<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Probabilistic QoS Constrained Robust Downlink Multiuser MIMO Transceiver Design with Arbitrarily Distributed Channel Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>He, X; Wu, YC</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>IEEE Transactions on Wireless Communications, 2013, v. 12 n. 12, p. 6292-6302</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/199094">http://hdl.handle.net/10722/199094</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>IEEE Transactions on Wireless Communications. Copyright © IEEE.; ©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Probabilistic QoS Constrained Robust Downlink Multiuser MIMO Transceiver Design with Arbitrarily Distributed Channel Uncertainty

Xin He and Yik-Chung Wu

Abstract—We study the robust transceiver optimization in downlink multiuser multiple-input multiple-output (MU-MIMO) systems aiming at minimizing transmit power under probabilistic quality-of-service (QoS) requirements. Owing to the unknown distributed interference, the channel estimation error obtained from the linear minimum mean square error (LMMSE) estimator can be arbitrarily distributed. Under this situation, the QoS requirements should account for the worst-case channel estimation error distribution. While directly finding the worst-case distribution is challenging, two methods are proposed to solve the robust transceiver design problem. One is based on the Markov’s inequality, while the other is based on a novel duality method. Two convergence-guaranteed iterative algorithms are proposed to solve the transceiver design problems. Furthermore, for the special case of MU multiple-input single-output (MISO) systems, the corresponding robust transceiver design problems are shown to be convex. Simulation results show that, compared to the non-robust method, the QoS requirement is satisfied by both proposed algorithms. Among the two proposed methods, the duality method shows a superior performance in transmit power, while the Markov method demonstrates a lower computational complexity. Furthermore, the proposed duality method results in less conservative QoS performance than the Gaussian approximated probabilistic robust method and bounded robust method.

Index Terms—LMMSE channel estimation, QoS, robust MU-MIMO transceiver design, arbitrarily distributed uncertainty.

I. INTRODUCTION

In downlink MU-MIMO systems, there are two commonly used transceiver design methodologies, aiming at different goals. The first one is to maximize or minimize a performance metric, such as capacity, mutual information or sum data mean square error (MSE) subject to a power constraint [1]-[4]. However, this framework does not take fairness into account in multiuser system. The other methodology is QoS and fairness based design: minimizing total transmit power subject to different QoS constraints [5]-[7]. Fairness among users is automatically introduced by defining different QoS requirements.

For QoS constraint, it can be in terms of MSE or bit error rate (BER). If the noise is Gaussian, BER is strongly related to the signal-to-interference-plus-noise ratio (SINR) through the Gaussian Q function, and the SINR in each data stream is related to the inverse of MSE [8], [9]. Therefore, the BER requirement on the substream can be transformed into the MSE requirement on the substream. However, in practice, owing to the interference from unintended users (e.g., co-channel interference in a cellular system or interference from secondary user in cognitive radio systems), or even hostile jamming signals, the distribution of the interference plus noise might be unknown. In this case, the relation between BER and MSE cannot be expressed explicitly, and directly using the BER as the QoS constraint is intractable.

In general, channel state information (CSI) is required for transceiver design [1]-[7]. In practice, CSI has to be estimated, and estimation error is unavoidable. Therefore, robust transceiver design, which takes the channel estimation uncertainty into consideration, is important. For existing works, the estimation errors are usually assumed to be bounded or Gaussian distributed. In [10]-[12], bounded channel estimation error is introduced in the QoS constraints in the multiuser transceiver design problem. Nevertheless, in general, the distribution of estimation error of a random variable is unbounded [13]. To tackle this problem, several approximation based probabilistic robust transceiver designs with Gaussian distributed channel uncertainty were proposed. In uplink SIMO system, [14] aims to minimize total users’ transmit power with individual probabilistic SINR constraints. However, the method is only suitable for users with single antenna. Furthermore, owing to the difference between uplink and downlink inter-user interference management, the method in [14] cannot be generalized to downlink system. For cognitive radio systems when communication is carried out between single antenna secondary users, a robust power control is proposed in [15] aiming at maximizing secondary users’ capacity subject to probabilistically constrained interference to the primary user. The geometric programming (GP) method was used to solve the power allocation problem. However, the GP method cannot be extended to transceiver matrices design in MU-MIMO system. In single-user MIMO system, a robust transmitter was designed in [16] aiming at maximizing average SINR while total SINR was probabilistically constrained. However, the framework of [16] cannot handle individual SINR constraint on each data stream. Only in a recent paper [17], a robust transmitter design in downlink MIMO system is proposed to minimize total transmit power subject to individual probabilistic SINR constraint for different users. However, there is no corresponding receiver design in [17].
In this paper, we consider a general scenario that the distribution of the interference plus noise is not known, and the distribution of the channel estimation error under interference cannot be modeled in prior. To the best of our knowledge, this is the first work designing probabilistic QoS based robust MU-MIMO transceiver when the channel uncertainty distribution is not known. In particular, we formulate the probabilistic QoS constraints under arbitrarily distributed channel estimation error into worst-case probabilistic constraints, and two methods are proposed to tackle this problem. One is based on the Markov’s inequality, which provides an upper bound for the worst-case probability. The other is based on a novel duality method, in which the worst-case probability problem is transformed into a deterministic finite constrained problem by using the Lagrange duality and S-Lemma, with strong duality guaranteed. For both proposed methods, the resulting non-convex problems are solved by convergence-guaranteed iterative algorithms between two convex subproblems. Furthermore, for the special case of MU-MISO systems, the robust transceiver design problems are shown to be convex, and thus global optimal solutions can be guaranteed. Simulation results show that the duality method has an excellent performance on the special case of MU-MISO systems, the robust transceiver algorithms between two convex subproblems. Furthermore, for convex problems are solved by convergence-guaranteed iterative algorithms.

In Section IV, and conclusions are drawn in Section V.

The downlink MU-MIMO system consists of one base station (BS) equipped with $N$ transmit antennas and $K$ active users, with the $k$th user equipped with $M_k$ antennas ($\sum_{k=1}^{K} M_k = M$). It is assumed that $L_k$ independent data streams are transmitted to the $k$th user and $\sum_{k=1}^{K} L_k = L$. In order to guarantee data recovery at the users, it is required that $L_k \leq M_k$ and $L \leq \min\{M, N\}$. Let $s_k$ be the $L_k \times 1$ data vector transmitted to the $k$th user, we have $\mathbb{E}(s_k s_k^H) = I_{L_k}$.

Let $G$ be the $N \times L$ precoding matrix at the BS, and $s = [s_1^T \cdots s_K^T]^T$ is the data vector for all active users, then the received $M_k \times 1$ signal at the $k$th user is

$$y_k = H_k G s_k + n_k,$$

where $H_k$ and $n_k$ are the $M_k \times N$ channel matrix and the received $M_k \times 1$ interference plus noise vector of the $k$th user, respectively. It is assumed that the interference plus noise can be arbitrarily distributed with its first two moments being known. Without loss of generality, we assume $\mathbb{E}(n_k) = 0$ and $\mathbb{E}(n_k n_k^H) = R_k$, and we write $n_k \sim A(0, R_k)$, where $A$ denotes an arbitrary distribution.

At the receiver, an $L_k \times M_k$ equalizer $F_k$ is used to process the received signal. The recovered $L_k \times 1$ data vector at the $k$th user is

$$\hat{s}_k = F_k H_k G s + F_k n_k.$$  

Since the transmitted data are independent and noise, the total MSE of the $k$th user’s data can be calculated as

$$\text{MSE}_k = \mathbb{E}_{n_k, s_k} \left[ \text{Tr}\{ (D_k s - \hat{s}_k) (D_k s - \hat{s}_k)^H \} \right]$$

$$= \|F_k^H H_k^T G - D_k\|^2_F + \text{Tr}(F_k^H R_k F_k^H),$$

where the matrix $D_k = \{0_{L_k \times \sum_{k=1}^{K-1} L_k}, I_{L_k}, 0_{L_k \times \sum_{k=K+1}^{K} L_k} \}$ is used to select the $k$th user’s data streams. The basic QoS based MU-MIMO transceiver design problem is [7]

$$\min_{g \in \mathbb{F}_1, \cdots, \mathbb{F}_K} \text{Tr}(G^H G^H)$$

s.t. $\text{MSE}_k(G, F_k) \leq \varepsilon_k, \quad k = 1, \cdots, K,$

where $\varepsilon_k$ is the required QoS target in terms of MSE. However, as shown in (4), in addition to the precoder $G$ and equalizer $F_k$, the MSE also depends on the channel realization $H_k$, which has to be estimated in practice.

**B. Channel Estimation Under Arbitrarily Distributed Interference**

During channel estimation, an $N \times L_t$ training matrix $S$ is transmitted from the BS, where $L_t \geq N$. Let $G = I_N$ during training, the received signal at the $k$th receiver with observation window length $L_t$ is

$$Y_k = H_k S + N_k^t,$$

where $N_k^t$ is an $M_k \times L_t$ matrix with arbitrarily distributed elements and represent interference plus noise at the training stage. In this paper, the well-known Kronecker model is used for the MIMO channel [18], i.e., $H_k = R_{t,k}^2 H_{w,k} R_{t,k}^2$, where $R_{t,k}$ and $R_{t}$ are the correlation matrix at the $k$th receiver and the BS, respectively, and $H_{w,k}$ has independent and identically distributed (i.i.d.) complex Gaussian elements distributed as $\mathcal{CN}(0, 1)$. After taking vectorization on both sides of (5), we have $\text{vec}(Y_k) = (S^T \otimes I_{M_k}) \text{vec}(H_k) + \text{vec}(N_k^t)$. With $y_k^t \triangleq \text{vec}(Y_k), h_k^t \triangleq \text{vec}(H_k)$ and $n_k^t \triangleq \text{vec}(N_k^t)$, we get

$$y_k^t = (S^T \otimes I_{M_k}) h_k + n_k^t,$$

where the first two moments of $n_k^t$ is assumed to be known, i.e., $n_k^t \sim A(0, \Phi_k)$. According to the Kronecker channel model, we have $h_k \sim \mathcal{CN}(0, R_{t,k}^2 \otimes R_{t,k})$.

In order to utilize the statistical information of the channel and the interference plus noise, we employ the linear minimum
mean squared error (LMMSE) estimator, which takes the expression
\[ \hat{h}_k = (\mathbf{R}_i^{-T} \otimes \mathbf{R}_{r_k}^{-1}) + (\mathbf{S}^* \otimes \mathbf{I}_{M_k}) \Phi_k^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{M_k})^{-1} \cdot (\mathbf{S}^* \otimes \mathbf{I}_{M_k}) \Phi_k^{-1} \cdot y_k. \]
It is easy to prove that the mean of the channel estimation error is zero, i.e., \( \mathbb{E}(\hat{h}_k - \hat{h}_k) = 0 \), and the covariance of the error is [13]
\[ \Sigma_k = (\mathbf{R}_i^{-T} \otimes \mathbf{R}_{r_k}^{-1}) + (\mathbf{S}^* \otimes \mathbf{I}_{M_k}) \Phi_k^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{M_k})^{-1}. \] (7)

Therefore, the channel estimation error can be modeled as
\[ \mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta_k, \] (8)
where \( \text{vec}(\hat{\mathbf{H}}_k) = \hat{\mathbf{h}}_k \). Since the distribution of the interference plus noise is unknown, we can only model the channel estimation uncertainty \( \Delta_k \) as arbitrarily distributed with its first two moments known, i.e., \( \text{vec}(\Delta_k) \sim \mathcal{A}(0, \Sigma_k). \)

After substituting (8) into (4), the total MSE of the \( k \)th user is
\[ \text{MSE}_k(\Delta_k) = ||\mathbf{F}_k(\hat{\mathbf{H}}_k + \Delta_k)\mathbf{G} - \mathbf{D}_k||_F^2 + \text{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H). \] (9)
Note that the distribution of the MSE depends on that of the \( \Delta_k \). Furthermore, \( \mathbf{G} \) and \( \mathbf{F}_k \) are unknown, and in general depends on the statistical information of \( \Delta_k \). Therefore, the distribution of the MSE cannot be obtained or approximated in advance.

**Remark 1.** Besides the estimation error, the quantization error is also unavoidable in frequency-division duplex (FDD) systems (e.g., channel is estimated at the users and then quantized and fed back to the BS). Channel direction information is usually quantized by Grassmannian codebook or random vector quantization (RVQ) [19], [20], while the channel quality information is quantized by entropy-maximizing method [21]. In general, accurate analytical statistical model of quantization error is intractable as it depends on the distribution of the estimated channel \( \hat{\mathbf{H}}_k \). For example, for the simple case of scalar quantization [12], only an error bound can be derived. For RVQ technique with i.i.d. Rayleigh distributed channel, only the error angle is modeled. Fortunately, empirical mean and variance of the quantization error can be obtained once the codebook is fixed. Since the channel estimation error is independent with the quantization error, the combined channel uncertainty can also be modeled using (8) with updated mean and variance. As the bias of the quantization error can be shifted to the estimated channel, the channel uncertainty can still be modeled as \( \text{vec}(\Delta_k) \sim \mathcal{A}(0, \Sigma_k). \)

### III. Robust MU-MIMO Transceiver Design

In multiuser system, different users may have different QoS requirements, which can be formulated as
\[ \text{Prob}\{\text{MSE}_k(\Delta_k) \leq \varepsilon_k\} \leq p_k, \]
where \( \varepsilon_k \) and \( p_k \) are the required QoS target and guaranteed probability at the \( k \)th receiver, respectively. After the QoS requirements are satisfied, it is crucial to save transmission power at the BS. Therefore, the probabilistic QoS constrained transceiver design problem can be formulated as
\[
\begin{align*}
\min_{\mathbf{G}, \mathbf{F}_k \ldots, \mathbf{F}_K} & \quad \text{Tr}(\mathbf{G}^H) \\
\text{s.t.} & \quad \sup_{\text{vec}(\Delta_k) \sim \mathcal{A}(0, \Sigma_k)} \text{Prob}\{\text{MSE}_k(\Delta_k) \geq \varepsilon_k\} \leq p_k, \quad \forall k.
\end{align*}
\] (P0)

In the QoS constraints, due to the arbitrarily distributed channel estimation errors, the supremum of the outage probability is used to guarantee the QoS performance is satisfied even at the worst case.

Note that the problem (P0) is a bilevel optimization problem, and the lower level problem involves finding the supremum of the outage probability. Owing to the unknown distribution of \( \Delta_k \), it is difficult to get an analytical solution for the lower level problem. Below, we consider two methods to solve this problem. The first one is to use the Markov’s inequality to eliminate the supremum. Another method is the proposed dual method, in which the lower level problem is transformed from a stochastic problem into a solvable deterministic problem.

#### A. The Markov’s Inequality Based Method

Since the distribution of the MSE is unknown, the Markov’s inequality [22] can be used to get an upper bound of the outage probability
\[
\text{Prob}\{\text{MSE}_k(\Delta_k) \geq \varepsilon_k\} \leq \frac{\mathbb{E}_k\{\text{MSE}_k(\Delta_k)\}}{\varepsilon_k},
\] (10)
where \( \mathbb{E}_k\{\text{MSE}_k(\Delta_k)\} \) can be derived as
\[
\begin{align*}
\mathbb{E}_k\{\text{MSE}_k\} &= \mathbb{E}_k\{||\mathbf{F}_k(\hat{\mathbf{H}}_k + \Delta_k)\mathbf{G} - \mathbf{D}_k||_F^2 + \text{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H)\}
\end{align*}
\] (11)
\[
= \mathbb{E}_k\{||\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{D}_k||_F^2 + \text{Tr}(\mathbf{R}_k \mathbf{F}_k^H)^2\}
\] (12)
\[
= ||\text{vec}\left((\mathbf{G}^T \otimes \mathbf{F}_k)\Sigma_k^2\right)_{\mathbf{F}_k}||_F^2.
\] (13)

Since the right hand side of (10) is independent of the exact distribution, the problem (P0) can be approximated as
\[
\begin{align*}
\min_{\mathbf{G}, \mathbf{F}_k \ldots, \mathbf{F}_K} & \quad \text{Tr}(\mathbf{G}^H) \\
\text{s.t.} & \quad ||\text{vec}\left((\mathbf{G}^T \otimes \mathbf{F}_k)\Sigma_k^2\right)_{\mathbf{F}_k}||_F^2 + \text{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H)^2 \leq p_k \varepsilon_k, \quad \forall k.
\end{align*}
\] (14)

From (14), it is obvious that the probabilistic robust approach becomes a statistically average approach [23], [24]. Although this problem is still nonconvex, it is observed that it becomes a convex problem of \( \mathbf{G} \) when all equalizer \( \mathbf{F}_k \) are fixed. Furthermore, when the precoder \( \mathbf{G} \) is fixed, the objective function is not related to the equalizer \( \mathbf{F}_k \). In order to provide a larger feasible space for the next round precoder design, the equalizer \( \mathbf{F}_k \) can be used to minimize the left side of the QoS constraints in (14). Therefore, the problem (14) can be solved by iterations between two convex subproblems as follows.

In the first subproblem, with the equalizer \( \mathbf{F}_k \) fixed, the optimum precoder \( \mathbf{G} \) can be obtained from the second-order cone programming (SOCP) problem
\[
\begin{align*}
\min_{\mathbf{G}} & \quad P \\
\text{s.t.} & \quad ||\text{vec}(\mathbf{G})||_2^2 \leq P
\end{align*}
\] (15)
where $P$ is a slack variable.

In the second subproblem, with the precoder $G$ fixed, the equalizer $F_k$ can be used to minimize the left side of the QoS constraint. Expressing the left side of the QoS constraint in (14) with the Frobenius norm, the equalizer $F_k$ can be updated from the following problem

$$\min_{F_k} \| G^T F_k \Sigma_k^{1/2} F_k^T + \| F_k \bar{H} G - D_k \|_F^2 + \| R_k^{1/2} F_k H_k^T \|_F^2. \quad (16)$$

Furthermore, writing the first term of the cost function (16) as a quadratic form,

$$\| (G^T F_k) \Sigma_k^{1/2} F_k^T \|_F^2 = \text{Tr} \left( \Sigma_k \left( (G^T G)^T \otimes (F_k H_k^T) \right) \right) \quad (17)$$

$$= \text{Tr} \left( \sum_{i=1}^N \sum_{j=1}^N g_{ij} \Sigma_{ij} \right) \quad (18)$$

where $g_{ij}$ is the $(i,j)_{th}$ element of the matrix $G^T G$, $\Sigma_{ij}$ is the $(i,j)_{th}$ subblock of the matrix $\Sigma_k$. Putting (18) into (16), and setting the derivative of the cost function (16) with respect to $F_k$ to zero, the optimum equalizer $F_k$ is

$$F_k = (\bar{H} G D_k H_k^T)^{-1} (\bar{H} G G^H H_k^T + R_k + \sum_{i=1}^N \sum_{j=1}^N g_{ij} \Sigma_{ij})^{-1}. \quad (19)$$

It is observed that the obtained equalizer is a conventional MMSE equalizer with additional regularization from the weighted uncertainty $\sum_{i=1}^N \sum_{j=1}^N g_{ij} \Sigma_{ij}$.

The iterative algorithm between the two subproblems is summarized in Table I. With a feasible initialization, the transmit power in each iteration obtained by the iterative algorithm decreases monotonically and converges. Similar proofs have been provided in [11].

With regard to the initialization, it is a common problem for the QoS based MU-MIMO transceiver design since a feasible initial transceiver pair is required [7] [11]. Conventionally, the receiver $F_k$ is initialized with an identity or a randomly generated matrix. However, these initializations are not guaranteed to satisfy the QoS constraints, and the precoder $G$ design in (15) may not exist. It is observed in (15) that if the elements of $F_k$ are small, we have a better chance of satisfying the QoS constraints. Based on this observation, it is suggested in [11] that scaling factors $1/a_k$ are introduced into the initial chosen equalizer in (15), and the scaling factors can be obtained from the following SOCP problem

$$\min_{\alpha_k, P} \| \alpha_k G \|_2 \leq P$$

$$\| \text{vec}(G^T F_k \Sigma_k^{1/2} F_k^T \text{vec}(F_k H_k^T G - D_k) \|_F^2 + \| R_k^{1/2} F_k H_k^T \|_F^2 \quad (20)$$

where $G_{k0}$ is the initial chosen equalizer. Compared to the precoder design with fixed equalizers in (15), the joint precoder and scalable equalizers ($G_{k}, \frac{1}{a_k} F_{k0}, \cdots, \frac{1}{a_k} F_{K0}$) in (20) has more degree of freedom and thus has a higher chance of satisfying the QoS constraints. If problem (20) is feasible, the result ($G_{k}, \frac{1}{a_k} F_{k0}, \cdots, \frac{1}{a_k} F_{K0}$) is used as the initial starting point for the iterative algorithm. Otherwise, another $F_{k0}$ with a different beamforming direction may be chosen (20) is solved again. In practice, the feasibility problem can also be mitigated by utilizing it as the user admission criterion if the number of active user is large [25]. However, the corresponding cross-layer design problem is beyond the scope of this paper.

### B. The Duality Based Method

In the MSE expression (9), the random variable $\Delta_k$ is weighted by unknown $F_k$ and $G$, whose values in general depend on $\Delta_k$. Therefore, the MSE is a sum of correlated elements. According to the generalized weak-convergence theorem [26], a sum of many random variables with dependence will tend to be distributed according to one of a small set of stable distributions. This means that although the channel estimation uncertainty $\Delta_k$ is arbitrarily distributed, the MSE in (9) is in fact not arbitrarily distributed. Therefore, the Markov’s inequality in (10) is quite loose [22], and the QoS and power saving performance of the Markov method is expected to be conservative. In this subsection, the exact solution of the lower level problem is derived by the proposed duality method.

Since the MSE is a function of the channel estimation uncertainty, let $\psi_k(x_k) = \text{MSE}_k(\Delta_k)$, where $x_k \triangleq \text{vec}(\Delta_k)$. The lower level problem $\sup_{\text{vec}(\Delta_k) \sim A(\Sigma_k)} \text{Prob}\{ \text{MSE}_k(\Delta_k) \geq \varepsilon_k \}$ can be reformulated as

$$\sup_{\psi_k(x_k) \geq \varepsilon_k} \text{Prob}\{ \psi_k(x_k) \geq \varepsilon \} \quad \text{s.t.} \quad \int f(x_k) dx_k = 1$$

$$E(x_k) = 0$$

$$E(x_k x_k^H) = \Sigma_k,$$

where $f(x_k)$ is the probability density function (PDF) of the vectorized uncertainty $x_k$.

### Transforming problem (21) into its dual problem

The Lagrangian of this problem is presented as

$$\mathcal{L}_k(f(x_k), \alpha_k, \eta_k, \Xi_k)$$

$$= \text{Prob}\{ \psi_k(x_k) \geq \varepsilon \} + \alpha_k \left( 1 - \int f(x_k) dx_k \right)$$

$$+ \eta_k (0 - E(x_k)) + \text{Tr} \left( \Xi_k^H (\Sigma_k - E(x_k x_k^H)) \right)$$

$$= \alpha_k + \text{Tr} (\Xi_k^H \Sigma_k) + \int f(x_k) dx_k$$

$$- \int (\alpha + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H)) f(x_k) dx_k$$

where $x_k \in \mathbb{C}^{N_k}$. This problem can be rewritten as

### Table I

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Set iteration number $j = 1$, initialize with a feasible transceiver set $(G[0], F[0], \cdots, F_{k0})$, define $P[0] = \text{Tr}(G[0]^H G[0])$</td>
</tr>
<tr>
<td>3.</td>
<td>Update $F[j]$ using (19)</td>
</tr>
<tr>
<td>4.</td>
<td>If $P[j - 1] - P[j] \leq \varepsilon$ (where $\varepsilon$ is a pre-defined threshold) then stop, otherwise increment $j$ and go to step 2</td>
</tr>
</tbody>
</table>
\[=\alpha_k + \text{Tr}(\Xi_k^H\Sigma_k) + \int \left(1 - \alpha_k - \eta_k^H x_k - \text{Tr}(\Xi_k^H x_k x_k^H)\right) \cdot f(x_k)dx_k \]

\[\psi_k(x_k) \geq \varepsilon_k\]

\[= \int \left(0 - \alpha_k - \eta_k^H x_k - \text{Tr}(\Xi_k^H x_k x_k^H)\right) \cdot f(x_k)dx_k. \quad (24)\]

\[\psi_k(x_k) \geq \varepsilon_k\]

where \(\alpha_k, \eta_k, \Xi_k\) are the Lagrangian multipliers, and \(\Sigma_k = \Xi_k^H\). With the implicit PDF constraint \(f(x_k) \geq 0\), the Lagrange dual function of the problem (21) is

\[g_k(\alpha_k, \eta_k, \Xi_k) = \sup_{f(x_k) \geq 0} L_k(\alpha_k, \eta_k, \Xi_k). \quad (25)\]

Note that the supremum of the first integral in (24) with the nonnegative PDF constraint is zero if \(\alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) > 1\), otherwise the supremum is infinity. Similarly, the supremum of the second integral with the nonnegative PDF constraint is also zero if \(\alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) \geq 0\), otherwise the supremum is infinity. Therefore, the Lagrange dual function \(g_k(\alpha_k, \eta_k, \Xi_k)\) is

\[g_k(\alpha_k, \eta_k, \Xi_k) = \begin{cases} 
\alpha_k + \text{Tr}(\Xi_k^H \Sigma_k) & \text{if } \alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) \geq 0, \forall x_k : \psi_k(x_k) < \varepsilon_k; \text{ and } \alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) > 1, \forall x_k : \psi_k(x_k) \geq \varepsilon_k \\
+\infty & \text{otherwise}
\end{cases}\]

(26)

Combining the two conditions that make \(g_k(\alpha_k, \eta_k, \Xi_k) = \alpha_k + \text{Tr}(\Xi_k^H \Sigma_k)\), the first condition can be replaced by \(\alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) \geq 0\) for all \(x_k \in \mathbb{C}^{NM_k}\). Therefore, the dual of the problem (21) can be formulated as

\[\min_{\alpha_k, \eta_k, \Sigma_k} \alpha_k + \text{Tr}(\Xi_k^H \Sigma_k) \quad \text{s.t. } \alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) \geq 0, \forall x_k : x_k \in \mathbb{C}^{NM_k} \]

\[\alpha_k + \eta_k^H x_k + \text{Tr}(\Xi_k^H x_k x_k^H) > 1, \forall x_k : \psi_k(x_k) \geq \varepsilon_k \]

\[\Xi_k = \Xi_k^H. \quad (27)\]

**Remark 2.** It is recognized that the problem (21) is known as the moment problem [27]. Since only the first two moments of the random vector \(x_k\) are used in (21), the feasible moment vector set of \(x_k\) is \(M = \{0, \Sigma_k|\Sigma_k \succeq 0\}\). According to the general theory on the moment problem [27, p. 812], strong duality holds between the primal moment problem (21) and its dual problem (27) when the moment vector of \(x_k\) is an interior point of \(M\), i.e., \(\Sigma_k \succ 0\).

**Remark 3.** From (7), the condition \(\Sigma_k \succ 0\) is equivalent to requiring that either the covariance matrix of the channel \(\text{vec}(\Phi_k^H \otimes R_k)\) or the interference plus noise covariance matrix \(\Phi_k\) is positive definite. Furthermore, when strong duality holds, the optimal value of (21) is equivalent to that of (27).

In order to get a compact form of (27), we define \(Z_k \triangleq \begin{bmatrix} \Xi_k^H & \frac{1}{2}\eta_k \\ \frac{1}{2}\eta_k^H & \alpha_k \end{bmatrix}, \Sigma_k \triangleq \begin{bmatrix} \Sigma_k & 0 \\ 0 & 1 \end{bmatrix}\) and \(u_k \triangleq [x_k^H, 1]^T\). Then we get

\[\min_{\Sigma_k} \text{Tr}(Z_k^H \Sigma_k) \quad \text{s.t. } u_k^H Z_k u_k \geq 0, \forall x_k : x_k \in \mathbb{C}^{NM_k} \]

\[u_k^H Z_k u_k - 1 > 0, \forall x_k : \psi_k(x_k) \geq \varepsilon_k \]

\[\Sigma_k = Z_k^H. \quad (28)\]

Note that the first and the third constraints of (28) can be combined into \(Z_k \succeq 0\). Furthermore, after replacing the MSE term \(\psi_k(x_k)\) with \(\text{MSE}_k(\Delta_k)\) in (9), the problem (28) becomes

\[\min_{Z_k} \text{Tr}(Z_k \Sigma_k) \quad \text{s.t. } Z_k \succeq 0 \]

\[u_k^H Z_k u_k - 1 > 0, \forall \Delta_k : \|F_k(\hat{H}_k + \Delta_k)G - D_k\|_F^2 + \text{Tr}(F_k R_k F_k^H) \geq \varepsilon_k. \quad (29)\]

Since (29) does not contain probabilistic constraint, it is a deterministic optimization problem. But it is still an infinite constrained problem.

**Transforming problem (29) into finite constrained problem**

We first reformulate the Frobenius norm as a spectral norm as follows

\[\|F_k(\hat{H}_k + \Delta_k)G - D_k\|_F^2 = \|\text{vec}(F_k(\hat{H}_k + \Delta_k)G - D_k)\|^2_z \quad (30)\]

\[= \|\text{vec}(F_k \Delta_k G) + \text{vec}(F_k \hat{H}_k G - D_k)\|^2_z \quad (31)\]

\[= \left\|G^T F_k \text{vec}(F_k \hat{H}_k G - D_k) \right\|^2 \quad (32)\]

\[u_k^H Q_k F_k u_k, \quad (33)\]

where \(\text{vec}(F_k \Delta_k G) = (G^T F_k) \text{vec}(\Delta_k)\) is used from (31) to (32) [28], and \(Q_k \triangleq [G^T F_k \text{vec}(F_k \hat{H}_k G - D_k)]\). Putting (33) into the condition of the second constraint of (29), we get the quadratic form

\[\|F_k(\hat{H}_k + \Delta_k)G - D_k\|_F^2 + \text{Tr}(F_k R_k F_k^H) - \varepsilon_k \]

\[= u_k^H Q_k - \text{diag}((0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H))) u_k \geq 0. \quad (34)\]

Therefore, the second constraint of the problem (29) can be transformed as

\[u_k^H (Z_k - \text{diag}((0, 1))) u_k > 0, \forall x_k : u_k^H (Q_k^H Q_k - \text{diag}((0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H)))) u_k \geq 0. \quad (36)\]

By using the S-Lemma in control theory [29, p. 23], (36) can be equivalently formulated as

\[\exists \lambda_k \geq 0 : Z_k - \text{diag}((0, 1)) - \lambda_k (Q_k^H Q_k - \text{diag}((0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H)))) \geq 0. \quad (37)\]

It is shown in Appendix A that the problem (P0) will be infeasible if \(\lambda_k = 0\) for any \(k \in \{1, 2, \cdots, K\}\). Therefore, we can omit \(\lambda_k = 0\) in (37). Letting \(\beta_k \triangleq 1/\lambda_k > 0\), the constraint (37) can be reformulated as

\[\exists \beta_k > 0 : \beta_k Z_k + \text{diag}((0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k)) - Q_k^H Q_k \geq 0. \quad (38)\]

According to the Schur’s Complement [30], (38) can be transformed into a linear matrix inequality (LMI) form of \(\beta_k\) as

\[\exists \beta_k > 0 : \beta_k Z_k + \text{diag}((0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k)) - Q_k^H Q_k \geq 0. \quad (39)\]
After replacing the second constraint of the problem (29) with its equivalent form (39), the problem (29) becomes

\[
\min_{\beta_k, Z_k} \text{Tr}(Z_k \Sigma_k)
\text{ s.t. } Z_k \geq 0, \quad \beta_k > 0
\]

\[
\begin{bmatrix}
\beta_k Z_k + \text{diag}([0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0.
\] (40)

Defining a new variable \( \tilde{Z}_k \triangleq \beta_k Z_k \), then the problem (40) further becomes an LMI problem

\[
\min_{\beta_k, \tilde{Z}_k} \text{Tr}(\tilde{Z}_k \Sigma_k)/\beta_k
\text{ s.t. } \tilde{Z}_k \geq 0, \quad \beta_k > 0
\]

\[
\begin{bmatrix}
\tilde{Z}_k + \text{diag}([0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0, \forall k.
\] (41)

Therefore, the stochastic problem (21) is transformed into the deterministic finite constrained problem (41).

**Transforming \((\mathbb{P}0)\) into a solvable single-level problem**

Since the strong duality holds between the primal and dual problem, after replacing the lower level problem of \((\mathbb{P}0)\) with (41), we get the following results.

**Proposition 1.** The original bilevel optimization problem \((\mathbb{P}0)\) is equivalent to the following single-level problem

\[
\min_{G, F_1, \ldots, F_K} \text{Tr}(GG^H)
\text{ s.t. } \text{Tr}(\tilde{Z}_k \Sigma_k)/\beta_k \leq p_k, \quad \beta_k > 0, \quad \tilde{Z}_k \succeq 0, \forall k
\]

\[
\begin{bmatrix}
\tilde{Z}_k + \text{diag}([0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0, \forall k.
\] (P1)

**Proof.** See Appendix B.

Since the term \( Q_k \) in problem \( P1 \) contains the product of \( G \) and \( F_k \), it is a nonconvex problem of \( G \) and \( F_k \). However, it can be solved by an iteration between two subproblems as follows.

In the first subproblem, when all equalizers \( F_k \) are fixed, the subproblem becomes a convex problem of \( G \), and the precoder design problem is

\[
\min_{G, \beta_1, \ldots, \beta_K} \text{Tr}(GG^H)
\text{ s.t. } \text{Tr}(\tilde{Z}_k \Sigma_k) \leq p_k \beta_k, \quad \beta_k > 0, \quad \tilde{Z}_k \succeq 0, \forall k
\]

\[
\begin{bmatrix}
\tilde{Z}_k + \text{diag}([0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0, \forall k.
\] (42)

This convex problem can be efficiently solved by the interior point method in [31, p. 561].

For the second subproblem, similar to the idea in the Markov method, the equalizer \( F_k \) is chosen to minimize the guaranteed data MSE \( \varepsilon_k \) to create a larger feasible space for the next round precoder design. Therefore, \( F_k \) is updated using the following problem

\[
\min_{F_k, \beta_k, Z_k, \varepsilon_k} \varepsilon_k
\text{ s.t. } \text{Tr}(\tilde{Z}_k \Sigma_k) \leq p_k \beta_k, \quad \beta_k > 0, \quad \tilde{Z}_k \succeq 0
\]

\[
\begin{bmatrix}
\tilde{Z}_k + \text{diag}([0, \varepsilon_k - \text{Tr}(F_k R_k F_k^H) - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0, \forall k.
\] (43)

In contrast to the problem (16), we cannot get a closed-form solution from (43). Furthermore, owing to the nonlinear term \( \text{Tr}(R_k F_k^H F_k) \) in the positive semidefinite constraint, this subproblem seems to be a nonconvex problem on \( F_k \). But since \( \text{Tr}(R_k F_k^H F_k) \) is a scalar and lies on the diagonal of the matrix \( \begin{bmatrix} \tilde{Z}_k + \text{diag}([0, \varepsilon_k - \text{Tr}(R_k F_k^H F_k) - \beta_k]) - \beta_k \end{bmatrix} Q_k^H (I_{LL_k}) \), by introducing a slack variable \( t_k \), the problem (43) is transformed into

\[
\min_{F_k, \beta_k, Z_k, \varepsilon_k, t_k} \varepsilon_k
\text{ s.t. } \text{Tr}(\tilde{Z}_k \Sigma_k) \leq p_k \beta_k, \quad \beta_k > 0, \quad \tilde{Z}_k \succeq 0
\]

\[
\begin{bmatrix}
\tilde{Z}_k + \text{diag}([0, \varepsilon_k - t_k - \beta_k]) - \beta_k
\end{bmatrix} Q_k^H (I_{LL_k}) \succeq 0, \forall k.
\] (44)
where the positive amplitude constraint $a_k > 0$ regularizes the set $\{(a_k, c_k)|a_k c_k \geq 1\}$ into a convex set, which can be represented as a LMI $\left[\frac{a_k}{1-c_k}, \frac{1}{1-c_k}\right] \geq 0$. Therefore, the initial transceiver pair $(G, \frac{1}{a_k} F_{1\alpha}, \cdots, \frac{1}{a_k} F_{K\alpha})$ can be efficiently obtained from the convex problem (46).

C. Convex Robust MU-MISO Transceiver Design

In MU-MISO systems, the proposed iterative algorithms converge, but in general it is not known whether the converged solution is the global optimal or not [1]-[7]. However, for the special case of MU-MISO systems, we will show that the robust probabilistic QoS constrained transceiver design problems are convex, thus global optimal solution is guaranteed.

1) Markov’s Inequality Based Method: Since each user only equipped with one antenna in the MU-MISO system, $L_k = M_k = 1, L = K$ and the $k^{th}$ user’s equalizer $F_k$ in (14) should be replaced by $1/a_k e^{j\theta_k}$. After multiplying $G_k$ on both sides of all inequalities in problem (14), it becomes

$$\min_{\alpha_1, \cdots, \alpha_K, \beta_1, \cdots, \beta_K} \text{Tr}(GG^H)$$

$$\text{s.t.} \quad \|\text{vec}(G^T \Sigma_k^{1/2})\|_2^2 + \|	ext{vec}(H_k G e^{j\theta_k} - a_k D_k)\|_2^2 \geq \frac{1}{a_k^2} |\rho_k|, \quad \forall k,$$

where $D_k = [0, \cdots, 1, \cdots, 0]$, with 1 appears at the $k^{th}$ position, is a $1 \times K$ basis vector in MU-MISO system. Now, define $G_k \triangleq G e^{j\theta_k}$ and $1 \leq a_k \leq \rho_k$, we get

$$\min_{G, \alpha_1, \cdots, \alpha_K} \text{Tr}(GG^H)$$

$$\text{s.t.} \quad \|	ext{vec}(G^T \Sigma_k^{1/2})\|_2^2 + \|	ext{vec}(H_k G_k e^{j\theta_k} - a_k D_k)\|_2^2 \geq \frac{1}{a_k^2} |\rho_k|, \quad \forall k,$$

and the global optimal transceiver $(G, \frac{1}{a_k}, \cdots, \frac{1}{a_k})$ for the Markov’s inequality based MU-MISO system can be obtained.

2) Duality Based Method: Similar to the above Markov method, letting $F_k \triangleq a_k e^{j\theta_k}$ and introducing $G$ to (30), we get

$$\min_{\alpha_1, \cdots, \alpha_K} \text{Tr}(GG^H)$$

$$\text{s.t.} \quad \left[\begin{array}{c}
\text{vec}(G^T \Sigma_k^{1/2}) \\
H_k G - a_k D_k, F_k^{1/2}
\end{array}\right] \leq a_k \rho_k, \quad \forall k,$$

and the global optimal transceiver $(G, \frac{1}{a_k}, \cdots, \frac{1}{a_k})$ for the Markov’s inequality based MU-MISO system can be obtained.

D. Implementation Consideration and Complexity Analysis

In practical systems, the proposed Markov and duality method can be used when the estimated downlink channels and the statistical information of the estimation error are obtained at the BS. In particular, since the downlink channel is the same as the uplink channel in the time-division duplex (TDD) system, the BS can estimate the uplink channels and then infer the downlink channels. Once these information are obtained, the BS can calculate the optimal precoder $G$ and all equalizers, and send the equalizer $F_k$ to the corresponding user.

For computational complexity, since the first subproblem of the Markov method is a SOCP problem, the complexity in each iteration is $O\left(n^2 \cdot \sum_{k=1}^K d_k\right)$ [32], where $n = NL$ is the number of unknown variables and $d_k = NLM_k L_k + L^2 + M_k^2$ is the vector length in the $k^{th}$ second order cone constraint. The complexity of the second subproblem is $O(M_k^2)$ owing to the dominant matrix inversion operation.

For the computational complexity of the duality method, since all subproblems are SDP problems, the complexity in each iteration is $O\left(n^2 \cdot \sum_{k=1}^K d_k^2\right)$ [32] where $n$ is the number of unknown variables and $d_k$ is the matrix dimension in the $k^{th}$ positive semidefinite constraint. For the first subproblem, $n = NL + \sum_{k=1}^K (N M_k + 1)^2 + K$ and $d_k = N M_k + 1 + LL_k$. Although the matrix dimension of the second subproblem is the same as that of the first subproblem, the variable number is reduced to $n = L_k M_k + (N M_k + 1)^2 + 3$ owing to the fact that all equalizers can be computed in parallel. Therefore, the computational complexity of the Markov method is lower than that of the duality method.

IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of the two proposed algorithms is illustrated by simulations. Unlike most existing works, the channel estimation is implemented first, and the estimated channel and statistical information of the estimation errors are taken into consideration in the transceiver design. The MIMO channel in the simulations has complex Gaussian entries with zero mean and uniform variance. The channel correlation matrices are taken as the exponential model, i.e., $[R_{ij}]_{ij} = \rho_\ell [i-j]$, $[R_{rk}]_{ij} = \rho_c [i-j]$, with the correlation coefficient $\rho_\ell = 0.2$ and $\rho_c = 0.5$. We assume the system is interference dominated, and the received interference at every antenna is i.i.d. with zero mean and variance 0.01. Since interference in future networks (e.g., Cognitive Radio [33] and Heterogeneous Network [34]) is often modeled by heavy-tailed distributions, it is assumed in this section that the distribution of interference is Laplace unless stated otherwise. The BS is equipped with four antennas and there are two active users, and each is equipped with two antennas, i.e., $N = 4, K = 2, M_k = 2$. Spatial multiplexing is implemented for the two users, i.e., $L_k = 2$. At the channel estimation stage, orthogonal Walsh sequences with length $L_c = 2N$ are employed in different transmit antennas. At the data detection stage, the required outage probability is $p_1 = p_2 = 0.15$, the QoS target for the
second user is $\varepsilon_2 = 0.3$ while that for the first user is varied between 0.05 and 0.3, and the threshold to terminate the iteration is $\varepsilon = 5 \times 10^{-4}$. All simulation results are averaged over $10^3$ randomly chosen channel realizations. For fair comparison in the transmit power performance, the randomly chosen channel realizations are feasible for all algorithms under consideration.

In Fig. 1, the convergence performance of the proposed algorithms with $F_{ko}$ being identity matrix and random matrix as initialization is compared. The QoS target for the first user is $\varepsilon_1 = 0.3$. It can be seen that for any particular proposed method, both initializations lead to the same transmit power level after convergence, but the method with identity matrix initialization converges slightly faster than the random one. Therefore, $F_{ko} = I$ is used for equalizer initialization in the subsequent simulations. Besides the convergence speed, from Fig. 1, it is further noted that the duality method achieves a lower transmission power level than the Markov method, as the Markov method involves a loose upper bound, and therefore is more conservative.

In Fig. 2, the histograms of the recovered first user’s data MSEs of the two proposed methods and the non-robust method using the basic formulation after (4) [7] is compared when the QoS target for the first user is $\varepsilon_1 = 0.15$. The non-robust method takes the estimated channel as the true channel, and the channel estimation error is ignored. It is observed that the residual MSEs of the two proposed robust methods are smaller than the prefixed QoS target, while that of the non-robust method exceeds the QoS target for about 70% of the channel realizations. Furthermore, it is also noted that the residual MSEs of the duality method are very close to the QoS target, while that of the Markov method are much smaller than the target. This shows the conservative nature of the Markov’s inequality based design.

The required transmit power of the two proposed algorithms to guarantee QoS performance is illustrated in Fig. 3. It is observed that over a wide range of QoS target for the first user, the duality method consistently requires a significantly lower transmission power than that of the Markov method, saving more than 8 dB in transmission power. This is owing to the loose Markov’s inequality which makes the approximated QoS requirements in (14) over conservative. The additional required transmit power of the duality method compared to the non-robust method is about 3 dB, which is the cost of the guaranteed QoS performance against the channel uncertainty. At Fig. 4, the required transmit power as a function of outage probability $p_1$ (with $\varepsilon_1 = 0.15$) is shown. It can be seen that similar conclusions can be drawn as in Fig. 3.

From Figs. 2, 3 and 4, it can be concluded that the proposed duality method guarantees the QoS requirement, while maintaining a low transmission power. While the Markov method also guarantee the QoS constraints, the duality method achieves a good balance between the QoS requirement and the power saving requirement. Notice that although the non-robust method seems to indicate a low transmit power, it does not represent a valid solution because the QoS constraints are not satisfied as shown in Fig. 2. Similar conclusions can be drawn...
in other simulation settings with different interference distributions, e.g., Gaussian and uniform distributed interference. The corresponding figures are not repeated in this paper.

One may argue that in order to have fair comparison, the transmit power of the non-robust method should be increased to the same level as the robust method. However, this is possible only after we executed the proposed robust transceiver design algorithm (i.e., after we obtained the required transmit power to guarantee the QoS requirements). Without the proposed algorithms, the only way to guarantee the QoS requirements for the non-robust method is to perform extensive simulations on the MSE with increasing transmit power level, until the QoS requirements are satisfied. This would take prohibitively significant amount of time, as the minimum required power level would also depend on numerous other factors, such as pilot pattern, pilot transmit power, total number of users, antenna number in BS and different users, number of data stream of each user, signal-to-noise ratio, and most importantly the interference plus noise distribution. Simulations have to be re-run if any one of the above parameters is changed. Therefore, this approach is almost impossible in practice.

The performance comparison between the Gaussian approximated robust method in [11] and the duality method is illustrated in Fig. 5. In the simulation, there are two users, both equipped with single antenna, and one BS equipped with two antennas. This corresponds to the MU-MISO case. The MSE target is 0.15 for both users (the corresponding SINR target can be calculated to be 5.66 [9]), and the outage probability is 0.15. As shown in Fig. 5, the residual MSE and SINR of the method in [11] are more spread out (conservative) than the proposed duality method, the reason is that the Bernstein-type inequality in [11] is a conservative approximation for linear and quadratic formulated problems with probabilistic constraint over a set of ambiguous uncertainties [35, p.91, p.118], while the proposed duality method is a tight solution for the specific problem. Also note that the method in [11] can only be applied to the MU-MISO system, while our proposed methods are applicable to the more general MU-MIMO system.

In Fig. 6, the cumulative distribution function (CDF) of the residual first user’s data MSE for the proposed duality method and the bounded robust method in [11] are compared. The interference distribution is assumed to be Gaussian, but not known to the design procedure. For the bounded method, since the interference distribution is unknown, the bound is heuristically selected as $3b$ and $0.3b$, where $b = \|\Sigma_k\|_F$. It can be seen that the QoS performance of the bounded method with bound $3b$ is too conservative, while the QoS requirement is not satisfied for the bound $0.3b$. Therefore, without the interference distribution information, there is no systematic way to choose an accurate bound to realize the probabilistic QoS target. On the other hand, if we assume the interference distribution is known to be Gaussian, the bound can be set as $1.44b$ to make this bound to cover 85% of the Gaussian channel estimation error. Only under this strong assumption, the bounded method has a comparable QoS performance to the proposed duality method, which assumes no knowledge on the interference distribution. However, even with the accurate bound $1.44b$, the distribution of the MSEs of the bounded method is more spread out than that of the duality method. This is because the bounded robust method only exploited the deterministic bound information of channel uncertainty, while the duality method fully utilized the first and second-order moment information. It can be concluded that the bounded robust method in [11] is not a practical way to realize the probabilistic QoS constrained transceiver design without distribution information.

In practical wireless systems, the transmission power is limited. Therefore, besides the feasibility problem at the initialization, the problem is also considered as infeasible if the required transmission power is larger than the peak power, resulting in two factors that affect the feasibility of the whole transceiver design problem. With the QoS target for the first user $\varepsilon_1 = 0.3$, the impact of the peak power constraint on the feasible rate of different methods is illustrated in Fig. 7. On the left hand side of Fig. 7, the rapidly increasing feasible rate with increasing peak power reveals that the peak power is the main limiting factor. On the other hand, on the right
hand side of Fig. 7, the flat feasible rate with increasing peak power reveals that the peak power is not the limiting factor, but the feasibility rate of initialization would become the limiting factor. For practical peak power around 1 to 2 (corresponds to SINR 20 to 23 dB), we observe that the feasible rate of the proposed duality method is close to that of the non-robust method, while that of the Markov method is close to zero owing to its high transmit power requirement.


cumulative Probability

Feasible Rate

Fig. 6. The CDF of residual MSE for the proposed duality method and the bounded robust method ($\mu_1 = \rho_2 = 0.15, \epsilon_1 = 0.15, \epsilon_2 = 0.3, b = \|\Sigma_k^1\|/\rho$).

Fig. 7. The feasible rate of different methods with different peak power constraints.

boundary of the constraint and the optimal value of problem (50) becomes $\text{Tr}(\text{diag}([0, 1]|\Sigma_k)) = 1$. Putting this optimal value into (F0), the resulted constraint in (F0) becomes $1 \leq p_k$, which is infeasible because the outage probability has to be smaller than one.

APPENDIX B

PROOF OF THE PROPOSITION 1

Owing to the strong duality between the supremum of the outage probability in (F0) and its dual problem (41), the problem (F0) is represented as

$$\min_{g, f_1,...,f_K} \text{Tr}(g g^H)$$

s.t. $\beta_k \geq 0$, $\tilde{Z}_k \geq 0$, $B_k(g, f_k, \beta_k, \tilde{Z}_k) \geq 0$

where $B_k(g, f_k, \beta_k, \tilde{Z}_k)$ represents the matrix in the last constraint of problem (41). Let $(G^0, F^0, \ldots, F^0_K)$ be any transceiver, then the constraints in problem (51) is

$$\beta_k > 0, \tilde{Z}_k \geq 0, B_k(G, F_k, \beta_k, \tilde{Z}_k) \geq 0,$$

$$\min \text{Tr}(Z_k \Sigma_k)/\beta_k \leq p_k, \forall k. \tag{52}$$

If there exist $\beta_k, \tilde{Z}_k$ such that

$$\beta_k > 0, \tilde{Z}_k \geq 0, B_k(G^0, F^0_k, \beta_k, \tilde{Z}_k) \geq 0,$$

$$\text{Tr}(Z_k \tilde{\Sigma}_k)/\beta_k \leq p_k, \forall k \tag{53}$$

is satisfied, then (52) is also satisfied. On the other hand, if (52) is satisfied, then there exist $\beta_k, \tilde{Z}_k$ such that (53) is satisfied. Therefore, the constraint (52) and (53) are equivalent, and the proposition is proved.

APPENDIX A

Replacing the second constraint of problem (29) with its equivalent form (37), (29) becomes

$$\min_{Z_k, \tilde{Z}_k} \text{Tr}(Z_k \Sigma_k)$$

s.t. $Z_k \geq 0, \lambda_k \geq 0, Z_k - \text{diag}([0, 1]) - \lambda_k (Q_k^H Q_k - \text{diag}([0, \epsilon_k - \text{Tr}(F_k R_k F_k^H)]) \geq 0. \tag{50}$$

If $\lambda_k = 0$, then the constraints of (50) reduces to $Z_k - \text{diag}([0, 1]) \geq 0$. Since $\Sigma_k \geq 0$, the optimal $Z_k$ occurs on the

REFERENCES


Xin He received the B.Eng. degree in Electronic Information from SiChuan University, Chengdu, China, in 2007, the M.Eng. degree in Information and Communication from Beijing University of Posts and Telecommunications, in 2010, and the M.Phil. (EEE) degree in 2012 from the University of Hong Kong (HKU). He is currently working toward the Ph.D. degree in HKU. His research interests are optimization and statistics for probabilistic inference and beamforming.