This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

Efficient Global Optimal Resource Allocation in Non-Orthogonal Interference Networks

Bho Matthiesen, *Student Member, IEEE,* and Eduard A. Jorswieck, *Senior Member, IEEE*

***Abstract*—Many resource allocation tasks are challenging global (i.e., non-convex) optimization problems. The main issue is that the computational complexity of these problems grows exponentially in the number of variables instead of polynomially as for many convex optimization problems. However, often the non-convexity stems only from a subset of variables. Conventional global optimization frameworks like monotonic optimization or DC programming [4] treat all variables as global variables and require complicated, problem specific decomposition approaches to exploit the convexity in some variables [5]. To overcome this challenge, we develop an easy-to-use algorithm that inherently differentiates between convex and non-convex variables, preserving the low computational complexity in the number of convex variables. Another issue with these widely used frameworks is that they may suffer from severe numerical problems. We discuss this issue in detail and provide a clear motivating example. The solution to this problem is to replace the traditional approach of finding an ε-approximate solution by the novel concept of *ε-essential feasibility*. The underlying algorithmic approach is called successive incumbent transcending (SIT) algorithm and builds the foundation of our developed algorithm. A further highlight of this algorithm is that it inherently treats fractional objectives making the use of Dinkelbach's iterative algorithm obsolete. Numerical experiments show a speed-up of four orders of magnitude over state-of-the-art algorithms and almost three orders of magnitude of additional speed-up over Dinkelbach's algorithm for fractional programs.**

***Index Terms*—Resource allocation, global optimization, successive incumbent transcending, essential feasibility, multi-way relay channel, simultaneous non-unique decoding, interference networks**

I. Introduction

Resource allocation is essential in most communication systems [6]. Practical systems usually use algorithms with no or only weak optimality guarantees for performance reasons. Nevertheless, asserting the quality of these algorithms requires the knowledge of the optimal solution to these problems. The general optimization problem

$$\max_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x}) \tag{P1}$$

with $\mathcal{C} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ covers a large class of resource allocation problems. A point $\boldsymbol{x}^* \in \mathcal{C}$ satisfying $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{C}$ is called a *global maximizer* of f. If $\boldsymbol{x}^*$ only satisfies this condition for all $\boldsymbol{x}$ in an open ε-neighborhood of $\boldsymbol{x}^*$ for some $\varepsilon > 0$, i.e., for all $\boldsymbol{x} \in \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon\} \cap \mathcal{C}$, it is called a *local minimizer*. The difficulty in obtaining a global optimal solution to (P1) is that all algorithms with polynomial computational complexity can at most obtain a local optimal solution. So, unless (P1) belongs to the class of optimization problems with the property that every local maximum is a global maximum,[1] solving (P1) has exponential computational complexity [7].

As an example, consider allocating the transmit power in an interference network. Albeit the capacity region of such an network is not known in general, the optimal decoder under the assumption of random codebooks[2] is known to be simultaneous non-unique decoding (SND) [8]. This leads to a global optimization problem where optimization is done jointly over the rates and powers. A close examination of this problem reveals that it is linear in the rate variables, i.e., for fixed power variables the problem can be solved in polynomial time [5], [9], [10]. Hence, the power variables are the only reason that the optimization problem is global and has exponential complexity. We call these variables *global* variables, while the remaining ones are named *non-global*.[3] The most popular solution approaches for global resource allocation problems are monotonic optimization and DC programming. Both frameworks treat all variables as global variables which often results in unnecessary high numerical complexity. Moreover, transforming typical resource allocation problems to fit into these frameworks often requires auxiliary variables, which, of course, further increases computational complexity. Instead, in this paper we present a novel framework that preserves the computational complexity of the non-global variables and does not require any auxiliary variables.

Another often neglected issue with these algorithms is the assumption of a robust feasible set, i.e., a set with no isolated points. If this assumption does not hold, which might be the case for resource allocation problems, it leads to serious numerical problems. We avoid this problem entirely by using

Parts of this paper were presented at the 19th IEEE International Workshop on Signal Processing Advances in Wireless Communications, 2018 [1], and the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing [2].

Bho Matthiesen is with the Chair for Communications Theory, Communications Laboratory, Technische Universität Dresden, Dresden, Germany (e-mail: bho.matthiesen@tu-dresden.de). Eduard A. Jorswieck was with Technische Universität Dresden and is now with the Department of Information Theory and Communication Systemes, TU Braunschweig, Germany (e-mail: e.jorswieck@tu-bs.de).

This work is supported in part by the German Research Foundation (DFG) in the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing," and under grant number JO 801/24-1.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author. The material includes the complete source code and is also available on GitHub [3].

[1] An important example are convex optimization problems where the objective of (P1) is a concave function and $\mathcal{C}$ is a convex set.

[2] The random codebooks are restricted to superposition coding and time sharing.

[3] A more precise definition is given in Section I-B.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

robust global optimization[4] [4], [14], [15]. The core idea is to shrink the feasible set by an infinitesimal amount and then solve a sequence of feasibility problems with a branch and bound (BB) procedure. This approach is called *successive incumbent transcending (SIT) scheme* and does not require any assumptions on the robustness of the feasible set $\mathcal{C}$ because it is designed to operate only on the accumulation points of $\mathcal{C}$. The result is a numerically much stabler procedure than could be obtained using classical monotonic or DC programming algorithms. Moreover, the SIT approach always provides a good feasible solution even if stopped prematurely. Instead, conventional algorithms usually outer approximate the solution rendering intermediate solutions almost useless (because they are infeasbile).

Fractional objectives, which occur, e.g., in the optimization of the energy efficiency (EE), can not be handled directly by monotonic optimization or DC programming. Instead, Dinkelbach's algorithm is used where the original problem is transformed into an auxiliary problem which is then solved several times with one of these frameworks [16]. However, this approach has several drawbacks. First, convergence to the optimal solution of the original problem is only guaranteed if the auxiliary problem is solved exactly. In practice, this algorithm also works well for approximate solutions but the numerical accuracy should be sufficiently high. Second, the auxiliary problem needs to be solved several times, and, finally, the stopping criterion is unrelated to the distance of the obtained approximate optimal value to the true optimum. Especially the first two are critical for global optimization since they increase the computation time significantly. Instead, our framework is able to deal directly with fractional objectives avoiding these problems entirely.

a) Related work: The SIT approach was developed by Hoang Tuy in [4], [14], [15] and, to the best of our knowledge, has not been adopted for resource allocation problems yet. However, the importance of robust feasible sets has been noted in [17] where beamforming in a cognitive radio network is solved with DC programming. In [18], the basic principle of the SIT approach is used to solve a multi-objective optimization problem (MOP).

Instead, decomposition approaches [19] are widely used. For example, in [20] the design of linear transceivers for multicarrier multiple-input multiple-output channels is considered. This challenging non-convex problem is solved by primal decomposition into a convex outer problem and inner problems with closed-form solutions. The authors of [21] combine successive convex approximation and primal decomposition to solve the sum rate maximization problem with Quality of Service (QoS) constraints for interfering broadcast channels with first order optimality. A distributed algorithm for coordinated beamforming in multicell multigroup multicast systems is developed in [22] based on primal decomposition and semidefinite relaxation. In [23] the partly convex-monotone structure of utility maximization problems in broadcast and interference channels is exploited via a BB procedure where branching is only performed over the global variables. This approach is similar to our framework but the problem setting is more specific and the optimization is over a convex set, i.e., the SIT approach is not needed. In [5], we solve a special case of the problems considered here by decomposing it into an inner linear and an outer monotonic program which is solved by the Polyblock algorithm [24].

Resource allocation for interference networks is mostly done under orthogonality assumptions to fall back to the noise-limited case. For example, in [25] monotonic optimization was first used to maximize the throughput in an interference network where interference is treated as noise. In [26] a monotonic optimization based framework for resource allocation in coordinated multi-cell systems is presented. Optimization of the EE in interference networks under the assumption that interference is treated as noise is considered in [16], [27]. In [28] energy-efficient coordinated beamforming in multi-cell, multi-user systems is considered under the assumption of realistic power consumption models. Energy-efficient resource allocation in OFDMA systems with and without wireless power transfer is studied in [29] and [30], respectively. This is, naturally, just a very incomplete list of papers dealing with resource allocation problems in interference networks under orthogonality constraints. Of course, the reasons for making orthogonality assumptions are manifold. One reason is surely that they are considerably easier to optimize than more involved systems relying on non-orthogonal communication schemes (e.g. SND). The algorithm we develop here is well suited for exactly these more involved resource allocation problems. For further references on resource allocation problems please refer to Section II-A.

A. Major Contributions

The key contributions of this paper are the following:

- The concept of essential feasibility is introduced and its relevance for resource allocation problems is shown with a simple example. The numerical problems of conventional global optimization algorithms are discussed and the SIT scheme is introduced as a remedy for the these issues.
- We extend the SIT algorithm as developed in [4] into an optimization framework able to deal with fractional objectives, non-convex variables, and multiple constraints. Specifically, we design an adaptive BB procedure that only branches over the global variables and leverages the power of commercially available state-of-the-art solvers to handle the non-global variables. To this end, we identify two different sets of sufficient conditions on the optimization problem. Fractional objectives are directly treated within the developed algorithm making the use of Dinkelbach's iterative algorithm obsolete.
- The framework is applied to optimize the throughput and energy efficiency of a multi-way relay channel (MWRC). Extensive numerical results show a speed-up over state-of-the-art algorithms of four orders of magnitude for throughput maximization and almost three

[4]There are, at least, two different meanings of "robust optimization:" the one discussed here that is robust against the effects of non-robust feasible sets and small changes in the tolerances, and the one that provides robustness against uncertainty in the input data [11], e.g., robust beamforming [12], or robust monotonic optimization [13].

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

orders of magnitude over Dinkelbach's Algorithm for EE maximization.
- Code and data are made publicly available on GitHub [3]. This allows other researchers to easily verify our results and adapt the SIT algorithm for their own research.

B. Notation & Preliminaries

A vector $\boldsymbol{x} \in \mathbb{R}^n$ with components $(x_1, \ldots, x_n)$ is said to *dominate* another vector $\boldsymbol{y} \in \mathbb{R}^n$, i.e., $\boldsymbol{y} \leq \boldsymbol{x}$, if $y_i \leq x_i$ for all $i = 1, \ldots, n$. For $\boldsymbol{a} \leq \boldsymbol{b}$, the set $[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \,|\, \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$ is called a *box*. A function $f : \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}$ is *increasing* if $f(\boldsymbol{x}') \leq f(\boldsymbol{x})$ whenever $\boldsymbol{x}' \leq \boldsymbol{x}$, and *decreasing* if $-f$ is increasing. It is called *mixed monotonic* if it is increasing in the variables $(x_i)_{i\in\mathcal{I}}$ and decreasing in $(x_i)_{i\in\{1,2,\ldots,n\}\setminus\mathcal{I}}$ for some index set $\mathcal{I}$. The functions $f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})$ are called *jointly mixed monotonic* if all functions are mixed monotonic with respect to the same index set $\mathcal{I}$. A *common minimizer (maximizer)* of the functions $f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})$ over the set $\mathcal{X}$ is any $\boldsymbol{x}^*$ that satisfies $\boldsymbol{x}^* \in \bigcap_{i=1}^n \arg\min_{\boldsymbol{x}\in\mathcal{X}} f_i(\boldsymbol{x})$ ($\boldsymbol{x}^* \in \bigcap_{i=1}^n \arg\max_{\boldsymbol{x}\in\mathcal{X}} f_i(\boldsymbol{x})$).

A set $\mathcal{G} \subseteq \mathbb{R}^n_{\geq 0}$ is said to be *normal* if for $0 \leq \boldsymbol{x}' \leq \boldsymbol{x}$, $\boldsymbol{x} \in \mathcal{G} \Rightarrow \boldsymbol{x}' \in \mathcal{G}$, and *normal in a box* $[\boldsymbol{a}, \boldsymbol{b}]$ if the previous implication only holds for $\boldsymbol{a} \leq \boldsymbol{x}' \leq \boldsymbol{x} \leq \boldsymbol{b}$. A set $\mathcal{H} \subseteq \mathbb{R}^n_{\geq 0}$ is called *conormal* if $\boldsymbol{x} + \mathbb{R}^n_{\geq 0} \subseteq \mathcal{H}$ whenever $\boldsymbol{x} \in \mathcal{H}$, and *conormal in a box* $[\boldsymbol{a}, \boldsymbol{b}]$ if for $\boldsymbol{b} \geq \boldsymbol{x}' \geq \boldsymbol{x} \geq \boldsymbol{a}$, $\boldsymbol{x} \in \mathcal{H} \Rightarrow \boldsymbol{x}' \in \mathcal{H}$ [4, Sec. 11.1.1]. Let $\mathcal{A} \subseteq \mathbb{R}^n$ and $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}$. Then, $\operatorname{proj}_{\boldsymbol{x}} \mathcal{A} = \{\boldsymbol{x} \,|\, (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A} \text{ for some } \boldsymbol{y}\}$, i.e., the projection of $\mathcal{A}$ onto the $\boldsymbol{x}$ coordinates; $\operatorname{diam} \mathcal{A}$ is the *diameter* of $\mathcal{A}$, i.e., the maximum distance between two points in $\mathcal{A}$; and $\mathcal{A}_{\tilde{\boldsymbol{x}}} = \{\boldsymbol{y} \,|\, (\tilde{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{A}\}$ is called the $\tilde{\boldsymbol{x}}$*-section* of $\mathcal{A}$.

Finally, consider (P1) and let its optimal value be v(P1). To make the previously introduced notion of global and non-global variables more precise, split $\boldsymbol{x}$ into two vectors $\boldsymbol{y} = (x_i)_{i\in\mathcal{I}}$ and $\boldsymbol{z} = (x_i)_{i\in\{1,2,\ldots,n\}\setminus\mathcal{I}}$ and consider a modified version of (P1) where we only optimize over $\boldsymbol{y}$ for some fixed $\boldsymbol{z}$, i.e., $\max_{\boldsymbol{y}\in\mathcal{C}_{\boldsymbol{z}}} f(\boldsymbol{y}, \boldsymbol{z})$. If there exists an algorithm to solve this optimization problem with computational complexity significantly less than required for solving the global part of (P1),[5] the variables $\boldsymbol{y}$ and $\boldsymbol{z}$ are denoted as *non-global* and *global* variables, respectively.

C. Outline

The remaining part of this paper is organized as follows. In the next section, we formally state the considered optimization problem and motivate it in the context of resource allocation problems in wireless interference networks. Section III introduces important mathematical preliminaries including the SIT scheme that forms the basis of the proposed algorithm. These developed principles are then applied in Section IV to construct Algorithm 2, one of the main contributions of this paper. In Section V, we apply the developed framework to a specific resource allocation problem that is used to benchmark our method against the state-of-the-art. Finally, we give our conclusions in Section VI.

[5]Ideally, the computational complexity for solving the non-global part of (P1) grows polynomially in the number of non-global variables. This is, for example, the case if (P1) is linear in $\boldsymbol{y}$ [9], [10]. It also holds in many cases were (P1) is convex in $\boldsymbol{y}$ [31]. However, there exist convex optimization problems that are not solvable in polynomial time [32, Sect. 5.4] but still have computational complexity significantly less than general non-convex optimization problems [33]. Nevertheless, it might be helpful to think of the terms "global" and "non-global" variables as synonyms for "non-convex" and "convex" variables, respectively.

II. Problem Statement

We consider the following global optimization problem

$$\begin{cases} \max\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & \dfrac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \\ \text{s. t.} & g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m, \end{cases} \tag{P2}$$

with global variables $\boldsymbol{x}$ and non-global variables $\boldsymbol{\xi}$. The functions $\{g_i^-(\boldsymbol{x})\}$ are required to have a common maximizer over every box $[\underline{\boldsymbol{x}}, \bar{\boldsymbol{x}}] \subseteq \mathcal{M}_0$ with $\mathcal{M}_0$ being a box enclosing the $\boldsymbol{x}$ dimensions of $\mathcal{C}$, i.e., $\mathcal{M}^0 \supset \operatorname{proj}_{\boldsymbol{x}} \mathcal{C}$. A sufficient condition for the existence of this common maximizer is that the functions $g_i^-(\boldsymbol{x})$ are jointly mixed monotonic. This includes the case where all functions are either increasing or decreasing. Further, we assume the functions $f^-(\boldsymbol{x},\boldsymbol{\xi})$, $g_i^+(\boldsymbol{x},\boldsymbol{\xi})$, $i = 1, \ldots, m$, to be lower semi-continuous (l.s.c.), the functions $f^+(\boldsymbol{x},\boldsymbol{\xi})$, $g_i^-(\boldsymbol{x})$, $i = 1, \ldots, m$, to be upper semi-continuous, and, without loss of generality (w.l.o.g.), $f^-(\boldsymbol{x},\boldsymbol{\xi}) > 0$.

The goal of this paper is to design a numerically stable BB procedure to solve (P2) that preserves the computational complexity in the non-global variables $\boldsymbol{\xi}$. This requires additional assumptions to those stated above. Specifically, we have identified two different sets of technical requirements that are stated below. Both of these cases contain conditions that depend on a constant γ which will hold the current best known value in the developed algorithm. We will discuss the domain of γ after the definition of both cases below.

Case A (DC problems): If $\mathcal{C}$ is a closed convex set and $\gamma f^-(\boldsymbol{x},\boldsymbol{\xi}) - f^+(\boldsymbol{x},\boldsymbol{\xi})$, $g_1^+(\boldsymbol{x},\boldsymbol{\xi})$, $\ldots$, $g_m^+(\boldsymbol{x},\boldsymbol{\xi})$ are jointly convex in $(\boldsymbol{x},\boldsymbol{\xi})$ for all γ, problem (P2) resembles a DC optimization problem but with fractional objective and additional non-DC variables. ◊

Case B (Separable problems): Let $\mathcal{C} = \mathcal{X} \times \Xi$ such that $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$ with Ξ being a closed convex set, and let each function of $(\boldsymbol{x},\boldsymbol{\xi})$ be separable in the sense that $h(\boldsymbol{x},\boldsymbol{\xi}) = h_x(\boldsymbol{x}) + h_\xi(\boldsymbol{\xi})$. Further, let the functions $\gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi})$, $g_{1,\xi}^+(\boldsymbol{\xi})$, $\ldots$, $g_{m,\xi}^+(\boldsymbol{\xi})$ be convex in $\boldsymbol{\xi}$ for all γ, and let the functions $\gamma f_x^-(\boldsymbol{x}) - f_x^+(\boldsymbol{x})$, $g_{1,x}^+(\boldsymbol{x})$, $\ldots$, $g_{m,x}^+(\boldsymbol{x})$ have a common minimizer over $\mathcal{X} \cap \mathcal{M}$ for every box $\mathcal{M} \subseteq \mathcal{M}_0$ and all γ. Finally, let the function $\gamma f_x^-(\boldsymbol{x}) - f_x^+(\boldsymbol{x})$ be either increasing for all γ with $\mathcal{X}$ being a closed normal set in some box, or decreasing for all γ with $\mathcal{X}$ being a closed conormal set in some box. ◊

Remark 1: Separable problems often lead to linear auxiliary optimization problems with less variables instead of, typically, convex problems for Case A. They usually have lower computational complexity than DC problems and, thus, if the problem at hand falls into both cases, it is usually favorable to consider it as a separable problem.

With both cases defined we can continue our discussion of γ. First, observe that γ only appears as a factor to $f^-(\boldsymbol{x},\boldsymbol{\xi})$. Thus,

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

its value is only relevant if (P2) is a fractional program, i.e., if $f^-(\boldsymbol{x},\boldsymbol{\xi})$ is not constant. In that case, the only relevant property of γ is its sign and whether it may change during the algorithm. For example, in Case A the function $\gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi})$ is convex if $f_\xi^+(\boldsymbol{\xi})$ is concave and $\gamma f_\xi^-(\boldsymbol{\xi})$ is convex. The latter is the case if $\gamma \geq 0$ and $f_\xi^-(\boldsymbol{\xi})$ is convex, or if $\gamma \leq 0$ and $f_\xi^-(\boldsymbol{\xi})$ is concave. Thus, in most cases, we should ensure that the sign of γ is constant. In general, γ may take values between some γ_0 and $v(P2)+\eta$ for some small $\eta > 0$. The lower end of the range γ_0 is either the objective value of (P2) for some preliminary known nonisolated feasible point $(\boldsymbol{x},\boldsymbol{\xi})$ or an arbitrary value satisfying $\gamma_0 \leq \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})}$ for all feasible $(\boldsymbol{x},\boldsymbol{\xi})$. This implies, e.g., that γ is non-negative if $f^+(\boldsymbol{x},\boldsymbol{\xi})$ is non-negative. Otherwise, it might be necessary to find a nonisolated feasible point such that $f^+(\boldsymbol{x},\boldsymbol{\xi}) \geq 0$ or transform the problem.

A. Application Example: Resource Allocation in Interference Channels

Determining an achievable rate region of a communication network usually involves two steps: first, characterizing the achievable rate region with information theoretical tools, and, second, finding Pareto-optimal resource allocations. The corresponding optimization problem in many Gaussian interference networks is

$$\begin{cases} \max\limits_{\boldsymbol{p},\boldsymbol{R}} & f(\boldsymbol{p},\boldsymbol{R}) \\ \text{s.t.} & \boldsymbol{a}_i^T\boldsymbol{R} \leq \log\left(1+\dfrac{\boldsymbol{b}_i^T\boldsymbol{p}}{\boldsymbol{c}_i^T\boldsymbol{p}+\sigma_i}\right), \quad i=1,\dots,n \\ & \boldsymbol{R} \geq 0, \quad \boldsymbol{p} \in [\boldsymbol{0},\boldsymbol{P}] \end{cases} \tag{P3}$$

for some performance function $f(\boldsymbol{p},\boldsymbol{R})$ and positive vectors $\boldsymbol{a}_i,\boldsymbol{b}_i,\boldsymbol{c}_i \geq 0$, $i=1,\dots,n$. The optimization variables $\boldsymbol{p}$ are the allocated transmit powers and $\boldsymbol{R}$ are the achievable transmission rates for asymptotically error-free communication. Usually, the vectors $\boldsymbol{b}_i$ and $\boldsymbol{c}_i$ represent the effective channel gain, σ_i is the variance of the Gaussian noise observed at receiver i, $\boldsymbol{a}_i$ is a sparse vector where the non-zeros entries are small integers (mostly ones), and $\boldsymbol{P}$ are the maximum transmit powers. Applications of this model include multi-cell communication systems [26], heterogeneous dense small-cell networks [34], cognitive radio [35], and digital subscriber line systems [36]. For the sake of simplicity, we assume the signal to interference plus noise ratio (SINR) in the right-hand side (RHS) expressions of the constraints to be linear fractions of the transmit powers. We note that this is not always the case and that our framework is not limited to this case.

The feasible set of (P3) belongs to the class of considered problems. We identify the global variables as $\boldsymbol{p}$ and the non-globals as $\boldsymbol{R}$. The rate constraints are equivalent to

$$\boldsymbol{a}_i^T\boldsymbol{R} + \log\left(\boldsymbol{c}_i^T\boldsymbol{p}+\sigma_i\right) - \log\left(\left(\boldsymbol{b}_i^T+\boldsymbol{c}_i^T\right)\boldsymbol{p}+\sigma_i\right) \leq 0.$$

Since $\boldsymbol{a}_i,\boldsymbol{b}_i,\boldsymbol{c}_i \geq 0$ for all i, the linear function $\boldsymbol{a}_i\boldsymbol{R}$ and the log-functions are increasing. Thus, we can identify $g_i^+(\boldsymbol{p},\boldsymbol{R})$ and $g_i^-(\boldsymbol{p})$ as:

$$\begin{aligned} g_i^+(\boldsymbol{p},\boldsymbol{R}) &\coloneqq \boldsymbol{a}_i^T\boldsymbol{R} - \log\left(\left(\boldsymbol{b}_i^T+\boldsymbol{c}_i^T\right)\boldsymbol{p}+\sigma_i\right) \\ g_i^-(\boldsymbol{p}) &\coloneqq -\log\left(\boldsymbol{c}_i^T\boldsymbol{p}+\sigma_i\right). \end{aligned} \tag{1}$$

The functions $g_i^-(\boldsymbol{p})$ are decreasing, and, thus, are jointly maximized over the box $[\underline{\boldsymbol{p}},\bar{\boldsymbol{p}}]$ by $\underline{\boldsymbol{p}}$. Further, $g_i^+(\boldsymbol{p},\boldsymbol{R})$ is separable in $\boldsymbol{p}$ and $\boldsymbol{R}$ with $g_{i,R}^+(\boldsymbol{R})$ being linear and $g_{i,p}^+(\boldsymbol{p})$ decreasing and convex in $\boldsymbol{p}$. Thus, depending on $f(\boldsymbol{p},\boldsymbol{R})$ problem (P3) qualifies for both, Cases A and B.

The complete Pareto boundary is characterized by all solutions to the MOP (P3) with the vector objective $f(\boldsymbol{p},\boldsymbol{R}) = [R_1,R_2,\dots,R_K]$ [37]. Several approaches exist to transform this into a scalar optimization problem. Two very popular are the scalarization and the rate profile approach[6] [39].

a) Weighted Sum Rate: In the scalarization approach, the weighted sum of the objectives is optimized, i.e., $f(\boldsymbol{p},\boldsymbol{R}) = f^+(\boldsymbol{p},\boldsymbol{R}) \coloneqq \boldsymbol{w}^T\boldsymbol{R}$. Obviously, with this objective, (P3) belongs to the class of separable problems since Case B is satisfied. Varying the weights between 0 and 1 with $\sum_k w_k = 1$ characterizes the convex hull of the Pareto boundary of the achievable rate region. This is by far the most widely used performance metric for power control in wireless communication systems due its clear operational meaning: with all weights $w_k = 1$, $f^+(\boldsymbol{p},\boldsymbol{R})$ is the total *throughput* in the network. Moreover, in several networks this condition leads to the characterization of the stability region. In this queueing theoretic setting, choosing the weights proportional to the queue lengths prioritizes longer queues and stabilizes the network [40], [41].

Many instances of the weighted sum rate maximization problem considered in the literature are either convex optimization problems [42]–[44], or have a rather good-natured rate region allowing to eliminate the rates $\boldsymbol{R}$ from the problem such that the resulting global optimization is only over the powers [25], [26], [45]. If neither is the case, variable reduction techniques are necessary to keep the computational complexity at a reasonable level. In [5], the problem at hand is transformed into a monotonic optimization problem [24] where the objective is a linear program with the transmit power as parameter. This approach is considerably slower, theoretically more involved, and less versatile than the proposed method.

b) Rate Profile Approach: The utility profile approach finds the intersection of a ray in the direction $\boldsymbol{w}$ and the Pareto boundary of the achievable rate region, i.e.,

$$\begin{cases} \max\limits_{\boldsymbol{p},\boldsymbol{R},t} & t \\ \text{s.t.} & \boldsymbol{a}_i^T\boldsymbol{R} \leq \log\left(1+\dfrac{\boldsymbol{b}_i^T\boldsymbol{p}}{\boldsymbol{c}_i^T\boldsymbol{p}+\sigma_i}\right), \quad i=1,\dots,n \\ & \boldsymbol{R} \geq t\boldsymbol{w}, \quad \boldsymbol{p} \in [\boldsymbol{0},\boldsymbol{P}] \end{cases} \tag{P4}$$

for some $\boldsymbol{w} \geq 0$. W.l.o.g. one can assume $\|\boldsymbol{w}\| = 1$. By varying the direction of the ray, the complete Pareto boundary can be characterized. It is easily verified that (P4) satisfies Case B.

Problem (P4) is the epigraph form of (P3) with $f(\boldsymbol{p},\boldsymbol{R}) = \min_k \frac{R_k}{w_k}$ as long as $\boldsymbol{w} \neq \boldsymbol{0}$. Since $f(\boldsymbol{p},\boldsymbol{R})$ is continuous and concave, it also satisfies Case B. Besides reducing the number of variables, this reformulation also makes it apparent that the rate profile approach is equivalent to the *weighted max-min fairness* performance function. For $\boldsymbol{w} = \boldsymbol{1}$, it puts very strong emphasis on the weakest user and often results in

[6]This approach is also known as rate balancing [38].

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

low spectral efficiency. However, in some cases this approach might be computationally less challenging than the scalarization approach as we will discuss below.

c) Global Energy Efficiency: The global energy efficiency (GEE) is the most widely used metric to measure the network energy efficiency, a key performance metric in 5G and beyond networks [27], [46]. It is defined as the benefit-cost ratio of the total network throughput and the associated power consumption, i.e.,

$$\mathrm{GEE} = \frac{\sum_k R_k}{\boldsymbol{\phi}^T \boldsymbol{p} + P_c}, \tag{2}$$

where $\boldsymbol{\phi} \geq 1$ are the inverses of the power amplifier efficiencies and P_c is the total circuit power necessary to operate the network. Using the GEE as performance metric results in a fractional programming problem [47]–[50]. Usually, it is solved iteratively with Dinkelbach's Algorithm [48] where, in each iteration, an auxiliary optimization problem is solved. This auxiliary problem generally has very similar properties to the weighted sum rate maximization problem, i.e., for interference networks it is a global optimization problem with exponential complexity. This inner problem must be solved several times with high numerical accuracy for the convergence guarantees of Dinkelbach's Algorithm to hold [16].

Instead, our proposed algorithm allows to solve these fractional programs directly resulting in significantly lower complexity. Consider the generic resource allocation problem (P3) with objective (2). We identify $f^+(\boldsymbol{p}, \boldsymbol{R}) = \sum_k R_k$ and $f^-(\boldsymbol{p}, \boldsymbol{R}) = \boldsymbol{\phi}^T \boldsymbol{p} + P_c$ and observe that both are linear functions. Hence, $\gamma f^-(\boldsymbol{x}, \boldsymbol{\xi}) - f^+(\boldsymbol{x}, \boldsymbol{\xi})$ is convex for all γ and Case A is satisfied. Of course, $f^+(\boldsymbol{p}, \boldsymbol{R})$ and $f^-(\boldsymbol{p}, \boldsymbol{R})$ are also separable, and because the GEE is non-negative, $\gamma \geq 0$ and $\gamma f_p^-(\boldsymbol{p})$ is increasing. But since $g_{i,p}^+(\boldsymbol{p})$ is decreasing, the functions $\gamma f_p^-(\boldsymbol{p})$, $g_{1,p}^+(\boldsymbol{p})$, $\ldots$, $g_{m,p}^+(\boldsymbol{p})$ do not have a common minimizer over $\mathcal{X} \cap \mathcal{M}$ for any box $\mathcal{M}$ and Case B does not apply. However, observe that instead of the choice in (1), we can also identify $g_i^+(\boldsymbol{p}, \boldsymbol{R})$ and $g_i^-(\boldsymbol{p})$ as

$$\begin{aligned} g_i^+(\boldsymbol{p}, \boldsymbol{R}) &\coloneqq \boldsymbol{a}_i^T \boldsymbol{R} + \log\big(\boldsymbol{c}_i^T \boldsymbol{p} + \sigma_i\big) \\ g_i^-(\boldsymbol{p}) &\coloneqq \log\big(\big(\boldsymbol{b}_i^T + \boldsymbol{c}_i^T\big) \boldsymbol{p} + \sigma_i\big) . \end{aligned} \tag{3}$$

The functions $g_i^+(\boldsymbol{p}, \boldsymbol{R})$ are separable in $\boldsymbol{p}$ and $\boldsymbol{R}$, and $g_{i,p}^+(\boldsymbol{p})$ and $g_i^-(\boldsymbol{p})$ are increasing in $\boldsymbol{p}$. Hence, $\gamma f_p^-(\boldsymbol{p})$, $g_{1,p}^+(\boldsymbol{p})$, $\ldots$, $g_{m,p}^+(\boldsymbol{p})$ have a common minimizer over $\mathcal{X} \cap \mathcal{M}$ for every box $\mathcal{M}$ and Case B applies. In general, (1) will result in tighter bounds and, thus, faster convergence. On the other hand, Case B has significantly lower numerical complexity than Case A (cf. Remark 1) which should compensate for this drawback. However, a final assessment is only possible through numerical experimentation which will be carried out in Section V.

d) Energy Efficiency Region: Similar to the GEE, the individual energy efficiency of link k is defined as the benefit-cost ratio of the link's throughput divided by the link's power consumption, i.e., $\mathrm{EE}_k = \frac{R_k}{\phi_k p_k + P_{c,k}}$. Analogue to the achievable rate region, there is also an achievable EE region whose Pareto boundary is the solution to the MOP (P3) with objective $f(\boldsymbol{p}, \boldsymbol{R}) = [\mathrm{EE}_1, \ldots, \mathrm{EE}_K]$ [51]. The same approaches as discussed earlier can be used to transform this MOP into a scalar optimization problem. In this case, the rate profile (or rather utility profile) approach should be favored because the scalarization approach leads to a sum-of-ratios problem. The sum-of-ratios problem is one of the most difficult fractional programs and known to be essentially $\mathcal{NP}$-hard [52], [53]. While for other EE problems (e.g., the GEE) first-order optimal solutions are usually observed to be globally optimal [16] this does not hold for the sum-of-ratios case [34].

Instead, with the utility profile approach the objective in (P3) is the weighted minimum EE $f(\boldsymbol{p}, \boldsymbol{R}) = \min_k \frac{\mathrm{EE}_k}{w_k}$. This is, as is the sum-of-ratios problem, a multi-ratio optimization problem but numerically less challenging. Another benefit of using the utility profile approach over the scalarization approach is that the EE region is non-convex and the scalarization approach only obtains the convex hull of the Pareto region [16]. The general approach to solving such an optimization problem is to use the Generalized Dinkelbach's Algorithm [16], [51], [54]. As in the GEE case, our proposed method solves this problem with much lower complexity. However, some minor modifications of the algorithm are necessary to deal with this kind of problem. Please refer to Section IV-C for more details.

e) Proportional Fairness: Another well known performance function is proportional fairness where the objective is the product of the rates, i.e., $\prod_k R_k$ [55]. The operating point achieved by this metric is usually almost as fair as the one obtained by max-min fairness but achieves significantly higher throughput. Observe that the objective is log-concave. Thus, we can determine the proportional fair operating point by solving (P3) with objective $f(\boldsymbol{p}, \boldsymbol{R}) = \sum_k \log(R_k)$.

III. Robust Approach to Global Optimization

In this section, we introduce some mathematical preliminaries essential for developing the proposed algorithm. This is done by means of a simpler problem than (P2) to ease the exposition. Specifically, we consider the optimization problem

$$\max_{\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]} \; f(\boldsymbol{x}) \quad \text{s.t.} \quad g_i(\boldsymbol{x}) \leq 0, \; i = 1, 2, \ldots, m \tag{P5}$$

where $f, g_1, g_2, \ldots, g_m$ are non-convex continuous real-valued functions and $\boldsymbol{a}, \boldsymbol{b}$ are real-valued vectors satisfying $\boldsymbol{a} \leq \boldsymbol{b}$. This is a general non-convex optimization problem with possibly quite complicated feasible set.

A. The Issue with ε-Approximate Solutions

Most current solution methods for this problem are devised to compute a solution $\bar{\boldsymbol{x}}(\varepsilon)$ of the ε-relaxed problem

$$\max_{\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]} \; f(\boldsymbol{x}) \quad \text{s.t.} \quad g_i(\boldsymbol{x}) \leq \varepsilon, \; i = 1, 2, \ldots, m, \tag{P6}$$

This solution $\bar{\boldsymbol{x}}(\varepsilon)$ is usually accepted as an *ε-approximate optimal solution* to (P5) since $\bar{\boldsymbol{x}}(\varepsilon)$ is almost feasible for small $\varepsilon > 0$ and tends to a feasible solution as $\varepsilon \to 0$. Since $f(\bar{\boldsymbol{x}}(\varepsilon))$ also tends to the optimal value $v(P5)$, $f(\bar{\boldsymbol{x}}(\varepsilon))$ should be close to $v(P5)$ for a sufficiently small $\varepsilon = \varepsilon_0$. The problem with this approach is that ε_0 is, in general, unknown and hard to determine. Thus, the obtained solution and $f(\bar{\boldsymbol{x}}(\varepsilon))$ can be quite far away from the true optimal value even for a small ε [14]. This is apparent from the following example.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

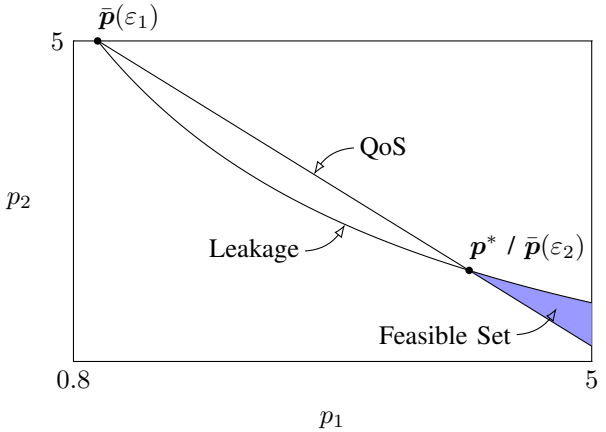

Fig. 1. Feasible set of Example 1 with optimal solution $\boldsymbol{p}^*$ and ε-approximate solutions $\bar{\boldsymbol{p}}(\varepsilon_1)$ and $\bar{\boldsymbol{p}}(\varepsilon_2)$ for $\varepsilon_1 = 10^{-3}$ and $\varepsilon_2 = 10^{-4}$. The ε_1-approximate solution is quite far away from the true optimum.

Example 1 (Issues with ε-approximate solutions): Consider a Gaussian 2-user single-input single-output (SISO) multiple-access channel (MAC) with channel gains h_1 and h_2, transmit powers $p_1 \leq P_1$ and $p_2 \leq P_2$, and a minimum total throughput of Q, i.e., $\log_2(1 + |h_1|^2 p_1 + |h_2|^2 p_2) \geq Q$. Transmitter i, $i = 1, 2$, is eavesdropped by a single antenna adversary over a channel g_i. The eavesdroppers are only able to overhear one of the transmitters and do not cooperate with each other [56]. The total information leakage to the eavesdroppers is limited by L, i.e., $\log_2(1 + |g_1|^2 p_1) + \log_2(1 + |g_2|^2 p_2) \leq L$. The transmit power of transmitter 1 should be minimized without violating these constraints. The resulting feasible set for $|h_1|^2 = |h_2|^2 = 10$, $|g_1|^2 = \frac{1}{2}$, $|g_2|^2 = 1$, $Q = \log_2(61)$, and $L = \log_2(8.99)$ is shown in Fig. 1. The true optimum solution is $\boldsymbol{p}^* = (4.00665, 1.99335)$ with $f(\boldsymbol{p}^*) = 4.00665$, while the ε-approximate solution for $\varepsilon_1 = 10^{-3}$ is $\bar{\boldsymbol{p}}(\varepsilon_1) = (0.995843, 5)$ with $f(\bar{\boldsymbol{p}}(\varepsilon_1)) = 0.995843$. This is, obviously, quite far away from both, the optimal solution and value. Instead, the ε-approximate solution obtained for $\varepsilon_2 = 10^{-4}$ is $\bar{\boldsymbol{p}}(\varepsilon_2) = (4.00541, 1.99417)$ with $f(\bar{\boldsymbol{p}}(\varepsilon_2)) = 4.00541$. $\Diamond$

Thus, even for well behaved problems the ε-relaxation approach might fail. A slightly better approach to the approximate optimal solution of (P5) is that of an *η-optimal solution*. A feasible vector $\bar{\boldsymbol{x}}$ is called η-optimal solution if it satisfies $f(\bar{\boldsymbol{x}}) \geq f(\boldsymbol{x}) - \eta$ for all feasible $\boldsymbol{x}$. The downside of this approach is that it does not converge in finitely many steps if the optimal solution is an isolated feasible point, i.e., a point at the center of a ball containing no other feasible points [14].

Example 2 (Isolated optimal solution): Consider Example 1 again. With $L = \log_2(9)$ there exists an isolated feasible point $\boldsymbol{p} = (1, 5)$ (close to $\bar{\boldsymbol{p}}(\varepsilon_1)$ in Fig. 1) that also happens to be the optimal solution. However, apart from the algorithmic difficulties in computing it, it might also be unstable under small pertubations of the data and is, thus, quite impractical from an engineering point of view. $\Diamond$

B. ε-Essential Feasibility

A common approach to this dilemma is to assume that the feasible set $\mathcal{F}$ is robust, i.e., it satisfies $\mathcal{F}^* = \mathrm{cl}(\mathrm{int}\,\mathcal{F})$ where cl and int denote the closure and interior, respectively. Unfortunately, this condition is generally very hard to check, so that, in practice, we have to deal with feasible sets where we do not know a priori whether they are robust or not.

This motivates the concept of ε-essential optimality developed by Tuy [4], [14], [15]. A solution $\boldsymbol{x}^* \in \mathcal{F}^*$ is called *essential optimal solution* of (P5) if $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ for all $x \in \mathcal{F}^*$. 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 (P5) is said to be *essential (ε, η)-optimal* if it satisfies

$$f(\boldsymbol{x}^*) + \eta \geq \sup\{f(\boldsymbol{x}) | \boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}],\ \forall i : g_i(\boldsymbol{x}) \leq -\varepsilon\}, \quad (4)$$

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

C. Successive Incumbent Transcending Scheme

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 (P5). The core problem in the SIT scheme is, given a real number γ, to check whether (P5) has a nonisolated feasible solution $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \geq \gamma$, or, else, establish that no such ε-essential feasible $\boldsymbol{x}$ exists. Given that this subproblem is solved within finitely many steps, Algorithm 1 converges to the global optimal solution within finitely many iterations. 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) point even if terminated prematurely. Instead, conventional algorithms usually outer approximate the solution rendering intermediate solutions almost useless.

Algorithm 1 SIT Algorithm [4, Sect. 7.5.1].

Step 0	Initialize $\bar{\boldsymbol{x}}$ with the best known nonisolated feasible solution and set $\gamma = f(\bar{\boldsymbol{x}}) + \eta$; otherwise do not set $\bar{\boldsymbol{x}}$ and choose $\gamma \leq f(\boldsymbol{x})\ \forall \boldsymbol{x} \in \mathcal{F}$.
Step 1	Check if (P5) has a nonisolated feasible solution $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \geq \gamma$; 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 $\bar{\boldsymbol{x}}$ is set, it is an essential (ε, η)-optimal solution; else Problem (P5) is ε-essential infeasible.

Consider the optimization problem

$$\min_{\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]} \max_{i=1,2,\ldots,m} g_i(\boldsymbol{x}) \quad \text{s.t.} \quad f(\boldsymbol{x}) \geq \gamma \quad (P7)$$

where we interchanged objective and constraints of (P5). Very often $f(x)$ has, or could be modified easily to have, nice properties like being concave or increasing, such that the feasible set of (P7) is *nice*, i.e., it is robust and a feasible point can be computed efficiently using an adaptive BB procedure [4, Prop. 6.2]. The following proposition, which is an adapted version of [4, Prop. 7.13], establishes a duality between (P5) and (P7) in the sense that the feasibility problem in Step 1 of Algorithm 1 is equivalent to solving (P7).

Proposition 1: For every $\varepsilon > 0$, the ε-essential optimal value of (P5) is less than γ if the optimal value of (P7) is greater than $-\varepsilon$.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

Proof (adapted from [4, Prop. 7.13]): If the optimal value of (P7) is greater than $-\varepsilon$, then any $\boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}]$ such that $g_i(\boldsymbol{x}) \leq -\varepsilon$, for all $i = 1, 2, \ldots, m$, must satisfy $f(\boldsymbol{x}) < \gamma$. Hence, by the compactness of the feasible set of (P5), $\max\{f(\boldsymbol{x}) | \boldsymbol{x} \in [\boldsymbol{a}, \boldsymbol{b}], \forall i : g_i(\boldsymbol{x}) \leq -\varepsilon\} < \gamma$. ■

Observe that every point $\boldsymbol{x}'$ in the feasible set of (P7) with objective value less or equal than zero is also a feasible point of (P5) with objective value better than γ. Thus, we can solve (P5) sequentially by solving (P7) with a BB method. Each time this BB algorithm finds a feasible point $\boldsymbol{x}'$ with objective value less or equal than zero (in (P7)), the current best value γ is updated with the objective value of $\boldsymbol{x}'$ in the original problem (P5) (plus the tolerance η). Then, the BB solver continues solving (P7) with updated feasible set until it either finds a new point to update γ or establishes that no solution to (P7) with objective value less or equal than $-\varepsilon$ exists. By virtue of Proposition 1, the last feasible point $\boldsymbol{x}'$ that was used to update γ is an (ε, η)-optimal solution of (P5). This observation will be formalized in Proposition 4.

In the next section, we return our attention to (P2) and use Algorithm 1 and Proposition 1 to solve it globally.

IV. Robust Global Resource Allocation

We now apply the theory developed in the previous section to the solution of (P2). Recall that a core idea is to exchange objective and constraints to obtain an optimization problem that is considerably easier to solve with a BB procedure than the original problem. This is mainly due to nice structural properties of the dual feasible set which facilitates an easy implementation of the feasibility checks required in BB methods and has no isolated points. These points are hard to compute and can lead to numerical instabilities.

Interchanging objective and constraints in (P2) leads to

$$\begin{cases} \min\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & \max\limits_{i=1,2,\ldots,m} \left(g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x})\right) \\ \text{s. t.} & \dfrac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma \end{cases}$$

or, equivalently,

$$\begin{cases} \min\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & \max\limits_{i=1,2,\ldots,m} \left(g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x})\right) \\ \text{s. t.} & \gamma f^-(\boldsymbol{x},\boldsymbol{\xi}) - f^+(\boldsymbol{x},\boldsymbol{\xi}) \leq 0 \end{cases} \tag{P8}$$

since $f^-(\boldsymbol{x},\boldsymbol{\xi}) > 0$ by assumption. The feasible set $\mathcal{D}$ of (P8) is nice as long as the conditions in Case A or B is satisfied. This is formally established in the proposition below.

Proposition 2: The feasible set of (P8) does not contain any isolated points if the conditions in Case A or B are satisfied.

Proof: Please refer to the appendix. ■

We now design a BB procedure to solve (P8). Together with the SIT scheme in Algorithm 1 this will result in a method to solve (P2) and is stated in Algorithm 2. The core idea of BB is to relax the feasible set and subsequently partition it such that lower bounds on the objective value can be determined easily. In our case, a rectangular subdivision procedure is a reasonable choice due to the required existence of a common maximizer of $\{g_i^-(\boldsymbol{x})\}$ over every box in the domain of (P8).

Since the BB procedure is supposed to only operate on the global variables $\boldsymbol{x}$, it partitions the $\boldsymbol{x}$-dimensions of $\mathcal{D}$ successively into boxes $\{\mathcal{M}_i\}$. Specifically, in iteration k the algorithm selects a box $\mathcal{M}^k = [\boldsymbol{p}^k, \boldsymbol{q}^k]$ with the lowest bound and bisects it via $(\boldsymbol{v}^k, j_k)$, i.e., $\mathcal{M}^k$ is replaced by

$$\begin{aligned} \mathcal{M}_-^k &= \{\boldsymbol{x} \,|\, p_{j_k}^k \leq x_{j_k} \leq v_{j_k}^k,\, p_i^k \leq x_i \leq q_i^k \ (i \neq j_k)\} \\ \mathcal{M}_+^k &= \{\boldsymbol{x} \,|\, v_{j_k}^k \leq x_{j_k} \leq q_{j_k}^k,\, p_i^k \leq x_i \leq q_i^k \ (i \neq j_k)\}. \end{aligned} \tag{5}$$

For each box $\mathcal{M}_i$, a lower bound $\beta(\mathcal{M}_i)$ for (P8) with additional constraint $\boldsymbol{x} \in \mathcal{M}_i$ is computed. We will discuss the computation of $\beta(\mathcal{M}_i)$ in the next subsection. For now it suffices to assume that it generates two points $\boldsymbol{x}^k \in \mathcal{M}^k$, $\boldsymbol{y}^k \in \mathcal{M}^k$ satisfying

$$(\boldsymbol{x}^k, \boldsymbol{\xi}^k) \in \mathcal{D}, \quad \min_{\boldsymbol{\xi}\in\mathcal{D}_{\boldsymbol{y}^k}} g(\boldsymbol{y}^k, \boldsymbol{\xi}) - \beta(\mathcal{M}^k) \to 0 \quad \text{as } \left\|\boldsymbol{x}^k - \boldsymbol{y}^k\right\| \to 0 \tag{6}$$

for some $\boldsymbol{\xi}^k$ and with

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

Thus, we can employ an adaptive bisection that exhibits much faster convergence than the common exhaustive subdivision, and choose the bisection parameters in (5) as $\boldsymbol{v}^k = \frac{1}{2}(\boldsymbol{x}^k + \boldsymbol{y}^k)$ and $j_k = \arg\max_j \left|y_j^k - x_j^k\right|$. For a formal proof of the convergence of such an adaptive BB procedure, we refer the interested reader to [4, Prop. 6.2].

A. Bounding

We now discuss the computation of lower bounds for (P8) that satisfy (6). First, we establish a lower bound on the objective of (P8).

Proposition 3: Let $\bar{\boldsymbol{x}}_{\mathcal{M}}^*$ be a common maximizer of $\{g_i^-(\boldsymbol{x})\}_{i=1,\ldots,m}$ over the box $\mathcal{M}$. Then, (P8)'s objective is lower bounded over $\mathcal{M}$ by

$$\max_{i=1,2,\ldots,m} \left\{g_i^+(\boldsymbol{x}, \boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}_{\mathcal{M}}^*)\right\} \tag{7}$$

This bound is tight at $\bar{\boldsymbol{x}}_{\mathcal{M}}^*$.

Proof: Please refer to the appendix. ■

With Proposition 3 we can determine the lower bound $\beta(\mathcal{M}_i)$ as the optimal value of

$$\begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{\xi}} & \max\limits_{i=1,2,\ldots,m} \left\{g_i^+(\boldsymbol{x}, \boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}_{\mathcal{M}_i}^*)\right\} \\ \text{s. t.} & \gamma f^-(\boldsymbol{x},\boldsymbol{\xi}) - f^+(\boldsymbol{x},\boldsymbol{\xi}) \leq 0 \\ & (\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{C},\ \boldsymbol{x} \in \mathcal{M}_i. \end{cases} \tag{P9}$$

This is a convex optimization problem if the conditions in Case A or B are satisfied and can be solved in polynomial time under very mild assumptions using standard tools [31]. For Case A, the objective of (P9) is a convex function because $g_i^-(\bar{\boldsymbol{x}}_{\mathcal{M}_i}^*)$ is constant and the feasible set is convex (cf. Proposition 2). Thus, (P9) is convex given Case A. For Case B, (P9) can be written as

$$\begin{cases} \min\limits_{\boldsymbol{x},\boldsymbol{\xi}} & \max\limits_{i=1,2,\ldots,m} \left\{g_{i,\xi}^+(\boldsymbol{\xi}) + g_{i,x}^+(\boldsymbol{x}) - g_i^-(\bar{\boldsymbol{x}}_{\mathcal{M}_i}^*)\right\} \\ \text{s. t.} & \gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi}) + \gamma f_x^-(\boldsymbol{x}) - f_x^+(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{\xi} \in \Xi,\ \boldsymbol{x} \in \mathcal{X} \cap \mathcal{M}_i. \end{cases} \tag{P10}$$

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

Let $\underline{\boldsymbol{x}}^*_{\mathcal{M}_i}$ be the common minimizer of $\gamma f_x^-(\boldsymbol{x}) - f_x^+(\boldsymbol{x})$, $g_{1,x}^+(\boldsymbol{x}), \ldots, g_{m,x}^+(\boldsymbol{x})$ over $\mathcal{X} \cap \mathcal{M}_i$. Then (P10) is equivalent to

$$\begin{cases} \min\limits_{\boldsymbol{\xi}} & \max\limits_{i=1,2,\ldots,m} \left\{ g_{i,\xi}^+(\boldsymbol{\xi}) + g_{i,x}^+(\underline{\boldsymbol{x}}^*_{\mathcal{M}_i}) - g_i^-(\bar{\boldsymbol{x}}^*_{\mathcal{M}_i}) \right\} \\ \text{s.t.} & \gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi}) + \gamma f_x^-(\underline{\boldsymbol{x}}^*_{\mathcal{M}_i}) - f_x^+(\underline{\boldsymbol{x}}^*_{\mathcal{M}_i}) \leq 0 \\ & \boldsymbol{\xi} \in \Xi. \end{cases} \tag{P11}$$

It is easy to see that $\underline{\boldsymbol{x}}^*_{\mathcal{M}_i}$ is the optimal solution of (P10) since it jointly minimizes the objective and the first constraint. Problem (P11) is convex due to the assumptions on Ξ and the remaining functions of $\boldsymbol{\xi}$ made in Case B.

Finally, for each $\mathcal{M}^k$, we identify the variables from (6) as $\boldsymbol{y}^k = \bar{\boldsymbol{x}}^*_{\mathcal{M}^k}$ and $(\boldsymbol{x}^k, \boldsymbol{\xi}^k)$ as the optimal solution of (P9) or $\boldsymbol{x}^k = \underline{\boldsymbol{x}}^*_{\mathcal{M}_i}$ if (P11) is solved.

B. The SIT Algorithm

The original SIT algorithm from [4, Sect. 7.5.2] does not distinguish between global and non-global variables. We extent it such that branching is only performed over the global variables and state-of-the-art commercially available solvers can be used for the non-global variables. This preserves the computational complexity in the non-global variables, and increases computational performance and numerical accuracy compared to self-crafted algorithms due to the high maturity of these industry-grade solvers. We also extend it to fractional objectives which removes the necessity for Dinkelbach's algorithm.

The BB procedure from the previous section solves (P8), but that is not exactly what is required by the SIT scheme. Instead, Algorithm 1 requires the implementation of

Step 1 Check if (P2) has a nonisolated feasible solution $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \geq \gamma$; otherwise, establish that no such ε-essential feasible $\boldsymbol{x}$ exists and go to Step 3.

This is accomplished by a modified version of the adaptive BB algorithm from the previous section. Consider the following proposition which is adapted from [4, Prop. 7.14] and leverages the simple observation in Proposition 1.

Proposition 4: Let $\varepsilon > 0$ be given. Either $g(\boldsymbol{x}^k, \boldsymbol{\xi}^*) < 0$ for some k and $\boldsymbol{\xi}^*$ or $\beta(\mathcal{M}^k) > -\varepsilon$ for some k. In the former case, $(\boldsymbol{x}^k, \boldsymbol{\xi}^*)$ is a nonisolated feasible solution of (P5) satisfying $\frac{f^+(\boldsymbol{x}^k,\boldsymbol{\xi}^*)}{f^-(\boldsymbol{x}^k,\boldsymbol{\xi}^*)} \geq \gamma$. In the latter case, no ε-essential feasible solution $(\boldsymbol{x}, \boldsymbol{\xi})$ of (P5) exists such that $\frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma$.

Proof: Straightforward adaption of [4, Prop. 7.14]. ■

Thus, an adaptive BB algorithm for solving (P7) with deletion criterion $\beta(M) > -\varepsilon$ and stopping criterion $\min_{\boldsymbol{\xi} \in \mathcal{D}_{\boldsymbol{x}^k}} g(\boldsymbol{x}^k, \boldsymbol{\xi}) < 0$ implements Step 1 in Algorithm 1: In the first case of Proposition 4 the incumbent feasible solution can be improved, in the latter case, if $\gamma = \frac{f^+(\bar{\boldsymbol{x}},\bar{\boldsymbol{\xi}})}{f^-(\bar{\boldsymbol{x}},\bar{\boldsymbol{\xi}})} + \eta$ for a given $\eta > 0$ and a nonisolated feasible solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\xi}})$, the incumbent $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\xi}})$ is an essential (ε, η)-optimal solution of (P5).

With the observation in Proposition 4, we can formally state the complete procedure for solving (P2) with global optimality in Algorithm 2. It is initialized in Step 0 where an initial box $\mathcal{M}^0 = [\boldsymbol{p}^0, \boldsymbol{q}^0]$ is required that contains the $\boldsymbol{x}$-dimensions of $\mathcal{C}$, i.e.,

$$p_i^0 \leq \min_{(\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{C}} x_i \qquad q_i^0 \geq \max_{(\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{C}} x_i. \tag{8}$$

The set $\mathscr{P}_k$ contains new boxes to be examined in Step 1, γ holds the current best value adjusted by the tolerance η, and $\mathscr{R}$ holds all boxes that are not yet eliminated. In Step 1 the bound is computed for each box in $\mathscr{P}_k$. If it is less than $-\varepsilon$, the box may contain a nonisolated feasible solution with objective value greater than γ and is added to $\mathscr{R}$. Then, in Step 3, the box with the smallest bound is taken out of $\mathscr{R}$. If the point $\boldsymbol{x}^k$ attaining the bound is feasible in the original problem (P2), it is a nonisolated feasible point and needs to be examined further: if the objective value for $\boldsymbol{x}^k$ is greater than the current best value $\gamma - \eta$, $\boldsymbol{x}^k$ is the new current best solution and γ is updated accordingly in Step 4. Irrespective of $\boldsymbol{x}^k$'s feasibility, the box selected in Step 3 is bisected via $(j_k, \boldsymbol{v}^k)$ adaptively. These new boxes are then passed to Step 1 and the algorithm repeats until $\mathscr{R}$ holds no more boxes which is checked in Step 2. Convergence of the algorithm is stated formally in the theorem below.

Theorem 1: Algorithm 2 converges in finitely many steps to the (ε, η)-optimal solution of (P2) or establishes that no such solution exists.

Proof: Please refer to the appendix. ■

Algorithm 2 SIT Algorithm for (P2)

Step 0 Initialize $\varepsilon, \eta > 0$ and $\mathcal{M}_0 = [\boldsymbol{p}^0, \boldsymbol{q}^0]$ as in (8), $\mathscr{P}_1 = \{\mathcal{M}_0\}$, $\mathscr{R} = \emptyset$, and $k = 1$. Initialize $\bar{\boldsymbol{x}}$ with the best known nonisolated feasible solution and set γ as described in Step 4; otherwise do not set $\bar{\boldsymbol{x}}$ and choose $\gamma \leq \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})}$ for all feasible $(\boldsymbol{x}, \boldsymbol{\xi})$.

Step 1 For each box $\mathcal{M} \in \mathscr{P}_k$:
- Compute $\beta(\mathcal{M})$. Set $\beta(\mathcal{M}) = \infty$ if (P9) (or (P11)) is infeasible.
- Add $\mathcal{M}$ to $\mathscr{R}$ if $\beta(\mathcal{M}) \leq -\varepsilon$.

Step 2 Terminate if $\mathscr{R} = \emptyset$: If $\bar{\boldsymbol{x}}$ is not set, then (P2) is ε-essential infeasible; else $\bar{\boldsymbol{x}}$ is an essential (ε, η)-optimal solution of (P2).

Step 3 Let $\mathcal{M}^k = \arg\min\{\beta(\mathcal{M}) \,|\, \mathcal{M} \in \mathscr{R}\}$. Let $\boldsymbol{x}^k$ be the optimal solution of (P9) for the box $\mathcal{M}^k$ (or $\underline{\boldsymbol{x}}^*_{\mathcal{M}^k}$ if (P11) is employed for bounding), and $\boldsymbol{y}^k = \bar{\boldsymbol{x}}^*_{\mathcal{M}^k}$. Solve the feasibility problem

$$\begin{cases} \text{find} & \boldsymbol{\xi} \in \mathcal{C}_{\boldsymbol{x}^k} \\ \text{s.t.} & g_i^+(\boldsymbol{x}^k, \boldsymbol{\xi}) - g_i^-(\boldsymbol{x}^k) \leq 0, \; i = 1, \ldots, m. \end{cases} \tag{P12}$$

If (P12) is feasible go to Step 4; otherwise go to Step 5.

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

$$\begin{cases} \min\limits_{\boldsymbol{\xi} \in \mathcal{C}_{\boldsymbol{x}^k}} & \dfrac{f^+(\boldsymbol{x}^k, \boldsymbol{\xi})}{f^-(\boldsymbol{x}^k, \boldsymbol{\xi})} \\ \text{s.t.} & g_i^+(\boldsymbol{x}^k, \boldsymbol{\xi}) - g_i^-(\boldsymbol{x}^k) \leq 0, \quad i = 1, 2 \ldots, m. \end{cases} \tag{P13}$$

If $\bar{\boldsymbol{x}}$ is not set or $\frac{f^+(\boldsymbol{x}^k,\boldsymbol{\xi}^*)}{f^-(\boldsymbol{x}^k,\boldsymbol{\xi}^*)} > \gamma - \eta$, set $\bar{\boldsymbol{x}} = \boldsymbol{x}^k$ and $\gamma = \frac{f^+(\boldsymbol{x}^k,\boldsymbol{\xi}^*)}{f^-(\boldsymbol{x}^k,\boldsymbol{\xi}^*)} + \eta$.

Step 5 Bisect $\mathcal{M}^k$ via $(\boldsymbol{v}^k, j_k)$ where $j_k \in \arg\max_j \{|y_j^k - x_j^k|\}$ and $\boldsymbol{v}^k = \frac{1}{2}(\boldsymbol{x}^k + \boldsymbol{y}^k)$ (cf. (5)). Remove $\mathcal{M}^k$ from $\mathscr{R}$. Let $\mathscr{P}_{k+1} = \{\mathcal{M}_-^k, \mathcal{M}_+^k\}$. Increment k and go to Step 1.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

Observe that the BB procedure is directly incorporated in Algorithm 1. Thus, except for the first cycle, the BB procedure is started from the boxes already in $\mathscr{R}$ instead of starting from scratch [14, Remark 2]. Also observe that $\mathscr{R}$ is a priority queue [57, Sect. 5.2.3], and that (P12) is a convex optimization problem. Due to the assumption of $\boldsymbol{\xi}$ being non-global variables, problem (P13) is efficiently solvable. For example, if $f^+(\boldsymbol{x},\boldsymbol{\xi})$ is non-negative and concave in $\boldsymbol{\xi}$, and $f^-(\boldsymbol{x},\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$ it is solvable with Dinkelbach's algorithm. Furthermore, if $f^-(\boldsymbol{x},\boldsymbol{\xi})$ is affine, $f^+(\boldsymbol{x},\boldsymbol{\xi})$ may also be negative. Please refer to [50], [51], [58] for more detailed treatments of fractional programming.

Remark 2: The original algorithm in [4] contains an additional reduction step before computing the bound of a box. This step is optional and should only be included if it speeds up the algorithm. A direct application of the approach outlined in [4] requires several convex optimization problems to be solved for each reduction. In our numerical experiments this slowed down the algorithm and is thus omitted. We leave the design of an efficient reduction procedure open for future work.

C. Extension: Pointwise Minimum

Algorithm 2 is easily extended to the case were the objective is the pointwise minimum of several functions, i.e.,

$$\begin{cases} \max\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & \min\limits_{j} \left\{ \dfrac{f_j^+(\boldsymbol{x},\boldsymbol{\xi})}{f_j^-(\boldsymbol{x},\boldsymbol{\xi})} \right\} \\ \text{s.t.} & g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x}) \leq 0, \quad i = 1,2,\ldots,m. \end{cases} \tag{P14}$$

In that case, all fractions in the minimum need to be greater or equal than γ and the dual problem (P8) becomes

$$\begin{cases} \min\limits_{(\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C}} & \max\limits_{i=1,2,\ldots,m} \left(g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\boldsymbol{x})\right) \\ \text{s.t.} & \gamma f_j^-(\boldsymbol{x},\boldsymbol{\xi}) - f_j^+(\boldsymbol{x},\boldsymbol{\xi}) \leq 0, \quad j = 1,2,\ldots \end{cases} \tag{P15}$$

Then, the branching and bounding procedures are easily adjusted to this extended dual problem (P15).

V. Numerical Evaluation

The application of any optimization framework to a specific optimization problem requires, in general, some transformation of the initial problem to bring it into a form suitable for the framework. In addition, modification of the problem often allows to reduce the computational complexity significantly.

In this section, we first present the system model of a specific interference network, namely the Gaussian MWRC, and two of its achievable rate region. We then formulate the resource allocation problems and discuss, based on Section II-A, the application of Algorithm 2. Subsequently, we employ Algorithm 2 to obtain throughput and GEE optimal resource allocations and compare the performance of Algorithm 2 to the state-of-the-art. We conclude this section by evaluating how the performance of Algorithm 2 scales with an increasing number of global and non-global variables in Section V-D.

A. System Model

We consider a 3-user SISO Gaussian MWRC [59], [60] with amplify-and-forward (AF) relaying, multiple unicast transmissions and no direct user-to-user links. Users are indexed by k, $k \in \mathcal{K} = \{1,2,3\}$, and the relay is node 0. User k transmits with power P_k over the channel h_k to the relay. The relay propagates the observed symbol back to the users with transmit power P_0 over the channels g_k, $k \in \mathcal{K}$. Each node observes the independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian noise with power N_k, $k \in \mathcal{K} \cup \{0\}$ and is subject to an average power constraint $\bar{P}_k$ on X_k, $k \in \mathcal{K} \cup \{0\}$. The message exchange is defined by the functions $q : \mathcal{K} \mapsto \mathcal{K}$ and $l : \mathcal{K} \mapsto \mathcal{K}$, where the receiver of node k's message is $q(k)$ and the user not interested in it is $l(k)$. Without loss of generality we assume $q(1) = l(3) = 2$, $q(2) = l(1) = 3$, and $q(3) = l(2) = 1$. Due to space constraints we refer the reader to [1], [5], [61] for a more detailed treatment of this system model.

The receiver uses SND which is the optimal decoder for interference networks when restricted to random codebooks with superposition coding and time sharing [8]. We consider two different codebook constructions: traditional single message encoding and Han-Kobayashi [62] inspired rate splitting.

1) Single Message: The achievable rate region is given below where $S_k = \frac{P_k}{N_0}$, $\boldsymbol{S} = (S_k)_{k\in\mathcal{K}}$, and $\bar{\boldsymbol{S}} = (\bar{S}_k)_{k\in\mathcal{K}}$. Observe that this region is strictly larger than previously published SND regions [5], [61] and includes treating interference as noise (IAN) as a special case. This is due to recent insights on SND decoders [8].

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

$$R_k \leq \log\left(1 + \frac{|h_k|^2 S_k}{\gamma_k(\boldsymbol{S})}\right) \tag{9}$$

or

$$R_k \leq \log\left(1 + \frac{|h_k|^2 S_k}{\delta_k(\boldsymbol{S})}\right) \tag{10a}$$

$$R_k + R_{l(k)} \leq \log\left(1 + \frac{|h_k|^2 S_k + \left|h_{l(k)}\right|^2 S_{l(k)}}{\delta_k(\boldsymbol{S})}\right) \tag{10b}$$

where $S_k \leq \bar{S}_k$, $\delta_k(\boldsymbol{S}) = 1 + \widetilde{g}_{q(k)}^{-1}\left(1 + \sum_{i\in\mathcal{K}} |h_i|^2 S_i\right)$ with $\widetilde{g}_k = |g_k|^2 \frac{\bar{P}_0}{N_k}$, and $\gamma_k(\boldsymbol{S}) = \delta_k(\boldsymbol{S}) + \left|h_{l(k)}\right|^2 S_{l(k)}$.

Proof: The proof follows along the lines of [8, Sect. II-A.] and is omitted due to space constraints. ■

Let $\mathcal{R}_{k,\text{IAN}}$ and $\mathcal{R}_{k,\text{SND}}$ be the regions defined by (9) and (10), respectively. Then, the rate region in Lemma 1 is

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

Since $\inf_{\boldsymbol{x}\in\bigcup_i \mathcal{D}_i} f(\boldsymbol{x}) = \min_i \inf_{\boldsymbol{x}\in\mathcal{D}_i} f(\boldsymbol{x})$, we can split the resource allocation problem for Lemma 1 into eight individual optimization problems. Each is easily identified as an instance of (P3) and solvable with Algorithm 2 using the initial box $\mathcal{M}_0 = [\boldsymbol{0}, \bar{\boldsymbol{S}}]$.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

2) Rate Splitting: Each message is divided 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 achievable rate region is given below in terms of the signal-to-noise ratios (SNRs) $S_k^c = \frac{P_k^c}{N_0}$ and $S_k^p = \frac{P_k^p}{N_0}$. Further, we define $S_k = S_k^c + S_k^p$, $\bar{S}_k = \frac{\bar{P}_k}{N_0}$, $\boldsymbol{S}^c = (S_k^c)_{k\in\mathcal{K}}$, $\boldsymbol{S}^p = (S_k^p)_{k\in\mathcal{K}}$, $\boldsymbol{S} = (\boldsymbol{S}^c, \boldsymbol{S}^p)$, and $\bar{\boldsymbol{S}} = (\bar{S}_k)_{k\in\mathcal{K}}$.

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

$$R_k \leq B_k, \tag{11a}$$
$$R_k + R_{q(k)} \leq A_k + D_{q(k)}, \tag{11b}$$
$$R_k + R_{q(k)} + R_{l(k)} \leq A_k + C_{q(k)} + D_{l(k)}, \tag{11c}$$
$$2R_k + R_{q(k)} + R_{l(k)} \leq A_k + C_{q(k)} + C_{l(k)} + D_k, \tag{11d}$$

and,
$$R_1 + R_2 + R_3 \leq C_1 + C_2 + C_3, \tag{11e}$$

with
$$A_k = \log\left(1 + \frac{|h_k|^2 S_k^p}{\gamma_k(\boldsymbol{S})}\right) \tag{12a}$$
$$B_k = \log\left(1 + \frac{|h_k|^2 (S_k^p + S_k^c)}{\gamma_k(\boldsymbol{S})}\right) \tag{12b}$$
$$C_k = \log\left(1 + \frac{|h_k|^2 S_k^p + |h_{l(k)}|^2 S_{l(k)}^c}{\gamma_k(\boldsymbol{S})}\right) \tag{12c}$$
$$D_k = \log\left(1 + \frac{|h_k|^2 (S_k^p + S_k^c) + |h_{l(k)}|^2 S_{l(k)}^c}{\gamma_k(\boldsymbol{S})}\right) \tag{12d}$$

where $S_k^c + S_k^p \leq \bar{S}_k$ and
$$\gamma_k(\boldsymbol{S}) = 1 + |h_{l(k)}|^2 S_{l(k)}^p + \tilde{g}_{q(k)}^{-1}\left(1 + \sum_{i\in\mathcal{K}} |h_i|^2 (S_i^c + S_i^p)\right),$$
with $\tilde{g}_k = |g_k|^2 \frac{\bar{P}_0}{N_k}$.

Proof: Please refer to the appendix. ■

Global optimal resource allocation for this scenario is a straightforward extension of (P3):

$$\begin{cases} \max\limits_{\boldsymbol{S},\boldsymbol{R}} & f(\boldsymbol{S},\boldsymbol{R}) \\ \text{s.t.} & \boldsymbol{a}_i^T \boldsymbol{R} \leq \sum\limits_j \log\left(1 + \dfrac{\boldsymbol{b}_{i,j}^T \boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S})}\right), \quad i = 1,\dots,n \\ & S_k^c + S_k^p \leq \bar{S}_k,\ k \in \mathcal{K}, \qquad \boldsymbol{R} \geq 0, \quad \boldsymbol{S} \geq 0 \end{cases} \tag{P16}$$

where $\kappa(i,j)$ maps from (i,j) to the correct $k \in \mathcal{K}$ and $\boldsymbol{a}_i, \boldsymbol{b}_{i,j} \geq 0$ are easily identified from Lemma 2. From the discussion in Section II-A it is apparent that (P16), depending on the identification of $g_i^+(\boldsymbol{S},\boldsymbol{R})$ and $g_i^-(\boldsymbol{S})$, satisfies both, Cases A and B, and is solvable with Algorithm 2 where the global and non-global variables are $\boldsymbol{S}$ and $\boldsymbol{R}$, respectively. However, we can reduce the number of global variables in (P16) from six to four: When using the identification $g_i^-(\boldsymbol{S}) := -\sum_j \log(\gamma_{\kappa(i,j)}(\boldsymbol{S}))$, the non-convexity due to $\boldsymbol{S}^c$ stems only from the sum $\sum_{k\in\mathcal{K}} |h_k|^2 S_k^c$ in $\gamma_k(\boldsymbol{S})$. Thus, if we replace this sum by an auxiliary variable y, the variables $\boldsymbol{S}^c$ can be treated as non-global. The resulting problem is

$$\begin{cases} \max\limits_{\boldsymbol{S},\boldsymbol{R},y} & f(\boldsymbol{S},\boldsymbol{R}) \\ \text{s.t.} & \boldsymbol{a}_i^T \boldsymbol{R} \leq \sum\limits_j \log\left(1 + \dfrac{\boldsymbol{b}_{i,j}^T \boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S}^p, y)}\right), \quad i = 1,\dots,n \\ & y \geq \sum\limits_{k\in\mathcal{K}} |h_k|^2 S_k^c \\ & S_k^c + S_k^p \leq \bar{S}_k,\ k \in \mathcal{K}, \qquad \boldsymbol{R} \geq 0, \quad \boldsymbol{S} \geq 0 \end{cases} \tag{P17}$$

where $(\boldsymbol{S}^p, y)$ are the global and $(\boldsymbol{S}^c, \boldsymbol{R})$ are the non-global variables. Since the computational complexity grows exponentially in the number of global variables, solving (P17) instead of (P16) reduces the complexity significantly. The drawback of this approach is that Case B is no longer satisfied and the bounding problem becomes convex instead of linear. However, the performance gain outweighs these, comparatively small, performances losses significantly. The proposition below formally states the equivalence of (P16) and (P17).

Proposition 5: Problems (P16) and (P17) are equivalent in the sense that v(P16) $=$ v(P17) if $f(\boldsymbol{S},\boldsymbol{R})$ is increasing in $\boldsymbol{R}$ for fixed $\boldsymbol{S}$.

Proof: Please refer to the appendix. ■

Finally, note that all constraints in (P17) except the first are linear and, thus, $\mathcal{C}$ is a convex set. The initial box $\mathcal{M}_0$ required by Algorithm 2 is easily identified as $[\boldsymbol{0}, \bar{\boldsymbol{S}}] \times [0, \sum_{k\in\mathcal{K}} |h_k|^2 \bar{S}_k]$.

B. Throughput

We employ Algorithm 2 to compute the throughput optimal resource allocation for the rate regions in Lemmas 1 and 2, i.e., the objective is $f(\boldsymbol{S},\boldsymbol{R}) = \boldsymbol{w}^T\boldsymbol{R}$ with $w_i = 1$ for all i and obviously fulfills Proposition 5. We assume equal maximum power constraints and noise power at the users and the relay, and no minimum rate constraint, i.e. $\underline{\boldsymbol{R}} = \boldsymbol{0}$. Channels are assumed reciprocal and chosen i.i.d. with circular symmetric complex Gaussian distribution, i.e., $h_k \sim \mathcal{CN}(0,1)$ and $g_k = h_k^*$. Results are averaged over 1,000 channel realizations. The precision of the objective value is $\eta = 0.01$, ε is chosen as 10^{-5}, and the algorithm is started with $\gamma = 0$.

Figure 2 displays the maximum throughput for SND as in Lemma 1 and rate splitting (RS) as in Lemma 2. Before [8], common wisdom was that neither 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 [8]. Figure 2 also includes results for "traditional" SND and IAN as obtained rate regions defined by (9) and (10), respectively. 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 11 % and 22 % at 10 dB, respectively, or 0.29 bpcu and 0.54 bpcu. Note that this gain is only achieved by allowing each

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

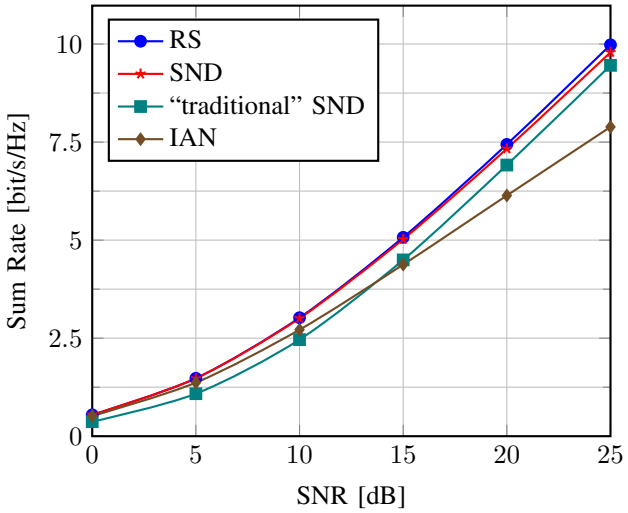

Fig. 2. Throughput in the MWRC with AF relaying and 1) RS 2) SND; 3) "traditional" SND; and 4) IAN. Averaged over 1000 i.i.d. channel realizations.

receiver to choose 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 is rather small, e.g., at 25 dB it is only 2 % (or 0.2 bpcu). However, depending on the channel realization we observed gains up to half a bit (or 4.6 %) at 20 dB. With spectrum being an increasingly scarce and expensive resource this occasional gain might very well justify the slightly higher coding complexity.

A popular approach to solving non-convex resource allocation problems with global optimality is to use monotonic optimization [16], [26]. As already pointed out in the introduction, the main challenge is that for SND the optimization is over the rates and powers (instead of just the powers for IAN). Thus, every approach that treats the rates as global variables will suffer from very long computation time. Instead, in [5] the problem is solved by decomposing it into an inner linear and an outer monotonic program. We use this approach as the state-of-the-art reference for our performance comparison. The inner linear problem is solved with Gurobi 8 [63] and the monotonic program with the Polyblock algorithm[7] [24] with a tolerance of 0.01. Results match those obtained with our SIT algorithm for "traditional" SND and, thus, are not displayed in Fig. 2. Run times for both algorithms are reported in Table I. About 5.4 % of the computations with the Polyblock algorithm did not complete within one week. For these, a run time of one week was assumed for the computation of the mean in Table I. Thus, the reported run times for the Polyblock algorithm are an underestimate. Nevertheless, we observe, on average, roughly 10,000× faster convergence for the SIT algorithm for low to medium SNRs. Interestingly, the median run time for high SNRs is lower for the Polyblock algorithm while the SIT is clearly faster on average.

C. Energy Efficiency

The EE as defined in (2) is computed with the same parameters as before except for the precision which is chosen as

[7]The Polyblock algorithm was implemented in C++ with similar techniques as in the implementation of Algorithm 2. Thus, performance differences should be mostly due to algorithmic differences.

TABLE I
MEAN AND MEDIAN RUN TIMES OF THROUGHPUT MAXIMIZATION FOR SND AND "TRADITIONAL" SND. FOR "TRADITIONAL" SND, BOTH THE SIT AND THE POLYBLOCK ALGORITHM (PA) [5] ARE EMPLOYED.

		SNR 0 dB	15 dB	30 dB
SND	Mean	385.78 s	351.92 s	5,335.48 s
	Median	75.51 s	29.16 s	2,494.54 s
"trad." SND	Mean	19.77 s	6.23 s	3,217.21 s
	Median	9.5 s	1.84 s	831.71 s
PA	Mean	$1.92 \cdot 10^5$ s	77,670.36 s	10,602.06 s
	Median	46,210.42 s	7,091.09 s	35.67 s

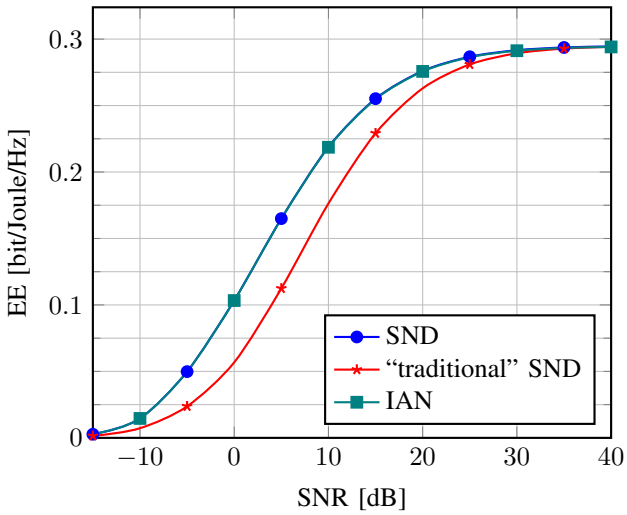

Fig. 3. Energy efficiency in the MWRC with AF relaying and 1) SND; 2) "traditional" SND; and 3) IAN. Averaged over 1000 i.i.d. channel realizations and computed with a precision of $\eta = 10^{-3}$.

$\eta = 10^{-3}$ for Fig. 3 and $\eta = 10^{-2}$ for Table II. Additionally, we assume the static circuit power $P_c = 1\,\mathrm{W}$, the power amplifier inefficiencies $\phi_i = 4$, and that the relay always transmits at maximum power. Results for SND, "traditional" SND, and IAN are displayed in Fig. 3. First, observe that the curves saturate starting from 30 dB as is common for EE maximization. In this saturation region, all three approaches achieve the same EE. However, for lower SNRs, IAN and, thus also SND, outperform "traditional" SND by 24 % on average at 10 dB. Of course, the EE performance depends quite a lot on the choice of real-world simulation parameters [64], [65], so further work is necessary to draw final conclusions in this regard.

The main point of this subsection, however, is the performance comparison of different computational approaches to the EE computation. Recall from Section II-A that there are two possible identifications of g_i^+ and g_i^- that result in the problem either belonging to Case A or Case B. As discussed before, identification (1) (resp. Case A) leads to tighter bounds than (3) but requires the solution of a convex optimization problem in the bounding step. Instead, (3) (resp. Case B) leads to a linear bounding problem which is considerably easier to solve. Table II summarizes mean and median computations times for both approaches. Each bounding problem is solved with the fastest solver available: the convex problem that stems

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

TABLE II
MEAN AND MEDIAN RUN TIMES OF EE COMPUTATION FOR "TRADITIONAL" SND AND DIFFERENT SOLVERS, ALL WITH PRECISION $\eta = 0.01$

		SNR 0 dB	20 dB	40 dB
Gurobi	Mean	5.1438 s	0.1771 s	0.155 s
	Median	3.2781 s	0.0762 s	0.06 s
Mosek	Mean	15.0453 s	0.1368 s	0.0323 s
	Median	10.2756 s	0.0219 s	0.019 s
Dinkelbach	Mean	377.1501 s	145.4181 s	36.969 s
	Median	162.811 s	23.027 s	16.9229 s

from Case A with Mosek [66] and the linear from Case B with Gurobi [63]. Observe, that Gurobi is, on average, thrice as fast as Mosek at 0 dB and Mosek is almost five times faster than Gurobi at 40 dB. However, these are only relative numbers. The average total computation time per channel realization is 8 s for Gurobi and 19 s for Mosek. So, despite the faster convergence speed of Case A (i.e., Mosek), Case B is much faster due to the lower computational complexity of the linear bounding problem.

The state-of-the-art approach to compute the EE is Dinkelbach's Algorithm [48], [51]. This requires the global solution of a sequence of auxiliary problems with very similar properties to the throughput maximization problem in the previous section. Here, we solve this inner problem with the fastest method available: the SIT algorithm developed in this paper (cf. Table I). Hence, the differences in the run time are solely due to the use of Dinkelbach's Algorithm. It can be observed from Table II that the inherent treatment of fractional objectives in our algorithm is always significantly faster (up to 800× on average at 20 dB) than Dinkelbach's Algorithm. Moreover, the obtained result is guaranteed to lie within an η-region around the true essential optimal value.

D. Benchmark of Algorithm 2

Above, we evaluated the performance of Algorithm 2 against the state-of-the-art on a real-world resource allocation problem. The goal of this subsection is to assess how Algorithm 2 scales with an increasing number of global and non-global variables. This is done by means of an idealized example. Namely, consider a K-user Gaussian interference channel with input-output relation

$$y_k = \sum_{j=1}^{K} h_{kj} x_j + z_k \tag{13}$$

where $h_{kj} \in \mathbb{C}$ is the (effective) channel gain from transmitter j to receiver k, x_j is the complex-valued channel input of transmitter j that is subject to an average power constraint $\bar{P}_j$, y_k is the received symbol at receiver k, and z_k is circularly-symmetric complex Gaussian noise with zero mean and power N_k. Consider just the k-th receiver which is interested in decoding x_k and assume it jointly decodes a subset $\mathcal{S}_k \subseteq \{1, 2, \ldots, K\}$ of messages and treats the remaining x_j, $j \in \mathcal{S}_k^c = \{1, 2, \ldots, K\} \backslash \mathcal{S}_k$, as noise. This is a MAC with capacity region

$$\mathcal{A}_k(\mathcal{S}_k) = \left\{ (R_j)_{j \in \mathcal{S}_k} \;\middle|\; \sum_{i \in \mathcal{T}} R_i \leq \log\left(1 + \frac{P_k(\mathcal{T})}{N_k + P_k(\mathcal{S}_k^c)}\right) \text{ for every } \mathcal{T} \subseteq \mathcal{S}_k \right\} \tag{14}$$

where $P_k(\mathcal{S}) = \sum_{j \in \mathcal{S}} |h_{kj}|^2 P_j$. The achievable rate region for the complete system then is

$$\mathcal{R} = \bigcap_{k=1}^{K} \mathcal{A}_k(\mathcal{S}_k). \tag{15}$$

For the benchmark, assume that receiver i jointly decodes its own message and the interfering message of transmitter $i+1 \mod K$. All other messages are treated as noise. Moreover, assume that $h_{ij} = 0$ for $i \neq j$, $j \neq i+1$, $j > \kappa$, and some positive κ. These assumptions allow to precisely control the complexities in the global and non-global variables. Thus, $\mathcal{S}_k = \{k, k+1 \mod K\}$ for all k and $\mathcal{S}_k^c = \{1, \ldots, \kappa\} \setminus \mathcal{S}_k$. Maximizing the throughput in this system requires the solution of

$$\begin{cases} \max\limits_{\boldsymbol{p}, \boldsymbol{R}} & \sum\limits_{k=1}^{K} R_k \\ \text{s.t.} & \boldsymbol{R} \in \mathcal{A}_k(\mathcal{S}_k), \text{ for all } k \\ & \boldsymbol{R} \geq 0, \quad 0 \leq \boldsymbol{p} \leq \bar{\boldsymbol{P}}. \end{cases} \tag{P18}$$

First, observe that this is a non-convex optimization problem due to $P_k(\mathcal{S}_k^c)$ in the denominators of the log-terms in (14). Further, observe that $\mathcal{S}_k^c \subseteq \{1, \ldots, \kappa\}$ for all $k = 1, \ldots, K$. Hence, we can identify the global variables as $(p_k)_{k=1}^{\kappa}$ and the non-global variables as $(p_k)_{\kappa+1}^{K}$ and $\boldsymbol{R}$. Moreover, the powers $(p_k)_{\kappa+1}^{K}$ only occur in the numerators of the log-terms in (14) and these log-terms are increasing functions in $(p_k)_{\kappa+1}^{K}$. Thus, $p_k = \bar{P}_k$ is the optimal solution for all $k = \kappa+1, \ldots, K$. Plugging-in this partial solutions in (P18), we obtain a simplified version of (P18) with κ global variables $(p_k)_{k=1}^{\kappa}$, K linear variables $\boldsymbol{R}$, and $3K$ inequality constraints. Thus, we can indepently control the number of global and non-global variables. Since the number of constraints grows linearly in the number of global variables and the bounding problem is a linear program, the computational complexity of solving (P18) should grow polynomially in the number of non-global variables.

This rationale is verified numerically in Fig. 4 where (P18) is solved for a fixed number of κ global variables and an increasing number of K total users in the system. We employ Algorithm 2 with Gurobi as inner solver and choose $\bar{P}_k = 1$ and $N_k = 0.01$ for all $k = 1, \ldots, K$, i.e., a transmit SNR of 20 dB, and i.i.d. $h_{kj} \sim \mathcal{CN}(0,1)$. Each data point is obtained by averaging over 1,000 independent channel realizations. Algorithm 2 was started with parameters $\eta = 0.01$, $\varepsilon = 10^{-5}$, and $\gamma = 0$. Since the abcissa is directly proportional to the number of non-global variables $(K - \kappa)$ and the number of inequality constraints $(3K)$, we expect to observe a polynomial growth of the run time. Indeed, Fig. 4 displays a roughly

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

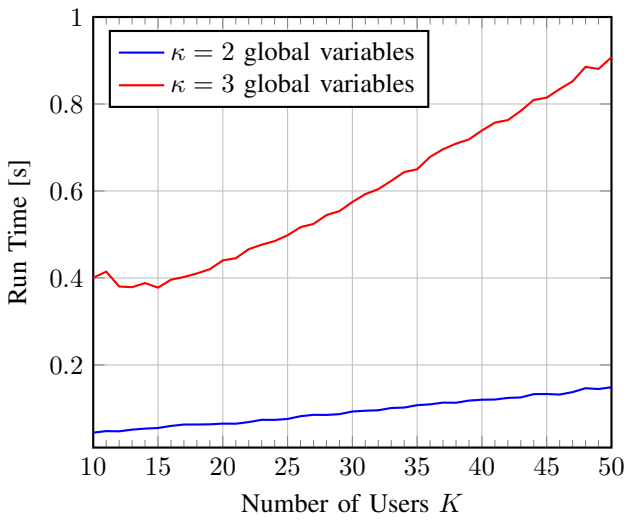

Fig. 4. Average run time of Algorithm 2 to solve (P18) as a function of the total number of users for fixed number of κ global variables. The abcissa is proportional to the number of non-global variables ($K - \kappa$). Averaged over 1,000 i.i.d. channel realizations.

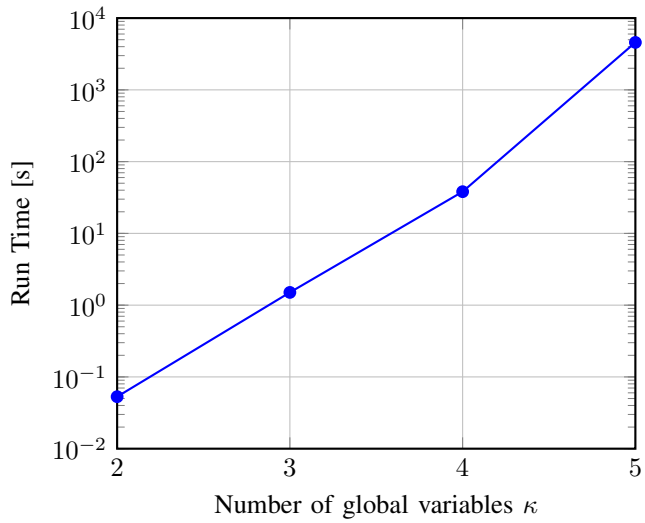

Fig. 5. Average run time of Algorithm 2 to solve (P18) as a function of the number of global variables κ for fixed number of $K = 7$ users. Averaged over 1,000 i.i.d. channel realizations.

linear growth expect for small K. This is because, for big K, the optimal power allocation converges to setting all global variables to zero. Instead, for small K these powers are generally non-zero and the global part of the optimization problem requires more time to converge to the optimal solution.

Finally, Fig. 5 verifies the exponential complexity in the number of global variables. With the same paramters as in Fig. 4, Fig. 5 displays the run time for a fixed number of $K = 7$ users over the number of global variables κ. Due to the logarithmical y-axis, exponential growth corresponds to (at least) a linear function in Fig. 5.

E. Implementation Details

Algorithm 2 was implemented in C++ and compiled with GCC 7.3. Linear optimization problems are solved with Gurobi 8 [63] and non-linear problems with Mosek 8.1 [66] using the kindly provided academic licenses. The complete source code is available on GitHub [3].

All computations were done on TU Dresden's Bull HPC-Cluster Taurus. Reported performance results were obtained on Intel Haswell nodes with Xeon E5-2680 v3 CPUs running at 2.50 GHz. We thank the Center for Information Services and High Performance Computing (ZIH) at TU Dresden for generous allocations of computer time.

VI. Conclusions

We established ε-essential feasibility as an important concept towards numerical stable global optimization algorithms and introduced the accompanying successive incumbent transcending (SIT) approach. Based on these concepts we built a novel global optimization framework tailored to resource allocation problems that preserves the computational complexity in the number of non-global variables, inherently supports fractional objectives, avoids numerical problems with non-robust feasible sets, and is four orders of magnitude faster than state-of-the-art algorithms. We applied the proposed scheme to a resource allocation problem for the Gaussian MWRC with AF relaying. Reproducible research and easy adoption of this algorithm is enabled by releasing the complete source code on GitHub [3].

Appendix

A. Proof of Proposition 2

First consider Case A. The level set $\mathcal{F} = \{\gamma f^-(\boldsymbol{x}, \boldsymbol{\xi}) - f^+(\boldsymbol{x}, \boldsymbol{\xi}) \leq 0\}$ is closed and convex [67, Thms. 4.6 & 7.1]. Hence, $\mathcal{D} = \mathcal{C} \cap \mathcal{F}$ is also closed and convex. Every nonempty closed convex set is robust [67, Thm. 6.3]. Thus, $\mathcal{D}$ is robust or (P8) is infeasible.

Now for Case B. The feasible set of (P8) is $\mathcal{D} = \{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\xi} \in \Xi : f_\xi(\boldsymbol{\xi}) + f_x(\boldsymbol{x}) \leq 0\}$ with $f_\xi(\boldsymbol{\xi}) = \gamma f_\xi^-(\boldsymbol{\xi}) - f_\xi^+(\boldsymbol{\xi})$ and $f_x(\boldsymbol{x}) = \gamma f_x^-(\boldsymbol{x}) - f_x^+(\boldsymbol{x})$. By assumption, $f_\xi(\boldsymbol{\xi})$ is an l.s.c. convex and $f_x(\boldsymbol{x})$ an l.s.c. increasing (decreasing) function. Further, $\mathcal{X}$ is normal (conormal) within a box and Ξ is convex. Observe that $\mathcal{D}$ is a convex set in $\boldsymbol{\xi}$ since for fixed $\boldsymbol{x}$, $f_x(\boldsymbol{x})$ is a constant, $\tilde{f}_\xi(\boldsymbol{\xi}) = f_\xi(\boldsymbol{\xi}) + \text{const.}$ is a convex function, and $\{\boldsymbol{\xi} : \tilde{f}_\xi(\boldsymbol{\xi}) \leq 0\}$ is a closed convex set [67, Thms. 4.6 & 7.1]. By the same argument, $\tilde{f}_x(\boldsymbol{x})$ is an increasing (decreasing) function and $\{\boldsymbol{x} : \tilde{f}_x(\boldsymbol{x}) \leq 0\}$ is a closed normal (conormal) set [4, Prop. 11.2]. Thus, $\mathcal{D}$ is normal (conormal) in a box in $\boldsymbol{x}$. Neither closed convex nor closed (co-)normal sets have any isolated feasible points [15]. Since $\mathcal{D}$ is either convex or (co-)normal in each coordinate the proposition is proven. ■

B. Proof of Proposition 3

For all $\boldsymbol{x} \in \mathcal{M}$, $g_i^-(\bar{\boldsymbol{x}}^*) \geq g_i^-(\boldsymbol{x})$ and hence also $g_i^+(\boldsymbol{x}, \boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}^*) \leq g_i^+(\boldsymbol{x}, \boldsymbol{\xi}) - g_i^-(\boldsymbol{x})$ with equality at $\boldsymbol{x} = \bar{\boldsymbol{x}}^*$. 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

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

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 $\min_i\{\underline{h}_i(\boldsymbol{y})\} = \min_i\{h_i(\boldsymbol{y})\}$. ■

C. Proof of Theorem 1

Convergence of the BB procedure is mostly established by Proposition 4 and [4, Prop. 6.2]. It remains to show that choosing $\beta(\mathcal{M}_i)$ as the optimal value of (P9) or (P11) satisfies (6) and that continuing the BB procedure after updating γ preserves convergence.

First, observe that neither (P9) nor (P11) relax the feasible set $\mathcal{D}$ of (P8). Hence, the first condition in (6) is always satisfied. Further,

$$\beta(\mathcal{M}^k) = \min_{\boldsymbol{\xi}\in\mathcal{D}_{\boldsymbol{x}^k}} \max_{i=1,2,\ldots,m} \left\{g_i^+(\boldsymbol{x}^k,\boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}^*_{\mathcal{M}_i})\right\}$$

and

$$\min_{\boldsymbol{\xi}\in\mathcal{D}_{\boldsymbol{v}^k}} g(\boldsymbol{v}^k,\boldsymbol{\xi}) = \min_{\boldsymbol{\xi}\in\mathcal{D}_{\boldsymbol{v}^k}} \max_{i=1,2,\ldots,m} \left(g_i^+(\bar{\boldsymbol{x}}^*_{\mathcal{M}^k},\boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}^*_{\mathcal{M}^k})\right)$$

Thus, $\beta(\mathcal{M}^k) \to \min_{\boldsymbol{\xi}\in\mathcal{D}_{\boldsymbol{v}^k}} g(\boldsymbol{v}^k,\boldsymbol{\xi})$ as $\left\|\boldsymbol{x}^k - \boldsymbol{v}^k\right\| = \left\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}^*_{\mathcal{M}^k}\right\| \to 0$ and (6) is satisfied.

Next, let $\{\gamma_k\}$ be the sequence of updated gammas, and observe that this sequence is increasing, i.e., $\gamma_{k+1} \geq \gamma_k$. Thus, the feasible sets of (P9) form a decreasing sequence of sets

$$\left\{\boldsymbol{x}\in\tilde{\mathcal{C}} \,\middle|\, \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma_{k+1}\right\} \subseteq \left\{\boldsymbol{x}\in\tilde{\mathcal{C}} \,\middle|\, \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma_{k}\right\}$$

with $\tilde{\mathcal{C}} = \{(\boldsymbol{x},\boldsymbol{\xi}) \,|\, (\boldsymbol{x},\boldsymbol{\xi})\in\mathcal{C},\, \boldsymbol{x}\in\mathcal{M}\}$. Therefore, the optimal value of the bound (P9) is increasing with k

$$\begin{aligned}&\min_{\boldsymbol{x},\boldsymbol{\xi}} \left\{\tilde{g}(\boldsymbol{x},\boldsymbol{\xi}) \,\middle|\, \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma_{k+1},\, (\boldsymbol{x},\boldsymbol{\xi})\in\tilde{\mathcal{C}}\right\} \\ \geq &\min_{\boldsymbol{x},\boldsymbol{\xi}} \left\{\tilde{g}(\boldsymbol{x},\boldsymbol{\xi}) \,\middle|\, \frac{f^+(\boldsymbol{x},\boldsymbol{\xi})}{f^-(\boldsymbol{x},\boldsymbol{\xi})} \geq \gamma_{k},\, (\boldsymbol{x},\boldsymbol{\xi})\in\tilde{\mathcal{C}}\right\}\end{aligned} \tag{16}$$

where $\tilde{g}(\boldsymbol{x},\boldsymbol{\xi}) = \max_{i=1,2,\ldots,m}\left\{g_i^+(\boldsymbol{x},\boldsymbol{\xi}) - g_i^-(\bar{\boldsymbol{x}}^*_{\mathcal{M}_i})\right\}$. Hence, every box eliminated due to the deletion criterion $\beta(\mathcal{M}) > -\varepsilon$ in a BB procedure with γ_k, would also be eliminated in a procedure with γ_{k+1}. It follows that the set holding the boxes $\mathscr{R}_k$ remains valid after updating γ. In particular, no box is eliminated prematurely. Thus, restarting the BB procedure after updating γ is not necessary. Convergence of Algorithm 2 is finite since the underlying BB procedure is finite and $\{\gamma_k\}$ is bounded. The same argument can be made for (P11). ■

D. Proof sketch for Lemma 2

Extend [61, Thm. 1] to Gaussian channels using the standard procedure in [68, 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. ■

E. Proof of Proposition 5

Let $\mathcal{F}$ be the feasible set of (P16) without the first constraint. Then, we have

$$\begin{aligned}&\max_{(\boldsymbol{S},\boldsymbol{R})\in\mathcal{F}} \left\{f(\boldsymbol{S},\boldsymbol{R}) \,|\, \forall i : \boldsymbol{a}_i^T\boldsymbol{R} \leq \sum_j \mathrm{C}\left(\frac{\boldsymbol{b}_{i,j}^T\boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S})}\right)\right\} \\ &= \max_{\substack{(\boldsymbol{S},\boldsymbol{R})\in\mathcal{F} \\ y=\sum_{k\in\mathcal{K}}|h_k|^2 S_k^c}} \left\{f(\boldsymbol{S},\boldsymbol{R}) \,|\, \forall i : \boldsymbol{a}_i^T\boldsymbol{R} \leq \sum_j \mathrm{C}\left(\frac{\boldsymbol{b}_{i,j}^T\boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S}^p,y)}\right)\right\} \\ &\leq \max_{\substack{(\boldsymbol{S},\boldsymbol{R})\in\mathcal{F} \\ y\geq\sum_{k\in\mathcal{K}}|h_k|^2 S_k^c}} \left\{f(\boldsymbol{S},\boldsymbol{R}) \,|\, \forall i : \boldsymbol{a}_i^T\boldsymbol{R} \leq \sum_j \mathrm{C}\left(\frac{\boldsymbol{b}_{i,j}^T\boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S}^p,y)}\right)\right\},\end{aligned}$$

since relaxing a constraint does not decrease the optimal value. Conversely, $\gamma_k(\boldsymbol{S}^p, y)$ is increasing in y and, thus, the RHSs of the constraints are decreasing in y. Since $f(\boldsymbol{S},\boldsymbol{R})$ and $\boldsymbol{a}_i^T\boldsymbol{R}$ are increasing in $\boldsymbol{R}$, the RHS should be as large as possible. Thus, the optimal y is as small as possible and

$$\begin{aligned}&\max_{\substack{(\boldsymbol{S},\boldsymbol{R})\in\mathcal{F} \\ y\geq\sum_{k\in\mathcal{K}}|h_k|^2 S_k^c}} \left\{f(\boldsymbol{S},\boldsymbol{R}) \,|\, \forall i : \boldsymbol{a}_i^T\boldsymbol{R} \leq \sum_j \mathrm{C}\left(\frac{\boldsymbol{b}_{i,j}^T\boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S}^p,y)}\right)\right\} \\ &\leq \max_{\substack{(\boldsymbol{S},\boldsymbol{R})\in\mathcal{F} \\ y=\sum_{k\in\mathcal{K}}|h_k|^2 S_k^c}} \left\{f(\boldsymbol{S},\boldsymbol{R}) \,|\, \forall i : \boldsymbol{a}_i^T\boldsymbol{R} \leq \sum_j \mathrm{C}\left(\frac{\boldsymbol{b}_{i,j}^T\boldsymbol{S}}{\gamma_{\kappa(i,j)}(\boldsymbol{S}^p,y)}\right)\right\}.\end{aligned}$$

This establishes Proposition 5. ■

References

[1] B. Matthiesen and E. A. Jorswieck, "Optimal resource allocation for non-regenerative multiway relaying with rate splitting," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Kalamata, Greece, Jun. 2018.

[2] ——, "Global energy efficiency maximization in non-orthogonal interference networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, May 2019.

[3] B. Matthiesen. (2018) Accompanying source code. [Online]. Available: https://github.com/bmatthiesen/efficient-global-opt

[4] H. Tuy, *Convex Analysis and Global Optimization*, ser. Springer Optimization and Its Applications. Springer, 2016.

[5] B. Matthiesen and E. A. Jorswieck, "Weighted sum rate maximization for non-regenerative multi-way relay channels with multi-user decoding," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*, Curaçao, Dutch Antilles, Dec. 2017.

[6] Z. Han and K. J. R. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*. Cambridge University Press, 2008.

[7] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, 3rd ed. Springer, 1996.

[8] 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.

[9] L. G. Khachiyan, "A polynomial algorithm in linear programming," *Doklady Academii Nauk SSSR*, vol. 244, pp. 1093–1096, 1979.

[10] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, vol. 4, no. 4, pp. 373–395, Dec. 1984.

[11] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[12] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[13] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, Jan. 2012.

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

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

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

[16] A. Zappone, E. Björnson, L. Sanguinetti, and E. A. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2844–2859, Jun. 2017.

[17] A. H. Phan, H. D. Tuan, and H. H. Kha, "D.C. iterations for SINR maximin multicasting in cognitive radio," in *Proc. 6th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Gold Coast, Australia, Dec. 2012.

[18] J. Xu, Z. Miao, and Q. Liu, "New method to get essential efficient solution for a class of D.C. multiobjective problem," in *Proc. Int. Conf. Comput., Mechatronics, Control, Electron. Eng. (CMCE)*, Changchun, China, Oct. 2010.

[19] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[20] D. P. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4661–4674, Dec. 2005.

[21] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[22] O. Tervo, H. Pennanen, D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Distributed optimization for coordinated beamforming in multicell multigroup multicast systems: Power minimization and SINR balancing," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 171–185, Jan. 2018.

[23] M. Rossi, A. M. Tulino, O. Simeone, and A. M. Haimovich, "Non-convex utility maximization in Gaussian MISO broadcast and interference channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011.

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

[25] L. P. Qian, Y. J. Zhang, and J. Huang, "MAPEL: Achieving global optimality for a non-convex wireless power control problem," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1553–1563, 2009.

[26] E. Björnson and E. A. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*, ser. Found. Trends Commun. Inf. Theory. Now Publishers, 2013, vol. 9, no. 2-3.

[27] A. Zappone, L. Sanguinetti, G. Bacci, E. A. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1668–1683, Apr. 2016.

[28] O. Tervo, A. Tölli, M. Juntti, and L.-N. Tran, "Energy-efficient beam coordination strategies with rate dependent processing power," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 6097–6112, Nov. 2017.

[29] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless information and power transfer: energy efficiency optimization in OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6352–6370, Dec. 2013.

[30] ——, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[31] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, ser. SIAM Studies in Applied Mathematics. SIAM, 1994.

[32] A. Ben-Tal and A. Nemirovskii, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, ser. MPS-SIAM Series on Optimization. SIAM, 2001.

[33] A. S. Nemirovskii and D. B. Yudin, *Problem complexity and method efficiency in optimization.* Wiley, 1983.

[34] B. Matthiesen, Y. Yang, and E. A. Jorswieck, "Optimization of weighted individual energy efficiencies in interference networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018.

[35] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[36] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002. [Online]. Available: http://dx.doi.org/10.1109/JSAC.2002.1007390

[37] L. A. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. Autom. Control*, vol. 8, no. 1, pp. 59–60, Jan. 1963.

[38] E. A. Jorswieck and H. Boche, "Rate balancing for the multi-antenna Gaussian broadcast channel," in *Proc. IEEE 7th Int. Symp. Spread Spectr. Techn. Appl.*, Prague, Czech Republic, Sep. 2002.

[39] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[40] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[41] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[42] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.

[43] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[44] E. A. Jorswieck and H. Boche, "On the performance optimization in multiuser MIMO systems," *Trans. Emerg. Telecommun. Technol.*, vol. 18, no. 3, pp. 287–304, Apr. 2007.

[45] E. A. Jorswieck and E. G. Larsson, "Monotonic optimization framework for the two-user MISO interference channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2159–2168, Jul. 2010.

[46] C. Isheden, Z. Chong, E. A. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[47] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Res. Logistics Quart.*, vol. 9, no. 3-4, pp. 181–186, 1962.

[48] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[49] S. Schaible and T. Ibaraki, "Fractional programming," *Eur. J. Operational Res.*, vol. 12, no. 4, pp. 325–338, Apr. 1983.

[50] S. Schaible, "Fractional programming," in *Handbook of global optimization*, R. Horst and P. M. Pardalos, Eds. Kluwer Academic Publishers, 1993.

[51] A. Zappone and E. A. Jorswieck, *Energy Efficiency in Wireless Networks via Fractional Programming Theory*, ser. Found. Trends Commun. Inf. Theory. Now Publishers, 2015, vol. 11, no. 3-4.

[52] R. W. Freund and F. Jarre, "Solving the sum-of-ratios problem by an interior-point method," *J. Global Optim.*, vol. 19, no. 1, pp. 83–102, 2001.

[53] S. Schaible and J. Shi, "Fractional programming: The sum-of-ratios case," *Optim. Methods, Softw.*, vol. 18, no. 2, pp. 219–229, 2003.

[54] J.-P. Crouzeix, J. A. Ferland, and S. Schaible, "An algorithm for generalized fractional programs," *J. Optim. Theory Appl.*, vol. 47, no. 1, pp. 35–49, 1985.

[55] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Operational Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[56] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, Jul. 2006.

[57] D. E. Knuth, *The Art of Computer Programming, Volume 3: Sorting and Searching*, 2nd ed. Addison-Wesley, 1997.

[58] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang, *Generalized Concavity*. Plenum Press, 1988.

[59] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, "The multiway relay channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 51–63, Jan. 2013.

[60] A. Chaaban and A. Sezgin, *Multi-way Communications: An Information Theoretic Perspective*, ser. Found. Trends Commun. Inf. Theory. Now Publishers, 2015, vol. 12, no. 3-4.

[61] B. Matthiesen and E. A. Jorswieck, "Instantaneous relaying for the 3-way relay channel with circular message exchanges," in *Proc. 49th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2015, pp. 475–479.

[62] 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.

[63] Gurobi Optimization. (2018) Gurobi optimizer 8.0.1. [Online]. Available: http://gurobi.com

[64] B. Matthiesen, A. Zappone, and E. A. Jorswieck, "Resource allocation for energy-efficient 3-way relay channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4454–4468, Aug. 2015.

[65] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO networks: Spectral, energy, and hardware efficiency*, ser. Found. Trends Signal Process. Now Publishers, 2017, vol. 11, no. 3-4.

[66] MOSEK ApS. (2017) MOSEK optimizer 8.1.0.34. [Online]. Available: http://mosek.com

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2019.2941068

[67] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
[68] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

Bho Matthiesen (S'12) received the Diplom-Ingenieur (M.S.) degree in electrical engineering from the Technische Universität Dresden, Germany, in 2012. Since May 2012, he has been a Research Associate with the Chair of Communications Theory at Technische Universitatät Dresden, where he is currently finishing his Ph.D. degree. His research interests are in the area of signal processing and communication theory, with a focus on global optimization methods for resource allocation. He serves as publications chair for ISWCS 2020.

Eduard A. Jorswieck (S'01 – M'03 – SM'08) was born in 1975 in Berlin, Germany. Since August 2019, he has been the head of the Chair for Communications Systems and Full Professor at Technische Universität Braunschweig, Germany. From 2008 until 2019, he was the head of the Chair of Communications Theory and Full Professor at Dresden University of Technology (TUD), Germany.

His main research interests are in the broad area of communications. He has published more than 110 journal papers, 13 book chapters, 3 monographs, and some 275 conference papers on these topics.

Dr. Jorswieck is member of the IEEE SAM Technical Committee since 2015. Since 2017, he serves as Editor-in-Chief of the EURASIP Journal on Wireless Communications and Networking. He serves currently on the editorial board for IEEE Transactions on Information Forensics and Security. In 2006, he received the IEEE Signal Processing Society Best Paper Award.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.