{\mathbf{r}_m^{n+1}\}$ is found by

$$\mathbf{r}_m^{n+1} = \arg \min_{\|\mathbf{r}_m\|=1} [\mathbf{r}_m^\dagger \mathbf{B}_m (\mathbf{p}^{n+1}) \mathbf{r}_m] y^{n+1} p_m^{-1} + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m (\mathbf{p}^{n+1}) \mathbf{r}_m} y^{n+1} p_m^{-1} - \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m. \quad (39)$$

This problem can be solved in a similar way described in Section III-A for the receiver beamforming optimization.

- 4) Go back to Step 2 until convergence.

Very importantly, we can see that if the optimal $y^{\text{opt}} \geq 1$, then (33) [and hence (14)] is feasible. Therefore, y^{opt} can serve as a feasibility indicator. On the other hand, it can be proved that y^n is a non-decreasing sequence as the iteration goes, so it converges. In order to show that this algorithm can ensure to find a feasible solution if there exists any, we also need to prove that this algorithm can achieve the optimal y for any given power P_T . The reason is that although y is non-decreasing, it is not necessarily a strictly increasing function of P_T , and when $P_T \rightarrow \infty$, y may remain in a local optima. In other words, if the above algorithm is suboptimal, there is a chance that the problem (14) is indeed feasible but the algorithm fails to find a feasible solution or it can only converge to some $y^\infty < 1$.

Theorem 4: The QoS-balancing algorithm converges to the global optimum of (37).

Proof: Note that the QoS-balancing algorithm converges to the stationary point $\{\mathbf{p}^*, \{\mathbf{r}_m^*\}, y^*\}$:

$$\begin{cases} \forall m : \mathbf{r}_m^* = \arg \min_{\|\mathbf{r}_m\|=1} \frac{[\mathbf{r}_m^\dagger \mathbf{B}_m (\mathbf{p}^*) \mathbf{r}_m] y^*}{p_m^*} \\ \quad + \frac{\sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m (\mathbf{p}^*) \mathbf{r}_m} y^*}{p_m^*} - \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m, \\ \forall m : p_m^* = y^* \frac{(\mathbf{r}_m^*)^\dagger \mathbf{B}_m (\mathbf{p}^*) \mathbf{r}_m^* + \sqrt{(\mathbf{r}_m^*)^\dagger \mathbf{C}_m (\mathbf{p}^*) \mathbf{r}_m^*}}{(\mathbf{r}_m^*)^\dagger \mathbf{A}_m \mathbf{r}_m^*}, \\ \sum_{m=1}^M p_m^* = P_T. \end{cases} \quad (40)$$

Given $y^*, \{\mathbf{p}^*, \{\mathbf{r}_m^*\}\}$ is clearly optimal for the problem

$$\min_{\mathbf{p} > \mathbf{0}, \{\mathbf{r}_m\}} \sum_{m=1}^M p_m \text{ s.t. } \frac{p_m \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m}} \geq y^*, \quad (41)$$

and $\sum_{m=1}^M p_m^* = P_T$. As a consequence, there exists no better solution $\tilde{y} > y^*$, unless the optimal solution $\tilde{\mathbf{p}}$ to the problem

$$\min_{\mathbf{p} > \mathbf{0}, \{\mathbf{r}_m\}} \sum_{m=1}^M p_m \text{ s.t. } \frac{p_m \mathbf{r}_m^\dagger \mathbf{A}_m \mathbf{r}_m}{\mathbf{r}_m^\dagger \mathbf{B}_m \mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m \mathbf{r}_m}} \geq \tilde{y}, \quad (42)$$

permits to have $\sum_{m=1}^M \tilde{p}_m > P_T$, which, however, contradicts the stationary property in (40). ■

D.2 Generating Feasible Initial Points: Based on the QoS-balancing algorithm, we provide a systematic way to obtain a feasible solution if there exists any, which we detail as follows.

- 1) Initialize the beamforming vectors by choosing standard solutions such as ZF or matched filter (or even randomly generated). If they give a feasible solution, then exit.
- 2) a) Initialize $n = 0$ and the total transmit power P_T^n .
b) With P_T^n , use the QoS-balancing algorithm to find the optimal balanced y^n .
c) If $y^n \geq 1$, then a feasible initial solution has been found and exit. Otherwise, update $n = n + 1$, and increase the power budget by $P_T^{n+1} = \theta P_T^n$ for some preset $\theta > 1$ and then go back to Step 2-b). If P_T^n is unreasonably large but still $y^n < 1$, (14) is declared to be infeasible, which means that in (14), the users' requirements (e.g., SINR, the service probabilities or even the number of users) have to be compromised or the transmission has to be postponed until the channels improve to a better state.

To conclude, as long as the reformulated problem (14) is feasible, we can always find a feasible initial solution using the above approach. Note, however, that (6) and (14) generally have different feasible regions. Therefore, whether or not the original problem (6) is feasible is not indicated by (14) although it is arguable that they should be equivalent in the asymptotic limit (as $M \rightarrow \infty$). More discussion will be given in Section VI.

V. FURTHER IMPROVEMENTS

A. Extraction of Rank-1 Beamforming Vectors

In Appendix B, we describe how the optimal $\{\mathbf{r}_m\}$ can be extracted from the higher-rank solution \mathbf{R}_m . However, it

turns out to be unnecessary to do so at each iteration of the algorithm. This can be seen from the power update (19) that

$$p_m^n = \frac{\text{trace}(\mathbf{B}_m(\mathbf{p}^n)\mathbf{r}_m\mathbf{r}_m^\dagger) + \sqrt{\text{trace}(\mathbf{C}_m(\mathbf{p}^n)\mathbf{r}_m\mathbf{r}_m^\dagger)}}{\text{trace}(\mathbf{A}_m\mathbf{r}_m\mathbf{r}_m^\dagger)} \quad (43)$$

and it depends only on $\mathbf{r}_m\mathbf{r}_m^\dagger$. As a result, the higher-rank solution \mathbf{R}_m obtained from (17) can be used to perform the power update. The rank-1 extraction is only needed if the steady-state beamforming solution is obtained after convergence.

B. Faster Convergence

In Section III-A, the fixed-point update of (19) is adopted. Note that before the algorithm converges, this simple update (19) does not guarantee the optimality of \mathbf{p}^n in terms of total or individual power minimization for fixed $\{\mathbf{r}_m^{n+1}\}$ at each iteration. On the other hand, we note that it is a second-order cone programming (SOCP) problem [27] and can be solved efficiently. Though SOCP can be used to find the optimal power vector, the following method is found to be much more efficient.

For a fixed \mathbf{r}_m , in Step 3 of the n th iteration in the proposed algorithm (see Section IV-A), perform the following update

$$p_m^{n+1}(i+1) = \frac{\mathbf{r}_m^\dagger \mathbf{B}_m(\mathbf{p}^{n+1}(i))\mathbf{r}_m + \sqrt{\mathbf{r}_m^\dagger \mathbf{C}_m(\mathbf{p}^{n+1}(i))\mathbf{r}_m}}{\mathbf{r}_m^\dagger \mathbf{A}_m\mathbf{r}_m} \quad (44)$$

and loop it over i until convergence, where we have used i to denote the inner iteration index, and $\mathbf{p}^{n+1}(0) = \mathbf{p}^n$. Proof of this method to find the optimal power follows from that of [28].

VI. SIMULATION RESULTS

A. Setup

Simulations are conducted to evaluate the performance of the proposed algorithm in i.i.d. MIMO Rayleigh flat-fading channels. The total transmit SNR (defined as $\frac{\sum_m \mathbb{E}[p_m]}{N_0}$) and the per-user transmit SNR (defined as $\frac{\sum_m \mathbb{E}[p_m]}{N_0 M}$), both averaged over many independent channel realizations and channel error conditions, are used as the performance measures. In addition, the users are assumed to have the same target SINR and probability constraint, $\gamma_m = \gamma = 10$ (dB), and $\varepsilon = \varepsilon_m \forall m$.

B. Benchmarks

We have considered the following three benchmarks.

- Optimal power control with ZF beamforming receivers—A straightforward benchmark is a two-step optimization, which first obtains the receive beamforming vectors using ZF based on the estimated CSIR $\{\tilde{\mathbf{h}}_m\}$ and then determines the power by (19) given the ZF vectors.
- Optimal power control with MMSE receivers—This benchmark integrates the power solution (19) with multiuser MMSE beamformers. The MMSE beamforming receivers are aimed at minimizing the expected MSE of

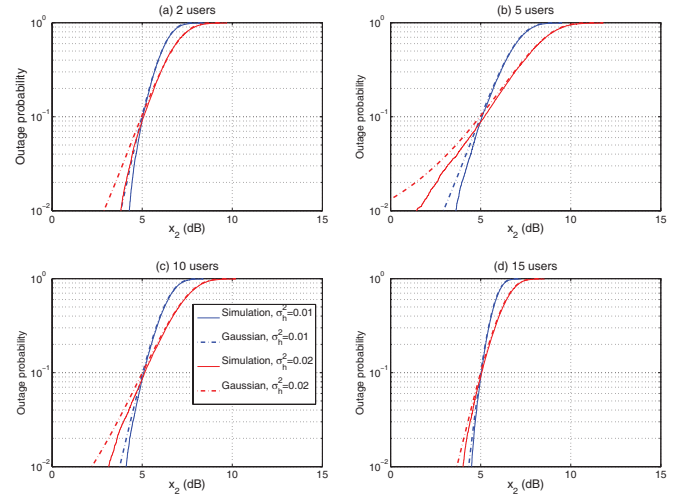


Fig. 2. Comparison of the outage probability of the actual and the Gaussian-approximated x_2 for 2, 5, 10 and 15-user systems using the proposed algorithm.

the received signal over the channel error. Based on (1) and (2), we have

$$\text{MSE}_m = (\mathbf{r}_m^\dagger \mathbf{h}_m - 1)^2 p_m + \sum_{\substack{n=1 \\ n \neq m}}^M |\mathbf{r}_m^\dagger \mathbf{h}_n|^2 p_n + \|\mathbf{r}_m\|^2 N_0 \quad (45)$$

and its expectation given by

$$\mathbb{E}_{\Delta h_m}[\text{MSE}_m] = \mathbf{r}_m^\dagger \left[\left(\sum_{n=1}^M \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_h^2 \mathbf{I} \right) p_n + N_0 \mathbf{I} \right] \mathbf{r}_m - 2p_m \mathbf{r}_m^\dagger \tilde{\mathbf{h}}_m - 2p_m \tilde{\mathbf{h}}_m^\dagger \mathbf{r}_m + 1. \quad (46)$$

Therefore, the MMSE receiver for a given \mathbf{p} is found as

$$\mathbf{r}_m = \varsigma \left[\sum_{n=1}^M \left(\tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_n^2 \mathbf{I} \right) p_n + N_0 \mathbf{I} \right]^{-1} \tilde{\mathbf{h}}_m, \quad (47)$$

where ς is chosen to ensure $\|\mathbf{r}_m\| = 1$.

This receiver structure depends upon \mathbf{p} , which makes the joint optimization complicated. In particular, we use the method in [13] to decouple the power control and the beamforming design by removing their dependence, so

$$\mathbf{r}_m = \varsigma \left[\sum_{n=1}^M \left(\tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_n^2 \mathbf{I} \right) + N_0 \mathbf{I} \right]^{-1} \tilde{\mathbf{h}}_m. \quad (48)$$

The modified MMSE receivers balance both the interference and estimation error, as opposed to the ZF receivers that handle the interference based on the estimated CSIR. The robust power allocation is also found by (19).

- Optimal power control with matched filtering receivers—Another possible solution is to use the matched filtering receivers in accordance with the estimated CSIR $\{\tilde{\mathbf{h}}_m\}$, and then apply the power allocation solution (19) with the matched filters. Nevertheless, in our simulations, we have never found any feasible power allocation given these receivers, indicating that it is utterly inappropriate to use

$$\tilde{\mathbf{H}}_{15} = \begin{bmatrix} 0.9431 - 1.3557i & 0.2708 + 1.1821i & 0.0045 + 0.2550i & 0.8760 + 0.5882i & -0.7468 + 0.4918i \\ -0.6740 + 0.2229i & -0.3033 + 0.2973i & -0.2186 + 0.4553i & -1.1853 - 0.0275i & -0.3517 + 0.4143i \\ 1.3621 + 0.6679i & -0.5634 + 0.5462i & 0.3538 - 1.0096i & 0.0792 - 0.4169i & 0.6617 + 0.5530i \\ -0.1495 + 0.9110i & 0.2966 - 0.5430i & 1.2275 - 0.1438i & -0.3575 + 0.6727i & 0.0064 - 1.0244i \\ -0.0124 + 0.5407i & 0.4324 + 0.3762i & -1.3414 - 0.2395i & -0.7082 - 0.2501i & 1.0989 - 0.5308i \\ 1.2977 - 0.5750i & 0.8983 - 1.0918i & -0.1065 - 0.5158i & -0.7037 + 0.8076i & -1.4050 + 0.5962i \\ 0.9500 + 0.1299i & 0.1112 - 0.8081i & -0.8161 - 0.5118i & 1.0693 - 0.3965i & -1.0864 + 0.8649i \\ -0.2017 - 0.6983i & -0.2032 + 0.1112i & 0.4924 + 1.3254i & 0.1921 + 0.1889i & -0.8103 + 0.5954i \\ 1.4123 + 0.7907i & -0.5359 - 1.3770i & -0.4949 - 0.8270i & 0.2337 - 0.6899i & 0.9965 - 0.5609i \\ -0.6075 - 0.2792i & 0.3006 - 0.3623i & 0.5273 - 0.1718i & -0.2368 - 1.2584i & -0.7211 + 0.5067i \\ -0.9675 - 0.1169i & -0.4242 + 0.8797i & -0.1928 - 0.2037i & 0.7291 + 0.8292i & -1.3508 - 0.0603i \\ 0.6420 - 0.1421i & 1.4356 - 0.4393i & -1.1321 - 0.5457i & 0.6662 - 0.1388i & -0.4982 + 0.8212i \\ -0.7088 + 1.1305i & -1.0695 - 0.2432i & 0.5166 - 0.1034i & 1.7100 - 0.1062i & 0.3129 + 0.6364i \\ 0.2023 + 0.2395i & -0.2328 - 0.9612i & -0.6651 - 1.3290i & -0.0221 + 1.4073i & -0.5343 + 0.2850i \\ 0.6790 + 0.2851i & 0.0204 + 0.1354i & -0.4058 - 0.3762i & 0.4700 - 0.2367i & 0.7614 - 0.1942i \\ -0.0625 - 0.0377i & 0.4782 - 0.4646i & -0.1957 + 0.4587i & -1.1319 - 1.0807i & 0.8706 - 0.8651i \\ 0.0312 - 1.4945i & 0.0904 - 0.6047i & -0.4896 + 0.2600i & 0.0405 + 1.2697i & 0.3250 - 0.1171i \\ -0.4916 + 0.2126i & 0.0059 + 1.3573i & -0.3633 + 0.8973i & -0.4904 - 0.6986i & -0.0842 + 0.4037i \\ -1.5534 + 0.3251i & -0.2069 - 0.2642i & -0.1282 + 0.4852i & -0.2299 + 1.6185i & -0.4848 - 0.2782i \\ 0.3809 + 0.1927i & -0.0954 + 1.0633i & -0.2399 - 0.6717i & 0.3467 - 0.7036i & 0.5993 + 0.1586i \\ -0.1395 - 0.8255i & -1.3769 - 0.5799i & -1.1823 - 0.9877i & 1.2967 - 0.8031i & 0.3276 - 0.8607i \\ 0.5493 - 1.1303i & 1.4493 - 1.7355i & -0.8690 + 0.0866i & -0.0743 - 0.6989i & -0.5922 + 0.8139i \\ -0.6674 - 0.8508i & 0.5501 - 0.1570i & 0.6852 - 1.5035i & -0.0456 + 0.1122i & -1.5837 + 0.6927i \\ 0.1027 - 0.6696i & -0.7018 - 0.6746i & -0.9971 + 0.8039i & 0.5193 + 0.4255i & -0.5817 - 0.2707i \\ -0.4619 + 1.2445i & 0.2241 + 0.1487i & -0.6598 + 1.1188i & -0.2984 + 0.2346i & -0.0694 - 0.5524i \\ -1.0336 + 0.7366i & 1.0809 + 0.2206i & -0.7389 + 0.2395i & 0.4590 - 0.3300i & 0.8135 - 0.0789i \\ -1.0332 + 0.1506i & -0.4176 + 0.8503i & -0.0291 - 0.4014i & -0.1787 + 0.4023i & 0.6017 - 0.4154i \\ -0.6661 - 0.7391i & 0.1445 + 0.6024i & -0.1734 + 0.5628i & 1.0679 - 1.0498i & -0.7523 - 0.8128i \\ -1.1304 - 0.8767i & -0.5853 - 0.3580i & -0.2376 + 0.3195i & 0.2630 + 0.0659i & 0.4919 + 0.3365i \\ 0.0350 + 1.5216i & 0.0030 + 0.0993i & 0.5496 + 0.5087i & -0.4131 - 0.0777i & 0.6070 - 0.8350i \\ -0.6044 - 0.6444i & 0.4888 - 0.9344i & -0.8595 + 0.6919i & 1.2401 - 0.7191i & -0.6088 - 0.5721i \\ -1.0000 - 1.3230i & -1.3409 + 0.0698i & 0.4299 - 0.6144i & -1.0547 - 0.7532i & 0.3427 + 0.2370i \\ 0.1340 + 0.1221i & 0.3820 + 0.6691i & -0.5722 - 0.3445i & 0.1004 + 0.0575i & -0.2418 - 0.0390i \\ -0.9731 - 0.1409i & 0.7623 - 0.3350i & -0.3765 + 0.3331i & -0.4730 - 1.0103i & 1.4052 + 0.2440i \\ -0.1999 - 0.5304i & -0.1878 + 0.0885i & 1.1668 - 0.5637i & 0.9502 - 0.7933i & -0.8923 + 1.1790i \\ 0.4093 - 0.0255i & 0.5975 + 0.1563i & -0.0153 - 0.0182i & 1.1464 + 0.3123i & -0.7579 + 0.7351i \\ -0.4971 - 0.4329i & -0.1250 - 0.1695i & -0.1552 + 0.5911i & 0.6383 + 0.4633i & -0.8433 - 0.0297i \\ 0.0893 + 0.0162i & -0.9801 + 0.3841i & -0.8618 + 0.8134i & -0.4123 - 0.4427i & 0.7184 + 0.1941i \\ -0.9496 - 0.3268i & -0.5094 - 0.2172i & 0.9089 + 0.5180i & 1.0674 - 1.2402i & -1.7653 + 0.0378i \\ 0.6252 + 0.9565i & 0.1116 - 0.6897i & -0.2026 + 1.7557i & -1.0443 - 0.9831i & -0.5372 - 0.0675i \\ -0.8533 + 0.7132i & -1.0939 - 0.1540i & -1.8381 - 0.9031i & 0.1751 + 1.0755i & 0.7877 + 0.8211i \\ -0.5100 - 1.4811i & 0.2825 + 0.5899i & -0.3330 - 0.0522i & 0.8729 + 0.1501i & 0.1068 + 0.1696i \\ -0.1196 + 1.0050i & 0.2121 - 0.3101i & 0.3144 + 0.4263i & 0.3170 - 0.3591i & -0.6796 - 0.7659i \\ -0.1975 + 0.7900i & -0.6318 + 1.5371i & -0.6784 + 0.2430i & -0.1772 - 0.8372i & 0.2544 - 0.4947i \\ -0.0554 - 0.4421i & 0.4493 + 0.1394i & -0.2974 + 0.4311i & -0.3095 + 0.1780i & 0.1767 + 0.6998i \end{bmatrix}$$

Fig. 3. The matrix $\tilde{\mathbf{H}}_{15}$.

purely single-user receivers without concerning the inter-user interferences or channel errors. Because of that, no numerical results are shown for this benchmark.

- Non-robust beamforming (NRB)—In this case, the beamforming vectors and power allocation are optimized based on the estimated CSIR without concerning the CSI errors.

C. Results

Figure 2 shows the impact of Gaussian approximation on the outage probability constraints. In particular, this is assessed by the simulated outage probabilities and the Gaussian approximated results against x_2 for 2, 5, 10 and 15-user SIMO systems with target SINR $\gamma_m = 5$ (dB) and outage requirement 10^{-2} . In this figure, the CDF results over the random channel error statistics are plotted for a given channel realization $\tilde{\mathbf{H}}_{\text{all-user}} \triangleq [\tilde{\mathbf{h}}_1 \cdots \tilde{\mathbf{h}}_M] \equiv \tilde{\mathbf{H}}_{15}(1 : M, 1 : M)$, which is given in Figure 3. Results for both $\sigma_h^2 = 0.01$ and $\sigma_h^2 = 0.02$ are shown to cover a wide range of channel estimation conditions.

As we can see, the CDF of $\sigma_h^2 = 0.01$ is sharper than that of $\sigma_h^2 = 0.02$ and this is because $\sigma_h^2 = 0.01$ leads to a smaller variance on x_2 , as is anticipated from (9b). Besides, it is observed that for small $\sigma_h^2 = 0.01$, the Gaussian approximated probability matches very well to the exact one, even if the number of users is as small as 2. For a larger $\sigma_h^2 = 0.02$, this agreement continues to be seen for 15-user systems. Importantly, for all the results we show, the outage probability is always guaranteed to be less than the required value 10^{-2} . As a final remark, we are indeed not expecting to see the improving accuracy of Gaussian approximation with the number of users in the figures, as the results are limited to

TABLE I
SERVICE PROBABILITY FOR 10-USER SIMO SYSTEMS.

	$\varepsilon = 0.85$		$\varepsilon = 0.99$		$\varepsilon = 0.9999$	
	Proposed	NRB	Proposed	NRB	Proposed	NRB
User 1	0.8631	0.0231	0.9831	0.0062	0.9989	0.0464
User 2	0.8579	0.0032	0.9816	0.0094	0.9980	0.0226
User 3	0.8612	0.0253	0.9792	0.0375	0.9987	0.0208
User 4	0.8599	0.0194	0.9843	0.0484	0.9983	0.0283
User 5	0.8571	0.0055	0.9879	0.0308	0.9996	0.0869
User 6	0.8629	0.0066	0.9825	0.0548	0.9981	0.0182
User 7	0.8629	0.0147	0.9817	0.0282	0.9980	0.0093
User 8	0.8631	0.0078	0.9825	0.0675	0.9987	0.0286
User 9	0.8589	0.0669	0.9837	0.0416	0.9990	0.0415
User 10	0.8638	0.03710	0.9797	0.0325	0.9982	0.0178

some particular channel realizations $\tilde{\mathbf{H}}_{15}$, though the tightness should improve in the statistical sense (if the effect of the channel realization is averaged out) from CLT.

Results in Table I provide the achieved service (or non-outage) probability for the proposed algorithm with 10 users and $\sigma_h^2 = 0.01$, when the probability requirements at SINR of 10 (dB) are $\varepsilon = 0.85, 0.99, 0.9999$. The simulation results were obtained by averaging over 10^6 independent channel realizations in the presence of channel errors. As can be observed, the probability constraints are closely met by the proposed algorithm whereas the non-robust approach attains only a very low service probability ($< 9\%$), which reveals the importance of considering the channel error statistics for beamforming and power allocation optimization for QoS provision.

The convergence of the proposed algorithm is illustrated by the results in Figure 4, which we show for a 10-user system. Results verify that the algorithm converges, in both the total transmit SNR and the individual users' SINR. Remarkably,

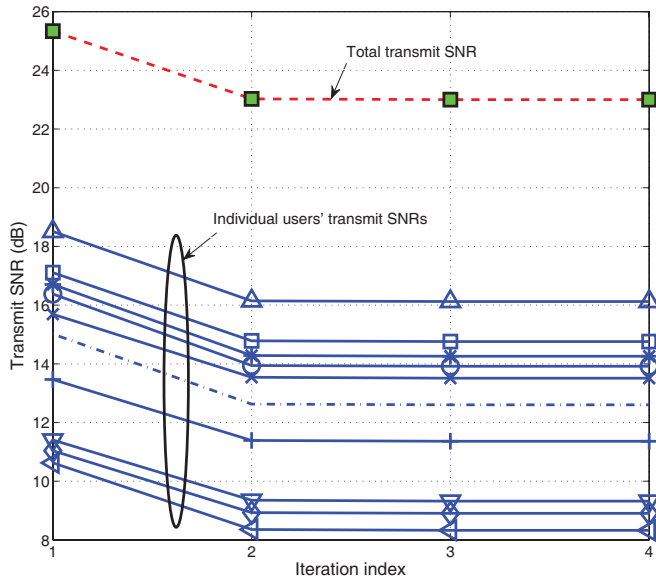


Fig. 4. Convergence behavior of the proposed algorithm for a 10-user system.

convergence is typically achieved in just a few iterations. Figure 5 studies the impacts of the channel error variance and the target SINR on the total transmit SNR against the service probability requirements for the proposed system with 10 users. In particular, the proposed algorithm and the optimal power control with ZF and MMSE beamforming receivers are compared. First, as expected, the required transmit SNR increases with the target SINR, the service probability and the channel error variance. In addition, there is a considerable SNR gap between the proposed algorithm and the ZF approach, which can be as large as 10 (dB) if the target SINR is 7 (dB), $\sigma_h^2 = 0.02$ and $\varepsilon = 0.99$, demonstrating the importance of joint receiver beamforming and power allocation optimization. We also note that for the case with $\varepsilon = 0.99$ and $\sigma_h^2 = 0.02$, the required transmit SNR for the ZF systems is nearly 30 (dB) while the proposed algorithm needs only slightly under 20 (dB). In addition, the performance of robust power control with MMSE receivers is also shown with the target SINR of 7 (dB) and $\varepsilon = 0.8 \sim 0.95$ (other results are not available due to the difficulty of finding feasible channel realizations). Results illustrate that for relatively small ε , it has about 1 (dB) gain as compared to the system with ZF receivers, while this gain vanishes and may require even more SNR than the ZF system if the channel error is severe.

In Figure 6, we study the impact of channel spatial correlation on the system performance. We assume that the (i, j) th entry of the channel correlation matrix is chosen to be $\rho^{|i-j|}$ with $\rho = 0.4$. Compared the results in Figures 6 & 5, it is easily seen that with channel correlation, more transmit SNR is required to attain the required individual users' QoS performance. Another observation we have is that when there exists channel correlation, there is a detrimental effect on the system feasibility for each particular channel realization. In addition, from the results in both Figures 5 and 6, it can be concluded that the proposed algorithm is much less sensitive to the channel estimation errors, as compared to the ZF and MMSE approaches.

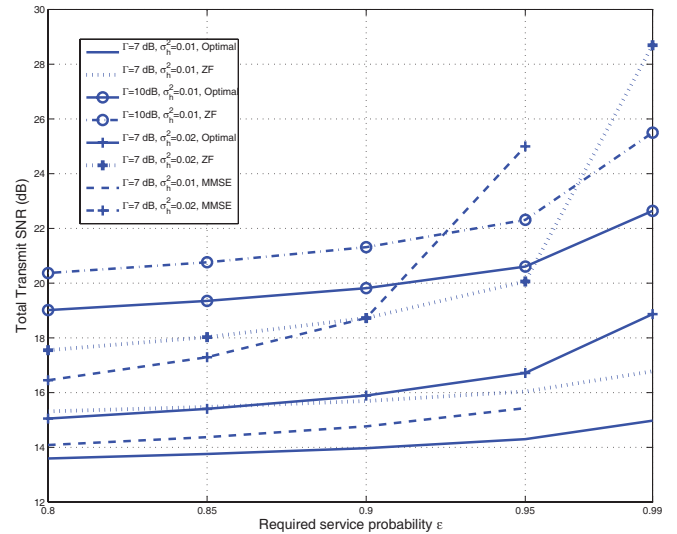


Fig. 5. Impacts of channel error variance, SINR targets and service probability requirements on the average total transmit SNR.

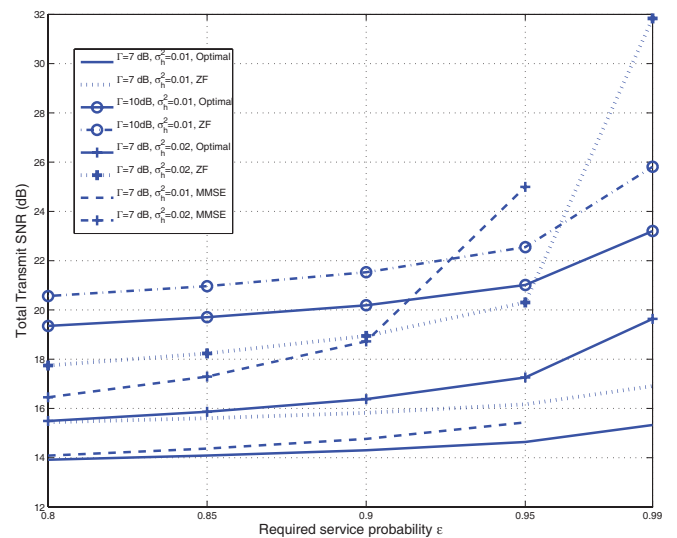


Fig. 6. Impacts of channel error variance, SINR targets and service probability requirements on the average total transmit SNR with channel spatial correlation.

Finally, the average transmit SNR per user results are provided for 2-user, 5-user, 10-user and 15-user systems in Figure 7. Results illustrate that the average transmit SNR per user increases with the target SINR and the channel error variance but decreases with the number of users, which indicates that multiuser diversity is in use to lower the required transmit SNR.

VII. CONCLUSION

This paper has investigated the robust beamforming design in the multiuser SIMO uplink with CSIR uncertainties at the base station receiver. The objective is to minimize each mobile user's transmit power by jointly optimizing the power allocation and the receive beamforming vectors subject to individual users' probability constraints on the output SINRs. Regarding the SMI as Gaussian, we proposed an iterative algorithm

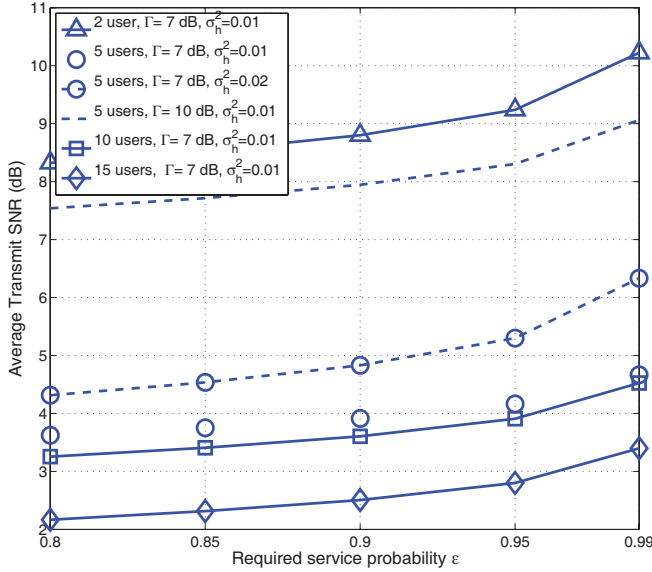


Fig. 7. Average transmit SNR per user against the required service probability of the proposed algorithm for various settings.

which is proved to converge to the globally optimal solution. Results have shown that the proposed method outperforms the power-only allocation with ZF and MMSE beamforming vectors significantly.

APPENDIX A: MEAN AND VARIANCE OF x_m

Define $t_m \triangleq p_m - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n$. Then, we have

$$\mu_{x_m} = \mathbf{r}_m^\dagger \left(p_m \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger \right) \mathbf{r}_m + \sigma_h^2 t_m \quad (49)$$

and (50) (see top of next page), which can be simplified to

$$\sigma_{x_m}^2 = \mathbf{r}_m^\dagger \left[\begin{array}{c} (2\sigma_h^2 \tilde{\mathbf{h}}_m \tilde{\mathbf{h}}_m^\dagger + \sigma_h^4 \mathbf{I}) p_m^2 \\ + \gamma_m^2 \sum_{\substack{n=1 \\ n \neq m}}^M (2\sigma_h^2 \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger + \sigma_h^4 \mathbf{I}) p_n^2 \end{array} \right] \mathbf{r}_m, \quad (51)$$

where $\|\mathbf{r}_m\|^2 = 1$ has been used and

$$\mathbf{E} \left[\mathbf{r}_m^\dagger \tilde{\mathbf{h}}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \mathbf{r}_m^\dagger \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \right] = 0, \quad (52)$$

$$\mathbf{E} \left[\mathbf{r}_m^\dagger \tilde{\mathbf{h}}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \mathbf{r}_m^\dagger \Delta \mathbf{h}_n \tilde{\mathbf{h}}_n^\dagger \mathbf{r}_m \right] = \sigma_h^2 \mathbf{r}_m^\dagger \tilde{\mathbf{h}}_n \tilde{\mathbf{h}}_n^\dagger \mathbf{r}_m, \quad (53)$$

$$\mathbf{E} \left[\mathbf{r}_m^\dagger \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \mathbf{r}_m^\dagger \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \right] = 2\sigma_h^4. \quad (54)$$

Note that $\mathbf{W}_n \triangleq \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger$ obeys the complex Wishart Distribution, i.e., $\mathbf{W}_n \sim \mathcal{CW}(\sigma_h^2 \mathbf{I}, 1)$. Therefore, (53) and (54) have been simplified by

$$\begin{aligned} \mathbf{E} \left[\Delta \mathbf{h}_n^\dagger \mathbf{r}_m \mathbf{r}_m^\dagger \Delta \mathbf{h}_n \right] &= \mathbf{E}[\text{trace}(\mathbf{r}_m \mathbf{r}_m^\dagger \mathbf{W}_n)] \\ &= \mathbf{r}_m^\dagger \text{trace}(\mathbf{E}[\mathbf{W}_n]) \mathbf{r}_m, \end{aligned} \quad (55)$$

$$\mathbf{E} \left[\mathbf{r}_m^\dagger \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \mathbf{r}_m^\dagger \Delta \mathbf{h}_n \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \right] = \mathbf{E} \left[\mathbf{r}_m^\dagger \mathbf{W}_n \mathbf{r}_m \mathbf{r}_m^\dagger \mathbf{W}_n \mathbf{r}_m \right], \quad (56)$$

which follow from the useful expectations for complex Wishart matrix $\mathbf{W}_n \sim \mathcal{CW}(\Sigma, k)$ in [29]:

$$\begin{cases} \mathbf{E}[\mathbf{W}_n] = k\Sigma, \\ \mathbf{E}[\mathbf{W}_n \mathbf{R} \mathbf{W}_n] = k^2 \Sigma \mathbf{R} \Sigma + k \text{trace}(\mathbf{R} \Sigma) \Sigma. \end{cases} \quad (57)$$

APPENDIX B: PROOF OF THEOREM 1

Note that part of the proof is based on the previous work with two inequality constraints in [30].

B.1 The Tightness of the SDP Relaxation

First note that (17) can be rewritten as

$$\max_{g, \mathbf{r}} g \quad \text{s.t.} \quad -\mathbf{r}^\dagger \mathbf{D} \mathbf{r} - g \geq 0, \quad \mathbf{r}^\dagger \mathbf{r} = 1, \quad \mathbf{r}^\dagger \mathbf{C} \mathbf{r} = t. \quad (58)$$

Using the extended S-Lemma in [25], [26], the following two claims are equivalent:

- (1) $-\mathbf{r}^\dagger \mathbf{D} \mathbf{r} - g \geq 0, \forall \mathbf{r} \in \mathbb{C}^{n_R \times 1}$ such that $\mathbf{r}^\dagger \mathbf{r} = 1$ and $\mathbf{r}^\dagger \mathbf{C} \mathbf{r} = t$.
- (2) There exist real λ_1, λ_2 , such that

$$\begin{bmatrix} -g + \lambda_1 + \lambda_2 t & \mathbf{0} \\ \mathbf{0} & -\mathbf{D} - \lambda_1 \mathbf{I} - \lambda_2 \mathbf{C} \end{bmatrix} \succeq \mathbf{0}. \quad (59)$$

As a result, (58) becomes a semi-definite problem

$$\max_{\lambda_1, \lambda_2, g} g \quad \text{s.t.} \quad \begin{bmatrix} -g + \lambda_1 + \lambda_2 t & \mathbf{0} \\ \mathbf{0} & -\mathbf{D} - \lambda_1 \mathbf{I} - \lambda_2 \mathbf{C} \end{bmatrix} \succeq \mathbf{0}. \quad (60)$$

Because of the fact that (17) and (60) are equivalent, they have the same objective value.

Now, consider the SDP relaxation of (17)

$$\min_{\mathbf{R} \succeq \mathbf{0}} -\text{trace}(\mathbf{D} \mathbf{R}) \quad \text{s.t.} \quad \text{trace}(\mathbf{R}) = 1, \quad \text{trace}(\mathbf{C} \mathbf{R}) = t. \quad (61)$$

Because \mathbf{R} is not necessarily rank-1, (61) will in general yield a lower bound for (17). However, we are going to show next that (61) is indeed the dual of (60).

Let the dual variable of (60) be

$$\mathbf{Q}_0 = \begin{bmatrix} Q_{11} & \times \\ \times & \mathbf{Q} \end{bmatrix} \succeq \mathbf{0} \in \mathbb{C}^{(N_r+1) \times (N_r+1)}, \quad (62)$$

where $Q_{11} \geq 0$ is the (1, 1)th element of \mathbf{Q}_0 and $\mathbf{Q} \succeq \mathbf{0}$. The Lagrangian is given by

$$\begin{aligned} L(\lambda, s_1, s_2, \mathbf{Q}_0) &= -\lambda - \text{trace} \left(\begin{bmatrix} -\lambda + s_1 + s_2 t & \mathbf{0} \\ \mathbf{0} & -\mathbf{D} - s_1 \mathbf{I} - s_2 \mathbf{C} \end{bmatrix} \mathbf{Q}_0 \right) \\ &= -\lambda - Q_{11}(-\lambda + s_1 + s_2 t) + \text{trace}(\mathbf{D} \mathbf{Q}) \\ &\quad + s_1 \text{trace}(\mathbf{Q}) + s_2 \text{trace}(\mathbf{C} \mathbf{Q}) \\ &= \lambda(Q_{11} - 1) + \text{trace}(\mathbf{D} \mathbf{Q}) + s_1(\text{trace}(\mathbf{Q}) - Q_{11}) \\ &\quad + s_2(\text{trace}(\mathbf{C} \mathbf{Q}) - t Q_{11}). \end{aligned} \quad (63)$$

The dual function is

$$\begin{aligned} g(\mathbf{Q}_0) &= \inf_{\lambda, s_1, s_2} L(\lambda, s_1, s_2, \mathbf{Q}_0) \\ &= \inf_{\lambda, s_1, s_2} \left[\lambda(Q_{11} - 1) + \text{trace}(\mathbf{D} \mathbf{Q}) \right. \\ &\quad \left. + s_1(\text{trace}(\mathbf{Q}) - Q_{11}) + s_2(\text{trace}(\mathbf{C} \mathbf{Q}) - t Q_{11}) \right]. \end{aligned} \quad (64)$$

$$\sigma_{x_m}^2 = \mathbf{E} \left[\begin{array}{c} p_m \left(\mathbf{r}_m^\dagger \tilde{\mathbf{h}}_m^\dagger \Delta \mathbf{h}_m^\dagger \mathbf{r}_m + \mathbf{r}_m^\dagger \Delta \mathbf{h}_m^\dagger \tilde{\mathbf{h}}_m^\dagger \mathbf{r}_m + \mathbf{r}_m^\dagger \Delta \mathbf{h}_m^\dagger \Delta \mathbf{h}_m^\dagger \mathbf{r}_m \right) \\ - \gamma_m \sum_{\substack{n=1 \\ n \neq m}}^M p_n \left(\mathbf{r}_m^\dagger \tilde{\mathbf{h}}_n^\dagger \Delta \mathbf{h}_n^\dagger \mathbf{r}_m + \mathbf{r}_m^\dagger \Delta \mathbf{h}_n^\dagger \tilde{\mathbf{h}}_n^\dagger \mathbf{r}_m + \mathbf{r}_m^\dagger \Delta \mathbf{h}_n^\dagger \Delta \mathbf{h}_n^\dagger \mathbf{r}_m \right) \end{array} \right]^2 - \sigma_h^4 t_m^2 \quad (50)$$

Apparently, $g(\mathbf{Q}_0)$ is unbounded unless

$$Q_{11} = 1, \text{trace}(\mathbf{Q}) = Q_{11} = 1, \text{and } \text{trace}(\mathbf{CQ}) = tQ_{11}. \quad (65)$$

Therefore, the following dual problem of (60) is obtained:

$$\min_{\mathbf{Q} \succeq \mathbf{0}} -\text{trace}(\mathbf{DQ}) \text{ s.t. } \text{trace}(\mathbf{Q}) = 1, \text{trace}(\mathbf{CQ}) = t, \quad (66)$$

which is exactly (61). Because both (61) and (60) are convex and feasible, the strong duality holds and they attain the same objective value. As a consequence, (17) and (61) also have the same objective value, which means that the relaxation problem (61) is exact and importantly, it has at least one rank-1 solution.

B.2 The Recovery of the Exact Rank-1 Solution

\mathbf{R} can be easily found by solving (61) but it is not necessarily rank-1. The remaining challenge is to find the optimal \mathbf{r} based on the solution of the dual problem of (61), derived as

$$\max_{\lambda_1, \lambda_2} -\lambda_1 - \lambda_2 t \text{ s.t. } \lambda_1 \mathbf{I} + \lambda_2 \mathbf{C} - \mathbf{D} \succeq \mathbf{0}. \quad (67)$$

The Karush-Kuhn-Tucker condition, which is both sufficient and necessary for the optimality of \mathbf{R} in (61) is

$$(\lambda_1 \mathbf{I} + \lambda_2 \mathbf{C} - \mathbf{D})\mathbf{R} = \mathbf{0}, \text{trace}(\mathbf{R}) = 1, \text{trace}(\mathbf{CR}) = t. \quad (68)$$

As such, the sufficient and necessary condition for the optimality of the rank-1 solution \mathbf{r} is

$$\mathbf{r}^\dagger (\lambda_1 \mathbf{I} + \lambda_2 \mathbf{C} - \mathbf{D})\mathbf{r} = 0, \mathbf{r}^\dagger \mathbf{C}\mathbf{r} = t, \text{ and } \|\mathbf{r}\| = 1. \quad (69)$$

In the following, \mathbf{r} that satisfies the conditions in (69) will be found. Suppose \mathbf{E} forms a basis for the null space of $\lambda_1 \mathbf{I} + \lambda_2 \mathbf{C} - \mathbf{D}$, then \mathbf{r} should have the structure $\mathbf{r} = \mathbf{E}\mathbf{w}$, where \mathbf{w} is a vector to be determined. Using (69), it follows that

$$\mathbf{w}^\dagger \mathbf{E}^\dagger \mathbf{C}\mathbf{E}\mathbf{w} = t. \quad (70)$$

Suppose $\mathbf{E}^\dagger \mathbf{C}\mathbf{E}\mathbf{w}_0 = \lambda_{\min} \mathbf{w}_0$ and $\mathbf{E}^\dagger \mathbf{C}\mathbf{E}\mathbf{w}_1 = \lambda_{\max} \mathbf{w}_1$, in which λ_{\max} and λ_{\min} are the maximal and minimal eigenvalues of $\mathbf{E}^\dagger \mathbf{C}\mathbf{E}$, respectively, and \mathbf{w}_0 and \mathbf{w}_1 are the corresponding eigenvectors. Without loss of generality, we can assume

$$\mathbf{w} = \sin(\theta) \mathbf{w}_0 + \cos(\theta) \mathbf{w}_1. \quad (71)$$

Substituting this into (70) gives

$$\mathbf{w}^\dagger \mathbf{E}^\dagger \mathbf{C}\mathbf{E}\mathbf{w} = \sin^2(\theta) \lambda_{\min} + \cos^2(\theta) \lambda_{\max} = t, \quad (72)$$

which has the solution

$$\sin^2(\theta) = \frac{\lambda_{\max} - t}{\lambda_{\max} - \lambda_{\min}}, \quad \cos^2(\theta) = \frac{t - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}. \quad (73)$$

As a result, the desired rank-1 solution is

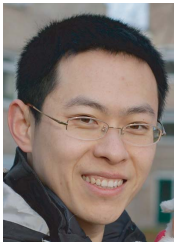
$$\mathbf{r} = \mathbf{E} \left(\sqrt{\frac{\lambda_{\max} - t}{\lambda_{\max} - \lambda_{\min}}} \mathbf{w}_0 + \sqrt{\frac{t - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}} \mathbf{w}_1 \right). \quad (74)$$

Note that this solution requires $\lambda_{\max} > \lambda_{\min}$. If $\lambda_{\max} = \lambda_{\min}$, \mathbf{w} is chosen to be either \mathbf{w}_0 or \mathbf{w}_1 . Up to now, we have provided a complete solution for (15). The main computational part is to find the solution of (67), which involves 2 scalar variables and its computation complexity is about $\mathcal{O}(n_R^{6.5})$ [31] per accuracy digit. The software package, SeDuMi [32], is available to solve the SDP problem and there is also a very convenient interface called YALMIP [33] for easy implementation.

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Tech. Rep., AT&T Bell Labs, 1995.
- [2] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.
- [3] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.
- [4] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875-1892, Sept. 2004.
- [5] F. R. Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1437-1450, Oct. 1998.
- [6] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook Antennas Wireless Commun.*, Aug. 2001.
- [7] Z. G. Pan, K. K. Wong, and T. S. Ng, "Generalized multiuser orthogonal space division multiplexing," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1-5, Nov. 2004.
- [8] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203-2214, May 2006.
- [9] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 35, no. 10, pp. 1365-1376, Oct. 1987.
- [10] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 386-398, Feb. 1987.
- [11] B. K. Chalise and A. Czylik, "Robust uplink beamforming based upon minimum outage probability criterion," in *Proc. IEEE Global Telecom. Conf.*, vol. 6, pp. 3974-3978, Nov. 2004, Dallas, TX, USA.
- [12] V. K. N. Lau and M. L. Jiang, "Downlink scheduling and rate adaptation design of multi-user, multiple-antenna base station with imperfect CSIT," in *Proc. IEEE Global Telecommun. Conf.*, vol. 5, pp. 5, Nov. 2005, St. Louis, MO, USA.
- [13] V. K. N. Lau and M. L. Jiang, "Rate quantization and cross-layer design of multiple-antenna base stations with transmit MMSE and imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2952-2960, Aug. 2007.
- [14] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: a probabilistically constrained approach," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1560-1570, Aug. 2006.
- [15] G. Zheng, K. K. Wong, and T. S. Ng, "Convergence analysis of downlink MIMO antenna systems using second-order cone programming," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, pp. 492-496, Sept. 2005, Dallas, TX, USA.
- [16] S. Kandukuri and S. Boyd, "Optimal power control in interference limited fading wireless channels with outage probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46-55, Jan. 2002.
- [17] J. Papandriopoulos, J. Evans, and S. Dey, "Optimal power control for Rayleigh-faded multiuser systems with outage constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2705-2715, Nov. 2005.
- [18] J. Papandriopoulos, J. Evans, and S. Dey, "Outage-based optimal power control for generalized multiuser fading channels," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 693-703, Apr. 2006.

- [19] S. Ulukus and R. Yates, "Adaptive power control and MMSE interference suppression," *ACM Wireless Networks*, vol. 4, no. 6, pp. 489-496, 1998.
- [20] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.
- [21] T. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. Allerton Conf. Commun., Control and Computing*, pp. 958-966, Sept. 1999, Monticello, IL, USA.
- [22] J. Cheng and T. Berger, "On the difference of two sums of independent generalized gamma random variables with applications to error performance analysis and outage probability evaluation," in *Proc. IEEE Int. Sym. Inform. Theory*, pp. 39, June 2003, Yokohama, Japan.
- [23] M. K. Simon and M. S. Alouini, "On the difference of two chi-square variates with application to outage probability computation," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1946-1954, Nov. 2001.
- [24] C. D. Perttunen, D. Jones, and B. Stuckman, "Lipschitzian optimization without the lipschitz constant," *J. Opt. Theory App.*, vol. 79, no. 1, pp. 157-181, Oct. 1993.
- [25] A. L. Fradkov and V. A. Yakubovich, "The S-Procedure and the duality relation in convex quadratic programming problems," *Vestnik Leningrad Univ Math*, pp. 155: 81-87, Oct. 1973.
- [26] I. Pólik and T. Terlaky, "A survey of the S-Lemma," to appear in *SIAM Review*.
- [27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [28] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1341-1348, Sept. 1995.
- [29] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Systems Signal Processing*, vol. 5, no. 4, pp. 263-279, July 1994.
- [30] A. Beck and Y. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 3, pp. 844-860, 2006.
- [31] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, ser. MPS-SIAM Series on Optimization, 2001.
- [32] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods Software*, vol. 11-12, pp. 625-653, Aug. 1999.
- [33] J. Löfberg, "YALMIP: a toolbox for modeling and optimization in MATLAB," in *Proc. Int. Sym. Computer Aided Control Systems Design*, pp. 284-289, Sept. 2004, Taipei, Taiwan.



Gan Zheng (S'05-M'09) received the BEng and the MEng degrees from Tianjin University, Tianjin, China, in 2002 and 2004, respectively, both in Electronic and Information Engineering, and the PhD degree in Electrical and Electronic Engineering from The University of Hong Kong, Hong Kong, in 2008. He is now a research associate at the Adastral Park Research Campus, University College London. His research interests center around the general area of wireless communications, with emphasis on beamforming design and robust optimization in multiuser

MIMO systems, cooperative network and cognitive radio systems.



Kai-Kit Wong (S'98-M'01-SM'08) received the BEng, the MPhil, and the PhD degrees, all in Electrical and Electronic Engineering, from the Hong Kong University of Science and Technology, Hong Kong, in 1996, 1998, and 2001, respectively. After graduation, he joined the Department of Electrical and Electronic Engineering, the University of Hong Kong as a Research Assistant Professor. From July 2003 to December 2003, he visited the Wireless Communications Research Department of Lucent Technologies, Bell-Labs, Holmdel, NJ, U.S. where

he was a Visiting Research Scholar studying optimization in broadcast MIMO channels. After that, he then joined the Smart Antennas Research Group of Stanford University as a Visiting Assistant Professor conducting research on overloaded MIMO signal processing. From 2005 to August 2006, he was with the Department of Engineering, the University of Hull, U.K., as a Communications Lecturer. Since August 2006, he has been with University College London Adastral Park Research Campus where he is a Senior Lecturer.

His current research interests center around MIMO antenna systems, cross-layer optimization in wireless multimedia networks, cognitive radio, cooperative communications and secrecy capacity analysis. Dr. Wong is a senior member of IEEE and is also on the editorial board of *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*. Dr. Wong won the IEEE Vehicular Technology Society Japan Chapter Award of the International IEEE Vehicular Technology Conference-Spring in 2000, and was also a co-recipient of the First Prize Paper Award in the IEEE Signal Processing Society Postgraduate Forum Hong Kong Chapter in 2004. In 2002 and 2003, he received, respectively, the SY King Fellowships and the WS Leung Fellowships from the University of Hong Kong. Also, he was awarded the Competitive Earmarked Research Grant Merit and Incentive Awards in 2003-2004.



Tung-Sang Ng (F'03) received the B.Sc.(Eng.) degree from The University of Hong Kong in 1972, the M.Eng.Sc. and Ph.D. degrees from the University of Newcastle, Australia, in 1974 and 1977, respectively, all in electrical engineering.

He worked for BHP Steel International and The University of Wollongong, Australia after graduation for 14 years before returned to The University of Hong Kong in 1991, taking up the position of Professor and Chair of Electronic Engineering. He was Head of Department of Electrical and Electronic

Engineering from 2000 to 2003 and Dean of Engineering from 2003 to 2007. His current research interests include wireless communication systems, spread spectrum techniques, CDMA and digital signal processing. He has published over 250 international journal and conference papers.

He was the General Chair of ISCAS'97 and the VP-Region 10 of IEEE CAS Society in 1999 & 2000. He was an Executive Committee Member and a Board Member of the IEE Informatics Divisional Board (1999-2001) and was an ordinary member of IEE Council (1999-2001).

He was awarded the Honorary Doctor of Engineering Degree by the University of Newcastle, Australia in 1997, the Senior Croucher Foundation Fellowship in 1999, the IEEE Third Millennium medal in 2000 and the Outstanding Researcher Award by The University of Hong Kong in 2003. He is a Fellow of IEEE, IEE and HKIE.