Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions

Raphael Hunger, David A. Schmidt, and Wolfgang Utschick

International Symposium on Information Theory 2007

June 24 – 29, 2007

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

Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions

Raphael Hunger, David A. Schmidt, and Wolfgang Utschick
Associate Institute for Signal Processing, Munich University of Technology, 80290 Munich, Germany
E-Mail: {hunger,dschmidt,utschick}@tum.de

***Abstract*— In this paper, we present a novel algorithm for determining the sum-rate optimal transmit covariance matrices for the MIMO broadcast channel. Instead of optimizing the covariances directly, our algorithm operates on the precoding matrices, i. e., the square roots of the covariances. As a result, no eigenvalue decompositions are required in the iterations, and the complexity per iteration is significantly lower. A look at the convergence over the required number of computations shows a visible advantage over the state-of-the-art sum power iterative waterfilling algorithm. Also, our algorithm allows us to find the optimal sum-rate for an arbitrarily limited number of data streams per user. Finally, with a simple modification, our algorithm can also be used for sum-MSE minimization.**

I. Introduction

We consider the *multiple input multiple output* (MIMO) broadcast channel with perfect channel state information at the transmitter. The optimal sum-rate for a given transmit power constraint is achieved by *dirty paper coding* [1]. Furthermore, the resulting sum-capacity is the same as that of a dual MIMO *multiple access channel* (MAC) with a sum power constraint [2]. As the problem of finding the optimal transmit covariances for the dual MAC is concave as opposed to the broadcast problem, and since the MAC covariances can be converted to the broadcast scenario by means of a simple procedure [2], the sum-capacity problem for the broadcast channel is usually addressed by means of the dual MAC.

A closed form solution for this problem does not exist for more than one user. In [3], an algorithm based on following the principal eigenmode of the gradient was introduced. A more sophisticated iterative algorithm was proposed in [4], making use of the idea of iterative waterfilling [5]. This algorithm in its simplest form is not necessarily always convergent; one method to solve this problem is to update only the covariances of one random user pair at a time [6], since for two users the algorithm always converges. Alternatively, convergence can be ensured with an iteration-invariant memory factor, however, at the cost of slower convergence [4]. In [7], a simple search for a near-optimal memory factor in each iteration was shown to significantly speed up convergence. The method in [7] so far presents the computationally most efficient way of finding the sum-capacity. Nonetheless, at each iteration several *eigenvalue decompositions* (EVDs) must be computed.

Another iterative covariance-optimizing method has recently been proposed in [8]. Here, the Lagrange-dual optimization problem is investigated. The resulting algorithm allows for a slightly less complex waterfilling operation, but consists of nested iterative loops and therefore has high, non-deterministic complexity per inner loop.

In this paper, we approach the problem from a different perspective: instead of maximizing the sum-rate by choice of the covariance matrices, we examine the precoding (or beamforming) matrices, i. e., the square roots of the covariances. This way, the optimization problem does not have any semidefiniteness constraints. We can then formulate an iterative algorithm based on the projected gradient ascent method. This algorithm does not require any eigenvalue decompositions and the computational complexity per iteration is significantly lower than that of the state-of-the-art iterative waterfilling method, as is shown in our detailed complexity analysis in Section IV. Furthermore, it allows us to limit the number of data streams for each user to a number smaller than the number of that user's antennas. This way, the complexity per iteration can be dramatically reduced, while the resulting precoders are in general suboptimal with respect to the sum-rate.

As shown in Section V, an interesting property of our sum-capacity algorithm is that simply by changing the exponent of a certain expression from one to two, we arrive at an algorithm for finding the optimal precoding matrices in terms of *minimum mean square error* (MMSE) when only linear operations are allowed at transmitter and receiver, a problem to which a closed form solution does not exist either. This algorithm has already been published in [9], and is shown to clearly outperform existing procedures for the MIMO MMSE problem in terms of complexity. Finally, in Section VI, we plot the convergence behavior of our novel sum-capacity algorithm compared to that of the state-of-the-art technique taking into account the lower complexity per iteration. It turns out that the proposed algorithm without eigenvalue decompositions clearly performs better in the examined scenarios.

Notation: $\mathbf{0}_m$ denotes the $m \times m$ zero matrix, $\mathbf{I}_{m\times n}$, where $m \geq n$, is the left $m \times n$ part of the $m \times m$ identity matrix. $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix. The operator $\succcurlyeq$ is the partial order induced by the convex cone of positive semidefinite matrices.

II. Dual MAC System Model

Instead of treating the problem of the broadcast channel directly, we examine the dual MAC as in [2], which is obtained by switching the roles of transmitters and receivers and taking the conjugate transpose of the channel matrices. In the dual MAC, the same sum power constraint is imposed as in the broadcast channel. As shown in [2], the sum-capacity of the broadcast channel is the same as that of its dual MAC. Furthermore, the capacity-achieving transmit covariances for the broadcast channel can be obtained from the respective MAC covariances by a simple set of transformations [2].

In the dual MAC, K decentralized multi-antenna users transmit the data vectors to one centralized receiver with N antennas. The precoding matrix of user k is denoted by $\boldsymbol{T}_k \in \mathbb{C}^{r_k \times B_k}$, where r_k denotes the number of transmit antennas of user k and B_k the number of precoded data streams. Assuming data vectors with identity covariance matrices, the transmit covariances evaluate to $\boldsymbol{Q}_k = \boldsymbol{T}_k \boldsymbol{T}_k^{\mathrm{H}} \in \mathbb{C}^{r_k \times r_k}\ \forall k$. The sum power constraint reads as $\sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{Q}_k) \leq P_{\mathrm{Tx}}$. We describe the transmission from user k to the receiver with the channel matrix $\boldsymbol{H}_k \in \mathbb{C}^{N \times r_k}$. At the receiver, white Gaussian noise with the covariance matrix $\sigma_\eta^2 \mathbf{I}_N$ is added to the signal.

III. Review of Existing Algorithms

The MAC sum-rate resulting from the transmit covariance matrices $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K$ can be expressed as

$$R(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K) = \log_2 \det \left[\sigma_\eta^2 \mathbf{I}_N + \sum\nolimits_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^{\mathrm{H}}\right] - N \log_2 \sigma_\eta^2. \quad (1)$$

The maximization of $R(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K)$ subject to the positive semidefiniteness of the covariances $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K$ and the sum power constraint $\sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{Q}_k) \leq P_{\mathrm{Tx}}$ is a concave maximization problem [4] for which the Karush-Kuhn-Tucker optimality conditions are not only necessary, but also sufficient [10]. In general, the optimization

$$\max_{\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K} R(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K) \quad \text{s.t.: } \boldsymbol{Q}_k \succcurlyeq \mathbf{0}_{r_k}\ \forall k \text{ and } \sum\nolimits_{k=1}^{K} \operatorname{tr}(\boldsymbol{Q}_k) \leq P_{\mathrm{Tx}} \quad (2)$$

cannot be solved in a closed form for $K \geq 2$, therefore iterative techniques must be applied. Jindal et al. [4] extended the iterative waterfilling algorithm with *individual* power constraints from [5] to the case in which there is a sum power constraint. To this end, the authors introduce effective channels, resulting from interpreting the other users' contributions as colored noise, and iteratively perform waterfilling on the blockdiagonal effective channel Gram.

We examine this procedure from another point of view: after some manipulations of the KKT conditions for (2), we find that the optimal $\check{\boldsymbol{Q}}_1, \ldots, \check{\boldsymbol{Q}}_K$ of (2) must fulfill the system of nonlinear equations

$$\operatorname{blkdiag}\{\check{\boldsymbol{Q}}_k\}_{k=1}^{K} = \left[\operatorname{blkdiag}\left\{\frac{\mathbf{I}_{r_k}}{\check{\mu}} - (\boldsymbol{H}_k^{\mathrm{H}} \check{\boldsymbol{X}}_k^{-1} \boldsymbol{H}_k)^{-1}\right\}_{k=1}^{K}\right]_{\perp} \quad (3)$$

where $\boldsymbol{X}_k := \sigma_\eta^2 \mathbf{I}_N + \sum_{i \neq k} \boldsymbol{H}_i \boldsymbol{Q}_i \boldsymbol{H}_i^{\mathrm{H}} \in \mathbb{C}^{N \times N}$ denotes the interference plus noise covariance matrix of user k and μ is the Lagrangian factor ensuring that the sum power constraint is fulfilled with equality. Checked variables $\check{(\cdot)}$ indicate that they fulfill the KKT conditions. The operator $(\cdot)_\perp$ performs the orthogonal projection of its Hermitian argument onto the cone of positive semidefinite matrices by setting all negative eigenvalues to zero [11]. This operator can be thought of as the multi-dimensional extension of the $\max(\cdot, 0)$ operator known from scalar waterfilling. However, above compact notation already tells us which eigenspace has to be chosen for the optimum transmit covariances, and the optimum waterfilling power allocation directly follows as well.

The system of nonlinear equations describing the KKT conditions is only an implicit set of equations for the optimum covariances $\check{\boldsymbol{Q}}_1, \ldots, \check{\boldsymbol{Q}}_K$ and cannot be solved explicitly. It can be thought of, however, as a fixed point equation, which Jindal et al. attempt to solve iteratively by means of the *Picard* (or direct) iteration: the covariances $\boldsymbol{Q}_k^{(\ell+1)}, k = 1, \ldots, K$ in iteration $\ell + 1$ are obtained by evaluating the right-hand side of the matrix-valued function (3) for the covariances of iteration ℓ, resulting in the iterative waterfilling algorithm:

$$\operatorname{blkdiag}\{\tilde{\boldsymbol{Q}}_k^{(\ell)}\}_{k=1}^{K} \leftarrow \left[\operatorname{blkdiag}\left\{\frac{\mathbf{I}_{r_k}}{\mu^{(\ell)}} - (\boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{X}_k^{(\ell),-1} \boldsymbol{H}_k)^{-1}\right\}_{k=1}^{K}\right]_{\perp}$$

$$\boldsymbol{Q}_k^{(\ell+1)} \leftarrow \tilde{\boldsymbol{Q}}_k^{(\ell)}\ \forall k$$

Depending on the channel realization and the number of users, the matrix-valued function on the right-hand side of (3) need not be a contraction mapping, i. e., the sequence obtained by repeatedly applying the function must not converge to the unique fixed point regardless of the initial value. Convergence to the optimum solution when applying the Picard iteration can only be guaranteed for $K \leq 2$ users, see [4], and indeed, this kind of iteration starts to oscillate in some cases. To overcome this problem, Jindal et al. introduced a 'smoothing' factor $\alpha^{(\ell)} \in (0; 1]$ representing the influence of $\tilde{\boldsymbol{Q}}_k^{(\ell)}$ on the covariances in iteration $\ell + 1$:

$$\boldsymbol{Q}_k^{(\ell+1)} \leftarrow \alpha^{(\ell)} \tilde{\boldsymbol{Q}}_k^{(\ell)} + \left(1 - \alpha^{(\ell)}\right) \boldsymbol{Q}_k^{(\ell)}\ \forall k. \quad (4)$$

This type of iteration is called *Mann* iteration, and the authors in [4] conservatively choose $\alpha^{(\ell)} = \frac{1}{K}\ \forall \ell$, which is proven to lead to convergence to the global optimum. This conservative iteration-invariant choice, however, leads to a rather low rate of convergence.

A clever enhancement of the modified iterative waterfilling algorithm was proposed by Böhnke et al. in [7]. Here, the step size $\alpha^{(\ell)}$ is assumed to be time variant and is optimized in each iteration with a simple line search technique: the sum-rate is evaluated for the temporary step sizes $\tilde{\alpha}^{(\ell)} = 0$, $\tilde{\alpha}^{(\ell)} = 1$, and $\tilde{\alpha}^{(\ell)} = \frac{1}{2}$. Based on these three samples, the behavior of the sum-rate over $\alpha^{(\ell)} \in (0; 1]$ is approximated by a parabola. If the apex of the parabola lies within this half-open interval, the step size $\alpha^{(\ell)}$ is chosen accordingly, otherwise $\alpha^{(\ell)} = 1$. Algorithm 1 shows the implementation for unit variance noise $\sigma_\eta^2 = 1$ according to [7].

In the following, we analyze the number of *floating point operations* (FLOPs) required per iteration, where both a scalar complex multiplication and a scalar complex addition count as one FLOP and comparisons etc. are neglected [12]. We assume $r_k = r$, i. e., all users have the same number of antennas, and neglect linear and quadratic terms.

- Line 2: computation and scaling of $\boldsymbol{H}_k \boldsymbol{H}_k^{\mathrm{H}}$, summation of K Hermitian matrices: $KN^2r + KNr + \frac{KN^2}{2}$ FLOPs
- Line 4: KN^2 FLOPs
- Line 5: Cholesky factorization of $\boldsymbol{X}_k = \boldsymbol{M}_k \boldsymbol{M}_k^{\mathrm{H}}$, calculation of $\boldsymbol{M}_k^{-1} \boldsymbol{H}_k$ via forward substitution, computation of the Gram: $\frac{KN^3}{3} + \frac{KN^2}{2} + KN^2r + KNr^2 + KNr - \frac{Kr^2}{2}$ FLOPs

Algorithm 1 Covariance-Based Rate Maximization [7]

1: $\boldsymbol{Q}_k \leftarrow \frac{P_{\text{Tx}}}{\sum_{k=1}^K r_k}\mathbf{I}_{r_k}\ \forall k$
2: $\boldsymbol{X} \leftarrow \mathbf{I}_N + \sum_{k=1}^K \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^{\text{H}}$
3: **repeat**
4: $\quad \boldsymbol{X}_k \leftarrow \boldsymbol{X} - \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^{\text{H}}\ \forall k$
5: $\quad \boldsymbol{W}_k\boldsymbol{\Lambda}_k\boldsymbol{W}_k^{\text{H}} \leftarrow \boldsymbol{H}_k^{\text{H}}\boldsymbol{X}_k^{-1}\boldsymbol{H}_k\ \forall k$
6: $\quad \boldsymbol{P}_k \leftarrow \max(\mu\mathbf{I}_{r_k} - \boldsymbol{\Lambda}_k^{-1}, \mathbf{0})$ s.t.: $\sum_{k=1}^K \operatorname{tr}(\boldsymbol{P}_k) = P_{\text{Tx}}$
7: $\quad \tilde{\boldsymbol{Q}}_k \leftarrow \boldsymbol{W}_k\boldsymbol{P}_k\boldsymbol{W}_k^{\text{H}}\ \forall k$
8: $\quad \tilde{\boldsymbol{X}} \leftarrow \mathbf{I}_N + \sum_{k=1}^K \boldsymbol{H}_k\tilde{\boldsymbol{Q}}_k\boldsymbol{H}_k^{\text{H}}$
9: $\quad$ Find optimum scalar α from parabola
10: $\quad \boldsymbol{Q}_k \leftarrow \alpha\tilde{\boldsymbol{Q}}_k + (1-\alpha)\boldsymbol{Q}_k\ \forall k$
11: $\quad \boldsymbol{X} \leftarrow \alpha\tilde{\boldsymbol{X}} + (1-\alpha)\boldsymbol{X}$
12: **until** convergence

Complete eigenvalue decomposition: $9r^3K$ FLOPs
Note that for $N \gtrapprox 3r$, it is cheaper to apply the matrix inversion lemma and to compute only a single inverse $\boldsymbol{X}^{-1}$ instead of K inverses $\boldsymbol{X}_k^{-1}$

Line 7: $Kr^3 + \frac{3Kr^2}{2}$ FLOPs

Line 8: Cholesky factorization of $\boldsymbol{Q}_k = \boldsymbol{L}_k\boldsymbol{L}_k^{\text{H}}$, computation of $\boldsymbol{L}_k^{\text{H}}\boldsymbol{H}_k^{\text{H}}$, computation of the Gram, summation of the K Hermitian matrices: $\frac{Kr^3}{3} + \frac{Kr^2}{2} + KNr^2 + KN^2r + KNr$ FLOPs
Note that when $N < \frac{4}{3}r$ it is more efficient not to calculate $\tilde{\boldsymbol{Q}}_k$ explicitly, but to start with $\boldsymbol{H}_k\boldsymbol{W}_k$ and then include $\boldsymbol{P}_k$ and compute the Gram

Line 9: three determinants of Hermitian $N \times N$ matrices via Cholesky factorization: N^3 FLOPs

Line 10: update $\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^{\text{H}}$: $\frac{N^2K}{2}$ FLOPs

In sum, we have $KN^2r + KNr + \frac{KN^2}{2}$ FLOPs for the initialization (lines 1 and 2), and $\frac{KN^3}{3} + N^3 + 2KN^2r + 2KN^2 + 2KNr^2 + \frac{31Kr^3}{3} + \frac{3Kr^2}{2} + 2KNr$ FLOPs per iteration (lines 4 through 11).

IV. Rate Maximization without Eigenvalue Decompositions

Since the covariance-based rate maximization algorithms require a projection onto the cone of positive semidefinite matrices, eigenvalue decompositions are inevitable. A complete eigenvalue decomposition, however, is computationally rather expensive: it requires about $9r^3 + \mathcal{O}(r^2)$ FLOPs for an $r \times r$ Hermitian matrix, see [13]. In the following, we present a rate maximization approach that is based on the precoding matrices instead of the transmit covariances and therefore does not require an EVD. To this end, we utilize the scaled projected gradient algorithm [14], which can be seen as an extension of the steepest ascent method to constrained optimization problems. The precoding matrices do not have to be positive semidefinite, the only constraint therefore is the sum power, which now reads as $\sum_{k=1}^K \|\boldsymbol{T}_k\|_{\text{F}}^2 \leq P_{\text{Tx}}$. Thus, the optimization (2) changes to

$$\max_{\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K} R(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K) \quad \text{s.t.: } \sum\nolimits_{k=1}^K \|\boldsymbol{T}_k\|_{\text{F}}^2 \leq P_{\text{Tx}}, \tag{5}$$

where the utility $R(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K)$ follows from (1) in conjunction with $\boldsymbol{Q}_k = \boldsymbol{T}_k\boldsymbol{T}_k^{\text{H}}\ \forall k$. The scaled complex gradient of the utility function w. r. t. the precoder of user k reads as

Algorithm 2 Precoding-Filter-Based Rate Maximization

$\boldsymbol{T}_k \leftarrow \sqrt{\frac{P_{\text{Tx}}}{\sum_{k=1}^K B_k}}\mathbf{I}_{r_k \times B_k}\ \forall k$, choose s_0, $s' \leftarrow 1$
$\boldsymbol{X} \leftarrow \sigma_\eta^2\mathbf{I}_N + \sum_{k=1}^K \boldsymbol{H}_k\boldsymbol{T}_k\boldsymbol{T}_k^{\text{H}}\boldsymbol{H}_k^{\text{H}}$
$last\_metric \leftarrow \log_2\det\boldsymbol{X}$
repeat
$\quad \delta\boldsymbol{T}_k \leftarrow \boldsymbol{H}_k^{\text{H}}\boldsymbol{X}^{-1}\boldsymbol{H}_k\boldsymbol{T}_k\ \forall k$ — compute gradients
$\quad \lambda \leftarrow \sqrt{\frac{P_{\text{Tx}}}{\sum_{k=1}^K \|\delta\boldsymbol{T}_k\|_{\text{F}}^2}}$ — length normalization factor
$\quad$ **repeat**
$\quad\quad \boldsymbol{T}_k' \leftarrow \boldsymbol{T}_k + \frac{s_0}{s'}\lambda\delta\boldsymbol{T}_k\ \forall k$ — unscaled precoders
$\quad\quad \boldsymbol{T}_k' \leftarrow \boldsymbol{T}_k'\frac{\sqrt{P_{\text{Tx}}}}{\sqrt{\sum_{i=1}^K \|\boldsymbol{T}_i'\|_{\text{F}}^2}}\ \forall k$ — projection
$\quad\quad \boldsymbol{X}' \leftarrow \sigma_\eta^2\mathbf{I}_N + \sum_{k=1}^K \boldsymbol{H}_k\boldsymbol{T}_k'\boldsymbol{T}_k'^{\text{H}}\boldsymbol{H}_k^{\text{H}}$
$\quad\quad new\_metric \leftarrow \log_2\det\boldsymbol{X}'$
$\quad\quad$ **if** $new\_metric \leq last\_metric$ **then**
$\quad\quad\quad s' \leftarrow s' + 1$ — reduce step-size
$\quad\quad$ **end if**
$\quad$ **until** $new\_metric > last\_metric$
$\quad \boldsymbol{T}_k \leftarrow \boldsymbol{T}_k'\ \forall k$, $\boldsymbol{X} \leftarrow \boldsymbol{X}'$, $last\_metric \leftarrow new\_metric$
until convergence

$$\ln 2 \cdot \frac{\partial R(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K)}{\partial \boldsymbol{T}_k^*} = \boldsymbol{H}_k^{\text{H}}\boldsymbol{X}^{-1}\boldsymbol{H}_k\boldsymbol{T}_k, \tag{6}$$

where $\boldsymbol{X} := \sigma_\eta^2\mathbf{I}_N + \sum_{i=1}^K \boldsymbol{H}_i\boldsymbol{T}_i\boldsymbol{T}_i^{\text{H}}\boldsymbol{H}_i^{\text{H}}$. Since we use the definition of complex derivatives from [15], we need to derive with respect to the complex conjugate, in order to get the direction of the steepest ascent. For high signal-to-noise ratios (SNRs), the squared Frobenius norm of (6) scales with the reciprocal SNR and is therefore very small, leading to unnecessarily small steps. In order to overcome this problem, we introduce a preconditioning scalar λ which rescales all gradients such that the 'length' of the total gradient is the same as that of the covariances themselves:

$$\lambda = \sqrt{\frac{P_{\text{Tx}}}{\sum_{k=1}^K \|\boldsymbol{H}_k^{\text{H}}\boldsymbol{X}^{(\ell),-1}\boldsymbol{H}_k\boldsymbol{T}_k^{(\ell)}\|_{\text{F}}^2}}.$$

The update rule for the precoders in iteration $\ell+1$ reads as

$$\boldsymbol{T}_k^{(\ell+1)} = \frac{\left(\mathbf{I}_{r_k} + s^{(\ell)}\lambda^{(\ell)}\boldsymbol{H}_k^{\text{H}}\boldsymbol{X}^{(\ell),-1}\boldsymbol{H}_k\right)\boldsymbol{T}_k^{(\ell)}}{\kappa^{(\ell)}}. \tag{7}$$

The step size $s^{(\ell)}$ is chosen adaptively and reduced as soon as the utility decreases. Setting $s^{(\ell)} := \frac{s_0}{s'^{(\ell)}}$ with fixed s_0 and incrementing $s'^{(\ell)}$ by one as soon as the utility decreases satisfies the sufficient properties $\sum_{\ell=0}^\infty s^{(\ell)} = \infty$ and $\sum_{\ell=0}^\infty s^{(\ell),2} < \infty$ to let the algorithm converge [14]. Clearly, the new precoders in step $\ell+1$ also need to satisfy the transmit power constraint. This orthogonal projection onto the ball with radius $\sqrt{P_{\text{Tx}}}$ is simply a division by

$$\kappa^{(\ell)} = \sqrt{\frac{\sum_{k=1}^K \|\boldsymbol{T}_k^{(\ell)} + s^{(\ell)}\lambda^{(\ell)}\boldsymbol{H}_k^{\text{H}}\boldsymbol{X}^{(\ell),-1}\boldsymbol{H}_k\boldsymbol{T}_k^{(\ell)}\|_{\text{F}}^2}{P_{\text{Tx}}}}.$$

Adaptively decreasing the step size ensures that the sequence in fact reaches a stationary point fulfilling the KKT conditions of (5) and does not run into an oscillation for $\ell \to \infty$. Recall that the covariance-based optimization (2) is a concave maximization problem and the stationary point fulfilling the KKTs of (2) is the global optimum. However, the utility function is not concave in the precoders. Furthermore, it can be shown that, when the precoders are used instead of the covariances, there are several additional rank-deficient solutions to the KKTs, which turn out to represent saddle-points and minima. Since the problem is concave on the set of covariances, it is clear that a steepest ascent algorithm on $\boldsymbol{Q}_k$ will lead to the global maximum. As each set of precoding matrices $\boldsymbol{T}_k$ corresponds to a set of covariance matrices, it is intuitively plausible that following the steepest ascent in the precoders will in general lead to the same result. From the second line in (7), however, we can see that the rank of $\boldsymbol{T}_k$ will never change in the course of a finite number of iterations, i. e., a rank-deficient initialization will never be able to reach the maximum, if the maximum has full rank. If the maximum is rank-deficient, on the other hand, and the initialization has full rank, the algorithm will reduce the rank asymptotically by increasing the condition number from step to step.

We can interpret the saddle-points corresponding to rank-deficient precoders as optimal solutions for the case in which users are not allowed to perform full multiplexing, but must transmit less data streams as they have transmit antennas. Indeed, this is another advantage of our approach of working on the precoders and not on the covariances: assume that a multi-antenna user is restricted to transmitting at most a single data stream. Then, its transmit covariance matrix is restricted to have rank one. Such a constraint is nonconvex and the existing algorithms are not able to handle this case. With our algorithm, on the other hand, we are free to restrict the number of data streams transmitted per user arbitrarily from 1 to r_k, simply by initializing the algorithm with tall precoding matrices. If at most a single stream per user is to be transmitted, we are likely to achieve a sum-rate below the maximum, the computational complexity of each iteration, on the other hand, is drastically reduced.

In Algorithm 2, the pseudo-code for our procedure is given for the case in which the number of data streams per user can be arbitrarily restricted. In line 1, we initialize all precoders with full rank and uniform power allocation, and the initial step size is chosen. In line 3, we compute the rate up to a constant summand, see (1). All gradients are computed in line 5, the normalization factor λ is computed in line 6. Lines 8 through 14 are repeatedly executed with decreasing step size, until the utility increases. First, the new unscaled precoders are determined in line 8 with the gradient ascent method, then line 9 performs the projection onto the power constraint set. If the sum-rate obtained by all $\boldsymbol{T}_k'$ is smaller than the one achieved by all $\boldsymbol{T}_k$ (which, as simulations show, occurs very seldomly), we reduce the step size in line 13 and try again. Otherwise, the precoders are updated.

Again, we analyze the computational complexity of the initialization and iterations of Alg. 2 assuming $B_k = r_k = r$.

- Line 2: computation and scaling of $\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}}$, summation of K Hermitian matrices: $KN^2r+KNr+\frac{KN^2}{2}$ FLOPs
- Line 3: determinant via Cholesky factorization of $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$: $\frac{N^3}{3}$ FLOPs
- Line 5: $\boldsymbol{L}^{-1}\boldsymbol{H}_k$ via forward substitution, calculation of Gram, multiplication from the right with $\boldsymbol{T}_k$: $KN^2r+KNr^2+KNr+2Kr^3-\frac{5Kr^2}{2}$ FLOPs
- Line 6: K Frobenius norms: $2Kr^2$ FLOPs
- Line 8: scaling and addition: $2Kr^2$ FLOPs
- Line 9: K Frobenius norms and scaling: $3Kr^2$ FLOPs
- Line 10: calculation of $\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}$, computation of the Gram, summation of K Hermitian matrices: $2KNr^2 + KN^2r$ FLOPs
- Line 11: determinant via Cholesky factorization of $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$, same as line 3

Altogether, $\frac{N^3}{3}+KN^2r+KNr+\frac{KN^2}{2}$ FLOPs are required for the initialization (lines 1 through 3), and $\frac{N^3}{3}+2KN^2r+3KNr^2+KNr+2Kr^3+\frac{9Kr^2}{2}$ FLOPs per iteration (lines 5 through 16).

V. Relation to Sum-MSE Minimization

In [9], a very similar algorithm was presented for the problem of sum-MSE minimization with linear precoding and equalization in a multi-user MIMO environment. In the following, we will briefly derive this MMSE algorithm and show how it differs from our sum-rate algorithm.

Again, we examine a dual MAC with a sum power constraint, according to the duality framework described in [9], [16]. The K transmitters employ the precoding matrices $\boldsymbol{T}_k$ and the receiver uses an equalization matrix to obtain estimates of the transmitted data vectors from the noisy received signal. Assuming that the receiver chooses the optimal receive filter, the sum-MSE $\varepsilon(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K)$ reads as [9]

$$\begin{aligned}\varepsilon(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_K) =&\sigma_\eta^2 \operatorname{tr}\Big[\underbrace{\Big(\sigma_\eta^2\mathbf{I}_N+\sum_{k=1}^K\boldsymbol{H}_k\boldsymbol{T}_k\boldsymbol{T}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}\Big)}_{=:\boldsymbol{X}}{}^{-1}\Big]\\ &+\sum_{k=1}^K B_k - N.\end{aligned} \tag{8}$$

Note that the matrix $\boldsymbol{X}$ is the same as in the previous sections and that the precoder-dependent part that is to be optimized reads as $\operatorname{tr}(\boldsymbol{X}^{-1})$, as opposed to $\log_2\det(\boldsymbol{X})$ for the sum-rate problem. For the projected gradient algorithm to be applied to the sum-MSE minimization problem, we compute the gradient of $\operatorname{tr}(\boldsymbol{X}^{-1})$, where we neglect the factor σ_η^2, as the gradient will be normalized anyway. The scaled gradient reads as

$$\frac{1}{\sigma_\eta^2}\frac{\partial\varepsilon(\boldsymbol{T}_1,\ldots,\boldsymbol{T}_k)}{\partial\boldsymbol{T}_k^*} = -\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{X}^{-2}\boldsymbol{H}_k\boldsymbol{T}_k, \tag{9}$$

which has the exact same structure as the gradient (6) that arises in the sum-rate maximization problem, with the exception that the matrix $\boldsymbol{X}$ is taken to the power of -2 instead of -1 and that the expression is preceded by a minus sign. We wish to perform a gradient *descent*, however, instead of an ascent, and therefore we subtract the scaled gradient. The update rules for the sum-MSE problem and the rate-maximization problem can now be unified to

$$\boldsymbol{T}_k^{(\ell+1)} = \frac{\boldsymbol{T}_k^{(\ell)} + s^{(\ell)}\lambda^{(\ell)}\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{X}^{(\ell),-\alpha}\boldsymbol{H}_k\boldsymbol{T}_k^{(\ell)}}{\kappa^{(\ell)}}, \tag{10}$$

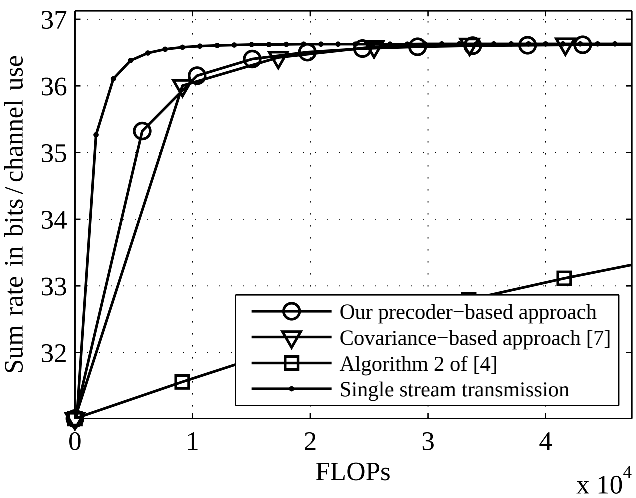

Fig. 1. Rate over FLOPs for $K = 15$ users, $N = 4$ receive antennas, and $r = 3$ transmit antennas per user. $P_{\text{Tx}} = 100$ and $\sigma_\eta^2 = 1$.

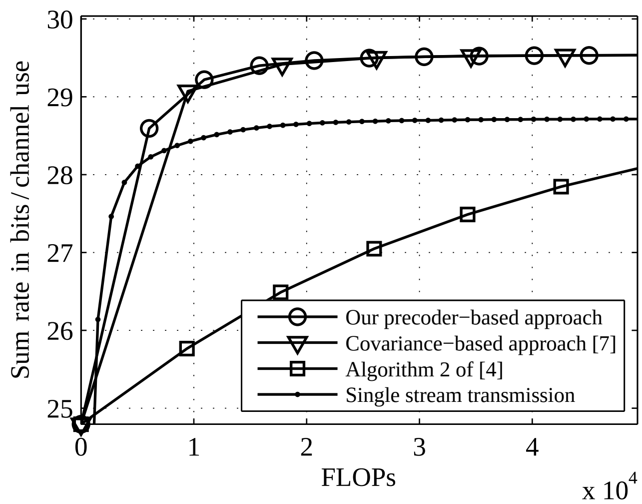

Fig. 2. Rate over FLOPs for $K = 6$ users, $N = 6$ receive antennas, and $r = 4$ transmit antennas per user. $P_{\text{Tx}} = 10$ and $\sigma_\eta^2 = 1$.

where $\alpha = 1$ maximizes the sum-rate and $\alpha = 2$ minimizes the MSE. Furthermore, we have to replace $\log_2 \det(\cdot)$ by $\mathrm{tr}[(\cdot)^{-1}]$ in lines 3 and 11 of Algorithm 2 and replace line 12 by $(-1)^\alpha new\_metric \geq (-1)^\alpha last\_metric$ and line 15 by $(-1)^\alpha new\_metric < (-1)^\alpha last\_metric$.

VI. Simulation Results

For the numerical simulations, we chose a scenario similar to the one in [7], however with all users having the same number of transmit antennas, in order to able to apply the FLOP counts from the end of Section III and Section IV. In Fig. 1, we look at a single representative channel realization with $K = 15$ users, $N = 4$ base-station antennas, and $r = 3$ antennas per user. The initial step size s_0 is 10, the transmit power P_{Tx} is set to 100, and the noise has unit variance. Our novel precoder-based approach with no restrictions on the number of data streams ($B_k = r_k \ \forall k$) is marked with circles, whereas the covariance based approach from [7] and Algorithm 2 from [4] are marked with triangles and squares, respectively. Since these algorithms uniformly allocate the power to all streams initially, both curves start at the same rate for zero FLOPs, i. e., without iteration. Although our novel approach has a slower convergence *per iteration* (each marker corresponds to a single iteration), the convergence over the number of FLOPs, which is a far more relevant metric, is slightly better than with the covariance based approach. Restricting the number of data streams to $B_k = 1 \ \forall k$ (dot marker curve) leads to a drastically reduced complexity, but reaches the maximum sum-rate for this particular configuration.

Another scenario is shown in Fig. 2, where $K = 6$ users transmit their data over $r_k = 4 \ \forall k$ transmit antennas each to a base station equipped with $N = 6$ antennas. Again, our precoder-based approach allowing the transmission of $B_k = r_k$ streams visibly outperformes the state-of-the-art algorithm from [7]. If we now restrict the number of streams to $B_k = 1 \ \forall k$, the sum-capacity of the channel cannot be reached.

Note that although we only presented two exemplary channel realizations, the qualitative behavior is somewhat similar over a wide range of system parameters.

References

[1] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of MIMO Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.

[3] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink Capacity Evaluation of Cellular Networks With Known-Interference Cancellation," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 802–811, June 2003.

[4] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. J. Goldsmith, "Sum Power Ierative Water-Filling for Multi-Antenna Gaussian Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[5] W. Yu and J. M. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels.," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.

[6] M. Codreanu, M. Juntti, and M. Latva-Aho, "Low-Complexity Iterative Algorithm for Finding the MIMO-OFDM Broadcast Channel Sum Capacity," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 48–53, January 2007.

[7] R. Böhnke, V. Kühn, and K. D. Kammeyer, "Fast Sum Rate Maximization for the Downlink of MIMO-OFDM Systems," in *Canadian Workshop on Information Theory (CWIT 2005)*, Montreal, Canada, June 2005.

[8] W. Yu, "Sum-Capacity Computation for the Gaussian Vector Broadcast Channel Via Dual Decomposition," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 754–759, February 2006.

[9] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver Design for Multi-User MIMO Systems," in *Proc. ITG/IEEE WSA 2006*, March 2006.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[11] N. J. Higham, "Matrix Nearness Problems and Applications," in *Applications of Matrix Theory*, M. J. C. Gover and S. Barnett, Eds. 1989, pp. 1–27, Oxford University Press.

[12] R. Hunger, "Floating Point Operations in Matrix-Vector Calculus," Tech. Rep. TUM-LNS-TR-05-05, Munich University of Technology, October 2005.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1991.

[14] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computations*, Prentice-Hall, 1989.

[15] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, second edition, 1991.

[16] W. Utschick and M. Joham, "On the Duality of MIMO Transmission Techniques for Multiuser Communications," in *Proc. EUSIPCO 2006*, September 2006, *Invited paper*.