Weighted Sum Rate Maximization for MIMO-OFDM Systems with Linear and Dirty Paper Precoding

Ronald Böhnke and Karl-Dirk Kammeyer

Department of Communications Engineering, University of Bremen, Germany
{boehnke, kammeyer}@ant.uni-bremen.de

Abstract

Many sophisticated resource allocation strategies are based on the maximization of the weighted sum of data rates for a given transmit power. While this problem can be easily solved for orthogonal multiple access schemes like TDMA, it is much more complicated if users are separated in space using multiple antennas at the base station due to the mutual coupling. In this paper, we propose a new projected conjugate gradient algorithm for the optimization of the transmit filters. The power constraint is taken into account in the calculation of the search direction by projecting the gradient onto a tangent hyperplane. Our method features excellent convergence properties when applied to dirty paper precoding, and it may also be used for the optimization of linear precoders.

1 Introduction

The multicarrier technique OFDM is well established for broadband wireless communication. In upcoming standards, it will be combined with multiple input multiple output (MIMO) schemes using more than one transmit and receive antenna simultaneously. In the downlink of a multi-user scenario, resources must be allocated subject to a total power constraint. A suitable optimization criterion is the sum of weighted data rates. Many advanced scheduling algorithms can be cast into this framework, e.g. taking the inverse average throughputs as weighting factors leads to the proportional fair policy [1], while choosing the queue lengths as measure for the priorities reduces the risk of buffer overflows [2]. A more sophisticated approach aims at the minimization of the average packet delay [3], and it is also possible to incorporate specific quality of service demands of different applications.

It was shown in [4] that all possible rate allocations for a given transmit power can be achieved by performing nonlinear dirty paper precoding at the base station. Interestingly, the same rate region results if the roles of transmitter and receivers are exchanged [5]. This duality also holds for purely linear processing [6] and its combination with nonlinear interference cancelation. Efficient optimization algorithms usually solve a certain problem for the uplink and transform the solution to the downlink afterwards. After introducing the system model in Section 2, we will discuss the properties of the weighted sum rate for linear and nonlinear precoding at the base station in Section 3. Existing algorithms for the solution of this optimization problem are reviewed in Section 4; some of these are only suited for special cases, e.g. equal priorities or mobile terminals with a single antenna. In Section 5, a novel projected conjugate gradient algorithm is presented, that converges extremely fast if dirty paper precoding is employed and may also be used for design of linear precoders. In contrast to [7], we optimize the transmit filters instead of the covariance matrices, as they do not have to be positive semidefinite. Additionally, the power constraint is taken into account in the calculation of the conjugate gradient search direction. Numerical results for a typical indoor scenario are presented in Section 6.

2 System Model

We consider a MIMO-OFDM system with N_U users and N_S subcarriers. The guard interval is assumed to be of sufficient length, i.e. longer than the channel impulse responses, but the corresponding rate and power loss is neglected. The base station is equipped with N_B antennas and each mobile terminal has N_M antennas. It is straightforward to generalize the results for mobiles with different number of antennas. Although we are mainly interested in the downlink, the duality of [5] allows us to focus on an equivalent uplink channel with the same power constraint instead, which is in general easier to handle. Then, the receive signal at subcarrier s is given by

$$\boldsymbol{y}[s] = \sum_{u=1}^{N_\mathrm{U}} \boldsymbol{H}_u[s]\,\boldsymbol{x}_u[s] + \boldsymbol{n}[s] \tag{1}$$

where $\boldsymbol{H}_u[s] \in \mathbb{C}^{N_\mathrm{B}\times N_\mathrm{M}}$ and $\boldsymbol{x}_u[s] \in \mathbb{C}^{N_\mathrm{M}}$ denote the dual channel matrix and transmit vector of user u, respectively, and $\boldsymbol{n}[s] \in \mathbb{C}^{N_\mathrm{B}}$ represents normalized additive white Gaussian noise with covariance matrix $\mathrm{E}\{\boldsymbol{n}[s]\,\boldsymbol{n}[s]^\mathrm{H}\} = \boldsymbol{I}$. Furthermore, the transmit signal $\boldsymbol{x}_u[s]$ follows from multiplying N_L uncorrelated unit variance Gaussian symbols with the filter matrix

$\boldsymbol{T}_u[s] \in \mathbb{C}^{N_\mathrm{M}\times N_\mathrm{L}}$, so the corresponding transmit covariance matrix becomes

$$\boldsymbol{Q}_u[s] = \boldsymbol{T}_u[s]\,\boldsymbol{T}_u^\mathrm{H}[s]\ , \tag{2}$$

and the power constraint can be expressed as

$$\sum_{u=1}^{N_\mathrm{U}}\sum_{s=1}^{N_\mathrm{S}} \mathrm{tr}(\boldsymbol{Q}_u[s]) = \sum_{u=1}^{N_\mathrm{U}}\sum_{s=1}^{N_\mathrm{S}} \big\|\boldsymbol{T}_u[s]\big\|_\mathrm{F}^2 \le N_\mathrm{S} P \tag{3}$$

with the average transmit power per subcarrier P. For notational brevity, we collect all transmit covariances and filters in the matrices $\boldsymbol{Q}$ and $\boldsymbol{T}$, respectively.

3 Weighted Sum Rate

Assume that a certain priority is associated with each user. Our goal is to assign resources such that the sum of data rates R_u weighted by the priorities w_u is maximized. Unfortunately, the rates can not be optimized directly, but only through the transmit covariance matrices $\boldsymbol{Q}$ or, equivalently, the filters $\boldsymbol{T}$. Hence, the optimization problem may be written as

$$\max_{\boldsymbol{Q}\text{ or }\boldsymbol{T}} \sum_{u=1}^{N_\mathrm{U}} w_u R_u \quad \text{subject to (3)} \tag{4}$$

Additionally, the data rates depend on the type of signal processing at the base station. In the following, the differences between linear and nonlinear methods are examined in more detail.

3.1 Dirty Paper Precoding

In the downlink, all rate tuples inside the capacity region can be achieved by nonlinear dirty paper precoding. The users are encoded one after another, and the known interference is presubtracted at the base station. This is exactly the counterpart to successive interference cancelation in the dual uplink.

We may assume without loss of generality that the users are decoded in the order $k = N_\mathrm{U}, \ldots, 1$. Then, the receive covariance matrix for the k-th user on subcarrier s becomes

$$\begin{aligned}\boldsymbol{\Phi}_k[s] &= \boldsymbol{I} + \sum_{u=1}^{k} \boldsymbol{H}_u[s]\,\boldsymbol{Q}_u[s]\,\boldsymbol{H}_u^\mathrm{H}[s] \\ &= \boldsymbol{\Phi}_{k-1}[s] + \boldsymbol{H}_k[s]\,\boldsymbol{Q}_k[s]\,\boldsymbol{H}_k^\mathrm{H}[s]\ ,\end{aligned} \tag{5}$$

where $\boldsymbol{\Phi}_{k-1}[s]$ with $\boldsymbol{\Phi}_0[s] = \boldsymbol{I}$ represents the effective noise including interference from subsequent users. Consequently, the corresponding data rate becomes

$$\begin{aligned}R_k^\mathrm{dp}[s] &= \log\det\big(\boldsymbol{I} + \boldsymbol{\Phi}_{k-1}^{-1}[s]\,\boldsymbol{H}_k[s]\,\boldsymbol{Q}_k[s]\,\boldsymbol{H}_k^\mathrm{H}[s]\big) \\ &= \log\det(\boldsymbol{\Phi}_k[s]) - \log\det(\boldsymbol{\Phi}_{k-1}[s])\ .\end{aligned} \tag{6}$$

The term containing $\boldsymbol{\Phi}_k[s]$ appears in $R_k^\mathrm{dp}[s]$ and $R_{k+1}^\mathrm{dp}[s]$ with different signs. Hence, with the definitions $w_{N_\mathrm{U}+1} = 0$ and $\Delta_k = w_k - w_{k+1}$, the weighted sum rate using dirty paper coding may be written as

$$\mathrm{WSR}^\mathrm{dp} = \frac{1}{N_\mathrm{S}} \sum_{s=1}^{N_\mathrm{S}}\sum_{k=1}^{N_\mathrm{U}} \Delta_k \log\det(\boldsymbol{\Phi}_k[s])\ . \tag{7}$$

For equal priorities, the decoding order is arbitrary. However, in general the users must be sorted according to their weighting factors such that $w_1 \ge \ldots \ge w_{N_\mathrm{U}}$ in order to maximize the weighted sum rate, i.e. the user with the highest priority is decoded last [7]. Then, $\Delta_k \ge 0\ \forall k$, and, since $\log\det(\cdot)$ is a concave function, (7) is concave with respect to $\boldsymbol{Q}$. Furthermore, the feasible set is convex. This assures that there exists a unique optimum.

Besides fulfilling the power constraint (3), the transmit covariance matrices additionally have to be positive semidefinite, which can be achieved by setting all negative eigenvalues of $\boldsymbol{Q}_u[s]$ to zero. To this end, an eigenvalue decomposition is required for all subcarriers of each user, unless the mobile terminals have only a single antenna. However, substituting (2) into (5), the weighted sum rate may alternatively be expressed in terms of the transmit filters $\boldsymbol{T}$, which can be chosen arbitrarily as long as (3) is satisfied. Furthermore, an orthogonal projection onto the power constraint corresponds to a simple scaling. From this point of view, it seems to be more appropriate to perform the optimization over $\boldsymbol{T}$ instead of $\boldsymbol{Q}$. Indeed, the algorithm presented in Section 5 is based on the transmit filters. For later reference, we need the complex gradient of WSR^dp with respect to $\boldsymbol{T}_u[s]$, which equals

$$\boldsymbol{G}_u[s] = \boldsymbol{H}_u^\mathrm{H}[s]\,\boldsymbol{\Psi}_u[s]\,\boldsymbol{H}_u[s]\,\boldsymbol{T}_u[s] \tag{8}$$

up to a scaling factor $2\log(e)/N_\mathrm{S}$. In (8), we introduced the shorthand notation

$$\boldsymbol{\Psi}_u[s] = \sum_{k=u}^{N_\mathrm{U}} \Delta_k\,\boldsymbol{\Phi}_k^{-1}[s] = \boldsymbol{\Psi}_{u+1}[s] + \Delta_u\,\boldsymbol{\Phi}_u^{-1}[s] \tag{9}$$

for the cumulative sum of the weighted inverse receive covariance matrices, that can be computed in a recursive manner, starting with $\boldsymbol{\Psi}_{N_\mathrm{U}}[s] = \Delta_{N_\mathrm{U}}\,\boldsymbol{\Phi}_{N_\mathrm{U}}^{-1}[s]$. Likewise, the matrix inversion lemma yields the update equation

$$\boldsymbol{\Phi}_k^{-1}[s] = \boldsymbol{\Phi}_{k-1}^{-1}[s] - \boldsymbol{B}_k[s]\,(\boldsymbol{I} + \boldsymbol{C}_k[s])^{-1}\,\boldsymbol{B}_k^\mathrm{H}[s] \tag{10}$$

with $\boldsymbol{A}_k[s] = \boldsymbol{H}_k[s]\,\boldsymbol{T}_k[s]$, $\boldsymbol{B}_k[s] = \boldsymbol{\Phi}_{k-1}^{-1}[s]\,\boldsymbol{A}_k[s]$, and $\boldsymbol{C}_k[s] = \boldsymbol{A}_k^\mathrm{H}[s]\,\boldsymbol{B}_k[s]$. For the important case where the maximum number of parallel data streams per user N_L is significantly smaller than the number of antennas at the base station N_B, (10) is much more efficient than directly inverting $\boldsymbol{\Phi}_k[s]$. In particular, for $N_\mathrm{L} = 1$, no explicit matrix inversion is required at all. To summarize, first the inverse matrices $\boldsymbol{\Phi}_k^{-1}[s]$ are calculated in the order $k = 1, \ldots, N_\mathrm{U}$, and afterwards the scaled gradients $\boldsymbol{G}_u[s]$ for $u = N_\mathrm{U}, \ldots, 1$. The intermediate results can be used to obtain the data rates $R_k^\mathrm{dp}[s] = \log\det(\boldsymbol{I} + \boldsymbol{C}_k[s])$ according to (6).

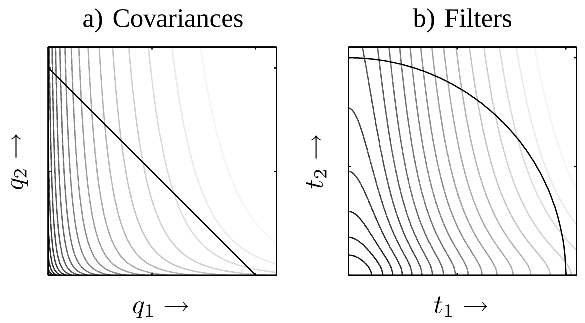

Fig. 1. Weighted sum rate with dirty paper precoding.

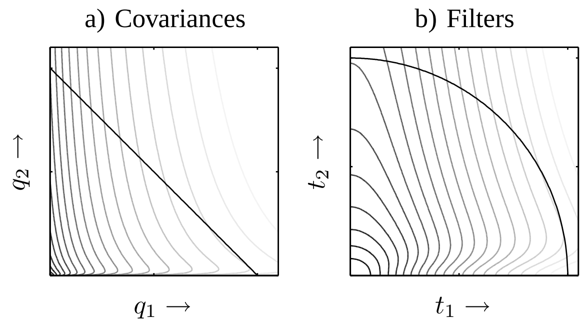

Fig. 2. Weighted sum rate with linear precoding.

Despite the apparent advantages of considering transmit filters instead of covariance matrices when trying to maximize the weighted sum rate, it should be noted that the objective function is no longer concave. A simple example for $N_\mathrm{U} = 2$ users with $N_\mathrm{M} = 1$ antenna each and $N_\mathrm{S} = 1$ subcarrier is depicted in Figure 1. User 1 has a higher priority $w_1 = 5\,w_2$, but user 2 has a better channel quality $\|\boldsymbol{H}_2\|_\mathrm{F}^2 = 10\,\|\boldsymbol{H}_1\|_\mathrm{F}^2$, and the total transmit power is set to $P = 20\,\mathrm{dB}$. When plotted versus the transmit powers $q_1 = |t_1|^2$ and $q_2 = |t_2|^2$, the weighted sum rate is concave, as mentioned before, and the power constraint is a straight line. Obviously, the absolute phase of the transmit filters is irrelevant, so we may assume t_u to be real and nonnegative. Although the function is not concave, it can be observed from Figure 1 that moving along the boundary in an ascent direction will ultimately lead to a stationary point, which must be a global optimum, because the mapping from $\boldsymbol{T}$ to $\boldsymbol{Q}$ is unique (but not vice versa). However, the transmit filters may not be initialized with zeros, as the gradients in (8) vanish in this case and one gets stuck at a suboptimal solution.

3.2 Linear Precoding

With linear precoding at the base station, all active users interfere with each other, so in the dual uplink, all other users contribute to the effective noise covariance matrix of user k, which will be denoted as

$$\begin{aligned}\boldsymbol{\Phi}_{\setminus k}[s] &= \boldsymbol{I} + \sum_{u \neq k} \boldsymbol{H}_u[s]\,\boldsymbol{Q}_u[s]\,\boldsymbol{H}_u^\mathrm{H}[s] \\ &= \boldsymbol{\Phi}_{N_\mathrm{U}}[s] - \boldsymbol{H}_k[s]\,\boldsymbol{Q}_k[s]\,\boldsymbol{H}_k^\mathrm{H}[s]\;, \qquad (11)\end{aligned}$$

with $\boldsymbol{\Phi}_{N_\mathrm{U}}[s]$ from (5). Similar to (10), the inverses can be updated according to

$$\boldsymbol{\Phi}_{\setminus k}^{-1}[s] = \boldsymbol{\Phi}_{N_\mathrm{U}}^{-1}[s] + \boldsymbol{B}_k[s]\,(\boldsymbol{I} - \boldsymbol{C}_k[s])^{-1}\,\boldsymbol{B}_k^\mathrm{H}[s] \qquad (12)$$

with $\boldsymbol{B}_k[s] = \boldsymbol{\Phi}_{N_\mathrm{U}}^{-1}[s]\,\boldsymbol{A}_k[s]$, and $\boldsymbol{A}_k[s]$ as well as $\boldsymbol{C}_k[s]$ like before. The data rate on subcarrier s is now given by

$$\begin{aligned}R_k^\mathrm{lin}[s] &= \log\det(\boldsymbol{\Phi}_{N_\mathrm{U}}[s]) - \log\det\big(\boldsymbol{\Phi}_{\setminus k}[s]\big) \\ &= -\log\det(\boldsymbol{I} - \boldsymbol{C}_k[s])\;. \qquad (13)\end{aligned}$$

In contrast to (6), there are no terms that may cancel each other in the weighted sum rate, so we just write

$$\mathrm{WSR}^\mathrm{lin} = \frac{1}{N_\mathrm{S}} \sum_{s=1}^{N_\mathrm{S}} \sum_{k=1}^{N_\mathrm{U}} w_k\, R_k^\mathrm{lin}[s]\;. \qquad (14)$$

Note that (14) is in general not concave with respect to the transmit covariance matrices, so there may be several local maxima on the boundary of the feasible set.

The gradient is slightly more complicated than before, as $\boldsymbol{T}_u[s]$ appears in all terms except $\boldsymbol{\Phi}_{\setminus u}[s]$. Substituting (13) into (14), we obtain after some simplifications

$$\boldsymbol{G}_u[s] = \boldsymbol{H}_u^\mathrm{H}[s]\,\Big(w_u\,\boldsymbol{\Phi}_{\setminus u}^{-1}[s] + \boldsymbol{\Psi}[s]\Big)\,\boldsymbol{A}_u[s]\;, \qquad (15)$$

where the scaling factor was again neglected and

$$\boldsymbol{\Psi}[s] = \sum_{k=1}^{N_\mathrm{U}} w_k\,\Big(\boldsymbol{\Phi}_{N_\mathrm{U}}^{-1}[s] - \boldsymbol{\Phi}_{\setminus k}^{-1}[s]\Big) \qquad (16)$$

does not depend on the user index, so it can be calculated in advance using (12).

Figure 2 shows a contour plot of the objective function for the same system parameters as in Section 3.1. At first sight, it seems to be quite similar to that in Figure 1 when plotted against the transmit powers. There is a maximum for approximately the same power allocation as before, but a closer look reveals that the global optimum is achieved for $q_2 = 0$, i.e. serving only the user with the higher priority and worse channel. This becomes more obvious if the weighted sum rate is expressed in terms of the transmit filters.

In Figure 3, the achievable data rates using linear and nonlinear precoding are compared. If $w_1 > w_2$, the optimal rate allocation with dirty paper coding lies on the curve for the decoding order $2 \rightarrow 1$, which is switched otherwise. Except for $w_1 = w_2$, where the order is arbitrary and even time-sharing between different orders is allowed, the rates change continuously with varying priorities. This is different for the linear precoder, as the achievable rate region is not convex. Once the relative weight of a user falls below a certain limit, he should be switched off abruptly in order to maximize the weighted sum rate.

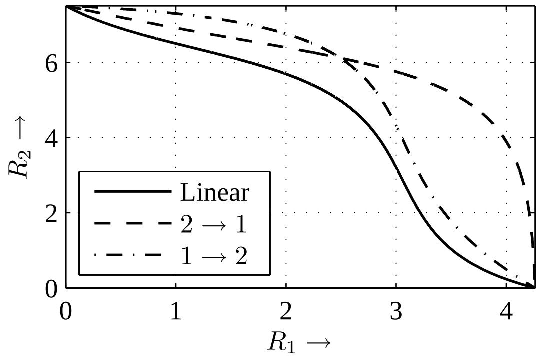

Fig. 3. Achievable rates for linear and dirty paper precoding with both decoding orders in the dual uplink.

Algorithm 1 Projected Conjugate Gradient Algorithm

1: $\boldsymbol{T}_u[s] = \sqrt{\frac{P}{N_\mathrm{U} N_\mathrm{L}}} \boldsymbol{I}_{N_\mathrm{M} \times N_\mathrm{L}} \ \forall u, s$
2: $\boldsymbol{S} = \boldsymbol{0}$, $\rho = 1$, $\alpha = 1$
3: **repeat**
4: Store $\boldsymbol{T}_\mathrm{old} = \boldsymbol{T}$, $\boldsymbol{S}_\mathrm{old} = \boldsymbol{S}$, $\rho_\mathrm{old} = \rho$
5: Calculate gradient $\boldsymbol{G}$ using (8) or (15)
6: Normalize gradient $\bar{\boldsymbol{G}} = \sqrt{\frac{N_\mathrm{S} P}{\|\boldsymbol{G}\|_\mathrm{F}^2}} \boldsymbol{G}$
7: Project gradient $\check{\boldsymbol{G}} = \bar{\boldsymbol{G}} - \frac{\mathrm{tr}(\boldsymbol{T}^H \bar{\boldsymbol{G}})}{\mathrm{tr}(\boldsymbol{T}^H \boldsymbol{T})} \boldsymbol{T}$
8: Calculate Frobenius norm $\rho = \|\check{\boldsymbol{G}}\|_\mathrm{F}^2$
9: Update search direction $\boldsymbol{S} = \check{\boldsymbol{G}} + \frac{\rho}{\rho_\mathrm{old}} \boldsymbol{S}_\mathrm{old}$
10: Step in search direction $\tilde{\boldsymbol{T}} = \boldsymbol{T}_\mathrm{old} + \alpha \boldsymbol{S}$
11: Normalize transmit filters $\boldsymbol{T} = \sqrt{\frac{N_\mathrm{S} P}{\|\tilde{\boldsymbol{T}}\|_\mathrm{F}^2}} \tilde{\boldsymbol{T}}$
12: **if** $\mathrm{WSR} - \mathrm{WSR}_\mathrm{old} < \delta/2$ **then**
13: Decrease step size α and set $\boldsymbol{S} = \check{\boldsymbol{G}}$
14: Go to step 10
15: **end if**
16: **until** desired accuracy reached

4 Existing Algorithms

Most publications related to the solution of (4) focus on nonlinear precoding due to the problems involved with the linear one. For the special case of equal priorities, an iterative waterfilling approach based on the eigenvalue decomposition of effective channel matrices was presented in [8] for individual power constraints per user. If only the total transmit power is fixed in each iteration and all users are updated simultaneously, it may be necessary to average the transmit covariance matrices over the current and previous [9]. The speed of convergence was significantly increased by a simple approximate line search in [10], which is referred to as state of the art in [11]. Instead, the covariance matrices of randomly selected user pairs are optimized in [12], while in [13] users are successively updated for fixed waterfilling level, which is then adjusted using a bisection search until the desired total transmit power is reached.

For arbitrary weighting factors, the basic concepts of [9] and [10] are adopted in [14] for single-antenna mobile terminals, while the dual decomposition principle of [13] is used in [15] for the same scenario in combination with OFDM. Note that the whole algorithm consists of three nested loops, where each has to be repeated until a certain accuracy is reached. Nevertheless, these waterfilling-based procedures are supposed to be significantly faster than the more general method of [16], that performs a line search along the principal eigenvector of the gradient. Recently, a projected conjugate gradient algorithm based on the transmit covariance matrices was proposed in [7], that was originally developed for pure sum rate maximization in [17]. As already mentioned in Section 3.1, eigenvalue decompositions are required for the projection of $\boldsymbol{Q}$ onto the feasible set. An additional drawback will be explained in the following section.

In [18] an alternating optimization of transmit and receive filters is performed based on the uplink-downlink duality. Even though a small number of iterations seems to be sufficient, it should be noted that a $N_\mathrm{U} N_\mathrm{L} \times N_\mathrm{U} N_\mathrm{L}$ matrix must be inverted for the filter conversion each time the transmit direction is reversed, so the complexity grows roughly cubic in the number of users. A simple projected gradient method based on the transmit filters $\boldsymbol{T}$ was used in [6] for the related weighted sum mean square error minimization problem, and lately also for sum rate maximization in [11], where it performed comparable to the approach of [10]. One problem of such gradient based strategies is the proper choice of the step size. In [11], it is decreased as soon as the objective shows no improvement in order to ensure convergence. However, a small step size results in slow speed of convergence. An advantage of the conjugate gradient method described next is the implicit adjustment of this step length.

5 Proposed Algorithm

We propose a projected conjugate gradient algorithm for the optimization of the transmit filters $\boldsymbol{T}$, so in contrast to [7], no eigenvalue decompositions are required. The whole procedure is outlined in Algorithm 1, where the iteration index has been omitted in order to simplify notation and emphasize the fact that most operations can be performed in place. Apart from the search direction, the method is similar to that in [11]. It may also be used for the design of linear precoders, although convergence to the global optimum can not be guaranteed in this case.

First, the gradient is calculated according to the type of precoding. The normalization ensures that the speed of convergence is almost independent of the transmit power, as the Frobenius norm of the gradient tends to be rather small for large P. A key feature of our algorithm is the projection of the gradient such that $\check{\boldsymbol{G}}$ is orthogonal to $\boldsymbol{T}$. In the optimum, the gradient

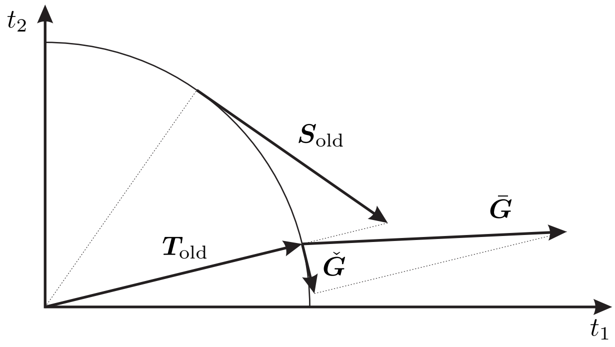

Fig. 4. Illustration of the gradient projection

is in general not zero, but perpendicular to the power constraint, i.e. $\|\check{\boldsymbol{G}}\|_{\text{F}}^2 = 0$ and $\|\boldsymbol{G}\|_{\text{F}}^2 \geq 0$. This is illustrated in Figure 4 for a simple example with two single-antenna terminals. If the projection was omitted as in [7], the old search direction $\boldsymbol{S}_{\text{old}}$ would always significantly contribute to the new one, and the algorithm may even try to step out of the optimum again. This can not happen with our modified Fletcher-Reeves deflection [19] based on the projected gradient $\check{\boldsymbol{G}}$. The updated search direction has two desirable properties: If $\check{\boldsymbol{G}}$ and $\boldsymbol{S}_{\text{old}}$ point into similar directions, they add up constructively and the effective step length is increased. On the other hand, oscillations around the optimum, that are quite common for steepest ascent methods, are prevented to a large extend by using the conjugate gradient.

The transmit filters are updated using $\boldsymbol{S}$ and normalized such that the power constraint is fulfilled. If the weighted sum rate does not show a sufficient improvement, the step size α is reduced by a factor 2 and the projected gradient is used as the new search direction. In general, the increase of the objective function should be related to the linear increment

$$\delta = 2\log(e)/N_{\text{S}} \cdot \text{Re}\left\{\text{tr}\left((\boldsymbol{T} - \boldsymbol{T}_{\text{old}})^{\text{H}}\, \boldsymbol{G}\right)\right\}, \quad (17)$$

in line 12, which is known as the Armijo condition for an approximate line search [7]. However, for dirty paper coding we usually observed even faster convergence by simply setting $\delta = 0$ as in [11].

6 Numerical Results

For the performance evaluation, we consider the scenario depicted in Figure 5. There are $N_{\text{U}} = 50$ users in a $20 \times 10\ \text{m}$ room, e.g. a large office or a class room. Each mobile station has $N_{\text{M}} = 4$ antennas arranged on a square with $10\ \text{cm}$ side length. The base station is located at the ceiling $3\ \text{m}$ above the mobiles and consists of two crossed linear arrays with four antennas each, so $N_{\text{B}} = 8$. The carrier frequency is $5.2\ \text{GHz}$, and the total bandwidth of $40\ \text{MHz}$ is divided into $N_{\text{S}} = 8$ subcarriers. The channel matrices are derived from a simple single-bounce model. To this

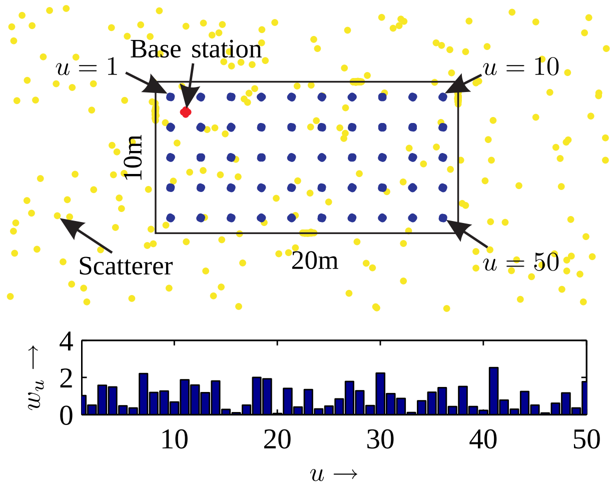

Fig. 5. Illustration of the simulation setup

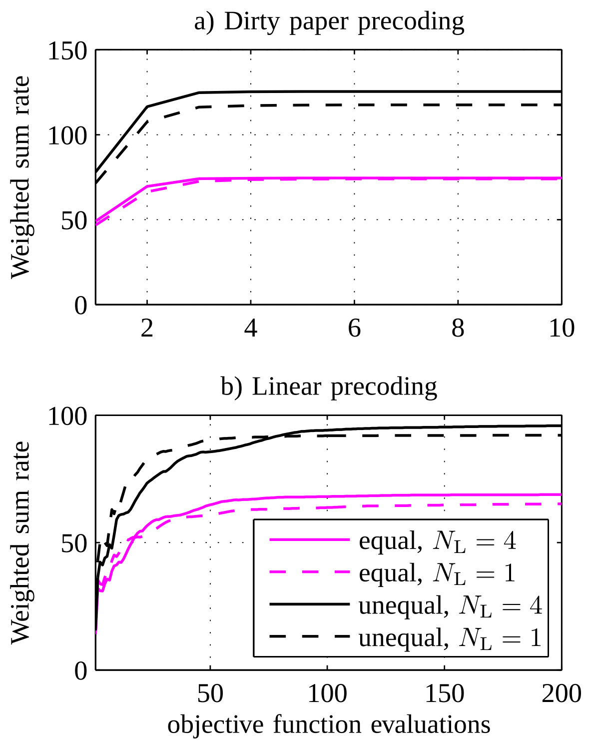

Fig. 6. Convergence of the proposed algorithm for different weight distributions and number of data streams.

end, 200 fixed scatter objects are randomly distributed, and additionally 25 scatterers are placed on each wall around the points of reflection for a certain user. The channel coefficients are normalized such that their average squared magnitude is equal to one, and the total transmit power per subcarrier is set to $P = 20\ \text{dB}$.

In Figure 6, the convergence behavior of our projected conjugate gradient algorithm is shown. The number of objective function evaluations is used as a measure for the complexity, because this requires an update of the inverse receive covariance matrices. In addition to the priorities given in Figure 5, we also

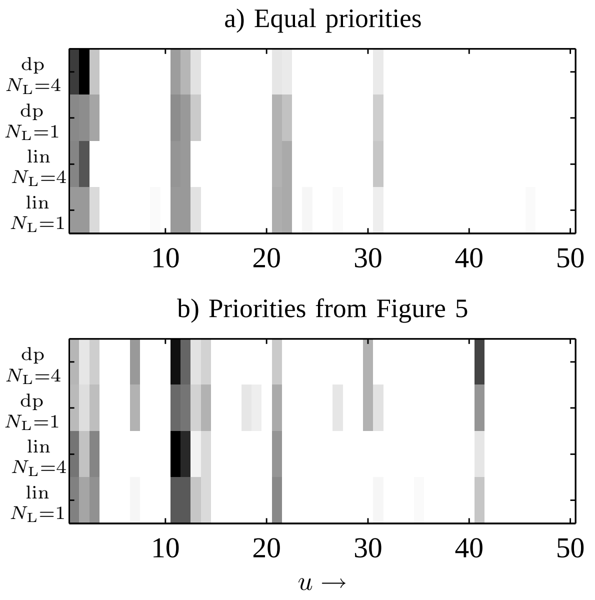

Fig. 7. Rate allocations for different transmit strategies.

simulated the pure sum rate using equal weighting factors, for comparison. Furthermore, the number of data streams is varied. For dirty paper precoding, the method always converges to the global optimum within about 4 iterations. Compared to the covariance-based sum rate maximization algorithm of [10], it takes less than half the run time to reach a certain accuracy for the given system parameters. Restricting the number of data streams per user to $N_{\mathrm{L}} = 1$ leads to a further speed-up without sacrificing much performance, especially for equal priorities. However, when applied to linear precoding, the convergence speed is much lower. The step size α must be decreased several times during the initial iterations to reach an improved objective function value. Although the algorithm does not necessarily find the global optimum, the resulting sum rate is quite close to that of nonlinear precoding. The weighted sum rate, on the other hand, is significantly smaller. This can be intuitively explained by looking at the rate allocations in Figure 7. For the pure sum rate, the active users are just chosen according to their channel qualities. However, the objective function linearly depends on the priorities, while the channels appear only in the logarithm. Thus, it is in general advisable to support high priority users, even if their channels are not particularly good. In our example, the users $u \in \{7, 30\}$ are relatively far away from the base station, but their weighting factors are fairly large. Using dirty paper coding, these users experience only interference from users with lower priorities, so they can still achieve comparatively high data rates. This is not possible with linear precoding anymore, because here all users interfere with each other.

References

[1] P. Viswanath, D.N.C. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[2] H. Boche and M. Wiczanowski, "Stability-Optimal Transmission Policy for Multiple Antenna Multiple Access Channel in the Geometric View," *Elsevier Signal Processing*, vol. 86, no. 8, pp. 1815–1833, Aug. 2006.

[3] C. Zhou and G. Wunder, "Delay and Throughput Optimal Scheduling for OFDM Broadcast Channels," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[4] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[5] 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, no. 10, pp. 2658–2668, Oct. 2003.

[6] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver Design for Multi-User MIMO Systems," in *Proc. ITG/IEEE Workshop on Smart Antennas*, Mar. 2006.

[7] J. Liu and Y.T. Hou, "Maximum Weighted Sum Rate of Multi-Antenna Broadcast Channels," in *submitted to IEEE GLOBECOM*, Washington, DC, USA, Nov. 2007.

[8] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi, "Iterative Water-filling for Gaussian Vector Multiple Access Channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–151, Jan. 2004.

[9] N. Jindal, W. Rhee, S. Vishwanath, S.A. Jafar, and A. Goldsmith, "Sum Power Iterative Water-filling for Multi-Antenna Gaussian Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

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

[11] R. Hunger, D.A. Schmidt, and W. Utschick, "Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

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

[13] W. Yu, "Sum-Capacity Computation for the Gaussian Vector Broadcast Channel via Dual Decomposition," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 754–759, Feb. 2006.

[14] M. Kobayashi and G. Caire, "An Iterative Water-Filling Algorithm for Maximum Weighted Sum-Rate of Gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640–1646, Aug. 2006.

[15] M. Kobayashi and G. Caire, "Iterative Waterfilling for Weighted Rate Sum Maximization In MIMO-OFDM Broadcast Channels," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hawaii, USA, Apr. 2007.

[16] 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, no. 5, pp. 802–811, June 2003.

[17] J. Liu, Y.T. Hou, and H.D. Sherali, "Conjugate Gradient Projection Approach for Multi-Antenna Gaussian Broadcast Channels," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[18] S. Shi, M. Schubert, and H. Boche, "Weighted Sum-Rate Optimization for Multiuser MIMO Systems," in *Proc. Conference on Information Sciences and Systems*, Baltimore, USA, Mar. 2007.

[19] M. Al-Baali, "Descent Property and Global Convergence of the Fletcher-Reeves Method with Inexact Line Search," *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121–124, Jan. 1985.