Set Squeezing Procedure for Quadratically Perturbed Chance-Constrained Programming

Xin He and Yik-Chung Wu, *Senior Member, IEEE*

***Abstract*—The set squeezing procedure, a new optimization methodology for solving chance-constrained programming problems under continuous uncertainty distribution, is proposed in this paper. The generally intractable chance constraints and unknown convexity are tackled by a novel analyses of local structure of the feasible set. Based on the newly discovered structure, it is proved that the set squeezing procedure converges and local optimality is guaranteed under mild conditions. Furthermore, efficient algorithms are derived for the set squeezing procedure under the widely used quadratically perturbed constraints. The developed method is applied to the mean squared error (MSE) based probabilistic transceiver design as an application example. Simulation results show that the MSE outage probability can be controlled tightly, which leads to lower transmit power, compared to the existing dominant safe approximation method and the bounded robust optimization method.**

***Index Terms*—Chance-constrained programming, probabilistic transceiver design and beamforming, outage probability constraint.**

I. Introduction

UNCERTAINTY exists in many practical optimization problems, such as classification with modeling uncertainty in machine learning [1], portfolio modeling in finance [2], Lyapunov stability problem with parameter uncertainty in control system [3], and the transceiver design with channel uncertainty in wireless communication system [4]. Two general methods are frequently used to tackle the uncertainty in optimization [5]. One is the bounded robust optimization which models the uncertainty to lie in a bounded set $\mathcal{U}$ and focuses on the worst-case, i.e., $\min\{h(\mathbf{w})|g(\mathbf{w},\mathbf{x}) \leq 0, \forall \mathbf{x} \in \mathcal{U}\}$[6]. The other is chance-constrained programming which imposes probabilistic constraint allowing certain outage probability, i.e., $\min\{h(\mathbf{w})|\Pr\{g(\mathbf{w},\mathbf{x}) \leq 0\} \geq 1-p\}$ [7]. If the probability density function (PDF) of the uncertainty is known, chance-constrained programming provides a more flexible control on the constraint than the bounded robust optimization.

Within the class of chance-constrained programming, the constraint function $g(\mathbf{w},\mathbf{x})$ being quadratic with respect to uncertainty $\mathbf{x}$ occurs in a wide range of research problems. For example,

1) For beamforming with channel uncertainty in wireless communications, $\mathbf{w}$ is the beamforming vector, $\mathbf{x}$ is the channel uncertainty in the channel state information $\mathbf{h} = \hat{\mathbf{h}} + \mathbf{x}$ with $\hat{\mathbf{h}}$ being the estimated channel. One beamforming target is to control the beamforming gain to exceed a certain threshold ε probabilistically, i.e., $\Pr\{|\mathbf{w}^H(\hat{\mathbf{h}} + \mathbf{x})|_2^2 \geq \varepsilon\} \geq 1-p$.
2) For classification with kernel matrix uncertainty in machine learning, $\mathbf{w}$ is the weighting vector to be designed, $\mathbf{X}$ is the uncertainty in the positive semidefinite kernel matrix $\mathbf{K} = \hat{\mathbf{K}} + \mathbf{X}\mathbf{X}^T$ with $\hat{\mathbf{K}}$ being the estimated kernel matrix from training data. The probabilistic constraint in support vector machine (SVM) is $\Pr(\mathbf{w}^T \mathrm{diag}(\mathbf{y})(\hat{\mathbf{K}} + \mathbf{X}\mathbf{X}^T)\mathrm{diag}(\mathbf{y})\mathbf{w} \leq \varepsilon) \geq 1-p$ [8], where $\mathbf{y}$ is the label of the training data.
3) For Lyapunov stability condition $\exists \mathbf{W} \succ 0 : \mathbf{X}^T\mathbf{W}\mathbf{X} - \mathbf{W} \prec 0$ in a discrete dynamic system $\mathbf{z}_{n+1} = \mathbf{X}\mathbf{z}_n$ [3], if the parameter $\mathbf{X}$ contains random uncertainty, the probabilistic stability condition is $\exists \mathbf{W} \succ 0 : \Pr\{\mathbf{X}^T\mathbf{W}\mathbf{X} - \mathbf{W} \prec 0\} \geq 1-p$.

Other applications involving quadratically perturbed optimization include truss topology design [9], quadratic controller synthesis [6] and multiuser transceiver design [4], [10], [11]. The application example in this paper is focused on the transceiver design problem in wireless communication.

However, the challenge of solving chance-constrained programming (CCP) problem is that the integration in the probability function usually does not admit closed-form expression, which hinders the convex analysis of the feasible set. The classic theory in chance-constrained programming reveals that the feasible set is a convex set if $g(\mathbf{w},\mathbf{x})$ is jointly quasiconvex on $\mathbf{w}$ and $\mathbf{x}$, and the PDF of $\mathbf{x}$ is logconcave [12]. Further convexity analysis in [13] shows that the feasible set is convex if the outage probability p is sufficiently small, $g(\mathbf{w},\mathbf{x})$ being a generalized convex function and the PDF of $\mathbf{x}$ is a generalized decreasing function. Other than the above two cases, the convexity or nonconvexity of the feasible set is generally unknown.

Existing methods for CCP problems are proposed under different specific conditions. Under ambiguous density information for uncertainty, the ϕ-divergence is used in [14] to describe the

Manuscript received November 19, 2019; revised August 12, 2020 and December 17, 2020; accepted December 18, 2020. Date of publication December 25, 2020; date of current version February 2, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Caroline Chaux. This work was supported by the National Natural Science Foundation of China under Grant 61807018. *(Corresponding author: Xin He.)*

Xin He is with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Department of Computer and Information Engineering, Hanshan Normal University, Chaozhou 521041, China (e-mail: hx66hx@gmail.com).

Yik-Chung Wu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong 999077 (e-mail: ycwu@eee.hku.hk).

Digital Object Identifier 10.1109/TSP.2020.3047200

uncertainty and a data-driven approach is utilized to solve the CCP problem. On the other hand, a convex safe approximation [15] is used to solve the CCP problem, where the uncertainty is only described by its first and second order moments as well as its support. Under discrete distributed uncertainty, the CCP problem with discrete perturbation is solved in [16], the Wasserstein distance is used in [17] to describe the ambiguous discrete uncertainty and a mixed integer programming method is used to solve the CCP problem. Under special constraint functions, a Monte Carlo based method [18] is proposed to solve the CCP problem, where the constraint function is a convex function of the decision variable, and tractable safe approximation methods are proposed in [19], [20] for CCP problems with linear perturbation. With accurate density information or empirical uncertainty samples, the sample average approximation method is proposed to solve the CCP problem in [21], [22]. Owing to the large number of samples, the large number of constraints and variables make the computation expensive. A projected stochastic subgradient algorithm is proposed in [23] to solve the CCP problem. However, owing to the large iteration number of the stochastic algorithm [23], the computation cost of the projection operation is expensive. In this paper, we focus on the CCP problem with quadratic continuous uncertainty, the methods with ambiguous uncertainties [14], [15], discrete uncertainties [16], [17] and special constraints [18]–[20] are not applicable. Furthermore, without considering the special structure of quadratic perturbation, the sample average approximation methods [21], [22] and the stochastic algorithm [23] are not efficient.

The leading approach for solving CCP problem with quadratic perturbation is safe approximation. For example, Bernstein-type inequality is an effective way to tackle the problem with constraint perturbed by Gaussian uncertainty [24], and has been used to solve probabilistic robust beamforming in downlink multiuser multiple-input single-output (MU-MISO) systems [11], [26], [27]. Furthermore, Vysochanskii-Petunin inequality is suitable for problems with PDF of the perturbed constraint function being unimodal, and it has been proposed for solving chance-constrained power control problem in wireless communications [28], [29]. Markov's inequality is another popular choice as it only requires the first and second order moment information of the uncertainty, and it has been applied to classification with missing data [30]. However, owing to the fact that the feasible set is restricted in these safe approximations, the obtained solution in general is conservative. Recently, a tight solution based on set squeezing procedure is obtained in the probabilistic signal-to-interference-noise-ratio (SINR) constrained beamforming [25]. However, the design and analysis in [25] are tailored to the specific problem of SINR constrained beamforming, which is a special case of the indefinite quadratically perturbed chance-constrained programming. In this paper, the set squeezing procedure is extended for obtaining a tight solution for the general quadratically perturbed chance-constrained programming.

In particular, the local structure of the feasible set of the chance-constrained programming problem is analyzed first. Based on the newly discovered feasible set structure, the properties of set squeezing procedure are established, including the convergence guarantee of the proposed procedure and the property of the converged solution. While the above analyses are valid for general CCP, we further consider the implementation of the set squeezing procedure in quadratically perturbed CCP. Then the set squeezing procedure is demonstrated in multiuser multiantenna systems, with probabilistic mean squared error (MSE) constraints, which is a special case of the definite quadratically perturbed chance-constrained programming. Simulation results show that the set squeezing procedure realizes the outage requirement tightly, and reduces transmit power compared to the safe approximation method.

The rest of this paper is organized as follows. In Section II, the structure of the feasible set of the chance-constrained programming is analyzed, and the property of the set squeezing procedure is established. In Section III, the computational aspect of the set squeezing procedure under quadratically perturbed constraint function is discussed. The application of the proposed set squeezing procedure to multiuser multiantenna system with probabilistic MSE requirements is detailed in Section IV. Simulation results are presented in Section V, and conclusions are drawn in Section VI.

Notation: In this paper, $\mathbb{E}(\cdot)$, $(\cdot)^T$, and $(\cdot)^H$ denote statistical expectation, transposition and Hermitian, respectively, while $\|\cdot\|_2$ denotes the norm of a vector. In addition, $\mathrm{Tr}(\cdot)$ and $\|\cdot\|_F$ refer to the trace and Frobenius norm of a matrix, respectively. The notations $\mathrm{vec}(\cdot)$ and $\otimes$ stand for the vectorization and Kronecker product, respectively. Symbol $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with vector $\mathbf{x}$ on its diagonal, and $\mathbf{I}_K$ is a $K \times K$ identity matrix. For a set $\mathcal{X}$, $\mathrm{int}(\mathcal{X})$ and $\partial(\mathcal{X})$ represent the interior and boundary of $\mathcal{X}$, respectively.

II. Feasible Set Analysis and Set Squeezing Procedure

A class of chance-constrained programming (CCP) is [7]

$$\begin{aligned} \min_{\mathbf{w}} \ & h(\mathbf{w}) \\ \text{s.t.} \ & \Pr\{g(\mathbf{w},\mathbf{x}) \le 0\} \ge 1-p, \end{aligned} \tag{1}$$

where $\mathbf{w}$ is the real-valued or complex-valued decision variable and $\mathbf{x}$ is the real-valued or complex-valued uncertainty random variable with continuous PDF $f(\mathbf{x})$ and support $\mathcal{X}_0$, $g(\mathbf{w},\mathbf{x})$ is continuously differentiable and $h(\mathbf{w})$ is a convex function. The feasible set of problem (1) is denoted as $\mathcal{W}_0$.

The difficulty of solving problem (1) is the integration in the probability function, which usually does not admit closed-form expression, and the leading approaches are to replace the probability function in (1) with its lower bounded probability inequalities, e.g., Bernstein inequality [1], [11], [26], Vysochanskii-Petunin inequality [29] and Markov's inequality [4]. However, these approaches only provide feasible solutions, without assurance of any property. In the following, we will reveal that for any given feasible candidate in $\mathcal{W}_0$, the result of the proposed procedure provides much more information than the simple feasible solution.

A. Feasible Set Analysis on Chance-Constrained Programming

For a given feasible solution $\mathbf{w}_i$ of problem (1), it satisfies $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq 0\} \geq 1 - p$. The probability can be reformulated as

$$\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq 0\} = \int_{\mathbf{x} \in \mathcal{X}_0, g(\mathbf{w}_i, \mathbf{x}) \leq 0} f(\mathbf{x}) d\mathbf{x}. \tag{2}$$

To simplify the notation, we define the set of $\mathbf{x}$ in the integration region of (2) to be

$$\mathcal{X}(\mathbf{w}_i) \triangleq \{\mathbf{x} | \mathbf{x} \in \mathcal{X}_0, \quad g(\mathbf{w}_i, \mathbf{x}) \leq 0\}. \tag{3}$$

We call $\mathcal{X}(\mathbf{w}_i)$ a support subset generated by $\mathbf{w}_i$. By using this support subset, we define another set

$$\mathcal{W}(\mathbf{w}_i) \triangleq \{\mathbf{w} | g(\mathbf{w}, \mathbf{x}) \leq 0, \quad \forall \mathbf{x} \in \mathcal{X}(\mathbf{w}_i)\}, \tag{4}$$

with the following property.

Property 1: $\mathbf{w}_i \in \mathcal{W}(\mathbf{w}_i) \subseteq \mathcal{W}_0$.

Proof: Since $\mathbf{w}_i$ satisfies $g(\mathbf{w}_i, \mathbf{x}) \leq 0$ for all $\mathbf{x} \in \mathcal{X}(\mathbf{w}_i)$ in (3), from the definition in (4), it is obvious that $\mathbf{w}_i \in \mathcal{W}(\mathbf{w}_i)$. Furthermore, any $\mathbf{w}$ in $\mathcal{W}(\mathbf{w}_i)$ is a feasible solution of (1) since

$$\begin{aligned}
&\Pr\{g(\mathbf{w}, \mathbf{x}) \leq 0\} \\
&= \int_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i)} \mathbb{I}(g(\mathbf{w}, \mathbf{x}) \leq 0) \cdot f(\mathbf{x}) d\mathbf{x} \\
&\quad + \int_{\mathbf{x} \in (\mathcal{X}_0) \mathcal{X}(\mathbf{w}_i)} \mathbb{I}(g(\mathbf{w}, \mathbf{x}) \leq 0) \cdot f(\mathbf{x}) d\mathbf{x} && (5) \\
&= \underbrace{\int_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i)} 1 \cdot f(\mathbf{x}) d\mathbf{x}}_{\geq 1-p} \\
&\quad + \underbrace{\int_{\mathbf{x} \in (\mathcal{X}_0) \mathcal{X}(\mathbf{w}_i)} \mathbb{I}(g(\mathbf{w}, \mathbf{x}) \leq 0) \cdot f(\mathbf{x}) d\mathbf{x}}_{\geq 0} && (6) \\
&\geq 1 - p, && (7)
\end{aligned}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Therefore, $\mathcal{W}(\mathbf{w}_i) \subseteq \mathcal{W}_0$. ■

That is to say, each feasible solution $\mathbf{w}_i$ of (1) can generate a set $\mathcal{W}(\mathbf{w}_i)$ which contains $\mathbf{w}_i$ itself, and therefore optimization over $\mathcal{W}(\mathbf{w}_i)$ might find better solution than $\mathbf{w}_i$. We call $\mathcal{W}(\mathbf{w}_i)$ a feasible subset of (1).

Now, observing from the feasible subset definition in (4), the coupling effect between support subset $\mathcal{X}(\mathbf{w}_i)$ and feasible subset $\mathcal{W}(\mathbf{w}_i)$ reveals that deleting some elements in the support subset $\mathcal{X}(\mathbf{w}_i)$ may enlarge the feasible subset $\mathcal{W}(\mathbf{w}_i)$. In particular, we consider the following squeezed support subset

$$\mathcal{X}(\mathbf{w}_i, q_i) \triangleq \{\mathbf{x} | \mathbf{x} \in \mathcal{X}_0, \quad g(\mathbf{w}_i, \mathbf{x}) \leq q_i\}, \tag{8}$$

where q_i is from the following definition,

Definition 1: Let q_i be the solution of $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq q_i\} = 1 - p$.

Since $\mathbf{w}_i$ is a feasible solution, we have $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq 0\} \geq 1 - p$. Compare this with the constraint in (8), and noticing that $g(\mathbf{w}_i, \mathbf{x})$ is a continuous random variable, there must exist $q_i \leq 0$ to make $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq q_i\} = 1 - p$ holds. Furthermore, since $q_i \leq 0$, we have

$$\mathcal{X}(\mathbf{w}_i, q_i) \subseteq \mathcal{X}(\mathbf{w}_i). \tag{9}$$

Now, we define another set generated from the squeezed support subset (8) as

$$\begin{aligned}
\mathcal{W}(\mathbf{w}_i, q_i) &\triangleq \{\mathbf{w} | g(\mathbf{w}, \mathbf{x}) \leq 0, \quad \forall \mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)\} && (10) \\
&= \{\mathbf{w} | \sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) \quad \leq 0\}. && (11)
\end{aligned}$$

The property of $\mathcal{W}(\mathbf{w}_i, q_i)$ can be established as follow.

Property 2: $\mathcal{W}(\mathbf{w}_i) \subseteq \mathcal{W}(\mathbf{w}_i, q_i) \subseteq \mathcal{W}_0$.

Proof: It can be shown with similar derivations from (5) to (7) that $\mathcal{W}(\mathbf{w}_i, q_i) \subseteq \mathcal{W}_0$. Furthermore, according to (9), the pointwise supremum function has following relationship

$$\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i)} g(\mathbf{w}, \mathbf{x}), \tag{12}$$

based on which we have $\{\mathbf{w} | \sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i)} g(\mathbf{w}, \mathbf{x}) \leq 0\} \subseteq \{\mathbf{w} | \sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) \leq 0\}$, and therefore

$$\mathcal{W}(\mathbf{w}_i) \subseteq \mathcal{W}(\mathbf{w}_i, q_i). \tag{13}$$

■

That is, the squeezed support subset $\mathcal{X}(\mathbf{w}_i, q_i)$ in (9) enlarges the corresponding feasible subset $\mathcal{W}(\mathbf{w}_i, q_i)$ in (13). Combining *Properties 1* and *2*, we have $\mathbf{w}_i \in \mathcal{W}(\mathbf{w}_i, q_i)$. Therefore, a solution at least as good as $\mathbf{w}_i$ can be found from $\mathbf{w}_{i+1} = \min\{h(\mathbf{w}) | \mathbf{w} \in \mathcal{W}(\mathbf{w}_i, q_i)\}$. With the new solution $\mathbf{w}_{i+1}$, we have

Definition 2: Let q_{i+1} be the solution of $\Pr\{g(\mathbf{w}_{i+1}, \mathbf{x}) \leq q_{i+1}\} = 1 - p$.

With the new solution $\mathbf{w}_{i+1}$ and q_{i+1}, the next feasible subset $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ is constructed as that in (10). A crucial question is if $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ contains new candidates, which makes further improvement possible? We answer this question in the following.

Theorem 1: If $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous on $\mathbf{w}$ and $\mathbf{w}_i \in \mathrm{int}(\mathcal{W}_0)$, then $\mathbf{w}_i \in \mathrm{int}(\mathcal{W}(\mathbf{w}_i, q_i))$.

Proof: Since $\mathbf{w}_i \in \mathrm{int}(\mathcal{W}_0)$, we have $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq 0\} > 1 - p$. Therefore, $q_i < 0$ is required to make $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq q_i\} = 1 - p$ hold. Due to the definition of $\mathcal{X}(\mathbf{w}_i, q_i) = \{\mathbf{x} | \mathbf{x} \in \mathcal{X}_0, \quad g(\mathbf{w}_i, \mathbf{x}) \leq q_i\}$, we have $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}_i, \mathbf{x}) \leq q_i < 0$. By the assumption that $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous, there exists a ball $\mathcal{B}(\mathbf{w}_i, r) = \{\mathbf{w} | \|\mathbf{w} - \mathbf{w}_i\| < r, r > 0\}$ with center $\mathbf{w}_i$ and radius r such that $|\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) - \sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}_i, \mathbf{x})| < -q_i$ for all $\mathbf{w} \in \mathcal{B}(\mathbf{w}_i, r)$, which implies $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) < 0$ for all $\mathbf{w} \in \mathcal{B}(\mathbf{w}_i, r)$. According to the definition of $\mathcal{W}(\mathbf{w}_i, q_i)$ in (11), we have $\mathcal{B}(\mathbf{w}_i, r) \subseteq \mathcal{W}(\mathbf{w}_i, q_i)$. Since $\mathbf{w}_i$ is the center of a ball with radius $r > 0$, $\mathbf{w}_i$ is an interior point of $\mathcal{W}(\mathbf{w}_i, q_i)$. ■

Theorem 1 can be used to reveal the inter-relationship between $\mathcal{W}(\mathbf{w}_i, q_i)$ and $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ as follows.

Theorem 2: Given $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous, if the feasible solution $\mathbf{w}_{i+1}$ is a boundary point of $\mathcal{W}(\mathbf{w}_i, q_i)$ and an interior point of $\mathcal{W}_0$, (i.e., $\mathbf{w}_{i+1} \in \partial(\mathcal{W}(\mathbf{w}_i, q_i))$ and $\mathbf{w}_{i+1} \in \mathrm{int}(\mathcal{W}_0)$), then $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}) \cap \mathcal{W}(\mathbf{w}_i, q_i) \neq \varnothing$ and $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}) \setminus \mathcal{W}(\mathbf{w}_i, q_i) \neq \varnothing$.

Proof: First, since $g(\mathbf{w}, \mathbf{x})$ is continuous, $g(\mathbf{w}, \mathbf{x})$ is lower semicontinuous (l.s.c.). Furthermore, since the pointwise supremum preserves the l.s.c. property, $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ is l.s.c. and its sublevel set $\{\mathbf{w}|\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x}) \leq c, c \in \mathbb{R}\}$ is closed [31, p. 31]. Therefore, feasible subsets $\mathcal{W}(\mathbf{w}_i)$ and $\mathcal{W}(\mathbf{w}_i, q_i)$ are closed. Since $\mathcal{W}(\mathbf{w}_i, q_i)$ is closed, $\mathbf{w}_{i+1} \in \partial(\mathcal{W}(\mathbf{w}_i, q_i))$ implies $\mathbf{w}_{i+1} \in \mathcal{W}(\mathbf{w}_i, q_i)$. Combining with the result $\mathbf{w}_{i+1} \in \mathcal{W}(\mathbf{w}_{i+1}) \subseteq \mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ owing to (13), we have $\mathbf{w}_{i+1} \in (\mathcal{W}(\mathbf{w}_i, q_i) \cap \mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}))$.

Second, $\mathbf{w}_{i+1} \in \text{int}(\mathcal{W}_0)$ implies $\mathbf{w}_{i+1} \in \text{int}(\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}))$ under continuous $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ owing to *Theorem 1*. Furthermore, $\mathbf{w}_{i+1} \in \text{int}(\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}))$ means there exists a ball $\mathcal{B}(\mathbf{w}_{i+1}, r_1)$ with center $\mathbf{w}_{i+1}$ and radius $r_1 > 0$ such that $\mathcal{B}(\mathbf{w}_{i+1}, r_1) \subseteq \mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$. On the other hand, $\mathbf{w}_{i+1} \in \partial(\mathcal{W}(\mathbf{w}_i, q_i))$ means there exists a point $\mathbf{w} \in \mathcal{B}(\mathbf{w}_{i+1}, r_2)$ such that $\mathbf{w} \notin \mathcal{W}(\mathbf{w}_i, q_i)$ for all $r_2 > 0$. Therefore, if we consider $r_2 < r_1$, then $\mathbf{w} \in (\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1}) \setminus \mathcal{W}(\mathbf{w}_i, q_i))$. ∎

Theorem 2 reveals the condition for $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ to contain new feasible candidate that does not belong to $\mathcal{W}(\mathbf{w}_i, q_i)$, which enables sequential optimizations to be presented next.

B. Set Squeezing Procedure and Its Properties

Based on the local structure of the feasible subsets of the chance-constrained programming, a set squeezing procedure can be used to solve (1). In particular, the set squeezing procedure is an iteration between the following two steps until convergence [25].

- ***P-step***: Update $\mathcal{X}(\mathbf{w}_i, q_i)$ in (8) by finding q_i such that $\Pr\{g(\mathbf{w}_i, \mathbf{x}) \leq q_i\} = 1 - p$.
- ***O-step***: Solve $\min\{h(\mathbf{w})|\mathbf{w} \in \mathcal{W}(\mathbf{w}_i, q_i)\}$, and denote the solution as $\mathbf{w}_{i+1}$. Increment i by one.

The P-step refers to probability evaluation, and the O-step refers to optimization. The inter-connected sequential feasible subsets enable the convergence of the set squeezing procedure as follows.

Definition 3: If $h(\mathbf{w}_{i+1}) \leq h(\mathbf{w}_i)$, then $\mathbf{w}_{i+1}$ is a descent solution from $\mathbf{w}_i$ at O-step.

Lemma 1: If $h(\mathbf{w})$ is bounded below or $\mathcal{W}_0$ is compact, and if $\mathbf{w}_{i+1}$ is a descent solution from $\mathbf{w}_i$ at O-step, the convergence of the set squeezing procedure is guaranteed.

Proof: Since $\mathbf{w}_i \in \mathcal{W}(\mathbf{w}_i)$ and $\mathcal{W}(\mathbf{w}_i) \subseteq \mathcal{W}(\mathbf{w}_i, q_i)$ in (13), $\mathbf{w}_i \in \mathcal{W}(\mathbf{w}_i, q_i)$ is established, which makes optimization descent from $\mathbf{w}_i$ possible. Furthermore, a descent solution from $\mathbf{w}_i$ results in $h(\mathbf{w}_{i+1}) \leq h(\mathbf{w}_i)$. Therefore, monotonic decreasing property is established for the set squeezing procedure. With a bounded below objective function or a compact feasible set, the convergence is guaranteed. ∎

Although the inter-connected sequential feasible subsets enable convergence of the set squeezing procedure, if the function $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ is not continuous, new feasible candidates are not guaranteed in sequential feasible subsets, which makes improvement in objective function value not guaranteed beyond the first iteration. On the other hand, if $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous, new feasible candidates are available in the next feasible subset $\mathcal{W}(\mathbf{w}_{i+1}, q_{i+1})$ according to *Theorem 2*, which leads to the nature of the converged solution described as follows.

Lemma 2: If Lemma 1 holds with local optimal solution at O-step and $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous, the limit solution $\mathbf{w}^\star$ is a local optimum of (1) or a boundary solution of (1).

Proof: Obviously, the limit solution $\mathbf{w}^\star$ lies in the boundary or the interior of $\mathcal{W}_0$. Let $q^\star$ satisfies $\Pr\{g(\mathbf{w}^\star, \mathbf{x}) \leq q^\star\} = 1 - p$ in (8). If the limit solution $\mathbf{w}^\star$ occurs in the interior of $\mathcal{W}_0$, (i.e., $\Pr\{g(\mathbf{w}^\star, \mathbf{x}) \leq 0\} > 1 - p$), we have $\mathbf{w}^\star \in \text{int}(\mathcal{W}(\mathbf{w}^\star, q^\star))$ owing to *Theorem 1*. Since $\mathbf{w}^\star$ is a local optimal solution at O-step, $\mathbf{w}^\star$ is a local optimum in the feasible subset $\mathcal{W}(\mathbf{w}^\star, q^\star)$. Combining with the facts that $\mathbf{w}^\star \in \text{int}(\mathcal{W}(\mathbf{w}^\star, q^\star))$ and $\mathcal{W}(\mathbf{w}^\star, q^\star) \subseteq \mathcal{W}_0$, the local neighborhood of $\mathbf{w}^\star$ in $\mathcal{W}(\mathbf{w}^\star, q^\star)$ is the same as that in $\mathcal{W}_0$, which implies $\mathbf{w}^\star$ is a local optimal solution of (1). ∎

Intuitively, the P-step enables new feasible candidates available in the inter-connected sequential feasible subsets if the function $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous, and the descent solutions at O-step ensure convergence. Note that inequality-based safe approximation methods only provide a feasible solution, while the proposed set squeezing procedure guarantees a local optimum or a boundary solution under mild conditions. Even without the local optimality, the monotonic decreasing property of the set squeezing procedure reveals whether there is possible improvement after just one iteration from the inequality based initialization. Furthermore, the above analyses are valid for a general CCP problem, and can be considered as the generalization of [25].

III. Set Squeezing Procedure for Quadratically Perturbed CCP

Although the proposed set squeezing procedure is conceptually simple, practical implementation might not be trivial. For example, the optimization problem in O-step is infinitely constrained, and in general is a difficult problem. Furthermore, obtaining the tradeoff between complexity and accuracy of the numerical integration in P-step is of interest from signal processing perspective. In this section, we present discussions on how these two steps can be realized for the popular class of $g(\mathbf{w}, \mathbf{x})$ being quadratic with respect to $\mathbf{x}$.

First, we present the sufficient condition to guarantee the function $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ being continuous, a condition critical for the set squeezing procedure to generate new feasible candidates in sequential optimizations.

Theorem 3: If the continuous function $g(\mathbf{w}, \mathbf{x})$ is a quadratic function with respect to $\mathbf{x}$, any one of the following conditions guarantees $\sup_{\mathbf{x}\in\mathcal{X}(\mathbf{w}_i,q_i)} g(\mathbf{w}, \mathbf{x})$ being continuous,

- The uncertainty support set $\mathcal{X}_0$ is unbounded.
- The boundary of $\mathcal{X}_0$ is characterized by a real-valued quadratic function in complex space.

Proof: See Appendix A ∎

The support set of many common random variables are unbounded, e.g., Gaussian, Laplace and t-distribution. Furthermore, the bounded uncertainty in optimization problem is usually modeled to lie within an ellipsoid [6], [32], which can be described by a quadratic function. Therefore, many uncertainty models in practical applications can be covered by *Theorem 3*. Next, we present specific details on the O-step and P-step.

A. Handling Infinite Constraints in O-Step

The optimization problem to be solved at the O-step is

$$\min_{\mathbf{w}} h(\mathbf{w})$$
$$\text{s.t. } g(\mathbf{w},\mathbf{x}) \leq 0, \quad \forall \mathbf{x} \in \mathcal{X}_0 : g(\mathbf{w}_i,\mathbf{x}) \leq q_i. \tag{14}$$

Problem (14) is in general a difficult one due to the infinite constraints. Complicated numerical methods exist for solving (14) if $g(\mathbf{w},\mathbf{x})$ is a quasiconvex function [34]. Fortunately, for the widely used quadratic function $g(\mathbf{w},\mathbf{x})$, the infinite constraints can be transformed into finite constraints.

First consider the case when the support of $\mathbf{x}$ is unbounded, i.e., $\mathcal{X}_0 = \mathbb{C}^n$. By using the S-lemma [35], [36], and due to $g(\mathbf{w},\mathbf{x})$ is in quadratic form with respect to $\mathbf{x}$, i.e., $g(\mathbf{w},\mathbf{x}) \triangleq \mathbf{x}_e^H \mathbf{A}(\mathbf{w})\mathbf{x}_e$, $g(\mathbf{w}_i,\mathbf{x}) \triangleq \mathbf{x}_e^H \mathbf{A}(\mathbf{w}_i)\mathbf{x}_e$ with $\mathbf{x}_e \triangleq [\mathbf{x}^T, 1]^T$, the constraint in (14) can be equivalently formulated as

$$\exists \lambda \geq 0 : \quad \lambda \cdot (\mathbf{A}(\mathbf{w}_i) - \text{diag}([\mathbf{0}, q_i])) - \mathbf{A}(\mathbf{w}) \succeq 0. \tag{15}$$

On the other hand, if the support of $\mathbf{x}$ is characterized by a quadratic function $\mathcal{X}_0 = \{\mathbf{x} | \mathbf{x}_e^H \mathbf{B} \mathbf{x}_e \leq 0, \mathbf{B}^H = \mathbf{B}, \mathbf{x}_e \triangleq [\mathbf{x}^T, 1]^T, \mathbf{x} \in \mathbb{C}^n\}$, the S-Lemma in complex space [37] allows the quadratic constraint in (14) be equivalently reformulated as

$$\exists \lambda \geq 0, \beta \geq 0 : \quad \lambda \cdot (\mathbf{A}(\mathbf{w}_i) - \text{diag}([\mathbf{0}, q_i])) + \beta \cdot \mathbf{B} - \mathbf{A}(\mathbf{w}) \succeq 0. \tag{16}$$

An example of using this idea was demonstrated in [25]. If the matrix function $\mathbf{A}(\mathbf{w})$ is a linear function of $\mathbf{w}$, (15) and (16) are linear matrix inequality (LMI) constraints. If $\mathbf{A}(\mathbf{w})$ is a quadratic function of $\mathbf{w}$, Schur complement can be used to further reformulate the quadratic matrix inequalities (15) and (16) into LMI. For other situations, nonlinear terms in the matrix inequality might be lifted to an LMI representation by introducing slack variables [36]. We will illustrate the use of this method in the application example in the next section.

B. Efficient Probability Evaluation in P-Step

For a given feasible solution $\mathbf{w}_i$, the support subset update is to find the quantile q_i such that

$$\Pr\{g(\mathbf{w}_i,\mathbf{x}) \leq q_i\} = 1 - p. \tag{17}$$

Equation (17) can be solved by bisection method if the cumulative distribution function (CDF) of $g(\mathbf{w}_i,\mathbf{x})$ is known. A straightforward method for approximating CDF of $g(\mathbf{w}_i,\mathbf{x})$ is to use Monte Carlo numerical methods. Generating samples from a given distribution is a well-studied topic in statistics, and a number of methods, such as inversion method and importance sampling, exist [39]. The details are not discussed here. Note that if the PDF of $\mathbf{x}$ is unknown, empirical samples from the uncertainty are needed for the Monte Carlo method.

However, the computational complexity of the Monte Carlo methods would be high if a very accurate q_i is required. Fortunately, saddlepoint approximation provides efficient and accurate probability evaluation if the cumulant-generating function (CGF) $k(t)$ of $g(\mathbf{w}_i,\mathbf{x})$ is known. For example, under $g(\mathbf{w}_i,\mathbf{x})$ being quadratic and $\mathbf{x}$ is Gaussian distributed, its CGF can be derived from Chi-squared distribution [40]. After obtaining the CGF, in the bisection procedure, a candidate q_i determines the saddlepoint t_0 from $k'(t_0) = q_i$. Then the second order saddlepoint approximation of the probability in (17) is [41, p. 53]

$$\Pr(g(\mathbf{w}_i,\mathbf{x}) \leq q_i) \simeq \Phi(u) + \phi(u) \cdot \left\{ \frac{1}{u} - \frac{1}{v} - v^{-1}\left(\frac{O_4}{8} - \frac{5}{24}(O_3)^2\right) + v^{-3} + \frac{O_3}{2v^2} - u^{-3} \right\}, \tag{18}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of standard normal distribution, $u = \text{sign}(t_0)\sqrt{2(t_0 \cdot q_i - k(t_0))}$, $v = t_0\sqrt{k''(t_0)}$, $O_n = k^{(n)}(t_0)/\{k''(t_0)\}^{n/2}$, with $n = \{3, 4\}$ are the normalized high order derivatives. Note that [44] proposed a different way to evaluate the probability in (17) and a closed-form solution is obtained for the case with central quadratic form. However, the application of [25] and our following application are in noncentral quadratic form.

Since both Monte Carlo and saddlepoint methods are numerical methods, we need to guarantee the realized outage probability $\hat{p} \triangleq 1 - \Pr(g(\mathbf{w}_i,\mathbf{x}) \leq q_i)$ to lie within a small interval ϵ from the target p with reliability $1 - \delta$. For Monte Carlo method, the number of independent samples required to guarantee $\Pr(|\hat{p} - p| \leq \epsilon) \geq 1 - \delta$ is $\frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$ [42, p. 114]. Taking $p - \epsilon$ as the modified outage target guarantees the realized outage probability $\hat{p} \in [p - 2\epsilon, p]$ with reliability $1 - \delta$. On the other hand, for saddlepoint method using (18), uniformity of relative error $\hat{p}/p$ is preserved over the entire range of support of $\mathbf{x}$ [40], and $\hat{p}/p$ can be computed if $g(\mathbf{w},\mathbf{x})$ is quadratic and $\mathbf{x}$ is Gaussian distributed [43]. Therefore, if the outage probability target is predistorted by the relative error, accurate outage probability can be achieved.

IV. Probabilistic MSE Constrained Multiantenna Transceiver Design

In order to illustrate the set squeezing procedure, transceiver design in the downlink multiuser multiantenna system is considered. In particular, the system consists of one base station (BS) equipped with N transmit antennas, and K active users with the kth user equipped with M_k antennas. L_k independent data streams are transmitted to the kth user and $\sum_{k=1}^{K} L_k = L$. To guarantee data recovery, it is required that $L_k \leq M_k$ and $L \leq N$. Let $\mathbf{G}$ be the $N \times L$ precoding matrix at BS, $\mathbf{H}_k$ and $\mathbf{F}_k$ are the $M_k \times N$ channel matrix and the $L_k \times M_k$ equalizer of the kth user, respectively. The received $M_k \times 1$ noise vector at the kth user is Gaussian distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ with $\mathbf{R}_k \succ 0$. With the transmitted symbols being independent with zero-mean and unit power, the MSE at the kth user can be derived as $\text{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) = \|\mathbf{F}_k\mathbf{H}_k\mathbf{G} - \mathbf{D}_k\|_F^2 + \text{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)$, where $\mathbf{D}_k = [\mathbf{0}_{L_k \times \sum_{k=1}^{k-1} L_k} \quad \mathbf{I}_{L_k} \quad \mathbf{0}_{L_k \times \sum_{k=k+1}^{K} L_k}]$ [4], [32].

Assuming the obtained channel information $\hat{\mathbf{H}}_k$ is perfect, the nonrobust MSE constrained transceiver design with MSE target ε_k is $\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^K}\{\|\mathbf{G}\|_F|\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\hat{\mathbf{H}}_k)\le\varepsilon_k,\forall k\}$ [45]. By modeling the channel uncertainty $\boldsymbol{\Delta}_k$ lying in a bounded region $\mathcal{U}_k$, the bounded robust optimization aims to guarantees the MSE requirement under all channel uncertainty, i.e., $\min\{\|\mathbf{G}\|_F|\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k=\hat{\mathbf{H}}_k+\boldsymbol{\Delta}_k)\le\varepsilon_k,\forall\boldsymbol{\Delta}_k\in\mathcal{U}_k,\forall k\}$ [32]. In this paper, we consider another class of transceiver design problems, where the PDF of $\boldsymbol{\Delta}_k$ is known, and the transceiver design aims at minimizing transmit power at the BS under probabilistic MSE constraints for different users

$$\begin{aligned}\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^K}\quad & \|\mathbf{G}\|_F\\ \text{s.t.}\quad & \Pr\{\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)\le\varepsilon_k\}\ge 1-p_k,\quad\forall k,\end{aligned}\tag{19}$$

where p_k is the MSE outage probability at the kth receiver. In the following, we consider the widely used unbiased linear channel estimators, where the vectorized channel uncertainty $\mathrm{vec}(\boldsymbol{\Delta}_k)$ can be modeled by Gaussian distributed $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma}_k)$ [46], [48]. Application of set squeezing procedure to (19) is illustrated as follows.

Initialization: Since $\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)$ and the MSE target ε_k are positive, by applying the Markov's inequality to the constraint of (19), a safe approximation to (19) is obtained as

$$\begin{aligned}\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^K}\quad & \|\mathbf{G}\|_F\\ \text{s.t.}\quad & \mathbb{E}_{\mathbf{H}_k}\{\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)\}/\varepsilon_k\le p_k,\quad\forall k.\end{aligned}\tag{20}$$

Taking expectation in (20) and expressing $\mathbf{F}_k\triangleq\frac{1}{a_k}\tilde{\mathbf{F}}_k$ with positive amplitude $1/a_k$ and receiver beamforming direction $\tilde{\mathbf{F}}_k$, the initial transceiver with fixed receiver beamforming direction $\tilde{\mathbf{F}}_k$ can be obtained from the second order cone programming (SOCP) problem

$$\begin{aligned}\min_{\mathbf{G},\{a_k\}_{k=1}^K}\quad & \|\mathbf{G}\|_F\\ \text{s.t.}\quad & \Big\|\Big[\mathrm{vec}\Big((\mathbf{G}^T\otimes\tilde{\mathbf{F}}_k)\boldsymbol{\Sigma}_k^{\frac{1}{2}}\Big)^T\quad \mathrm{vec}(\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G}-a_k\mathbf{D}_k)^T\\ & \mathrm{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})^T\Big]\Big\|_2\le a_k\sqrt{p_k\varepsilon_k},\quad\forall k.\end{aligned}\tag{21}$$

Popular choices of $\tilde{\mathbf{F}}_k$ are identity matrix and singular vector matrix of $\hat{\mathbf{H}}_k$ [32], [45]. If (21) is not feasible for the above two choices, random $\tilde{\mathbf{F}}_k$ can be used in a trial-and-error fashion.

P-step: For a given feasible solution $(\mathbf{G}^{[i]},\{\mathbf{F}_k^{[i]}\}_{k=1}^K)$, according to (17) the P-step is to find the quantile $q_k^{[i]}$ such that

$$\begin{aligned}&\Pr\Big(\|\mathbf{F}_k^{[i]}\mathbf{H}_k\mathbf{G}^{[i]}-\mathbf{D}_k\|_F^2+\mathrm{Tr}(\mathbf{F}_k^{[i]}\mathbf{R}_k(\mathbf{F}_k^{[i]})^H)\le q_k^{[i]}\Big)\\ &=1-p_k,\end{aligned}\tag{22}$$

which can be solved by probability evaluation with bisection candidate $q_k^{[i]}\in[\mathrm{Tr}(\mathbf{F}_k^{[i]}\mathbf{R}_k(\mathbf{F}_k^{[i]})^H),\varepsilon_k]$. Since the channel uncertainty $\mathrm{vec}(\boldsymbol{\Delta}_k)$ is Gaussian distributed, the CGF of the random variable $\|\mathbf{F}_k^{[i]}\mathbf{H}_k\mathbf{G}^{[i]}-\mathbf{D}_k\|_F^2$ can be obtained. In particular, let the singular value decomposition of $((\mathbf{G}^{[i]})^T\otimes\mathbf{F}_k^{[i]})(\boldsymbol{\Sigma}_k/2)^{\frac{1}{2}}=\mathbf{U}[\mathrm{diag}([\sigma_1,\ldots,\sigma_{LL_k}]),\mathbf{0}_{LL_k\times(NM_k-LL_k)}]\mathbf{V}^H$ with descending singular values, it is shown in *Appendix IX* that the CGF of $\|\mathbf{F}_k^{[i]}\mathbf{H}_k\mathbf{G}^{[i]}-\mathbf{D}_k\|_F^2$ is

$$k(t)=\sum_{j=1}^{LL_k}\left(\frac{|\eta_j|^2\sigma_j^2t}{1-2\sigma_j^2t}-\ln(1-2\sigma_j^2t)\right),\tag{23}$$

with its domain $(-\infty,1/(2\sigma_1^2))$, and η_j is the jth element of the vector $[\mathbf{I}_{LL_k},\mathbf{0}_{LL_k\times(NM_k-LL_k)}]\mathbf{V}^H(\boldsymbol{\Sigma}_k/2)^{-1/2}\mathrm{vec}(\hat{\mathbf{H}}_k)-\mathrm{diag}([1/\sigma_1,\ldots,1/\sigma_{LL_k}])\mathbf{U}^H\mathrm{vec}(\mathbf{D}_k)$. Then, the saddlepoint t_0 is calculated through $k'(t_0)=q_k^{[i]}-\mathrm{Tr}(\mathbf{F}_k^{[i]}\mathbf{R}_k(\mathbf{F}_k^{[i]})^H)$ by bisection in the domain $t_0\in(-\infty,1/(2\sigma_1^2))$. Note that the uniqueness of the saddlepoint is guaranteed by the fact that the CGF $k(t)$ is strictly convex, i.e., $k'(t)$ is strictly monotonically increasing. Finally, the probability in (22) is evaluated using equation (18).

O-step: With a feasible transceiver pair $(\mathbf{G}^{[i]},\{\mathbf{F}_k^{[i]}\}_{k=1}^K)$ and the quantile $q_k^{[i]}$ obtained in the P-step, the corresponding problem in O-step is

$$\begin{aligned}\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^K}\quad & \|\mathbf{G}\|_F\\ \text{s.t.}\quad & \mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)\le\varepsilon_k,\forall\mathbf{H}_k:\\ & \mathrm{MSE}_k(\mathbf{G}^{[i]},\mathbf{F}_k^{[i]},\mathbf{H}_k)\le q_i^{[i]},\forall k.\end{aligned}\tag{24}$$

Since $\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)$ is a quadratic function of $\mathbf{H}_k$, by making use of (15), it is shown in *Appendix X* that the infinite constrained problem (24) can be equivalently transformed into the following finite constrained problem

$$\begin{aligned}\min_{\mathbf{G},\{a_k,\tilde{\mathbf{F}}_k,c_k,\lambda_k\}_{k=1}^K}\quad & \|\mathbf{G}\|_F\\ \text{s.t. }\lambda_k\ge 0,\ & \begin{bmatrix}\lambda_k\mathbf{A}^{[i]}+\mathrm{diag}([\mathbf{0},a_k\varepsilon_k-c_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & a_k\mathbf{I}_{LL_k}\end{bmatrix}\succeq 0,\ \forall k\\ a_k>0,\quad & \|[2\mathrm{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})^T,a_k-c_k]\|_2\le a_k+c_k,\quad\forall k,\end{aligned}\tag{25}$$

where $\mathbf{Q}_k\triangleq[\mathbf{G}^T\otimes\tilde{\mathbf{F}}_k\quad\mathrm{vec}(-a_k\mathbf{D}_k)]$, $\mathbf{A}_k^{[i]}=(\mathbf{Q}_k^{[i]})^H\mathbf{Q}_k^{[i]}-\mathrm{diag}([\mathbf{0},q_k^{[i]}-\mathrm{Tr}(\mathbf{F}_k^{[i]}\mathbf{R}_k(\mathbf{F}_k^{[i]})^H)])$ with $\mathbf{Q}_k^{[i]}=[(\mathbf{G}^{[i]})^T\otimes\mathbf{F}_k^{[i]}\quad\mathrm{vec}(-\mathbf{D}_k)]$.

The product of $\mathbf{G}$ and $\tilde{\mathbf{F}}_k$ in $\mathbf{Q}_k$ makes (25) a nonconvex problem. However, the bilinear relationship of $\mathbf{G}$ and $\tilde{\mathbf{F}}_k$ enables the conventional alternating optimization between $\mathbf{G}$ and all $\mathbf{F}_k=\tilde{\mathbf{F}}_k/a_k$.[1] With $(\mathbf{G}^{[i]},\{\mathbf{F}_k^{[i]}\}_{k=1}^K)$ as initialization, the obtained transmit power $\|\mathbf{G}\|_F^2$ in each alternating iteration decreases monotonically and converges [32], [45], and a descent solution from $(\mathbf{G}^{[i]},\{\mathbf{F}_k^{[i]}\}_{k=1}^K)$ at O-step is guaranteed.

Summary and Further Properties: The proposed set squeezing procedure for the transceiver design problem starts with an initialization, and follows an iteration between (22) at P-step and (25) at O-step until the difference between successive transmit power is smaller than a pre-defined threshold. Since descent

[1] Without loss of generality, we can set $a_k=1,\forall k$.

solution is guaranteed at O-step, according to *Lemma 1*, the set squeezing procedure converges.

For the special case of MU-MISO systems, all k users are equipped with single antenna. In this case, the factor $1/a_k$ captures the amplitude information and $\tilde{\mathbf{F}}_k$ reduces to a complex scalar, which only causes a phase rotation of $\mathbf{G}$, and does not change the optimum of (25) [4], [10]. Therefore, without loss of generality, we can fix $\tilde{\mathbf{F}}_k = 1, \forall k$, and the optimal transceiver $(\mathbf{G}, \{\frac{1}{a_k}\}_{k=1}^K)$ can be efficiently solved from the resulting semidefinite programming (SDP) problem. Since optimal solution can be obtained from SDP problem in the O-step, and it can be proved by contradiction that local optimal solutions of (19) cannot occur at the interior of the feasible set of (19), according to *Lemma 2*, the limit solution of the set squeezing procedure lies on the boundary of the feasible set of problem (19) in MU-MISO systems.

For MU-MIMO system, although *Lemma 2* cannot be directly used to establish the property of the limit solution due to lack of guarantee of local optimal solution at O-step, but thanks to the special structure of the transceiver design problem, we can obtain the following result.

Proposition 1: The limit solution of the set squeezing procedure activates all users' constraints in problem (19).

Proof: With fixed equalizers $\{\mathbf{F}_k\}_{k=1}^K$, let the optimal precoder of convex subproblem (24) be $\mathbf{G}^\star$. If the kth user's constraint of (24) is not active with precoder $\mathbf{G}^\star$, the kth user's beamforming vectors (the $1 + \sum_{m=1}^{k-1} L_m$ to $\sum_{m=1}^{k} L_m$ column vectors of $\mathbf{G}^\star$) can be scaled down until the kth user's constraint is active. This power scaling operation reduces the objective function and other users' MSEs strictly, which contradicts the optimality of $\mathbf{G}^\star$. Therefore, the optimal precoder $\mathbf{G}^\star$ for given $\{\mathbf{F}_k\}_{k=1}^K$ activates all K user's constraints in subproblem (24).

With fixed precoder $\mathbf{G}$, optimization variable in (24) now only appears in the constraints, and the problem to be solved becomes to minimize the guaranteed MSEs $\bar{\varepsilon}_k$ in convex problem $\min\{\bar{\varepsilon}_k | \mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) \leq \bar{\varepsilon}_k, \quad \forall \mathbf{H}_k : \mathrm{MSE}_k(\mathbf{G}^{[i]}, \mathbf{F}_k^{[i]}, \mathbf{H}_k) \leq q_k^{[i]}\}$. With the obtained optimum $\{\bar{\varepsilon}_k\}_{k=1}^K$, if any inequality $\{\bar{\varepsilon}_k \leq \varepsilon_k\}_{k=1}^K$ is not active, the solution $(\mathbf{G}, \{\mathbf{F}_k\}_{k=1}^K)$ does not activate all K users' constraints in (24), which makes the next optimal precoder with fixed equalizers $\{\mathbf{F}_k\}_{k=1}^K$ reduce transmit power strictly, as discussed in the first paragraph of this proof. Therefore, the alternating optimization of $\mathbf{G}$ and $\{\mathbf{F}_k\}_{k=1}^K$ converges to a solution $(\mathbf{G}^{[i+1]}, \{\mathbf{F}_k^{[i+1]}\}_{k=1}^K)$ which activates all K constraints in (24).

If the solution $(\mathbf{G}^{[i+1]}, \{\mathbf{F}_k^{[i+1]}\}_{k=1}^K)$ does not activate the kth constraint in the original problem (19), according to *Theorem 1*, $(\mathbf{G}^{[i+1]}, \{\mathbf{F}_k^{[i+1]}\}_{k=1}^K)$ would not activate the corresponding constraint in the next O-step subproblem (24). From the discussion in the first paragraph of this proof, we obtain $\|\mathbf{G}^{[i+2]}\|_F < \|\mathbf{G}^{[i+1]}\|_F$. Therefore, the set squeezing procedure would generate transceiver solutions with strictly decreasing transmit power in successive O-steps, as long as any constraint in (19) is not active. That is, the limit solution activates all constraints in (19). ∎

Note that the optimal solution of problem (19) activates all users' chance constraints. Under this condition, a calibration method based on the safe approximation (21) might be conceived.[2] In particular, given a safe approximation solution, its realized MSE outage is smaller than the outage target, the parameter p_k in problem (21) can be kept increasing, e.g., $\bar{p}_k = 2^n p_k$ with $n = 1, 2, \cdots$, such that the realized outage of the tuned safe approximation (21) with parameter $\bar{p}_k$ is larger than the outage target. Then bisection is applied between the last two parameter $\bar{p}_k$ until the realized outage of the tuned safe approximation (21) reaches the outage target.

Since the realized outage of the tuned safe approximation will bounce up and down around the outage target, while that of the proposed method is monotonically increasing to the outage target, the proposed method is expected to converge quicker than the calibration method. Therefore, the iteration number of the proposed method would be smaller than that of the calibration method. Note that the precondition to use the calibration method is that the optimal solution activates all the chance constraints, which is the case in the transceiver design problem. However, in other applications, this precondition might not hold, and the convergence of the calibration method may not be guaranteed.

For the computational complexity in each iteration, the computational complexity of the proposed method in P-step is slightly larger than that of the calibration method owing to several extra calculations in (18), the major complexity differences come from the optimization in O-step, where the computational complexity of the proposed SDP method is $\mathcal{O}(N^6)$, which is larger than the calibration method (SOCP problem) with complexity order $\mathcal{O}(N^{4.5})$.

V. Simulation Results and Discussions

In this section, the performance of the set squeezing procedure in MU-MIMO transceiver design is illustrated. In the simulation, unless specified otherwise, the BS is equipped with four antennas and there are two active users, i.e., $N = 4, K = 2$. The MIMO channel model is the widely used Kronecker model $\mathbf{H}_k = \mathbf{\Psi}_{rk}^{\frac{1}{2}} \mathbf{H}_w \mathbf{\Psi}_t^{\frac{1}{2}}$, where $\mathbf{H}_w$ has complex Gaussian entries with zero mean and unit variance, $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_{rk}$ are the correlation matrices at the BS and the kth user respectively. Channel correlation matrices are taken as the exponential model $[\mathbf{\Psi}_{t,k}]_{ij} = \rho_{t,k}^{|i-j|}$ and $[\mathbf{\Psi}_{r,k}]_{ij} = \rho_{r,k}^{|i-j|}$, the correlation coefficient of the first user is set as $\rho_{t,1} = 0.2, \rho_{r,1} = 0.5$, while that of other users are set as $\rho_{t,k} = 0.3, \rho_{r,k} = 0.6$. The received Gaussian noise at every antenna is independent and identically distributed with zero mean and variance $\delta_n^2 = 0.01$, i.e., $\mathbf{R}_k = \delta_n^2 \mathbf{I}_{M_k}$. Of the two users, the second user is equipped with one antenna ($M_2 = 1$) with fixed MSE requirement $\varepsilon_2 = 0.2$ and $p_2 = 10\%$, while the setting for the first user is varied to test the set squeezing procedure under different scenarios. All simulation results are averaged over 10^3 random channel realizations. The optimization problems are solved on a laptop PC with Intel i7 CPU (3.6 GHz) and 16 GB RAM, using the parser CVX and solver SDPT3 [47].

For the set squeezing procedure, the bisection accuracy in finding the quantile $q_k^{[i]}$ is 0.01%, and the bisection accuracy

[2]The calibration method was suggested by an anonymous reviewer.

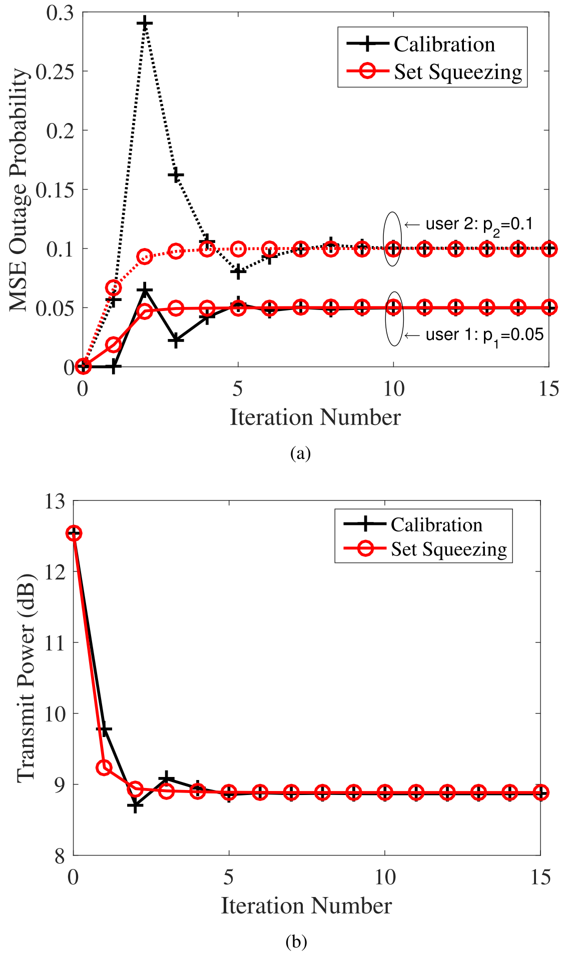


Fig. 1. Convergence performance of different methods under one channel realizations.

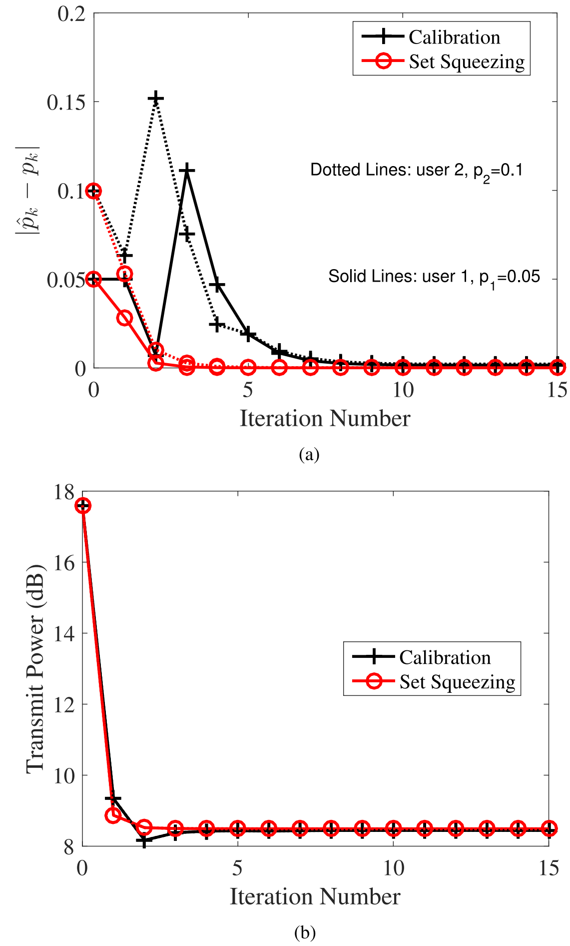


Fig. 2. Convergence performance of different methods under multiple channel realizations.

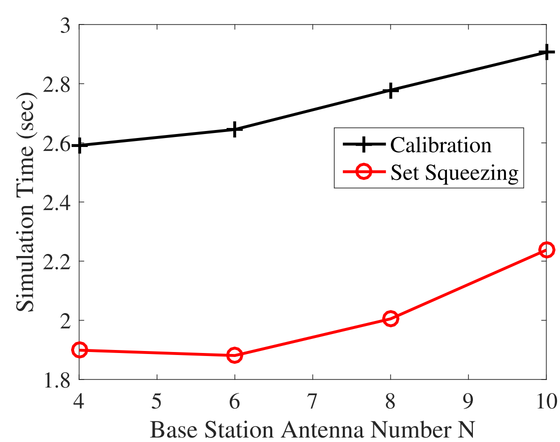


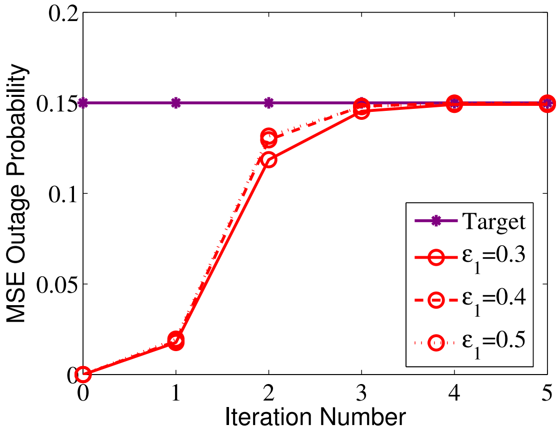
Fig. 3. Simulation times of different methods.

for the saddle point t_0 is 10^{-8}. To backoff the relative error of the saddlepoint method, all outage targets are predistorted as $p_k/1.015$ [43]. The relative power difference $|\|\mathbf{G}^{[i+1]}\|_F^2 - \mathbf{G}^{[i]}\|_F^2|/\|\mathbf{G}^{[i]}\|_F^2 \leq 5 \times 10^{-3}$ is set as the termination criterion.

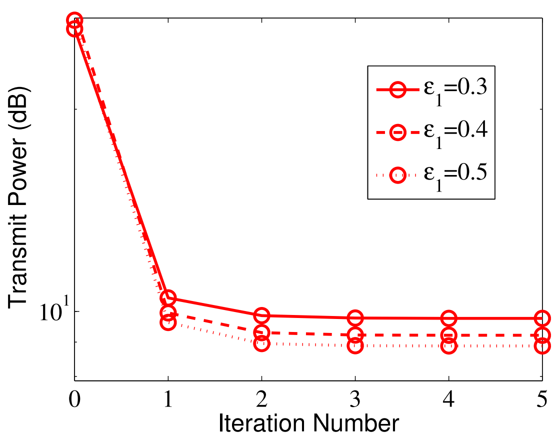
A. *Gaussian Channel Estimation Error*

In time division duplex system, owing to the channel reciprocity property, the BS takes the estimated uplink channel as the downlink channel. Therefore, the channel uncertainty comes from the estimation error, which is usually modelled as Gaussian distribution. With the linear minimum mean square error channel estimator, the channel estimation error covariance matrix is $\mathbf{\Sigma}_k = ((\mathbf{\Psi}_t^{-T} \otimes \mathbf{\Psi}_{rk}^{-1}) + \frac{P_t^2}{\delta_n^2}\mathbf{I}_{NM_k})^{-1}$ [4], [48]. In the following, the pilot-to-noise ratio is set as $\frac{P_t^2}{\delta_n^2} = 10^2$ (i.e., 20 dB).

The convergence performance of the set squeezing procedure is compared with the calibration method at Fig. 1 under one channel realization. With $p_1 = 5\%, \varepsilon_1 = 0.1$, Fig. 1(a) shows that the realized MSE outage probability ($\hat{p}_k$) of the set squeezing procedure monotonically approaches to the outage target, while that of the calibration method oscillates around the outage target. Under multiple channel realizations, Fig. 2(a) shows that the proposed set squeezing procedure converges quicker than the calibration method. Note that the transmit power oscillation behavior of Fig. 1(b) is averaged out in Fig. 2(b) under multiple channel realizations. Owing to the quick convergence speed of the proposed method, its simulation time (the average time between the initialization and the converged solution) is smaller than that of the calibration method as shown in Fig. 3. Note that the results with "Iteration Number 0" in Figs. 1 and 2 are the initialization results [4].



(a)



(b)

Fig. 4. Convergence performance of the set squeezing procedure for multiple data streams.

With multiple data streams in the second user, $M_1 = L_1 = 2$ and $p_1 = 15\%$, the convergence performance of the set squeezing procedure is illustrated at Fig. 4. Fig. 4(a) shows that the outage probabilities gradually approach the outage target, irrespective to the MSE requirement value ε_1. Furthermore, it is noticed that the outage probabilities is very close to the 15% outage target at the second iteration, the remaining space to reduce the transmit power is small, and therefore the transmit powers in Fig. 4(b) decrease slowly after the second iteration. Furthermore, from Fig. 4(b), it is observed that the proposed set squeezing procedure reduces the transmit powers significantly in the first iteration, owing to the large feasible subset provided by the set squeezing procedure.

With eight transmit antennas and four users, i.e., $N = 8, K = 4, \{M_k = 1\}_{k=1}^4, \varepsilon_1 = 0.1, \{\varepsilon_k = 0.2\}_{k=2}^4$, the convergence performance of the set squeezing procedure is illustrated at Fig. 5. Fig. 5 shows that the realized outage probabilities of all four users converge to their outage targets within three iterations. Figs. 2, 4 and 5 reveal that the proposed method converges quickly under different settings.

Next, we compare the performance of the set squeezing procedure to that of non-robust method [45] and safe approximation method. In particular, the safe approximation based on

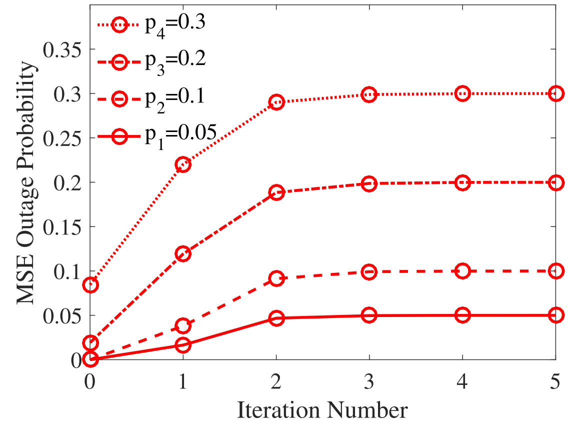
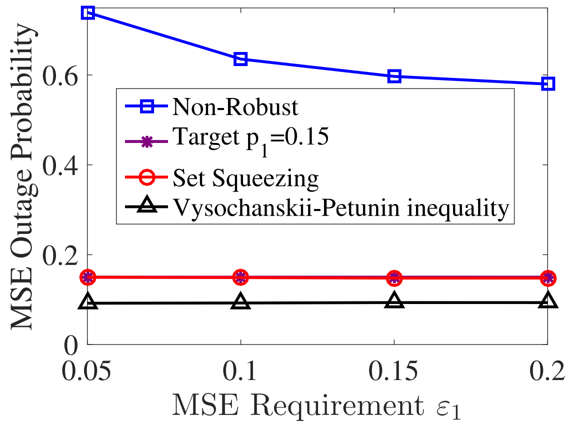


Fig. 5. The convergence performance of the set squeezing procedure for four users.

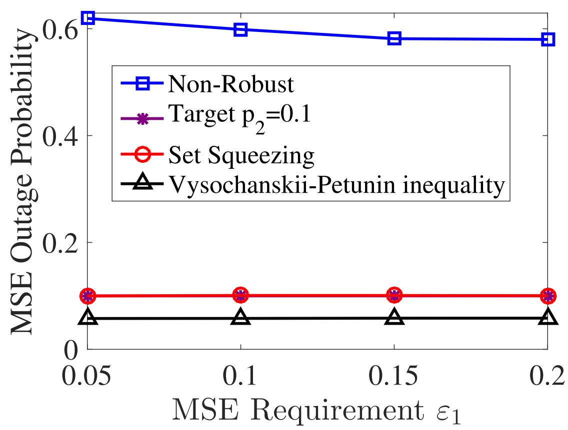
Vysochanskii-Petunin (V-P) inequality [29] was chosen for comparison since it is specially designed for Gaussian uncertainty, the proposed procedure uses the feasible solution in [29] as the initial solution. For fair comparison, the channel realizations are chosen to be feasible for the proposed and the other two methods, and the first user is equipped with single antenna ($M_1 = 1$), as V-P method is only valid for MU-MISO system. In Fig. 6, the converged MSE outage probability (Fig. 6(a) for user 1 and Fig. 6(b) for user 2) and transmit power are shown versus ε_1 with $p_1 = 15\%$ and $p_2 = 10\%$. It is observed from Figs. 6(a) and 6(b) that the MSE outage probability targets for both users are realized tightly by the set squeezing procedure over a wide range of MSE requirement ε_1, while the non-robust method fails to satisfy the outage requirements and the V-P method reaches conservative probabilities of 9% and 5% for the first user and the second user, respectively. As a result of the tightly controlled outage performance from the set squeezing procedure, 0.5 to 1 dB transmit power gain is achieved compared to the V-P method as shown in Fig. 6(c). Since the set squeezing procedure in this example takes the feasible solution of the V-P method as the initial solution, both method have the same feasibility rates, which are around 78% and 98% for $\varepsilon_1 = 0.05$ and $\varepsilon_1 = 0.2$, respectively. On the other hand, in Fig. 7, MSE outage probability (for user 1 only, results for user 2 is similar to Fig. 6(b)) and transmit power are shown versus outage requirement p_1 with $\varepsilon_1 = 0.2$. It can be observed that similar conclusions can be drawn as in Fig. 6, i.e., the set squeezing procedure realizes the outage target tightly, and achieves about 1 dB transmit power saving compared to the V-P method.

B. Uniform Channel Quantization Error

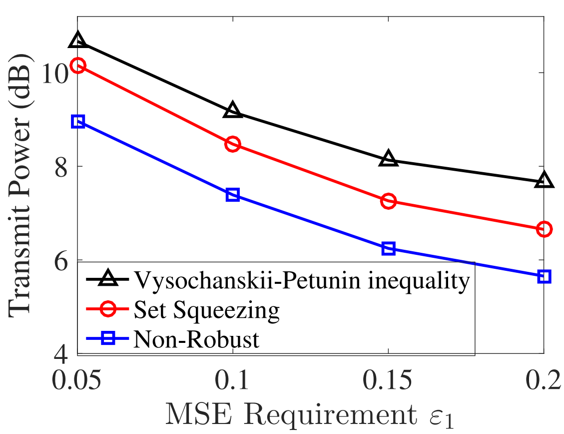
In frequency division duplex system, the base station takes the quantized channel from the mobile user as the downlink channel. Therefore, the channel uncertainty mainly comes from the quantization error. Under high-resolution quantization, it is common to assume that the quantization error is uniform distributed in a sphere [49], [50]. In particular, the channel uncertainty is modelled to lie in a sphere $\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|_F^2 \leq r_k^2$ with radius $r_k = \sqrt{2NM_k}b$, where b is the error bound per dimension in bounded



(a)



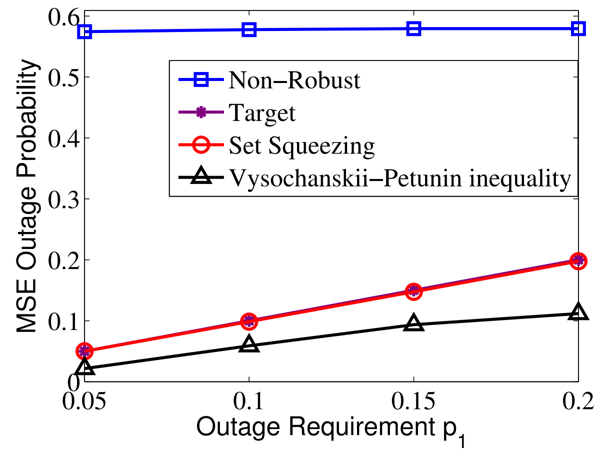
(b)



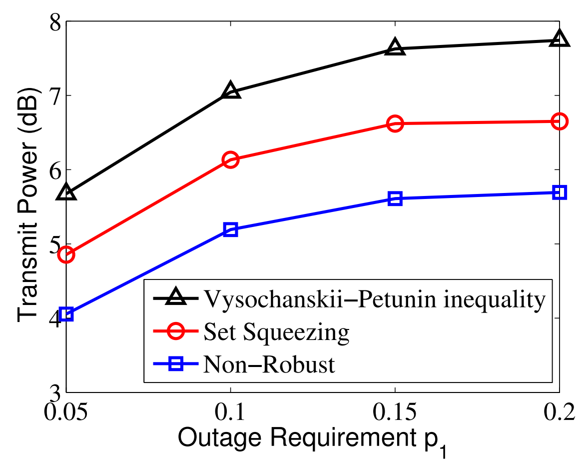
(c)

Fig. 6. Achieved MSE outage probability and transit power versus different MSE requirements.

robust transceiver design [32]. Under this model, the support of the random channel in (16) is $\{\mathbf{H}_k | \mathbf{x}_e^H \mathbf{B}_k \mathbf{x}_e \leq 0, \mathbf{x}_e = [\text{vec}(\mathbf{H}_k)^T, 1]^T, \mathbf{B}_k = [\begin{smallmatrix} \mathbf{I}_{NM_k} & -\text{vec}(\hat{\mathbf{H}}_k) \\ -\text{vec}(\hat{\mathbf{H}}_k)^H & \|\hat{\mathbf{H}}_k\|_F^2 - r_k^2 \end{smallmatrix}]\}$, and the LMI in (25) becomes $[\begin{smallmatrix} \beta_k \mathbf{B}_k + \lambda_k \mathbf{A}^{[i]} + \text{diag}([\mathbf{0}, a_k \varepsilon_k - c_k]) & \mathbf{Q}_k^H \\ \mathbf{Q}_k & a_k \mathbf{I}_{LL_k} \end{smallmatrix}] \succeq 0$ with extra slack variable $\beta_k \geq 0$. Since saddlepoint method cannot be applied to this case, Monte Carlo method is used in the P-step. In particular, 2^{16} uniform quantization error samples are



(a)



(b)

Fig. 7. Achieved MSE outage probability and transit power versus different outage targets under $\varepsilon_1 = 0.2$.

generated, which guarantees the probability evaluation accuracy within 1% from the true value with reliability 99.999%. Two data streams are transmitted to the first user with two antennas ($M_1 = 2$), and the MSE requirement for the first user is $\varepsilon_1 = 0.4$.

Fig. 8 compares the performance of the non-robust method [45], bounded robust method [32] and the set squeezing procedure (with [32] as initialization) with b ranges from 0.01 to 0.21. It is observed from Fig. 8(a) that the proposed method guarantees MSE target to be fulfilled tightly under different channel uncertainty bounds, while the MSE outage probability of the non-robust method are larger than 52% and increases with the increasing quantization error norm. For the bounded robust method, it achieves a strictly zero outage, but at the cost of very high transmit power as shown in Fig. 8(b). On the other hand, the transmit power performance of the proposed method in Fig. 8(b) shows a 5 to 12 dB saving from the bounded robust method, and is close to that of non-robust method. In summary, the non-robust method fails to satisfy the MSE requirement and the bounded robust method is conservative owing to the worst-case target, while the proposed set squeezing procedure provides a flexible balance between stabilizing MSE requirement and saving power.

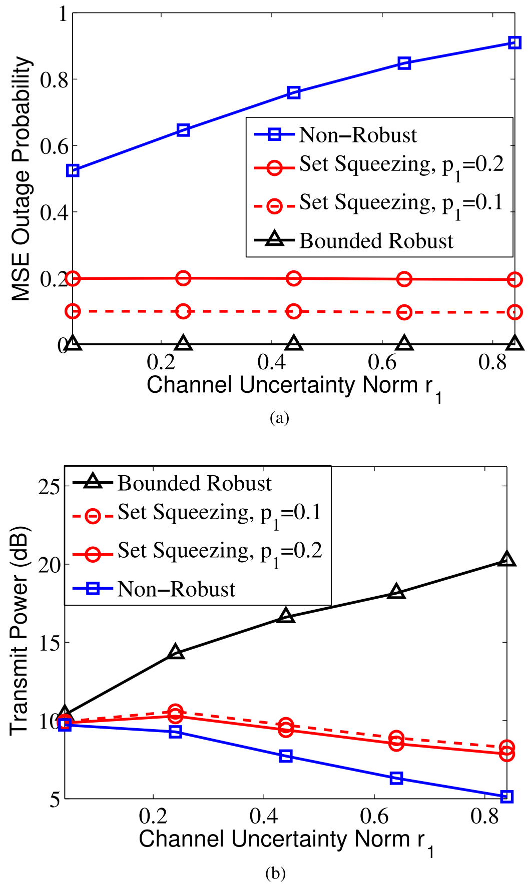


Fig. 8. Achieved MSE outage probability and transit power versus different channel uncertainty error bounds.

VI. Conclusion

In this paper, by exploiting the local structure of the feasible set, the set squeezing procedure was established to solve the general quadratically perturbed chance constrained programming problem. With the convergence and local optimality of the set squeezing procedure guaranteed under mild conditions, the set squeezing procedure overcome the conservative nature of the safe approximation method and the bounded robust optimization method. With probabilistic MSE constrained transceiver design as an example, simulation results validated that the MSE outage can be controlled tightly, which leads to lower transmit power while achieving the target MSE outage probability.

Appendix A
Proof of the Theorem 3

We will write the generic quadratic function in $\mathbf{x}$ as a quadratic form in $\mathbf{x}_e \triangleq [\mathbf{x}^T, 1]^T$, so that the continuous quadratic function is $g(\mathbf{w}, \mathbf{x}) \triangleq \mathbf{x}_e^H \mathbf{A}(\mathbf{w}) \mathbf{x}_e$, where $\mathbf{A}(\mathbf{w})$ is a finite dimension matrix with elements continuously depend on $\mathbf{w}$. If the support set $\mathcal{X}_0$ is unbounded, by using S-Lemma [35], [36], the feasible subset $\mathcal{W}(\mathbf{w}_i, q_i)$ in (10) can be generated from a linear matrix inequality (LMI) $\mathbf{\Phi}(\lambda, \mathbf{w}) \succeq 0$ with $\mathbf{\Phi}(\lambda, \mathbf{w}) = \lambda(\mathbf{A}(\mathbf{w}_i) - \mathrm{diag}([\mathbf{0}, q_i])) - \mathbf{A}(\mathbf{w})$ and $\lambda \geq 0$. The LMI requirement is equivalent to $\inf_{\lambda \geq 0}\{-\min \mathrm{eig}[\mathbf{\Phi}(\lambda, \mathbf{w})]\} \leq 0$, where $\mathrm{eig}[\cdot]$ is the eigenvalues of $\mathbf{\Phi}(\lambda, \mathbf{w})$. Compare with (11), we obtain $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) = \inf_{\lambda \geq 0}\{-\min \mathrm{eig}[\mathbf{\Phi}(\lambda, \mathbf{w})]\}$. Note that the pointwise supremum operation preserves the lower semicontinuous (l.s.c.) property [31, p. 23], $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is l.s.c.. Furthermore, since all eigenvalues of $\mathbf{\Phi}(\lambda, \mathbf{w})$ with fixed λ depend continuously on $\mathbf{w}$ [33, p. 130], the pointwise minimum of finite continuous functions is still a continuous function, i.e., $\min \mathrm{eig}[\mathbf{\Phi}(\lambda, \mathbf{w})]$ with fixed λ depends continuously on $\mathbf{w}$. Since the pointwise infimum operation preserves the upper semicontinuous (u.s.c.) property, $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) = \inf_{\lambda \geq 0}\{-\min \mathrm{eig}[\mathbf{\Phi}(\lambda, \mathbf{w})]\}$ is u.s.c.. Therefore, $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is both l.s.c. and u.s.c., which implies $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ is continuous.

On the other hand, if the boundary function of $\mathcal{X}_0$ is a real-valued quadratic function in complex space, $\mathcal{X}_0 = \{\mathbf{x} | \mathbf{x}_e^H \mathbf{B} \mathbf{x}_e \leq 0, \mathbf{B}^H = \mathbf{B}, \mathbf{x}_e \triangleq [\mathbf{x}^T, 1]^T, \mathbf{x} \in \mathbb{C}^n\}$. It can be shown using S-Lemma in complex space [37] that $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x}) = \inf_{\lambda \geq 0, \beta \geq 0}\{-\min \mathrm{eig}[\mathbf{\Phi}(\lambda, \beta, \mathbf{w})]\}$ with $\mathbf{\Phi}(\lambda, \beta, \mathbf{w}) = \lambda(\mathbf{A}(\mathbf{w}_i) - \mathrm{diag}([\mathbf{0}, q_i])) + \beta \mathbf{B} - \mathbf{A}(\mathbf{w})$, and the continuous property of $\sup_{\mathbf{x} \in \mathcal{X}(\mathbf{w}_i, q_i)} g(\mathbf{w}, \mathbf{x})$ still holds with similar proof as above.

Appendix B
Finding the CGF of $\|\mathbf{F}_k^{[i]} \mathbf{H}_k \mathbf{G}^{[i]} - \mathbf{D}_k\|_F^2$

With the estimated channel $\hat{\mathbf{H}}_k$, and for circularly-symmetric jointly-complex-Gaussian (CSCG) channel uncertainty $\mathrm{vec}(\mathbf{\Delta}_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$, the random channel is $\mathbf{h}_k = \mathrm{vec}(\mathbf{H}_k) \sim \mathcal{CN}(\mathrm{vec}(\hat{\mathbf{H}}_k), \mathbf{\Sigma}_k)$. Then

$$
\begin{aligned}
&\|\mathbf{F}_k^{[i]} \mathbf{H}_k \mathbf{G}^{[i]} - \mathbf{D}_k\|_F^2 \\
&= \|((\mathbf{G}^{[i]})^T \otimes \mathbf{F}_k^{[i]}) \mathbf{h}_k - \mathrm{vec}(\mathbf{D}_k)\|_2^2, \qquad (26) \\
&= \|((\mathbf{G}^{[i]})^T \otimes \mathbf{F}_k^{[i]}) \bar{\mathbf{\Sigma}}_k^{\frac{1}{2}} \cdot \ddot{\mathbf{h}}_k - \mathrm{vec}(\mathbf{D}_k)\|_2^2, \qquad (27)
\end{aligned}
$$

where $\bar{\mathbf{\Sigma}}_k = \mathbf{\Sigma}_k/2$, and decorrelation $\ddot{\mathbf{h}}_k = \bar{\mathbf{\Sigma}}_k^{-\frac{1}{2}} \mathbf{h}_k$ ensures independent between elements in $\ddot{\mathbf{h}}_k$, i.e., $\ddot{\mathbf{h}}_k \sim \mathcal{CN}(\bar{\mathbf{\Sigma}}_k^{-\frac{1}{2}} \mathrm{vec}(\hat{\mathbf{H}}_k), 2\mathbf{I}_{NM_k})$. Now, we perform singular value decomposition

$$
\begin{aligned}
&((\mathbf{G}^{[i]})^T \otimes \mathbf{F}_k^{[i]}) \bar{\mathbf{\Sigma}}_k^{\frac{1}{2}} \\
&= \mathbf{U} \left[\underbrace{\mathrm{diag}([\sigma_1, \ldots, \sigma_{LL_k}])}_{\mathbf{D}_\sigma} \quad \mathbf{0}_{LL_k \times (NM_k - LL_k)} \right] \mathbf{V}^H, \qquad (28)
\end{aligned}
$$

where the singular values $\{\sigma_j\}_{j=1}^{LL_k}$ are ordered in descending order, and $\mathbf{U}, \mathbf{V}$ are unitary matrices. Since there are L independent data streams transmitted and L_k independent data stream received at the kth user, feasible solution $((\mathbf{G}^{[i]})^T \otimes \mathbf{F}_k^{[i]})$ must have rank LL_k, which implies all singular values in (28) are positive. Putting (28) into (27), we have

$$
\begin{aligned}
&\|\mathbf{F}_k^{[i]} \mathbf{H}_k \mathbf{G}^{[i]} - \mathbf{D}_k\|_F^2 \\
&= \|[\mathbf{D}_\sigma, \mathbf{0}_{LL_k \times (NM_k - LL_k)}] \mathbf{V}^H \ddot{\mathbf{h}}_k - \mathbf{U}^H \mathrm{vec}(\mathbf{D}_k)\|_2^2 \qquad (29) \\
&= \Big\| \mathbf{D}_\sigma \Big([\mathbf{I}_{LL_k}, \mathbf{0}_{LL_k \times (NM_k - LL_k)}] \\
&\quad \times \mathbf{V}^H \ddot{\mathbf{h}}_k - \mathbf{D}_\sigma^{-1} \mathbf{U}^H \mathrm{vec}(\mathbf{D}_k) \Big) \Big\|_2^2. \qquad (30)
\end{aligned}
$$

Since the unitary transformed term $\mathbf{V}^H\ddot{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{V}^H\bar{\boldsymbol{\Sigma}}_k^{-\frac{1}{2}}\text{vec}(\hat{\mathbf{H}}_k), 2\mathbf{I}_{NM_k})$ is an independent and identically distributed complex Gaussian vector, the statistical representation of (30) is

$$\|\mathbf{F}_k^{[i]}\mathbf{H}_k\mathbf{G}^{[i]} - \mathbf{D}_k\|_F^2 \sim \sum_{j=1}^{LL_k} \sigma_j^2 \chi^2_{(|\eta_j|^2,2)}, \tag{31}$$

which is a weighted sum of independent noncentral chi-squared variables $\chi^2_{(|\eta_j|^2,2)}$ with degrees of freedom two, and η_j is the jth element of the vector $[\mathbf{I}_{LL_k}, \mathbf{0}_{LL_k\times(NM_k-LL_k)}]\cdot \mathbf{V}^H(\boldsymbol{\Sigma}_k/2)^{-1/2}\text{vec}(\hat{\mathbf{H}}_k) - \text{diag}([1/\sigma_1, \ldots, 1/\sigma_{LL_k}])\mathbf{U}^H\text{vec}(\mathbf{D}_k)$.

Finally, since the moment-generating function of $\chi^2_{(|\eta_j|^2,2)}$ is $\exp(|\eta_j|^2 t/(1-2t))/(1-2t)$ with domain $2t < 1$, the CGF of the sum of independent random variables in (31) is

$$k(t) = \sum_{j=1}^{LL_k}\left(\frac{|\eta_j|^2\sigma_j^2 t}{1-2\sigma_j^2 t} - \ln(1-2\sigma_j^2 t)\right), \tag{32}$$

with domain $(-\infty, 1/(2\sigma_1^2))$, owing to the reason that σ_1 is the largest singular value.

Appendix C
Transforming Optimization Problem (24) Into (25)

Using the factorized equalizer $\mathbf{F}_k \triangleq \frac{1}{a_k}\tilde{\mathbf{F}}_k$ with $a_k > 0$, multiplying a_k on both sides of the inequality in the kth constraint of (24), we get

$$\begin{aligned} &a_k > 0, a_k\text{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) \le a_k\varepsilon_k,\\ &\forall \mathbf{H}_k : \text{MSE}_k(\mathbf{G}^{[i]}, \mathbf{F}_k^{[i]}, \mathbf{H}_k) \le q_k^{[i]}. \end{aligned} \tag{33}$$

Putting the MSE expression into the first MSE inequality of (33), we have

$$\begin{aligned} &a_k\text{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) - a_k\varepsilon_k\\ &= a_k\left\|\frac{1}{a_k}\tilde{\mathbf{F}}_k\mathbf{H}_k\mathbf{G} - \mathbf{D}_k\right\|_F^2 + \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H) - a_k\varepsilon_k \end{aligned} \tag{34}$$

$$\begin{aligned} &= \mathbf{u}_k^H\left(\frac{1}{a_k}\mathbf{Q}_k^H\mathbf{Q}_k\right.\\ &\quad\left. - \text{diag}\left(\left[\mathbf{0}_{1\times NM_k}, a_k\varepsilon_k - \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)\right]\right)\right)\mathbf{u}_k, \end{aligned} \tag{35}$$

where $\mathbf{u}_k \triangleq [\text{vec}(\mathbf{H}_k)^T, 1]^T$ and $\mathbf{Q}_k \triangleq [\mathbf{G}^T \otimes \tilde{\mathbf{F}}_k \;\; \text{vec}(-a_k\mathbf{D}_k)]$. Similarly, with the MSE expression applied to the second MSE inequality of (33), we have $\mathbf{u}_k^H\mathbf{A}_k^{[i]}\mathbf{u}_k \le 0$, where $\mathbf{A}_k^{[i]} = (\mathbf{Q}_k^{[i]})^H\mathbf{Q}_k^{[i]} - \text{diag}([\mathbf{0}, q_k^{[i]} - \text{Tr}(\mathbf{F}_k^{[i]}\mathbf{R}_k(\mathbf{F}_k^{[i]})^H)])$ with $\mathbf{Q}_k^{[i]} = [(\mathbf{G}^{[i]})^T \otimes \mathbf{F}_k^{[i]} \;\; \text{vec}(-\mathbf{D}_k)]$. Therefore, (33) becomes

$$\begin{aligned} &a_k > 0, \mathbf{u}_k^H\left(\frac{1}{a_k}\mathbf{Q}_k^H\mathbf{Q}_k\right.\\ &\quad\left. - \text{diag}\left(\left[\mathbf{0}, a_k\varepsilon_k - \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)\right]\right)\right)\mathbf{u}_k \le 0,\\ &\forall \mathbf{u}_k : \quad \mathbf{u}_k^H\mathbf{A}_k^{[i]}\mathbf{u}_k \le 0. \end{aligned} \tag{36}$$

Now, (36) is in the same form as the constraint in (14) with unbounded support set. Therefore, applying (15), equation (36) is transformed as

$$\begin{aligned} &\exists \lambda_k \ge 0 : \quad a_k > 0,\\ &\lambda_k\mathbf{A}^{[i]} + \text{diag}\left(\left[\mathbf{0}, a_k\varepsilon_k - \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)\right]\right)\\ &\quad - \frac{1}{a_k}\mathbf{Q}_k^H\mathbf{Q}_k \succeq 0. \end{aligned} \tag{37}$$

By using the Schur complement, (37) can be equivalently formulated as

$$\begin{aligned} &\exists \lambda_k \ge 0 :\\ &\begin{bmatrix} \lambda_k\mathbf{A}^{[i]} + \text{diag}([\mathbf{0}, a_k\varepsilon_k - \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & a_k\mathbf{I}_{LL_k} \end{bmatrix}\\ &\succeq 0, \quad a_k > 0. \end{aligned} \tag{38}$$

Introducing a slack variable c_k with $c_k \ge \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)$, the feasible set $(\mathbf{G}, a_k, \tilde{\mathbf{F}}_k, \lambda_k)$ of (38) is lifted to a higher dimension set $(\mathbf{G}, a_k, \tilde{\mathbf{F}}_k, \lambda_k, c_k)$ as follows

$$\begin{aligned} &\exists \lambda_k \ge 0 : \quad \begin{bmatrix} \lambda_k\mathbf{A}^{[i]} + \text{diag}([\mathbf{0}, a_k\varepsilon_k - c_k]) & \mathbf{Q}_k^H\\ \mathbf{Q}_k & a_k\mathbf{I}_{LL_k} \end{bmatrix} \succ 0,\\ &\qquad a_k > 0, \quad c_k \ge \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H). \end{aligned} \tag{39}$$

Note that any feasible solution $(\mathbf{G}, a_k, \tilde{\mathbf{F}}_k, \lambda_k)$ in ((39)) satisfy the constraint in (38). Furthermore, any feasible solution in (38) must satisfy the constraint in (39) with $c_k = \frac{1}{a_k}\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H)$. Therefore, the lifting process does not change the feasible set of (38) on $(\mathbf{G}, a_k, \tilde{\mathbf{F}}_k, \lambda_k)$.

Since the noise covariance matrix $\mathbf{R}_k \succ 0$, $\text{Tr}(\tilde{\mathbf{F}}_k\mathbf{R}_k\tilde{\mathbf{F}}_k^H) \ge 0$. Therefore, the last two constraints of (39) are equivalent to

$$a_k > 0, \quad c_k \ge 0, \quad a_kc_k \ge \|\text{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})\|_2^2, \tag{40}$$

which can be expressed as the following SOCP form

$$a_k > 0, \quad \|[2\text{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})^T, a_k - c_k]\|_2 \le a_k + c_k. \tag{41}$$

References

[1] A. Ben-Tal, S. Bhadra, C. Bhattacharyya, and J. S. Nath, “Chance constrained uncertain classification via robust optimization,” *Math. Program.*, vol. 127, no. 1, pp. 145–173, 2011.

[2] A. Prékopa, K. Yodaa, M. M. Subasib, “Uniform quasi-concavity in probabilistic constrained stochastic programming,” *Operations Res. Lett.*, vol. 39, no. 3, pp. 188–192, 2011.

[3] M. C. de Oliveira, J. Bernussou, and J. C. Geromel, “A new discrete-time robust stability condition,” *Syst. Control Lett.*, vol. 37, no. 4, pp. 261–265, 1999.

[4] X. He and Y. C. Wu, “Probabilistic QoS constrained robust downlink multiuser MIMO transceiver design with arbitrarily distributed channel uncertainty,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6292–6302, Dec. 2013.

[5] V. Gabrela, C. Murata, and A. Thiele, “Recent advances in robust optimization: An overview,” *Eur. J. Oper. Res.*, vol. 235, no. 3, pp. 471–483, Jun. 2014.

[6] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton Series in Applied Mathematics, Princeton, NJ, USA: Princeton Univ. Press, 2009.

[7] A. Nemirovski, “On safe tractable approximations of chance constraints,” *Eur. J. Oper. Res.*, vol. 219, no. 3, pp. 707–718, 2012.

[8] A. Ben-Tal, S. Bhadra, C. Bhattacharyya, A. Nemirovski, "Efficient methods for robust classification under uncertainty in kernel matrices," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 2923–2954, 2012.

[9] A. Ben-Tal and A. Nemirovski, "Robust truss topology design via semidefinite programming," *SIAM J. Optim.*, vol. 7, no. 4, pp. 991–1016, 1997.

[10] N. Vučić and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 714–725, Feb. 2009.

[11] K.-Y. Wang, T. H. Chang, W.-K. Ma, A. M.-C. So, and C.-Y. Chi, "Probabilistic SINR constrained robust transmit beamforming: A bernsteintype inequality based conservative approach," in *Proc. ICASSP*, pp. 3080–3083, 2011.

[12] A. Prékopa, *Stochastic Programming*. Dordrecht, The Netherlands: Kluwer, 1995.

[13] R. Henrion and C. Strugarek, "Convexity of chance constraints with independent random variables," *Comput. Optim. Appl.*, vol. 41, no. 2, pp. 263–276, 2008.

[14] R. Jiang, Y. Guan, "Data-driven chance constrained stochastic program," *Math. Program.*, vol. 158, no. 1-2, pp. 291–327, 2016.

[15] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, vol. 137, no. 1–2, pp. 167–198, 2013.

[16] J. Luedtke, "A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support," *Math. Program.*, vol. 146, no. 1-2, pp. 219–244, 2014.

[17] W. Xie, "On distributionally robust chance constrained programs with wasserstein distance," *Math. Program.*, 2019. [Online]. Available: https://doi.org/10.1007/s10107-019-01445-5

[18] L. Jeff Hong, Y. Yang, and L. Zhang, "Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach," *Operations Res.*, vol. 59, no. 3, pp. 617–630, 2011.

[19] W. Chen, M. Sim, J. Sun, and C. P. Teo, "From CVaR to uncertainty set: Implications in joint chance-constrained optimization," *Operations Res.*, vol. 58, no. 2, pp. 470–485, 2010.

[20] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2007.

[21] Y. Cao and V. M. Zavala, "A sigmoidal approximation for chance-constrained nonlinear programs," 2020, *arXiv:2004.02402*.

[22] A. Peña-Ordieres, J. R. Luedtke, and A. Wachter, "Solving chance-constrained problems via a smooth sample-based nonlinear approximation," 2019, *arXiv:1905.07377*.

[23] R. Kannan, and J. Luedtke, "A stochastic approximation method for chance-constrained nonlinear programs," 2018, *arXiv:1812.07066*.

[24] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," 2009, *arXiv:0909.3595*.

[25] X. He and Y.-C. Wu, "Tight probabilistic SINR constrained beamforming under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3490–3505, Jul. 2015.

[26] K.-Y. Wang, N. Jacklin, Z. Ding, and C.-Y. Chi, "Robust MISO transmit optimization under outage-based QoS constraints in two-tier heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1883–1897, Apr. 2013.

[27] K.-Y. Wang, A. M. So, T. Chang, W. Ma, and C. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.

[28] K.-L. Hsiung and R. Chen, "Power allocation with outage probability constraints in lognormal fading wireless channels: A relaxation approach," in *Proc. 39th Southeastern Symp. Syst. Theory*, 2007, pp. 173–175.

[29] N. Vučić and H. Boche, "A tractable method for chance-constrained power control in downlink multiuser MISO systems with channel uncertainty," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 346–349, May 2009.

[30] P. K. Shivaswamy, C. Bhattacharyya, and A. J. Smola, "Second order cone programming approaches for handling missing and uncertain data," *J. Mach. Learn. Res.*, vol. 7, no. 7, pp. 1283–1314, 2006.

[31] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*. London, U.K.: Springer, 2013.

[32] N. Vučić, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[33] P. D. Lax, *Linear Algebra and Its Applications*. Hoboken, NJ, USA: Wiley, 2007.

[34] M. A. Goberna and M. A. López (Eds.), *Semi-Infinite Programming*. Dordrecht, The Netherlands: Kluwer, 2001.

[35] V. A. Yakubovich, "S-procedure in nonlinear control theory," *Vestnik Leningrad Univ.*, vol. 4, pp. 73–93, 1977.

[36] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. (Studies in Applied Mathematics), vol. 15, Philadelphia, PA, USA: SIAM, 1994.

[37] A. L. Fradkov and V. A. Yakubovich, "The s-procedure and a duality relations in nonconvex problems of quadratic programming," (in Russian), *Vestnik Leningrad Univ. Math.*, vol. 5, no. 1, pp. 101–109, 1979.

[38] I. Pólik and T. Terlaky, "A survey of the S-lemma," *SIAM Rev.*, vol. 49, no. 3, pp. 371–418, 2007.

[39] M. T. Tan, G. L. Tian, and K. W. Ng, *Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation*. Chapman & Hall/CRC, 2009.

[40] J. L. Jensen, *Saddlepoint Approximations*. Oxford, U.K.: Oxford Univ. Press, 1995.

[41] R. W. Butler, *Saddlepoint Approximations With Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[42] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*. London, U.K.: Springer-Verlag, 2013.

[43] R. W. Butler and M. S. Paolella, "Uniform saddlepoint approximations for ratios of quadratic forms," *Bernoulli*, vol. 14, no. 1, pp. 140–154, 2008.

[44] T. Y. Al-Naffouri, M. Moinuddin, N. Ajeeb, B. Hassibi and A. L. Moustakas, "On the distribution of indefinite quadratic forms in Gaussian random variables," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 153–165, Jan. 2016.

[45] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3702–3712, Aug. 2008.

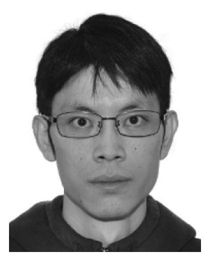
[46] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

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

[48] E. Bjornson and B. Ottersten, "Training-based Bayesian MIMO channel and channel norm estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 2701–2704, 2009.

[49] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[50] S. Borodachov and Y. Wang, "Lattice quantization error for redundant representations," *Appl. Comput. Harmonic Anal.*, vol. 27, no. 3, pp. 334–341, 2009.



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 the Beijing University of Posts and Telecommunications, Beijing, China, in 2010, and the Ph.D. degree in electrical and electronics engineering (EEE) from The University of Hong Kong, Hong Kong, in 2015. From 2015 to 2017, he was a Research Fellow with the Department of EEE, Nanyang Technological University, Singapore. From 2017 to 2018, he worked as a Research Assistant Professor with the College of Innovation and Entrepreneurship, Southern University of Science and Technology, Shenzhen, China. He currently works with Shenzhen University and Hanshan Normal University. His research interests include optimization and statistics for communication beamforming and radar waveform design.

Yik-Chung Wu (Senior Member, IEEE) received the B.Eng. in electrical and electronics engineering (EEE) and M.Phil. degrees from The University of Hong Kong (HKU), in 1998 and 2001, respectively, and the Ph.D. degree from Texas A&M University, College Station, in 2005. From 2005 to 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a member of Technical Staff. Since 2006, he has been with HKU, currently as an Associate Professor. He was a Visiting Scholar with Princeton University, in Summer 2015 and in Summer 2017. His research interests include general areas of signal processing, machine learning, and communication systems. He was an Editor for the IEEE COMMUNICATIONS LETTERS and the IEEE TRANSACTIONS ON COMMUNICATIONS, and is currently an Editor of *Journal of Communications and Networks*.