On Spectral Efficiency for Multiuser MISO Systems under Imperfect Channel Information

Quang-Doanh Vu, Le-Nam Tran, *Senior Member, IEEE*, and Markku Juntti, *Fellow, IEEE*

***Abstract*—We consider downlink transmission whereby a multiantenna base station simultaneously transmits data to multiple single-antenna users. We focus on slow flat fading channel where the channel state information is imperfect, the channel estimation error is unbounded and its statistics are known. The aim is to design beamforming vectors such that the sum rate is maximized under the constraints on probability of successful transmission for each user and maximum transmit power. The optimization problem is intractable due to the chance constraints. To this end, we propose an efficient solution drawn upon stochastic optimization. In particular, we first use the step function and its smooth approximation to get an approximate nonconvex stochastic program of the considered problem. We then develop an iterative procedure to solve the stochastic program based on the stochastic successive convex approximation framework. The numerical results show that the proposed solution can achieve remarkable sum rate gains compared to the conventional one.**

***Index Terms*—MISO, beamforming, imperfect channel estimation, sum rate, stochastic optimization.**

I. Introduction

Improving spectral efficiency is one of the main tasks in wireless communications because the demand on data traffic is always growing while the spectrum resource is scarce. Multiuser multi-antenna transmissions is one of the key techniques to achieve high spectral efficiency. The efficiency of multiuser multi-antenna relies on the availability of channel state information (CSI), i.e., the performance would degrade with the presence of uncertain CSI [1]. In reality, getting perfect (i.e., error-free) CSI is very difficult [2]. Thus it is essential to take into account the CSI uncertainty in beamforming design.

There are two main CSI uncertainty models, namely, deterministic model and stochastic model [3]. The deterministic model uses deterministically-bounded additive uncertainty sets, i.e., the actual channel is assumed to lie in prespecified bounded sets centered at the estimated channels. This model is useful for the systems where the channels are estimated and quantized at receivers, which are then fed back to transmitters, e.g., the frequency division duplex (FDD) systems where the channel reciprocity of uplink and downlink does not hold. On the other hand, the stochastic model assumes that the error parts are random variables [4]. This model is appropriate for the systems with uplink–downlink reciprocity (e.g., TDD systems) where transmitters can estimate users' channels via the pilot signals received on the uplink.

Herein, we focus on the stochastic CSI error model and on the long-term performance. The reason for considering the stochastic model is that it can properly represent several types of CSI error such as the error caused by the minimum mean squared error estimator [5] or time delay in reciprocity-based channel estimation [3], [4]. For the long-term performance, a popular approach is focusing on sum ergodic data rate [6]–[8]. This approach can achieve good average performance, but, it is inapplicable to slow fading channel and delay-sensitive users since the codeword should be sufficiently long to experience all possible fading states [9, Chap. 4]. Alternatively, for such applications, the outage-based approach, i.e., transmission with some outage probability, is more suitable [9], [10]. The challenge of such an approach is to handle the chance constraints resulting from guaranteeing some predefined outage probability threshold. A possible solution is to approximate the chance constraints using bounded uncertainty sets, then to solve the approximate problem using worst-case robust optimization [4], [11]–[13]. However, this approach is often too conservative, possibly leading to poor performance [12]. For the problem of power minimization, there are other approaches to deal with chance constraints based on the Bernstein-type inequality, decomposition-based large deviation inequality [12], or offset approximation [14]. However, these approaches are not applicable to the sum rate maximization problem considered in this paper because the data rates are the design parameters.

Recently, stochastic optimization was successfully used for the problem of power minimization with outage constraint in multiuser downlink multiple-input single-output (MISO) systems [15, Example 2]. Therein, it was numerically shown that the stochastic optimization approach can outperform other existing methods in terms of power efficiency. However, it remains to be seen whether such an optimization approach can lead to an efficient solution for the spectral efficiency maximization problem, since the structures of which are different from those in [15, Example 2]. One of the goals in this paper is to understand the potential spectral efficiency gains that an stochastic approach can provide over related existing methods, which was not reported in [15] .

In this paper, we consider multiuser downlink MISO sys-

This work was supported in part by the Academy of Finland under projects FURMESFuN (Grant 310898), 6Genesis Flagship (Grant 318927), WiConIE (Grant 297803) and EERA (Grant 332362), and in part by Science Foundation Ireland under Grant 17/CDA/4786.

Quang-Doanh Vu was with Centre for Wireless Communications, University of Oulu, Finland. He is now with the Mobile Networks, Nokia, 90650 Oulu, Finland. Email: vuquangdoanh@gmail.com.

Le-Nam Tran is with School of Electrical and Electronic Engineering, University College Dublin, Ireland. Email: nam.tran@ucd.ie.

Markku Juntti is with Centre for Wireless Communications, University of Oulu, FI-90014, Finland. Email: markku.juntti@oulu.fi.

tems where the CSI is imperfect. We aim at designing beamforming vectors at the base station (BS) so that the spectral efficiency is maximized under the outage probability constraints for each user. Inspired by the recent success of stochastic optimization on the problem of power minimization with outage constraint [15], and outage minimization [16], we develop an efficient solution to the considered problem based on the stochastic successive convex approximation framework [15]. Towards this end, the outage probability constraints are approximated using the Heaviside step function and its smooth approximation. This leads to a stochastic nonconvex program, which is then solved by an iterative procedure following the stochastic successive convex approximation framework. At each iteration of the algorithm, only conic quadratic programs are solved with the computational cost is $\mathcal{O}(M^{3.5})$, where M is the number of antennas at the BS. Finally, we provide extensive numerical results demonstrating that the proposed algorithm can significantly outperform the existing method using bounded uncertainty set and robust optimization.

Notation: Bold lower and upper case letters represent vectors and matrices, respectively; $\|\mathbf{a}\|$ represents the ℓ_2 norm of $\mathbf{a}$; $|a|$ represents the absolute value of a; $\mathbb{C}^{a\times b}$ represents the space of complex matrices of dimensions given in superscript; $\mathcal{CN}(0,a)$ denotes a complex Gaussian random variable with zero mean and variance a; $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ represent real and image part of the argument, respectively; $\mathbb{E}\{\cdot\}$ denote the expectation operator; $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$ stand for normal transpose and Hermitian of $\mathbf{A}$, respectively; $\mathbf{I}_M$ represents an $M\times M$ identity matrix; $\Pr\{A\}$ denotes probability of event A; ∇f denote the gradient of f. Notation $\langle\mathbf{a},\mathbf{b}\rangle$ denotes the inner product of the two vectors, and notation $[\mathbf{x};\mathbf{y}]$ stands for the vector obtained by staking vertically $\mathbf{x}$ and $\mathbf{y}$. Other notations are defined at their first appearance.

II. System Model and Problem Formulation

We consider a system where a BS equipped with M antennas simultaneously transmits data to U single-antenna users. Let $\mathbf{h}_k\in\mathbb{C}^{1\times M}$ denote the channel (row) vector between the BS to user k, and $\mathbf{v}_k\in\mathbb{C}^{M\times 1}$ denote the beamforming vector for user k. Then the signal received at user k is

$$y_k=\mathbf{h}_k\mathbf{v}_kx_k+\mathbf{h}_k\Big(\textstyle\sum_{t=1,t\neq k}^{U}\mathbf{v}_tx_t\Big)+z_k \quad (1)$$

where x_k is the normalized transmitted data symbol for user k, i.e., $\mathbb{E}\{|x_k|\}=1$, $\forall k$, and $z_k\sim\mathcal{CN}(0,\sigma_k^2)$ is the additive white Gaussian noise (AWGN) at user k.

We suppose that the channel is flat and slow fading, and the CSI at the BS is imperfect while the users has perfect CSI. In particular, the channel is modeled as [4], [12]

$$\mathbf{h}_k=\bar{\mathbf{h}}_k+\boldsymbol{\theta}_k$$

where $\bar{\mathbf{h}}_k\in\mathbb{C}^{1\times M}$ is the estimated channel vector, and $\boldsymbol{\theta}_k\in\mathbb{C}^{1\times M}$ is a channel error random vector following probability distribution $\mathcal{P}_k$, which is assumed to be known [4], [12]. Then the signal-to-interference-plus-noise ratio (SINR) at user k is

$$\gamma_k(\mathbf{v};\boldsymbol{\theta}_k)=\frac{\mathbf{v}_k^{\mathrm{H}}\mathbf{H}_k(\boldsymbol{\theta}_k)\mathbf{v}_k}{\sum_{t=1,t\neq k}^{U}\mathbf{v}_t^{\mathrm{H}}\mathbf{H}_k(\boldsymbol{\theta}_k)\mathbf{v}_t+\sigma_k^2} \quad (2)$$

where $\mathbf{v}\triangleq[\mathbf{v}_1;...;\mathbf{v}_U]$, and $\mathbf{H}_k(\boldsymbol{\theta}_k)\triangleq(\bar{\mathbf{h}}_k+\boldsymbol{\theta}_k)^{\mathrm{H}}(\bar{\mathbf{h}}_k+\boldsymbol{\theta}_k)$. Let $r_k=\log(1+\mu_k)$, $\mu_k\geq 0$, be the data rate for user k. Due to the channel estimation error, the transmission rate is determined such that the probability of successfully transmission is high. Let us denote by $\phi\in(0,1)$ the predefined successful threshold transmission for a user. Then the problem of maximizing the sum rate reads

$$\underset{\mathbf{v},\boldsymbol{\mu}}{\text{maximize}}\ \textstyle\sum_{k=1}^{U}\log(1+\mu_k) \quad (3a)$$

$$\text{subject to}\ \Pr\{\gamma_k(\mathbf{v};\boldsymbol{\theta}_k)\geq\mu_k\}\geq\phi,\forall k \quad (3b)$$

$$||\mathbf{v}||^2\leq\bar{P},\mu_k\geq 0,\ \forall k \quad (3c)$$

where $\boldsymbol{\mu}\triangleq[\mu_1,...,\mu_U]$, and $\bar{P}$ is the maximum transmit power at the BS. Note that the quantity $\log(1+\mu_k)$ may not be achievable due to the imperfect CSI. We explain in detail how to compute the achievable rate at the text corresponding to Fig. 2 in Section IV. As discussed, the approaches based on the ergodic capacity might achieve better average performance [6]–[8], [17]. However, they are not suitable to quasi-static channels considered in this paper [9].

The challenge in solving (3) is due to the chance constraint (3b) which is computationally intractable [18]. For μ_k is fixed which is the case for the problem of power minimization subject to the outage constraints, we can find tractable approximations of (3b) when $\boldsymbol{\theta}_k$ has a circularly symmetric complex Gaussian distribution, by using sphere bounding, the Bernstein-type inequality, or decomposition-based large deviation inequality together with semidefinite relaxation [12]. However, for problem (3), these approaches are inapplicable, since μ_k is a variable. Popular existing approaches to solving (3) are based on robust optimization [19]. In particular, sphere bounding is used to turn (3b) into semi-infinite constraints [4]. Then a tractable safe approximation is achieved by treating error vector $\boldsymbol{\theta}_k$ at the numerator and the denominator of $\gamma_k(\mathbf{v}_k;\boldsymbol{\theta}_k)$ independently [11], [19], [20]. However, the application of the such approach is restricted to the specific case of distribution of $\boldsymbol{\theta}_k$, i.e. circularly symmetric complex Gaussian distribution. Also, the performance of such a conservative method is significantly degraded. We present below an efficient solution to (3) by means of stochastic optimization.

III. Proposed Algorithm based on Stochastic Optimization

For the ease of exposition, we convert (3) into the real-domain. Let us introduce some notations $\tilde{\mathbf{v}}_k=[\mathrm{Re}(\mathbf{v}_k);\mathrm{Im}(\mathbf{v}_k)]$, $\tilde{\mathbf{v}}=[\tilde{\mathbf{v}}_1,...,\tilde{\mathbf{v}}_U]$, $\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)=[\mathrm{Re}(\mathbf{H}_k(\boldsymbol{\theta}_k)),-\mathrm{Im}(\mathbf{H}_k(\boldsymbol{\theta}_k));\mathrm{Im}(\mathbf{H}_k(\boldsymbol{\theta}_k)),\mathrm{Re}(\mathbf{H}_k(\boldsymbol{\theta}_k))]$. Then, we can rewrite the SINR at user k as

$$\tilde{\gamma}_k(\tilde{\mathbf{v}};\boldsymbol{\theta}_k)=\frac{\tilde{\mathbf{v}}_k^{\mathrm{T}}\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)\tilde{\mathbf{v}}_k}{\sum_{t=1,t\neq k}^{U}\tilde{\mathbf{v}}_t^{\mathrm{T}}\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)\tilde{\mathbf{v}}_t+\sigma_k^2}.$$

A. Approximate Stochastic Program of (3)

As the first step, we approximate (3) into a stochastic program. To do so, we rewrite the chance constraint (3b) using the Heaviside step function. In particular, we have

$$\mathbb{1}\left(f_k(\tilde{\mathbf{v}},\mu_k;\boldsymbol{\theta}_k)\right)=\begin{cases}1 & \text{if } f_k(\tilde{\mathbf{v}},\mu_k;\boldsymbol{\theta}_k)\geq 0\\ 0 & \text{otherwise}\end{cases}$$

where $f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = \tilde{\mathbf{v}}_k^{\mathrm{T}} \tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k) \tilde{\mathbf{v}}_k - \mu_k \left(\sum_{t=1, t\neq k}^{U} \tilde{\mathbf{v}}_t^{\mathrm{T}} \tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k) \tilde{\mathbf{v}}_t + \sigma_k^2 \right)$. Then constraint (3b) is equivalently rewritten as [21, Lemma 1.3]

$$\mathbb{E}_{\boldsymbol{\theta}_k \sim \mathcal{P}_k} \left\{ \mathbb{1}\left(f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \right) \right\} \geq \phi.$$

The Heaviside step function is discontinuous at zero. A common way to overcome this issue is to use a smooth approximation [15], [16]. We note that there are several smooth approximations of the Heaviside step function. Here, we use the logistic function as the smooth approximation function, since it is simple, continuously differentiable and efficient [16], which is given by [22]

$$\begin{aligned} &\mathbb{1}\left(f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \right) \approx \\ &\qquad g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = 0.5 + 0.5 \tanh \kappa (f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k))) \end{aligned}$$

where κ is the accuracy parameter, i.e., a larger κ corresponds to a sharper transition. In practice, parameter κ should be tuned for striking the balance between the approximation accuracy and numerical stability. Finally, the approximate stochastic program of (3) is

$$\begin{aligned} &\underset{\tilde{\mathbf{v}}, \boldsymbol{\mu}}{\text{maximize}} \ \textstyle\sum_{k=1}^{U} \log(1 + \mu_k) && \text{(4a)} \\ &\text{subject to } \mathbb{E}_{\boldsymbol{\theta}_k \sim \mathcal{P}_k} \left\{ g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \right\} \geq \phi, \forall k && \text{(4b)} \\ &\qquad \tilde{\mathbf{v}}^{\mathrm{T}} \tilde{\mathbf{v}} \leq \bar{P}, \mu_k \geq 0, \ \forall k. && \text{(4c)} \end{aligned}$$

which is a nonconvex constrained stochastic optimization problem.

B. The Iterative Procedure

We now apply the stochastic successive convex approximation framework to achieve an efficient solution to (4) [15], which is an iterative procedure in nature. In each iteration, a new sample of channel errors is generated, then the expectations of random functions are replaced by their convex surrogate ones which contain the information of the past as well as new samples. After that the resulting approximate convex problem is solved to determine the direction in which the parameters are updated.

Let us consider the iteration n where the sample of channel error vectors $\{\boldsymbol{\theta}_k^{(n)}\}_k$ are randomly generated following probability distribution $\{\mathcal{P}_k\}_k$. Let $(\tilde{\mathbf{v}}^{(n)}, \boldsymbol{\mu}^{(n)})$ be the current iterate. As a key step, we form the convex approximation of (4b) using the recursive formula. In particular, we first construct the sample surrogate function of $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k^{(n)})$ at point $(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)})$. To do so, we have the derivative of $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)$ given as $\nabla g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \triangleq [\nabla_{\tilde{\mathbf{v}}} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k); \nabla_{\mu_k} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)]$ where

$$\nabla_{\tilde{\mathbf{v}}_k} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = 2 \hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k) \tilde{\mathbf{v}}_k \qquad (5)$$

$$\nabla_{\tilde{\mathbf{v}}_t} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = -2 \mu_k \hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k) \tilde{\mathbf{v}}_t, \forall t \neq k \qquad (6)$$

$$\begin{aligned} &\nabla_{\mu_k} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = \\ &- \hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \left(\textstyle\sum_{t=1, t\neq k}^{U} \tilde{\mathbf{v}}_m^{\mathrm{T}} \tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k) \tilde{\mathbf{v}}_t + \sigma_k^2 \right) \qquad (7) \end{aligned}$$

and $\hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = 0.5 \kappa (1 - \tanh^2 \kappa (f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)))$. Then the sample surrogate function is obtained by the first order approximation of $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k^{(n)})$ at point $(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)})$ as

$$\begin{aligned} \tilde{g}_k(\tilde{\mathbf{v}}, \mu_k; \tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k^{(n)}) &= g_k(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k^{(n)}) \\ &+ \left\langle \nabla g_k(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k^{(n)}), [\tilde{\mathbf{v}}; \mu_k] - [\tilde{\mathbf{v}}^{(n)}; \mu_k^{(n)}] \right\rangle \\ &- \rho_k || [\tilde{\mathbf{v}}; \mu_k] - [\tilde{\mathbf{v}}^{(n)}; \mu_k^{(n)}] ||^2 \qquad (8) \end{aligned}$$

where $\rho_k > 0$; here the quadratic regularizing term is added to make $\tilde{g}_k(\tilde{\mathbf{v}}, \mu_k; \tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k^{(n)})$ strongly concave [15, Assumption 4]. Finally, we have the convex approximation of (4b) given as

$$\begin{aligned} G_k^{(n)}(\tilde{\mathbf{v}}, \mu_k) &\triangleq (1 - \alpha_k^{(n)}) G_k^{(n-1)}(\tilde{\mathbf{v}}, \mu_k) \\ &+ \alpha_k^{(n)} \tilde{g}_k(\tilde{\mathbf{v}}, \mu_k; \tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k^{(n)}) \qquad (9) \end{aligned}$$

in which $\alpha_k^{(n)} \in (0, 1]$ is a weighting factor. The important property of the formula in (9) is that, with suitably chosen $\{\alpha_k^{(n)}\}$ (specified later in Alg. 1), we can have $G_k^{(n)}(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}) \to \mathbb{E}_{\boldsymbol{\theta}_k \sim \mathcal{P}_k} \left\{ g_k(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k) \right\}$ and $\nabla G_k^{(n)}(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}) \to \nabla \mathbb{E}_{\boldsymbol{\theta}_k \sim \mathcal{P}_k} \left\{ \nabla g_k(\tilde{\mathbf{v}}^{(n)}, \mu_k^{(n)}; \boldsymbol{\theta}_k) \right\}$ when $n \to \infty$ [15, Proposition 1].

By the arguments above, we arrive at the convex approximate problem solved at iteration n given as

$$\begin{aligned} &\underset{\tilde{\mathbf{v}}, \boldsymbol{\mu}}{\text{maximize}} \ \textstyle\sum_{k=1}^{U} \log(1 + \mu_k) && \text{(10a)} \\ &\text{subject to } G_k^{(n)}(\tilde{\mathbf{v}}, \mu_k) \geq \phi, \forall k && \text{(10b)} \\ &\qquad \tilde{\mathbf{v}}^{\mathrm{T}} \tilde{\mathbf{v}} \leq \bar{P}, \mu_k \geq 0, \ \forall k. && \text{(10c)} \end{aligned}$$

If problem (10) is infeasible, the following problem is solved to determine the update direction

$$\underset{\tilde{\mathbf{v}}, \boldsymbol{\mu}, t}{\text{maximize}} \ t \ \text{ subject to } \ G_k^{(n)}(\tilde{\mathbf{v}}, \mu_k) \geq \phi + t, \forall k, \text{(10c)}. \quad (11)$$

We note that problem (11) is always feasible and the objective is bounded. Let $(\tilde{\mathbf{v}}^{\star}, \boldsymbol{\mu}^{\star})$ denote the optimal solution of (10) or (11), then the next iterate $(\tilde{\mathbf{v}}^{(n+1)}, \boldsymbol{\mu}^{(n+1)})$ is given by

$$\tilde{\mathbf{v}}^{(n+1)} = (1 - \beta^{(n)}) \tilde{\mathbf{v}}^{(n)} + \beta^{(n)} \tilde{\mathbf{v}}^{\star}, \qquad (12)$$

$$\boldsymbol{\mu}^{(n+1)} = (1 - \beta^{(n)}) \boldsymbol{\mu}^{(n+1)} + \beta^{(n)} \boldsymbol{\mu}^{\star} \qquad (13)$$

where $\beta^{(n)}$ is the update step size. In summary, the proposed algorithm is outlined in Alg. 1.

C. Convergence Analysis

We discuss the convergence of Alg. 1 following closely the arguments in [15]. First, we justify that problem (3) satisfies the conditions on the feasible set and expectation functions presented in [15, Assumption 1], which are required to establish the convergence. Let us consider the set $\mathcal{S}_B = \{(\tilde{\mathbf{v}}, \boldsymbol{\mu}) | 0 \leq \mu_k \leq B, \forall k, ||\tilde{\mathbf{v}}||^2 \leq \bar{P}\}$ where $B > 0$. Given some finite B, $\mathcal{S}_B$ is compact and convex. We note that with B is large enough, we can equivalently rewrite (4) as $\underset{(\tilde{\mathbf{v}}, \boldsymbol{\mu}) \in \mathcal{S}_B}{\text{maximize}} \ \sum_{k=1}^{U} \log(1 + \mu_k)$ subject to (4b). This can also be applied in (10) and (11). In addition, we can observe from

Algorithm 1 The proposed algorithm

1: **Initialization**: an initial point $(\tilde{\mathbf{v}}^{(0)}, \boldsymbol{\mu}^{(0)})$, set $n = 0$, $0.5 < \tau_1 < \tau_2 \leq 1$.
2: **repeat**
3: Generate sample $\{\boldsymbol{\theta}_k^{(n)}\}_k$ independently following $\{\mathcal{P}_k\}_k$.
4: Update $\alpha_k^{(n)} = (1+n)^{-\tau_1}, \forall k$, and $\beta^{(n)} = (1+n)^{-\tau_2}$.
5: Update $G_k^{(n)}(\tilde{\mathbf{v}}, \mu_k)$ following (9).
6: Compute $(\tilde{\mathbf{v}}^{\star}, \boldsymbol{\mu}^{\star})$ (by solving (10) or (11))
7: Compute $(\tilde{\mathbf{v}}^{(n+1)}, \boldsymbol{\mu}^{(n+1)})$ following (12) and (13)
8: $n := n + 1$
9: **until** stopping criteria is met

(5), (6), and (7) that $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)$ is continuously differentiable on $\mathcal{S}_B$. Moreover, we recall that $0 \leq g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \leq 1$. So $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)$ is uniformly bounded. Also, we have

$$\begin{aligned} ||\nabla_{\tilde{\mathbf{v}}_k} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)|| &\leq \kappa \sqrt{\bar{P}} ||\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)|| \\ ||\nabla_{\tilde{\mathbf{v}}_t} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)|| &\leq \kappa B \sqrt{\bar{P}} ||\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)||, \forall t \neq k \\ |\nabla_{\mu_k} g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)| &\leq 0.5\kappa(\bar{P}||\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)|| + \sigma_k^2) \end{aligned}$$

In practice, the magnitude of channel gain is bounded and so is $||\tilde{\mathbf{H}}_k(\boldsymbol{\theta}_k)||$. Consequently, the first derivative of $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)$ is uniformly bounded. Similarly, we have

$$\begin{aligned} \nabla^2 g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) = \nabla \hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)(\nabla f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k))^{\mathrm{T}} \\ + \hat{g}_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k) \nabla^2 f_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k). \quad (14) \end{aligned}$$

From this, we can determine a finite bound for each term in the right hand side of (14), leading to the property that second order derivative of $g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)$ is uniformly bounded. We skip the details for the sake of brevity.

Now, let $\{(\tilde{\mathbf{v}}^{(n)}, \boldsymbol{\mu}^{(n)})\}$ be the iterates generated by Alg. 1. According Theorem 1 in [15], if the step size $\beta^{(0)}$ is significantly small and $(\tilde{\mathbf{v}}^{(0)}, \boldsymbol{\mu}^{(0)})$ is a feasible point of (4), then every limiting point of $\{(\tilde{\mathbf{v}}^{(n)}, \boldsymbol{\mu}^{(n)})\}$ which satisfies the Slater condition is almost surely a stationary point of (4). For our considered problem, we can get a feasible point $(\tilde{\mathbf{v}}^{(0)}, \boldsymbol{\mu}^{(0)})$ of (4) by generating values of elements in $\boldsymbol{\mu}^{(0)}$ sufficiently small.

D. Computational Complexity

In each iteration, problem (10) needs to be solved. We note that the objective function of (10) can be equivalently rewritten as a geometric mean, i.e. $(\prod_{k=1}^{U}(1+\mu_k))^{1/U}$. Thus, problem (10) can be represented by a conic quadratic program [23]. Consequently, the worst case computational cost for solving (10) by a generic interior point method solver is $\mathcal{O}(M^{3.5})$ [24]. In the case that problem (10) is infeasible, problem (11) is solved. The computational cost order for (11) is same as that for (10).

In the systems where the number of antennas at the BS is large, using an interior point solver to solve (10) may be computationally expensive. A possible approach overcoming this issue is to use the alternating direction method of multipliers for parallel processing, i.e., splitting the large-scale problem (10) into smaller-scale subproblems which can be computed in parallel [25]. Another possible approach is to adopt first-order methods, whose per-iteration computational complexity is cheap, since only first-order information is used [26].

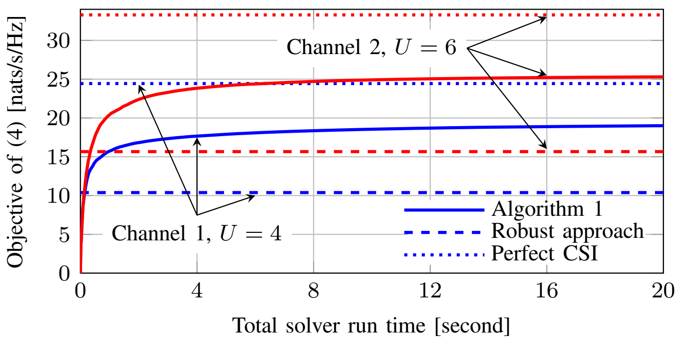
IV. Numerical Results

In this section, we numerically evaluate the performance of Alg. 1. We first consider simulation configuration based on that in [12], [15]. In particular, the estimated channel vectors are generated as $\bar{\mathbf{h}}_k \sim \mathcal{CN}(0, (1-\delta^2)\mathbf{I}_M), \forall k$; the channel error vectors are set as $\boldsymbol{\theta}_k^{(n)} \sim \mathcal{CN}(0, \delta^2\mathbf{I}_M), \forall k$ where δ^2 represents the estimate error level which is specified in the experiment. Unless otherwise stated, we take the parameters as follows. The noise variance is $\sigma_k^2 = 1, \forall k$. The maximum transmit power at the BS is $\bar{P} = 20$ dB. The number of antennas at the BS is $M = 20$. The probability of successful transmission is set as $\phi = 0.9$. We take $\tau_1 = 0.9$, $\tau_2 = 0.91$, and $\rho_k = 10^{-4}, \forall k$. The accuracy parameter is $\kappa = 0.85$. The number of iterations of Alg. 1 is set as 3000. The convex subproblems are solved using the solver MOSEK (version 9.1) [27]. The codes are executed on a 64-bit laptop with 16 Gbyte RAM and Intel CORE i5.

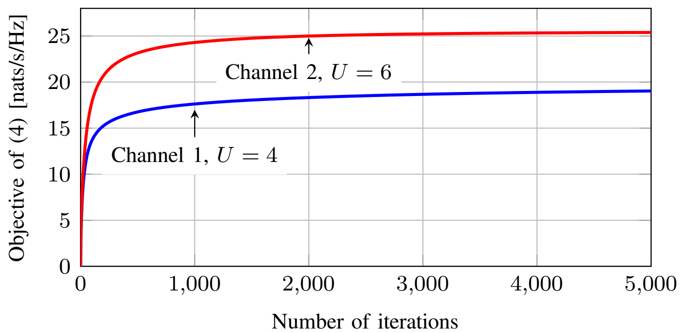
We consider the solution based on the robust optimization as a baseline scheme. In particular, given ϕ, constraint (3b) is approximated by guaranteeing that the transmission is successful for all channel error vectors $\boldsymbol{\theta}_k$ inside the ball $\mathcal{B}_k = \{\boldsymbol{\theta}_k | \lambda \geq ||\boldsymbol{\theta}_k||^2\}$ where $\lambda = \frac{\delta^2}{2}\Phi^{-1}_{\chi^2_{2M}}(\phi)$ and $\Phi^{-1}_{\chi^2_{2M}}$ is the inverse cumulative distribution function of the Chi-square random variable with $2M$ degrees of freedom [4]. With $\{\mathcal{B}_k\}_k$, the robust-based method in [11] is used to obtained the solution. We note that the robust approach based on [11] is also an iterative procedure (developed based on alternating optimization). There are two semidefinite programs needed to be solved in each iteration [11, Theorem 6]. The worst case computational cost solving these two programs by a generic interior point solver with respect to the number of transmit antennas are $\mathcal{O}(M^{4.5})$ and $\mathcal{O}(M^{2.5})$. Thus the computational complexity of the robust approach scales faster with M compared to Alg. 1. We also provide the result for the ideal case that the CSI is perfectly known, for which we use the solution developed in [28].

In Fig. 1, we plot the value of objective function of (4) achieved by Alg. 1 over the total run time of the solver (Fig. 1(a)) and number of iteration (Fig. 1(b)) for two random channel realizations. We also provide the achieved sum rate of the perfect CSI, and robust approach solutions. As we can see that Alg. 1 can converge to a higher sum rate compared to the conventional robust approach. Here, the beamforming vectors and transmit data rate are achieved by Alg. 1 are feasible for all users in the two channels.

In Fig. 2, we numerically investigate the average achieved sum rate of Alg. 1 with different level of CSI uncertainty. For each setting, the considered schemes are averaged over 2000 random channel realizations. For calculating sum rate, if the obtained transmit rate and beamforming vectors are infeasible for a user, then the rate corresponding to the user is zero (i.e. treated as unsuccessful transmission). We can observe from the



(a) The achieved objective value over total solver run time



(b) The achieved objective value over number of iterations

Figure 1. The achieved objective value of (4) by Alg. 1 over the total solver run time and number of iterations with two channel realizations. We take $\delta^2 = 0.01$.

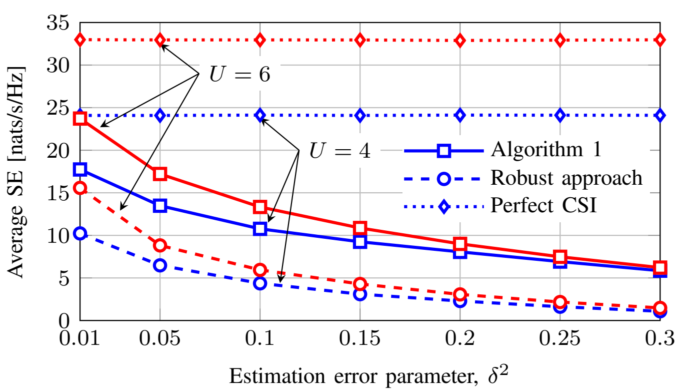


Figure 2. The average achieved spectral efficiency (SE) of the considered schemes over different level of channel estimation error.

figure that Alg. 1 significantly outperforms the robust approach in term of spectral efficiency in all cases of considered network settings. This is because the robust approach is conservative resulting from the worst case design. Specifically, the problem formulated using the robust optimization approach is still intractable. In order to have an efficient solution, a tractable approximation of the problem was employed. Following the

Table I
THE MAXIMUM OUTAGE PROBABILITY (IN PERCENT) AMONG THE USERS CORRESPONDING TO THE RESULTS IN FIG. 2.

δ^2	0.01	0.05	0.1	0.15	0.2	0.25	0.3
$U = 4$	0.5	8.7	9.3	6.4	4.0	2.80	1.75
$U = 6$	0.2	2.55	2.3	1.5	1.25	1.0	0.95

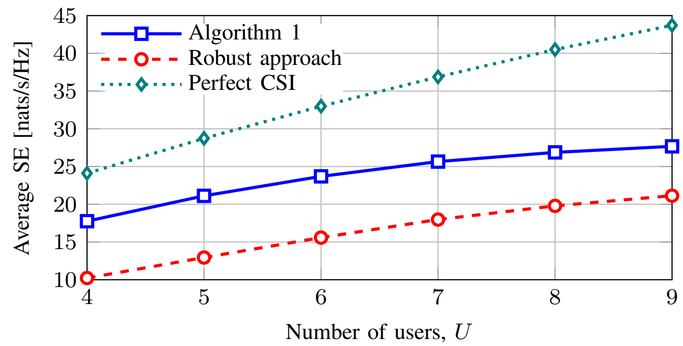


Figure 3. The average achieved spectral efficiency (SE) of the considered schemes with different numbers of users. We take $\delta^2 = 0.01$.

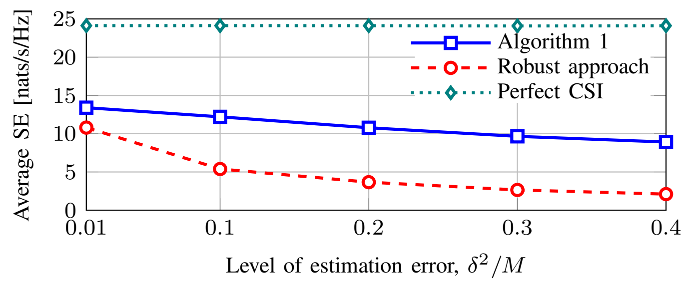


Figure 4. The average achieved spectral efficiency (SE) of the considered schemes with the channel error vectors $\{\boldsymbol{\theta}_k\}$ are uniformly distributed in predefined balls. We take $U = 4$.

concept of worst-case robust optimization, the approximation is obtained using conservative bounds, which likely degrades the performance of the approach. We can also observe from the figure the impact of the level of channel estimation error on the spectral efficiency: the larger the estimation error variance the smaller the achieved spectral efficiency. This is reasonable since more resource is required in order to compensate the uncertainty when the estimation error variance increases.

In Table I, we show the outage probability corresponding to the results in Fig. 2 to numerically verify whether the probability constraints in (3b) are satisfied by Alg. 1. Since there are multiple users, only the largest outage probability (among the users) for each system setting is provided. We can see that the outage probability corresponding to Alg. 1 is less than 10% in all cases. This means the constraints in (3b) are satisfied since we set $\phi = 0.9$. Also, we can observe that the the largest outage probability is not the same for different values of δ^2 and U. This is because the outage constraints are approximated by functions $\{g_k(\tilde{\mathbf{v}}, \mu_k; \boldsymbol{\theta}_k)\}_k$ which depend on U and $\{\boldsymbol{\theta}_k\}_k$. Thus, for a given κ, the outage probability could change when U and δ^2 vary.

Fig. 3 shows the performance of the considered schemes as the function of the number of users with $\delta^2 = 0.01$. We can observe that the performance of all considered schemes increase when U increases, which comes from the multiuser diversity gain. Again, we can observe that Alg. 1 achieves significant spectral efficiency gains compared to the robust approach with all considered numbers of users.

In Fig. 4, we consider another channel estimation error model which is a result of the quantization [29]. In particular,

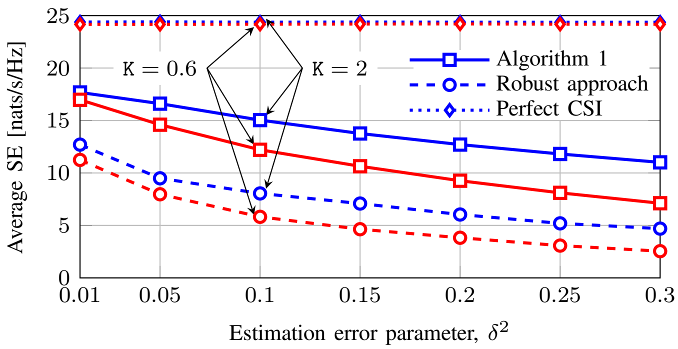


Figure 5. The average achieved spectral efficiency of the considered schemes over Rician fading over different level of channel estimation error.

the channel error vector $\boldsymbol{\theta}_k$ is assumed to be uniformly distributed in the ball $\tilde{\mathcal{B}}_k = \{\boldsymbol{\theta}_k | \delta^2 \geq ||\boldsymbol{\theta}_k||^2\}$. Consequently, the ball corresponding to outage parameter ϕ considered in the baseline scheme is $\mathcal{B}_k = \{\boldsymbol{\theta}_k | \delta^2 \sqrt[M]{\phi} \geq ||\boldsymbol{\theta}_k||^2\}$ [29, Eq. (24)]. We take $\tau_1 = 0.92$, $\tau_2 = 0.96$. Fig. 4 shows the average spectral efficiency of the considered schemes with different levels of estimation error. Similar to the estimation error with Gaussian distribution, we can observe in Fig. 4 that Alg. 1 outperforms the robust approach in all cases. Also, the performance of both schemes reduce when δ^2 increases.

We now evaluate the performance of Alg. 1 over Rician fading. In particular, let $\mathtt{K}$ denote the power ratio between the line-of-sight (LoS) and non-line-of-sight (NLoS) components. Let us denote by $\check{\mathbf{h}}_k$ the mean corresponding to the LoS component, which is generated as $[\check{\mathbf{h}}_k]_m = \sqrt{\frac{\mathtt{K}}{1+\mathtt{K}}} e^{j(m-1)\pi \sin \varphi_k}$ where $\varphi_k \in [0, 2\pi)$ represents the angle-of-departure. Then, the estimated channel vectors are generated as $\bar{\mathbf{h}}_k \sim \mathcal{CN}(\check{\mathbf{h}}_k, \frac{1-\delta^2}{1+\mathtt{K}} \mathbf{I}_M), \forall k$, and the estimation error vectors are set as $\boldsymbol{\theta}_k^{(n)} \sim \mathcal{CN}(0, \frac{\delta^2}{1+\mathtt{K}} \mathbf{I}_M), \forall k$ [5]. We observe that, given δ^2 with larger $\mathtt{K}$, Alg. 1 and the robust approach achieves better performance. This is because the uncertainty level reduces with $\mathtt{K}$. Also, similar to the Rayleigh fading, the proposed algorithm outperforms the existing solution in all cases of δ^2 and $\mathtt{K}$.

V. Conclusion

We have investigated the multiuser downlink MISO systems where the channel uncertainty is modeled following some probability distribution. In particular, we have focused on designing beamforming vectors and transmit data rate such that the spectral efficiency is maximized under the constraints on maximum transmit power at the BS and the outage probability for each user. Towards a tractable formulation, we have approximated the chance constraints using smooth approximate function of the Heaviside step function, then arrived at a nonconvex stochastic program. After that we have developed an iterative procedure to find efficient solutions based on the stochastic convex approximation framework. Finally, we have provided extensive numerical results showing that the proposed approach can achieve better spectral efficiency compared to the existing approach using robust optimization.

References

[1] M. B. Shenouda and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, Dec. 2007.

[2] N. Jindal, "Mimo broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[3] M. Payaro, A. Pascual-Iserte, and M. A. Lagunas, "Robust power allocation designs for multiuser and multiantenna downlink communication systems through convex optimization," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1390–1401, Sep. 2007.

[4] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[5] O. Özdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with spatially correlated Rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, Jan. 2019.

[6] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, Jun 2016.

[7] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, "A parallel decomposition method for nonconvex stochastic multi-agent optimization problems," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2949–2964, June 2016.

[8] L. You, J. Xiong, A. Zappone, W. Wang, and X. Gao, "Spectral efficiency and energy efficiency tradeoff in massive MIMO downlink transmission with statistical CSIT," *IEEE Trans. Signal Process.*, vol. 68, pp. 2645–2659, April 2020.

[9] A. Goldsmith, *Wireless Communications*, A. Goldsmith, Ed. Cambridge, 2005.

[10] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, October 2008.

[11] A. Tajer, N. Prasad, and X. Wang, "Robust linear precoder design for multi-cell downlink transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 235–251, Jan 2011.

[12] K. Y. Wang, A. M. C. So, T. H. Chang, W. K. Ma, and C. Y. 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.

[13] H. Vaezy, M. J. Omidi, and H. Yanikomeroglu, "Energy efficient precoder in multi-user MIMO systems with imperfect channel state information," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 669–672, June 2019.

[14] M. Medra, Y. Huang, and T. N. Davidson, "Offset-based beamforming: A new approach to robust downlink transmission," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 70–82, Jan. 2019.

[15] A. Liu, V. K. N. Lau, and B. Kananian, "Stochastic successive convex approximation for non-convex constrained stochastic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4189–4203, Aug 2019.

[16] Y. Shi, A. Konar, N. D. Sidiropoulos, X. Mao, and Y. Liu, "Learning to beamform for minimum outage," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5180–5193, Oct. 2018.

[17] M. M. Zhao, Y. Cai, M. J. Zhao, Y. Xu, and L. Hanzo, "Robust joint hybrid analog-digital transceiver design for full-duplex mmwave multicell systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4788–4802, 2020.

[18] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1574–1609, 2009.

[19] S. K. Joshi, U. L. Wijewardhana, M. Codreanu, and M. Latva-aho, "Maximization of worst-case weighted sum-rate for miso downlink systems with imperfect channel knowledge," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3671–3685, Oct 2015.

[20] M. F. Hanif, L.-N. Tran, A. Tölli, M. Juntti, and S. Glisic, "Efficient solutions for weighted sum rate maximization in multicellular networks with channel uncertainties," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5659–5674, Nov. 2013.

[21] A. R. Meyer, "Expectation & variance," https://www.cs.princeton.edu/courses/archive/fall06/cos341/handouts/variance-notes.pdf, 2006.

[22] N. Kyurkchiev and S. Markov, "On the Hausdorff distance between the Heaviside step function and Verhulst logistic function," *Journal of Mathematical Chemistry*, vol. 54, pp. 109–119, 2016.

[23] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Soc. Ind. Appl. Math. (SIAM), 2001.

[24] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193–228, Nov. 1998.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[26] S. Bubeck, "Convex optimization: Algorithms and complexity," *Found. Trends Mach. Learn.*, vol. 8, no. 3-4, pp. 231–357, 2015.

[27] MOSEK ApS, 2020, [Online]. Available: www.mosek.com.

[28] L.-N. Tran, M. Hanif, A. Tölli, and M. Juntti, "Fast converging algorithm for weighted sum rate maximization in multicell MISO downlink," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 872–875, Dec 2012.

[29] M. Botros Shenouda, T. N. Davidson, and L. Lampe, "Outage-based design of robust Tomlinson-Harashima transceivers for the MISO downlink with QoS requirements," *Signal Processing*, vol. 93, no. 12, pp. 3341–3352, 2013.