Utility Maximization Strategies in the Multi-User MIMO Downlink

Johannes Brehmer and Wolfgang Utschick

International Workshop on Cross Layer Design (IWCLD)

September 2007

Technische Universität München
Associate Institute for Signal Processing
http://www.msv.ei.tum.de

Utility Maximization Strategies in the Multi-User MIMO Downlink

Johannes Brehmer and Wolfgang Utschick
Associate Institute for Signal Processing
Technische Universität München
{brehmer, utschick}@tum.de

***Abstract*— The problem of maximizing utility over the capacity region of the MIMO broadcast channel is addressed. While a direct solution is not possible, it is discussed how decomposition techniques can be used to find an optimum solution. Two decompositions are considered: A standard dual decomposition and a gradient projection-based decomposition.**

I. Introduction

The optimization of the physical layer of multiuser MIMO systems has recently achieved wide attention, usually based on physical layer performance metrics such as sum rate or weighted sum rate. Utility functions represent a generic model for capturing the properties of upper layers. Optimizing the parameters of the physical layer with respect to system utility is, inherently, a cross-layer optimization problem. There exists a large amount of literature on utility maximization for wireless networks, see, e.g., [1], [2]. These network-oriented works usually consider a large number of nodes with a simple physical layer setup, and focus on computationally efficient and distributed resource allocation strategies for large networks. In contrast, the multiuser MIMO downlink corresponds to a more involved physical layer setup, but also allows for a centralized solution at the transmitter.

From an information theoretic viewpoint, the multiuser MIMO downlink is a MIMO broadcast channel [3], [4]. Despite the fact that the capacity region of the MIMO BC is convex, a direct solution of the *downlink utility maximization* (DUM) is not possible, mainly due to the fact that with respect to the physical layer parameters, the problem is non-convex, and convexity is only achieved by a convex hull operation. This work discusses how decomposition techniques can be used to solve the DUM problem.

The first decomposition is the standard dual decomposition, which has been used, e.g., in the context of network utility maximization [2], in optimizing the transmission strategy in a MIMO BC under QoS targets [5], and in a recent work for utility maximization in the MIMO BC [6]. Application of the dual decomposition to the MIMO BC is summarized. Moreover, special emphasis is put on the computation and construction of maximizing rate vectors that lie in time-sharing regions.

The second decomposition under consideration, denoted as *iterative efficient set approximation* (IEA), has recently been proposed by the authors [7]. It is based on the idea to view the DUM problem as an optimization over a manifold, and to solve the DUM problem with a gradient projection algorithm, where the projections are on the manifold. In this work, IEA is applied to utility maximization in the MIMO BC. In particular, the issue of recovering time-sharing solutions is addressed, which was left as an open problem in [7].

II. Problem Setup

A MIMO broadcast channel with K receivers is considered. The transmitter has N transmit antennas, while receiver k is equipped with M_k receive antennas. The transmitter sends independent information to each of the receivers.

The received signal at receiver k is given by

$$\boldsymbol{y}_k = \boldsymbol{H}_k \sum_{i=1}^{K} \boldsymbol{x}_i + \boldsymbol{\eta}_k,$$

where $\boldsymbol{H}_k \in \mathbb{C}^{M_k \times N}$ is the channel to receiver k and $\boldsymbol{x}_k \in \mathbb{C}^N$ is the signal transmitted to receiver k. Furthermore, $\boldsymbol{\eta}_k$ is the circularly symmetric complex Gaussian noise at receiver k, with $\boldsymbol{\eta}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{1}_{M_k})$.

Let $\boldsymbol{Q}_k$ denote the transmit covariance matrix of user k. The total transmit power has to satisfy the power constraint $\operatorname{tr}\left(\sum_{k=1}^{K} \boldsymbol{Q}_k\right) \leq P_{\text{tr}}$. Accordingly, with $\boldsymbol{Q} = \left(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K\right)$ the set of feasible transmit covariance matrices is given by

$$\mathcal{Q} = \left\{ \boldsymbol{Q} : \boldsymbol{Q}_k \in \mathbb{H}^{N \times M_k}, \boldsymbol{Q}_k \geq \boldsymbol{0}, \operatorname{tr}\left(\sum_{k=1}^{K} \boldsymbol{Q}_k \right) \leq P_{\text{tr}} \right\}.$$

where $\mathbb{H}^{N \times M}$ denotes the set of Hermitian $N \times M$ matrices.

As proved in [3], capacity is achieved by *dirty paper coding* (DPC). Let π denote the encoding order, i.e., $\pi : \{1, \ldots, K\} \to \{1, \ldots, K\}$ is a permutation and $\pi(i)$ is the index of the user which is encoded at the i-th position. Let Π denote the set of all possible permutations on $\{1, \ldots, K\}$.

For fixed $\boldsymbol{Q}$ and π, an achievable rate vector is given by $\boldsymbol{r}(\boldsymbol{Q}, \pi) = (r_1(\boldsymbol{Q}, \pi), \ldots, r_K(\boldsymbol{Q}, \pi))$, with

$$r_{\pi(i)} = \log \frac{|\boldsymbol{1} + \boldsymbol{H}_{\pi(i)} (\sum_{j \geq i} \boldsymbol{Q}_{\pi(j)}) \boldsymbol{H}_{\pi(i)}^{\text{H}}|}{|\boldsymbol{1} + \boldsymbol{H}_{\pi(i)} (\sum_{j > i} \boldsymbol{Q}_{\pi(j)}) \boldsymbol{H}_{\pi(i)}^{\text{H}}|}.$$

Let $\mathcal{R}(\pi)$ denote the set of achievable rate vectors for a fixed order π:

$$\mathcal{R}(\pi) = \{ \boldsymbol{r}(\boldsymbol{Q}, \pi) : \boldsymbol{Q} \in \mathcal{Q} \}.$$

The set of rate vectors achievable by certain choices of $\boldsymbol{Q}$ and π is given by

$$\mathcal{R} = \bigcup_{\pi \in \Pi} \mathcal{R}(\pi).$$

Finally, the capacity region of the MIMO BC is defined as the convex hull of $\mathcal{R}$:

$$\mathcal{C} = \operatorname{co}(\mathcal{R}).$$

The capacity region contains the rate vectors that are achievable (with time-sharing). For transmission, one actually has to choose a rate vector from the capacity region. The decision which rate vector to choose is based on a maximization of some performance metric over the set of achievable rates. In the following, it is assumed that the properties of the upper layers are summarized in a system utility function $u : \mathbb{R}_{0,+}^K \to \mathbb{R}$. Moreover, it is assumed that the system utility function is differentiable, concave, and strictly monotonically increasing. Here, strict monotonicity implies that

$$\boldsymbol{r} > \boldsymbol{r}' \Rightarrow u(\boldsymbol{r}) > u(\boldsymbol{r}'), \tag{1}$$

where for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K$, with $K > 1$, order relations are component wise (Pareto order), i.e.,

$$\begin{aligned} \boldsymbol{x} > \boldsymbol{y} \ &\Rightarrow x_k \geq y_k, k = 1, \ldots, K, \\ &\exists k : x_k > y_k. \end{aligned}$$

The DUM problem for the MIMO BC is given by

$$\max_{\boldsymbol{r} \in \mathcal{C}} u(\boldsymbol{r}). \tag{2}$$

In the following, different strategies to solve (2) are discussed.

III. Time-sharing regions

Before discussing utility-optimum rate allocation, some important properties of the structure of the MIMO BC capacity region are summarized.

In general, $\mathcal{R} \subset \mathcal{C}$, i.e., the capacity region contains rate vectors which do not correspond to a specific choice of $\boldsymbol{Q}$ and π. The set of these rate vectors is achieved by time-sharing between multiple vectors in $\mathcal{R}$.

Let $\mathcal{E}$ denote the (Pareto efficient) boundary of the capacity region $\mathcal{C}$:

$$\mathcal{E} = \left\{ \boldsymbol{r} \in \mathcal{C} : \nexists \boldsymbol{r}' \in \mathcal{C} \quad \text{with} \quad \boldsymbol{r}' > \boldsymbol{r} \right\}. \tag{3}$$

The efficient set $\mathcal{E}$ contains the largest rate vectors (under the partial Pareto order). Due to the strict monotonicity of the system utility, a maximizer of the DUM problem will always be element of the efficient set. As a consequence, the properties of the efficient set play an important role in all proposed methods for solving the DUM problem.

Due to the convexity of $\mathcal{C}$, the set $\mathcal{E}$ can be written as

$$\mathcal{E} = \bigcup_{\boldsymbol{\lambda} > \boldsymbol{0}} \mathcal{E}(\boldsymbol{\lambda}) \quad \text{with} \quad \mathcal{E}(\boldsymbol{\lambda}) = \operatorname*{argmax}_{\boldsymbol{r} \in \mathcal{C}} \boldsymbol{\lambda}^{\text{T}} \boldsymbol{r}.$$

As noted before, some points in $\mathcal{C}$ are (in general) only achievable by time-sharing. Naturally, the same holds for the efficient boundary. For the MIMO BC, it is known that $\mathcal{E}(\boldsymbol{\lambda})$ is a time-sharing region if at least two of the entries in $\boldsymbol{\lambda}$ are equal. Accordingly, define a set Λ_{ts} as follows:

$$\Lambda_{\text{ts}} = \{ \boldsymbol{\lambda} \geq \boldsymbol{0} : (\exists (k, i) : i \neq k, \lambda_k = \lambda_i) \}.$$

If $\boldsymbol{\lambda} \in \Lambda_{\text{ts}}$, the set $\mathcal{E}(\boldsymbol{\lambda})$ is a convex combination of a finite number of points, i.e.,

$$\mathcal{E}(\boldsymbol{\lambda}) = \{ \boldsymbol{r} = \boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda}) \boldsymbol{\alpha}, \alpha_k \geq 0, \|\boldsymbol{\alpha}\|_1 = 1 \},$$

where the columns of the matrix $\boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda})$ are the corner points of the time-sharing region. Note that each corner point $\boldsymbol{r}_{\text{cp},k} = \boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda}) \boldsymbol{e}_k$ is an element of $\mathcal{R}$, i.e., there exist a covariance matrix $\boldsymbol{Q}$ and an order π such that

$$\boldsymbol{r}_{\text{cp},k} = \boldsymbol{r}(\boldsymbol{Q}, \pi).$$

Finally, it is known that $\boldsymbol{\lambda}$ is normal to $\mathcal{E}(\boldsymbol{\lambda})$, i.e.,

$$(\boldsymbol{r} - \boldsymbol{r}')^{\text{T}} \boldsymbol{\lambda} = 0, \ \boldsymbol{r}, \boldsymbol{r}' \in \mathcal{E}(\boldsymbol{\lambda}).$$

Consequently, if $\boldsymbol{\lambda} \in \Lambda_{\text{ts}}$, the affine hull of $\mathcal{E}(\boldsymbol{\lambda})$ is given by

$$\begin{aligned} \operatorname{aff}(\mathcal{E}(\boldsymbol{\lambda})) &= \{\boldsymbol{r} = \boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda}^*)\boldsymbol{\alpha}, \|\boldsymbol{\alpha}\|_1 = 1\} \\ &= \left\{\boldsymbol{r}_{\text{cp}} + \boldsymbol{v}, \boldsymbol{v} \in \operatorname{null} \boldsymbol{\lambda}^{\mathrm{T}}\right\}, \end{aligned} \tag{4}$$

where $\boldsymbol{r}_{\text{cp}}$ denotes an arbitrary corner point of $\mathcal{E}(\boldsymbol{\lambda})$.

According to (2) the objective is to find the rate vector $\boldsymbol{r}^* \in \mathcal{C}$ which maximizes utility. In fact, the main interest is not merely in $\boldsymbol{r}^*$ but also in the physical layer parameter setup that provides $\boldsymbol{r}^\star$. Two cases have to be distinguished:

1) $\boldsymbol{r}^* \in \mathcal{R}$: In this case, no time-sharing is needed. The solution is determined by the physical layer parameters $\boldsymbol{Q}^*$ and π^* such that $\boldsymbol{r}^* = \boldsymbol{r}(\boldsymbol{Q}^*, \pi^*)$.
2) $\boldsymbol{r}^* \notin \mathcal{R}$: In this case, the optimum rate vector lies in a time-sharing region. In order to fully determine a solution, it is required to identify a set of corner points $\left\{\boldsymbol{r}_{\text{cp},k}\right\}$ and a set of weights $\{\alpha_k\}$ such that

$$\boldsymbol{r}^* = \sum_{k=1}^{W} \alpha_k \boldsymbol{r}_{\text{cp},k}, \alpha_k \geq 0, \sum_{k=1}^{W} \alpha_k = 1,$$

where W denotes the number of corner points required to construct $\boldsymbol{r}^*$. In addition, for each corner point $\boldsymbol{r}_{\text{cp},k}$, the corresponding covariance matrix and encoding order are needed.

IV. Direct Solution

Due to the convex hull operation, it is not possible to solve (2) directly. A first approach may be to divide the problem into two subproblems: First, solve

$$\max_{\boldsymbol{Q},\pi} u(\boldsymbol{r}(\boldsymbol{Q}, \pi)) \quad \text{s.t.} \quad \boldsymbol{Q} \in \mathcal{Q}, \pi \in \Pi. \tag{5}$$

Second, enumerate all corner points of all time-sharing regions, and solve

$$\max_{\boldsymbol{\lambda} \in \Lambda_{\text{ts}}} \max_{\boldsymbol{\alpha}} u(\boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda})\boldsymbol{\alpha}) \quad \text{s.t.} \quad \alpha_k \geq 0, \|\boldsymbol{\alpha}\|_1 = 1. \tag{6}$$

If solutions for the two subproblems are available, the optimum solution can be found by maximizing over the maximum utility of the two solutions. In general, however, solving the first subproblem already represents a major difficulty. This is due to the fact that problem (5) is (in general) non-convex. Moreover, for $K > 2$ there are in general infinitely many time-sharing regions. As a result, a direct solution of the MIMO BC DUM problem is in general not feasible.

A special case is given by $u(\boldsymbol{r}) = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{r}$, i.e., if system utility is given by a weighted sum of rates. In this case, the optimum encoding order can be inferred from the weight $\boldsymbol{\lambda}$ [8]. Denote this optimum order by $\pi^*(\boldsymbol{\lambda})$. Then

$$\max_{\boldsymbol{Q} \in \mathcal{Q}} \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{r}(\boldsymbol{Q}, \pi^*(\boldsymbol{\lambda})) \tag{7}$$

can be solved as a convex optimization problem in the dual MAC [8]. In addition, if $\boldsymbol{\lambda} \in \Lambda_{\text{ts}}$, the maximum weighted sum rate is achieved by the corner points of $\mathcal{E}(\boldsymbol{\lambda})$. That is, time-sharing regions need not be considered for weighted sum rate maximization, and an optimum rate vector can be found by solving (7).

V. Dual Decomposition

The DUM problem is first modified by introducing additional variables:

$$\max_{\boldsymbol{s} \geq \boldsymbol{0}} u(\boldsymbol{s}) \quad \text{s.t.} \quad \boldsymbol{s} \leq \boldsymbol{r}, \boldsymbol{r} \in \mathcal{C}. \tag{8}$$

After introducing the Lagrangian

$$L(\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{\lambda}) = u(\boldsymbol{s}) + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{r} - \boldsymbol{s})$$

the dual function is given by

$$g(\boldsymbol{\lambda}) = g_{\mathrm{A}}(\boldsymbol{\lambda}) + g_{\mathrm{P}}(\boldsymbol{\lambda}),$$

with $\boldsymbol{\lambda} > \boldsymbol{0}$ (the cases $\lambda_k = 0$ can be excluded) and

$$g_{\mathrm{A}}(\boldsymbol{\lambda}) = \max_{\boldsymbol{s} \geq \boldsymbol{0}} u(\boldsymbol{s}) - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{s}, \quad \text{and} \tag{9}$$

$$g_{\mathrm{P}}(\boldsymbol{\lambda}) = \max_{\boldsymbol{r} \in \mathcal{C}} \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{r}. \tag{10}$$

For a fixed $\boldsymbol{\lambda}$, the optimization is decomposed into two subproblems (9) and (10). Subproblem (10) is a weighted sum rate maximization. As discussed in Section IV, weighted sum rate maximization can be solved as a convex optimization problem over the set of feasible covariance matrices.

The optimum dual variable is found by minimizing the dual function with respect to $\boldsymbol{\lambda}$. Note that every $\boldsymbol{r}' \in \mathcal{E}(\boldsymbol{\lambda})$ yields a subgradient of g at $\boldsymbol{\lambda}$. In other words, the existence of time-sharing regions implies that g is non-differentiable. The minimization of g can be carried out using any of the well-known methods for non-differentiable convex optimization, such as subgradient methods, cutting plane methods, or the ellipsoid method [9]. Note that any any of these methods requires the computation of a subgradient of g at iterates $\boldsymbol{\lambda}^{(k)}$. As any point in $\mathcal{E}(\boldsymbol{\lambda}^{(k)})$ yields a subgradient, it is sufficient to only consider the corner points of $\mathcal{E}(\boldsymbol{\lambda}^{(k)})$ – and these can be provided by (7). In other words, the dual problem can be solved without explicit knowledge of $\mathcal{C}$.

Clearly, determining an (approximately) optimum dual solution is only an intermediate step. In the end, a primal solution $(\boldsymbol{s}^*, \boldsymbol{r}^*)$ is required. Let $\boldsymbol{\lambda}^*$ denote the optimum dual variable. Clearly,

$$\boldsymbol{r}^* \in \mathcal{E}(\boldsymbol{\lambda}^*).$$

If $\mathcal{E}(\boldsymbol{\lambda}^*)$ contains a single element, the optimum rate as well as the optimum covariance matrix can be determined from (7). In contrast, if $\boldsymbol{\lambda}^* \in \Lambda_{\text{ts}}$, the optimum primal solution lies in a time-sharing segment, and $\mathcal{E}(\boldsymbol{\lambda}^*)$ is not a singleton set. In order to obtain a primal solution in the time-sharing case, standard methods for primal recovery could be employed. Alternatively, a straightforward way to recover a primal solution $(\boldsymbol{s}^*, \boldsymbol{r}^*)$ is provided by exploiting the linear structure of $\mathcal{E}(\boldsymbol{\lambda}^*)$ in the time-sharing case. Given the set of corner points of $\mathcal{E}(\boldsymbol{\lambda}^*)$, the optimum rate can be found by solving

$$\max_{\boldsymbol{\alpha}} u(\boldsymbol{R}_{\text{cp}}(\boldsymbol{\lambda})\boldsymbol{\alpha}) \quad \text{s.t.} \quad \alpha_k \geq 0, \|\boldsymbol{\alpha}\|_1 = 1. \tag{11}$$

Enumerating all corner points of $\mathcal{E}(\boldsymbol{\lambda}^*)$, however, is rather tedious, especially for larger K. In fact, $\boldsymbol{r}^*$ is a convex combination of at most K corner points. Therefore, we propose to first compute $\boldsymbol{r}^*$, and then enumerate the corner points of $\mathcal{E}(\boldsymbol{\lambda}^*)$ until a set of corner points is found whose convex combination contains $\boldsymbol{r}^*$.

The optimum rate $\boldsymbol{r}^*$ is found by optimizing utility over the affine hull of $\mathcal{E}(\boldsymbol{\lambda}^*)$. While knowledge of all corner points is required to specify $\mathcal{E}(\boldsymbol{\lambda}^*)$, the affine hull of $\mathcal{E}(\boldsymbol{\lambda}^*)$ requires knowledge of a single corner point only, see (4). We have that $\boldsymbol{r}^* \in \mathcal{E}(\boldsymbol{\lambda}^*)$. Due to convexity of the original problem, this also implies that $\boldsymbol{r}^*$ is a maximizer of

$$\max_{\boldsymbol{r}} u(\boldsymbol{r}) \quad \text{s.t.} \quad \boldsymbol{r} \in \operatorname{aff}(\mathcal{E}(\boldsymbol{\lambda}^*)). \tag{12}$$

By introducing a basis $\boldsymbol{B}$ of $\operatorname{null}((\boldsymbol{\lambda}^*)^{\mathrm{T}})$, problem (12) can be written as a simple unconstrained convex program:

$$\max_{\boldsymbol{\mu}} u(\boldsymbol{r}_{\text{cp}} + \boldsymbol{B}\boldsymbol{\mu}).$$

The optimum primal solution then follows as $\boldsymbol{r}^* = \boldsymbol{r}_{\text{cp}} + \boldsymbol{B}\boldsymbol{\mu}^*$ and $\boldsymbol{s}^* = \boldsymbol{r}^*$.

VI. Iterative Efficient Set Approximation

In the previous section, it was discussed how a dual decomposition can be used to solve the DUM problem in the MIMO BC. We repeatedly made use of the fact that the optimum rate vector is an element of the efficient set. This observation motivates an approach that directly operates on the efficient set in the search for an optimum solution. In [7], we proposed an alternative decomposition, denoted as *iterative efficient set approximation* (IEA), which is based on this idea. The IEA decomposition can also be interpreted as a gradient projection method, where projection is on the efficient set of rate vectors [10].

IEA solves the following problem:

$$\max_{\boldsymbol{r} \in \mathcal{E}} u(\boldsymbol{r}). \tag{13}$$

The efficient set is a $K-1$ dimensional submanifold with boundary of $\mathbb{R}^K$. For simplicity, in the following it is assumed that the optimum rate point $\boldsymbol{r}^*$ does not lie on the boundary of $\mathcal{E}$ and considerations are limited to the interior[1] $\tilde{\mathcal{E}}$ of $\mathcal{E}$.

In case of the MIMO BC, a closed-form global parameterization of $\tilde{\mathcal{E}}$ is not available. As a consequence, IEA works with local parameterizations of $\tilde{\mathcal{E}}$ and solves (13) in an iterative manner.

For $\boldsymbol{r} \in \tilde{\mathcal{E}}$, let $\mathcal{T}_{\boldsymbol{r}}$ denote the tangent space of $\tilde{\mathcal{E}}$ at $\boldsymbol{r}$, defined by the property that $\{\boldsymbol{r} + \boldsymbol{v}, \boldsymbol{v} \in \mathcal{T}_{\boldsymbol{r}}\}$ is the unique supporting hyperplane of $\mathcal{E}$ at $\boldsymbol{r}$. For the capacity region of the MIMO BC, the supporting hyperplane of $\mathcal{E}$ at $\boldsymbol{r}$ is unique for $\boldsymbol{r} \in \tilde{\mathcal{E}}$, i.e., $\tilde{\mathcal{E}}$ is a differentiable manifold. Moreover, from the properties of the weighted sum-rate maximization it follows that

$$\boldsymbol{r} \in \mathcal{E}(\boldsymbol{\lambda}) \Rightarrow \mathcal{T}_{\boldsymbol{r}} = \operatorname{null}(\boldsymbol{\lambda}^{\mathrm{T}}).$$

Let $\boldsymbol{B}(\boldsymbol{r}) \in \mathbb{R}^{K \times K-1}$ denote an orthonormal basis of the tangent space at $\boldsymbol{r}$. Then,

$$\hat{\mathcal{E}}_{\boldsymbol{r}} = \left\{\boldsymbol{r} + \boldsymbol{B}(\boldsymbol{r})\boldsymbol{\mu}, \boldsymbol{\mu} \in \mathbb{R}^{K-1}\right\},$$

represents a first-order approximation of $\mathcal{E}$ around $\boldsymbol{r}$.

The IEA algorithm consists of three main steps:

1) Using the tangent space at the current iterate $\boldsymbol{r}^{(k)}$, a first order approximation of $\mathcal{E}$ at $\boldsymbol{r}^{(k)}$ is communicated to the application layer.
2) The application layer uses this approximation to determine a gradient update in the tangent space. In fact, the application layer considers the problem

   $$\max_{\boldsymbol{\mu}} u\left(\boldsymbol{r}^{(k)} + \boldsymbol{B}(\boldsymbol{r}^{(k)})\boldsymbol{\mu}\right),$$

   and performs the first iteration of a gradient ascent method, starting at $\boldsymbol{\mu} = \boldsymbol{0}$, yielding the update

   $$\tilde{\boldsymbol{\mu}}^{(k)} = t\boldsymbol{B}(\boldsymbol{r}^{(k)})^{\mathrm{T}} \nabla u(\boldsymbol{r}^{(k)}), \tag{14}$$

   with a stepsize $t > 0$.
3) The application layer update results in a new point

   $$\tilde{\boldsymbol{r}}^{(k+1)} = \boldsymbol{r}^{(k)} + \boldsymbol{B}(\boldsymbol{r}^{(k)})\tilde{\boldsymbol{\mu}}^{(k)},$$

[1] $\tilde{\mathcal{E}} = \{\boldsymbol{r} \in \mathcal{E} : r_k > 0, \forall k\}$

which in general lies outside the efficient set. In the final step of each iteration, the point requested by the application layer is projected onto the efficient set, yielding the next iterate $\boldsymbol{r}^{(k+1)} = \mathrm{P}_{\mathcal{E}}(\tilde{\boldsymbol{r}}^{(k+1)})$.

These steps are repeated until the optimum rate $\boldsymbol{r}^*$ is found.

There exist different possibilities to project $\tilde{\boldsymbol{r}}^{(k+1)}$ on $\mathcal{E}$. Due to the nature of $\mathcal{E}$, a Euclidean projection on $\mathcal{E}$ seems prohibitive. Instead, a projection orthogonal to the tangent space is employed, as in [10]. Let $\boldsymbol{n}$ denote the unit-norm vector that is orthogonal to $\mathcal{T}_{\boldsymbol{r}}$ and points away from $\mathcal{C}$. To project $\tilde{\boldsymbol{r}}$ on $\mathcal{E}$, the following problem is solved:

$$\max_{x,\boldsymbol{r}} \; x \quad \text{s.t.} \quad \tilde{\boldsymbol{r}} + x\boldsymbol{n} \leq \boldsymbol{r}, \; \boldsymbol{r} \in \mathcal{C}. \tag{15}$$

The Lagrangian is given by

$$L(x, \boldsymbol{r}, \boldsymbol{\lambda}) = x + \boldsymbol{\lambda}^T(\boldsymbol{r} - \tilde{\boldsymbol{r}} - x\boldsymbol{n}).$$

The dual function follows as

$$\begin{aligned} g(\boldsymbol{\lambda}) &= \sup_{\substack{x\in\mathbb{R}\\ \boldsymbol{r}\in\mathcal{C}}} \left(x(1 - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{n}) + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{r} - \tilde{\boldsymbol{r}})\right) \\ &= \begin{cases} +\infty, \; \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{n} \neq 1, \\ \max_{\boldsymbol{r}\in\mathcal{C}} \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{r} - \tilde{\boldsymbol{r}}), \; \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{n} = 1. \end{cases} \end{aligned} \tag{16}$$

Note that for $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{n} = 1$, again a weighted sum-rate maximization problem is to be solved.

Let $\boldsymbol{r}^*(\boldsymbol{\lambda})$ denote an optimizer of the weighted sum-rate maximization in (16). The optimum dual variable $\boldsymbol{\lambda}$ is found by solving

$$\min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{r}^*(\boldsymbol{\lambda}) - \tilde{\boldsymbol{r}}) \quad \text{s.t.} \quad \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{n} = 1. \tag{17}$$

Again, the existence of time-sharing regions leads to a non-differentiable cost function, and the minimization with respect to $\boldsymbol{\lambda}$ can be carried out using any of the aforementioned methods for non-differentiable convex optimization.

Let $\boldsymbol{\lambda}^*$ denote the optimum dual variable. As for the dual decomposition, again two cases have to be distinguished:

1) If $\boldsymbol{\lambda}^* \in \Lambda_{\text{ts}}$, then $\mathcal{E}(\boldsymbol{\lambda}^*)$ contains a single element and the projection of $\tilde{\boldsymbol{r}}^{(k+1)}$ on $\mathcal{E}$ is given by $\boldsymbol{r}^{(k+1)} = \boldsymbol{r}^*(\boldsymbol{\lambda}^*)$.
2) In the time-sharing case, the primal solution can again be recovered by exploiting the linear structure of $\mathcal{E}(\boldsymbol{\lambda}^*)$. By replacing $\mathcal{C}$ by $\mathrm{aff}(\mathcal{E}(\boldsymbol{\lambda}^*))$ in (15), a set of linear equations results, where $\boldsymbol{r}_{\text{cp}}$ is again an arbitrary corner point of $\mathcal{E}(\boldsymbol{\lambda}^*)$:

$$\tilde{\boldsymbol{r}} + x\boldsymbol{n} = \boldsymbol{r}_{\text{cp}} + \boldsymbol{B}\boldsymbol{\mu}. \tag{18}$$

Then $\boldsymbol{r}^{(k+1)} = \boldsymbol{r}_{\text{cp}} + \boldsymbol{B}\boldsymbol{\mu}^*$, where $\boldsymbol{\mu}^*$ solves (18).

Even if the iterates $\boldsymbol{r}^{(k)}$ lie in a time-sharing region, it is not necessary to identify the corresponding corner points. Only for the last iterate, i.e., the optimum rate vector $\boldsymbol{r}^*$ the corner points and time-sharing coefficients are identified in the same manner as for the dual decomposition.

VII. Conclusions

Due to the inherent non-convexity with respect to the physical layer parameters, the utility maximization problem in the MIMO BC cannot be solved directly. A solution can be obtained by applying decomposition techniques. Application of a dual decomposition and of a gradient-projection based projection were discussed. While both decomposition techniques provide the optimum solution, they operate in a very different manner. While the dual decomposition solves the DUM problem via the dual, the IEA method performs a gradient ascent on the set of efficient rate vectors.

References

[1] S. Stanczak, M. Wiczanowski, and H. Boche, *Resource Allocation in Wireless Networks - Theory and Algorithms*, ser. Lecture Notes in Computer Science (LNCS 4000). Springer-Verlag, 2006.

[2] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: a mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, 2007.

[3] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, 2006.

[4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, 2003.

[5] T. Michel and G. Wunder, "Achieving QoS and Efficiency in the MIMO Downlink with Limited Power," in *ITG/IEEE Int. Workshop on Smart Antennas (WSA)*, Vienna, February 2007.

[6] J. Liu and Y. T. Hou, "Cross-layer optimization of MIMO-based mesh networks with Gaussian vector broadcast channels," April 2007, *Submitted to IEEE Transactions on Information Theory. Preprint available on arXiv.* [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:0704.0967

[7] J. Brehmer and W. Utschick, "A decomposition of the downlink utility maximization problem," in *Conference on Information Sciences and Systems (CISS)*, Baltimore, USA, March 2007.

[8] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 802–811, 2003.

[9] J.-L. Goffin and J.-P. Vial, "Convex nondifferentiable optimization: a survey focussed on the analytic center cutting plane method," Logilab, Geneva, Switzerland, Tech. Rep. 99.02, 1999.

[10] D. G. Luenberger, "The gradient projection method along geodesics," *Management Science*, vol. 18, pp. 620–631, July 1972.