Optimal Resource Allocation for Non-Regenerative Multiway Relaying with Rate Splitting

Bho Matthiesen and Eduard A. Jorswieck
Institute of Communication Technology, Technische Universität Dresden, Germany
Email: {bho.matthiesen, eduard.jorswieck}@tu-dresden.de

***Abstract*—Optimal resource allocation in interference networks requires the solution of non-convex optimization problems. Except from treating interference as noise (IAN) one usually has to optimize jointly over the achievable rates and transmit powers. This non-convexity is normally only due to the transmit powers while the rates are linear. Conventional approaches like the Polyblock Algorithm treat all variables equally and, thus, require a two layer solver to exploit the linearity in the rates and keep the computational complexity at a reasonable level. In this paper, we develop a branch and bound algorithm that exploits most of the problem structure and, compared to previous algorithms, has significantly better performance, improved numerical stability and provides a feasible solution even if terminated prematurely. We employ this novel algorithm to study throughput optimal power allocation in a multi-way relay channel with simultaneous non-unique decoding (SND) and rate splitting (RS) encoders. We evaluate the performance gains of RS over "pure" SND and IAN numerically. While SND often achieves significantly higher throughput than IAN, the benefits of rate splitting are not that pronounced on average and largely depend on the channel condition.**

***Index Terms*—Resource allocation, global optimization, multi-way relay channel, rate splitting, simultaneous non-unique decoding, interference networks**

I. Introduction

Recent results show that simultaneous non-unique decoding (SND) is an optimal decoder for general interference networks under the restriction to random codebooks with superposition coding and time sharing [1]. Before, common wisdom was that neither treating interference as noise (IAN) nor SND dominates the other rate-wise, with IAN generally better in noise limited scenarios and SND superior when interference is the limiting factor. This misconception is due to an longstanding oversight in the SND proof that was clarified in [1]. The major obstacle in implementing SND is the joint decoding whose direct implementation requires quite complex multiuser sequence detection. This issue, however, is subject to ongoing research. For example, in [2] commercially available point-to-point codes are used to achieve the whole SND region asymptotically. These major advances towards the implementation of theoretically optimal decoders brings the encoding into focus. For Gaussian channels, the most common approach is to use a single random Gaussian codebook per message, albeit there might be better choices for IAN.

We can improve upon this codebook by employing rate splitting (RS) and superposition coding. In this approach, each encoder splits its message into two parts, a "common" and a "private" and encodes it with a Gaussian codebook. It then transmits a symbol-wise superposition of both codewords. Each receiver recovers all parts of its desired message and the "common" messages of the interfering transmitters using SND. This coding scheme known as Han-Kobayashi coding is the most powerful available coding scheme for the interference channel [3], [4]. Its theoretical advantage over the non-RS approach is that IAN and joint decoding can be used jointly on the same message, while, otherwise, the decoder has to commit to only one of these strategies.

The goal of this paper is to develop a procedure to obtain global optimal transmit power allocations for a 3-user multi-way relay channel (MWRC) [5] with amplify-and-forward (AF) relaying and RS encoders [6]. This algorithm is then employed to evaluate the rate gain over non-RS encoders numerically. Specifically, we solve the optimization problem $\max_{\boldsymbol{R},\boldsymbol{P}}\{\boldsymbol{w}^T\boldsymbol{R} \,|\, \boldsymbol{R} \in \mathcal{R}(\boldsymbol{P}),\ \boldsymbol{P} \leq \bar{\boldsymbol{P}}\}$ for some $\boldsymbol{w} \in \mathbb{R}^n_{\geq 0}$ where $\mathcal{R}(\boldsymbol{P})$ denotes the achievable rate region. The major challenge here, as for most interference networks, is that this problem is not convex and, hence, the computational complexity of solving it grows exponentially in the number of variables [7]. Most research in the wireless resource allocation community focuses on rectangular rate regions that occur, e.g., when IAN is employed. Due to the simple structure of those rate regions, the rates can be eliminated analytically and numerical optimization over them is not necessary. Instead, the rate region at hand is more involved making it almost impossible to eliminate the rates analytically. Moreover, with RS each transmitter requires the allocation of two powers. Thus, we have three times as many optimization variables per transmitter as when simply using IAN. However, a careful examination of the optimization problem reveals that the non-convexity lies only in a few variables. Exploiting this structural property is key to solving the problem with reasonable complexity. In [8], we show how to achieve this within the widely used monotonic optimization framework [9]. This is, however, neither the most elegant nor efficient approach. Instead, we develop an optimization framework based on recent advancements of global optimization theory [10] that exploits most of the

This work is supported in part by the German Research Foundation (DFG) in the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing."

We thank the Center for Information Services and High Performance Computing (ZIH) at TU Dresden for generous allocations of computer time.

structure present in the problem at hand and is applicable to a wide range of resource allocation problems. The proposed algorithm is guaranteed to obtain a numerically stable solution and exhibits faster convergence and less numerical complexity than our earlier approach in [8].

II. System Model & Problem Statement

We use the system model from [8] and consider a 3-user single-input single-output (SISO) MWRC where the users communicate in multiple unicast transmissions via an AF relay. Gaussian channels with quasi-static block flat fading, full-duplex transmission, SND at the receivers and no direct user-to-user links are assumed. Users are indexed by k, $k \in \mathcal{K} = \{1, 2, 3\}$ and the relay is node 0.

The relay receives the signal $Y_0 = \sum_{k\in\mathcal{K}} h_k X_k + Z_0$, with X_k the channel input at node $k \in \mathcal{K}$ with power P_k, h_k the channel coefficient from user k to the relay, and Z_0 the independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian noise with power N_0 observed at the relay. The relay adjusts the power of the received symbol and broadcasts it back to the users, i.e., $X_0 = \alpha Y_0$ where $\alpha = \sqrt{P_0 / (\sum_{k\in\mathcal{K}} |h_k|^2 P_k + N_0)}$ is chosen such that the relay's transmit power is P_0.

User $k \in \mathcal{K}$ receives the signal $Y_k = g_k X_0 + Z_k$, with g_k the channel coefficient from the relay to user k, and Z_k the i.i.d. zero mean circularly symmetric complex Gaussian noise with power N. The channel inputs are subject to an average power constraint $\bar{P}_k$ on X_k, $k \in \mathcal{K} \cup \{0\}$. The message exchange is defined as follows. Let the receiver of node k's message be $q(k)$ and the user not interested in it $l(k)$. Then we consider without loss of generality $q(1) = l(3) = 2$, $q(2) = l(1) = 3$, and $q(3) = l(2) = 1$.

1) Rate Splitting: Encoder k, $k \in \mathcal{K}$, employs RS to divide its message into a common message to be decoded by all receivers and a private part that is treated as additional noise by unconcerned receivers. These messages are then encoded by individual Gaussian codebooks with powers P_k^c and P_k^p and linearly superposed to be transmitted in a single codeword with power $P_k = P_k^c + P_k^p$. The receiver first removes its self-interference from the observed signal and then employs SND to decode simultaneously for its desired message and non-uniquely for the cloud centers of the interfering message, while treating the interfering satellite codeword as noise.

The achievable rate region for this coding scheme in discrete memoryless multi-way relay channel is computed in [6] and extended to Gaussian channels below. Since the achievable rates do not depend on the absolute values of P_k^c, P_k^p and N_0 but on their ratio, we state all results in terms of the signal-to-noise ratio (SNR). Define the cloud and satellite transmit SNRs as $S_k^c = \frac{P_k^c}{N_0}$ and $S_k^p = \frac{P_k^p}{N_0}$, respectively, their sum $S_k = S_k^c + S_k^p$, the maximum transmit SNR $\bar{S}_k = \frac{\bar{P}_k}{N_0}$, and the corresponding vectors $\boldsymbol{S}^c = (S_1^c, S_2^c, S_3^c)$, $\boldsymbol{S}^p = (S_1^p, S_2^p, S_3^p)$, $\boldsymbol{S} = (\boldsymbol{S}^c, \boldsymbol{S}^p)$, and $\bar{\boldsymbol{S}} = (\bar{S}_1, \bar{S}_2, \bar{S}_3)$.

Lemma 1: A rate triple (R_1, R_2, R_3) is achievable for the Gaussian MWRC with AF relaying if, for all $k \in \mathcal{K}$,

$$R_k \leq B_k, \tag{1a}$$
$$R_k + R_{q(k)} \leq A_k + D_{q(k)}, \tag{1b}$$
$$R_k + R_{q(k)} + R_{l(k)} \leq A_k + C_{q(k)} + D_{l(k)}, \tag{1c}$$
$$2R_k + R_{q(k)} + R_{l(k)} \leq A_k + C_{q(k)} + C_{l(k)} + D_k, \tag{1d}$$

and,
$$R_1 + R_2 + R_3 \leq C_1 + C_2 + C_3, \tag{1e}$$

with
$$A_k = \log\left(1 + \frac{|h_k|^2 S_k^p}{\gamma_k(\boldsymbol{S})}\right) \tag{2a}$$
$$B_k = \log\left(1 + \frac{|h_k|^2 (S_k^p + S_k^c)}{\gamma_k(\boldsymbol{S})}\right) \tag{2b}$$
$$C_k = \log\left(1 + \frac{|h_k|^2 S_k^p + |h_{l(k)}|^2 S_{l(k)}^c}{\gamma_k(\boldsymbol{S})}\right) \tag{2c}$$
$$D_k = \log\left(1 + \frac{|h_k|^2 (S_k^p + S_k^c) + |h_{l(k)}|^2 S_{l(k)}^c}{\gamma_k(\boldsymbol{S})}\right) \tag{2d}$$

where $S_k^c + S_k^p \leq \bar{S}_k$ and

$$\gamma_k(\boldsymbol{S}) = 1 + |h_{l(k)}|^2 S_{l(k)}^p + \widetilde{g}_{q(k)}^{-1}\left(1 + \sum_{i\in\mathcal{K}} |h_i|^2 (S_i^c + S_i^p)\right),$$

with $\widetilde{g}_k = |g_k|^2 \frac{\bar{P}_0}{N_k}$.

Proof sketch: Extend [6, Thm. 1] to Gaussian channels using the standard procedure in [11, Sect. 3.4.1]. Evaluate it with Gaussian inputs $U_k \sim \mathcal{CN}(0, P_k^c)$ and $X_k = U_k + V_k$ with $V_k \sim \mathcal{CN}(0, P_k^p)$, and $\mathbb{E}[X_0^2] = P_0$ to obtain the rate expressions above with $\widetilde{g}_k = |g_k|^2 \frac{P_0}{N_k}$. The achievable rates are increasing in P_0. Thus, $P_0 = \bar{P}_0$ is rate-optimal. ■

2) Single Message: Encoder k, $k \in \mathcal{K}$, uses a Gaussian codebook with power P_k to transmit its message. Upon observing y_k^n the receiver uses the SND typicality decoder to recover its desired message. The achievable rate region is given below where $S_k = \frac{P_k}{N_0}$, $\boldsymbol{S} = (S_1, S_2, S_3)$, and $\bar{\boldsymbol{S}} = (\bar{S}_1, \bar{S}_2, \bar{S}_3)$. Observe that this region is strictly larger than previously published SND regions [6], [8] and includes IAN as a special case. This is due to the new insights on SND decoders in [1].

Lemma 2: A rate triple (R_1, R_2, R_3) is achievable for the Gaussian MWRC with AF and SND if, for each $k \in \mathcal{K}$,

$$R_k \leq \log\left(1 + \frac{|h_k|^2 S_k}{\gamma_{l(k)}(\boldsymbol{S})}\right) \tag{3}$$

or

$$R_k \leq \log\left(1 + \frac{|h_k|^2 S_k}{\delta_{l(k)}(\boldsymbol{S})}\right) \tag{4a}$$
$$R_k + R_{l(k)} \leq \log\left(1 + \frac{|h_k|^2 S_k + |h_{l(k)}|^2 S_{l(k)}}{\delta_{l(k)}(\boldsymbol{S})}\right) \tag{4b}$$

where $S_k \leq \bar{S}_k$, $\gamma_k(\boldsymbol{S})$ and $\widetilde{g}_k$ as in Lemma 1, and $\delta_k(\boldsymbol{S}) = \gamma_k(\boldsymbol{S}) - |h_{l(k)}|^2 S_{l(k)}^p$.

Proof: The proof follows along the lines of [1, Sect. II-A.] and is omitted due to space constraints. ■

Remark 1: Let $\mathcal{R}_{k,\text{IAN}}$ and $\mathcal{R}_{k,\text{SND}}$ be the regions defined by (3) and (4), respectively. Then, the rate region in Lemma 2 is

$$\mathcal{R} = \bigcap_{k\in\mathcal{K}} (\mathcal{R}_{k,\text{IAN}} \cup \mathcal{R}_{k,\text{SND}}) = \bigcup_{\boldsymbol{d}\in\{\text{IAN},\text{SND}\}^{|\mathcal{K}|}} \bigcap_{k\in\mathcal{K}} \mathcal{R}_{k,d_k}. \tag{5}$$

A. Problem Statement

The optimal power allocation that maximizes the weighted system throughput and characterizes the Pareto boundary of these achievable rate regions is the solution to

$$\begin{cases} \max_{\boldsymbol{R},\boldsymbol{S}} & \sum_{k\in\mathcal{K}} w_k R_k \\ \text{s.t.} & \boldsymbol{R} \in \mathcal{R}(\boldsymbol{S}), \quad \boldsymbol{R} \geq \underline{\boldsymbol{R}} \\ & \boldsymbol{S} \geq \boldsymbol{0}, \quad S_k \leq \bar{S}_k,\ k \in \mathcal{K} \end{cases} \tag{P1}$$

for all $\boldsymbol{w} \in \mathbb{R}^3_{\geq 0} \setminus \{\boldsymbol{0}\}$ where $\mathcal{R}(\boldsymbol{S})$ is one of the rate regions defined in Lemmas 1 and 2 and $\underline{\boldsymbol{R}} \geq \boldsymbol{0}$ are minimum rate requirements. This is a non-convex optimization problem because the right-hand sides (RHSs) of the rate constraints are non-concave functions. Instead, they belong to the classes of difference of convex (DC) and difference of increasing functions since

$$\log\left(1 + \frac{f(\boldsymbol{S})}{\gamma_k(\boldsymbol{S})}\right) = \log(f(\boldsymbol{S}) + \gamma_k(\boldsymbol{S})) - \log(\gamma_k(\boldsymbol{S})). \tag{6}$$

Following the general rule in global optimization to exploit as much structure in the problem as possible, we strive to exploit both these properties in our proposed algorithm. This, and the fact that (P1) is a linear program (LP) for fixed $\boldsymbol{S}$ prohibits the use of of-the-shelve DC progamming or monotonic optimization algorithms. Instead, we develop a robust branch and bound (BB) algorithm inspired by the successive incumbent transcending (SIT) algorithm for DC programming [10], [12].

III. Robust Global Optimization

A. Review of Robust Global Optimization

The following explanations are a summary of [12, Sect. 7.5] and [10], [13] with some minor modifications tailored for our specific application. Consider the general non-convex optimization problem

$$\min_{\boldsymbol{x}\in[\boldsymbol{a},\boldsymbol{b}]} f(\boldsymbol{x}) \quad \text{s.t.} \quad g_i(\boldsymbol{x}) \leq 0,\ i = 1, 2, \ldots, m \tag{P2}$$

where $f, g_1, g_2, \ldots, g_m$ are non-convex continuous real-valued functions. Most current solution methods for this problem are devised to find a (ε, η)-approximate optimal solution, i.e., a solution $\bar{\boldsymbol{x}}$ that satisfies $f(\bar{\boldsymbol{x}}) - \eta \leq f(\boldsymbol{x}^*)$ where $\boldsymbol{x}^*$ is the optimal solution of (P2) and $g_i(\bar{\boldsymbol{x}}) \leq \varepsilon$ for all i and some sufficiently small $\varepsilon > 0$. The problem with this approach is that such a solution $\bar{\boldsymbol{x}}$ might be infeasible and also quite far away from the real solution $\boldsymbol{x}^*$ if ε is not sufficiently small. Unfortunately, it is often not known in practice how small "sufficiently small" is so as to guarantee a correct approximate optimal solution.

If (P2) has a global optimal solution that is an isolated feasible solution, i.e., a point $\boldsymbol{x}$ at the center of a ball containing no other feasible points than $\boldsymbol{x}$, things are even worse. In that case, a slight change in the data or tolerances ε, η might cause a drastic change in the obtained solution (including the optimal value). Besides the numerical difficulties to compute such a solution, it is not even desirable to obtain one since resource allocation problems in wireless communications usually deal with data that is subject to measurement and estimation errors. Thus, a common practice is to assume a robust feasible set, i.e., a feasible set containing no isolated feasible solutions. Unfortunately, this assumption is generally very hard to check for a given problem. Instead, a practical algorithm should compute the best non-isolated feasible solution without knowing a priori whether the feasible set is robust or not.

Let $\mathcal{D}$ be the feasible set of (P2) and $\mathcal{D}^* = \operatorname{cl}(\operatorname{int}\mathcal{D})$ the set of nonisolated feasible points of $\mathcal{D}$, where cl and int denote the closure and interior, respectively. A solution $\boldsymbol{x}^* \in \mathcal{D}^*$ is called essential optimal solution of (P2) if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $x \in \mathcal{D}^*$. A point $\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]$ satisfying $g_i(\boldsymbol{x}) \leq -\varepsilon$ for all i and some $\varepsilon > 0$ is called ε-essential feasible and a solution of (P2) is said to be essential (ε, η)-optimal if it satisfies

$$f(\boldsymbol{x}^*) - \eta \leq \inf\{f(\boldsymbol{x}) | \boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}],\ \forall i : g_i(\boldsymbol{x}) \leq -\varepsilon\} \tag{7}$$

for some $\eta > 0$. Clearly, for $\varepsilon, \eta \to 0$ an essential (ε, η)-optimal solution is a nonisolated feasible point which is optimal.

The robust approach to global optimization employed here uses the SIT scheme in Algorithm 1 to generate a sequence of nonisolated feasible solutions converging to an essential optimal solution of (P2). The core problem in the SIT scheme is, given a real number γ, to check whether (P2) has a nonisolated feasible solution $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \leq \gamma$, or, else, establish that no such ε-essential feasible $\boldsymbol{x}$ exists. Given that this subproblem is solved within finitely many steps, Algorithm 1 is also finite. Apart from the improved numerical stability and convergence, the SIT algorithm has another very desirable feature: it provides a good nonisolated feasible (but possibly suboptimal) solution even if terminated prematurely. Instead, conventional algorithms usually outer approximate the solution rendering intermediate solutions almost useless.

Algorithm 1 SIT Algorithm [12, Sect. 7.5.1].

Step 0	Initialize $\bar{\boldsymbol{x}}$ with best known nonisolated feasible solution; otherwise choose $\bar{\boldsymbol{x}}$ such that $f(\bar{\boldsymbol{x}}) - \eta > f(\boldsymbol{x})\ \forall \boldsymbol{x} \in \mathcal{D}$. Set $\gamma_0 = \gamma = f(\bar{\boldsymbol{x}}) - \eta$.
Step 1	Check if (P2) has a nonisolated feasible solution $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \leq \gamma$; otherwise, establish that no such ε-essential feasible $\boldsymbol{x}$ exists and go to Step 3.
Step 2	Update $\bar{\boldsymbol{x}} \leftarrow \boldsymbol{x}$ and $\gamma \leftarrow f(\bar{\boldsymbol{x}}) - \eta$. Go to Step 1.
Step 3	Terminate: If $\gamma \neq \gamma_0$, $\bar{\boldsymbol{x}}$ is an essential (ε, η)-optimal solution; else Problem (P2) is ε-essential infeasible.

Now, let f be a convex and g_i be DC functions and consider the following dual problem to (P2)

$$\min_{\boldsymbol{x}\in[\boldsymbol{a},\boldsymbol{b}]} \max_{i=1,2,\ldots,m} g_i(\boldsymbol{x}) \quad \text{s.t.} \quad f(\boldsymbol{x}) \leq \gamma \tag{P3}$$

where objective and constraints are interchanged. In contrast to (P2), Problem (P3) has a convex feasible set and, thus, no isolated feasible solutions. Moreover, computing a feasible solution can be done at cheap cost using an adaptive BB procedure [12, Prop. 6.2] that partitions the feasible set into boxes. For each partition M, a lower bound $\beta(M)$ for (P3) with the additional constraint $\boldsymbol{x} \in M$ is computed. In each iteration of the BB algorithm two points $\boldsymbol{x}^k \in M_k$, $\boldsymbol{y}^k \in M_k$ satisfying

$$f(\boldsymbol{x}^k) \leq \gamma, \quad g(\boldsymbol{y}^k) - \beta(M_k) \to 0 \text{ as } k \to \infty \tag{8}$$

are generated, where M_k is the box with the best bound $\beta(M)$. The following proposition links the solution of (P3) to the original problem (P2).

Proposition 1 ([12, Prop. 7.14]): Let $\varepsilon > 0$ be given. Either $g(\boldsymbol{x}^k) < 0$ for some k or $\beta(M_k) > -\varepsilon$ for some k. In the former case, $\boldsymbol{x}^k$ is a nonisolated feasible solution of (P2) satisfying $f(\boldsymbol{x}^k) \leq \gamma$. In the latter case, no ε-essential feasible solution $\boldsymbol{x}$ of (P2) exists such that $f(\boldsymbol{x}) \leq \gamma$ (so, if $\gamma = f(\bar{\boldsymbol{x}}) - \eta$ for a given $\eta > 0$ and a nonisolated feasible solution $\bar{\boldsymbol{x}}$ then $\bar{\boldsymbol{x}}$ is an essential (ε, η)-optimal solution).

Thus, an adaptive BB algorithm for solving (P3) with deletion criterion $\beta(M) > -\varepsilon$ and stopping criterion $g(\boldsymbol{x}^k) < 0$ implements Step 1 in Algorithm 1.

B. Application to Resource Allocation Problems

We now return our attention to the original problem (P1) and extend the SIT algorithm from [12]. As noted before, the devised algorithm should only branch over the non-convex variables. Thus, we formulate the optimization problem in terms of the non-convex variables $\boldsymbol{x}$ and convex variables $\boldsymbol{\xi}$:

$$\begin{cases} \min\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & f(\boldsymbol{x},\boldsymbol{\xi}) \\ \text{s.t.} & g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x}) \leq 0, \quad i = 1,2,\ldots,m, \end{cases} \tag{P4}$$

where $\mathcal{C}$ is a convex set, f, g_i^+ are convex, and g_i^- are convex and decreasing functions.[1] The essential step in solving (P4) is the computation of the bound $\beta(M)$ for the dual problem. A common approach is to use a convex underestimator for $\max_{i=1,2,\ldots,m} g_i(\boldsymbol{x})$ that is tight at a point $\boldsymbol{y}^k \in M$ in order to satisfy (8).

Proposition 2: The objective of (P4)'s dual is underestimated on the box $M = [\boldsymbol{p}, \boldsymbol{q}]$ by

$$\max_{i=1,2,\ldots,m} \left\{ g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{p}) \right\}. \tag{9}$$

This lower bound is tight at $\boldsymbol{p}$.

Proof: Since $g_i^-(\boldsymbol{x})$ is monotonically decreasing, $g_i^-(\boldsymbol{p}) \geq g_i^-(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{M}$, and each term of (9) underestimates the corresponding term in the objective of (P4)'s dual and is tight for the point $\boldsymbol{p}$. It remains to show that for all real-valued functions $h_1, h_2, \ldots, \underline{h}_1, \underline{h}_2, \ldots$ satisfying $h_i \geq \underline{h}_i$ and $h_i(\boldsymbol{y}) = \underline{h}_i(\boldsymbol{y})$ for all i and some point $\boldsymbol{y}$, $\max_i\{h_i\} \geq \max_i\{\underline{h}_i\}$ and $\max_i\{h_i(\boldsymbol{y})\} = \max_i\{\underline{h}_i(\boldsymbol{y})\}$ holds.

Consider the case with two functions and assume that $\max\{h_1, h_2\} < \max\{\underline{h}_1, \underline{h}_2\}$. Since $h_i \geq \underline{h}_i$, this can only hold if $\max\{h_1, h_2\} = h_1$ and $\max\{\underline{h}_1, \underline{h}_2\} = \underline{h}_2$ or vice versa. This implies $h_1 \geq h_2 \geq \underline{h}_2$ which contradicts the assumption. The generalization to arbitrarily many functions follows by induction. Finally, if $\underline{h}_i(\boldsymbol{y}) = h_i(\boldsymbol{y})$ for all i, then $\min_i\{\underline{h}_i(\boldsymbol{y})\} = \min_i\{h_i(\boldsymbol{y})\}$. ■

Thus, a bound $\beta(M)$ for the optimal value of (P4)'s dual over the box $M = [\boldsymbol{p}, \boldsymbol{q}]$ is the solution to

$$\begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{\xi},t} & t \\ \text{s.t.} & f(\boldsymbol{x},\boldsymbol{\xi}) \leq \gamma \\ & g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{p}) \leq t, \quad i = 1,2,\ldots,m \\ & (\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{C}, \ \boldsymbol{x} \in M \end{cases} \tag{P5}$$

[1] Note that g_i^+ and g_i^- are convex functions while the logarithms, which they might represent, are concave functions. Thus, if the logarithm is increasing, g_i^- must be decreasing (cf. (P8)).

This is obviously a convex optimization problem since $g_i^-(\boldsymbol{p})$ is a constant now. Thus, we can use standard tools from convex optimizations to obtain a solution in polynomial time [14].

The final algorithm from [12, Sect. 7.5.2] is extended in Algorithm 2. Its convergence follows from [12, Theorem 7.10]. The required initial box $M_0 = [\boldsymbol{p}^0, \boldsymbol{q}^0]$ contains the $\boldsymbol{x}$-dimensions of $\mathcal{C}$, i.e.,

$$p_i^0 = \min_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} x_i \qquad q_i^0 = \max_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} x_i, \tag{10}$$

and $\mathcal{C}_{\boldsymbol{x}}$ denotes the $\boldsymbol{x}$-section of $\mathcal{C}$ defined as $\mathcal{C}_{\boldsymbol{x}} = \{(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}) \in \mathcal{C} \,|\, \boldsymbol{x} = \tilde{\boldsymbol{x}}\}$. Problems (P6) and (P7) in Steps 3 and 4 of Algorithm 2 are convex.

Algorithm 2 SIT Algorithm for (P4)

Step 0 Initialize $\varepsilon, \eta > 0$ and $M_0 = [\boldsymbol{p}^0, \boldsymbol{q}^0]$ as in (10), $\gamma = \infty$, $\mathscr{P}_1 = \{M_0\}$, $\mathscr{R} = \emptyset$, and $k = 1$.

Step 1 For each box $M \in \mathscr{P}_k$:
- Compute $\beta(M)$ with (P5). Set $\beta(M) = \infty$ if (P5) is infeasible.
- Add M to $\mathscr{R}$ if $\beta(M) \leq -\varepsilon$.

Step 2 Terminate if $\mathcal{R} = \emptyset$: If $\gamma = \infty$, then (P4) is ε-essential infeasible; else $\bar{\boldsymbol{x}}$ is an essential (ε, η)-optimal solution of (P4).

Step 3 Let $M_k = \arg\min\{\beta(M) \,|\, M \in \mathscr{R}\} = [\boldsymbol{p}^k, \boldsymbol{q}^k]$. Let $\boldsymbol{x}^k$ be the optimal solution of (P5) for the box M_k, and $\boldsymbol{y}^k = \boldsymbol{p}^k$. Solve the feasibility problem

$$\begin{cases} \text{find} & \boldsymbol{\xi} \in \mathcal{C}_{\boldsymbol{x}^k} \\ \text{s.t.} & g_i^+(\boldsymbol{x}^k, \boldsymbol{\xi}) - g_i^-(\boldsymbol{x}^k) \leq 0, \ i = 1, \ldots, m. \end{cases} \tag{P6}$$

If (P6) is feasible go to Step 4; otherwise go to Step 5.

Step 4 $\boldsymbol{x}^k$ is a nonisolated feasible solution satisfying $f(\boldsymbol{x}^k, \boldsymbol{\xi}) \leq \gamma$ for some $\boldsymbol{\xi} \in \mathcal{C}_{\boldsymbol{x}^k}$. Let $\boldsymbol{\xi}^*$ be the solution to

$$\begin{cases} \min\limits_{\boldsymbol{\xi}\in\mathcal{C}_{\boldsymbol{x}^k}} & f(\boldsymbol{x}^k, \boldsymbol{\xi}) \\ \text{s.t.} & g_i^+(\boldsymbol{x}^k, \boldsymbol{\xi}) - g_i^-(\boldsymbol{x}^k) \leq 0, \quad i = 1, 2 \ldots, m. \end{cases} \tag{P7}$$

If $\gamma = \infty$ or $\gamma < \infty$ and $f(\boldsymbol{x}^k, \boldsymbol{\xi}^*) < \gamma + \eta$, set $\bar{\boldsymbol{x}} = \boldsymbol{x}^k$ and $\gamma = f(\boldsymbol{x}^k, \boldsymbol{\xi}^*) - \eta$.

Step 5 Bisect M_k via (v^k, j_k) where $j_k \in \arg\max_j\{|y_j^k - x_j^k|\}$ and $v^k = \frac{1}{2}(x_{j_k}^k + y_{j_k}^k)$, i.e., let

$$M_k^- = \{\boldsymbol{x} \,|\, p_{j_k}^k \leq x_i \leq v^k, \ p_i^k \leq x_i \leq q_i^k \ (i \neq j_k)\}$$
$$M_k^+ = \{\boldsymbol{x} \,|\, v^k \leq x_i \leq q_{j_k}^k, \ p_i^k \leq x_i \leq q_i^k \ (i \neq j_k)\}.$$

Remove M_k from $\mathscr{R}$. Let $\mathscr{P}_{k+1} = \{M_k^-, M_k^+\}$. Increment k and go to Step 1.

C. Solution of Problem (P1)

1) Rate Splitting (Lemma 1): We already observed that (P1) is a LP for fixed $\boldsymbol{S}$, which leaves us with six non-convex variables $\boldsymbol{S}$. The non-convexity of (P1) stems from the negative $\log(\gamma_k(\boldsymbol{S}))$-terms on the RHS of $\mathcal{R}(\boldsymbol{S})$. The variables $\boldsymbol{S}^c$ only appear in $\gamma_k(\boldsymbol{S})$ as the sum $\sum_{k\in\mathcal{K}} |h_k|^2 S_k^c$. Thus, we can reduce the number of non-convex variables by introducing a (non-convex) auxiliary variable $y = \sum_{k\in\mathcal{K}} |h_k|^2 S_k^c$. Then, $\gamma_k(\boldsymbol{S}^p, y) = 1 + |h_{l(k)}|^2 S_{l(k)}^p + \widetilde{g}_{q(k)}^{-1}\left(1 + y + \sum_{i\in\mathcal{K}} |h_i|^2 S_i^p\right)$. Due to the exponential complexity in the number of non-convex variables this substitution reduces the computation time significantly. To bring (P1) in a form matching (P4), we introduce some additional definitions. Consider the constraints

(1a)–(1e) and number them by $i = 1, 2, \dots, 13$. The left-hand side of each inequality can be written as $\boldsymbol{a}_i^T \boldsymbol{R}$, while each term in the RHS is equivalent to $l_j^+(\boldsymbol{S}, y) - l_j^-(\boldsymbol{S}^p, y)$ for some $j = 1, \dots, 12$ (cf. (6)). Thus, the RHSs of (1a)–(1e) can be written as $L_i^+(\boldsymbol{S}, y) - L_i^-(\boldsymbol{S}^p, y)$ with $L_i^+(\boldsymbol{S}, y) = \sum_{j \in \mathcal{I}_i} l_j^+(\boldsymbol{S}, y)$ and $L_i^-(\boldsymbol{S}, y) = \sum_{j \in \mathcal{I}_i} l_j^-(\boldsymbol{S}^p, y)$ where $\mathcal{I}_i$ contains the corresponding indices. Then, (P1) is equivalent to

$$\begin{cases} \max_{\boldsymbol{R}, \boldsymbol{S}, y} & \boldsymbol{w}^T \boldsymbol{R} \\ \text{s.t.} & \boldsymbol{a}_i^T \boldsymbol{R} - L_i^+(\boldsymbol{S}, y) + L_i^-(\boldsymbol{S}^p, y) \leq 0, \quad i = 1, 2, \dots \\ & y = \sum_{k \in \mathcal{K}} |h_k|^2 S_k^c \\ & S_k^c + S_k^p \leq \bar{S}_k, \ k \in \mathcal{K}, \qquad \boldsymbol{R} \geq \underline{\boldsymbol{R}}, \quad \boldsymbol{S} \geq \boldsymbol{0} \end{cases} \tag{P8}$$

and we can easily identify $\boldsymbol{x} = (\boldsymbol{S}^p, y)$, $\boldsymbol{\xi} = (\boldsymbol{R}, \boldsymbol{S}^c)$, $f(\boldsymbol{x}, \boldsymbol{\xi}) = \boldsymbol{w}^T \boldsymbol{R}$, $g_i^+(\boldsymbol{x}, \boldsymbol{\xi}) = a_i^T \boldsymbol{R} - L_i^+(\boldsymbol{S}, y)$, $g_i^-(\boldsymbol{x}) = -L_i^-(\boldsymbol{S}^p, y)$, and $\mathcal{C}$ as the set defined by the last three lines of (P8). The initial box M_0 required by Algorithm 2 is identified as $[\boldsymbol{0}, \bar{\boldsymbol{S}}] \times [0, \sum_{k \in \mathcal{K}} |h_k|^2 \bar{S}_k]$.

2) Single Message (Lemma 2): Observe that $\inf_{\boldsymbol{x} \in \bigcup_i \mathcal{D}_i} f(\boldsymbol{x}) = \min_i \inf_{\boldsymbol{x} \in \mathcal{D}_i} f(\boldsymbol{x})$. Thus, we can split the resource allocation problem for Lemma 2 into eight individual optimization problems as indicated in Remark 1. Each is easily identified as an instance of (P1) and solvable with Algorithm 2 using the initial box $M_0 = [\boldsymbol{0}, \bar{\boldsymbol{S}}]$.

IV. Numerical Evaluation

We employ Algorithm 2 to answer two questions numerically: How does the extended SND region improve upon IAN and the "traditional" SND region, and how much can we gain over single message SND by using rate splitting? For this purpose we compute the optimal power allocations for the Gaussian MWRC with AF relaying considered in Section II with tolerances $\varepsilon = \eta = 0.01$. Results for "traditional" SND and IAN are published in [8] and [15], respectively. The simulation setup is identical to these publications: We assume equal maximum power constraints and noise power at the users and the relay, no minimum rate constraints, i.e. $\underline{\boldsymbol{R}} = \boldsymbol{0}$, reciprocal channels chosen randomly and independently with circular symmetric complex Gaussian distribution, i.e., $h_k \sim \mathcal{CN}(0, 1)$ and $g_k = h_k^*$. Results are averaged over 800 i.i.d. channel realizations.

Figure 1 displays the results. First, observe that, in accordance with conventional wisdom, neither "traditional" SND nor IAN dominates the other. Instead, "extended" SND clearly dominates the other two where the gain is solely due to allowing each receiver to either use IAN or "traditional" SND. The average gain of SND over the other two is approximately 12 % and 18 % at 10 dB, respectively, or 0.32 bpcu and 0.48 bpcu. Note that this gain is only achieved by allowing each receiver to chose between IAN and joint decoding which does not result in higher decoding complexity than "traditional" SND.

The average gain observed for RS over single message SND and shown in Fig. 1 is rather small, e.g., at 25 dB it is only 0.2 bpcu. However, depending on the channel realization we observed gains up to half a bit (or 3.5 %) at 10 dB. With spectrum being an increasingly scarce resource this occasional gain might justify the slightly higher coding complexity.

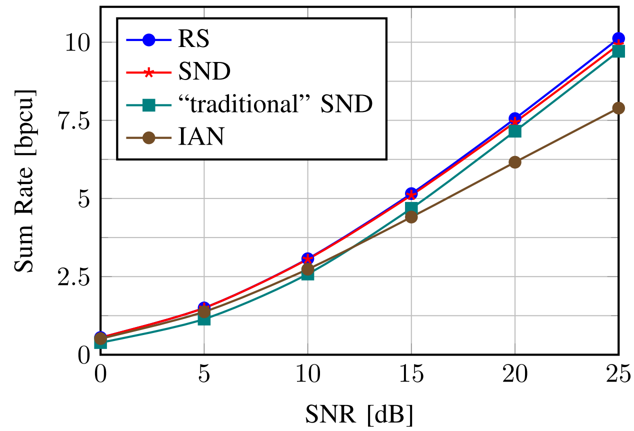

Fig. 1. Throughput in the MWRC with AF relaying and 1) RS 2) SND; 3) "traditional" SND; and 4) IAN. Averaged over 800 i.i.d. channel realizations.

V. Conclusions

We evaluated the merits of SND in MWRCs with and without rate splitting numerically and compared it with IAN and restricted SND. This required a novel global optimization algorithm to keep the numerical complexity at a reasonable level. Although the channel model is quite specific, the developed algorithm is applicable to a wide variety of non-convex optimization problems.

References

[1] B. Bandemer, A. El Gamal, and Y.-H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, Oct. 2015.

[2] L. Wang, Y.-H. Kim, C.-Y. Chen, H. Park, and E. Sasoglu, "Sliding-window superposition coding: Two-user interference channels," Jan. 2017. [Online]. Available: http://arxiv.org/abs/1701.02345

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

[4] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.

[5] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, "The multiway relay channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, Jan. 2013.

[6] B. Matthiesen and E. A. Jorswieck, "Instantaneous relaying for the 3-way relay channel with circular message exchanges," in *49th Asilomar Conf. on Signals, Syst., and Comput.*, Pacific Grove, CA, Nov. 2015.

[7] M. Hong and Z.-Q. Luo, "Signal processing and optimal resource allocation for the interference channel," in *Academic Press Library in Signal Processing*, N. D. Sidiropoulos, F. Gini, R. Chellappa, and S. Theodoridis, Eds. Elsevier, 2014, vol. 2, ch. 8, pp. 409–469.

[8] B. Matthiesen and E. A. Jorswieck, "Weighted sum rate maximization for non-regenerative multi-way relay channels with multi-user decoding," in *IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process (CAMSAP)*, Curaçao, Dec. 2017.

[9] H. Tuy, "Monotonic optimization: Problems and solution approaches," *SIAM J. Optimization*, vol. 11, no. 2, pp. 464–494, Feb. 2000.

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

[11] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[12] H. Tuy, *Convex Analysis and Global Optimization*. Springer, 2016.

[13] ——, "Robust solution of nonconvex global optimization problems," *J. Global Optim.*, vol. 32, no. 2, pp. 307–323, Jun. 2005.

[14] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.

[15] B. Matthiesen and E. A. Jorswieck, "Global sum rate optimal resource allocation for non-regenerative 3-way relay channels," in *Atto-Nets Workshop, IEEE Int. Conf. Ubiquitous Wireless Broadband*, 2017.