GLOBALLY OPTIMAL BEAMFORMING FOR RATE SPLITTING MULTIPLE ACCESS

Bho Matthiesen[⋆], *Yijie Mao*[†], *Petar Popovski*[‡⋆], *and Bruno Clerckx*[†]

[⋆] University of Bremen, Deptartment of Communications Engineering, Otto-Hahn-Allee 1, 28359 Bremen, Germany
[†] Imperial College London, Deptartment of Electical and Electronic Engineering, London, United Kingdom
[‡] Aalborg University, Department of Electronic Systems, 9220 Aalborg, Denmark

ABSTRACT

We consider globally optimal precoder design for rate splitting multiple access in Gaussian multiple-input single-output downlink channels with respect to weighted sum rate and energy efficiency maximization. The proposed algorithm solves an instance of the joint multicast and unicast beamforming problem and includes multicast- and unicast-only beamforming as special cases. Numerical results show that it outperforms state-of-the-art algorithms in terms of numerical stability and converges almost twice as fast.

Index Terms— rate splitting, global optimization, resource allocation, energy efficiency, interference networks

1. INTRODUCTION

Rate splitting multiple access (RSMA) is a powerful non-orthogonal transmission and robust interference management strategy for beyond 5G communication networks [1–3]. The key idea is to split each message into common and private parts and transmit them by superposition coding [4]. The common message is decoded by multiple users, while the private message is only decoded by the corresponding user employing successive interference cancellation (SIC). This approach allows arbitrary combinations of joint decoding and treating interference as noise by flexibly adjusting the message split. Recent results show that RSMA outperforms existing multiple access schemes such as space division multiple access, power-domain non-orthogonal multiple access, orthogonal multiple access, and multicasting in terms of weighted sum rate (WSR) [2, 5, 6] and energy efficiency (EE) [6, 7].

This paper treats the important question of downlink multiple-input single-output (MISO) beamforming for RSMA with respect to WSR and EE maximization. The corresponding optimization problem is related to joint multicast and unicast precoding that is known to be NP-hard [8, 9]. Existing works on RSMA focus on suboptimal strategies to obtain computationally tractable algorithms [2, 6, 7, 10–13]. While several globally optimal algorithms for unicast beamforming [14, 15] and multicast beamforming [16] exist, joint solution methods are scarce. In particular, the procedure in [17] solves the power minimization problem and [18] maximizes the WSR for joint multicast and unicast beamforming. All these methods are based on branch and bound (BB) in combination with the second-order cone (SOC) transformation in [19]. However, as this transformation moves the complexity into the feasible set, pure BB methods are prone to numerical problems, see Section 3. Instead, in this paper we design a successive incumbent transcending (SIT) BB algorithm to solve this beamforming problem with improved numerical stability and faster convergence. To the best of the authors knowledge, this is the first globally optimal solution algorithm for an instance of the joint unicast and multicast problem with respect to EE maximization. It is also the first global optimization method specifically targeted at RSMA.

This work is supported in part by the German Research Foundation (DFG) under grant EXC 2077 (University Allowance), by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under grants EP/N015312/1 and EP/R511547/1, and by the North-German Supercomputing Alliance (HLRN).

2. SYSTEM MODEL & PROBLEM STATEMENT

Consider the downlink in a wireless network where an M antenna base station (BS) serves K single-antenna users. The received signal at user k, $k \in \mathcal{K} = \{1, \dots, K\}$, for each channel use is $y_k = \boldsymbol{h}_k^H \boldsymbol{x} + n_k$, where the transmit signal $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$ is subject to an average power constraint P, $\boldsymbol{h}_k$ is the complex-valued channel from the BS to user k, and n_k is circularly symmetric complex white Gaussian noise with unit power at user k.

The transmitter employs 1-layer rate splitting [2,10], i.e., it splits the message W_k intended for user k into a common part $W_{c,k}$ and a private part $W_{p,k}$. Then, the common messages are combined into a single message W_c and these $K + 1$ messages are encoded with independent Gaussian codebooks into $s_c, s_1, \dots, s_K$, each having unit power. These symbols are combined with linear precoding into the transmit signal $\boldsymbol{x} = \boldsymbol{p}_c s_c + \sum_{k\in\mathcal{K}} \boldsymbol{p}_k s_k$. The BS is subject to an average power constraint, i.e., $\|\boldsymbol{p}_c\|^2 + \sum_{k\in\mathcal{K}} \|\boldsymbol{p}_k\|^2 \leq P$.

Each receiver uses SIC to first recover s_c and then s_k, treating all other messages as noise. Asymptotic error free decoding of W_c and $W_{p,k}$ is possible if the rates of these messages satisfy $R_c \leq \log(1 + \gamma_{c,k})$ and $R_{p,k} \leq \log(1 + \gamma_{p,k})$, with signal to interference plus noise ratios (SINRs)

$$\gamma_{c,k} = \frac{|\boldsymbol{h}_k^H \boldsymbol{p}_c|^2}{\sum_{j\in\mathcal{K}} |\boldsymbol{h}_k^H \boldsymbol{p}_j|^2 + 1}, \quad \gamma_{p,k} = \frac{|\boldsymbol{h}_k^H \boldsymbol{p}_k|^2}{\sum_{j\in\mathcal{K}\setminus k} |\boldsymbol{h}_k^H \boldsymbol{p}_j|^2 + 1}. \tag{1}$$

The rate R_c is shared across the users, where user k is allocated a portion C_k corresponding to the rate of $W_{c,k}$ such that $\sum_{k\in\mathcal{K}} C_k = R_c$. Then, the total rate of user k is $R_k = C_k + R_{p,k}$.

Observe that this system model includes multi-user linear precoding and multicast beamforming as special cases.

2.1. Problem Statement

We consider the following resource allocation problem under minimum rate R_k^{th} quality of service constraints

$$\max_{\substack{\boldsymbol{p}_1,\dots,\boldsymbol{p}_K,\\ \boldsymbol{p}_c, \mathbf{c}, \boldsymbol{\gamma}_c, \boldsymbol{\gamma}_p}} \quad \frac{\sum_{k\in\mathcal{K}} u_k \left(C_k + \log(1 + \gamma_{p,k})\right)}{\mu\left(\|\boldsymbol{p}_c\|^2 + \sum_{k\in\mathcal{K}} \|\boldsymbol{p}_k\|^2\right) + P_c} \tag{2a}$$

$$\text{s.t.} \quad \gamma_{c,k} \text{ and } \gamma_{p,k} \text{ as in (1)} \tag{2b}$$

$$\sum\nolimits_{k'\in\mathcal{K}} C_{k'} \leq \log(1 + \gamma_{c,k}), \forall k \in \mathcal{K} \tag{2c}$$

$$C_k \geq \max\left\{0,\, R_k^{th} - \log(1+\gamma_{p,k})\right\}, \forall k \in \mathcal{K} \quad \text{(2d)}$$

$$\|\boldsymbol{p}_c\|^2 + \sum\nolimits_{k\in\mathcal{K}} \|\boldsymbol{p}_k\|^2 \leq P \quad \text{(2e)}$$

with nonnegative weight vector $\mathbf{u} = [u_1, \ldots, u_K] \neq \mathbf{0}$, nonnegative power amplifier inefficiency μ, and positive static circuit power consumption P_c. This problem has two operational meanings: With unit weights, it maximizes the EE and, with $\mu = 0$, $P_c = 1$, it maximizes the WSR.

The following problem is equivalent to (2) and will be solved by the developed algorithm:

$$\max_{\substack{\boldsymbol{p}_c,\boldsymbol{p}_1,\ldots,\boldsymbol{p}_K,\\ \mathbf{c},\boldsymbol{\gamma}_p,s,\boldsymbol{d},\boldsymbol{e}}} \frac{\sum_{k\in\mathcal{K}} u_k\left(C_k + \log(1+\gamma_{p,k})\right)}{\mu\left(\|\boldsymbol{p}_c\|^2 + \sum_{k\in\mathcal{K}}\|\boldsymbol{p}_k\|^2\right) + P_c} \quad \text{(3a)}$$

$$\text{s.t.} \quad \sqrt{\gamma_{p,k}}\left(\sum\nolimits_{j\in\mathcal{K}\setminus k} |\mathbf{h}_k^H \mathbf{p}_j|^2 + 1\right)^{1/2} \leq \mathbf{h}_k^H \mathbf{p}_k \quad \text{(3b)}$$

$$\sqrt{s}\left(\sum\nolimits_{j\in\mathcal{K}} |\mathbf{h}_1^H \mathbf{p}_j|^2 + 1\right)^{1/2} \leq \mathbf{h}_1^H \mathbf{p}_c \quad \text{(3c)}$$

$$\sqrt{s}\left(\sum\nolimits_{j\in\mathcal{K}} |\mathbf{h}_k^H \mathbf{p}_j|^2 + 1\right)^{1/2} \leq d_k, \forall k > 1 \quad \text{(3d)}$$

$$(e_k, d_k) \in \mathcal{C}, \forall k > 1 \quad \text{(3e)}$$

$$\Re\{\boldsymbol{h}_k^H \boldsymbol{p}_k\} \geq 0, \quad \Im\{\boldsymbol{h}_k^H \boldsymbol{p}_k\} = 0 \quad \text{(3f)}$$

$$\Re\{\boldsymbol{h}_1^H \boldsymbol{p}_c\} \geq 0, \quad \Im\{\boldsymbol{h}_1^H \boldsymbol{p}_c\} = 0 \quad \text{(3g)}$$

$$\forall k > 1: d_k \geq 0,\ e_k = \mathbf{h}_k^H \mathbf{p}_c \quad \text{(3h)}$$

$$\sum\nolimits_{k\in\mathcal{K}} C_k \leq \log(1+s) \quad \text{(3i)}$$

$$\text{(2d) and (2e)} \quad \text{(3j)}$$

with $$(e, d) \in \mathcal{C} = \{e \in \mathbb{C}, d \in \mathbb{R} : d \leq |e|\}. \quad (4)$$

A crucial observation is that this problem is a second-order cone program (SOCP) for fixed s, $\boldsymbol{\gamma}_p$, except for constraint (3h). Hence, the nonconvexity of (2) is only due to the SINR expressions and not due to the beamforming vectors. We will exploit this partial convexity in the final algorithm to limit the numerical complexity.

Proposition 1. *Problems* (2) *and* (3) *have the same optimal value and every solution of* (3) *also solves* (2).

Proof. Omitted due to space constraints. Use the SOC reformulation from [19] for the SINRs, with additional auxiliary variables for the multicast beamformer $\boldsymbol{p}_c$ [16]. □

3. GLOBALLY OPTIMAL BEAMFORMING

Problem (3) is an NP-hard nonconvex optimization problem due to the multicast beamforming [8] and the power allocation in the private messages [9]. Previous global optimization algorithms for similar problems rely on BB procedures with SOCP bounding [14, 15, 17, 18]. However, this either leads to an infinite algorithm or requires the additional solution of several SOCPs to obtain a feasible point in each iteration [14] which is required to obtain a finite algorithm. Moreover, the auxiliary SOCP that is solved in every iteration of the BB procedure is numerically challenging and leads to problems even with commercial state-of-the-art solvers like Mosek [20]. This can be alleviated by the modified auxiliary problem in [14, §2.2.2] but this approach greatly increases convergence times. Instead, we design an algorithm based on the SIT scheme [21–24] and combine it with a branch reduce and bound (BRB) procedure. The resulting algorithm is numerically stable, has proven finite convergence, also solves EE maximization, and is the first global optimization algorithm specifically designed for RSMA. Practically, it outperforms algorithms for similar problems as will be verified in Section 4.

To better illustrate the core principles of SIT, consider the general optimization problem

$$\max\nolimits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{D}} \ f(\boldsymbol{x},\boldsymbol{\xi}) \quad \text{s.t.} \quad g_i(\boldsymbol{x},\boldsymbol{\xi}) \leq 0,\ i = 1,\ldots,n \quad (5)$$

with continuous, real valued functions $f, g_1, \ldots, g_n$ and nonempty feasible set. Further, assume that f is concave,[1] $g_1, \ldots, g_n$ are convex in $\boldsymbol{\xi}$ for fixed $\boldsymbol{x}$, and $\mathcal{D}$ is a closed convex set. Depending on the structure of $g_1, \ldots, g_n$ in $\boldsymbol{x}$, this problem might be quite hard to solve for BB methods [23, 25].[2] Instead, consider the problem

$$\min\nolimits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{D}} \ \max\nolimits_i\{g_i(\boldsymbol{x},\boldsymbol{\xi})\} \quad \text{s.t.} \quad f(\boldsymbol{x},\boldsymbol{\xi}) \geq \delta \quad (6)$$

that is obtained from (5) by exchanging the objective and constraints. If the optimal value of (6) is less than or equal to zero, the optimal value of (5) is greater than or equal to δ. Instead, if the optimal value of (6) is greater than zero, the optimal value of (5) is less than δ [22, Prop. 7]. Hence, the optimal solution of (5) can be obtained by solving a sequence of (6) with increasing δ. Since the feasible set of (6) is closed and convex, it can be solved much easier by BRB than (5) [22].

The SIT and BRB procedures can be integrated into a single BRB algorithm that solves (6) with low precision and updates δ whenever a point $\boldsymbol{x}^k$ feasible in (5) is encountered that achieves an objective value $f(\boldsymbol{x}^k) > \delta$. This BRB procedure relaxes the feasible set and subsequently partitions it in such a way that upper and lower bounds on the minimum objective value of (6) can be computed efficiently for each partition element. In particular, we use rectangular subdivision and define the initial box as $\mathcal{M}_0 = [\boldsymbol{r}^0, \boldsymbol{s}^0] = \{\boldsymbol{x} : r_i^0 \leq x_i \leq s_i^0\}$ satisfying $\mathcal{M}_0 \supseteq \operatorname{proj}_{\boldsymbol{x}} \mathcal{D}$. The algorithm subsequently partitions the relaxed feasible set $\mathcal{M}_0$ into smaller boxes and stores the current partition of $\mathcal{M}_0$ in $\mathscr{R}_k$. In iteration k, the algorithm selects a box $\mathcal{M}^k = [\boldsymbol{r}^k, \boldsymbol{s}^k]$ and bisects it into two new subrectangles. For each of these new boxes, a lower bound on the objective value is computed using a bounding function $\beta(\mathcal{M})$ that computes a lower bound on the objective value of (6) with additional constraint $\boldsymbol{x} \in \mathcal{M}$. If this problem is infeasible, then $\beta(\mathcal{M}) = \infty$. To ensure convergence, the bounding needs to be consistent with branching, i.e., $\beta(\mathcal{M})$ has to satisfy

$$\beta(\mathcal{M}) - \min_{\substack{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{F},\\ \boldsymbol{x}\in\mathcal{M}}} \max_i \{g_i(\boldsymbol{x},\boldsymbol{\xi})\} \to 0 \ \text{as} \ \max_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{M}} \|\boldsymbol{x}-\boldsymbol{y}\| \to 0, \quad (7)$$

and a dual feasible point $\boldsymbol{x}^k \in \operatorname{proj}_{\boldsymbol{x}} \mathcal{F} \cap \mathcal{M}_k$ is required, where $\mathcal{F} = \{\boldsymbol{x} \in \mathcal{D} : f(\boldsymbol{x}) \geq \delta\}$ is the feasible set of (6).

The following lemma is essential to establish the convergence of the SIT procedure. It follows that it can be incorporated in a BB procedure with pruning criterion $\beta(\mathcal{M}) < -\varepsilon$ and termination criterion $0 > \min_{\boldsymbol{\xi}} \ g(\boldsymbol{x}^k, \boldsymbol{\xi}) \ \text{s.t.} \ (\boldsymbol{x}^k, \boldsymbol{\xi}) \in \mathcal{F}$.

Lemma 1 ([24, Prop. 5.9])**.** *Let $\varepsilon > 0$ be given and define $g(\boldsymbol{x},\boldsymbol{\xi}) = \max_i\{g_i(\boldsymbol{x},\boldsymbol{\xi})\}$. Either $g(\boldsymbol{x}^k,\boldsymbol{\xi}^*) < 0$ for some k and $(\boldsymbol{x}^k,\boldsymbol{\xi}) \in \mathcal{F}$, or $\beta(\mathcal{M}_k) > -\varepsilon$ for some k. In the former case, $(\boldsymbol{x}^k,\boldsymbol{\xi}^*)$ is a nonisolated feasible solution of* (5) *satisfying $f(\boldsymbol{x}^k,\boldsymbol{\xi}^*) \geq \delta$. In the latter case, no ε-essential feasible solution $(\boldsymbol{x},\boldsymbol{\xi})$ of* (5) *exists such that $f(\boldsymbol{x},\boldsymbol{\xi}) \geq \delta$.*

Next, we design a suitable bounding procedure that satisfies (7).

3.1. Bounding Procedure

The SIT dual should contain all of the problem's nonconvexity in the objective function. Following the discussion in Section 2.1, the

[1] Although this assumption does not hold for (3), the approach is still applicable since the sole purpose of this assumption is to obtain a convex feasible set in (6).

[2] This is also true for outer approximation methods [25].

nonconvexity in (3) is due to (3b)–(3e). We obtain the SIT dual as

$$\min_{\substack{\boldsymbol{p}_c,\boldsymbol{p}_1,\dots,\boldsymbol{p}_K,\\ \mathbf{c},\boldsymbol{\gamma}_p,s,\boldsymbol{d},\boldsymbol{e}}} \quad \max\Big[\sqrt{s}\left(\sum\nolimits_{j\in\mathcal{K}}|\mathbf{h}_1^H\mathbf{p}_j|^2+1\right)^{1/2}-\mathbf{h}_1^H\mathbf{p}_c,$$
$$\max_{k>1}\left\{\sqrt{s}\left(\sum\nolimits_{j\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{p}_j|^2+1\right)^{1/2}-d_k\right\},$$
$$\max_{k\in\mathcal{K}}\left\{\sqrt{\gamma_{p,k}}\left(\sum\nolimits_{j\in\mathcal{K}\setminus k}|\mathbf{h}_k^H\mathbf{p}_j|^2+1\right)^{1/2}-\mathbf{h}_k^H\mathbf{p}_k\right\},$$
$$\max_{k>1}\left\{d_k-|e_k|\right\}\quad\Big] \tag{8a}$$
$$\text{s.t.}\quad \frac{\sum_{k\in\mathcal{K}}u_k\left(C_k+\log(1+\gamma_{p,k})\right)}{\mu\left(\|\boldsymbol{p}_c\|^2+\sum_{k\in\mathcal{K}}\|\boldsymbol{p}_k\|^2\right)+P_c}\geq\delta \tag{8b}$$
$$\text{(3f)–(3j)}. \tag{8c}$$

Observe that (8b) is equivalent to the SOC

$$\sum_{k\in\mathcal{K}}u_k\left(C_k+\log(1+\gamma_{p,k})\right)\geq\delta\Bigg(\mu\Bigg(\|\boldsymbol{p}_c\|^2+\sum_{k\in\mathcal{K}}\|\boldsymbol{p}_k\|^2\Bigg)+P_c\Bigg)$$

since the denominator in (8b) is positive.

A bounding function $\beta(\mathcal{M})$ that satisfies (7) is required. First, observe that the objective of (8) is increasing in $(\boldsymbol{\gamma}_p, s)$. Hence, a lower bound on $[\underline{\boldsymbol{\gamma}}_p, \bar{\boldsymbol{\gamma}}_p]\times[\underline{s},\bar{s}]$ is obtained by setting $\boldsymbol{\gamma}_p=\underline{\boldsymbol{\gamma}}_p$ and $s=\underline{s}$ in the objective. Next, smoothen the objective of (8) by using the epigraph form with auxiliary variable t, and convert the pointwise maximum expressions to smooth constraints. Then, the new constraints $t\geq d_k-|e_k|$, for $k>1$, are equivalent to $(e_k, d_k-t)\in\mathcal{C}$. This set $\mathcal{C}$ is nonconvex. Consistent bounding of this set is obtained using argument cuts [16], i.e., introduce auxiliary variables $\alpha_k\in[0,2\pi]$, $k>1$, and add the constraint $\angle e_k=\alpha_k$. The variables $\boldsymbol{\alpha}$ are included in the nonconvex variables handled by the BRB solver. Then, a lower bound on the objective value of (8) over the box $[\underline{\boldsymbol{\alpha}},\bar{\boldsymbol{\alpha}}]$ is obtained by replacing the constraints $d_k\leq|e_k|,\ \angle e_k\in[\underline{\alpha}_k,\bar{\alpha}_k]$, with their convex envelope. For $\bar{\alpha}_k-\underline{\alpha}_k\leq\pi$, this is

$$\sin(\underline{\alpha}_k)\Re\{e_k\}-\cos(\underline{\alpha}_k)\Im\{e_k\}\leq 0 \tag{9a}$$
$$\sin(\bar{\alpha}_k)\Re\{e_k\}-\cos(\bar{\alpha}_k)\Im\{e_k\}\geq 0 \tag{9b}$$
$$a_k\Re\{e_k\}+b_k\Im\{e_k\}\geq(d_k-t)(a_k^2+b_k^2) \tag{9c}$$

and $(e_k,d_k)\in\mathbb{C}\times\mathbb{R}$ otherwise [16, Prop. 1], where $a_k=\frac{1}{2}\left(\cos(\underline{\alpha}_k)+\cos(\bar{\alpha}_k)\right)$, and $b_k=\frac{1}{2}\left(\sin(\underline{\alpha}_k)+\sin(\bar{\alpha}_k)\right)$.

The resulting bounding problem depends on $\boldsymbol{\gamma}_p$ and s only through to the constraints (2d), (3i), (8b), and $(\boldsymbol{\gamma}_p, s,\boldsymbol{\alpha})\in\mathcal{M}$. These can be transformed into affine functions of $(\boldsymbol{\gamma}_p, s)$ by substituting $s'=\log(1+s)$ and $\gamma'_{p,k}=\log(1+\gamma_{p,k})$. Then, these constraints are equivalent to

$$\sum_{k\in\mathcal{K}}u_k\left(C_k+\gamma_{p,k}\right)\geq\delta\Big(\mu\Big(\|\boldsymbol{p}_c\|^2+\sum_{k\in\mathcal{K}}\|\boldsymbol{p}_k\|^2\Big)+P_c\Big) \tag{10a}$$
$$\sum_{k\in\mathcal{K}}C_k\leq s,\quad C_k\geq\max\left\{0, R_k^{th}-\gamma_{p,k}\right\},\ \forall k\in\mathcal{K} \tag{10b}$$
$$s\in[\log(1+\underline{s}),\log(1+\bar{s})] \tag{10c}$$
$$\gamma_{p,k}\in[\log(1+\underline{\gamma}_{p,k}),\log(1+\bar{\gamma}_{p,k})],\quad\forall k\in\mathcal{K} \tag{10d}$$

and the final bounding problem is the SOCP

$$\min_{\substack{\boldsymbol{p}_c,\boldsymbol{p}_1,\dots,\boldsymbol{p}_K,\\ \mathbf{c},\boldsymbol{\gamma}_p,s,\boldsymbol{d},\boldsymbol{e},t}} t \tag{11a}$$
$$\text{s.t.}\quad\sqrt{\underline{\gamma}_{p,k}}\Bigg(\sum_{j\in\mathcal{K}\setminus k}|\mathbf{h}_k^H\mathbf{p}_j|^2+1\Bigg)^{1/2}\leq t+\mathbf{h}_k^H\mathbf{p}_k \tag{11b}$$
$$\sqrt{\underline{s}}\left(\sum\nolimits_{j\in\mathcal{K}}|\mathbf{h}_1^H\mathbf{p}_j|^2+1\right)^{1/2}\leq t+\mathbf{h}_1^H\mathbf{p}_c \tag{11c}$$
$$\sqrt{\underline{s}}\Bigg(\sum_{j\in\mathcal{K}}|\mathbf{h}_k^H\mathbf{p}_j|^2+1\Bigg)^{1/2}\leq t+d_k,\forall k>1 \tag{11d}$$
$$\forall k\in\mathcal{I}_{\mathcal{M}}:\text{(9a)–(9c)} \tag{11e}$$
$$\text{(2e), (3f)–(3h), (10a)–(10d)} \tag{11f}$$

where $\mathcal{I}_{\mathcal{M}}=\left\{k\in\mathcal{K}:k>1\wedge\max_{\underline{\alpha},\bar{\alpha}\in\mathcal{M}}|\bar{\alpha}_k-\underline{\alpha}_k|\leq\pi\right\}$. The bound $\beta(\mathcal{M})$ takes the optimal value of (11) if it is feasible. Otherwise, $\beta(\mathcal{M})=\infty$ otherwise.

3.2. Feasible Point

A dual feasible point is obtained from the solution $(\boldsymbol{\gamma}_p^\star, s^\star,\boldsymbol{e}^\star,\dots)$ of (11) as $(\boldsymbol{\gamma}_p^k, s^k,\boldsymbol{\alpha}^k)$ with $\gamma_{p,i}^k=2^{\gamma_{p,i}^\star}-1$, for $i\in\mathcal{K}$, $s^k=2^{s^\star}-1$ and $\boldsymbol{\alpha}^k\in\operatorname{proj}_{\boldsymbol{\alpha}}\mathcal{M}_k=[\underline{\boldsymbol{\alpha}}^k,\bar{\boldsymbol{\alpha}}^k]$. Numerical experiments show that the obvious choice $\alpha_i^k=\angle e_i^\star$ leads to very slow convergence. A much faster alternative is $\alpha_i^k=\arg\min_{\alpha\in\{\underline{\alpha}_i^k,\bar{\alpha}_i^k\}}|\alpha-\angle e_i^\star|$. This point is primal feasible if the optimal value of

$$\min_{\substack{\boldsymbol{p}_1,\dots,\boldsymbol{p}_K,\\ \boldsymbol{p}_c,\mathbf{c},\boldsymbol{d},\boldsymbol{e},t}} t\quad\text{s.t.}\quad\text{(8b), (3f)–(3j)}|_{\boldsymbol{\gamma}_p=\boldsymbol{\gamma}_p^k,s=s^k} \tag{12a}$$
$$\text{(11b)–(11d)}|_{\underline{\boldsymbol{\gamma}}_p=\boldsymbol{\gamma}_p^k,\underline{s}=s^k} \tag{12b}$$
$$\forall i>1:(e_i,d_i-t)\in\mathcal{C},\angle e_i=\alpha_i^k \tag{12c}$$

is less than or equal to zero. This is an SOCP since (12c) is affine.

Denote the optimal solution of (12) as $(t^*,\boldsymbol{c}^*,\boldsymbol{y}^*)$. It can be shown that the primal objective value of $(\boldsymbol{c}^*,\boldsymbol{y}^*)$ is greater than or equal to δ. This value can be further increased without impairing primal feasibility by updating $\boldsymbol{c}^*$ with the solution of the linear program $\max_{\boldsymbol{c}}\sum_{k\in\mathcal{K}}u_kC_k\ \text{ s.t. (2d), (3i), (8b)}|_{\boldsymbol{y}^*}$.

3.3. Reduction Procedure

The convergence criterion (7) implies that the quality of the bound $\beta(\mathcal{M})$ improves as the diameter of $\mathcal{M}$ shrinks. Since tighter bounds lead to faster convergence, it is beneficial to reduce the size of $\mathcal{M}$ prior to bounding if possible at low computational cost. To ensure convergence to the global solution, it is important that the reduced box $\mathcal{M}'\subseteq\mathcal{M}$ still contains all solution candidates.

Consider the box $\mathcal{M}=[\underline{\boldsymbol{\gamma}}_p,\bar{\boldsymbol{\gamma}}_p]\times[\underline{s},\bar{s}]\times[\underline{\boldsymbol{\alpha}},\bar{\boldsymbol{\alpha}}]$. Due to monotonicity, a necessary condition for the feasibility of (8) over $\mathcal{M}$ is that (2d), (3i), (8b) hold for $\bar{\boldsymbol{\gamma}}_p,\bar{s},\bar{\boldsymbol{\alpha}}$. Clearly, (2d) and (3i) can only hold if

$$\sum\nolimits_{k\in\mathcal{I}}\Big(R_k^{th}-\log(1+\bar{\gamma}_{p,k})\Big)-\log(1+\bar{s})\leq 0 \tag{13}$$

with $\mathcal{I}=\{k\in\mathcal{K}:R_k^{th}-\log(1+\bar{\gamma}_{p,k})>0\}$. Similarly, a necessary condition for (8b) to hold is

$$\max_{k\in\mathcal{K}}\{u_k\}\log(1+\bar{s})+\sum\nolimits_{k\in\mathcal{K}}u_k\log(1+\bar{\gamma}_{p,k})\geq\delta W \tag{14}$$

with $W=\left(\mu\left(\min\|\boldsymbol{p}_c\|^2+\sum_{k\in\mathcal{K}}\min\|\boldsymbol{p}_k\|^2\right)+P_c\right)$, where the minimum is such that $\boldsymbol{\gamma}_p\in\mathcal{M}$. This can be relaxed as $\min_{\boldsymbol{p}_c,\dots,\boldsymbol{p}_K}\|\boldsymbol{p}_\kappa\|^2\quad\text{s.t.}\quad\underline{\gamma}_{p,\kappa}\leq|\mathbf{h}_\kappa^H\mathbf{p}_\kappa|^2$. From the Karush-Kuhn-Tucker conditions, the optimal value of this problem is obtained as $\underline{\gamma}_{p,\kappa}\|\boldsymbol{h}_\kappa\|^{-2}$. Similarly, a lower bound for $\min\|\boldsymbol{p}_c\|^2$ is obtained as $\underline{s}\max_k\|\boldsymbol{h}_k\|^{-2}$. Hence,

$$W=\mu\Big(\underline{s}\max_k\|\boldsymbol{h}_k\|^{-2}+\sum\nolimits_{k\in\mathcal{K}}\underline{\gamma}_{p,k}\|\boldsymbol{h}_k\|^{-2}\Big)+P_c. \tag{15}$$

Conditions (13) and (14) can be used to reduce $\mathcal{M}$ and as a preliminary feasibility check before bounding. For the reduction, let $\mathcal{M}'=[\underline{\boldsymbol{\gamma}}'_p,\bar{\boldsymbol{\gamma}}'_p]\times[\underline{s}',\bar{s}']\times[\underline{\boldsymbol{\alpha}},\bar{\boldsymbol{\alpha}}]$ and consider (14). Every dual feasible $\gamma_{p,\kappa}\in\mathcal{M}$ satisfies $W\delta\leq U-u_\kappa\log(1+\bar{\gamma}_{p,\kappa})+$

$u_\kappa \log(1 + \gamma_{p,\kappa})$, where U is the right-hand side of (14). Hence, every dual feasible $\gamma_{p,\kappa}$ satisfies $\gamma_{p,\kappa} \geq 2^{\frac{W\delta - U}{u_\kappa}}(1 + \bar{\gamma}_{p,\kappa}) - 1$. Similarly, let V be the left-hand side of (13). From this condition, we see that every dual feasible $\gamma_{p,\kappa}$ satisfies $\gamma_{p,\kappa} \geq 2^V(1+\bar{\gamma}_{p,\kappa})-1$, for $\kappa \in \mathcal{I}$, and $\gamma_{p,\kappa} \geq 2^{V+R_\kappa^{th}} - 1$, for $\kappa \notin \mathcal{I}$. Thus, the lower bound for $\gamma_{p,k}$ can be reduced to $\underline{\gamma}'_{p,k} = \max\{\underline{\gamma}_{p,k}, \underline{\gamma}''_{p,k}\}$ without losing feasible solution candidates, where $\underline{\gamma}''_{p,k} = 2^{\max\{\frac{W\delta - U}{u_k},\, V\}}(1 + \bar{\gamma}_{p,k}) - 1$ if $k \in \mathcal{I}$, and $\max\{\, 2^{\frac{W\delta - U}{u_k}}(1 + \bar{\gamma}_{p,\kappa}), 2^{V+R_k^{th}}\} - 1$ otherwise. Likewise, the lower bound s can be reduced to $\underline{s}' = \max\{\underline{s},\, 2^{\max\left\{\frac{W\delta - U}{\max_{k\in\mathcal{K}}\{u_k\}},\, V\right\}}(1 + \bar{s}) - 1\}$.

Let W' be as in (15), evaluated at $(\underline{s}', \underline{\gamma}'_p)$, and consider (14) again. With a similar argument as before, the upper bound of $\mathcal{M}'$ can be reduced to $\bar{\gamma}'_{p,k} = \min\left\{\bar{\gamma}_{p,k},\ \underline{\gamma}'_{k,p} + (\delta\mu)^{-1}\|\boldsymbol{h}_k\|^2(U - \delta W')\right\}$ and $\bar{s}' = \min\left\{\bar{s},\ \underline{s}' + (\delta\mu)^{-1}\min_k \|\boldsymbol{h}_k\|^2(U - \delta W')\right\}$. Observe that this reduction procedure may lead to $\mathcal{M}' = \emptyset$.

3.4. Algorithm and Convergence

The complete algorithm is stated in Algorithm 1. It is essentially a BRB procedure [24, 25] that solves the SIT dual of (3) and updates the constant δ whenever a primal feasible point is encountered.

The initial box in Step 0 is computed as $\mathcal{M}_0 = [\mathbf{0}, \bar{\boldsymbol{\gamma}}_p^0] \times [0, \bar{s}^0] \times [0, 2\pi]^{K-1}$ with $\bar{\gamma}_{p,k} = P\|\boldsymbol{h}_k\|^2$ and $\bar{s} = \min_{k\in\mathcal{K}} P\|\boldsymbol{h}_k\|^2$. The set $\mathscr{R}_k$ holds the current partition of the feasible set, δ_k is the current best value adjusted by the tolerance η, and $\bar{\boldsymbol{x}}^k$ is the current best solution (CBS). In Step 1, the next box is selected as $\mathcal{M}_k$ and bisected into $\mathscr{P}_k$. These boxes are reduced according to Section 3.3 in Step 2. In Step 3, bounds for each reduced box are computed, infeasibility is detected, and dual feasible points are obtained from the bounding problem. For each of these points, primal feasibility is checked in Step 4. If feasible, a feasible point is recovered as in Section 3.2 and the corresponding primal objective value is computed. If necessary, the CBS and δ_k are updated in Step 5. Boxes that cannot contain primal ε-essential feasible solutions are pruned in Step 6. The algorithm is terminated in Step 7.

Theorem 1. *Alg. 1 converges in finitely many steps to a* (ε, η)*-optimal solution of* (3) *or establishes that no such solution exists.*

Proof. Omitted due to space constraints. □

Algorithm 1 SIT Algorithm for (3)

Step 0 **(Initialization)** Set $\varepsilon, \eta > 0$. Let $k = 1$ and $\mathscr{R}_0 = \{\mathcal{M}_0\}$. If an initial feasible solution $\boldsymbol{y}^0 = (\boldsymbol{p}_c^0, \dots, \boldsymbol{p}_K^0)$ is available, set $\delta_0 = \eta + v(2)|_{\boldsymbol{y}_0}$ and initialize $\bar{\boldsymbol{x}}^0 = (\boldsymbol{\gamma}_p^0, s^0, \boldsymbol{\alpha}^0)$ from (1), $s^0 = \min_{k\in\mathcal{K}} \gamma_{c,k}^0$, and $\alpha_k^0 = \angle \boldsymbol{h}_k^H \boldsymbol{p}_c^0$. Otherwise, do not set $\bar{\boldsymbol{x}}^0$ and choose $\delta_0 = 0$.

Step 1 **(Branching)** Let $\mathcal{M}_k = [\boldsymbol{r}^k, \boldsymbol{s}^k] \in \arg\min\{\beta(\mathcal{M}) \,|\, \mathcal{M} \in \mathscr{R}_{k-1}\}$. Bisect $\mathcal{M}_k$ into

$$\mathcal{M}^- = \{\boldsymbol{x} : r_j \leq x_j \leq v_j, r_i \leq x_i \leq s_i\ (i \neq j)\}$$
$$\mathcal{M}^+ = \{\boldsymbol{x} : v_j \leq x_j \leq s_j, r_i \leq x_i \leq s_i\ (i \neq j)\}$$

where $j_k \in \arg\max_j s_j^k - r_j^k$ and $\boldsymbol{v}^k = \frac{1}{2}(\boldsymbol{s}^k + \boldsymbol{r}^k)$. Set $\mathscr{P}_k = \{\mathcal{M}_-^k, \mathcal{M}_+^k\}$.

Step 2 **(Reduction)** Replace each box in $\mathscr{P}_k$ with $\mathcal{M}'$ as in Section 3.3.

Step 3 **(Bounding)** For each reduced box $\mathcal{M} \in \mathscr{P}_k$, solve (11). If infeasible, set $\beta(\mathcal{M}) = \infty$. Otherwise, set $\beta(\mathcal{M})$ to the optimal value of (11) and obtain a dual feasible point $\boldsymbol{x}(\mathcal{M})$ as in Section 3.2.

Step 4 **(Feasible Point)** For each $\mathcal{M} \in \mathscr{P}_k$, if $\beta(\mathcal{M}) \leq 0$ solve (12) for $\boldsymbol{x}(\mathcal{M})$ and denote the optimal value as $t(\boldsymbol{x}(\mathcal{M}))$. If $t(\boldsymbol{x}(\mathcal{M})) \leq 0$, $\boldsymbol{x}(\mathcal{M})$ is primal feasible. Recover $\boldsymbol{x}'(\mathcal{M})$ from the solution of (12) with $\boldsymbol{\gamma}'_p$, s' as in Step 0 and $\alpha'_k = \angle e_k^*$, $k > 1$, where $\boldsymbol{e}^*$ is from the optimal solution of (12). Update $\boldsymbol{c}^*$ as in Section 3.2 and compute the primal objective value $f(\mathcal{M})$. If $\beta(\mathcal{M}) > 0$ or $t(\boldsymbol{x}(\mathcal{M})) > 0$, set $f(\mathcal{M}) = -\infty$.

Step 5 **(Incumbent)** Let $\mathcal{M}' \in \arg\min\{f(\mathcal{M}) : \mathcal{M} \in \mathscr{P}_k\}$. If $f(\mathcal{M}') > \delta_{k-1} - \eta$, set $\bar{\boldsymbol{x}}^k = \boldsymbol{x}'(\mathcal{M}')$ and $\delta_k = f(\mathcal{M}') + \eta$. Otherwise, set $\bar{\boldsymbol{x}}^k = \bar{\boldsymbol{x}}^{k-1}$ and $\delta_k = \delta_{k-1}$.

Step 6 **(Pruning)** Delete every $\mathcal{M} \in \mathscr{P}_k$ with $\beta(\mathcal{M}) > -\varepsilon$ and collect the remaining sets in $\mathscr{P}'_k$. Set $\mathscr{R}_k = \mathscr{P}'_k \cup (\mathscr{R}_{k-1} \setminus \{\mathcal{M}_k\})$.

Step 7 **(Termination)** Terminate if $\mathscr{R} = \emptyset$: If $\bar{\boldsymbol{x}}^k$ is not set, then (3) is ε-essential infeasible; else $\bar{\boldsymbol{x}}^k$ is an essential (ε, η)-optimal solution of (3). Otherwise, update $k \leftarrow k + 1$ and return to Step 1.

	$K = 2$	$K = 3$	$K = 4$
Alg. 1	0.175 s / 0.099 s	4.579 s / 1.959 s	334.8 s / 126.3 s
BB	0.173 s / 0.091 s	7.605 s / 2.606 s	—
BB2	42.41 s / 2.380 s	158.5 s / 12.42 s	704.1 s / 265.8 s

Table 1. Mean / median run times to obtain the optimal solution. Problem instances where not all algorithms converged are ignored.

4. NUMERICAL EVALUATION

As most numerical problems of similar state-of-the-art algorithms arise from the multiple unicast beamforming problem, i.e., where $\boldsymbol{p}_c = \mathbf{0}$, we evaluate the performance of the algorithm for this case. In particular, we have generated 100 random i.i.d. channel realizations and solved (2) for $u_k = 1$, $\mu = 0$, $P_c = 0$, $R_k^{th} = 0$, $\frac{P}{\text{dB}} = -10, -5, \dots, 20$, and $K = M \in \{2, 3, 4\}$. This results in 700 problem instances per K. As baseline comparison and verification, we chose the straightforward BB implementation of this problem [14, 15] ("BB") and its variant with modified bounding problem from [14, §2.2.2] ("BB2"). For $K = 2$, BB2 stalled in 364 problem instances, while the other algorithms solved all problems. For $K = 3$, BB2 stalled 146 times and BB failed 13× due to numerical problems of the convex solver. Finally, for $K = 4$, BB did not solve a single problem instance due to numerical issues and BB2 stalled in 27 instances. Moreover, Algorithm 1 and BB2 did not solve the problem withing 60 minutes in 4 and 60 instances, respectively. Average computation times on a single core of an Intel Cascade Lake Platinum 9242 CPU are reported in Table 1. It can be observed that the proposed Algorithm 1 is more efficient than the two baseline algorithms especially when more users are in the system. Moreover, the joint beamforming problem, i.e., with $\boldsymbol{p}_c \neq \mathbf{0}$, was solved by Algorithm 1 for $K = 2$ with mean and median run times of 942 s and 2786 s. However, 23 instances were not solved within 12 hours.

Observe from the discussion in Section 3 that the complexity scales with $O(\exp(2K))$ in the number of users and polynomially in the number of antennas M. Hence, no noticeable changes in the reported run times are to be expected by varying M.

5. CONCLUSIONS

We developed the first global optimization algorithm to solve MISO downlink beamforming for RSMA with respect to WSR and EE maximization. This problem is an instance of joint multicast and unicast beamforming and also solves these problems separately. The algorithm is numerically stable and outperforms state-of-the-art multiple unicast beamforming algorithms considerably.

6. REFERENCES

[1] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.

[2] Y. Mao, B. Clerckx, and V. O. K. 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. 133, May 2018.

[3] Y. Mao and B. Clerckx, "Beyond dirty paper coding for multi-antenna broadcast channel with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6775–6791, Nov. 2020.

[4] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[5] 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, pp. 349–353, Mar. 2020.

[6] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, Dec. 2019.

[7] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2018, pp. 1–5.

[8] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, June 2006.

[9] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[10] 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.

[11] Y. Mao, B. Clerckx, J. Zhang, V. O. K. Li, and M. Arafah, "Max-min fairness of K-user cooperative rate-splitting in MISO broadcast channel with user relaying," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6362–6376, Oct. 2020.

[12] Z. Li, C. Ye, Y. Cui, S. Yang, and S. Shamai, "Rate splitting for multi-antenna downlink: Precoder design and practical implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1910–1924, Aug. 2020.

[13] H. Fu, S. Feng, W. Tang, and D. W. K. Ng, "Robust secure resource allocation for downlink two-user MISO rate-splitting systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, June 2020.

[14] E. Björnson and E. A. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*, vol. 9 of *FnT Commun. Inf. Theory*, Now, Boston, MA, USA, 2013.

[15] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[16] C. Lu and Y.-F. Liu, "An efficient global algorithm for single-group multicast beamforming," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3761–3774, July 2017.

[17] Y.-F. Liu, C. Lu, M. Tao, and J. Wu, "Joint multicast and unicast beamforming for the MISO downlink interference channel," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, July 2017.

[18] E. Chen, M. Tao, and Y.-F. Liu, "Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6265–6279, Sept. 2018.

[19] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Comput.*, 1999, pp. 987–996.

[20] MOSEK ApS, "MOSEK optimizer 9.2.26," 2020.

[21] H. Tuy, "Robust solution of nonconvex global optimization problems," *J. Global Optim.*, vol. 32, no. 2, pp. 307–323, June 2005.

[22] H. Tuy, "$\mathcal{D}(\mathcal{C})$-optimization and robust global optimization," *J. Global Optim.*, vol. 47, no. 3, pp. 485–501, Oct. 2009.

[23] B. Matthiesen and E. A. Jorswieck, "Efficient global optimal resource allocation in non-orthogonal interference networks," *IEEE Trans. Signal Process.*, vol. 67, no. 21, pp. 5612–5627, Nov. 2019.

[24] B. Matthiesen, *Efficient Globally Optimal Resource Allocation in Wireless Interference Networks*, Ph.d. thesis, Technische Universität Dresden, Dresden, Germany, Nov. 2019.

[25] H. Tuy, *Convex Analysis and Global Optimization*, vol. 110 of *Springer Optim. Appl.*, Springer-Verlag, New York; Berlin, Germany; Vienna, Austria, 2 edition, 2016.