

Linköping University Post Print

**Monotonic Optimization Framework for
theTwo-User MISO Interference Channel**

Eduard A. Jorswieck and Erik G. Larsson

N.B.: When citing this work, cite the original article.

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Eduard A. Jorswieck and Erik G. Larsson, Monotonic Optimization Framework for theTwo-User MISO Interference Channel, 2010, IEEE Transactions on Communications, (58), 7, 2159-2169.

<http://dx.doi.org/10.1109/TCOMM.2010.07.090068>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-58551>

Monotonic Optimization Framework for the Two-User MISO Interference Channel

Eduard A. Jorswieck and Erik G. Larsson

Abstract—Resource allocation and transmit optimization for the multiple-antenna Gaussian interference channel are important but difficult problems. The spatial degrees of freedom can be exploited to avoid, align, or utilize the interference. In recent literature, the upper boundary of the achievable rate region has been characterized. However, the resulting programming problems for finding the sum-rate, proportional fair, and minimax (egalitarian) operating points are non-linear and non-convex.

In this paper, we develop a non-convex optimization framework based on monotonic optimization by outer polyblock approximation. First, the objective functions are represented in terms of differences of monotonic increasing functions. Next, the problems are reformulated as maximization of increasing functions over normal constraint sets. Finally, the idea to approximate the constraint set by outer polyblocks is explained and the corresponding algorithm is derived. Numerical examples illustrate the advantages of the proposed framework compared to an exhaustive grid search approach.

Index Terms—Resource allocation, interference channel, multiple-antenna systems, non-convex optimization.

I. INTRODUCTION

INTERFERENCE channels (IFC) consist of at least two transmitters and two receivers. The first transmitter wants to transfer information to the first receiver and the second transmitter to the second receiver, respectively. This happens at the same time on the same frequency causing interference at the receivers. Information-theoretic studies of the IFC have a long history [1]–[3]. These references have provided various achievable rate regions, which are generally larger in the more recent papers than in the earlier ones. However, the capacity region of the general IFC remains an open problem. For certain limiting cases, for example when the interference is weak or very strong, respectively, the sum capacity is known [3], [4]. The sum-rate capacity for scalar IC under weak interference is obtained in [5]–[7]. If the interference is weak, it can simply be treated as additional noise. For very strong interference, the interference can be decoded and subtracted by treating the useful signals as noised at both receivers. [8] is the first paper

Paper approved by N. Jindal, the Editor for MIMO Techniques of the IEEE Communications Society. Manuscript received February 1, 2009; revised July 28, 2009.

E. Jorswieck is with the Dresden University of Technology, Communications Laboratory, Chair of Communication Theory, Dresden, Germany (e-mail: eduard.jorswieck@tu-dresden.de).

E. G. Larsson is with Linköping University, Dept. of Electrical Engineering (ISY), Division of Communication Systems, Linköping, Sweden (e-mail: erik.larsson@isy.liu.se).

This work was supported in part by the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF). E. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Parts of the material in this paper have been presented at the Allerton 2008 and ICASSP 2009 conferences.

Digital Object Identifier 10.1109/TCOMM.2010.07.090068

that considers the capacities of MIMO IFCs. [9] presents a numerical method to compute a lower bound for the sum-rate capacity of MIMO IFCs. The capacity regions and sum-rate capacities for MIMO IFC under strong and weak interference is obtained in [10] and in the low interference regime in [11]. In [12], the rate region of the single-input single-output (SISO) IFC was characterized in terms of convexity and concavity. The MIMO IFC was also studied from a non-cooperative game-theoretic point of view in [13].

The IFC is a building block in many communication systems, for example for ad-hoc networks and cognitive radio. It also specializes to scenarios with cooperation either at the transmitter or at the receiver side, leading to for instance, the multiple-access channel (MAC) and the broadcast channel (BC). For system design it is important to analyze the achievable rate region of the general Gaussian IFC (as will be defined in Section II) and to design transmit strategies that operate at certain operating points.

An explicit parameterization of the Pareto boundary for the K -user Gaussian MISO IFC, for the case when all multiuser interference is treated as additive Gaussian noise at the receivers, was derived in [14]. For the special case of two users, any point in the rate region can be achieved by choosing beamforming vectors that are linear combinations of the zero-forcing (ZF) and the maximum-ratio transmission (MRT) beamformers. Hence, all important (i.e., Pareto-efficient) operating points can be expressed by two real-valued parameters between zero and one $0 \leq \lambda = [\lambda_1, \lambda_2] \leq 1$.

In the current work, we build on the parameterization in [14] and focus on the maximum sum-rate operating point, the proportional-fair operating point and the max-min rate point. The corresponding optimization problems are non-convex problems which are difficult to solve directly. In particular, the max-min problem is non-smooth and therefore derivative-based (gradient) optimization approaches cannot be applied. A suboptimal iterative algorithm based on alternating projection was proposed in [15]. In general, this algorithm converges to a local optimum. Therefore, we are interested in formulating a unified non-convex optimization framework which takes as much as possible of the problem structure into account, and which is able to find the global optimum of the problems.

The main contribution of this work is the development of a *systematic approach* to solve the non-convex optimization problem. In contrast to exhaustive search methods, such as a grid search, the proposed approach has the advantage that it can achieve a given accuracy. In order to develop our systematic optimization algorithm, we perform the following steps:

- 1) We review the framework of monotonic optimization

and difference of monotonic functions (d.m.) maximization, and adapt it to the problem at hand (Section III).

- 2) We analyze the properties of the achievable rates as functions of λ_1 and λ_2 (Section IV-A).
- 3) We reformulate the programming problems as difference of increasing functions optimization problems (Section IV-B) and as a monotonic optimization problem on standard form (Section IV-C). Once on the standard form, we apply the polyblock optimization method of [16]. The resulting method converges to the global optimum within a given accuracy in a finite amount of time.

All theoretical results and the proposed algorithms are illustrated by numerical simulations in Section V. The results show the advantages of the monotonic optimization framework compared to simple exhaustive grid searches. The paper is concluded in Section VI.

Notation: We use standard notation for matrices, vectors, and their operators. Vectors are written in lowercase boldface (\mathbf{x}), matrices in capital boldface (\mathbf{X}). Transpose, Hermitian transpose, matrix inverse, and the conjugate of a matrix or vector are denoted by \mathbf{X}^T , \mathbf{X}^H , \mathbf{X}^{-1} , and \mathbf{X}^* . The set of non-negative (positive) real vectors of dimension n is denoted by \mathbb{R}_+^n (\mathbb{R}_{++}^n). The ℓ -2 (Euclidean) norm is denoted by $\|\mathbf{x}\|$. All inequalities are component-wise if not otherwise stated. More notation and definitions will be introduced when they are needed.

II. SYSTEM MODEL AND SUMMARY OF RECENT RESULTS

A. System model and transceiver structure

The system model of the MISO IFC and the corresponding transceiver structure is standard in the literature, and we describe it briefly in what follows. We consider two independent wireless systems that operate in the same spectral band. The first system consists of a transmitter TX₁ that wants to convey information to a receiver RX₁. The second system consists of another transmitter TX₂ that wants to transmit information to a receiver RX₂. The systems share the same spectrum, so the communications between TX₁ → RX₁ and TX₂ → RX₂ take place simultaneously on the same channel. Thus RX₁ will hear a superposition of the signals transmitted from TX₁ and TX₂, and conversely RX₂ will also receive the sum of the signals transmitted by both transmitters. This setup is recognized as an interference channel (IFC) [1]–[3]. In the setup we consider, TX₁ and TX₂ have n transmit antennas each, that can be used with full phase coherency. RX₁ and RX₂, however, have a single receive antenna each. Hence our problem setup constitutes a multiple-input single-output (MISO) IFC [8]. See Figure 1.

We assume that transmission consists of scalar coding followed by beamforming, and that all propagation channels are frequency-flat. In [17], it is shown that any Pareto-optimal transmit covariance matrix has rank one, i.e. single-stream beamforming is sufficient. This leads to the following basic model for the matched-filtered, symbol-sampled complex baseband data received at RX₁ and RX₂:

$$y_1 = \mathbf{h}_{11}^T \mathbf{w}_1 s_1 + \mathbf{h}_{21}^T \mathbf{w}_2 s_2 + e_1$$

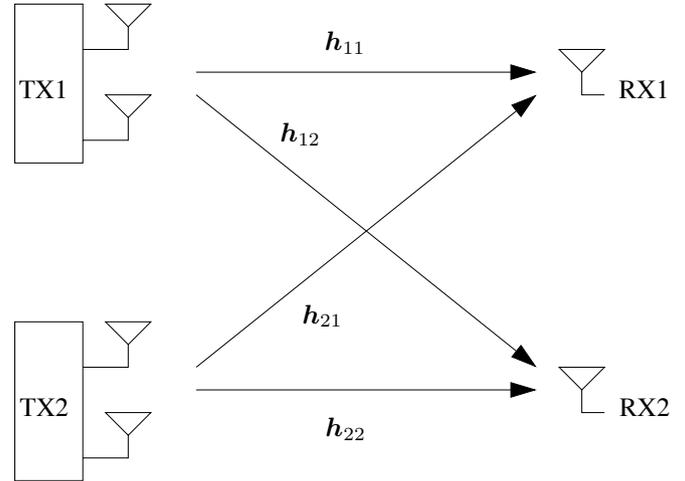


Fig. 1. The two-user MISO interference channel under study, illustrated for $n = 2$ transmit antennas.

$$y_2 = \mathbf{h}_{22}^T \mathbf{w}_2 s_2 + \mathbf{h}_{12}^T \mathbf{w}_1 s_1 + e_2$$

where s_1 and s_2 are transmitted symbols, \mathbf{h}_{ij} is the complex-valued $n \times 1$ channel-vector between TX _{i} and RX _{j} , and \mathbf{w}_i is the beamforming vector used by TX _{i} . The variables e_1 , e_2 are noise terms which we model as i.i.d. complex Gaussian with zero mean and variance σ^2 . We assume that each base station can use the transmit power P , but that power cannot be traded between the base stations.¹ Without loss of generality, we shall take $P = 1$. This gives the power constraint $\|\mathbf{w}_i\|^2 \leq 1$, $i = 1, 2$. Throughout, we define the signal-to-noise ratio (SNR) as $1/\sigma^2$. Various schemes that we will discuss require that the transmitters (TX₁ and TX₂) have different forms of channel state information (CSI). However, at no point we will require phase coherency between the base stations.

B. Recent, related results

The following beamformers are well known in literature and their operational meaning in a game-theoretic framework is studied in [18]. The MRT beamforming vectors are given by

$$\mathbf{w}_1^{\text{MRT}} = \frac{\mathbf{h}_{11}^*}{\|\mathbf{h}_{11}\|} \quad \text{and} \quad \mathbf{w}_2^{\text{MRT}} = \frac{\mathbf{h}_{22}^*}{\|\mathbf{h}_{22}\|}.$$

The ZF beamformers are given by

$$\mathbf{w}_1^{\text{ZF}} = \frac{\Pi_{\mathbf{h}_{12}^*}^\perp \mathbf{h}_{11}^*}{\|\Pi_{\mathbf{h}_{12}^*}^\perp \mathbf{h}_{11}^*\|} \quad \text{and} \quad \mathbf{w}_2^{\text{ZF}} = \frac{\Pi_{\mathbf{h}_{21}^*}^\perp \mathbf{h}_{22}^*}{\|\Pi_{\mathbf{h}_{21}^*}^\perp \mathbf{h}_{22}^*\|} \quad (1)$$

for TX₁ and TX₂, respectively, where $\Pi_{\mathbf{X}}^\perp = \mathbf{I} - \mathbf{X}(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$ denotes orthogonal projection onto the orthogonal complement of the column space of \mathbf{X} .

The following Theorem is proved in [15].

Theorem 1: Any point on the Pareto boundary of the rate

¹For simplicity of the exposition, we assume that both base stations operate under the same power constraint. At some additional expense of notation, our algorithms can be extended to the case of different power constraints.

region is achievable with the beamforming strategies

$$\begin{aligned} \mathbf{w}_1(\lambda_1) &= \frac{\lambda_1 \mathbf{w}_1^{\text{MRT}} + (1 - \lambda_1) \mathbf{w}_1^{\text{ZF}}}{\|\lambda_1 \mathbf{w}_1^{\text{MRT}} + (1 - \lambda_1) \mathbf{w}_1^{\text{ZF}}\|} \quad \text{and} \\ \mathbf{w}_2(\lambda_2) &= \frac{\lambda_2 \mathbf{w}_2^{\text{MRT}} + (1 - \lambda_2) \mathbf{w}_2^{\text{ZF}}}{\|\lambda_2 \mathbf{w}_2^{\text{MRT}} + (1 - \lambda_2) \mathbf{w}_2^{\text{ZF}}\|} \end{aligned} \quad (2)$$

for some $0 \leq \lambda_1, \lambda_2 \leq 1$.

The achievable rates as functions of the parameter vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$ read

$$\begin{aligned} R_1(\boldsymbol{\lambda}) &= \log \left(1 + \frac{|\mathbf{w}_1^T(\lambda_1) \mathbf{h}_{11}|^2}{\sigma^2 + |\mathbf{w}_2^T(\lambda_2) \mathbf{h}_{21}|^2} \right) \\ R_2(\boldsymbol{\lambda}) &= \log \left(1 + \frac{|\mathbf{w}_2^T(\lambda_2) \mathbf{h}_{22}|^2}{\sigma^2 + |\mathbf{w}_1^T(\lambda_1) \mathbf{h}_{12}|^2} \right). \end{aligned} \quad (3)$$

Theorem 1 shows that the ZF and MRT beamformers stand out because all interesting (Pareto-optimal) beamforming vectors are linear combinations of them. The ZF and MRT beamformers also have another interesting property: they yield the sum-rate point at high and low SNRs. More precisely, we have the following two theorems, which were first shown in [19]²:

Theorem 2: At high SNR, ZF is sum-rate optimal. More precisely

$$\lim_{\sigma \rightarrow 0} \arg \max_{\|\mathbf{w}_1\|^2 \leq 1, \|\mathbf{w}_2\|^2 \leq 1} \{R_1(\mathbf{w}_1, \mathbf{w}_2) + R_2(\mathbf{w}_1, \mathbf{w}_2)\} = (\mathbf{w}_1^{\text{ZF}}, \mathbf{w}_2^{\text{ZF}}).$$

(Here $R_i(\mathbf{w}_1, \mathbf{w}_2)$ denote the rates as functions of the beamforming vectors.)

Theorem 3: At low SNR, MRT is sum-rate optimal. More precisely

$$\lim_{\sigma \rightarrow \infty} \arg \max_{\|\mathbf{w}_1\|^2 \leq 1, \|\mathbf{w}_2\|^2 \leq 1} \{R_1(\mathbf{w}_1, \mathbf{w}_2) + R_2(\mathbf{w}_1, \mathbf{w}_2)\} = (\mathbf{w}_1^{\text{MRT}}, \mathbf{w}_2^{\text{MRT}}).$$

Theorems 2 and 3 are intuitively appealing, but their proofs are nontrivial; see [19].

C. Problem statement

We are interested in efficient algorithms for finding the following operating points:

- 1) The weighted sum-rate point:

$$\max_{0 \leq \omega \leq 1} \{\omega R_1(\boldsymbol{\lambda}) + (1 - \omega) R_2(\boldsymbol{\lambda})\} \quad (4)$$

where ω , $0 \leq \omega \leq 1$ is a weighting factor.

- 2) The proportional-fairness operating point:

$$\max_{0 \leq \boldsymbol{\lambda} \leq 1} \{R_1(\boldsymbol{\lambda}) \cdot R_2(\boldsymbol{\lambda})\}. \quad (5)$$

- 3) The max-min optimal point (egalitarian solution):

$$\max_{0 \leq \boldsymbol{\lambda} \leq 1} \min\{R_1(\boldsymbol{\lambda}), R_2(\boldsymbol{\lambda})\}. \quad (6)$$

²In [19] there is a misprint in the proof of Lemma 2 of Appendix I (page 712). The definition of the function $f_\sigma(\alpha_1, \alpha_2)$ in equation (18) should be $f_\sigma(\alpha_1, \alpha_2) = (1 + (\psi_1/\sigma)^2)(1 + (\psi_2/\sigma)^2) \times 2^{-R_\sigma(\alpha_1, \alpha_2)} = \dots$. The next row should indicate that 'the function 2^{-x} is strictly decreasing' as well as that $(1 + (\psi_1/\sigma)^2)(1 + (\psi_2/\sigma)^2)$ is a positive constant. Finally, the equation on the following row should read 'arg max' on the left-hand side. The authors acknowledge the help of Junwei Zhang and Danyo Danev for pointing this out and correcting it.

All three optimization problems (4), (5), and (6) are non-linear and non-convex. In [15] we proposed an iterative algorithm for solving them, based on cyclic optimization. However, this algorithm does not necessarily converge to the global optimum. Among algorithms that we are aware of up to this point, only an exhaustive grid search over $\boldsymbol{\lambda} \in [0, 1]^2$ could guarantee that the global optimum is approximatively found. In the following two sections, we propose a new approach that finds the global solution to the problems (4), (5) and (6) to within a given accuracy and in a finite number of steps. This is our main contribution.

Before we proceed, we note that [20] derives an algorithm called MAPEL to solve the problem of weighted sum-rate maximization for the single-antenna flat-fading interference channel. There, the non-convex problem is first transformed into a multiplicative linear fractional programming (MLFP) problem. This type of problem is one particular instance of a non-convex programming problem which can be solved using the framework of monotonic optimization [16, Section 8.1].

In contrast to the power allocation problem treated in [20], the beamforming problems in (4), (5), and (6) cannot be expressed as MLFP problems. This is so because the signal and interference power terms in (3), for example $|\mathbf{w}_1^T(\lambda_1) \mathbf{h}_{11}|^2$, are not affine in $\boldsymbol{\lambda}$. However, our proposed algorithm and the methods in [20] stand on a common ground as both problems can be solved by using the monotonic optimization framework.

III. PRELIMINARIES: MONOTONIC OPTIMIZATION

A. Increasing functions and normal sets

At first, we need the basic concepts of *increasing functions* and *normal sets*. This material is contained partly in [16]. However, we need the notion of *strictly increasing function* and therefore we provide a complete presentation and some alternative proofs.

Definition 1: For two vectors $\mathbf{x}', \mathbf{x} \in \mathbb{R}^n$ we write $\mathbf{x}' \geq \mathbf{x}$ and say that \mathbf{x}' dominates \mathbf{x} if $x'_i \geq x_i$ for all $i = 1, \dots, n$. We write $\mathbf{x}' > \mathbf{x}$ and say that \mathbf{x}' strictly dominates \mathbf{x} if $x'_i > x_i$ for all $i = 1, \dots, n$.

Note that the domination only induces a partial ordering because not all vectors can be compared. For example, if $\mathbf{x} = [1, 2]$ and $\mathbf{x}' = [2, 1]$ then we have neither $\mathbf{x} \geq \mathbf{x}'$ nor $\mathbf{x}' \geq \mathbf{x}$. The order in Definition 1 can be used to define the class of order preserving functions as follows.

Definition 2: A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be increasing on \mathbb{R}_+^n if $f(\mathbf{x}) \leq f(\mathbf{x}')$ whenever $0 \leq \mathbf{x} \leq \mathbf{x}'$. The function is said to be increasing in the box $[a, b]^n \subset \mathbb{R}_+^n$ if $f(\mathbf{x}) \leq f(\mathbf{x}')$ whenever $a\mathbf{1} \leq \mathbf{x} \leq \mathbf{x}' \leq b\mathbf{1}$. A function is said to be strictly increasing if for $\mathbf{x}' \geq \mathbf{x} \geq 0$ and $\mathbf{x}' \neq \mathbf{x}$ follows that $f(\mathbf{x}') > f(\mathbf{x})$. (Here $\mathbf{1} = [1, \dots, 1]^T$.)

Many functions encountered in resource allocation problems are increasing in the sense of Definition 2. For example, the sum-rate capacity of a multiple-access channel (MAC) [21] is increasing in the vector of powers allocated to the users: $f(\mathbf{p}) = \log \left(1 + \text{SNR} \cdot \sum_{k=1}^K p_k a_k \right)$ where a_1, \dots, a_K are squared channel gains.

If the domain of these increasing functions is a so-called normal set (to be defined next), we will later obtain a characterization of the set on which the maximum is achieved.

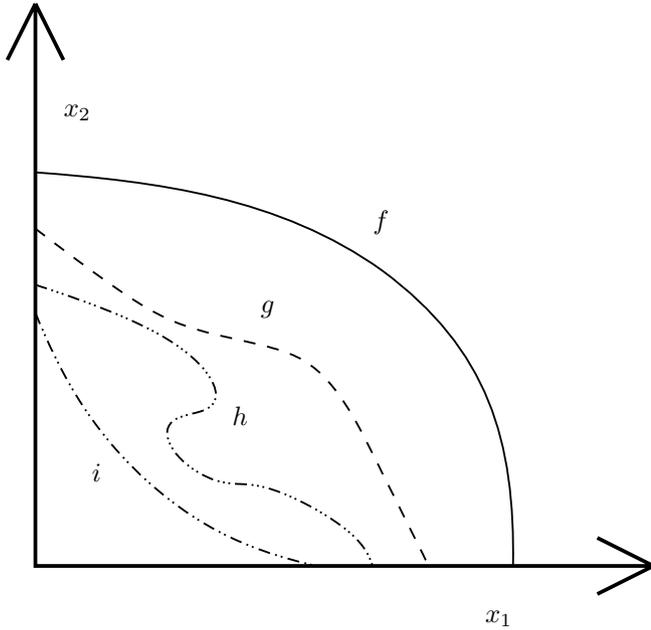


Fig. 2. Example of sets that are convex, normal, and neither convex nor normal.

A set G is said to be *normal* if for all $\mathbf{x} \in G$ all points in the box $[\mathbf{0}, \mathbf{x}]$ are also in G . See Figure 2. More precisely:

Definition 3: A set $G \subset \mathbb{R}_+^n$ is called normal if for any two points $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_+^n$ such that $\mathbf{x}' \leq \mathbf{x}$, if $\mathbf{x} \in G$, then $\mathbf{x}' \in G$, too.

The empty set, the singleton $\{0\}$, and \mathbb{R}_+^n are special normal sets. We refer to them as trivial normal subsets of \mathbb{R}_+^n .

In Figure 2, the set induced by f is convex and normal, and the sets g and i are normal but not convex. However, the set induced by h is neither convex nor normal.

For the characterization of the maximum of an increasing function over a normal set, we need the notion of *upper boundary*.

Definition 4: A point $\mathbf{y} \in \mathbb{R}_+^n$ is called an upper boundary point of a bounded closed normal set \mathcal{D} if $\mathbf{y} \in \mathcal{D}$ while the set $K_{\mathbf{y}} = \mathbf{y} + \mathbb{R}_+^n = \{\mathbf{y}' \in \mathbb{R}_+^n | \mathbf{y}' > \mathbf{y}\}$ lies outside \mathcal{D} , i.e.

$$K_{\mathbf{y}} \subset \mathbb{R}_+^n \setminus \mathcal{D}.$$

The set of upper boundary points of \mathcal{D} is called the upper boundary of \mathcal{D} and it is denoted by $\partial^+ \mathcal{D}$.

The following result shows that the maximum of a strictly increasing function over a normal set is always achieved on the upper boundary of the normal set. The statement is somewhat weaker than Proposition 7 in [16]. However, for our purposes we only need the following version and provide an alternative proof by contradiction.

Proposition 1: The global maximum of a strictly increasing function $f(\mathbf{x})$ over a normal set \mathcal{D} , if it exists, is attained on $\partial^+ \mathcal{D}$.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point where $f(\mathbf{x})$ attains a local maximum and that $\mathbf{x} \notin \partial^+ \mathcal{D}$. Then there exists a $\mathbf{y} \in \mathcal{D}$ with $\mathbf{x} \leq \mathbf{y}$ and for which at least one component in \mathbf{y} is larger than in \mathbf{x} . Since $f(\mathbf{x})$ is strictly increasing, we must have $f(\mathbf{x}) < f(\mathbf{y})$. ■

B. Monotonic optimization and outer polyblock approximation

We next give some background on monotonic optimization. A monotonic optimization problem on the standard form [22] is given by

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{D} \quad (7)$$

where \mathcal{D} is a normal, but not necessarily convex set. We assume that \mathcal{D} is normalized such that the smallest box containing \mathcal{D} is the unit box.

The main difficulty involved in solving the problem (7) is that the constraint set is non-convex. However, using the polyblock approach it turns out that there is an interesting duality between optimization of increasing functions over normal sets and optimization of convex functions over convex sets [23].

From Proposition 1 we know that the maximum of $f(\mathbf{x})$ over \mathcal{D} is attained at the upper boundary $\partial^+ \mathcal{D}$. The main idea to solve the non-convex optimization problem (7) is to approximate $\partial^+ \mathcal{D}$ by polyblocks since the global maximum lies on the upper boundary.

Definition 5: A set $P \subset \mathbb{R}_+^n$ is called a polyblock if it is the union of a finite number of boxes.

The polyblock P is generated by a set of vertices T . The minimal set of vertices consists of only proper vertices, i.e., vertices which are not dominated by any other vertex in T . It follows that for all $\mathbf{z}, \mathbf{z}' \in T$ with $\mathbf{z} \neq \mathbf{z}'$ it does not hold $\mathbf{z} > \mathbf{z}'$ or $\mathbf{z} < \mathbf{z}'$. Another important consequence of Proposition 1 is that the maximum of an increasing function over a polyblock is achieved at a proper vertex.

The main idea of the outer polyblock algorithm is as follows: Construct a nested sequence of polyblocks which approximate the normal set \mathcal{D} from above

$$P_1 \supset P_2 \supset \dots \supset \mathcal{D} \quad \text{such that} \quad \max_{\mathbf{x} \in P_k} f(\mathbf{x}) \searrow \max_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x}) \quad (8)$$

where $x_k \searrow x$ means that $x_k \rightarrow x$ when $k \rightarrow \infty$ and that $x_k \geq x_l \geq x$ for all $l \geq k$. The main steps of the outer polyblock algorithm are described next. Define the maximizer at iteration k as

$$\tilde{\mathbf{x}}^{(k)} \in \arg \max_{\mathbf{x} \in T_k} f(\mathbf{x}), \quad (9)$$

where T_k is the minimal vertex set of P_k . The first step is to construct the nested sequence in (8), i.e., to construct a new polyblock P_{k+1} contained in $P_k \setminus \{\tilde{\mathbf{x}}^{(k)}\}$ but still containing \mathcal{D} . This step is motivated in Propositions 17 and 18 in [16]. However, we provide an alternative description for convenience and completeness.

Let the set of vertices in step k be $T_k = \{\mathbf{x}_1^{(k)}, \dots, \mathbf{x}_{K^{(k)}}^{(k)}\}$. Denote $\bar{\mathbf{x}}^{(k)}$ as the unique intersection point of $\partial^+ \mathcal{D}$ and $\delta \tilde{\mathbf{x}}^{(k)}$ with $\delta \in [0, 1]$. Then the set of (not necessarily minimal) vertices in step $k+1$ is constructed as follows

$$T_{k+1} = T_k \setminus \{\tilde{\mathbf{x}}^{(k)}\} \cup \bigcup_{\nu=1}^n \{\tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_{\nu}^{(k)}] \mathbf{e}_{\nu}\} \quad (10)$$

where \mathbf{e}_n is the n th column of the identity matrix. The construction of the vertices in step $k+1$ is illustrated in Figure 3. The new vertices are $\mathbf{x}_{+1}^{(k)} = \tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_{\nu}^{(k)}] \mathbf{e}_1$,

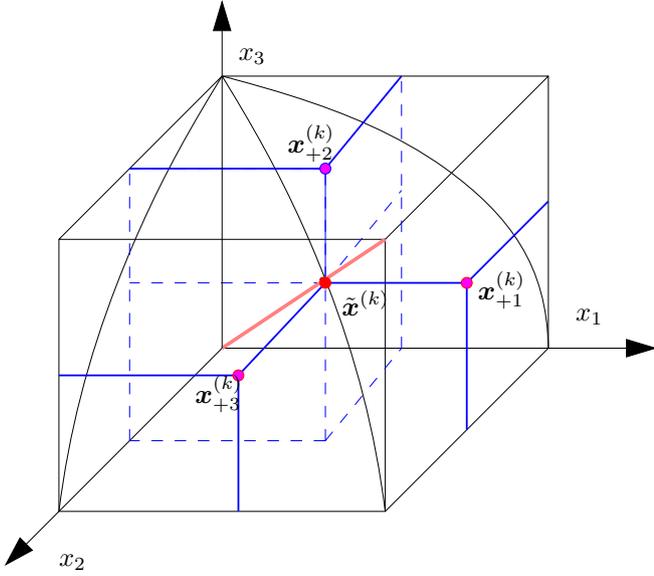


Fig. 3. Construction of vertices in step $k + 1$ in (10).

$\mathbf{x}_{+2}^{(k)} = \tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_2^{(k)}]e_2$, and $\mathbf{x}_{+3}^{(k)} = \tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_3^{(k)}]e_3$ as described in (10). Let P_k and P_{k+1} be the polyblocks induced by the minimal set of vertices T_k and T_{k+1} , respectively.

Proposition 2: The constructed polyblocks P_k and P_{k+1} fulfill

$$\mathcal{D} \subset P_{k+1} \subset P_k \setminus \{\tilde{\mathbf{x}}^{(k)}\}. \quad (11)$$

Proof: In order to show the second relation $P_{k+1} \subset P_k \setminus \{\tilde{\mathbf{x}}^{(k)}\}$ it suffices to verify that for each vertex $\mathbf{x}_{\ell}^{(k)}$ in T_k with $1 \leq \ell \leq K^{(k)}$ there exists a vertex $\mathbf{x}_{\bar{\ell}}^{(k+1)}$ in T_{k+1} such that $\mathbf{x}_{\ell}^{(k)} \geq \mathbf{x}_{\bar{\ell}}^{(k+1)}$. Let $\bar{\ell}$ be the index which belongs to the maximal vertex in T_k , i.e., $\bar{\ell} = \arg \max_{1 \leq \ell \leq K^{(k)}} f(\mathbf{x}_{\ell}^{(k)})$. For all $\ell \neq \bar{\ell}$ there is a $\mathbf{x}_{\bar{\ell}}^{(k+1)} = \mathbf{x}_{\ell}^{(k)}$ because T_{k+1} still contains these non-maximum vertices (or after removing of all dominated vertices they are dominated). For $\bar{\ell}$ there are n new vertices $\mathbf{x}_{\nu}^{(k+1)} = \{\tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_{\nu}^{(k)}]e_{\nu}\}$ for $1 \leq \nu \leq n$ with $\mathbf{x}_{\nu}^{(k+1)} \geq \mathbf{x}_{\bar{\ell}}^{(k)}$.

In order to prove the first relation $\mathcal{D} \subset P_{k+1}$ we have to find for all boundary points $\mathbf{d} \in \partial^+ \mathcal{D}$ a vertex $\mathbf{x}_{\ell}^{(k+1)}$ in T_{k+1} such that $\mathbf{d} \leq \mathbf{x}_{\ell}^{(k+1)}$. For all upper boundary points $\mathbf{d} \in \partial^+ \mathcal{D}$ with $\mathbf{d} \leq \mathbf{x}_{\bar{\ell}}^{(k)}$ and $\ell \neq \bar{\ell}$ we find immediately the corresponding vertex in T_{k+1} because these vertices were not removed in (10). For the upper boundary points $\mathbf{d} \in \partial^+ \mathcal{D}$ for which $\mathbf{d} \leq \mathbf{x}_{\bar{\ell}}^{(k)}$ we find one of the n new vertices $\mathbf{x}_{\nu}^{(k+1)} = \{\tilde{\mathbf{x}}^{(k)} - [\tilde{x}_{\nu}^{(k)} - \bar{x}_{\nu}^{(k)}]e_{\nu}\}$ for $1 \leq \nu \leq n$ such that $\mathbf{d} \leq \mathbf{x}_{\nu}^{(k+1)}$. ■

Finally, we can remove all dominated vertices of T_{k+1} to obtain the minimal vertex set needed for the next step $k + 2$.

C. Outer polyblock algorithm and stopping criteria

The general outer polyblock algorithm is described in Algorithm 1. There are three stopping criteria ϵ -, and η -accuracy reached, or maximum number of steps exceeded.

In Line 7, the search for the intersection point is a scalar optimization problem in $0 \leq \delta \leq 1$ and simple Newton

Result: Solve optimization problem (7)

Input: Constraint set \mathcal{D} , accuracies ϵ and η .

```

1 initialization: Set  $T = \mathbf{1}$ ,  $k = 1$ ;
2 while  $\epsilon, \eta$ -accuracy and maximum number of steps is not
   reached do
3    $\mathbf{x}^{(k)} = \arg \max\{f(\mathbf{x}) | \mathbf{x} \in T, \mathbf{x} \geq \epsilon \mathbf{1}\}$ ;
4   if  $\mathbf{x}^{(k)} \in \mathcal{D}$  then
5      $\mathbf{x}^* = \mathbf{x}^{(k)}$  is  $\epsilon$ -optimal solution;
6   else
7     Compute the intersection point  $\mathbf{y}^{(k)}$  of  $\partial^+ \mathcal{D}$  with
        $\delta \mathbf{x}^{(k)}$  with  $0 \leq \delta \leq 1$ ;
8      $\bar{\mathbf{y}}^{(k)} = \arg \max\{f(\bar{\mathbf{y}}^{(k-1)}), f(\mathbf{y}^{(k)})\}$ ;
9     if  $f(\bar{\mathbf{y}}^{(k)}) \geq f(\mathbf{x}^{(k)}) - \eta$  then
10       $\mathbf{x}^* = \bar{\mathbf{y}}^{(k)}$  is an  $(\epsilon, \eta)$ -approximate solution of
        (7);
11    else
12      Compute  $n$  extreme points of the rectangle
         $[\mathbf{y}^{(k)}, \mathbf{x}^{(k)}]$  that are adjacent to  $\mathbf{x}^{(k)}$ :
13       $\mathbf{x}^{(k),i} = \mathbf{x}^{(k)} - (\mathbf{x}_i^{(k)} - \mathbf{y}_i^{(k)})e^i$  for  $1 \leq i \leq n$ ;
14       $Z = [T \setminus \{\mathbf{x}^{(k)}\}] \cup \{\mathbf{x}^{(k),1}, \dots, \mathbf{x}^{(k),n}\}$ ;
         $T$  is obtained from  $Z$  after dropping all
        vectors which are dominated by others;
15    end
16  end
17   $k = k + 1$ ;
18 end

```

Output: Solution \mathbf{x}^* to (7)

Algorithm 1: Generalized outer polyblock algorithm

methods could be used. In the implementation, we used Bolzano's bisection procedure as suggested in [22, Section 8] to compute the intersection point.

Given a tolerance $\epsilon > 0$, denote

$$\mathcal{D}^{\epsilon} = \{\mathbf{x} \in \mathcal{D} | x_i \geq \epsilon, \quad i = 1, \dots, n\}.$$

Assuming ϵ is small but positive such that $\mathcal{D}^{\epsilon} \neq \emptyset$, a global solution of the problem

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{D}^{\epsilon} \quad (12)$$

will be called an ϵ -optimal solution of (7). A solution $\bar{\mathbf{x}} \in \mathcal{D}$, such that $f(\bar{\mathbf{x}})$ differs from the optimal value of (12) by at most $\eta > 0$, will be referred to as an (ϵ, η) -approximate optimal solution of (12).

Since it is not guaranteed that the algorithm stops within any fixed number of steps K , i.e., after K steps neither ϵ - nor η -accuracy might be reached, we additionally set a maximum number of steps. However, Theorem 1 in [22] shows that the algorithm terminates after a finite number of steps. Therefore, ϵ and η could be also increased until the algorithms converges in a number of K steps.

IV. SOLUTION BY MONOTONIC OPTIMIZATION

The monotonic optimization framework described above is now applied to the problem statements from Section II-C. First, the properties of our objective functions are analyzed and next the programming problems are reformulated in standard form as described in (7).

A. Properties of the achievable rates

In this section, we show that the atom functions of the individual user rates in (3) are strictly increasing. Thus, the user rates can be expressed as the difference of two strictly increasing functions. For future use, let us define the following quantities

$$\begin{aligned}\gamma_{11}^2 &= \|\mathbf{h}_{11}\|^2, & \gamma_{12}^2 &= \|\Pi_{\mathbf{h}_{12}}^\perp \mathbf{h}_{11}\|^2 \\ \gamma_{22}^2 &= \|\mathbf{h}_{22}\|^2, & \gamma_{21}^2 &= \|\Pi_{\mathbf{h}_{21}}^\perp \mathbf{h}_{22}\|^2.\end{aligned}\quad (13)$$

Obviously, it holds

$$\gamma_{11} \geq \gamma_{12} \quad \text{and} \quad \gamma_{22} \geq \gamma_{21}.\quad (14)$$

Define further the functions

$$f_1(\boldsymbol{\lambda}) = \log(\sigma^2 + |\mathbf{w}_1^T(\lambda_1)\mathbf{h}_{11}|^2 + |\mathbf{w}_2^T(\lambda_2)\mathbf{h}_{21}|^2) \quad (15)$$

$$f_2(\boldsymbol{\lambda}) = \log(\sigma^2 + |\mathbf{w}_2^T(\lambda_2)\mathbf{h}_{22}|^2 + |\mathbf{w}_1^T(\lambda_1)\mathbf{h}_{12}|^2) \quad (16)$$

$$g_1(\boldsymbol{\lambda}) = \log(\sigma^2 + |\mathbf{w}_1^T(\lambda_1)\mathbf{h}_{12}|^2) \quad (17)$$

$$g_2(\boldsymbol{\lambda}) = \log(\sigma^2 + |\mathbf{w}_2^T(\lambda_2)\mathbf{h}_{21}|^2). \quad (18)$$

Finally, $f(\boldsymbol{\lambda}) = f_1(\boldsymbol{\lambda}) + f_2(\boldsymbol{\lambda})$ and $g(\boldsymbol{\lambda}) = g_1(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda})$.

Lemma 1: The functions $f_1(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}), f(\boldsymbol{\lambda})$ as well as $g_1(\boldsymbol{\lambda}), g_2(\boldsymbol{\lambda}), g(\boldsymbol{\lambda})$ are strictly increasing, i.e., monotonically increasing in λ_1 and λ_2 .

Proof: All six functions depend on λ_1 or λ_2 via the following terms

$$\begin{aligned}\alpha_1(\lambda_1) &= |\mathbf{w}_1^T(\lambda_1)\mathbf{h}_{11}|^2 \\ &= \frac{|(\lambda_1\mathbf{w}_1^{\text{MRT}} + (1-\lambda_1)\mathbf{w}_1^{\text{ZF}})^T \mathbf{h}_{11}|^2}{\|\lambda_1\mathbf{w}_1^{\text{MRT}} + (1-\lambda_1)\mathbf{w}_1^{\text{ZF}}\|^2} \\ &= \frac{\left(\lambda_1\|\mathbf{h}_{11}\| + \frac{(1-\lambda_1)}{\|\Pi_{\mathbf{h}_{12}}^\perp \mathbf{h}_{11}\|} \mathbf{h}_{11}^H \Pi_{\mathbf{h}_{12}}^\perp \mathbf{h}_{11}\right)^2}{\lambda_1^2 + (1-\lambda_1)^2 + 2\lambda_1(1-\lambda_1) \frac{\|\mathbf{h}_{11}^H \Pi_{\mathbf{h}_{12}}^\perp \mathbf{h}_{11}\|}{\|\mathbf{h}_{11}\|}} \\ &= \frac{\lambda_1^2 \gamma_{11}^2 + (1-\lambda_1)^2 \gamma_{12}^2 + 2\lambda_1(1-\lambda_1) \gamma_{11} \gamma_{12}}{\lambda_1^2 + (1-\lambda_1)^2 + 2\lambda_1(1-\lambda_1) \frac{\gamma_{12}}{\gamma_{11}}} \\ &= \frac{(\lambda_1 \gamma_{11} + (1-\lambda_1) \gamma_{12})^2}{1 - 2\lambda_1(1-\lambda_1)(1 - \frac{\gamma_{12}}{\gamma_{11}})}.\end{aligned}\quad (19)$$

Similarly, we obtain

$$\begin{aligned}\alpha_2(\lambda_2) &= |\mathbf{w}_2^T(\lambda_2)\mathbf{h}_{22}|^2 \\ &= \frac{(\lambda_2 \gamma_{22} + (1-\lambda_2) \gamma_{21})^2}{1 - 2\lambda_2(1-\lambda_2)(1 - \frac{\gamma_{21}}{\gamma_{22}})},\end{aligned}\quad (20)$$

$$\begin{aligned}\beta_1(\lambda_1) &= |\mathbf{w}_1^T(\lambda_1)\mathbf{h}_{12}|^2 \\ &= \frac{\lambda_1^2 \gamma_{11}^2}{1 - 2\lambda_1(1-\lambda_1)(1 - \frac{\gamma_{12}}{\gamma_{11}})}\end{aligned}\quad (21)$$

$$\begin{aligned}\beta_2(\lambda_2) &= |\mathbf{w}_2^T(\lambda_2)\mathbf{h}_{21}|^2 \\ &= \frac{\lambda_2^2 \gamma_{22}^2}{1 - 2\lambda_2(1-\lambda_2)(1 - \frac{\gamma_{21}}{\gamma_{22}})}.\end{aligned}\quad (22)$$

Next, the first derivatives with respect to λ_1 or λ_2 are computed directly as

$$\begin{aligned}\frac{d\alpha_1(\lambda_1)}{d\lambda_1} &= \frac{1}{(\sim)^2} \cdot (2(\lambda_1(\gamma_{11} - \gamma_{12}) + \gamma_{12})\gamma_{11}(\gamma_{11} - \gamma_{12}) \\ &\quad \cdot (\gamma_{11}(1-\lambda_1) + \gamma_{12}(1-\lambda_1))) \geq 0\end{aligned}\quad (23)$$

where the last inequality follows from (14). The monotonicity of $\alpha_2(\lambda_2)$ follows similarly. The first derivatives of $\beta_1(\lambda_1)$ with respect to λ_1 is given by

$$\frac{d\beta_1(\lambda_1)}{d\lambda_1} = \frac{2\lambda_1\gamma_{11}^3(\gamma_{11}(1-\lambda_1) + \lambda_1\gamma_{12})}{(\sim)^2} \quad (24)$$

where (\sim) in (23) and (24) is given by $\gamma_{11} - 2\lambda_1\gamma_{11} + 2\lambda_1\gamma_{12} + 2\lambda_1^2\gamma_{11} - 2\lambda_1^2\gamma_{12}$.

Since $f(\lambda_1, \lambda_2)$ and $g(\lambda_1, \lambda_2)$ can be expressed as

$$\begin{aligned}f(\boldsymbol{\lambda}) &= \log(\sigma^2 + \alpha_1(\lambda_1) + \beta_2(\lambda_2)) \\ &\quad + \log(\sigma^2 + \alpha_2(\lambda_2) + \beta_1(\lambda_1)) \\ g(\boldsymbol{\lambda}) &= \log(\sigma^2 + \beta_2(\lambda_2)) + \log(\sigma^2 + \beta_1(\lambda_1))\end{aligned}\quad (25)$$

the result in Lemma 1 follows from (23) and (24). \blacksquare

B. Reformulation as d.m. problems

It is shown in [16] that the class of d.m. functions is rich, i.e., it does not only contain the sum or product of R_1 and R_2 but also other combinations including minimization and maximization. The d.m. property is invariant under certain transformations, as detailed in the following proposition.

Proposition 3 (Prop. 19 in [16]): If $\mu_1(\mathbf{x}), \dots, \mu_m(\mathbf{x})$ are d.m. then

- 1) for any $\alpha_i \in \mathbb{R}$ the function $\sum_{i=1}^m \alpha_i \mu_i(\mathbf{x})$ is also d.m.;
- 2) the function $\max\{\mu_1(\mathbf{x}), \dots, \mu_m(\mathbf{x})\}$ as well as $\min\{\mu_1(\mathbf{x}), \dots, \mu_m(\mathbf{x})\}$ is also d.m.

Based on this result, the next three corollaries show that the weighted sum-rate maximization problem in (4) as well as the proportional fair rate maximization problem in (5) and the max-min problem in (6) are d.m. programming problems.

Corollary 1: The maximum weighted sum-rate problem for weight $0 \leq \omega \leq 1$

$$\max_{0 \leq \lambda \leq 1} \omega R_1(\boldsymbol{\lambda}) + (1-\omega)R_2(\boldsymbol{\lambda})$$

is a d.m. programming problem.

Proof: The result follows as a corollary from Lemma 1 because the objective function can be rewritten as

$$\begin{aligned}&\omega R_1(\boldsymbol{\lambda}) + (1-\omega)R_2(\boldsymbol{\lambda}) \\ &= \omega[f_1(\boldsymbol{\lambda}) - g_2(\boldsymbol{\lambda})] + (1-\omega)[f_2(\boldsymbol{\lambda}) - g_1(\boldsymbol{\lambda})] \quad (26) \\ &= \underbrace{\omega f_1(\boldsymbol{\lambda}) + (1-\omega)f_2(\boldsymbol{\lambda})}_{\text{mon. incr.}} - \underbrace{(\omega g_2(\boldsymbol{\lambda}) + (1-\omega)g_1(\boldsymbol{\lambda}))}_{\text{mon. incr.}}.\end{aligned}$$

Corollary 2: The proportional-fair rate maximization problem

$$\max_{0 \leq \lambda \leq 1} R_1(\boldsymbol{\lambda})R_2(\boldsymbol{\lambda})$$

is a d.m. programming problem.

Proof: We start again with the expression for the rates from above $R_1 = f_1(\boldsymbol{\lambda}) - g_2(\boldsymbol{\lambda})$ and $R_2 = f_2(\boldsymbol{\lambda}) - g_1(\boldsymbol{\lambda})$ with strictly increasing $f_1, f_2, g_1,$ and g_2 . Expand the product $R_1 R_2$ to obtain

$$\begin{aligned}R_1 R_2 &= (f_1(\boldsymbol{\lambda}) - g_2(\boldsymbol{\lambda}))(f_2(\boldsymbol{\lambda}) - g_1(\boldsymbol{\lambda})) \\ &= \underbrace{f_1(\boldsymbol{\lambda})f_2(\boldsymbol{\lambda}) + g_1(\boldsymbol{\lambda})g_2(\boldsymbol{\lambda})}_{\text{mon.incr.}}\end{aligned}$$

$$-\underbrace{(f_1(\boldsymbol{\lambda})g_1(\boldsymbol{\lambda}) + f_2(\boldsymbol{\lambda})g_2(\boldsymbol{\lambda}))}_{\text{mon.incr.}} \quad (27)$$

which is again the difference of two monotonic functions. ■

From Corollaries 1 and 2 it can be observed that any linear combination and polynomial in $f_1(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}), g_1(\boldsymbol{\lambda})$, and $g_2(\boldsymbol{\lambda})$ can be expressed by expanding and collecting positive and negative parts as a d.m. function.

The following decomposition shows how to deal with the max-min problem in (6). The minimum of R_1 and R_2 can be written as

$$\begin{aligned} & \min(R_1(\boldsymbol{\lambda}), R_2(\boldsymbol{\lambda})) \quad (28) \\ &= \min(f_1(\boldsymbol{\lambda}) - g_2(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}) - g_1(\boldsymbol{\lambda})) \\ &= \min(f_1(\boldsymbol{\lambda}) + g_1(\boldsymbol{\lambda}) - (g_1(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda})), \\ & \quad f_2(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda}) - (g_1(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda}))) \\ &= \underbrace{\min(f_1(\boldsymbol{\lambda}) + g_1(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda}))}_{\text{mon.incr.}} \\ & \quad - \underbrace{(g_1(\boldsymbol{\lambda}) + g_2(\boldsymbol{\lambda}))}_{\text{mon.incr.}}. \end{aligned}$$

The minimum of the d.m. functions is itself a d.m. function (see e.g. Proposition 3).

Corollary 3: The max-min problem in (6)

$$\max_{0 \leq \boldsymbol{\lambda} \leq \mathbf{1}} \min(R_1(\boldsymbol{\lambda}), R_2(\boldsymbol{\lambda}))$$

is a d.m. programming problem.

Observe that the negative d.m. part of the max-min function in (28) is equal to the negative d.m. part of the sum rate function in (26). The difference is only in the positive d.m. function part.

C. Reformulation as monotonic optimization problems in standard form

We have seen above that the three problems of interest can be formulated as the following general d.m. problem

$$\max_{\boldsymbol{\lambda} \in [0,1]^2} \phi(\boldsymbol{\lambda}) - \psi(\boldsymbol{\lambda}) \quad (29)$$

with strictly increasing functions ϕ and ψ . The way forward that we propose here is to transform the problem in (29) to a domain where the parameter space has larger dimension but where the constraints are normal. After this transformation has been performed, the polyblock algorithm can be used to solve the optimization problems.

Specifically, we substitute $\psi(\boldsymbol{\lambda}) = \psi(\mathbf{1})(1 - t)$ in (29) where the range of t depends on $\boldsymbol{\lambda}$ and obtain the equivalent programming problem with $\mathbf{x} = [\lambda_1, \lambda_2, t]$

$$\max \underbrace{\phi(\mathbf{x}) + \psi(\mathbf{1})(x_3 - 1)}_{\Phi(\mathbf{x})} \quad \text{s.t. } \mathbf{x} \in \mathcal{D} \quad (30)$$

with constraint set

$$\mathcal{D} = \{\mathbf{x} \in \mathbb{R}_+^3 : x_1 \leq 1, x_2 \leq 1, x_3 \leq 1 - \frac{\psi(x_1, x_2)}{\psi(\mathbf{1})}\}. \quad (31)$$

Note that the function $\Phi(\mathbf{x})$ is strictly increasing and it holds that $\psi(\mathbf{1}) \geq 0$.

Lemma 2: The set \mathcal{D} defined in (31) is normal.

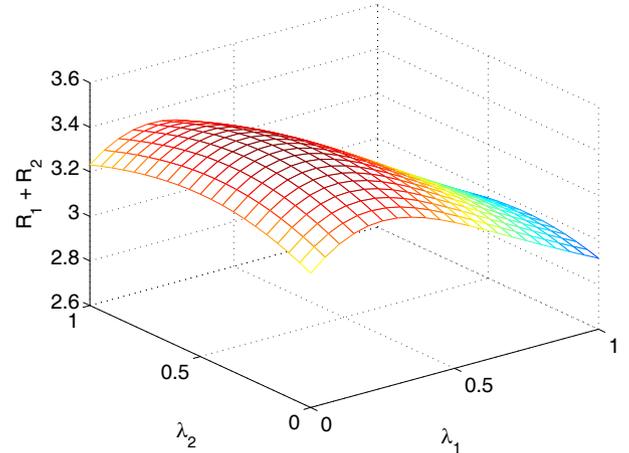


Fig. 4. Sum-rate $R_1 + R_2$ over $0 \leq \boldsymbol{\lambda} \leq \mathbf{1}$.

Proof: Choose the vector $\mathbf{x} \in \mathcal{D}$ and choose any vector $0 \leq \mathbf{y} \leq \mathbf{x}$. Next, we verify that $\mathbf{y} \in \mathcal{D}$ directly: $0 \leq y_1 \leq x_1 \leq 1$ and $0 \leq y_2 \leq x_2 \leq 1$. Since ψ is strictly increasing in x_1, x_2 , we have $\psi(x_1, x_2) \geq \psi(y_1, y_2)$ and thus from $y_3 \leq x_3$ follows that $\mathbf{y} \in \mathcal{D}$, too. ■

Furthermore, the constraint set is compact, bounded, and connected. The programming problem in (29) corresponds exactly to the problem (7). Therefore, we can apply the outer polyblock approximation algorithm shown in Algorithm 1 to solve all three problems, the weighted sum-rate maximization in (4), the proportional fair problem in (5), and the max-min problem in (6).

V. ILLUSTRATIONS

First, the solution by Algorithm 1 of the weighted sum-rate maximization problem (4) is illustrated in the next subsection. Then, the solution by Algorithm 1 of the proportional fair maximization problem in (5) is illustrated in Section V-B. Finally, the solution by Algorithm 1 of the max-min programming problem in (6) is illustrated in Section V-C.

The three examples are organized as follows. We choose a fixed but random channel scenario. First, we plot the objective functions $(R_1 + R_2), R_1 \cdot R_2$, and $\min(R_1, R_2)$ over λ_1, λ_2 . Next, we show the region \mathcal{D} and the approximation by the outer polyblock algorithm. Finally, we show the achievable rate region and the operating point for the max-min solution.

A. Weighted sum-rate maximization

For this example, the channel realization ($n_T = 3$) is given by

$$\begin{aligned} \mathbf{h}_{11} &= [0.0937 + 1.1175i; 1.1264 + 0.0556i; 0.7201 + 0.4820i], \\ \mathbf{h}_{12} &= [-0.7245 + 0.3036i; -0.8728 - 0.0395i; 0.2042 + 0.2601i] \\ \mathbf{h}_{21} &= [-0.3288 - 1.4935i; 0.2623 + 0.9598i; 0.5150 + 0.7231i], \\ \mathbf{h}_{22} &= [0.7339 - 0.2231i; -0.2756 - 1.0983i; -0.9767 - 0.5006i]. \end{aligned}$$

We operate at an SNR of 0 dB. In Figure 4, the objective function of the problem (4) is illustrated.

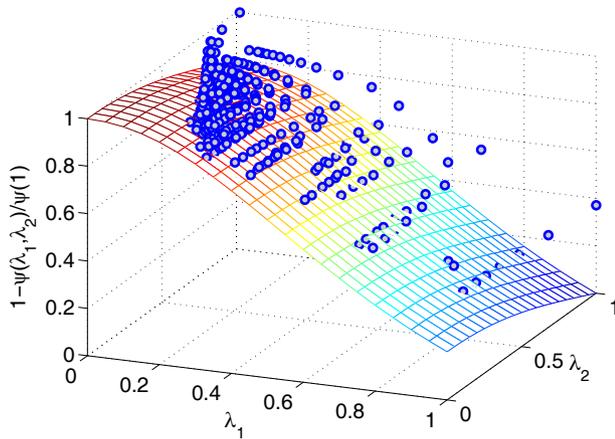


Fig. 5. Constraint set \mathcal{D} and vertices of outer polyblock approximation after 400 iterations.

In Figure 5, we plot the upper boundary of \mathcal{D} . The function on the z-axis $1 - \frac{\psi(\boldsymbol{\lambda})}{\psi(\mathbf{1})}$ is non-convex but well approximated by the outer polyblock algorithm.

The solution found by Algorithm 1 achieves the individual rates $R_1(\boldsymbol{\lambda}^*) = 1.891$ and $R_2(\boldsymbol{\lambda}) = 1.5713$ and thus a sum-rate of 3.4623. A 20×20 grid search (which corresponds to 400 function evaluations) gives the optimum as $(R_1 + R_2) = 3.4619 < (R_1(\boldsymbol{\lambda}) + R_2(\boldsymbol{\lambda}))$. This shows the advantage of the polyblock algorithm compared to a grid search for one sample channel realization.

Additionally, we computed the average performances of the outer polyblock algorithm and of the grid-search method for 1000 channel randomly chosen realizations, in order to show that the proposed algorithm performs well on the average, too. Here, the average sum-rate achieved with the outer polyblock algorithm (using at most 100 steps) is 4.374. A 10×10 grid search (100 function evaluations) gives an average sum-rate of 4.364.

B. Proportional fair maximization

For this example, the channel realization ($n_T = 2$) is given by

$$\begin{aligned} \mathbf{h}_{11} &= [0.5524 + 0.5810i; 0.4023 + 0.1878i], \\ \mathbf{h}_{12} &= [-0.8399 - 0.6974i; -1.5573 + 0.3667i] \\ \mathbf{h}_{21} &= [0.2315 - 0.0152i; 0.1655 + 0.7099i], \\ \mathbf{h}_{22} &= [-0.6697 + 0.8385i; -0.2648 + 0.7466i]. \end{aligned}$$

We operate at an SNR of 5 dB. In Figure 6, the objective function of the problem (5) is illustrated.

In Figure 7, we show the upper boundary of \mathcal{D} . The function on the z-axis $1 - \frac{\psi(\boldsymbol{\lambda})}{\psi(\mathbf{1})}$ is non-convex but well approximated by the outer polyblock algorithm.

The solution found by Algorithm 1 achieves the individual rates $R_1(\boldsymbol{\lambda}^*) = 1.0498$ and $R_2(\boldsymbol{\lambda}) = 2.1345$ and thus a product-rate of 2.2407. A 20×20 grid search (which corresponds to 400 function evaluations) gives the optimum as $(R_1 \cdot R_2) = 2.2402 < (R_1(\boldsymbol{\lambda}) \cdot R_2(\boldsymbol{\lambda}))$. This shows again the advantage of the polyblock algorithm compared to a grid search.

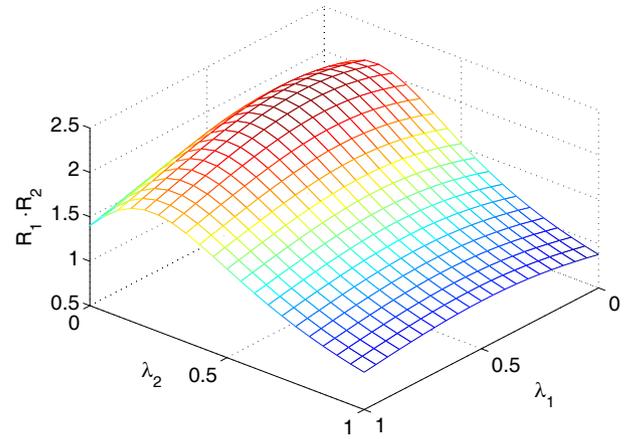


Fig. 6. Product of the rates ($R_1 \cdot R_2$) over $0 \leq \boldsymbol{\lambda} \leq 1$.

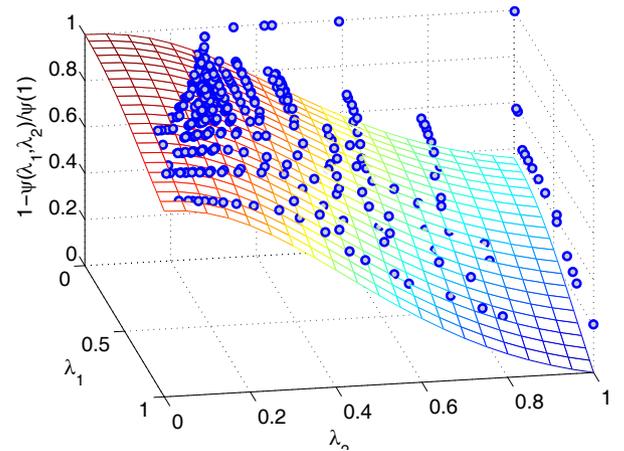


Fig. 7. Constraint set \mathcal{D} and vertices of the outer polyblock approximation after 400 iterations.

Additionally, just like in Section V-A, we compared the average performance of the outer polyblock algorithm to that of a grid search. The average was computed over 1000 channel realizations. Here, the average product-rate achieved with the outer polyblock algorithm (using maximally 100 steps) is 4.202. A 10×10 grid search (100 function evaluations) gives an average product rate of 4.234. Hence, the polyblock algorithm performs well here, too.

C. Max-min rate problem

In this scenario, the channel realization ($n_T = 2$) is given by

$$\begin{aligned} \mathbf{h}_{11} &= [-0.3059 - 0.0886i; -1.1777 - 0.2034i], \\ \mathbf{h}_{12} &= [-0.8107 - 0.8409i; 0.8421 + 0.0266i] \\ \mathbf{h}_{21} &= [0.2314 + 0.1320i; 0.1235 - 0.5132i], \\ \mathbf{h}_{22} &= [-0.4160 + 0.0964i; 1.5437 - 0.0806i]. \end{aligned}$$

We operate at an SNR of 0 dB. In Figure 8, the objective function of the problem in (6) is shown. The non-smooth

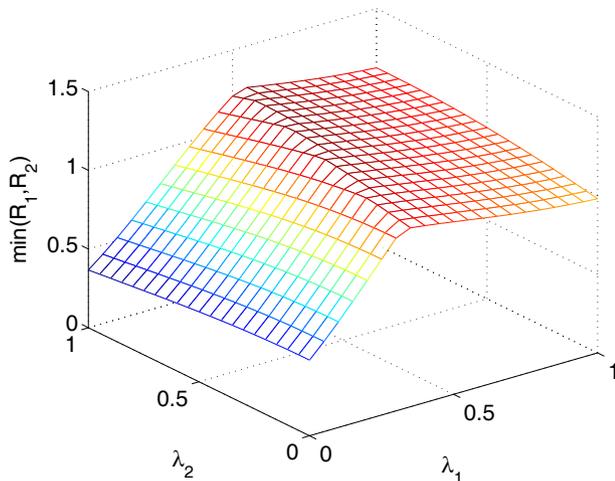


Fig. 8. Minimum of rates function $\min(R_1, R_2)$ over $0 \leq \lambda \leq 1$.

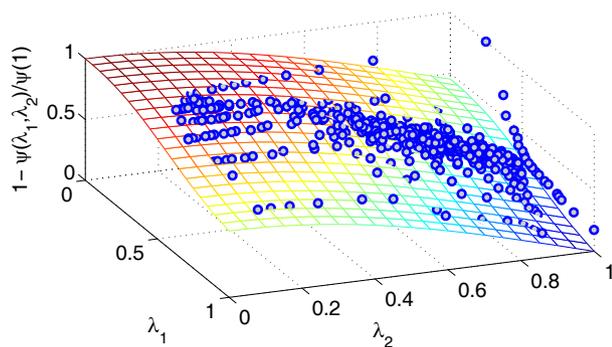


Fig. 9. Constraint set \mathcal{D} and vertices of outer polyblock approximation after 400 iterations.

function is clearly non-convex. Derivative-based approaches have difficulties handling this type of non-smooth functions.

However, the outer polyblock algorithm operates on the constraint set which is implicitly given in (28). The upper boundary of \mathcal{D} is shown in Figure 9. The function on the z-axis, $1 - \frac{\psi(\lambda)}{\psi(\mathbf{1})}$, is non-convex but smooth.

Also shown in Figure 9 are the vertices of the outer polyblock algorithm after 400 iterations. It can be observed that the constraint set is closely approximated. In particular, many vertices are tested close around the optimum, which is computed at $\lambda^* = [0.749831, 0.466272]$. Finally, in Figure 10, the achievable rate region and the solution of the polyblock algorithm are shown. Additionally, the angle bisector is shown for reference.

The solution found by Algorithm 1 achieves the individual rates $R_1(\lambda^*) = 1.2538$ and $R_2(\lambda^*) = 1.2551$. A 20×20 grid search (which uses 400 function evaluations) gives the optimum as $\min(R_1, R_2) = 1.2493 < \min(R_1(\lambda), R_2(\lambda))$. This shows again the advantage of the polyblock algorithm compared to a grid search.

We also compared the average performance (over 1000 random channel realization) of the polyblock algorithm to the performance of a grid search. The average minimax rate

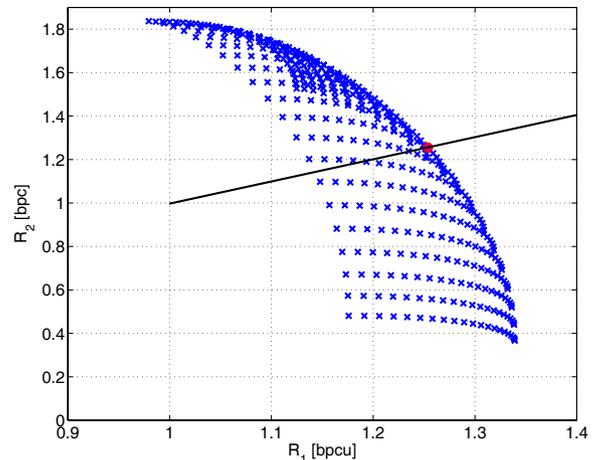


Fig. 10. Achievable rate region and operating point found by the polyblock algorithm.

achieved with the outer polyblock algorithm (terminating after at most 100 steps) is 1.826. A 10×10 grid search (100 function evaluations) gives an average minimax rate of 1.776.

D. Discussion

All three examples above show that the polyblock algorithm provides a solution which is better than what is produced by a simple grid search. In terms of complexity, the outer polyblock algorithm performs well. The main computational time is to find the intersection point between the line to the current best vertex and the upper boundary of the constraint set. The removal of dominated vertices is efficiently implemented according to [24, Proposition 4.2]. The interpretation of the outer polyblock algorithm in terms of the branch, cut, and bound framework shows that it is in fact part of a much larger framework for solving global optimization problems [25, Chapter 4, Theorem IV.1].

In [16], some hints for implementation are provided. In particular, one issue is related to the growth of the number of vertices in T_k . First, this might lead to storage problems and second, the complexity of the exhaustive search to find the best vertex in (9) also increases. A remedy to this problem is to restart the algorithm whenever $|T_k| > L$, where L is a fixed number.

The main advantage of the proposed approach is that it provides a structured and constructive way to solve the non-convex optimization problems associated with the computation of the sum-rate, proportional-fair and minimax operating points. The framework can be applied to other scenarios and systems as well. For example, it has been applied in [26] to the optimization of transmit strategies for the MISO broadcast channel, and later in [27] to the optimization problems for the MIMO broadcast channel.

VI. CONCLUSIONS

We have proposed a solution to the problem of optimal resource allocation and transmit beamforming for the two-user MISO interference channel. We developed a general

framework for determining the maximum sum-rate, maximum proportional fairness, and maximum minimum rate (a.k.a. the egalitarian solution) operating points using monotonic optimization and an outer polyblock approximation. To achieve a suitable representation, we exploited the monotonicity properties of the user rates as functions of the beamforming weights. Our approach is systematic compared to exhaustive search algorithms and numerical results suggest that the outer polyblock algorithm performs well compared to alternative approaches, in particular compared to a grid search.

REFERENCES

- [1] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, pp. 805-814, 1974.
- [2] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, pp. 60-70, 1978.
- [3] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, pp. 49-60, 1981.
- [4] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Trans. Inf. Theory*, vol. 27, pp. 786-788, 1981.
- [5] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689-699, Feb. 2009.
- [6] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620-643, Feb. 2009.
- [7] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: sum capacity in the low interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, p. 3032-3050, June 2009.
- [8] S. Vishwanath and S. A. Jafar, "On the capacity of vector Gaussian interference channels," *IEEE ITW*, 2004.
- [9] X. Shang, B. Chen, and M. J. Gans, "On the achievable sum rate for MIMO interference channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4313-4320, 2006.
- [10] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector Gaussian interference channels," submitted to *IEEE Trans. Inf. Theory*, July 2009.
- [11] V. S. Annapureddy and V. V. Veeravalli, "Sum capacity of MIMO interference channels in the low interference regime," submitted to *IEEE Trans. Inf. Theory*, Sep. 2009.
- [12] M. Charafeddine, A. Sezgin, and A. Paulraj, "Rate region frontiers for n-user interference channel with interference as noise," in *Proc. Allerton*, 2007.
- [13] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: a unified view," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1089-1103, 2008.
- [14] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292-5296, Oct. 2008.
- [15] E. Jorswieck and E. G. Larsson, "The MISO interference channel from a game-theoretic perspective: a combination of selfishness and altruism achieves pareto optimality," in *Proc. ICASSP*, 2008.
- [16] H. Tuy, "Monotonic optimization: problems and solution approaches," *SIAM J. Optimization*, vol. 11, pp. 464-494, 2000.
- [17] X. Shang, B. Chen, and H. V. Poor, "Multi-user MISO interference channels with single-user detection: optimality of beamforming and the achievable rate region," submitted to *IEEE Trans. Inf. Theory*, July 2009.
- [18] E. G. Larsson and E. A. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1059-1069, 2008.
- [19] E. G. Larsson, D. Danev, and E. A. Jorswieck, "Asymptotically optimal transmit strategies for the multiple antenna interference channel," in *Proc. Allerton*, 2008.
- [20] L. P. Qian, Y. J. Zhang, and J. Huang, "Mapel: achieving global optimality for a non-convex power control problem," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1553-1563, 2009.
- [21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [22] A. Rubinov, H. Tuy, and H. Mays, "An algorithm for monotonic global optimization problems," *Optimization*, vol. 49, pp. 205-221, 2001.
- [23] H. Tuy, *Convex Analysis and Global Optimization*. Kluwer Academic Publishers, 1998.
- [24] H. Tuy, F. Al-Khayyal, and P. T. Thach, *Essays and Surveys in Global Optimization*. Springer US, 2005, ch. Monotonic, pp. 39-78.
- [25] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*. Springer, 1996.
- [26] E. A. Jorswieck and E. G. Larsson, "Linear precoding in multiple antenna broadcast channels: efficient computation of the achievable rate region," in *Proc. IEEE Workshop Smart Antennas*, pp. 21-28, 2008.
- [27] J. Brehmer and W. Utschick, "Nonconcave utility maximization in the MIMO broadcast channel," *EURASIP J. Adv. Signal Process.*, 2008.



Eduard A. Jorswieck (S'01-M'05-SM'08) received his Diplom-Ingenieur degree and Doktor-Ingenieur (Ph.D.) degree, both in electrical engineering and computer science from the Berlin University of Technology (TUB), Germany, in 2000 and 2004, respectively. He was with the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute (HHI) Berlin, from 2001 to 2006. In 2006, he joined the Signal Processing Department at the Royal Institute of Technology (KTH) as a post-doc and became an Assistant Professor in 2007. Since

February 2008, he has been the head of the Chair of Communications Theory and Full Professor at Dresden University of Technology (TUD), Germany.

His research interests are within the areas of applied information theory, signal processing and wireless communications. He is senior member of IEEE and elected member of the IEEE SPCOM Technical Committee. From 2008-2010 he serves as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS. In 2006, he was co-recipient of the IEEE Signal Processing Society Best Paper Award.



Erik G. Larsson is Professor and Head of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. He joined LiU in September 2007. He has previously been Associate Professor (Docent) at the Royal Institute of Technology (KTH) in Stockholm, Sweden, and Assistant Professor at the University of Florida and the George Washington University, USA.

His main professional interests are within the areas of wireless communications and signal processing. He has published some 60 journal papers on these topics, he is co-author of the textbook *Space-Time Block Coding for Wireless Communications* (Cambridge Univ. Press, 2003) and he holds 10 patents on wireless technology.

He is Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and has been an editor for the the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is a member of the IEEE Signal Processing Society SAM and SPCOM technical committees.