Beamforming Design for Rate Splitting MIMO

Ticao Zhang and Shiwen Mao
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
Email: tzz0031@tigermail.auburn.edu, smao@ieee.org

2021 IEEE International Conference on Communications Workshops (ICC Workshops) | 978-1-7281-9441-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCWorkshops50388.2021.9473811

Abstract—**In this paper, an overloaded downlink rate splitting multiuser multiple input single output (MU-MISO) system is considered. To combat interference, we propose a beamforming method based on rate-splitting (RS), which divides a beamforming vector into a common vector and a private vector. A successive convex approximate (SCA) based optimization approach is then proposed to achieve high max-min fairness performance. The proposed approach is compared with the conventional weighted minimum mean square error (WMMSE) based approach in terms of convergence speed and max-min fairness (MMF) rate performance. Moreover, to adapt to the stochastic nature of wireless channel, a stochastic successive convex approximate (SSCA) based approach is proposed. The algorithm is based on approximating the original non-convex optimization problem with their convex surrogate function and performs convex optimization. Compared with the conventional sample average approximation (SAA) method, the proposed scheme does not need to collect a large number of samples for solving the optimization problem. Simulation results show that the proposed SCA based RS scheme converges faster than the conventional WMMSE based approach and outperforms the WMMSE based scheme in terms of the MMF rate performance. The proposed SSCA based approach can achieve a satisfactory performance with a significantly lower complexity.**

Index Terms—**Max-min fairness, multiple input multiple output (MIMO), rate splitting, successive convex approximation, stochastic optimization.**

I. Introduction

With the rapid development of advanced multimedia applications, such as virtual reality (VR) and 360 degree video, the next-generation wireless networks must deliver high spectral efficiency and support high connectivity requirements. Recently, Rate-splitting multiple access (RSMA) has been recognized as a promising non-orthogonal transmission technique for interference management in cellular networks. With RSMA, the information intended for each receiver is split into a *common part* and a *private part*. The common parts are encoded into one data stream and delivered to *all* the UEs. The private part, intended for the corresponding UE, can be decoded with successive interference cancellation (SIC) after the decoding of the common parts. The flexibility of managing interference to be partially decoded and partially treated as noise, makes RSMA a highly promising technique and a new frontier for the PHY layer of 6G [1], [2].

The concept of rate splitting dates back to the 1980s [3], when the rate region for a two-user single input single output (SISO) interference channel was analyzed. Later on, the benefit of RS has been demonstrated in several multiple input single output (MISO) broadcast channel settings [4]–[7]. Conventional multiple access interference management techniques, such as space division multiple access (SDMA) and non-orthogonal multiple access (NOMA), are actually the two extreme cases of RS. SDMA treats all the signals delivered to other UEs as interference, while NOMA requires to completely decode the signal delivered to the weaker UEs. Compared with SDMA and NOMA, RS offers extra degrees of freedom by dividing each message into a common part and a private part [1], [7]. As a result, RSMA can achieve a tradeoff and outperform both NOMA and SDMA in terms of spectrum efficiency in a wide range of networks [8], [9].

In [7], the performance of MMF rate for RS scheme was analyzed using degrees of freedom (DoF) analysis. It was shown that RS could solve the rate-saturation issue and achieve better MMF rate performance in both underloaded and overloaded wireless systems. In [8], the author studied the robust beamforming design for RS in terms of the max-min rate (MMF) performance. The concept of RS was later investigated for robust secure beamforming [10] in a two user system and was shown to outperforms the existing NOMA scheme. The energy efficiency performance of RS was recently studied in both the multicell system [11] and the intelligent reflecting surface (IRS) assisted networks [12] with an SCA and semidefinite programming (SDR) based approach.

In this paper, we study a downlink MU-MISO system with RS from a fairness standpoint and investigate the precoder design in terms of MMF rate. This is a well-known non-convex optimization problem. So far, most of the existing works leverage the equivalency between the weighted mean square error (WMMSE) and the achievable rate expressions to decompose the non-convex problem into two sub-problems: (i) optimal precoder design for given MMSE weights, and (ii) obtaining the optimal MMSE weights based on a given precoder. Alternative optimization is used to find a stationary solution iteratively. Some prior works studied the energy efficiency problem in RS with SCA and SDR based approaches [10]–[13]. To the best of our knowledge, there has been no work in the literature that studies both SDR/SCA and WMMSE in a joint optimization framework with RS. In this work, we compare their performance in terms of MMF. Our work shows that the SCA based approach provides a tighter approximation bound, converges faster, and achieves a better performance.

In practice, it is difficult to obtain perfect channel state information (CSI); only partial CSI is available. To address the channel estimation error, in this work, we also formulate a stochastic optimization problem and incorporate the stochastic successive convex approximate (SSCA) framework [14] into our solution algorithm. The stochastic non-convex constraint is approximated by a properly designed convex surrogate convex function and the beamforming vector is updated in an online fashion. Compared with the offline sample average approximate (SAA) algorithm, SSCA does not require a sample collection phase to obtain a sufficiently large number of channel samples. Hence, SSCA yields a faster convergence speed with a lower complexity.

This paper is organized as follows. The system model and problem formulation are described in Section II. The proposed solution is presented in Section III for the perfect CSI case and in Section IV for the imperfect CSI case. Simulation results are analyzed in Section V and Section VI concludes the paper.

II. System Model and Problem Formulation

A. Rate splitting model with IRS

We consider a downlink MISO communication system, which consists of a BS with M antennas, and K single antenna user equipments (UEs). In rate splitting multiple access, the message X_k intended for UE k is split into a private part $X_{\mathrm{p},k}$ and a common part $X_{\mathrm{c},k}$. The private parts, i.e., $X_{\mathrm{p},1}$, $X_{\mathrm{p},2}$, ..., $X_{\mathrm{p},K}$, are encoded independently into Gaussian data symbol streams, denoted as $[s_1, s_2, ..., s_K]^T \in \mathbb{C}^{K\times 1}$. Meanwhile, the common parts of all UEs, i.e., $X_{\mathrm{c},1}, X_{\mathrm{c},2}, ..., X_{\mathrm{c},K}$ are combined into a common message X_{c}, which is encoded into a common stream s_{c} with a public codebook known to all the UEs. As a result, the combined symbols are grouped in a vector $\mathbf{s} = [s_{\mathrm{c}}, s_1, s_2, ..., s_K]^T \in \mathbb{C}^{(K+1)\times 1}$. Each signal is assumed to have zero mean and unit variance, i.e., $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{K+1}$.

The set of UEs is denoted as $\mathcal{K} = \{1, 2, ..., K\}$. At the transmitter, the precoding matrix for all UEs is denoted as $\mathbf{W} = [\mathbf{w}_{\mathrm{c}}, \mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathbb{C}^{M\times(K+1)}$ where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the precoding matrix for UE $k \in \mathcal{K}$ for the private data and $\mathbf{w}_{\mathrm{c}}$ is the precoding matrix for the common message s_{c}. Then, the transmit signal is $\mathbf{W}\mathbf{s}$.

We use $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ to denote the direct channel from the BS to UE k. Since the channel between the BS and UE k might be blocked, we assume channel $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ has Rayleigh fading. Then for UE $k \in \mathcal{K}$, the received signal is given by

$$y_k = \mathbf{h}_k^H \mathbf{W}\mathbf{s} + n_k, \quad \forall k \in \mathcal{K}. \tag{1}$$

where $n_k \sim \mathcal{CN}(0, \sigma_0^2)$ is the additive white Gaussian noise (AWGN) at UE k.

At the receiver end, each UE first decodes the common stream by treating all the private streams as noise. Then its private message is decoded by removing the decoded common stream with successive interference cancellation (SIC). The decoding SINR for the common message and the private message for UE k at the receiver is given, respectively, by

$$\gamma_{\mathrm{c},k} = \frac{|\mathbf{h}_k^H \mathbf{w}_{\mathrm{c}}|^2}{\sum_{i\in\mathcal{K}} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_0^2}, \quad \forall k \in \mathcal{K} \tag{2}$$

$$\gamma_{\mathrm{p},k} = \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i\in\mathcal{K}, i\neq k} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_0^2}, \quad \forall k \in \mathcal{K}. \tag{3}$$

Under Gaussian signaling, the achievable rates of UE k in decoding the common and private messages are given by

$$R_{\mathrm{c},k} = \log_2(1 + \gamma_{c,k}) \tag{4}$$

$$R_{\mathrm{p},k} = \log_2(1 + \gamma_{p,k}). \tag{5}$$

In order to ensure that all UEs can decode the common message stream, the actual rate of the common message stream, which we denote as R_{c}, is constrained by each of the $R_{\mathrm{c},k}$, i.e.,

$$R_{\mathrm{c}} = \min_{k\in\mathcal{K}} R_{\mathrm{c},k}. \tag{6}$$

According to the RS decoding principle, the actual data stream R_{c} is shared by all UEs. By denoting $r_{\mathrm{c},k}$ as the general common rate allocated to UE k, we have

$$\sum_{k\in\mathcal{K}} r_{\mathrm{c},k} \leq R_{\mathrm{c}}, \quad r_{\mathrm{c},k} \geq 0. \tag{7}$$

After removing the impact of the common data stream, each UE decodes its own private message. Finally, the overall achievable data rate for UE k is given by

$$R_k = r_{\mathrm{c},k} + R_{\mathrm{p},k}, \, \forall k \in \mathcal{K}. \tag{8}$$

B. Problem formulation

We aim to maximize the minimum achievable rate of all UEs by performing beamforming at the BS. Specifically, this problem can be mathematically formulated as

$$\max_{\substack{\{\mathbf{w}_i\}_{i\in\mathcal{M}},\\ \{r_{\mathrm{c},k}\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} R_k \tag{9a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{M}} ||\mathbf{w}_k||^2 \leq P_{\max} \tag{9b}$$

$$\sum_{k\in\mathcal{K}} r_{\mathrm{c},k} \leq R_{\mathrm{c},k}, \quad \forall k \in \mathcal{K} \tag{9c}$$

$$r_{\mathrm{c},k} \geq 0, \tag{9d}$$

where $P_{\max}$ is the total transmit power at the BS and $\mathcal{M} = \mathcal{K} \cup \{\mathrm{c}\}$ denotes the combined set. In Problem (9), constraint (9b) denotes the power constraint and constraint (9c) can be obtained from (4) and (6).

In practice, when the perfect CSI is unknown, we model the estimated channel as $\widehat{\mathbf{h}}_k = \mathbf{h}_k + \mathbf{e}_k \in \mathbb{C}^{M\times 1}$, with $\mathbf{e}_k \in \mathbb{C}^{M\times 1}$ being the channel estimation error. Suppose we have N channel observations in total. Then the sampled channel observation set can be denoted as

$$\mathcal{H} = \{\mathcal{H}^n, \forall 1 \leq n \leq N | \mathcal{H}^n = [\widehat{\mathbf{h}}_1^n, \widehat{\mathbf{h}}_2^n, ..., \widehat{\mathbf{h}}_K^n]\}. \tag{10}$$

The BS may improve the average (or ergodic) max-min rate for all UEs under channel estimation error. Specifically, the beamforming problem with imperfect CSI can be formulated as the following stochastic optimization problem.

$$\max_{\substack{\{\mathbf{w}_i\}_{i\in\mathcal{M}},\\ \{r_{c,k}\}_{k\in\mathcal{K}}}} \quad \min_{k\in\mathcal{K}} \mathbb{E}[R_k] \tag{11a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{M}} ||\mathbf{w}_k||^2 \leq P_{\max} \tag{11b}$$

$$\sum_{k\in\mathcal{K}} r_{c,k} \leq \mathbb{E}[R_{c,k}], \quad \forall k\in\mathcal{K} \tag{11c}$$

$$r_{c,k} \geq 0, \tag{11d}$$

where the expectations are taken w.r.t. all the channel realizations in $\mathcal{H}$.

III. Perfect CSI Case

In this section, we assume that the channel state is perfectly known. The problem is a deterministic non-convex optimization. The non-convexity comes from the expression of $R_{\mathrm{p},k}$ and $R_{\mathrm{c},k}$. To be specific, as can be seen from (5), the expression of $R_{\mathrm{p},k}$ involves the logarithmic operation of $\gamma_{\mathrm{p},k}$, which is concave. The inner function $\gamma_{\mathrm{p},k}$, as shown in (3) involves a quadratic-over-linear operation of $\mathbf{w}_i$, which is convex. As a result, the combined expression of $R_{\mathrm{p},k}$ is neither concave nor convex.

So far, most of the research uses the WMMSE framework to find a stationary solution, which utilizes the relation between mutual information and MMSE [15], [16]. In this section, we propose a different scheme, which is based on the successive convex approximate (SCA) method.

A. Equivalent reformulation

Concerning the complex expression of $R_{\mathrm{p},k}$ and $R_{\mathrm{c},k}$, we introduce auxiliary variables $\gamma_{\mathrm{p},k}$, $\forall k\in\mathcal{K}$ and $\gamma_{\mathrm{c},k}$, $\forall k\in\mathcal{K}$ and s. The equivalent problem of (9) can be obtained as

$$\max_{\substack{\{\mathbf{w}_i\}_{i\in\mathcal{M}},s,\\ \{\gamma_{\mathrm{p},k},\gamma_{\mathrm{c},k},r_{\mathrm{c},k}\}_{k\in\mathcal{K}}}} \quad s \tag{12a}$$

$$\text{s.t.} \quad r_{\mathrm{c},k} + \log_2(1+\gamma_{\mathrm{p},k}) \geq s, \forall k\in\mathcal{K} \tag{12b}$$

$$\gamma_{\mathrm{p},k} \leq \frac{|\mathbf{h}_k^H\mathbf{w}_k|^2}{\sum_{i\in\mathcal{K},i\neq k}|\mathbf{h}_k^H\mathbf{w}_i|^2+\sigma_0^2}, \forall k\in\mathcal{K} \tag{12c}$$

$$\sum_{k\in\mathcal{K}} r_{\mathrm{c},k} \leq \log_2(1+\gamma_{\mathrm{c},k}), \forall k\in\mathcal{K} \tag{12d}$$

$$\gamma_{\mathrm{c},k} \leq \frac{|\mathbf{h}_k^H\mathbf{w}_\mathrm{c}|^2}{\sum_{i\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{w}_i|^2+\sigma_0^2}, \forall k\in\mathcal{K} \tag{12e}$$

$$(9b), (9d).$$

B. SCA-based reformulation

To facilitate the implementation of successive convex approximation (SCA), we propose to replace the the non-convex part in constraints (12c) and (12e) with their corresponding lower-bound surrogate functions. In particular, we introduce new auxiliary variables p_k, c_k and denote

$$I_k^{\mathrm{p}} = \sum_{i\in\mathcal{K},i\neq k}|\mathbf{h}_k^H\mathbf{w}_i|^2+\sigma_0^2, \forall k\in\mathcal{K} \tag{13}$$

$$I_k^{\mathrm{c}} = \sum_{i\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{w}_i|^2+\sigma_0^2, \forall k\in\mathcal{K}. \tag{14}$$

Then, we have

$$(12c) \iff \gamma_{\mathrm{p},k} \leq \frac{|\mathbf{h}_k^H\mathbf{w}_k|^2}{I_k^{\mathrm{p}}} \tag{15}$$

$$(12e) \iff \gamma_{\mathrm{c},k} \leq \frac{|\mathbf{h}_k^H\mathbf{w}_\mathrm{c}|^2}{I_k^{\mathrm{c}}}. \tag{16}$$

Now we use the following inequality to approximate the quadratic-over-linear part of the form x^2/y $(y>0)$ via its first order approximations as [17]

$$\frac{x^2}{y} \geq \frac{2x^t}{y^t}x - \left(\frac{x^t}{y^t}\right)^2 y, \tag{17}$$

where t is the iteration number and (x^t, y^t) are the values of (x,y) in iteration t.

According to (17), the right-hand-sides of (12c) and (12e) can be approximated via a lower bound linear function in each iteration, i.e.,

$$\gamma_{\mathrm{p},k} \leq \frac{2\mathrm{Re}(p_k^t p_k)}{I_k^{\mathrm{p},t}} - \left(\frac{p_k^t}{I_k^{\mathrm{p},t}}\right)^2 I_k^{\mathrm{p}}, \forall k\in\mathcal{K} \tag{18a}$$

$$\gamma_{\mathrm{c},k} \leq \frac{2\mathrm{Re}(c_k^t c_k)}{I_k^{\mathrm{c},t}} - \left(\frac{c_k^t}{I_k^{\mathrm{c},t}}\right)^2 I_k^{\mathrm{c}}, \forall k\in\mathcal{K}, \tag{18b}$$

where $p_k^t = \mathbf{h}_k^H\mathbf{w}_k^t$, $p_k = \mathbf{h}_k^H\mathbf{w}_k$, $c_k^t = \mathbf{h}_k^H\mathbf{w}_\mathrm{c}^t$ and $c_k = \mathbf{h}_k^H\mathbf{w}_\mathrm{c}$. Here $\mathbf{w}_k^t$ and $\mathbf{w}_\mathrm{c}^t$ are the obtained solution in the last iteration. The optimization problem can then be reformulated as the following convex optimization problem.

$$\max_{\substack{\{\mathbf{w}_i\}_{i\in\mathcal{M}},s,\\ \{\gamma_{\mathrm{p},k},\gamma_{\mathrm{c},k}\}_{k\in\mathcal{K}}}} \quad s \tag{19a}$$

$$\text{s.t.} \quad (12b), (12d), (18a), (18b), (9b), (9d).$$

This way, Problem (19) becomes a convex optimization problem. In each iteration of the SCA algorithm, we solve Problem (19) with convex optimization tools such as CVX, and iteratively update the parameters until convergence is reached. This algorithm is guaranteed to converge, since in each iteration, the bounds on the right-hand-sides of (12c) and (12e) become tighter and the objective function value keeps increasing. Due to the power constraint, the objective function value will converge to a stationary point of problem (9). The SCA-based algorithm is stated in Algorithm 1.

C. Complexity analysis

The complexity of the proposed algorithm comes from solving Problem (19) since all the other steps are just parameter updating operations. Note that Problem (19) belongs to second order cone programming (SOCP) problems. According to [11],

Algorithm 1 SCA-based Beamfoming

Input: $\mathbf{h}_k, P_{\max}$
Output: $\mathbf{w}_i^*$, $\forall i \in \mathcal{M}$ and $r_{\mathrm{c},k}$.

1: Initialize iteration index $t = 0$, generate the initial beamforming vector $\mathbf{w}_i^0$, $\forall i \in \mathcal{M}$, p_k^t , c_k^t, and obtain I_{p}^t and I_{c}^t. Set the convergence tolerance ϵ;
2: **while** $\frac{|s^t - s^{t-1}|}{|s^{t-1}|} \leq \epsilon$ **do**
3: Obtain the solution $r_{\mathrm{c},k}$ and $\{\mathbf{w}_i^t\}$ by solving Problem (19) under given $\{p_k^{t-1}, c_k^{t-1}, I_k^{\mathrm{p},t-1}, I_k^{\mathrm{c},t-1}\}$;
4: Update parameters $\{p_k^t, c_k^t, I_k^{\mathrm{p},t}, I_k^{\mathrm{c},t}\}$ from $\{\mathbf{w}_i^t\}$ and compute the objective value s^t;
5: $t = t + 1$;
6: **end while**

[18], the complexity will be $\mathcal{O}([KM]^{3.5})$ using general CVX solvers, such as MOSEK. Suppose the number of iterations needed for Algorithm 1 to converge is I. Then the overall complexity of Algorithm 1 will be $\mathcal{O}(I[KM]^{3.5})$).

IV. Imperfect CSI Case

In this section, we aim to solve Problem (11) under imperfect CSI. The challenge comes not only from the non-convexity nature of the constraints, but also from the stochastic nature of the constraints (i.e., the expectation operation in the constraints). Unlike the perfect CSI case where the problem is deterministic, with imperfect CSI, it is much more challenging to find an stationary point due to the randomness of channel realizations. In this section, we proposed a stochastic successive convex approximation (SSCA) method to find a feasible solution to Problem (11).

First of all, we reformulate the stochastic version of our problem in a similar manner as we did in Section III.

$$\max_{\substack{\{\mathbf{w}_i\}_{i\in\mathcal{M}}, s,\\ \{\gamma_{\mathrm{p},k}, \gamma_{\mathrm{c},k}, r_{\mathrm{c},k}\}_{k\in\mathcal{K}}}} \quad s \tag{20a}$$

$$\text{s.t.} \quad s - r_{\mathrm{c},k} - \log_2(1 + \gamma_{\mathrm{p},k}) \leq 0, \forall k \in \mathcal{K} \tag{20b}$$

$$\gamma_{\mathrm{p},k} - \mathbb{E}\left[\frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{i\in\mathcal{K}, i\neq k} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_0^2}\right] \leq 0, \forall k \in \mathcal{K} \tag{20c}$$

$$\sum_{k\in\mathcal{K}} r_{\mathrm{c},k} - \log_2(1 + \gamma_{\mathrm{c},k}) \leq 0, \forall k \in \mathcal{K} \tag{20d}$$

$$\gamma_{\mathrm{c},k} - \mathbb{E}\left[\frac{|\mathbf{h}_k^H \mathbf{w}_{\mathrm{c}}|^2}{\sum_{i\in\mathcal{K}} |\mathbf{h}_k^H \mathbf{w}_i|^2 + \sigma_0^2}\right] \leq 0, \forall k \in \mathcal{K} \tag{20e}$$

$$(9b), (9d).$$

where the expectation is taken with respect to all the channel samples.

Suppose the constraints of our problem (20b)–(20e) and (9b), (9d) are denoted as $f_i(\mathbf{w}) \leq 0, i = 1, 2, ..., 6$, where $\mathbf{w} = \{\mathbf{w}_i\}_{i\in\mathcal{M}}$. The idea is to replace the constraint functions $f_i(\mathbf{w})$ of Problem (11) with their convex surrogate functions $\bar{f}_i(\mathbf{w}), i = 1, 2, ..., 6$. If the constraint is already convex, we do not need change it further. Specifically, at time slot n, we obtain a new channel realization $\mathcal{H}^n$. The surrogate function $\bar{f}_i(\mathbf{w})$ is updated based on $\mathbf{w}^{n-1}$ and $\mathcal{H}^n$. To guarantee the convergence of the algorithm, the surrogate function has to be properly designed. Based on the surrogate function, an optimal solution $\overline{\mathbf{w}}^n$ is obtained by solving the relaxed problem with the surrogate constraint. Then $\mathbf{w}$ is updated according to

$$\mathbf{w}^{n+1} = (1 - \gamma^n)\mathbf{w}^n + \gamma^n \overline{\mathbf{w}}^n, \tag{21}$$

where γ^n is a decreasing sequence satisfying $\gamma^n \to 0$, $\sum_n \gamma^n = \infty$, and $\sum_n (\gamma^n)^2 \leq \infty$.

When a new channel estimate $\mathcal{H}^n$ is obtained, a surrogate function approximating constraint $f_i(\mathbf{w})$ is constructed as

$$\bar{f}_i^n(\mathbf{w}) = (1 - \gamma^n)\bar{f}_i^{n-1}(\mathbf{w}) + \gamma^n \hat{f}_i^n(\mathbf{w}; \mathcal{H}^n), \tag{22}$$

where $\bar{f}_i^0 = 0$ and the function $\hat{f}_i(\mathbf{w}; \mathcal{H}^n)$ is given by

$$\begin{aligned} \hat{f}_i(\mathbf{w}; \mathcal{H}^n) = f_i(\mathbf{w}; \mathcal{H}^n) + \mathrm{Re}\left(\frac{\partial f_i}{\partial \mathbf{w}}^T (\mathbf{w} - \mathbf{w}^{n-1})\right) \\ + \frac{\tau_i}{2}||\mathbf{w} - \mathbf{w}^{n-1}||_2^2, \end{aligned} \tag{23}$$

where τ_i, $\forall i$ is any positive constant to ensure convergence.

In Problem (20), constraints (20b), (12d), (9b), and (9d) are all convex. We only need to deal with the non-convex constraints (12c) and (12e) by finding their convex surrogate function. Fortunately, we have already construct a surrogate function in (18).

The optimal beamforming vector is obtained by solving the following problem.

$$\begin{aligned} \overline{\mathbf{w}}^n = \ & \arg\max_{\mathbf{w}} \quad s \\ & \text{s.t.} \quad \bar{f}_i(\mathbf{w}; \mathcal{H}^n) \leq 0, \forall i. \end{aligned} \tag{24}$$

This stochastic problem is solved in an online fashion by repeatedly solving the constructed surrogate problem. The surrogate constraint function is constantly rectified by real-time samples and hence will approximate the original expectations in the stochastic problem. We leave the convergence proof in our future work. The SSCA-based online algorithm is given in Algorithm 2.

Algorithm 2 SSCA-based Online Beamfoming

Input: $\{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_N\}, P_{\max}$
Output: $\mathbf{w}_i^*$, $\forall i \in \mathcal{M}$ and $r_{\mathrm{c},k}$.

1: Initialize time index $n = 1$ and generate the initial beamforming vector $\mathbf{w}_i^0$, $\forall i \in \mathcal{M}$ and $\{\gamma^n\}$;
2: **while** $n \leq N$ **do**
3: A channel observation $\mathcal{H}^n$ is obtained;
4: Solve (24) and obtain solution $\overline{\mathbf{w}}^n$;
5: Obtain the common rate allocation $\{r_{\mathrm{c},k}\}$ based on $\overline{\mathbf{w}}^n$ and $\mathcal{H}^n$, and compute the MMF rate s^n;
6: Update the beamforming vector $\mathbf{w}^n$ according to (21);
7: $n = n + 1$;
8: **end while**

V. Simulation Results

In this section, we present our extensive simulation results to evaluate the effectiveness of the proposed method. We assume that the channel follows a Rayleigh fading with zero mean and unit variance. The channel estimation error also follows a Rayleigh distribution with zero mean and variance $\sigma_e^2 = 0.02$ mW. The noise power is set to $\sigma^2 = 0.01$mW. For the implementation of the proposed SSCA in the imperfect CSI case, the parameters are chose as $\gamma^n = \frac{1}{(1+n)^{0.9}}$.

We compare the performance of the proposed method with the WMMSE method used in [8], [19] in terms of the max-min fairness performance. Note that the sum rate performance can be easily obtained by slightly modifying the constraints and objective function of the proposed framework. The results are averaged over 100 channel realizations. The threshold value ϵ for both the (S)SCA and WMMSE algorithms is set to 10^{-3}. For the imperfect CSI case, we compare the proposed SSCA algorithm with the sample average approximation (SAA) based WMMSE algorithm proposed in [6].

A. Perfect CSI case

First of all, we investigate the MMF performance in the perfect CSI case. We compare our rate splitting scheme with the spectrum division multiple access (SDMA) scheme where all the signals from other UEs are treated as interference. Note that SDMA is a special case of our rate splitting scheme by allocating zero power to the common beamforming vector. Hence the proposed SCA-based algorithm can be easily adapted to deal with the SDMA case. Meanwhile, the problem can also be solved via the WMMSE approach proposed in [8]. Note that the WMMSE algorithm leverages the relationship between rate and MMSE to find a suboptimal beamforming vector in an iterative way. We compare the performance of the proposed SCA algorithm with the conventional WMMSE algorithm in terms of two schemes, RS and SDMA.

1) Convergence performance: The convergence performance comparison is shown in Fig. 1. It can be seen that the proposed SCA based algorithm converges in about 10 iterations while the WMMSE based algorithm converges in about 15-20 iterations for rate splitting (RS) access. Moreover, the SCA-based scheme achieves a slightly higher MMF rate than the MMSE-based scheme for RS access. The SDMA scheme also converges quickly, however, the achieved MMF rate is less than half of that of the RS scheme. It can be seen that the RS scheme generally achieves a better MMF performance especially when the number of UEs is larger than the number of transmit antennas. This confirms the degree analysis in [19]. It is shown that the RS approach achieves a multiplexing gain of $\frac{1}{1+K-M}$ when $K \leq M$ in terms of max-min fairness while SDMA achieves zero gain in this case.

2) MMF performance: We then compare the rate-splitting scheme with the benchmark scheme SDMA by varying the total transmit power budget. As shown in Fig. 2, the MMF rate of the RS scheme increases with the power budget smoothly while the MMF rate performance of the SDMA scheme almost stays in a horizontal line. Meanwhile, the proposed SCA-based algorithm outperforms the WMMSE based approach, especially when the power budget becomes larger.

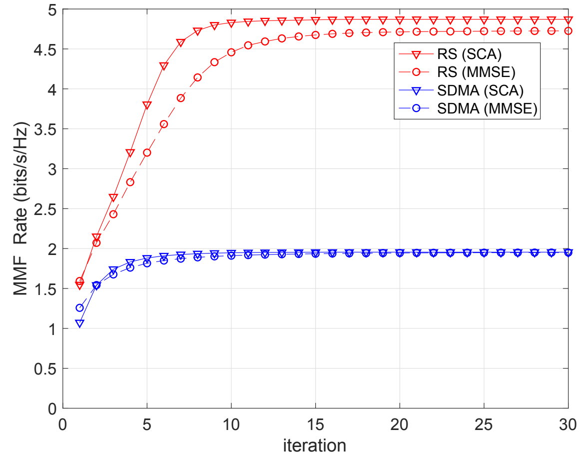

Fig. 1: A rate splitting MISO system with RS ($M = 3, K = 4$): convergence performance.

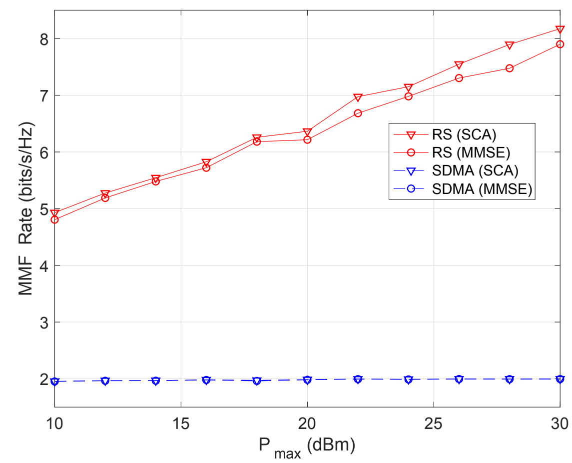

Fig. 2: A rate splitting MISO system with RS ($M = 3, K = 4$): MMF performance.

B. Imperfect CSI case

For the imperfect CSI case, we compare the performance of the proposed SSCA algorithm with the sample average approximate (SAA) based WMMSE algorithm for the RS scheme. We perform 50 channel estimations and the channel estimation error follows an i.i.d. Rayleigh distribution. The ergodic MMF rate of the proposed SSCA algorithm is compared with that of the SAA based algorithm in Table I. It can be seen that the SAA based algorithm can achieve an average of 4.4418 bits/s/Hz in terms of the MMF rate performance, which is slightly better than the proposed SSCA based approach. However, the proposed SSCA algorithm runs much faster than the SAA based algorithm. As shown in Fig. 3. The SAA-based algorithm requires on average 15-20 CPU time while the SSCA algorithm only requires 0-5 CPU time.

This is because the proposed SSCA algorithm is an *online* algorithm. It only requires the currently estimated channel state and the beamforming vector is updated in an online fashion.

TABLE I: Average MMF Performance Comparison

Method	Average MMF rate (bits/s/Hz)
WMMSE (SAA)	4.4418
SSCA	4.2725

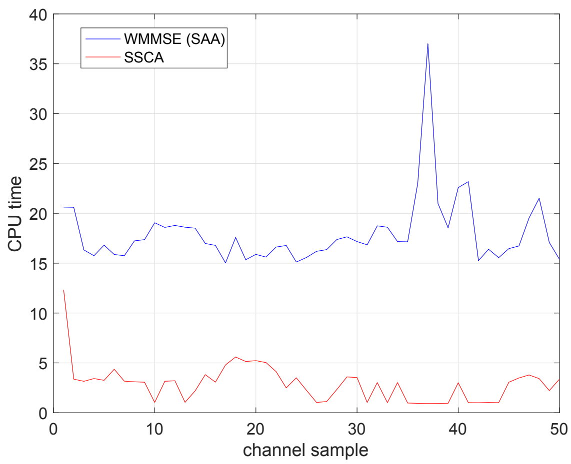

Fig. 3: A rate splitting MISO system with RS ($M = 3, K = 4$ and $P = 10$dBm): CPU time.

In contrast, the SAA algorithm is an *offline* method, which requires a channel sample collection phase to obtain a sufficiently large number of channel samples before calculating the optimal beamforming vector. Although for each sample channel collection, the SAA algorithm only needs to run once to find the beamforming vector, it is not adaptive when new channel samples are observed. The proposed SSCA algorithm does not need to collect a large number of samples for the random state before the stochastic optimization, hence it is much more flexible and faster. When the dimension of the channel or the size of collected channel samples becomes large, the proposed online algorithm can save more time and greatly reduces the complexity.

VI. Conclusions

In this paper, we studied the MMF oriented beamforming design in a downlink MIMO system with rate splitting multiple access. The formulated problem was non-convex and an SCA based algorithm was developed to find a KKT point of the considered problem. Simulation results showed that the proposed algorithm converged faster and achieved a better MMF rate performance than the conventional WMMSE approach. Compared with the SDMA scheme, the RS scheme could provide better MMF rates, which demonstrated that RS was a promising access technology for next-generation wireless networks. Moreover, to tackle the channel estimation error, we also formulated a stochastic beamforming design problem. An SSCA algorithm was proposed to obtain an effective solution in an online fashion. Compared with conventional algorithms, it does not require channel sample collection phase and can achieve a satisfactory performance with a lower complexity. We leave the extension to large-scale networks and the relevant convergence analysis to our future work.

Acknowledgments

This work is supported in part by the US NSF under Grants ECCS-1923717 and CNS-1822055.

References

[1] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, “Rate-splitting multiple access: A new frontier for the PHY layer of 6G,” arXiv preprint arXiv:2006.01437, June 2020. [Online]. Available: https://arxiv.org/abs/2006.01437

[2] Y. Mao, B. Clerckx, and V. O. Li, “Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming sdma and noma,” *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–54, 2018.

[3] T. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 49–60, 1981.

[4] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[5] C. Hao, Y. Wu, and B. Clerckx, “Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach,” *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, Sept. 2015.

[6] H. Joudeh and B. Clerckx, “Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach,” *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

[7] H. Joudeh and B. Clerckx, “Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7276–7289, Nov. 2017.

[8] H. Joudeh and B. Clerckx, “Robust transmission in downlink multiuser MISO systems: A rate-splitting approach,” *IEEE Trans. Sig. Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[9] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, “Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, Mar. 2019.

[10] H. Fu, S. Feng, W. Tang, and D. W. K. Ng, “Robust secure beamforming design for two-user downlink MISO rate-splitting systems,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8351–8365, Dec. 2020.

[11] J. Zhang, J. Zhang, Y. Zhou, H. Ji, J. Sun, and N. Al-Dhahir, “Energy and spectral efficiency tradeoff via rate splitting and common beamforming coordination in multicell networks,” *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7719–7731, Dec. 2020.

[12] Z. Yang, J. Shi, Z. Li, M. Chen, W. Xu, and M. Shikh-Bahaei, “Energy efficient rate splitting multiple access (RSMA) with reconfigurable intelligent surface,” in *Proc. ICC’20 Workshops*, Virtual Conference, June 2020, pp. 1–6.

[13] Y. Mao, B. Clerckx, and V. O. Li, “Energy efficiency of rate-splitting multiple access, and performance benefits over sdma and noma,” in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2018, pp. 1–5.

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

[15] D. Guo, S. Shamai, and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[16] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[17] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[18] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA: SIAM, 2001.

[19] B. Clerckx, Y. Mao, R. Schober, E. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, “Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques,” arXiv preprint arXiv:2101.04802, Jan. 2021.