Quadratically Perturbed Chance Constrained Programming with Fitted Distribution: t-Distribution vs. Gaussian

Xin He

Abstract

For chance-constrained programming (CCP) with non-Gaussian uncertainty, the optimization is generally intractable owing to the complicated probability density function (PDF). Using a simple fitted distribution with Kullback-Leibler (KL) divergence to represent the PDF mismatch is a systematic way to tackle CCP with non-Gaussian uncertainty. However, the essential difficulty of this methodology is to choose the fitted PDF, which should be close to the true PDF and make the resulting CCP problem solvable. In this paper, among the well-known PDFs, t-distribution rather than Gaussian is selected to be the fitted PDF. After the CCP with non-Gaussian uncertainty is transformed into CCP with a fitted PDF, the property of the regularized outage probability is analysed first. Then, the unimodal distributional properties are established for the quadratic form under t or Gaussian perturbation. Finally, based on the unimodal property and the regularized outage analysis, the analytical condition to make the safe approximation with fitted t-distribution have larger feasible set than the safe approximation with fitted Gaussian is obtained. With transceiver design under Logistic and Gaussian mixture channel uncertainties as an example, simulation results validate the less conservativeness property of the CCP with fitted t-distribution, compared to the CCP with fitted Gaussian and the classic moment method.

Index Terms

Chance-constrained programming, strongly unimodal, t-distribution, convex safe approximation.

Xin He is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. Emails: hexin@eee.hku.hk, hx66hx@gmail.com.

I. Introduction

Nowadays, more and more optimization related decision-makings are taking the parameter uncertainty into account. Examples include the classification problem with model uncertainty in machine learning [1], the financial portfolio design under future price uncertainty [2], and the beamforming or transceiver design under wireless channel uncertainty [3], [4]. For those problems, chance constrained programming (CCP) is an important method for decision-making under uncertainty.

Since the probability integration in the chance constraint is generally intractable, even under Gaussian uncertainty, the leading approach is the probability inequality based safe approximation [5]. For instance, Bernstein approximation has been applied to the chance constrained support vector machine with input feature uncertainty [6]. Vysochanskii-Petunin inequality has been proposed for solving chance constrained power control problem in wireless communications [7], [8]. Furthermore, Bernstein-type inequality in [9] has been used to solve the probabilistic beamforming problems [10]–[12]. However, those inequality based safe approximation methods are not guaranteed to provide safe approximations for non-Gaussian uncertainties.

Under non-Gaussian uncertainties, the moment method [13] is a systematic method to handle quadratically perturbed CCP if the moment information is known. The moment method is proposed for the probabilistic mean square error (MSE) constrained transceiver design in [3]. Furthermore, the Chebyshev inequality has been applied to classification with missing data in [14]. However, owing to the fact that the probability density function (PDF) is ignored in the moment method, the obtained solution is conservative. In order to exploit the PDF information, using a simple PDF to fit the complicated PDF with Kullback-Leibler (KL) divergence to represent PDF mismatch is proposed in [2], [13]. However, which PDF should be used for fitting is not discussed in [2], [13]. The KL divergence based CCP has been applied to the robust equalizer design [15]–[17], the linearly perturbed CCP [18], [19], and Gaussian is taken as the fitted PDF. However, Gaussian is a light-tailed distribution and is not flexible to fit non-Gaussian distributions. The fitted PDF should be flexible to match the true PDF and make the resulting CCP problem solvable. The t-distribution is well known for its flexibility to fit other distributions, but the safe approximation of the CCP with quadratical t-perturbation is revealed in this paper.

In this paper, after the CCP with a complicated PDF is transformed into CCP with a fitted simple PDF, the regularized outage probability is analysed first. Next, the unimodal property of the quadratic form under t-distribution or Gaussian perturbation is revealed. Finally, based on the unimodal property and the analysis of the regularized outage probability, the analytical condition to make safe approximation with fitted t-distribution have larger feasible set than that with fitted Gaussian is obtained. With probabilistic MSE constrained transceiver design under non-Gaussian channel uncertainty as an example, simulation results show the less conservativeness property of the CCP with fitted t-distribution, compared with the CCP with fitted Gaussian and the moment method.

The rest of this paper is organized as follows. In Section II, the reformulation of the KL divergence based CCP and the analysis of the regularized outage probability are investigated. In Section III, the invariant unimodal property of the quadratic form under t-distribution and Gaussian perturbation is established. The feasible sets of safe approximation with fitted t-distribution and Gaussian are compared in Section IV, and the application in transceiver design are shown in Section V. Simulation results are presented in Section VI, and conclusions are drawn in Section VII.

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. The notation $\overset{d}{=}$ means the two sides have the same distribution. The direct sum $\bigoplus_{j=1}^{2} \mathbf{A}_j = \mathbf{A}_1 \oplus \mathbf{A}_2$ creates a block diagonal matrix from square matrices $\{\mathbf{A}_j\}$, and $\mathbf{A}^{\dagger}$ donates the Moore-Penrose inverse.

II. Problem Formulation and Analysis

A conventional optimization problem can be described as $\min \{h(\mathbf{w}) | g(\mathbf{w}, \mathbf{x}) \geq \varepsilon\}$, where $\mathbf{w}$ is the decision variable and $\mathbf{x}$ is a parameter. If the parameter $\mathbf{x}$ contains uncertainty, which is described by a PDF $f_0(\mathbf{x})$, the chance-constrained programming (CCP) problem is formulated as

$$\min \{h(\mathbf{w}) | \mathbb{P}_{f_0(\mathbf{x})}\{g(\mathbf{w}, \mathbf{x}) \geq \varepsilon\} \leq p\}, \tag{1}$$

where $\mathbb{P}_{f_0(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} := \int_{g(\mathbf{w},\mathbf{x})\geq\varepsilon} f_0(\mathbf{x})d\mathbf{x}$, ε is the constraint requirement and $p \in (0,1)$ is the outage probability.

Since the integration in $\mathbb{P}_{f_0(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\}$ is generally intractable, the general method is to find an upper bound for $\mathbb{P}_{f_0(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\}$ as a safe approximation. However, since the available safe approximations are majorly focused on some simple PDF. e.g., Gaussian and Uniform, the safe approximation for $\mathbb{P}_{f_0(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\}$ with a complicated PDF $f_0(\mathbf{x})$ is not available. One method to solve this problem is to use a simple PDF $f_1(\mathbf{x})$ to fit the complicated PDF $f_0(\mathbf{x})$ and measure the mismatch between the two PDFs by the Kullback-Leibler (KL) divergence

$$d = D(f_0(\mathbf{x})\|f_1(\mathbf{x})) = \int \ln \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} f_0(\mathbf{x})d\mathbf{x}. \tag{2}$$

Therefore, the complicated PDF $f_0(\mathbf{x})$ lies in the set $\mathcal{F} = \{f(\mathbf{x})|D(f(\mathbf{x})\|f_1(\mathbf{x})) \leq d\}$, and a safe approximation of problem (1) is [13]

$$\begin{aligned} \min_{\mathbf{w}} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \sup_{f(\mathbf{x})\in\mathcal{F}} \mathbb{P}_{f(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} \quad \leq p. \end{aligned} \tag{P0}$$

For the bilevel optimization problem (P0), its lower level problem is formulated as

$$\begin{aligned} \sup_{f(\mathbf{x})} \quad & \mathbb{P}_{f(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} \\ \text{s.t.} \quad & \textstyle\int \ln \frac{f(\mathbf{x})}{f_1(\mathbf{x})} f(\mathbf{x})d\mathbf{x} \leq d \\ & \textstyle\int f(\mathbf{x})d\mathbf{x} = 1. \end{aligned} \tag{3}$$

Note that similar KL divergence restricted optimizations are formulated in many areas, e.g., the general CCP [2], [13], the linearly perturbed CCP [18], [19], robust estimation [15], robust equalization [16] and hypothesis test [17]. With a similar Lagrange multiplier method but a more rigorous approach based on the calculus of variation, as derived in *Appendix A*, the dual problem of (3) is

$$\begin{aligned} \min_{\eta} \quad & \eta d + \eta \ln\left(1 + (e^{1/\eta} - 1) \cdot \mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\}\right) \\ \text{s.t.} \quad & \eta > 0. \end{aligned} \tag{4}$$

Lemma 1. *The strong duality holds between the primal problem* (3) *and the dual problem* (4) *if* $d > 0$.

Proof: See Appendix B. ■

Putting (4) into (P0) with $d > 0$, the strong duality between (3) and (4) makes the original bilevel problem (P0) is equivalently transformed into a single level problem

$$\begin{aligned} \min_{\mathbf{w},\eta} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \eta d + \eta \ln\left(1 + (e^{1/\eta} - 1) \cdot \mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\}\right) \leq p \\ & \eta > 0. \end{aligned} \tag{5}$$

Since $\eta > 0$, the condition $e^{1/\eta} - 1 > 0$ is obtained and (5) is equivalently transformed into

$$\begin{aligned} \min_{\mathbf{w},\eta} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} \leq \tfrac{e^{p/\eta - d} - 1}{e^{1/\eta} - 1} \\ & \eta > 0. \end{aligned} \tag{6}$$

From (6), it is observed that the variables $\mathbf{w}$ and η are separable and the largest feasible set of $\mathbf{w}$ in (6) is obtained when $\frac{e^{p/\eta - d} - 1}{e^{1/\eta} - 1}$ achieves its maximum. Therefore, the problem (6) is equivalent to

$$\begin{aligned} \min_{\mathbf{w},} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} \leq \phi(p,d), \end{aligned} \tag{P1}$$

where $\phi(p,d) := \max\left\{\frac{e^{p/\eta - d} - 1}{e^{1/\eta} - 1} | \eta > 0\right\}$ is a regularized outage probability. Owing to the similar KL divergence restricted optimization in (3), the linearly perturbed CCP problems in [18], [19] also have similar formulation as (P1).

Therefore, a safe approximation of CCP problem (1) with complicated PDF $f_0(\mathbf{x})$ can be obtained from the CCP problem (P1) with the fitted PDF $f_1(\mathbf{x})$ and a regularized outage probability $\phi(p,d)$. Before discussing which PDF $f_1(\mathbf{x})$ should be selected in Section III, the property of the regularized outage probability $\phi(p,d)$ is studied as follows.

Remark 1. When $d = 0$, the PDF $f_1(\mathbf{x}) = f_0(\mathbf{x})$ and the problem (P0) becomes

$$\min\{h(\mathbf{w}) | \mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x}) \geq \varepsilon\} \leq p\}. \tag{7}$$

Therefore, by comparing (P1) and (7), $\lim_{d \to 0^+} \phi(p,d) = p$ is obtained.

A. Finding the Regularized Probability $\phi(p,d)$

Although the regularized outage probability $\phi(p,d)$ needs to be calculated from the nonconvex problem $\max\left\{\frac{e^{p/\eta - d} - 1}{e^{1/\eta} - 1} | \eta > 0\right\}$, its global optimum solution can be obtained by exploring the special structure. In order to facilitate the derivative calculation, $z = e^{1/\eta} - 1 > 0$ is introduced and the optimization problem is reformulated as $\max\left\{\frac{e^{-d}(z+1)^p-1}{z}|z>0\right\}$. The new objective function is introduced as

$$\psi(z)=\frac{e^{-d}(z+1)^p-1}{z}. \tag{8}$$

The first and second derivative of $\psi(z)$ are

$$\psi'(z)=\frac{e^{-d}(z+1)^{p-1}[(p-1)z-1]+1}{z^2}, \tag{9}$$

$$\psi''(z)=\frac{\psi_1(z)}{z^3}, \tag{10}$$

$$\psi_1(z)=e^{-d}(z+1)^{p-2}[(p-1)(p-2)z^2+(4-2p)z+2]-2. \tag{11}$$

With $d>0$ and $p\in(0,1)$, it can be calculated from (10) that $\lim\limits_{z\to 0^+}\psi''(z)=-\infty$ and $\lim\limits_{z\to+\infty}\psi''(z)=0^+$. Furthermore, since $\psi''(z)$ is continuous in $(0,+\infty)$, there must exist some point $z^\diamond>0$ such that $\psi''(z^\diamond)=0=\psi_1(z^\diamond)$. From (11), it can be calculated that

$$\psi_1'(z)=e^{-d}p(p-1)(p-2)z^2(z+1)^{p-3}. \tag{12}$$

Since $p\in(0,1)$, the derivative $\psi_1'(z)>0$ is guaranteed for all $z>0$. Therefore, $\psi_1(z)$ is monotonically increasing in $(0,+\infty)$. Together with the obtained result $\psi_1(z^\diamond)=0$, $\psi_1(z)<0$ in $(0,z^\diamond)$ and $\psi_1(z)>0$ in $(z^\diamond,+\infty)$ are obtained. Putting this result into (10), $\psi''(z)<0$ in $(0,z^\diamond)$ and $\psi''(z)>0$ in $(z^\diamond,+\infty)$ are obtained. Therefore, $\psi(z)$ is concave in $(0,z^\diamond)$ and convex in $(z^\diamond,+\infty)$. This implies $\psi'(z)$ is monotonically decreasing in $(0,z^\diamond)$ and monotonically increasing in $(z^\diamond,+\infty)$.

From (9), it can be derived that $\lim\limits_{z\to 0^+}\psi'(z)=+\infty$ and $\psi'(z)\underset{z\to+\infty}{=}0^-$. The monotonic property of $\psi'(z)$ and its limit conditions reveal that $\psi'(z)$ is monotonically decreasing from $+\infty$ to a negative number in $(0,z^\diamond]$ and monotonically increasing from that negative number to 0^- in $[z^\diamond,+\infty)$. Therefore, there exist a unique point $\bar{z}\in(0,+\infty)$ such that $\psi'(\bar{z})=0$, and $\bar{z}$ is the global maximum point of $\psi(z)$. The result can be summarized as follows.

Theorem 1. *The regularized outage probability $\phi(p,d)=\psi(\bar{z})=\frac{e^{-d}(\bar{z}+1)^p-1}{\bar{z}}$, where $\bar{z}$ is the unique root of $\psi'(\bar{z})=0$ by bisection in $(0,z_2]$ with $\psi'(z_2)<0$.*

B. Analysis of the Regularized Probability $\phi(p,d)$

Since different fitted PDF $f_1(\mathbf{x})$ results in different KL divergence d, the relationship between $\phi(p,d)$ and d provides an important criterion for selecting $f_1(\mathbf{x})$. From *Theorem 1*, the relationship between $\phi(p,d)$ and $\psi(\bar{z})$ is restated as

$$\phi(p,d) = \psi(\bar{z}) = \frac{e^{-d}(\bar{z}+1)^p - 1}{\bar{z}}. \tag{13}$$

Since the optimal solution $\bar{z}$ occurs at $\psi'(\bar{z}) = 0$, i.e.,

$$e^{-d}(\bar{z}+1)^{p-1}[(p-1)\bar{z} - 1] + 1 = 0, \tag{14}$$

the relationship between d and the optimal solution $\bar{z}$ is

$$d = (p-1)\ln(\bar{z}+1) + \ln((1-p)\bar{z}+1). \tag{15}$$

Therefore, the derivative

$$\frac{\partial d}{\partial \bar{z}} = (1-p)(\frac{1}{(1-p)\bar{z}+1} - \frac{1}{\bar{z}+1}) \tag{16}$$

$$= \frac{p(1-p)\bar{z}}{(\bar{z}+1)((1-p)\bar{z}+1)} > 0 \tag{17}$$

reveals the optimal solution $\bar{z}$ is monotonically increasing with d. With those intermediate results in (14)-(17), the relationship between $\phi(p,d)$ and d is obtained as follows.

Theorem 2. *The regularized outage probability $\phi(p,d)$ is a monotonically decreasing convex function of d, and the range of $\phi(p,d)$ is $(0,p]$.*

Proof: From (14), we have

$$e^{-d}(\bar{z}+1)^p = e^{-d}(\bar{z}+1)^{p-1}p\bar{z} + 1. \tag{18}$$

Putting (18) into (13), the regularized outage probability is

$$\phi(p,d) = e^{-d}(\bar{z}+1)^{p-1}p. \tag{19}$$

According to (19) and using the result of (17),

$$\frac{\partial \phi}{\partial d} = -e^{-d}(\bar{z}+1)^{p-1}p + e^{-d}(p-1)(\bar{z}+1)^{p-2}p\frac{\partial \bar{z}}{\partial d} \tag{20}$$

$$= \frac{-e^{-d}(\bar{z}+1)^p}{\bar{z}} < 0, \tag{21}$$

$$\frac{\partial^2 \phi}{\partial d^2} = \frac{e^{-d}(\bar{z}+1)^p}{\bar{z}} - \frac{e^{-d}(\bar{z}+1)^{p-1}(p\bar{z} - \bar{z} - 1)}{\bar{z}^2}\frac{\partial \bar{z}}{\partial d} \tag{22}$$

$$= \frac{e^{-d}(\bar{z}+1)^p}{\bar{z}}(1 + \frac{((1-p)\bar{z}+1)^2}{\bar{z}^2 p(1-p)}) > 0. \tag{23}$$

Therefore, (21) and (23) reveal that $\phi(p,d)$ is a monotonically decreasing function and convex function, respectively. Together the result $\phi(p,0)=p$ from *Remark 1*, $\phi(p,d)\le p$ is obtained. Furthermore, (19) reveals that $\phi(p,d)>0$ for all $p\in(0,1), d\ge 0$ and $\bar{z}>0$. Therefore, the range of $\phi(p,d)$ is $(0,p]$. ■

In order to enlarge the feasible set in the safe approximation (P1), Theorem 2 reveals that the KL divergence d should be small enough. This coincides with the common sense that the fitted PDF $f_1(\mathbf{x})$ should be close to the original PDF $f_0(\mathbf{x})$. However, the conservativeness and tractability of problem (P1) are also determined by the property of $\mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x})\ge\varepsilon\}$, which will be discussed in the following section.

III. Invariant Unimodal Property of the Quadratic Form

The distributional property of $g(\mathbf{w},\mathbf{x})$ is essential to solve the problem (P1). The distributional property of $g(\mathbf{w},\mathbf{x})$ is determined by the formulation of $g(\mathbf{w},\mathbf{x})$ and the property of the fitted PDF $f_1(\mathbf{x})$, which are specified as follows.

Note that most of the beamforming and transceiver design problems involve quadratic functions, which are also widely used in machine learning [1], [20] and control problems [21]. Therefore, in this paper, the constraint function $g(\mathbf{w},\mathbf{x})$ is restricted to be a quadratic form of $\mathbf{x}$, i.e.,

$$g(\mathbf{w},\mathbf{x})=\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} \text{ with } \big(\mathbf{A}(\mathbf{w})\big)^H=\mathbf{A}(\mathbf{w}). \tag{24}$$

Then, the general CCP in (P1) is specified as a quadratically perturbed CCP

$$\begin{aligned}\min_{\mathbf{w}}\quad & h(\mathbf{w})\\ \text{s.t.}\quad & \mathbb{P}_{f_1(\mathbf{x})}\{\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}\ge\varepsilon\}\quad\le\phi(p,d).\end{aligned}\tag{25}$$

Note that any Hermitian matrix $\mathbf{A}(\mathbf{w})$ can be decomposed as $\mathbf{A}(\mathbf{w})=\mathbf{U}\mathrm{Diag}([\lambda_1,\cdots,\lambda_N])\mathbf{U}^H$, where $\mathbf{U}$ is an unitary matrix and $\{\lambda_i\}_{i=1}^N$ are in descending order. Therefore, the quadratic form is represented as

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}=\mathbf{x}^H\cdot\mathbf{U}\mathrm{Diag}([\lambda_1,\cdots,\lambda_N])\overbrace{\mathbf{U}^H\cdot\mathbf{x}}^{\mathbf{y}} \tag{26}$$

$$=\sum_{i=1}^N\lambda_i|y_i|^2 \tag{27}$$

where y_i is the i^{th} element of $\mathbf{y}=\mathbf{U}^H\mathbf{x}$.

Since $\mathbf{U}$ is a function of $\mathbf{w}$ and the decision variable $\mathbf{w}$ is unknown, the distribution family of $\mathbf{y} = \mathbf{U}^H\mathbf{x}$ is also unknown in general. Therefore, it is crucial to find an invariant stochastic representation for (27) regardless of the decision variable $\mathbf{w}$. Since $\mathbf{U}$ is an unitary matrix, an invariant stochastic representation of (27) can be found if the fitted PDF $f_1(\mathbf{x})$ is an elliptically symmetric distribution. The elliptically symmetric distributions include Gaussian, t-distribution, Kotz, etc. [22]. The invariant stochastic representation of (24) with elliptical distribution $f_1(\mathbf{x})$ is discussed as follows.

A. Fitted Distribution $f_1(\mathbf{x})$ is t-distribution or General Gaussian

Definition 1. The complex multivariate Gaussian: In the complex field, let $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{C})$, where $\boldsymbol{\Sigma} = \mathbb{E}(\mathbf{n}\mathbf{n}^H) \succeq 0$ is the covariance matrix and $\mathbf{C} = \mathbb{E}(\mathbf{n}\mathbf{n}^T) \succeq 0$ is the relation matrix [23]. The general complex multivariate Gaussian is

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{n}, \tag{28}$$

where $\boldsymbol{\mu}$ is a $N \times 1$ vector. This is denoted as $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$.

Definition 2. The complex multivariate t-distribution: Let $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{C})$, $s \sim \frac{1}{\nu}\chi^2(\nu)$ is a scaled chi-square distribution with $\nu > 0$, and s is independent with $\mathbf{n}$. The complex multivariate t-distribution with ν degrees of freedom (DoF) is represented as [24]

$$\mathbf{x} = \boldsymbol{\mu} + s^{-1/2}\mathbf{n}, \tag{29}$$

which is denoted as $\mathbf{x} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$.

The major theorem of this paper is presented as follows.

Theorem 3. *With complex multivariate t-distribution $\mathbf{x} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$, the quadratic form in (24) is invariant unimodal under any decision variable $\mathbf{w} \in \{\mathbf{w} \in \mathbb{C}^N | \mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}\}$.*

Proof: With the multivariate t-distribution $\mathbf{x} = \boldsymbol{\mu} + s^{-1/2}\mathbf{n}$, the quadratic form in (24) becomes

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} = \mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n}/s + 2\mathrm{Re}\{\mathbf{n}^H\mathbf{A}(\mathbf{w})\boldsymbol{\mu}\}/s^{1/2} + \boldsymbol{\mu}^H\mathbf{A}(\mathbf{w})\boldsymbol{\mu}. \tag{30}$$

In the right hand side of (30), the first and second term are correlated, which hinders further statistical analysis. Under the condition $\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$, the quadratic form (30) is simplified into

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} \tag{31}$$

$$= s^{-1}\mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n} \tag{32}$$

$$= s^{-1}\begin{bmatrix}\mathrm{Re}(\mathbf{n})\\ \mathrm{Im}(\mathbf{n})\end{bmatrix}^T \begin{bmatrix}\mathrm{Re}(\mathbf{A}(\mathbf{w})) & -\mathrm{Im}(\mathbf{A}(\mathbf{w}))\\ \mathrm{Im}(\mathbf{A}(\mathbf{w})) & \mathrm{Re}(\mathbf{A}(\mathbf{w}))\end{bmatrix} \overbrace{\begin{bmatrix}\mathrm{Re}(\mathbf{n})\\ \mathrm{Im}(\mathbf{n})\end{bmatrix}}^{\mathbf{n}_e} \tag{33}$$

Since the covariance matrix of $\mathbf{n}_e$ in (33) is $\boldsymbol{\Sigma}_e = \frac{1}{2}\begin{pmatrix}\mathrm{Re}(\boldsymbol{\Sigma}+\mathbf{C}) & \mathrm{Im}(-\boldsymbol{\Sigma}+\mathbf{C})\\ \mathrm{Im}(\boldsymbol{\Sigma}+\mathbf{C}) & \mathrm{Re}(\boldsymbol{\Sigma}-\mathbf{C})\end{pmatrix}$ [23], the normalized real Gaussian can be denoted as $\dot{\mathbf{n}}_e := \boldsymbol{\Sigma}_e^{-1/2}\mathbf{n}_e \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, by using the singular value decomposition (SVD), (33) is reformulated as

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}$$

$$= s^{-1}\dot{\mathbf{n}}_e^T \underbrace{(\boldsymbol{\Sigma}_e^{1/2})^T \begin{bmatrix}\mathrm{Re}(\mathbf{A}(\mathbf{w})) & -\mathrm{Im}(\mathbf{A}(\mathbf{w}))\\ \mathrm{Im}(\mathbf{A}(\mathbf{w})) & \mathrm{Re}(\mathbf{A}(\mathbf{w}))\end{bmatrix} \boldsymbol{\Sigma}_e^{1/2}}_{\mathbf{U}_t\mathrm{Diag}([\lambda_1^t,\cdots,\lambda_{2N}^t])\mathbf{U}_t^T} \dot{\mathbf{n}}_e \tag{34}$$

$$= s^{-1}\dot{\mathbf{n}}_e^H \cdot \mathbf{U}_t\mathrm{Diag}([\lambda_1^t,\cdots,\lambda_{2N}^t]) \overbrace{\mathbf{U}_t^T \cdot \dot{\mathbf{n}}_e}^{\mathbf{y}^t} \tag{35}$$

$$= s^{-1}\sum_{i=1}^{2N}\lambda_i^t|y_i^t|^2. \tag{36}$$

Since $\dot{\mathbf{n}}_e \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is spherically symmetric distributed, the orthogonal matrix $\mathbf{U}_t$ makes $\mathbf{y}^t = \mathbf{U}_t^T \cdot \dot{\mathbf{n}}_e \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Therefore, the stochastic representation of the quadratic form (36) is

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} \overset{d}{=} s^{-1}\sum_{i=1}^{2N}\lambda_i^t\chi_i^2(1), \tag{37}$$

which is a weighted sums of independent chi-squares with one DoF divided by a scaled chi-square distribution. Since $\chi^2(1) \overset{d}{=} \mathrm{Gamma}(1/2, 2)$ and any Gamma random variable is self-decomposable [25, p. 1367], $\chi^2(1)$ is self-decomposable. According to the definition of self-decomposable [25, p. 1358], $\chi^2(1) \overset{d}{=} c\chi^2(1) + \epsilon$, where $c \in (0, 1)$ and ϵ is independent with $\chi^2(1)$. Therefore, for any real number λ_i^t, $\lambda_i^t\chi^2(1) \overset{d}{=} c(\lambda_i^t\chi^2(1)) + \lambda_i^t\epsilon$ is obtained. This result implies $\lambda_i^t\chi^2(1)$ is self-decomposable. Since the self-decomposable distribution is closed under convolution [25, p. 1358], $\sum_{i=1}^{2N}\lambda_i^t\chi_i^2(1)$ is self-decomposable. The self-decomposable property of $\sum_{i=1}^{2N}\lambda_i^t\chi_i^2(1)$ implies that $\sum_{i=1}^{2N}\lambda_i^t\chi_i^2(1)$ is unimodal [26].

Note that $s^{-1}\sim\frac{\nu}{\chi^2(\nu)}$ and $\frac{1}{\chi^2(\nu)}$ is an inverted-chi-squared distribution with ν degrees of freedom, the PDF of $\frac{1}{\chi^2(\nu)}$ is [27, pp. 119]

$$f_s(x)=\frac{2^{-\nu/2}}{\Gamma(\nu/2)}x^{-\nu/2-1}e^{-1/(2x)}. \tag{38}$$

Therefore, $\log(f_s(e^u))=\log(\frac{2^{-\nu/2}}{\Gamma(\nu/2)})-u(\nu/2+1)-e^{-u}/2$ is obtained. Furthermore, the derivation $(\log(f_s(e^u)))''=-e^{-u}/2<0$ reveals that the inverted-chi-squared distribution $\frac{1}{\chi^2(\nu)}$ is multiplicative strongly unimodal (MSU) [28, pp. 210]. Since the degenerate ν is MSU and the MSU property is closed under multiplicative convolution [28, pp. 208], $s^{-1}\sim\nu\cdot\frac{1}{\chi^2(\nu)}$ is MSU.

Since s is independent with $\mathbf{n}$ in (29), $h_1(s)$ and $h_2(\mathbf{n})$ are also independent for any continuous function $h_1(\cdot)$ and $h_2(\cdot)$. Therefore, s^{-1} is independent with $\sum_{i=1}^{2N}\lambda_i^t|y_i^t|^2$. Note that the unimodal property of $\sum_{i=1}^{2N}\lambda_i^t|y_i^t|^2$ is proved, the MSU property of s^{-1} ensures $s^{-1}\sum_{i=1}^{2N}\lambda_i^t|y_i^t|^2$ in (37) is unimodal [28, pp. 207]. ■

Corollary 1. With $\mathbf{x}\sim\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{C})$, the quadratic form in (24) is unimodal under any decision variable $\{\mathbf{w}\in\mathbb{C}^N|\mathbf{A}(\mathbf{w})\boldsymbol{\mu}=\mathbf{0}\}$.

Proof: See Appendix C ■

When the decision variable $\mathbf{w}$ and the random variable $\mathbf{x}$ are changed from the complex field to the real field, $\mathbf{C}=\boldsymbol{\Sigma}$ and the quadratic form in (24) become

$$g(\mathbf{w},\mathbf{x})=\mathbf{x}^T\mathbf{A}(\mathbf{w})\mathbf{x}\text{ with }\big(\mathbf{A}(\mathbf{w})\big)^T=\mathbf{A}(\mathbf{w}). \tag{39}$$

The real Gaussian and t-distribution are denoted as $\mathbf{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\mathbf{x}\sim t_\nu(\boldsymbol{\mu},\boldsymbol{\Sigma})$, respectively.

Corollary 2. In the real field, $\mathbf{x}\sim t_\nu(\boldsymbol{\mu},\boldsymbol{\Sigma})$ or $\mathbf{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, the real quadratic form in (39) is unimodal under any decision variable $\{\mathbf{w}\in\mathbb{R}^N|\mathbf{A}(\mathbf{w})\boldsymbol{\mu}=\mathbf{0}\}$.

Proof: See Appendix D ■

B. Fitted Distribution $f_1(\mathbf{x})$ is $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{C})$ with $\mathbf{C}=\mathbf{0}$

When the relation matrix $\mathbf{C}=\mathbf{0}$, the Gaussian random variable $\mathbf{n}\sim\mathcal{N}_{\mathbb{C}}(\mathbf{0},\boldsymbol{\Sigma},\mathbf{C})$ in (28) is known as circularly-symmetric complex Gaussian (CSCG) and denoted as $\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma})$ [29].

Therefore, the normalized CSCG is $\ddot{\mathbf{n}} := (\mathbf{\Sigma}/2)^{-1/2}\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, 2\mathbf{I})$, where the real and imaginary part of $\ddot{\mathbf{n}}$ are independently and identically distributed as $\mathcal{N}(0,1)$. Denote the normalized Gaussian as $\ddot{\mathbf{x}} = (\mathbf{\Sigma}/2)^{-\frac{1}{2}}\mathbf{x} \sim \mathcal{CN}((\mathbf{\Sigma}/2)^{-\frac{1}{2}}\boldsymbol{\mu}, 2\mathbf{I}_N)$, and decompose the Hermitian matrix $((\mathbf{\Sigma}/2)^{\frac{1}{2}})^H\mathbf{A}(\mathbf{w})(\mathbf{\Sigma}/2)^{\frac{1}{2}} = \mathbf{U}_g\mathrm{Diag}([\lambda_1^g, \cdots, \lambda_N^g])\mathbf{U}_g^H$, the quadratic Gaussian form is reformulated as

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} = \ddot{\mathbf{x}}^H \cdot ((\mathbf{\Sigma}/2)^{\frac{1}{2}})^H\mathbf{A}(\mathbf{w})(\mathbf{\Sigma}/2)^{\frac{1}{2}} \cdot \ddot{\mathbf{x}} \tag{40}$$

$$= \ddot{\mathbf{x}}^H \cdot \mathbf{U}_g\mathrm{Diag}([\lambda_1^g, \cdots, \lambda_N^g])\overbrace{\mathbf{U}_g^H \cdot \ddot{\mathbf{x}}}^{\mathbf{y}^g} \tag{41}$$

$$= \sum_{i=1}^N \lambda_i^g |y_i^g|^2, \tag{42}$$

where y_i^g is the i^{th} element of $\mathbf{y}^g = \mathbf{U}_g^H\ddot{\mathbf{x}}$. Since the normalized Gaussian $\ddot{\mathbf{x}} \sim \mathcal{CN}((\mathbf{\Sigma}/2)^{-\frac{1}{2}}\boldsymbol{\mu}, 2\mathbf{I}_N)$, the unitary matrix $\mathbf{U}_g$ makes $\mathbf{y}^g \sim \mathcal{CN}(\mathbf{U}_g^H(\mathbf{\Sigma}/2)^{-\frac{1}{2}}\boldsymbol{\mu}, 2\mathbf{I}_N)$. Therefore, the real and imaginary part of $\mathbf{y}^g$ are independent unit-variance Gaussian variables. Subsequently, the stochastic representation of (42) becomes

$$\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} \overset{d}{=} \sum_{i=1}^N \lambda_i^g \chi_i^2(2, |\eta_i|^2), \tag{43}$$

which is a weighted sums of independent noncentral chi-square distribution with DoF two, and η_i is the i^{th} element of $\mathbf{U}_g^H(\mathbf{\Sigma}/2)^{-\frac{1}{2}}\boldsymbol{\mu}$.

Although the eigenvalue λ_i^g and the noncentrality parameter η_i are all coupled with the decision variable $\mathbf{w}$, the invariant distributional property of (43) is stated as follows.

Theorem 4. *With $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{C})$ and the relation matrix $\mathbf{C} = \mathbf{0}$, the quadratic form in* (24) *is invariant unimodal under any decision variable $\mathbf{w} \in \mathbb{C}^N$.*

Proof: It is known that the density function of $\chi_i^2(2, |\eta_i|^2)$ is logconcave [30], i.e.,

$$\big(\log F'(x)\big)'' = \frac{F'''(x)F'(x) - (F''(x))^2}{(F'(x))^2} \leq 0, \quad \forall x > 0, \tag{44}$$

where $F(x)$ denotes the distribution function of $\chi_i^2(2, |\eta_i|^2)$. The logconcave condition implies $\chi_i^2(2, |\eta_i|^2)$ is strongly unimodal [31, p. 20].

If $\lambda_i^g > 0$, let the distribution function of $\lambda_i^g\chi_i^2(2, |\eta_i|^2)$ be denoted as $F_p(x) = \Pr\{\lambda_i^g\chi_i^2(2, |\eta_i|^2) \leq$

$x\} = F(x/\lambda_i^g)$ with domain $x \in (0, +\infty)$. Therefore, we have

$$\big(\log F_p'(x)\big)'' = \big(\log(F'(x/\lambda_i^g)/\lambda_i^g)\big)'' \tag{45}$$

$$= \frac{F'''(x/\lambda_i^g)F'(x/\lambda_i^g) - (F''(x/\lambda_i^g))^2}{(\lambda_i^g F'(x/\lambda_i^g))^2} \leq 0, \quad \forall x > 0, \lambda_i^g > 0, \tag{46}$$

which reveals that the density function $F_p'(x)$ is also logconcave, and consequently $\lambda_i^g \chi_i^2(2, |\eta_i|^2)$ with $\lambda_i^g > 0$ is strongly unimodal.

If $\lambda_i^g = 0$, the random variable $\lambda_i^g \chi_i^2(2, |\eta_i|^2)$ degenerates into a constant, which is also strongly unimodal [31, p. 17].

If $\lambda_i^g < 0$, let the distribution function of $\lambda_i^g \chi_i^2(2, |\eta_i|^2)$ be denoted as $F_n(x) = \Pr\{\lambda_i^g \chi_i^2(2, |\eta_i|^2) \leq x\} = 1 - F(x/\lambda_i^g)$ with domain $x \in (-\infty, 0)$. Therefore, we have

$$\big(\log F_n'(x)\big)'' = \big(\log(-F'(x/\lambda_i^g)/\lambda_i^g)\big)'' \tag{47}$$

$$= \frac{F'''(x/\lambda_i^g)F'(x/\lambda_i^g) - (F''(x/\lambda_i^g))^2}{(\lambda_i^g F'(x/\lambda_i^g))^2} \leq 0, \quad \forall x < 0, \lambda_i^g < 0, \tag{48}$$

which reveals that $\lambda_i^g \chi_i^2(2, |\eta_i|^2)$ with $\lambda_i^g < 0$ is strongly unimodal.

Since all N independent random variables in $\{\lambda_i^g \chi_i^2(2, |\eta_i|^2)\}_{i=1}^N$ are strongly unimodal and the strongly unimodal distribution is closed under convolution [31, p. 17], the quadratic form in (43) is strongly unimodal regardless of the decision variable $\mathbf{w}$. Finally, strongly unimodality implies unimodality.

■

Remark 2. Except for the t-distribution and Gaussian, the essential difficulty to obtain the unimodal property for other elliptical distributions is the square nonlinear transformation. Taking the Kotz distribution as an example, since there is no available stochastic representation for the squared Kotz, it is difficult to analyse the distributional property of the quadratic form.

IV. Safe Approximations

After obtaining the unimodal distributional properties of the quadratic form, safe approximations of the quadratically perturbed CCP

$$\begin{aligned} \min_{\mathbf{w}} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \mathbb{P}_{f_1(\mathbf{x})}\{\mathbf{x}^H \mathbf{A}(\mathbf{w})\mathbf{x} \geq \varepsilon\} \quad \leq \phi(p, d) \end{aligned} \tag{49}$$

with $\varepsilon > 0$ can be obtained. Under the fitted t-distribution and Gaussian, the feasible set comparison and the convex formulation are discussed as follows.

A. Density Fitting

The original uncertainty $\mathbf{x}$ is described by a density function $f_0(\mathbf{x})$ or observed samples, its mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and relation matrix $\mathbf{C}$ are assumed to be known.

If the fitted PDF $f_1(\mathbf{x})$ is selected as Gaussian $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{C}_g)$, its optimal parameter is obtained by minimizing the KL divergence

$$\min_{\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{C}_g} \quad \int \ln \frac{f_0(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})}{f_1(\mathbf{x}|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{C}_g)} f_0(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C}) d\mathbf{x}. \tag{50}$$

The optimal solution of (50) is the classic moment matching result, i.e., $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{C}_g) = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$.

On the other hand, if the fitted PDF $f_1(\mathbf{x})$ is selected as t-distribution $t_\nu(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \mathbf{C}_t)$, the closed-form solution is not available. However, it is also natural to use moment matching method to estimate the parameters in t-distribution. It is known that the first and second moments of $t_\nu(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \mathbf{C}_t)$ are $\boldsymbol{\mu}_t$ and $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}_t$, respectively [32]. Therefore, by matching the first and second moments of t-distribution to that of $f_0(\mathbf{x})$, the parameter of the fitted t-distribution is $(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \mathbf{C}_t) = (\boldsymbol{\mu}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}, \frac{\nu-2}{\nu}\mathbf{C})$ with $\nu > 2$.

B. Feasible Set Comparison

If $\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}$ is an unimodal random variable, the Vysochanskii-Petunin (V-P) inequality [31, p. 29] provides an upper bound for the outage

$$\Pr\{\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x} \geq \varepsilon\} \leq \Pr\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}| \geq \varepsilon\}, \text{with } \varepsilon > 0 \tag{51}$$

$$\leq \max\{\frac{r^*\mathbb{E}_{\mathbf{x}}\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}|^r\} - \varepsilon^r}{(r^*-1)\varepsilon^r}, \frac{\pi_r\mathbb{E}_{\mathbf{x}}\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}|^r\}}{\varepsilon^r}\}, \tag{52}$$

where $\pi_r = (r/(r+1))^r$, $r > 0$ and r^* is the root of the equation $r^*(r^* - r - 1)^r = r^r$ with $r^* > (r+1)$. Replacing the outage in (49) with the upper bound in (52), a unified safe approximation for (49) is

$$\begin{aligned} \min_{\mathbf{w}} \quad & h(\mathbf{w}) \\ \text{s.t.} \quad & \mathbb{E}_{\mathbf{x}}\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}|^r\} \leq R_r(\phi(p,d)) \cdot \varepsilon^r, \end{aligned} \tag{53}$$

where

$$R_r(\phi(p,d)) = \begin{cases} (1+\frac{1}{r})^r\phi(p,d) & \text{if } \phi(p,d) \in (0, 1/\{r^*[(1+1/r)^r - 1] + 1\}] \\ \big((r^*-1)\phi(p,d) + 1\big)/r^* & \text{if } \phi(p,d) \in [1/(r^*[(1+1/r)^r - 1] + 1), 1) \end{cases}. \tag{54}$$

When the fitted $f_1(\mathbf{x})$ is $t_\nu(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}, \frac{\nu-2}{\nu}\mathbf{C})$, the condition $\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$ guarantees the unimodal property of $\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}$ in *Theorem 3*. Therefore, the V-P inequality enables a feasible subset of (49) become

$$\mathcal{W}_t := \{\mathbf{w}|\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}, \mathbb{E}_{\mathbf{x}}\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}|^r\} \leq R_r(\phi(p, d_t))\varepsilon^r\}, \tag{55}$$

where d_t is the KL divergence between $f_0(\mathbf{x})$ and the fitted t-distribution. Since $\mathbf{x} \sim t_\nu(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}, \frac{\nu-2}{\nu}\mathbf{C})$, the random variable can be represented as $\mathbf{x} = \boldsymbol{\mu} + s^{-1/2} \cdot \sqrt{\frac{\nu-2}{\nu}}\mathbf{n}$ with $s \sim \frac{1}{\nu}\chi^2(\nu)$ and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{C})$. Since $\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{n}$ is independent with s, the expectation in (55) is

$$\mathbb{E}\{|\mathbf{x}^H\mathbf{A}(\mathbf{w})\mathbf{x}|^r\} = \mathbb{E}\{(\frac{\nu}{\chi^2(\nu)})^r \cdot |\frac{\nu-2}{\nu}\mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n}|^r\} \tag{56}$$

$$= \mathbb{E}\{(\frac{\nu}{\chi^2(\nu)})^r\} \cdot \mathbb{E}\{|\frac{\nu-2}{\nu}\mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n}|^r\}. \tag{57}$$

Note that the first term in (57) with $\nu > 2r$ is

$$\mathbb{E}\{(\frac{\nu}{\chi^2(\nu)})^r\} = \int_0^\infty (\nu x)^r \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{-\nu/2-1} e^{-1/(2x)} dx \tag{58}$$

$$= (\nu)^r \frac{2^{-\nu/2}}{\Gamma(\nu/2)} \cdot \frac{\Gamma(\nu/2 - r)}{(1/2)^{\nu/2-r}}, \tag{59}$$

$$= \prod_{j=1}^r \frac{\nu}{\nu - 2j}. \tag{60}$$

Therefore, putting (57) and (60) into (55), the set $\mathcal{W}_t$ is reformulated as

$$\mathcal{W}_t := \{\mathbf{w}|\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}, \mathbb{E}_{\mathbf{n}}\{|\mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n}|^r\} \leq \prod_{j=1}^r \frac{\nu-2j}{\nu-2} R_r(\phi(p, d_t))\varepsilon^r\}. \tag{61}$$

When the fitted $f_1(\mathbf{x})$ is $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$, the random variable can be represented as $\mathbf{x} = \boldsymbol{\mu} + \mathbf{n}$ and the feasible subset of (49) is

$$\mathcal{W}_g := \{\mathbf{w}|\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}, \mathbb{E}_{\mathbf{n}}\{|\mathbf{n}^H\mathbf{A}(\mathbf{w})\mathbf{n}|^r\} \leq R_r(\phi(p, d_g))\varepsilon^r\}, \tag{62}$$

where d_g is the KL divergence between $f_0(\mathbf{x})$ and the fitted Gaussian. Next, the feasible sets of the safe approximations with fitted t-distribution and fitted Gaussian are compared as follows.

Theorem 5. *If* $R_r(\phi(p, d_g)) < \prod_{j=1}^r \frac{\nu-2j}{\nu-2} R_r(\phi(p, d_t))$*, then* $\mathcal{W}_g \subset \mathcal{W}_t$*. If* $\nu \to \infty$*, then* $\mathcal{W}_g = \mathcal{W}_t$*.*

Proof: First, since $\varepsilon^r > 0$, the condition $\prod_{j=1}^r \frac{\nu-2j}{\nu-2} R_r(\phi(p, d_t)) > R_r(\phi(p, d_g))$ leads to $\prod_{j=1}^r \frac{\nu-2j}{\nu-2} R_r(\phi(p, d_t))\varepsilon^r > R_r(\phi(p, d_g))\varepsilon^r$. Together with the definition in (61) and (62), the conclusion $\mathcal{W}_g \subset \mathcal{W}_t$ is obtained. Second, with fixed r, $\lim_{\nu\to\infty} \prod_{j=1}^r \frac{\nu-2j}{\nu-2} = 1$. Furthermore, since $\lim_{\nu\to\infty} t_\nu(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}, \frac{\nu-2}{\nu}\mathbf{C}) \overset{d}{=} \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$, the condition $\nu \to \infty$ leads to $d_t = d_g$. Therefore, $\lim_{\nu\to\infty} \prod_{j=1}^{r} \frac{\nu-2j}{\nu-2} R_r(\phi(p, d_t)) = R_r(\phi(p, d_g))$ leads to the conclusion $\mathcal{W}_t = \mathcal{W}_g$. ■

Corollary 3. When $r = 1$, $d_g > d_t$ implies $\mathcal{W}_g \subset \mathcal{W}_t$, and a smaller d_t implies a larger $\mathcal{W}_t$.

Proof: When $r = 1$, according to *Theorem 5*, $\mathcal{W}_g \subset \mathcal{W}_t$ if $R_1(\phi(p, d_g)) < R_1(\phi(p, d_t))$. According to *Theorem 2*, $\phi(p, d)$ is a monotonically decreasing function of d. Furthermore, the function $R_1(\cdot)$ in (54) is a monotonically increasing function. Therefore, the composition function $R_1(\phi(p, d))$ is a monotonically decreasing function of d, which makes the condition $R_1(\phi(p, d_t)) > R_1(\phi(p, d_g))$ can be further simplified into $d_t < d_g$. ■

Theorem 5 reveals the condition to make safe approximation with fitted t-distribution and a appropriately selected DoF to have larger feasible subset than that with fitted Gaussian. Furthermore, the same mathematical formulation in (61) and (62) reveal the computation method and complexity of the two cases will be the same. Efficient computation method for safe approximation with feasible set (61) and (62) is discussed as follows.

Remark 3. Except for the moment matching method in Section IV-A, the fitted t-distribution can also be parameterized as $t_v(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, c\mathbf{C})$ with $c > 0$. With similar derivations from (55) to (61), the optimal parameters come from $\max\{\prod_{j=1}^{r} \frac{\nu-2j}{\nu c} R_r(\phi(p, d_t)) | c > 0, \nu > 2r\}$. This problem is intractable, owing to the reason that d_t can not be expressed as an analytic form of c and ν. Furthermore, the two parameters selection will be complicated. The parameter selection problem will be much more difficult for the general parametriation $t_v(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \mathbf{C}_t)$.

C. Convex Formulation for Quadratic Function $\mathbf{A}(\mathbf{w})$

In order to efficiently solve the safe approximation with feasible set (61) and (62), the matrix function $\mathbf{A}(\mathbf{w})$ must have a specific form and the widely used quadratic form is illustrated in this paper.

For the positive semidefinite quadratic form $\mathbf{A}(\mathbf{w}) = \mathbf{L}^H(\mathbf{w})\mathbf{L}(\mathbf{w})$, where $\mathbf{L}(\mathbf{w})$ is an affine function of $\mathbf{w}$. Note that the condition $\mathbf{L}^H(\mathbf{w})\mathbf{L}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$ is equivalent to $\mathbf{L}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$. When $r = 1$, the feasible subset $\mathcal{W}_t$ in (61) can be represented as

$$\mathcal{W}_t = \{\mathbf{w} | \mathbf{L}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}, \mathrm{Tr}\left(\mathbf{L}^H(\mathbf{w})\mathbf{L}(\mathbf{w})\boldsymbol{\Sigma}\right) \le R_1(\phi(p, d_t))\varepsilon\}. \tag{63}$$

The the linear constraint and the convex quadratic constraint reveal that $\mathcal{W}_t$ in (63) is a convex set. When $\nu \to \infty$, the set $\mathcal{W}_t$ is degenerated to $\mathcal{W}_g$. When $r \geq 2$, the high order nonlinear form in $\mathbb{E}_{\mathbf{n}}\{|\mathbf{n}^H \mathbf{L}^H(\mathbf{w})\mathbf{L}(\mathbf{w})\mathbf{n}|^r\}$ is a nonconvex function of $\mathbf{w}$ in general. On the other hand, the indefinite quadratic form of $\mathbf{A}(\mathbf{w})$ makes $\mathbb{E}_{\mathbf{n}}\{|\mathbf{n}^H \mathbf{A}(\mathbf{w})\mathbf{n}|^r\}$ become a nonconvex function of $\mathbf{w}$.

V. Application in Wireless Communication System

In order to illustrate the proposed CCP with fitted distributions, the probabilistic MSE constrained multiuser transceiver design under non-Gaussian channel uncertainty is taken as an example. The downlink multiuser system under consideration 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. L_k independent data streams are transmitted to the k^{th} user and $\sum_{k=1}^{K} L_k = L$. 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 k^{th} user, respectively. In order to successfully detect the data, the condition $L \leq N$ and $L_k \leq M_k$ are needed. The received $M_k \times 1$ interference vector at the k^{th} user has zero mean and covariance matrix $\mathbf{R}_k \succeq 0$. The beamforming design aiming at minimizing transmit power at BS with guaranteed probabilistic MSE requirements is formulated as [3], [33]

$$\begin{aligned} &\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^{K}} \|\mathbf{G}\|_F \\ &\text{s.t.} \quad \mathbb{P}_{f_0(\mathbf{H}_k)}\{\underbrace{\|\mathbf{F}_k\mathbf{H}_k\mathbf{G}-\mathbf{D}_k\|_F^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H)}_{\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k)} \geq \varepsilon_k\} \leq p_k, \forall k, \end{aligned} \tag{64}$$

where $\mathbf{D}_k = [\mathbf{0}_{L_k \times \sum_{k=1}^{k-1} L_k} \ \ \mathbf{I}_{L_k} \ \ \mathbf{0}_{L_k \times \sum_{k=k+1}^{K} L_k}]$, ε_k and p_k are the MSE requirement and the outage probability of the k^{th} user, respectively. With an estimated channel $\hat{\mathbf{H}}_k$, the channel is modelled as $\mathbf{H}_k = \hat{\mathbf{H}}_k + \boldsymbol{\Delta}_k$, where the channel uncertainty $\mathrm{vec}(\boldsymbol{\Delta}_k)$ has zero mean, covariance matrix $\boldsymbol{\Sigma}_k$ and relation matrix $\mathbf{C}_k$. The PDF of the channel is represented as $f_0(\mathbf{H}_k)$.

By matching the first and second moments of $f_1(\mathbf{H}_k)$ to that of $f_0(\mathbf{H}_k)$, the fitted t-distribution is $t_\nu(\mathrm{vec}(\hat{\mathbf{H}}_k), \frac{\nu-2}{\nu}\boldsymbol{\Sigma}_k, \frac{\nu-2}{\nu}\mathbf{C}_k)$ with $\nu > 2$ and the fitted Gaussian is $\mathcal{N}_{\mathbb{C}}(\mathrm{vec}(\hat{\mathbf{H}}_k), \boldsymbol{\Sigma}_k, \mathbf{C}_k)$. With fitted distribution $f_1(\mathbf{H}_k)$ and the KL divergence $d_k = D(f_0(\mathbf{H}_k)\|f_1(\mathbf{H}_k))$, a safe approximation of (64) is

$$\begin{aligned} &\min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^{K}} \|\mathbf{G}\|_F \\ &\text{s.t.} \quad \mathbb{P}_{f_1(\mathbf{H}_k)}\{\mathrm{MSE}_k(\mathbf{G},\mathbf{F}_k,\mathbf{H}_k) \geq \varepsilon_k\} \leq \phi(p_k, d_k), \quad \forall k. \end{aligned} \tag{65}$$

Corollary 4. If the fitted distribution is general complex Gaussian or t-distribution, the k^{th} user's MSE is unimodal under the transceiver pair satisfying $\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} = \mathbf{D}_k$. If the fitted distribution is complex Gaussian with zero relation matrix, the k^{th} user's MSE is invariant unimodal under any transceiver pair.

Proof: See *Appendix E.* ∎

If the fitted channel uncertainty distribution is selected as t-distribution, the condition $\{\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} = \mathbf{D}_k\}_{k=1}^K$ guarantees the unimodal property of the MSE. Therefore, by using the result of (61) with $r = 1$, a safe approximation of (65) is

$$\begin{aligned} \min_{\mathbf{G},\{\mathbf{F}_k\}_{k=1}^K} \quad & \|\mathbf{G}\|_F \\ \text{s.t.} \quad & \|(\mathbf{G}^T \otimes \mathbf{F}_k)\mathbf{\Sigma}_k^{1/2}\|_2^2 \leq R_1(\phi(p_k, d_k^t))\big(\varepsilon_k - \|\mathbf{F}_k\mathbf{R}_k^{1/2}\|_F^2\big), \quad \forall k \\ & \mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} = \mathbf{D}_k, \quad \forall k, \end{aligned} \tag{T0}$$

where $d_k^t = D(f_0(\mathbf{H}_k)\|t_\nu(\text{vec}(\hat{\mathbf{H}}_k), \frac{\nu-2}{\nu}\mathbf{\Sigma}_k, \frac{\nu-2}{\nu}\mathbf{C}_k))$. In order to enlarge the feasible set of (T0), as shown in *Corollary 3*, the DoF ν should be selected to minimize the divergence d_k^t. With a classic reformulation technique in [33], let $\mathbf{F}_k = \tilde{\mathbf{F}}_k/a_k$ with fixed beamforming direction $\tilde{\mathbf{F}}_k$ and $a_k > 0$, the problem (T0) becomes a convex problem

$$\begin{aligned} \min_{\mathbf{G},\{a_k\}_{k=1}^K} \quad & \|\mathbf{G}\|_F \\ \text{s.t.} \quad & \Big\|\Big[\text{vec}\Big((\mathbf{G}^T \otimes \tilde{\mathbf{F}}_k)\mathbf{\Sigma}_k^{\frac{1}{2}}\Big)^T \quad \sqrt{R_1(\phi(p_k, d_k^t))}\text{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})^T\Big]\Big\|_2 \leq a_k\sqrt{R_1(\phi(p_k, d_k^t))\varepsilon_k}, \quad \forall k, \\ & \tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} = a_k\mathbf{D}_k, \quad \forall k. \end{aligned} \tag{T1}$$

If the fitted channel uncertainty distribution is selected as CSCG, no restriction is needed for the unimodality and a convex safe approximation of (65) is

$$\begin{aligned} \min_{\mathbf{G},\{a_k\}_{k=1}^K} \quad & \|\mathbf{G}\|_F \\ \text{s.t.} \quad & \Big\|\Big[\text{vec}\Big((\mathbf{G}^T \otimes \tilde{\mathbf{F}}_k)\mathbf{\Sigma}_k^{\frac{1}{2}}\Big)^T \quad \sqrt{R_1(\phi(p_k, d_k^g))}\text{vec}(\tilde{\mathbf{F}}_k\mathbf{R}_k^{\frac{1}{2}})^T \\ & \quad \text{vec}(\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} - a_k\mathbf{D}_k)^T\Big]\Big\|_2 \leq a_k\sqrt{R_1(\phi(p_k, d_k^g))\varepsilon_k}, \quad \forall k, \end{aligned} \tag{G1}$$

where $d_k^g = D(f_0(\mathbf{H}_k)\|\mathcal{CN}(\text{vec}(\hat{\mathbf{H}}_k), \mathbf{\Sigma}_k))$. By comparing the formulation between the general complex Gaussian (t_∞) in (T1) and the special CSCG in (G1), the difference is whether the beamforming direction is fixed in $\{\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} = a_k\mathbf{D}_k\}_{k=1}^K$ or not.

Remark 4. If the linear constraints $\{\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} = a_k\mathbf{D}_k\}_{k=1}^K$ are feasible, the condition $\{\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} = a_k\mathbf{D}_k\}_{k=1}^K$ are equivalent to

$$\mathbf{G} := \mathbf{V}\mathbf{\Lambda}^{\dagger}\mathbf{U}^H \bigoplus_{k=1}^{K} a_k\mathbf{I}_{L_k}, \tag{66}$$

where the matrices $\mathbf{V}$, $\mathbf{\Lambda}$ and $\mathbf{U}$ come from the SVD $[(\tilde{\mathbf{F}}_1\hat{\mathbf{H}}_1)^T, \cdots, (\tilde{\mathbf{F}}_K\hat{\mathbf{H}}_K)^T]^T = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$. On the other hand, if any linear constraint in $\{\tilde{\mathbf{F}}_k\hat{\mathbf{H}}_k\mathbf{G} = a_k\mathbf{D}_k\}_{k=1}^K$ is infeasible, (T1) is infeasible. Therefore, putting $\mathbf{G}$ of (66) into (T1), the equality constraints in (T1) are eliminated and (T1) is simplified to a convex problem with only K real scalar variables rather than a $N \times L$ complex matrix.

Remark 5. By using the result of (61) with $r = 2$, the safe approximation of (65) is a nonconvex problem of $\mathbf{G}$. However, the unimodal property in *Corollary 4* enables the method in [8] can be adopted. In particular, the $f(p_k)$ in [8] is replaced by $f(1 - \phi(p_k, d_k^g))$ for the fitted CSCG case, and $f(p_k)$ is replaced by $\sqrt{\frac{\nu-4}{\nu-2}} f(1 - \phi(p_k, d_k^t))$ for the fitted t-distribution case.

VI. Simulation Results and Discussion

In this section, by using the proposed CCP with fitted distributions, the performance of the probabilistic transceiver design under non-Gaussian channel uncertainty is illustrated. In the simulation, the BS is equipped with four antennas and there are two single antenna users, i.e., $N = 4, K = L = 2, L_k = M_k = 1$. The Logistic distribution and Gaussian mixture are taken as examples for the non-Gaussian channel uncertainties [34], [35]. Furthermore, the non-Gaussian channel uncertainty has zero mean with covariance matrix $\mathbf{\Sigma}_k = 0.01\mathbf{I}_{NM_k}$ and zero relation matrix, and the elements in the channel uncertainty are independent. The covariance matrix of the received interference-plus-noise at the k^{th} user is $\mathbf{R}_k = 0.01\mathbf{I}_{M_k}$. The elements in $\hat{\mathbf{H}}_k$ are generated from independent standard CSCG, and there are 10^3 random channel realizations for the feasibility rate test. The MSE requirements are fixed as $\varepsilon_2 = 0.2$ and $p_1 = p_2 = 10\%$, while the MSE target of the first user is varied under different scenarios.

A. Logistic Channel Uncertainty

In this subsection, with outage target $p_1 = 10\%$, the proposed CCP with fitted distributions are tested under Logistic channel uncertainty.

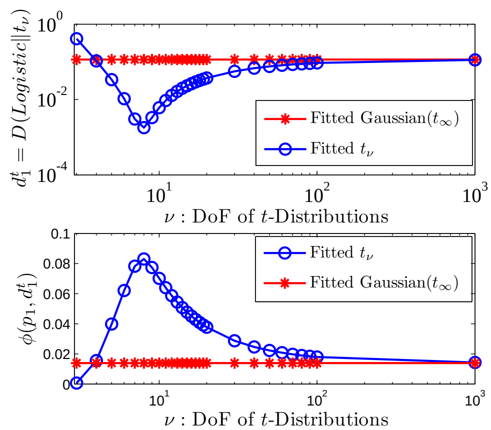


Fig. 1. The KL divergence d_1^t and the regularized outage probability $\phi(p_1, d_1^t)$ versus DoF of fitted t-distribution.

In Fig. 1, the KL divergences between the Logistic distribution and the fitted t-distribution $D(\text{Logistic}||t_\nu)$ with DoF varied from 3 to 10^3 and the corresponding regularized outage probabilities are shown. Firstly, the reversed curves in the KL divergence and the regularized outage probability of Fig. 1 show that the regularized outage probability is a monotonic decreasing function of the KL divergence, which validates the conclusion in *Theorem 2*. Secondly, it is observed that the KL divergences $D(\text{Logistic}||t_\nu)$ with $\nu \geq 4$ are always smaller than $D(\text{Logistic}||\text{Gaussian})$, and the regularized outage probabilities under t_ν with $\nu \geq 4$ are always larger than that of the fitted Gaussian. Therefore, the t-distribution is more flexible than Gaussian to fit the Logistic distribution, and the t-distribution with DoF 8 is a good candidate to fit the Logistic channel uncertainty. Lastly, it is interesting to find that the KL divergence and the regularized outage probability under the fitted Gaussian is similar to that of the fitted t_{10^3}, which is owing to the reason that Gaussian is a limit case of t-distribution.

In Fig. 2, the feasibility rate of the fitted t-distribution and Gaussian are compared under MSE target $\varepsilon_1 = 0.1$. It is observed that the feasibility rates of fitted t-distribution with $\nu \geq 4$ are always larger than that of the fitted Gaussian. Since the feasibility rate is a direct indicator for the largeness of the feasible set, by combining the KL divergence comparison results in Fig. 1 and the feasible rate comparison results in Fig. 2, the conclusion in *Corollary 3* is verified. Fig. 2 also shows that the feasibility rate of the safe approximation with $r = 2$ is larger than that with $r = 1$, the reason is that $R_2(\phi(p = 0.1, d)) > R_1(\phi(p = 0.1, d))$ as implied in (54).

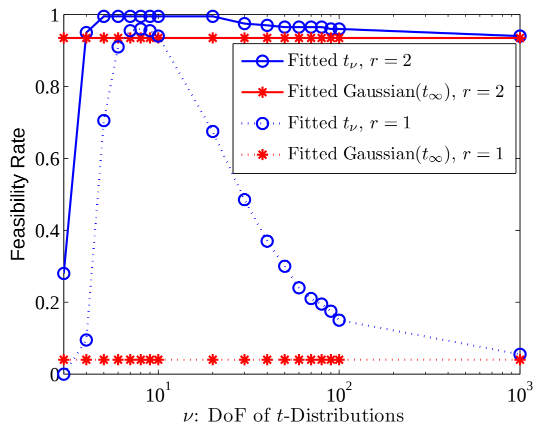


Fig. 2. The feasibility rate versus DoF of fitted t-distribution.

Therefore, the following safe approximations with fitted distribution are all implemented with $r = 2$.

In Fig. 3, the feasibility rate of the proposed CCP with fitted t-distribution (t_8), CCP with fitted Gaussian and the moment method [3] are compared under different MSE requirements. It is observed from Fig. 3 that the feasibility of the CCP with fitted t_8 and Gaussian are larger than that of the moment method. The reason is that the PDF are ignored in the moment method, while the distributional property of the fitted distribution and the PDF mismatch are utilized in the proposed methods. Furthermore, Fig. 3 also shows that the feasibility rate of the CCP with fitted t_8 is larger that of CCP with fitted Gaussian. The reason is that t_8 is closer to the Logistic distribution in terms of KL divergence than Gaussian as shown in Fig. 1.

B. Gaussian Mixture Channel Uncertainty

The Logistic distribution is a classic heavy-tailed distribution, the Gaussian mixture distribution with adjustable tail property is illustrated in this subsection. The density of the Gaussian mixture is $\frac{1}{2}(f_a+f_b)$, where the CSCG density functions f_a and f_b have zero mean with covariance matrix $0.02\rho\mathbf{I}_{NM_k}$ and $0.02(1-\rho)\mathbf{I}_{NM_k}$, respectively.

In Fig. 4, the KL divergences between the Gaussian mixture distribution with $\rho \in \{0.6, 0.7, 0.8\}$ and the fitted t-distribution are shown, and the corresponding regularized outage probability are also shown. Among different Gaussian mixture settings in Fig. 4, the KL divergence between

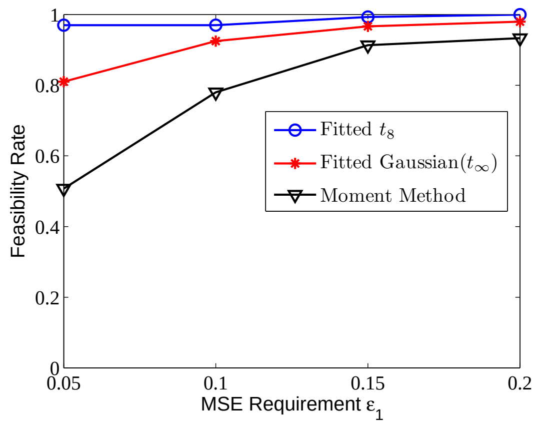
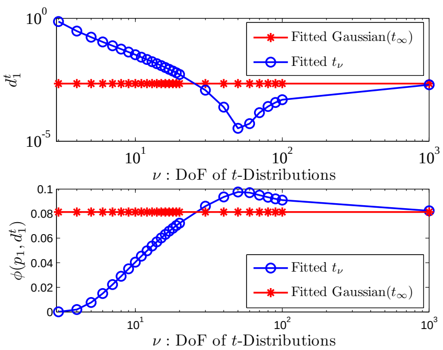


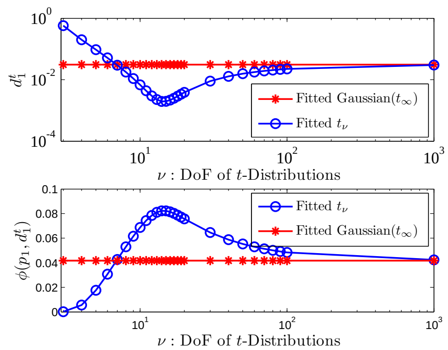
Fig. 3. The feasibility rate of different methods.

Gaussian mixture with $\rho = 0.6$ and the fitted Gaussian is the smallest, while that between Gaussian mixture with $\rho = 0.8$ and the fitted Gaussian is the largest. The reason is that the considered Gaussian mixture is close to the Gaussian distribution $\mathcal{CN}(\mathbf{0}, 0.01\mathbf{I}_{NM_k})$ when ρ is close to 0.5, while the considered Gaussian mixture approaches heavy-tailed distribution when ρ approaches 1. Fig. 4 also reveals that $D(\frac{1}{2}(f_a + f_b)||t_\nu)$ with many νs are smaller than $D(\frac{1}{2}(f_a + f_b)||\mathcal{CN}(\mathbf{0}, 0.01\mathbf{I}_{NM_k}))$, which validates the flexibility of t-distribution. Furthermore, it is not difficult to find that the t-distributions t_{50}, t_{15} and t_7 are good candidates to fit the Gaussian mixture with $\rho = 0.6, \rho = 0.7$ and $\rho = 0.8$, respectively.

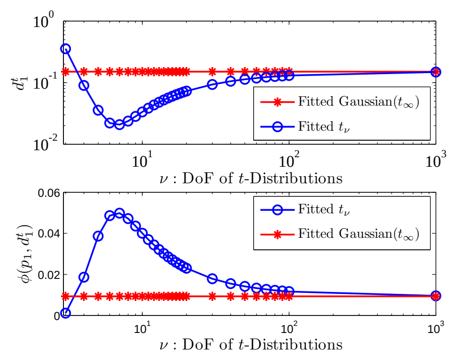
In Fig. 5, the feasibility rates of the proposed CCP with fitted t-distribution, CCP with fitted Gaussian and the moment method [3] are compared under Gaussian mixture channel uncertainty. It is found that the feasibility rate of the CCP with fitted t-distribution and Gaussian are always larger than that of the moment method. The underlying reason is that the proposed methods explore the PDF information, while the moment method ignores the density information. Furthermore, Fig. 5 also shows that the feasibility rate of CCP with fitted t-distribution are larger that that of CCP with fitted Gaussian, which validates the main conclusion that CCP with fitted t-distribution has larger feasible set than CCP with fitted Gaussian.



(a) Gaussian mixture with $\rho = 0.6$



(b) Gaussian mixture with $\rho = 0.7$

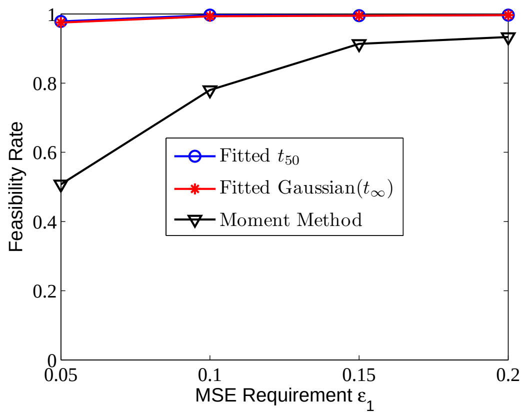


(c) Gaussian mixture with $\rho = 0.8$

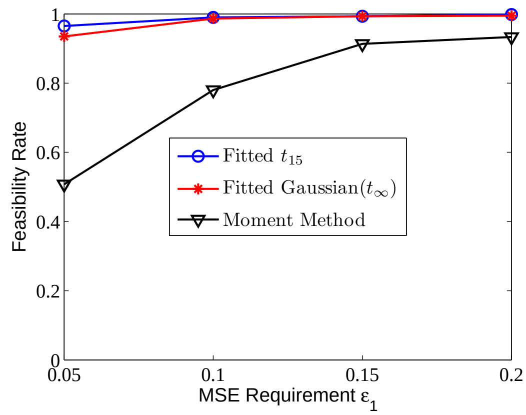
Fig. 4. The KL divergence d_1^t and the regularized outage probability $\phi(p_1, d_1^t)$ versus DoF of fitted t-distribution

VII. Conclusions

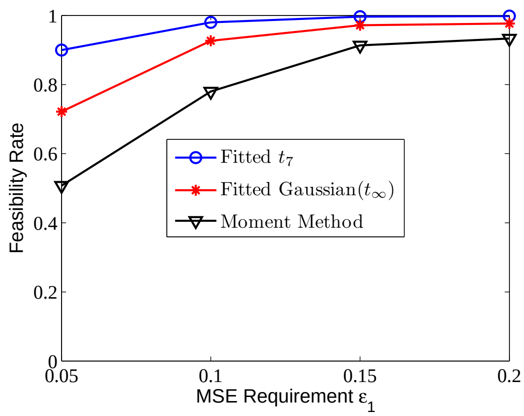
For the CCP problem with non-Gaussian uncertainty, a CCP problem with fitted distribution and a regularized outage probability provides a unified safe approximation. In this paper, the t-distribution rather than Gaussian has been proposed to be the fitted distribution for the quadratically perturbed CCP. The regularized probability was analysed first, and the unimodal property of the quadratical form under t-distribution or Gaussian perturbation was established. Based on these analyses, the analytical condition for the CCP with fitted t-distribution have larger feasible set than the CCP with fitted Gaussian was provided. With Logistic and Gaussian mixture as examples of non-Gaussian uncertainties, simulation results show the less conservative performance of the CCP with fitted t-distribution, compared to the CCP with fitted Gaussian and the classic moment method.



(a) Gaussian mixture with $\rho = 0.6$



(b) Gaussian mixture with $\rho = 0.7$



(c) Gaussian mixture with $\rho = 0.8$

Fig. 5. The feasibility rate of different methods.

APPENDIX A

After introducing the Lagrange multiplier $\eta \geq 0, \lambda \in \mathbb{R}$, the Lagrangian of the problem (3) is

$$L(f(\mathbf{x}), \eta, \lambda) = \int_{g(\mathbf{w},\mathbf{x})\geq\varepsilon} f(\mathbf{x})d\mathbf{x} + \eta(d - \int \ln \frac{f(\mathbf{x})}{f_1(\mathbf{x})} f(\mathbf{x})d\mathbf{x}) + \lambda(1 - \int f(\mathbf{x})d\mathbf{x}) \tag{67}$$

$$= \eta d + \lambda + \int (\mathbf{1}_{\mathcal{G}}(\mathbf{x}) - \lambda - \eta \ln \frac{f(\mathbf{x})}{f_1(\mathbf{x})}) f(\mathbf{x})d\mathbf{x}, \tag{68}$$

where $\mathbf{1}_{\mathcal{G}}(\mathbf{x})$ is an indicator function of the set $\mathcal{G} = \{\mathbf{x} | \ g(\mathbf{w}, \mathbf{x}) \geq \varepsilon\}$, i.e.,

$$\mathbf{1}_{\mathcal{G}}(\mathbf{x}) = \begin{cases} 1 & \forall \mathbf{x} : g(\mathbf{w}, \mathbf{x}) \geq \varepsilon \\ 0 & \forall \mathbf{x} : g(\mathbf{w}, \mathbf{x}) < \varepsilon \end{cases}. \tag{69}$$

With the implicit PDF constraints $f(\mathbf{x}) \geq 0$, the Lagrange dual function of the problem (3) is

$$J(\eta, \lambda) = \sup_{f(\mathbf{x}) \geq 0} L(f(\mathbf{x}), \eta, \lambda), \tag{70}$$

which can be solved by exploring the first and second derivatives of $L(f(\mathbf{x}), \eta, \lambda)$.

Since the Lagrangian $L(f(\mathbf{x}), \eta, \lambda)$ is a functional of the density function $f(\mathbf{x})$, by using the calculus of variations, the first variation of the Lagrangian is

$$\begin{aligned}
\delta L(f(\mathbf{x}); v(\mathbf{x}), \eta, \lambda) &= \lim_{\tau \to 0} \frac{1}{\tau} \left(L(f(\mathbf{x}) + \tau v(\mathbf{x}), \eta, \lambda) - L(f(\mathbf{x}), \eta, \lambda) \right) && (71)\\
&= \lim_{\tau \to 0} \int (\mathbf{1}_{\mathcal{G}}(\mathbf{x}) - \lambda - \eta \ln \frac{f(\mathbf{x}) + \tau v(\mathbf{x})}{f_1(\mathbf{x})}) v(\mathbf{x}) d\mathbf{x} && (72)\\
&- \lim_{\tau \to 0} \int \eta v(\mathbf{x}) \ln (1 + \frac{\tau v(\mathbf{x})}{f(\mathbf{x})})^{\frac{f(\mathbf{x})}{\tau v(\mathbf{x})}} d\mathbf{x} && (73)\\
&= \int \left(\mathbf{1}_{\mathcal{G}}(\mathbf{x}) - \eta - \lambda - \eta \ln \frac{f(\mathbf{x})}{f_1(\mathbf{x})} \right) v(\mathbf{x}) d\mathbf{x}, && (74)
\end{aligned}$$

where $v(\mathbf{x})$ is an arbitrary function. Similarly, the second variation of the Lagrangian is

$$\begin{aligned}
\delta^2 L(f(\mathbf{x}); v(\mathbf{x}), \eta, \lambda) &= -\eta \int \frac{v^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} && (75)\\
&\leq 0. && (76)
\end{aligned}$$

Therefore, the Lagrangian $L(f(\mathbf{x}), \eta, \lambda)$ is a concave functional of $f(\mathbf{x})$. Then the optimal solution is obtained by setting (74) to zero, and since $v(\mathbf{x})$ in (74) is an arbitrary function, we get

$$\mathbf{1}_{\mathcal{G}}(\mathbf{x}) - \eta - \lambda - \eta \ln \frac{f(\mathbf{x})}{f_1(\mathbf{x})} = 0. \tag{77}$$

From (77), it is observed that $\eta \neq 0$, i.e., the optimal solution is happened on the boundary of KL divergence constraint owing to the complementary slackness condition, otherwise the real Lagrange multiplier λ in (77) becomes indicator function. Therefore, the optimal solution of (70) is

$$f(\mathbf{x}) = f_1(\mathbf{x}) \exp(\frac{\mathbf{1}_{\mathcal{G}}(\mathbf{x}) - \eta - \lambda}{\eta}). \tag{78}$$

Note that the nonnegative PDF constraint is automatically satisfied. Although the same solution in (78) was available in [15]–[19], they did not prove the optimality by showing the concave property in (76). Furthermore, the issue of $\eta \neq 0$ is also not addressed in [15]–[19].

After substituting (78) into (70), it becomes

$$J(\eta,\lambda)=\eta d+\lambda+\eta\exp(\frac{-\eta-\lambda}{\eta})\left(\exp(\frac{1}{\eta})\int_{g(\mathbf{w},\mathbf{x})\geq\varepsilon}f_1(\mathbf{x})d\mathbf{x}+\int_{g(\mathbf{w},\mathbf{x})<\varepsilon}f_1(\mathbf{x})d\mathbf{x}\right) \tag{79}$$

$$=\eta d+\lambda+\eta\exp(\frac{-\eta-\lambda}{\eta})\left(1+(\exp(\frac{1}{\eta})-1)\mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x})\geq\varepsilon\}\right) \tag{80}$$

The dual problem of (3) is

$$\begin{aligned}\min_{\eta,\lambda}\quad & J(\eta,\lambda)\\ \text{s.t.}\quad & \eta>0.\end{aligned} \tag{81}$$

Since the dual problem is a convex problem and there is no constraint on λ, the optimal λ must satisfy the first order condition

$$\frac{\partial J(\eta,\lambda)}{\partial\lambda}=0, \tag{82}$$

and the optimal λ is

$$\lambda=\eta\ln(1+(e^{1/\eta}-1)\mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x})\geq\varepsilon\})-\eta. \tag{83}$$

After substituting the optimal λ into (81), the dual problem becomes

$$\begin{aligned}\min_{\eta}\quad & J(\eta)=\eta d+\eta\ln\left(1+(e^{1/\eta}-1)\mathbb{P}_{f_1(\mathbf{x})}\{g(\mathbf{w},\mathbf{x})\geq\varepsilon\}\right)\\ \text{s.t.}\quad & \eta>0.\end{aligned} \tag{84}$$

Appendix B
Proof of the Theorem 1

First, the first variation of the KL divergence is

$$\delta D(f(\mathbf{x});v(\mathbf{x})\|f_1(\mathbf{x}))=\lim_{\tau\to0}\frac{1}{\tau}\left(D(f(\mathbf{x})+\tau v(\mathbf{x}),f_1(\mathbf{x}))-D(f(\mathbf{x}),f_1(\mathbf{x}))\right) \tag{85}$$

$$=\lim_{\tau\to0}\int v(\mathbf{x})\ln(1+\frac{\tau v(\mathbf{x})}{f(\mathbf{x})})^{\frac{f(\mathbf{x})}{\tau v(\mathbf{x})}}d\mathbf{x}+\lim_{\tau\to0}\int(\ln\frac{f(\mathbf{x})+\tau v(\mathbf{x})}{f_1(\mathbf{x})})v(\mathbf{x})d\mathbf{x} \tag{86}$$

$$=\int(1+\ln\frac{f(\mathbf{x})}{f_1(\mathbf{x})})v(\mathbf{x})d\mathbf{x}. \tag{87}$$

Similarly, the second variation of the KL divergence is

$$\delta^2D(f(\mathbf{x});v(\mathbf{x})\|f_1(\mathbf{x}))=\int\frac{v^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} \tag{88}$$

$$\geq 0. \tag{89}$$

Therefore, the $D(f(\mathbf{x})\|f_1(\mathbf{x}))$ is a convex functional of $f(\mathbf{x})$. From problem (3), it is observed the cost functional and the functional in the second constraint are linear functionals of $f(\mathbf{x})$. Therefore, the primal problem (3) is a convex problem, and strong duality holds between primal and its dual if the primal problem is strictly feasible and its optimal value is finite [36, p. 224]. This two conditions can be proved as follows. Firstly, since the feasible solution $f(\mathbf{x}) = f_1(\mathbf{x})$ makes $D(f(\mathbf{x})\|f_1(\mathbf{x})) = 0$, the condition $d > 0$ makes the primal problem strictly feasible. Secondly, the primal problem is upper bounded by 1. Therefore, $d > 0$ guarantees the strong duality.

Appendix C

Since $s \sim \frac{1}{\nu}\chi^2(\nu)$ is a scaled chi-square distribution, its moment information can be calculated as

$$\mathbb{E}(s) = \nu/\nu = 1, \tag{90}$$

$$\mathbb{E}(s - \mathbb{E}(s))^2 = 2\nu/\nu^2 = 2/\nu. \tag{91}$$

As $\nu \to \infty$, the fixed mean in (90) and the vanishing variance in (91) reveals $\lim_{\nu\to\infty} s = \lim_{\nu\to\infty} \chi^2(\nu)/\nu = 1$. Together with the definition in (28) and (29), the result $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C}) \overset{d}{=} t_\infty(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{C})$ is obtained. Therefore, according to Theorem 3, the unimodal property for Gaussian quadratic form also holds for $\mathbf{w} \in \{\mathbf{w} \in \mathbb{C}^N | \mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}\}$.

Appendix D

With real t-distributed $\mathbf{x} = \boldsymbol{\mu} + s^{-1/2}\mathbf{n}$, the real quadratic form in (39) becomes

$$\mathbf{x}^T\mathbf{A}(\mathbf{w})\mathbf{x} = \mathbf{n}^T\mathbf{A}(\mathbf{w})\mathbf{n}/s + 2\mathbf{n}^T\mathbf{A}(\mathbf{w})\boldsymbol{\mu}/s^{\frac{1}{2}} + \boldsymbol{\mu}^T\mathbf{A}(\mathbf{w})\boldsymbol{\mu}. \tag{92}$$

With $\mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}$ and similar derivations in (34)-(37), the stochastic representation of (92) is

$$\mathbf{x}^T\mathbf{A}(\mathbf{w})\mathbf{x} = \mathbf{n}^T\mathbf{A}(\mathbf{w})\mathbf{n}/s \tag{93}$$

$$\overset{d}{=} s^{-1}\sum_{i=1}^{N} \lambda_i^n \chi_i^2(1), \tag{94}$$

where $\{\lambda_i^n\}_{i=1}^N$ are the eigenvalues of $(\boldsymbol{\Sigma}^{1/2})^T\mathbf{A}(\mathbf{w})\boldsymbol{\Sigma}^{1/2}$. Since the representation of (94) is the same as that of (37), the real t-perturbed quadratic form in (39) is unimodal under the decision variable $\mathbf{w} \in \{\mathbf{w} \in \mathbb{R}^N | \mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}\}$. Furthermore, since $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \overset{d}{=} t_\infty(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as shown in

Appendix C, the unimodal property for real Gaussian perturbed quadratic form in (39) also holds for $\mathbf{w} \in \{\mathbf{w} \in \mathbb{R}^N | \mathbf{A}(\mathbf{w})\boldsymbol{\mu} = \mathbf{0}\}$.

Appendix E

With $\mathbf{h}_k = \mathrm{vec}(\mathbf{H}_k), \mathbf{d}_k = \mathrm{vec}(\mathbf{D}_k)$, the MSE is reformulated as

$$\mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) = \|\mathbf{F}_k\mathbf{H}_k\mathbf{G} - \mathbf{D}_k\|_F^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H) \tag{95}$$

$$= \|(\mathbf{G}^T \otimes \mathbf{F}_k)\mathbf{h}_k - \mathbf{d}_k\|_2^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H). \tag{96}$$

Since the size of matrix $\mathbf{G}^T \otimes \mathbf{F}_k$ is $LL_k \times NM_k$ and $LL_k \leq NM_k$, the rank of $\mathbf{G}^T \otimes \mathbf{F}_k$ lies in the range $[0, LL_k]$. Denote the rank of $\mathbf{G}^T \otimes \mathbf{F}_k$ as n_k, and $n_k \in [0, LL_k]$ is obtained.

If $\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} = \mathbf{D}_k$, then $n_k = LL_k$ and the MSE in (96) can be reformulated as

$$\mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) = \|(\mathbf{G}^T \otimes \mathbf{F}_k)(\mathbf{h}_k - (\mathbf{G}^T \otimes \mathbf{F}_k)^\dagger \mathbf{d}_k)\|_2^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H). \tag{97}$$

Note that the first term of (97) is a quadratic form of the new random variable $\mathbf{h}_k - (\mathbf{G}^T \otimes \mathbf{F}_k)^\dagger \mathbf{d}_k$, its mean is $\mathrm{vec}(\hat{\mathbf{H}}_k) - (\mathbf{G}^T \otimes \mathbf{F}_k)^\dagger \mathbf{d}_k$. If $\mathbf{h}_k$ is fitted as t-distribution, according to *Theorem 3*, the MSE in (97) is unimodal if $(\mathbf{G}^T \otimes \mathbf{F}_k) \cdot \big(\mathrm{vec}(\hat{\mathbf{H}}_k) - (\mathbf{G}^T \otimes \mathbf{F}_k)^\dagger \mathbf{d}_k\big) = 0$, i.e. $\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} = \mathbf{D}_k$. On the other hand, if $\mathbf{h}_k$ is fitted as $\mathcal{N}_{\mathbb{C}}(\mathrm{vec}(\hat{\mathbf{H}}_k), \boldsymbol{\Sigma}_k, \mathbf{C}_k)$ with $\mathbf{C}_k = \mathbf{0}$, according to *Theorem 4*, the MSE in (97) is unimodal under any transceiver pair with $n_k = LL_k$.

When $n_k < LL_k$, the MSE reformulation in (97) can not be obtained. However, the unimodal property of MSE under CSCG channel uncertainty can still be proved as follows.

If $n_k \in [1, LL_k - 1]$, by using the singular value decomposition, the MSE in (96) is reformulated as

$$\begin{aligned}
&\mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) \\
&= \|[\mathbf{U}\Big[\mathrm{Diag}([\sigma_1, ..., \sigma_{n_k}, \mathbf{0}]) \ \mathbf{0}_{LL_k \times (NM_k - LL_k)}\Big]\mathbf{V}^H \cdot \mathbf{h}_k - \mathbf{d}_k\|_2^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H), && (98) \\
&= \|\Big[\mathrm{Diag}([\sigma_1, ..., \sigma_{n_k}, \mathbf{0}]) \ \mathbf{0}_{LL_k \times (NM_k - LL_k)}\Big]\mathbf{V}^H \cdot \mathbf{h}_k - \underbrace{\mathbf{U}^H\mathbf{d}_k}_{\mathbf{e}_k}\|_2^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H), && (99) \\
&= \|\Big[\mathrm{Diag}([\sigma_1, ..., \sigma_{n_k}]) \ \mathbf{0}_{n_k \times (NM_k - LL_k)}\Big]\mathbf{V}^H \cdot \mathbf{h}_k - \mathbf{e}_k[1:n_k]\|_2^2 \\
&\quad + \|\mathbf{e}_k[n_k + 1 : LL_k]\|_2^2 + \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_k\mathbf{F}_k^H), && (100)
\end{aligned}$$

where $\mathbf{e}_k[1:n_k]$ represents the first n_k elements of $\mathbf{e}_k$. If the fitted channel $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathrm{vec}(\hat{\mathbf{H}}_k), \boldsymbol{\Sigma}_k, \mathbf{0})$, (100) can be formulated as a quadratic form under CSCG perturbation. Therefore, according to

Theorem 4, the MSE in (100) is unimodal. On the other hand, if $n_k = 0$, the MSE in (96) becomes

$$\mathrm{MSE}_k(\mathbf{G}, \mathbf{F}_k, \mathbf{H}_k) = \|\mathbf{d}_k\|_2^2 + \mathrm{Tr}(\mathbf{F}_k \mathbf{R}_k \mathbf{F}_k^H), \tag{101}$$

which is a deterministic variable and unimodal.

Therefore, for any transceiver pair with $n_k \in [0, LL_k]$, the MSE in (96) with $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathrm{vec}(\hat{\mathbf{H}}_k), \mathbf{\Sigma}_k, \mathbf{0})$ is unimodal.

References

[1] A. Ben-Tal, S. Bhadra, C. Bhattacharyya, A. Nemirovski, "Efficient methods for robust classification under uncertainty in kernel matrices", *Journal of Machine Learning Research*, vol. 13, pp. 2923-2954, 2012.

[2] E. Erdoğan and G. Iyengar, "Ambiguous chance constrained problems and robust optimization," *Mathematical Programming*, pp. 37-61, 2004.

[3] 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, 2013.

[4] X. He and Y.C. Wu, "Tight probabilistic SINR constrained beamforming under channel uncertainties," to appear in *IEEE Trans. Signal Processing*, 2015.

[5] A. Nemirovski, "On safe tractable approximations of chance constraints", *European Journal of Operational Research*, vol. 219, pp. 707-718, 2012.

[6] 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.

[7] 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. System Theory*, pp. 173-175, 2007.

[8] N. Vučić and H. Boche, "A Tractable Method for Chance-Constrained Power Control in Downlink Multiuser MISO Systems With Channel Uncertainty," *IEEE Signal Processing Letters,* vol. 16, no. 5, pp. 346-349, 2009.

[9] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," 2009, preprint, available on http://arxiv.org/abs/0909.3595.

[10] 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.

[11] 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, 2013.

[12] W. Xu, A. Tajer, X. Wang, and S. Alshomrani, "Power allocation in MISO interference channels with stochastic CSIT," *IEEE Trans. Wireless Commun.*, vol.13, no.3, pp. 1716-1727, 2014.

[13] J. Smith, "Generalized chebyshev inequalities: theory and applications in decision analysis," *Oper. Res.*, vol. 43, pp. 807-825, 1995.

[14] P. K. Shivaswamy, C. Bhattacharyya, and A. J. Smola, "Second order cone programming approaches for handling missing and uncertain data," *Journal of Machine Learning Research*, vol. 7, pp. 1283-1314, 2006.

[15] B.C. Levy, R. Nikoukhah, "Robust least-squares estimation with a relative entropy constraint," *IEEE Trans. Information Theory,* vol. 50, no. 1, pp. 89-104, 2004.

[16] Y. Guo, B.C. Levy, "Robust MSE equalizer design for MIMO communication systems in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1840-1852, 2006.

[17] B. C. Levy, "Robust hypothesis testing with a relative entropy tolerance" *IEEE Trans. Information Theory*, vol. 55, pp. 413-421, 2009.

[18] L. E. Ghaoui, M. Oks, F. Oustry, "Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach," *Oper. Res.*, vol. 51, no. 4, pp. 543-556, 2003.

[19] R. Jiang and Y. Guan, "data-driven chance constrained stochastic program," *Math. Program.* 2015.

[20] 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.

[21] M. C. de Oliveira, J. Bernussou, and J. C. Geromel, "A new discrete-time robust stability condition", *Syst. Contr. Lett.*, vol. 37, pp. 261-265, 1999.

[22] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall, 1990.

[23] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637-2640, 1996.

[24] E. Ollila, J. Eriksson, and V. Koivunen, "Complex elliptically symmetric random variables-Generation, characterization, and circularity tests," *IEEE Trans. Signal Process.,* vol. 59, no. 1, pp. 5869, 2011.

[25] A. M. Iksanov, Z. J. Jurek, and B. M. Schreiber, "A new factorization property of the selfdecomposable probability measures," *Ann. Probab.*, vol. 32, no. 2, pp. 1356-1369, 2004.

[26] M. Yamazato, "Unimodality of infinitely divisible distribution functions of class L," *Ann. Probab.*, vol. 6, no. 4, pp. 523-531, 1978.

[27] J.M. Bernardo, A.F.M. Smith, *Bayesian Theory*, Wiley, 1993.

[28] I. Cuculescu and R. Theodorescu, "Multiplicative strong unimodality," *Austral. & New Zealand J. Statist.* vol. 40, no. 2, pp. 205-214, 1998.

[29] R. G. Gallager, "Circularly-Symmetric Gaussian random vectors," 2008.

[30] S. Das Gupta and S. K. Sarkar, "On TP2 and log-concavity," in: Y. L. Tong (Ed.), *Inequalities in Statistics and Probability*, Lecture Notes-monograph series, Institute of Mathematical Statistics, vol. 5, pp. 54-58, 1984.

[31] S. Dharmadhikari, K. Joag-Dev, *Unimodality, Convexity and Applications*, Academic Press, San Diego, CA, 1988.

[32] S. Kotz, S. Nadarajah, *Multivariate t Distributions and Their Applications*. Cambridge Univ. Press. 2004.

[33] 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, 2009.

[34] R. W. Heath Jr., M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using Poisson Point Processes" *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4114-4126, 2013.

[35] A. Rabbachin, T. Q.S. Quek, H. Shin, and M. Z. Win, "Cognitive network interference" *Selected Areas in Communications, IEEE Journal on*, Vol. 29, NO. 2, pp. 480-493, 2011.

[36] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, 1969.